Visual Security Evaluation for Video Encryption

Lingling Tong[1,2], Feng Dai[1], Yongdong Zhang[1]and Jintao Li[1]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China
[2]Graduate University of Chinese Academy of Sciences, Beijing, 100049, China
{tonglingling,fdai,zhyd,jtli}@ict.ac.cn

ABSTRACT
Video encryption plays an important role in data security guarantee, which is increasingly important with the development of multimedia technology. A great deal of effort has been made in recent years to develop video encryption methods. However, few studies focus on visual security evaluation, which has significant impact in measuring the effectiveness of these methods. In this paper, a new metric for video encryption is proposed, which evaluates visual security based on color and edge features of original and cipher-videos. The metric is easy to be incorporated into video encryption system for visual security based encryption decision. In addition, subjective tests for visual security assessment have been fully carried out. Experiments show that the proposed metric has better correlation with subjective results than others.

Categories and Subject Descriptors
K.6.5 [**Security and Protection**]

General Terms
Algorithms, Design, Experimentation, Measurement, Security

Keywords
Color moments, Gradient direction histogram, Video encryption, Visual security evaluation

1. INTRODUCTION
Video data security management has attracted extensive attention and been an important research area for multimedia applications, such as video-on-demand and video surveillance. As an effective solution to guarantee video data security, various video encryption algorithms have been proposed in recent years [1]-[5]. From the perspective of multimedia content protection, the basic and the most important requirement is security. Regarding video encryption, the security includes two aspects of cryptographic security and visual security [3]. Cryptographic security refers to the security against cryptographic attacks. While visual security refers to intelligibility of the encrypted video, which reflects the distortion degree of cipher-videos.

MM'10, October 25–29, 2010, Firenze, Italy.

Actually, visual security is an important issue for video encryption. First, visual security is a key factor in measuring encryption algorithm performance, as it directly reflects the intelligibility degree of cipher-videos, which is immediately related to the time attackers need for decryption. Attackers will quickly resolve the original video if they can obtain more information from cipher-images [6]. On the other hand, various applications have different visual security requirements. For applications like pay-per-view videos or video-on-demand, "perceptual encryption" does make sense for potential users, which requires that the videos are still partially perceptible after encryption [4]. While for some applications such as video conference, very chaotic cipher-images are required. Based on above analysis, efficient visual security evaluation is desired in two aspects. First, it is needed by image/video encryption systems to measure the effectiveness of encryption methods, and further optimize algorithms or parameter settings. Second, it could be employed in guiding the encryption algorithm selection for applications with different visual security requirements.

Unfortunately, less attention has been paid to visual security evaluation in video encryption studies till now. Almost all existing visual security evaluation methods adopt subjective approach, in which several decoded cipher-images are provided for the observers to assess [3]-[5]. It is too inconvenient and time-consuming for practical usage. On the other hand, some studies choose objective ways for visual security evaluation, although few reasonable metrics have been proposed. The well-known video quality assessment metric Peak Signal-to-Noise Ratio (PSNR) is extended to visual security evaluation in [3] for its simplicity. However, PSNR does not perform as well for visual security evaluation, which is confirmed by our experiment later. Recently, three objective metrics are proposed in [6], which include structural similarity (SSIM) [7], image entropy and spatial correlation. However, it is found that SSIM failed in measuring badly blurred images [8], while cipher-images are just very blurred. In addition, the accuracy of these metrics consistent with subjective results is not convincing for only two sequences' evaluation results are given.

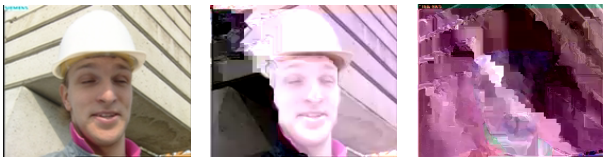
In this paper, a Local Feature Based Visual Security (LFBVS) metric is proposed to quantify cipher-video distortion degree. This metric is based on the fact that video frames can be represented well by their local features. Thus features which are sensitive to Human Visual System (HVS) can be used to evaluate cipher-video visual security. For different distortion levels, cipher-images features are changed to different extents. That is, the less similarity original and cipher-videos have on these features, the higher visual security of cipher-videos is.

