A Fast Selective Video Encryption Using Alternate Frequency Transform

Ashutosh Kharb
Department of ECE , USIT, *New Delhi, India.110006*
Seema
*Department of CSE,BMIET, Sonipat, Haryana, India.131001
Ravindra Purwar
GGSIPU, Sonipat, New Delhi, India.110006

Abstract— With commercialization of multimedia data over public networks security of multimedia data is a challenging issue. Further multimedia data is generally very large therefore it requires efficient compression to save transmission cost. In this manuscript a modified 4 point butterfly method is proposed to compute DCT for encoding of frames in video data. It has been experimentally compared with existing technique based on parameters like PSNR, compression ratio, execution time of each frame, time taken for evaluating DCT method. Also it has been shown theoretically that the proposed technique take lesser time than the existing method.

Keywords: DCT, motion estimation, selective encryption, spatial compression, video encoding.

I. INTRODUCTION

With commercialization of multimedia data over public networks security of multimedia data is a challenging issue. Further multimedia data is generally very large therefore it requires efficient compression to save transmission cost. In this manuscript a modified 4 point butterfly method is proposed to compute DCT for encoding of frames in video data. It has been experimentally compared with existing technique based on parameters like PSNR, compression ratio, execution time of each frame, time taken for evaluating DCT method. Also it has been shown theoretically that the proposed technique take lesser time than the existing method.

Now a days public networks like internet is heavily used for various multimedia based applications like video on demand, video conferencing, pay per TV etc. as the data size in such applications is very large in comparison to text data it is necessary to compress the data before transmission. Digital video signals get compressed using some coding standards MPEG 1-4, H.264 / AVC before transmission over the wired or wireless channel. These standards do not provide security to the multimedia data. So, various encryption schemes are proposed to secure the data. Traditional solution to [1,2] provide confidentiality is to scramble the data in frequency or temporal domain but these days these techniques are vulnerable to attacks. Another way is to encrypt either uncompressed data or to compressed data (bit stream level) using the conventional cryptosystems like DES and AES that works on the blocks of data therefore known as block ciphers. These procedures provide highest security but also require high processing time that is undesirable for real time applications. Also the video data is voluminous than text data so this results in a decrease in speed. Also the information density is lower in multimedia data than text data so whole video data encryption is unnecessary. Hence the focus shifts from complete encryption schemes to the partial or selective encryption schemes that provides lower computational costs and increases speed by reducing the processing time. The basic concept of partial encryption is to select the most important [i]coefficients and encrypt them with conventional cryptographic ciphers. The non selected coefficients are sent to the transmission channel with no encryption. Since selected coefficients are protected it is impossible for an attacker to recover any information from these coefficients.

The rest of the paper is organized as follows. In section 2,we discuss the basic concept of video compression. Section 3, introduces the partial video encryption technique. In section 4, proposed modified technique is discussed. The results of experiments are detailed in section 5, where we present comparison results with Yengs et al algorithm [1].

Finally in section 6, conclusions are drawn and future studies are explored.

II. BASIC CONCEPT OF VIDEO COMPRESSION

A brief introduction to the process of video compression is given in this section. Video compression comprises of two levels. Firstly, *spatial compression* takes place when there is high correlation between pixels (samples) of same frames and is equivalent to that of JPEG compression. And then *Temporal compression* is used to remove temporal redundancy between adjacent blocks by using the concept of motion estimation.

Video sequence is a collection of group of pictures or still images called frames. There are three such types of frames.

I frame (intra frame): This is the first frame that represents the beginning of a scene and followed by P and B frames. Spatial compression process is only applied to I frame.

P frame (predicted frame): This frame is predicted by the past reconstructed frame.

B frame (bidirectional frame): These frames are predicted from the I frame and P frames.

The general sequence of frames in a GOP can be illustrated as in figure1:

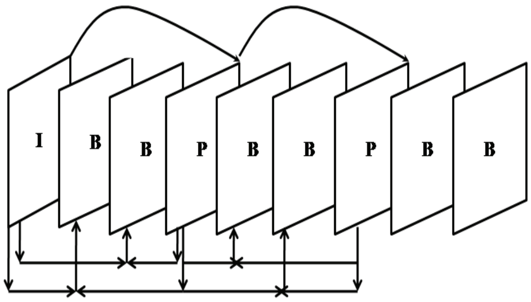


Figure 1: A sequence of GOP

The overall video compression process can be depicted as in figure 2. The main components of compression are:

Transform encoding

Quantization

Motion compensation and estimation

Zigzag reordering and RLE (Run Length Encoding)

Entropy encoding.

