COMPRESSION AND ENCRYPTION OF VIDEO DATA FOR EFFICIENT STORAGE/TRANSMISSION

MANOJ K. MISHRA, S. MUKHOPADHYAY AND G.P. BISWAS
Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, India
Email: {manojmishra.ism, msusant2001, gpbiswas}@gmail.com

ABSTRACT: This paper proposes a compression cum encryption technique of video data for secure storage with lesser memory requirement, and/or transmission over public network with lesser traffic than the requirements with clear videos. Initially, a group of pictures (GOP) is used to form an accordion matrix and then DCT transformation is applied to each 8×8 block for generating *DC* and *AC* values. The *DC*s are only then encrypted using RC4 byte stream cipher and send the same to the receiver. The receiver similarly uses RC4 to extract the *DC* values on decryption and the output is used for the approximation of the corresponding *AC* values. A smoothing operation is then applied for reconstructing the accordion matrix, which is finally used to reconstruct input GOP. The proposed scheme has been simulated on different video images, and it has been found that expected performance is achieved.

KEYWORDS: Video Encryption/Decryption, DCT, RC4, DH Protocol, PSNR, SSIM.

INTRODUCTION

Video data has become an indispensable and integrated part of modern communication system as it can convey any message or information with wider, deeper and quicker impact. Almost all the popular fields like news, infotainment, entertainment, education, sports, e-classroom, medical instrumentation, video conferencing etc involve significant amount of video data processing and handling. For high quality video data the volume of the data is enormous and is a big burden for the channel transmitting the video signal for on-line/off line cast. In most cases the data is to be made secure from the group of unauthorized users, in some cases the video data could be intended to be confidential. All these requirements are full filled by applying compression and encryption techniques on the video data at the transmitting end and decompression and decryption at the receiving end. The primary objective of video compression in most video applications is to reduce the amount of video data for storing or transmission purposes without affecting the visual quality. The desired video performances depend on applications requirements, in terms of quality, disks capacity and required bandwidth.

Basically, motion estimation based encoders are the most widely used in video compression. Such encoder exploits inter frame correlation to provide more efficient compression. However, motion estimation process is computationally intensive; its real time implementation is difficult and costly [1]-[5].

We have used 3D to 2D transformation [6], which turns the spatial temporal correlation of the video into high spatial correlation, as a result we have obtained one picture instead of each group of pictures eventually with high spatial correlation. Thus, the de-correlation of the resulting pictures by the DCT makes efficient energy compaction, and therefore produces a high video compression ratio. As, Joint Photographic Experts Group (JPEG) has selected DCT for its baseline

coding technique in which all DCT coefficients [4] are utilized in compressing the image/picture. In another example, RVEA algorithm [5] encrypts important DCT coefficients *DC* together with first few *AC*s, and shuffles significant bits between coefficients to achieve higher compression as well as security strength. Similarly in Zig-Zag algorithm [8] supplies a random permutation list to replace the original Zig-Zag ordered DCT coefficients of a block to a 1 × 64 vector to permute the compressed data sequence to make the coefficient unwise. In other example, Zeng and Lie [9] extended this permutation to a segment of macro-blocks in which each segment DCT coefficients of the same frequency are randomly shuffled within band. Generally significant energy of the image is represented by *DC* coefficient of image block. The estimation methods for lossy image/video compression are very few [7, 10, 11]. Gonzales et al. [7] described a technique which estimates a few low frequency coefficients precisely. As, *AC* coefficients of the DCT of image block represent less energy. Hence, compression on the video data (accordion matrix) can be achieved by not transmitting these coefficients. However, they can be estimated from *DC* coefficients of neighbouring DCT blocks at receiver end as a resultant it gives an error know as blocking artifacts, which removed using the existing smoothing operation [12–14].

In this paper, we have proposed DCT-based compression of the accordion matrix developed using group of pictures (GOP) and a light-weight RC4 based encryption of low bit videos. For the key management, the cookie-based Diffie-Hellman key exchange protocol is used for generating authenticated shared key to be utilized in RC4. The rest of this paper is organized as follows. A preliminary part containing accordion formation, RC4, Diffie-Hellman technique etc is given in section II. The proposed technique for the compression and encryption of video data are provided in section III. In section IV, some simulation results of the proposed scheme are given. Finally, the concluding remarks are presented in section V.