The rest of this paper is organized as follows. In section 2, we discuss the requirements and features of visual security evaluation. In section 3, details about the proposed LFBVS

metric are described. Performance of the metric is evaluated in section 4. Finally, we draw some conclusions in section 5.

2. VISUAL SECURITY FOR VIDEO ENCRYPTION

Visual security evaluation is used to measure the unidentifiable degree of encrypted videos. Generally speaking, an encrypted video is regarded as of high visual security if the distortion of cipher-images is too chaotic to be understood [3].



(a) (b) (c)

Figure 1. Foreman encrypted with different visual security: (a) Original image, (b) Low security and (c) High security

Figure 1 shows foreman sequence encrypted with methods in [3]. Figure.1 (b) is got from video encrypted with transform coefficients (TC), while TC, intra prediction mode (IPM) and motion vector difference (MVD) are all encrypted to produce Figure.1(c). Obviously, it can be found that the distortion in Figure.1(c) makes video scene more difficult to be identified.

Discussion with the test subjects who rated the visual security of encrypted videos reveals the characters of cipher-video: 1) cipher-images have severe changes not only in color or luminance, but also in edge and contour information. 2) Compared with color or luminance, HVS is more sensitive to the distortion of edge and contour information from the viewpoint of visual security (Figure.1(c)). 3) For a video image, as long as some part can be identified, the visual security of the entire image is considered to be bad.

While earlier research works have been reported on video quality assessment [8][9], they have addressed the problem of getting difference in details between decoded and reference image. The problem we address in this paper has very different requirements which demands different approaches.

3. LOCAL FEATURE BASED VISUAL SECURITY METRIC

The framework of the proposed visual security system is shown in Figure 2. The calculation of LFBVS mainly involves two steps: First, we extract the color and edge related features both in original and cipher-video, generate a set of feature vectors. Second, these features are combined to local-based value, frame-based value and finally a single visual security assessment value for the entire video.

3.1 Local Feature Extraction

Based on cipher-video characters discussed above, color and edge features are selected to form our metric, both of which are measured using local regions. The images in original and cipher-video are all divided into 16×16 non-overlapping blocks or MacroBlock (MB), which is conducive to integrate with the MB-based video codec. For color features, we calculate the first two color moments [10] of each MB as shown in equation (1).

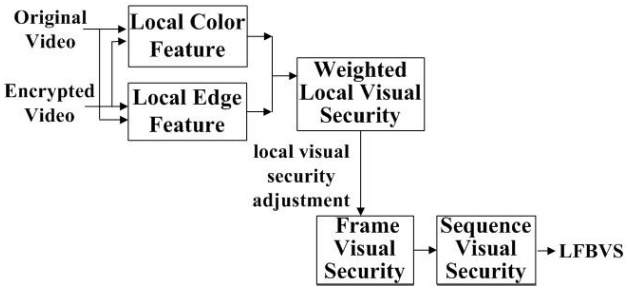


Figure 2. Proposed visual security evaluation framework

$$\mu(m,n)_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij}\ ,\ \sigma(m,n)_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij}-\mu(m,n)_i)\right)^{1/2} \quad (1)$$

Where p_{ij} is the i-th color channel at the j-th pixel in MB (m,n), N is the pixel number in an MB. Finally the color feature vector $CF(m,n)=\{\mu(m,n)_i, \sigma(m,n)_i \mid i=1,2,3\}$ for MB at the position of (m,n) are obtained. As luminance is more important to HVS than other channels, only luminance value is used for feature extraction in this paper.

There are many ways to get edge information, such as local gradients. Different operators correspond to different gradient calculation methods. In this paper, 8-dimension gradient orientation histogram similar to [11] is calculated for each MB in original and cipher image as follows:

1) For each pixel $p_{i,j}$, the operator as in equation (2) is used to obtain vertical and horizontal component of edge magnitude. Where $L(i, j)$ is the luminance value of pixel in the position (i, j) as shown in Figure 3.

$$\begin{aligned} \Delta x(i,j) &= L(i+1,j) - L(i-1,j) \\ \Delta y(i,j) &= L(i,j+1) - L(i,j-1) \end{aligned} \quad (2)$$