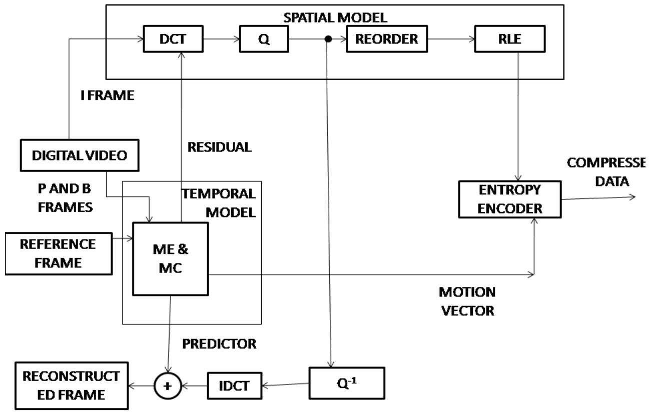


Figure 2: General Block Diagram of Video Compression [4]

Pixels in a video exhibit a certain level of correlation with the adjacent or neighboring pixels in the same frame and in the neighboring frames. The correlation in consecutive frames within a video is high. So in transform encoding phase a transformation from spatial (correlated) domain to uncorrelated domain takes place. This phase results in a transformation that maintains the relative relationship between the pixels but the redundancies are revealed.

Some of transforms that can be used [3] are image based transform (DWT (this is best suited for still images)), block based transform (DCT, KLT etc). The choice of transform depends on following factors:

- The data in transformed domain should be uncorrelated and compact (most of the energy should be concentrated into small number of values)
- Transform should be reversible.
- Transform should be computationally tractable.

The block based transform are best suited for compressing the block based motion compensated residuals.

The 1-D DCT (unitary transform) is applied on 1 D sample values and can be evaluated using the formula

$$y(x) = c(x)\sum_{n=0}^{N-1} f(n)\cos\frac{(2n+1)\pi x}{2N} \quad \text{............ (1)}$$

Where,

c(x) = $\frac{\sqrt{1}}{N}$ $, x = 0;$

and, $c(x) = \frac{\sqrt{2}}{N}$ $, x \neq 0;$

and, IDCT (inverse DCT) can be evaluated as,

$$f(n) = \sum_{x=0}^{N-1} c(x) y(x) \cos\left(\frac{\pi(2n+1)x}{2N}\right) \text{............... (2)}$$

for n=0, 1, 2, …………, N-1.

The first value at x=0 is known as DC coefficient that is the average value of the pixels, as at x=0,

$$y(0) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n) \text{........................ (3)}$$

and , all other coefficients are known as AC coefficients.

Similarly, 2D DCT (DCT-II) is used for calculating a 2D sample sequence and is given as in equation 4

$$y(x,y) = c(x)c(y) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos\frac{(2j+1)\pi y}{2N} \cos\frac{(2i+1)\pi x}{2N} \text{...... (4)}$$

In other form,

$$Y = AXA^T \text{........................ (5)}$$

where, X is a block of N x N samples and A is known as transform matrix.
Equation 4 can be viewed as applying successively 1D DCT twice once for column values and than for row values or vice versa. This property of DCT is known as separability.

Quantization: After the transform encoding the transformed coefficients are quantized to reduce the number of bits required for encoding. A quantizer maps a signal with a range of values *X* to a quantized signal with a reduced range of values *Y*. The quantizers can be broadly classified as scalar or vector quantizer. A *scalar quantizer* maps one sample of the input signal to one quantized output value and a *vector quantizer* maps a group of input samples (a 'vector') to a group of quantized values.