PRELIMINARIES

Accordion Representation

The video data is essentially a temporal sequence of frames with the presence of both intra frame and inter frame redundancies. The accordion matrix [6] is basically a 2D representation of a group of frames where the column of the accordion matrix comes from the column of the frames in a periodic fashion. If *F* (x, y, t) represents video frames with spatial coordinate *x, y* and temporal coordinate *t*, the accordion matrix *A* is constructed from *F* (*x, y, t*) as illustrated in Fig.1.

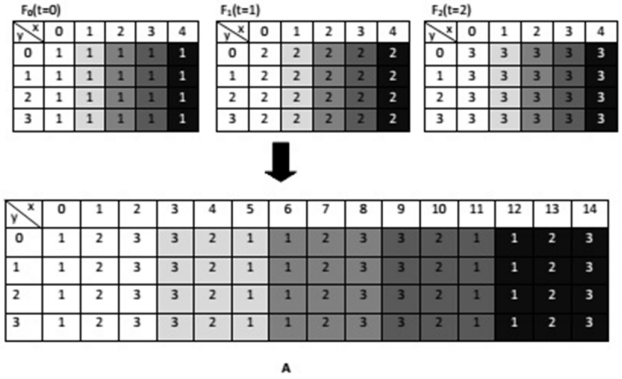

Fig.1 Accordion Matrix Formation

Encryption And Decryption Using Rc4

The mechanism of converting plaintext to ciphertext is called encryption or enciphering and restoring plaintext from ciphertext is known as decryption or deciphering, which result in making the confidentiality of the important messages. Based on the type cryptosystems, both the encryption and decryption algorithms require secret key(s) for securing plaintext against unauthorized access. In this work, the RC4 stream cipher for encrypting video stream is used for confidentiality of the video data and a brief description of the same is provided now.

The RC4 is a byte stream cipher designed by R. Rivest [15, 21] in 1984 for encrypting RSA data, and it has subsequently received well acceptance from cryptography community and applied in many real life applications including WEP (Wired Equivalent Privacy), GSM security[22] etc. It consists of three phases as briefly provided below:

Initialisation: It initialises a state vector *S* with 256 bytes and assigns a key vector *K* with 256 byte-key. Note that if the key-size is less than 256 bytes, the given secret key is repeated until the required size is made. For initialisation, the following algorithm is followed:

for $i = 0$ to 255
{ $S[i] = i$; $K[i]$ = Key[i % key-length]}, where Key is key-array having size of key-length

Permutation: The *S* vector is permutated or swapped using *K* vector for complete modification of the S vector. The permuting process is given below:

$j = 0$
for $i = 0$ to 255
{ ($j = j + S[i] + K[i]$) % 256; swap ($S[i]$, $S[j]$}

Key stream generation: This phase generates a new key byte for encrypting each plaintext byte and it is repeated until entire given input is encrypted. The process for the key generation is provided below:

$i = (i + 1)$ % 256; // at starting, $i = j = 0$
$j = (j + S[i]$ % 256;
swap ($S[i]$, $S[j]$);
$k = (S[S[i] + S[j])$ % 256; // k is the key-byte

Encryption and Decryption: In RC4, the logical XOR operation between plaintext and key bytes is simply used for encryption while during decryption, the same XOR operation is used between ciphertext and key bytes. The encryption and decryption processes are given as follows:

Encryption: $C = P \oplus k$
Decryption: $P = C \oplus k$, where P and C are plaintext and ciphertext bytes, respectively.

Since a common secret key between sender and receiver is required in RC4 for proper encryption/decryption of the plaintexts, a key management technique must be used a priori for successful implementation of the RC4. For this, the Diffie-Hellman (DH) secret key exchange

technique with cookies is to be used for authenticated secret key establishment between the sender and receiver of the video images. Note that a cookie is generated cryptographic hashing of (i) unique identity like IP-addresses, (ii) a secret number that is known to the cookie generator and (iii) a time-stamp. For the sake of clarity, the cookie-based DH protocol is provided below, where a large prime number P and a generator g corresponding the multiplicative group formed by P are assumed to be known to the participants publicly.