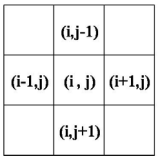


Figure 3. Pixel map in the image

2) Calculated the gradient amplitude $A(i,j)$ and direction $\theta(i, j)$ for each pixel (i,j) in an MB as shown in equation (3). Thus each pixel in the MB has an edge vector including amplitude and direction.

$$\begin{aligned} A(i,j) &= \sqrt{(\Delta x(i,j))^2 + (\Delta y(i,j))^2} \\ \theta(i,j) &= a\tan 2(\Delta y(i,j)/\Delta x(i,j)) \end{aligned} \quad (3)$$

3) The entire direction (0^o~360^o) is divided into 8 discrete directions as shown in Figure 4. Quantify the direction of each pixel as one of the 8 discrete directions.

4) Calculate the sum of all the pixel's amplitudes with the same direction in each MB and form the gradient direction histogram ($GDHis(m,n)$) for MB (m,n) finally.

3.2 Visual Security Evaluation

After obtaining the local feature vectors of the original and cipher-images, the local visual security (LVS) of cipher MB(m,n) is first calculated by equation (4).

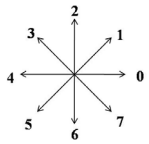


Figure 4. The 8 directions for quantification

$$LVS(m,n) = \alpha CD(m,n) + \beta ED(m,n) \quad \alpha + \beta = 1 \qquad (4)$$

Where α, β are constant between 0 and 1 to adjust the importance of color and edge distortion to visual security results. In our implantation, α and β are set to 0.2 and 0.8, respectively. $CD(m,n)$ is the color distortion of MB (m,n) ,defined as (5).

$$CD(m,n) = \sum_{k=1}^{3}(w_k * \sum_{i=1}^{2} a_i * \frac{d(CF(m,n)_{org}^{ki}, CF(m,n)_{cip}^{ki})}{255}) \qquad (5)$$

Where $CF(m,n)_{org}^{ki}$ and $CF(m,n)_{cip}^{ki}$ are the i-th moment of k-th channel in color feature vector of MB (m,n) in the original and cipher-images, respectively. w_k , a_i are constant between 0 and 1. w_k is the weight for the k-th channel for adjusting importance of different color channels and $\sum_{i=1}^{3} w_i = 1$. Generally, larger value will be assigned to luminance channel as it is more important for HVS. a_i is the weight for the i-th moment and $\sum_{i=1}^{2} a_i = 1$. $d(\cdot)$ is the L1 distance. Similarly, $ED(m,n)$ is the edge distortion metric calculated as follows:

$$ED(m,n) = \frac{\sum_{k=0}^{7} d(GDHis(m,n)_{org}^{k}, GDHis(m,n)_{cip}^{k})}{\sum_{k=0}^{7} \max(GDHis(m,n)_{org}^{k}, GDHis(m,n)_{cip}^{k})} \qquad (6)$$

Where $GDHis(m,n)_{org}^{k}$ and $GDHis(m,n)_{cip}^{k}$ are the k-th component in the original and cipher gradient direction histogram of MB (m,n), respectively. $d(\cdot)$ is the L1 distance and $max(\cdot)$ is the maximum value of the two elements.

In order to provide an overall evaluation about the cipher-image visual security, LVS values are needed to pool into a single visual security metric. An interesting discovery through the subjective visual security assessment is that the visibility of some part in the cipher-image will greatly decrease visual security of the entire frame. As the correlation effect of these visible parts may make people to identify the original scene from the encrypted image. Thus, simply pooling the LVS of each MB by averaging would not result a good overall metric and a mechanism is necessary to put more emphasis on the MBs with lower visual security in the image.

Here we define $\{LVS_i \mid i=1,2,\ldots,n\}$ be the MB visual security collection in a frame. The LVS in an image is first ordered from large to small $\{LVS_{(i)} \mid i=1,2,\ldots,n\}$ such as $LVS_{(1)} \geq LVS_{(2)} \ldots \geq LVS_{(n)}$, where n is the MB number in a frame. Then the frame visual security (FVS) is:

$$FVS = \frac{\sum_{i=1}^{n} \lambda_i LVS_{(i)}}{\sum_{i=1}^{n} \lambda_i} \qquad (7)$$

Where λ_i is the weight to the i-th ranked LVS value, which is obtained as follow:

$$\lambda_i = \exp[(\frac{i}{n} - 1)/\varepsilon] \qquad (8)$$

The weight enhances exponentially as the rank goes down, and the speed of enhancing is controlled by the parameter ε . In our experiment, ε is set to 0.5. Thus the lowest LVS value is given a weight of 1, which means poor local visual security has more effect on the visual security of the whole frame.