Motion estimation and compensation: this phase is the heart of temporal compression where the encoding side estimates the motion in the current frame with respect to a previous or future frame. A motion compensated image for the current frame is then created from the blocks of image from the reference frame. The motion vectors for blocks used for motion estimation are transmitted, as well as the difference of the compensated image with the current frame is also encoded. The main purpose of motion estimation based video compression is to save on bits by sending encoded difference images which have less energy and can be highly compressed as compared to sending a full frame. This is the most computationally expensive operation in the entire compression process.
The matching of one block with another is based on the output of a cost function. The block that results in the least cost is one that matches the closest to current block. There are various cost functions, of which the most popular and less computationally expensive is Mean Absolute Difference (MAD) given by equation (6). Another cost function is Mean Squared Error (MSE) given by equation (7).

$$MAD = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C_{ij-} R_{ij} \text{.................................... (6)}$$

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (C_{ij} - R_{ij})^2 \text{......................... (7)}$$

where N is the side of the macro block, C_{ij} and R_{ij} are the pixels being compared in current macro block and reference macro block, respectively.
Peak-Signal-to-Noise-Ratio (PSNR) given by equation (8) characterizes the motion compensated image that is created by using motion vectors and macro blocks from the reference frame.

$$PSNR = 10 * \log_{10} \left(\frac{(\text{peak to peak value of original data})^2}{MSE} \right) \text{..... (8)}$$

Zigzag reordering and RLE: Quantized transform coefficients are required to be encoded as compactly as possible prior to storage and transmission. In a transform-based image or video encoder, the output of the quantizer is a sparse array containing a few nonzero coefficients and a large number of zero-valued coefficients. Reordering (to group together nonzero coefficients) and efficient representation of zero coefficients are applied prior to entropy encoding.
The significant DCT coefficients of a block of image or residual samples are typically the 'low frequency'

positions around the DC (0, 0) coefficient. The nonzero DCT coefficients are clustered around the top-left (DC) coefficient and the distribution is roughly symmetrical in the horizontal and vertical directions. After quantization, the DCT coefficients for a block are reordered to group together nonzero coefficients, enabling efficient representation of the remaining zero-valued quantized coefficients. The optimum reordering path (scan order) depends on the distribution of nonzero DCT coefficients. For a typical frame block scan order is a zigzag starting from the DC (top-left) coefficient as shown in figure 3.

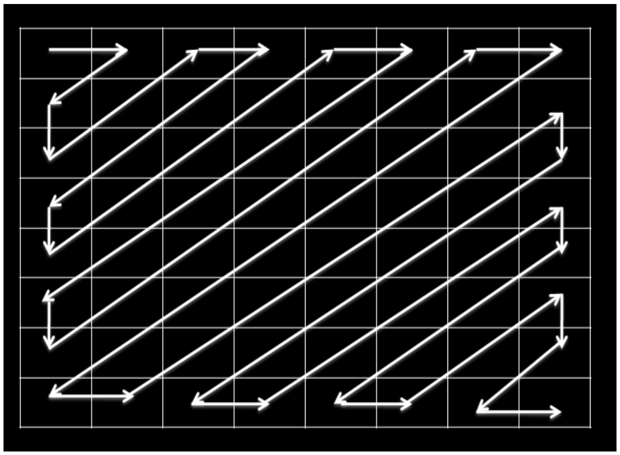


Figure 3: Zig Zag Reordering

Starting with the DC coefficient, each quantized coefficient is copied into a one-dimensional array. Nonzero coefficients tend to be grouped together at the start of the reordered array, followed by long sequences of zeros.

The output of the reordering process is an array that typically contains one or more clusters of nonzero coefficients near the start, followed by strings of zero coefficients. Higher-frequency DCT coefficients are very often quantized to zero and so a reordered block will usually end in a run of zeros.

III. PARTIAL VIDEO ENCRYPTION USING ALTERNATE TRANSFORM [3]

It focuses on 4x4 block of data. This scheme incorporates more transform rather than only one that explained in section 2, the general method for calculating the DCT. These new transforms are as efficient as DCT encoding of residual frames. The new unitary transforms can be derived from 1-D DCT for N = 4 sample values using equation (1),

For N=4,

$$c(0) = \sqrt{\frac{1}{4}} = \frac{1}{2} \text{..} (8)$$

and, $c(1) = c(2) = c(3) = \sqrt{\frac{2}{4}} = \frac{1}{\sqrt{2}}$ ....(9)

$$c(0) = \sin\frac{\pi}{4}\ \cos\frac{\pi}{4} \text{...................} (10)$$

$$c(1) = c(2) = c(3) = \sin\frac{\pi}{4} \text{.............} (11)$$

Due to symmetric property of cosine function,

$$\sin\frac{\pi}{4} = \cos\frac{\pi}{4} \text{..............................} (12)$$

$$\cos\frac{\pi}{8} = \cos\frac{3\pi}{8} \text{..............................} (13)$$

Using the above relations 1D DCT can be represented in a structure known as butterfly approach. The junction represents the addition operation and the number on line represents the multiplication operation.