I (Initiator):

1. Send Cookie-I to *R* (Responder)
2. On receiving *Cookie-I* and *Cookie-R* from responder, generates $X = g^x \ mod \ P$ and send the following to *R*
   Cookie-*I*, Cookie-*R*, X
3. On receiving Cookie-*I*, Cookie-*R*, $Y = g^y \bmod P$, Initiator generates the following authenticated shared secure key *K* as
   $$K = g^{xy} \bmod P$$

R (Responder):

1. Send Cookie-*I* and Cookie-*R* to *I* (Initiator)
2. On receiving Cookie-*I*, generates $Y = g^y \bmod P$ and send the following to *I*
   Cookie-I, Cookie-R, $Y = g^y \bmod P$
3. On receiving Cookie-*I*, Cookie-*R*, $X = g^x \bmod P$, Responder generates the following authenticated shared secure key *K* as
   $$K = g^{xy} \bmod P$$

AC Coefficient Prediction

Gonzales et al.[7] has proposed a technique which predicts a few low frequency *AC* coefficients for discrete cosine transform. The *AC* prediction method uses de-quantised *DC* values of a 3 × 3 neighborhood of 8 × 8 blocks to estimate the *AC* coefficients for the center block. Fig.2 shows the neighborhood of DCT blocks used in predicting the fifth block. At least five *AC* coefficients are required to reduce blocking artifacts. The transform coefficients in a 3 × 3 array of blocks in which each block containing an 8 × 8 transform coefficients, is modeled by a two-dimensional second degree polynomial [11, 15] of the form:

DC_1	DC_2	DC_3
DC_4	DC_5	DC_6
DC_7	DC_8	DC_9

Fig. 2 Set of DC Values

$$P(x, y) = A_1(x^2y^2) + A_2(x^2y) + A_3(xy^2) + A_4(x^2) + A_5(xy) + A_6(y^2) + A_7(x) + A_8(y) + A_9 \quad (1)$$

A quadratic surface, given by eq. (1) is used to model the shape of a local surface [10, 23]. Coefficients (A_1 to A_9) can be determined by finding the best fit to the surface determined by nine *DC* values in an array of 3 × 3 blocks. The nine DC coefficients A_1 through A_9 are uniquely determined by imposing the constraint that the mean of $P(x, y)$ over each of the nine blocks must

yield the correct *DC*-values. Low frequency *AC* coefficients required to reproduce the quadratic surface can be calculated by a set of equations expressed in terms of nine *DC* values. The prediction formulae for the first five un-quantized *AC* coefficients are shown below:

$$AC(1, 2) = 1.13885 * (DC4 - DC6) \div 8$$

$$AC(2, 1) = 1.13885 * (DC2 - DC8) \div 8$$

$$AC(1, 3) = 0.27881 * (DC4 + DC6 - 2 * DC5) \div 8$$

$$AC(3, 1) = 0.27881 * (DC2 + DC8 - 2 * DC5) \div 8$$

$$AC(2, 2) = 0.16213 * (DC1 + DC9 - DC3 - DC7) \div 8$$

PROPOSED VIDEO COMPRESSION-ENCRYPTION TECHNIQUE

Accordion matrix is often compressed using DCT on each 8 × 8 block. These DCT transform coefficients can be classified into two groups, *DC* and *AC*. The *DC* coefficient is the mean value of a block. All other coefficients describe the variation around this *DC* value and these are referred to as *AC* coefficients. Most of the signal energy of the accordion matrix block is compacted in the *DC* component of the DCT, and the remaining energy is distributed diminishingly in the *AC* components in zigzag scan order. The schematic diagram of the proposed method is shown in Fig.3.