Finally, all FVS values are averaged to a single overall video visual security metric LFBVS. It can be observed that higher LFBVS value means higher visual security, and vice versa. Note that LFBVS is also applicable to grayscale videos, in which case only the pixel intensity is used for local feature extraction.

4. EXPERIMENTS

4.1 Encrypted Video Materials

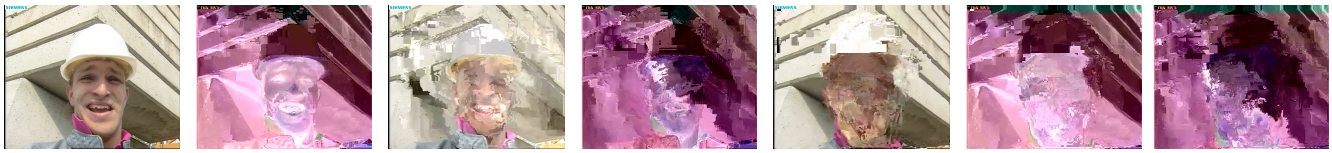
To obtain encrypted videos for visual security assessment, the video encryption scheme proposed in [3] is implemented on JM15.1 reference software. Specifically, the IPM, MVD and TC are selected for encryption. For a detailed description of the encrypted methods, we would like to refer the reader to [3]. Ten standard video sequences with different combinations of motion, texture and object are used, which include Foreman, Mobile, Carphone, Paris, News, Flowers, Bus, Mother&Daughter, Hall and Stefan. The sequences are each 10 seconds in duration and in CIF format. In order to obtain encrypted video with different visual security degrees, the IPM, MVD and TC encryption are implemented separately, with their combinations simultaneously. Thus a total of 70 distorted videos are used in our experiments.

4.2 Subjective Assessment

In this section, we introduce the subjective assessment process to obtain visual security mean opinion score (VSMOS) values of every encrypted video. The similar stimulus continuous quality evaluation (SSCQE) method in ITU-BT.500 [12] is used for subjective visual security assessment. 15 non-expert college students have evaluated the visual security of cipher-videos following the above guideline. The distortion is divided into five degree of "Excellent", "Good", "Fair", "Poor" and "Bad". Different from video quality assessment, "Good" means the video is totally distorted and no information can be got from cipher-video, while "Bad" means there are enough information to make out the whole scene, that is the visual security of this cipher-video is low. The 95% confidence intervals for subjective scores are around 0.045 for the VSMOS on a 0 to1 scale. The VSMOS for a sequence is calculated as the average of all scores obtained for the sequence.

4.3 Performance Analysis

Performance of the proposed visual security metric depends on how well it correlates with the subjective results. Inspired by the performance evaluation methods for video quality assessment [13], we assess the proposed metric performance through three quantitative metrics. The first metric is the Pearson's correlation coefficient (PCC) which relates to prediction accuracy of the proposed metric with respect to subjective results. The second one is spearman rank order correlation coefficient (SROCC). It reflects the degree to which the metric's predictions agree with relative magnitudes of subjective quality ratings. Finally, the outlier ratio (OR) is used to measure the prediction consistency. We compare LFBVS to PSNR and SSIM based visual security assessment metric in [6]. For SSIM, the image is partitioned into verlapping



(a) (b) (c) (d) (e) (f) (g)

Figure 5. Images of Foreman encrypted with different methods: (a) Original, (b) TC encryption, PSNR= 7.7488, SSIM=0.0303, FVS=0.3871, (c)IPM encryption, PSNR=14.6884,SSIM=0.5216,FVS=0.2534, (d)TC&IPM encryption, PSNR=7.8123,SSIM=0.102, FVS=0.4564, (e) MVD encryption, PSNR= 14.9693, SSIM=0.3212, FVS=0.2267, (f) TC&MVD encryption, PSNR=8.1902, SSIM=0.234, FVS=0.4323, (g) TC&MVD&IPM encryption, PSNR=7.458, SSIM=0.181, FVS=0.5569