The proposed video compression-encryption technique contains the following steps:

Sender-side:

1. Decomposition of the input video in groups of pictures (GOP)
2. Accordion representation (ACC) of the GOP
3. Decomposition into 8 × 8 blocks and apply of DCT
4. Collection of DC values and encryption using RC4
5. Transmission of the encrypted (ciphertext) video to the receiver

Receiver-side:

1. Decryption and extraction of DC values using RC4
2. Prediction of *AC* values using *DC* values
3. Apply of IDCT (Inverse DCT) and IACC (Inverse ACC)
4. Reconstruction of GOP using reconstructed accordion matrix
5. Apply the smoothing operation over the reconstructed GOP

Encryption of *DC*'s

This section of the paper describes the proposed encryption scheme on the *DC* of the discrete cosine transform. Though we can use any of the algorithms, like DES (Feistel Structure), RC4, RC5 for the encryption. In this work, we have used RC4 [21] to encrypt the *DC* coefficients. As we know that *DC* values in itself are sufficient to leak the video information. So it is unwise to ignore the security of its. Since the *DC* component carries most of the energy, it is the most

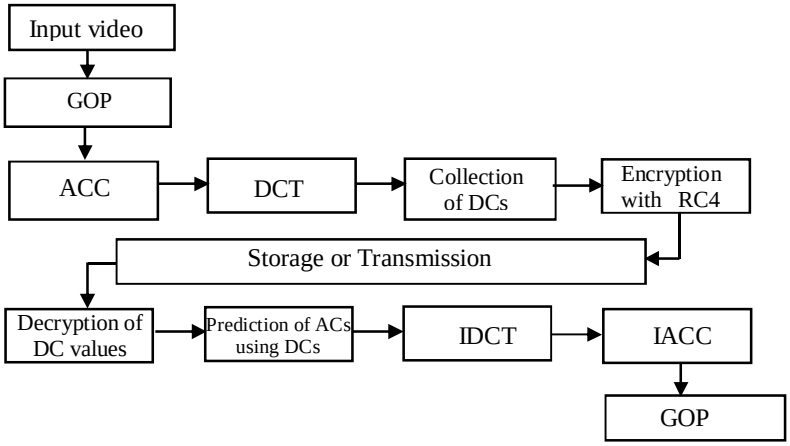

Fig. 3 Schematic Diagram of the Proposed Method

significant value of the block. For the encryption of *DC* coefficients, we have collected the *DC* coefficient values and encrypt them using RC4 algorithm. For obtaining the *DC* coefficients, the recipient can directly apply general RC4 decryption.

PERFORMANCE ANALYSIS

We start by studying the performances of the proposed method with different *N* values, it's pointed out that *N* presents the number of frames of the video cube that forms the "*A*" frame. The best compression rate is obtained with *N* =8. The compression of accordion representation process starts with breaking up it into 8 × 8 block using the DCT.

Simulation Results (for performance parameters)

Our proposed method consists of simple accordion compression using DCT in which encrypted *DC*s values are transmitted over the network and on the receiver side value of *AC*s are predicted using *DC*s values. The *DC* values would be obtained by the decryption process. If you compare our encryption process to DES [15] -[20], it is a lightweight encryption algorithm as it only considers the *DC* values and left the all corresponding *AC*s.

The following experiment results in Table.1 have proved our assumption. The algorithms are implemented and tested using matlab. Basically, natural image/frame signals are highly structured: Their pixels exhibit strong dependencies, especially when they are spatially proximate, and these dependencies carry important information about the structure of the objects in the visual scene. The visual quality gives a subjective evaluation of the proposed method. For objective evaluation we furnish below a few quantitative measures [24] namely: mean square error (MSE), peak signal to noise ratio (PSNR), structural similarity index measure (SSIM).

The simplest and most widely used full-referenced quality metric is the mean squared error (MSE), computed as:

$$MSE = \frac{1}{mn}\sum_{y=1}^{m}\sum_{x=1}^{n}[A(x,y) - A^{-1}(x,y)]^2 \qquad (2)$$

Table1. Performance of Proposed Algorithm

Video Sequence	Compression Ratio(*CR*)	Frame Number	MSE	PSNR	SSIM
Miss America	64:1	017	10.0390	23.6557	0.9781
		018	10.1527	23.7841	0.9676
		019	10.3412	23.5218	0.9684
		020	10.3512	23.0532	0.9674
		021	10.1462	23.5218	0.9617
		022	10.1310	23.0532	0.9606
		023	10.3010	23.5218	0.9682
		024	10.1120	23.0532	0.9678
Foreman	64:1	01	15.8520	21.4630	0.9421
		02	15.8986	21.0487	0.9323
		03	15.9383	21.1314	0.9317
		04	15.9783	21.0536	0.9308
		05	15.9217	21.1314	0.9317
		06	15.9321	21.2586	0.9312
		07	15.9781	21.1314	0.9310
		08	15.9756	21.2014	0.9308

and

$$PSNR = 20 \times \log_2 \left(\frac{255}{\sqrt{MSE}}\right) \tag{3}$$

Here, $A(x, y)$ is the original accordion frame, $A^{-1}(x, y)$ is the reconstructed accordion frame from the compressed data and m, n are the dimensions of the accordion matrix. However, MSE and PSNR do not have a strong correlation with human visual system (HVS). Rather the structural similarity (SSIM) approach provides an alternative and complementary way to tackle the problem of image quality assessment. Mathematically, (SSIM) can be defined as:

$$SSIM\ (A(x, y), A^{-1}(x, y)) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4}$$

Here, μ_x, σ_x^2 and σ_{xy} be the mean of $A(x, y)$, the variance of $A(x, y)$, and the covariance of $A(x, y)$ and $A^{-1}(x, y)$, respectively. Approximately, μ_x and σ_x can be viewed as estimates of the luminance and contrast of $A(x, y)$, and σ_{xy} measures the tendency of $A(x, y)$ and $A^{-1}(x, y)$ to vary together. We have applied the SSIM indexing algorithm for frame quality assessment using a sliding window approach. We have estimated the compression ratio (CR), MSE, PSNR and SSIM of the reconstructed images, which are shown in Table.1. So, we can observe that our method is comparable with Tarek et al. [6] and in some example of video sequence it shows the better result.

Simulation Results (for two videos)

A number of experiments have been conducted in order to study the performances of our method; we had chosen two different kinds of benchmark gray level video sequences having a dimension of 144 × 176 corresponding to each of frame. In the following, we have summarized the experimental

results with some analysis and comments. Fig.4 shows one of the source video frames, which we have used for our proposed work. It is a clip from a Foreman sequence which contains some objects with fast movement, so it is a challenging sample for the experiment. A movie video clip from Miss America movie sequence is also used to test the proposed algorithm. In our experiment, basically we are interested only for the compression and encryption of the accordion image sequence. An original frame and test results are shown in Fig. 4(a, b) and Fig. 5(a, b) corresponding to Foreman and Miss America sequence. After encryption, the accordion matrix became a highly disguised and it is next to impossible to predict any information from that as the of RC4 is highly secured.

Fig. 4(a) Original Foreman Sequences

Fig. 4(b) Reconstructed Foreman Sequences

Fig. 5(a) Original Miss America Sequences

Fig. 5(b) Reconstructed Miss America Sequences

CONCLUSION

The video data have high temporal redundancies between frames and the same has not been exploited directly by recent video compression techniques. In the proposed work, we have suggested a new video compression and encryption method, which exploits objectively the temporal redundancy through accordion matrix up to a very high extent as well as encrypts *DC* values of the DCT transformation using RC4 stream ciphering technique. As a result, not only the compression efficiency is enhanced, but also the encryption computation is improved. Since key management in cryptography is a challenging task, the further improvement of the proposed video

data compression-encryption technique would be done by incorporating key management method. Finally, the performance analysis presented through simulation on different video data justifies our claim and supports for the practical applications of the proposed video compression-encryption technique.

REFERENCES

I. Agi and L. Gong, "An empirical study of secure MPEG video transmission," in Proc. ISOC Symposium on Network and Distributed Systems Security (SNDSS96), pp. 137-144, 1996.