11×11 blocks. And the constants C_2 and C_3 are the same value as that used in reference [7]. For LFBVS, w_1 in equation (5) is set to 1 indicating only luminance value is used for calculation. a_1 and a_2 are both set as 0.5. Some cipher-images of Foreman sequence encrypted by different methods are shown in Figure 5. The PSNR, SSIM and FVS are all calculated for these cipher-images. It can be observed that the images with nearly identical PSNR values may have obviously different visual security. Comfortingly, the proposed visual security metric seems to correlate well with subjective evaluation. As all metrics are calculated based on luminance information in our experiment, similar visual security assessment results can be obtained for grayscale images.

Table 1 shows the PCC, SROCC and OR between objective metrics and VSMOS when all video sequences are included. The encouraging results also demonstrate the superior performance of the proposed metric.

The proposed metric is based on cipher-video inherent features, which is independent of video encryption algorithm. Thus LFBVS can be applied to a wider spectrum of encryption methods, and we believe it will be of great significance to future video encryption research and application.

Table 1. Performance comparison of different metrics

Metric	PCC	SROCC	OR
1/PSNR	0.752	0.739	0.169
1-SSIM	0.673	0.568	0.272
LFBVS	0.868	0.835	0.066

5. CONCLUSION

In this paper, we propose a local feature based visual security metric for video encryption. The color and edge related features are extracted for evaluation, which follows the fact that HVS is likely to construct the video scene based on these two types of important information. Thus poor similarity of these features between original and cipher-videos signifies high visual security. As we select features and parameters that are suited to describe cipher-videos distortion, prediction accuracy of the proposed metric versus subjective assessment scores is better than other objective assessment metrics. Experiments through precise subjective tests show the effectiveness of the proposed metric.

6. ACKNOWLEDGMENTS

This work was supported by the National Nature Science Foundation of China (60802028), the National Basic Research Program of China (973 Program, 2007CB311100), the Beijing New Star Project on Science & Technology (2007B071) and the Co-building Program of Beijing Municipal Education Commission.

7. REFERENCES

[1] W. Zeng and S. Lei, "Efficient frequency domain video scrambling for content access control", in. *Proc. of ACM Multimedia*, pp.285-294, 1999.

[2] H.J. Lee, "Low complexity controllable scrambler /descrambler for H.264/AVC in compressed domain", in. *Proc. of ACM Multimedia*, Santa Barbara, CA, Oct. 2006.

[3] S. Lian, Z. Liu, and Z. Ren, "Secure advanced video coding based on selective encryption algorithms", *IEEE Trans. Consumer Electronics*, vol. 52, no. 2, pp. 621-629, 2006

[4] S. Li, G. Chen, A. Cheung, B. Bhargava and K.T. Lo,"On the Design of Perceptual MPEG-Video Encryption Algorithms", *IEEE Trans. CSVT*, vol. 17, no. 2, 2007.

[5] Sohn, H., De Neve, W., and Ro, Y., "Region-of-interest scrambling for scalable surveillance video using JPEG XR", in. *Proc. of ACM Multimedia*, Beijing, China, 2009.

[6] Y Yao, Z Xu and W Li, "Visual security assessment for video encryption", *Communications and Networking* in China, 2008.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing,* vol. 13, no. 4, pp. 600-612, Apr. 2004.

[8] GH Chen, CL Yang and SL Xie, "Gradient-Based Structural Similarity for Image Quality Assessment", in *Proc. IEEE Int. Conf. Image Processing*, 2006.

[9] A Bhat, I Richardson, S Kannangara, "A new perceptual quality metric for compressed video", in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, 2009.

[10] Stricker M. and Orengo M. "Similarity of color images", In Proc. SPIE Storage and Retrieval for Image and Video Databases, 1995, Vol. 2420, pp.381-392.

[11] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *IJCV*, 60(2):91–110, 2004.

[12] ITU-R BT.500 Methodology for the Subjective Assessment of the Quality for Television Pictures, ITU-R Std., Rev. 11, June 2002.

[13] Video Quality Experts Group, "Final Report from the VQEG on the validation of Objective Models of Video Quality Assessment, Pase II", *www.vpeg.org,* August 2003.