G. Liu, T. Ikenaga, S. Goto, and T. Baba, "A selective video encryption scheme for MPEG compression standard," IEICE Trans. Fundamentals, vol. E89-A, no. 1, pp. 194-202, 2006.

B. Bhargava, C. Shi, and S.-Y. Wang, "MPEG video encryption algorithms," Multimedia Tools and Applications, vol. 24, no. 1, pp. 57-79, 2004.

M. S. Kankanhalli and T. T. Guan, "Compressed-domain scrambler/ descrambler for digital video," IEEE Trans. Consumer Eletronics, vol. 48, no. 2, pp. 356-365, 2002.

C. Shi and B. Bhargava, "Light-weight MPEG video encryption algorithm," in Proc. Int. Conference on Multimedia (Multimedia98, Shaping the Future), pp. 5561,

Tarek. Ouni, Walid. Ayedi, and Mohamed. Abid., "New Low Complexity DCT based Video Compression Method," IEEE Tele communications, ICT, pp. 202-207, Marrakech, Morocco, 2009.

C.A.Gonzales, L.Allman, T.Mccarthy and P.Wendt, "DCT coding for motion video storage using adaptive arithmetic coding," Signal Processing: Image Communication. Elsevier, pp. 145-154, 1990.

L. Tang, "Methods for encrypting and decrypting MPEG video data efficiently," in Proc. of ACM Multimedia, 1996.

Wenjun Zeng and Shawmin Lei, "Efficient frequency domain selective scrambling of digital video," in Proc. of the IEEE Transactions on Multimedia, pp. 118-129, 2002.

Gopal Lakhani, "Image fitting using arctan for JPEG AC coefficient prediction," Accepted paper, IEEE, 2008.

K. Veeraswamy, S. Srinivas Kumar, "Adaptive AC-Coefficient Prediction for Image Compression and Blind Watermarking ," Journal of Multimedia, Vol. 3, No. 1, MAY 2008.

J, Singh, S, Singh, D, Singh and Moin Uddin, "A signal adaptive filter for blocking effect reduction of JPEG compressed images," Int. Journal of Electronics and Communication, vol.65, pp. 827-839, 2011.

S, Liu, A, C, Bovik, "Efficient DCT-domain blind measurement and reduction blocking artifact," IEEE Transactions on Circuits and Systems for Video Technology, vol.12, pp. 1139-1149, 2002.

R. C. Gonzalez and R. E. Woods, "Digital Image Processing," Prentice-Hall, Upper Saddle River, NJ, second ed., 2002.

William Stallings, "Cryptography and network security: Principles and practice", Prentice Hall, Upper Saddle River, New Jersey, 2003.

W. Zeng, H. Yu, and C. Y. Lin, Eds., "Multimedia Security Technologies for Digital Rights Management" Orlando,Florida: Academic Press,Inc.,2006.

A. Uhl and A. Pommer, "Image and Video Encryption: From Digital Rights Management to Secured Personal Communication," Springer, Advances in Information Security, vol. 15, 2005.

B. Furht, E. Muharemagic,and D. Socek, Eds., "Multimedia Encryption and Watermarking," New York: Springer, 2005.

B. Furht, D. Socek, and A. M. Eskicioglu, “Fundamentals of multimedia encryption techniques,” in Multimedia Security Handbook, B. Furht and D. Kirovski, Eds. CRC Press, LLC, ch. 3, pp. 93 - 131, 2004.

BorkoFurht and DarkoKirovski, “Multimedia Encryption Techniques,” Multimedia Security Handbook, 2004.

I. Mantin, “Analysis of the Stream Cipher RC4”. Master thesis, The Weizmann Institute of Science, 2001.

C-C Lo, and Y-J Chen, “Secure Communication Mechanisms for GSM Networks,” IEEE Transactions on Consumer Electronics, Vol.45, No.4, pp.1074-1080, Nov. 1999.

“Information technology- digital compression and coding of continuous tone still images requirements and guide lines,” CCITT, T.81.

Wang, Z., Bovik, A. C., Sheikh, H. et al.: ”Image quality assessment: From error measurement to structural similarity,” IEEE Trans. Image Processing, vol. 13, Jan. 2004.