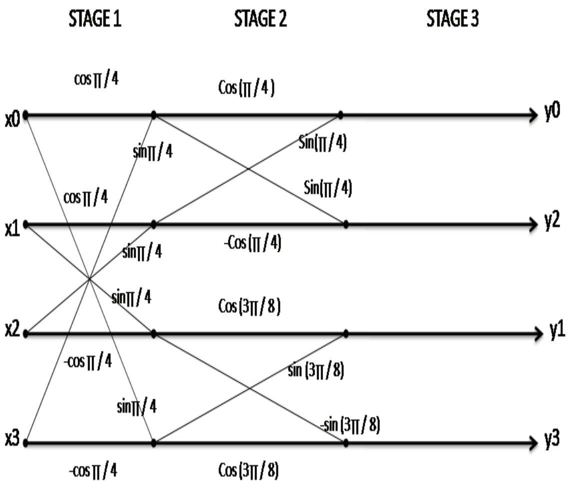


Figure 4: 1-D 4 POINT DCT METHOD [3]

[3] The flow graph consists of three stages. A plane based rotation of $\frac{\pi}{4}$ stage 1 and plane based rotation of $\frac{\pi}{4}$ and $\frac{3\pi}{8}$ at stage 2 and a permutation. New unitary transforms can be created by keeping stage 1 and 3 unchanged and changing the rotation angle at stage 2 as shown in figure 3.2.2 below, by varying the angles from $\frac{\pi}{4}$ to $\frac{\pi}{4} + \alpha 1$ and from $\frac{3\pi}{8}$ to $3\frac{\pi}{8} + \alpha 2$. Range of $\alpha 1\ and\ \alpha 2\ is - \frac{\pi}{4}\ to\ \frac{\pi}{4}$.

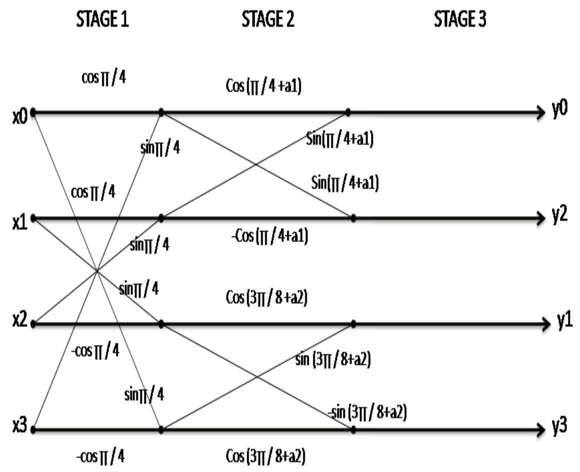


Figure 5: 4 POINT DCT INCORPORATING ROTATION ANGLES

[3] The scheme shows highest EPE (Energy Packing Efficiency) for highly correlated data (I frames) is when both a1 and a2 are set to zero. For weaker correlation between data (P and B frames) maximum EPE is shown at a1= $\frac{\pi}{8}$ and a2= - $\frac{\pi}{8}$

An encryption algorithm consists of two parts. First one is key generation and second is encryption using that key. This process is proposed for residual data only.

For the purpose of key generation RC4 key generator is used.

Steps for partial video encryption using ADE are as follows:

Design 2 $^{M-1}$ transform tables.

Repeat for each frame.

Initialize the RC4 key generator by a random 128 bit key.

For an input residual block of size 4*4 get M bit from the RC4.

Chose a transform table and apply it on input block based on M-1 bits.

Mth bit is used to encrypt the sign of DC component as change the sign of DC component if Mth bit is "1".

IV. MODIFIED ALTERNATE FAST DCT METHOD

The ADE scheme described in section 3 results in a increase in computational time for transform encoding as compared to general DCT method described in section 2, and hence a decrease in speed. On comparing equations (1) and the butterfly structure of fig.3 it is concluded that general DCT requires three additions and five multiplications operations for computing DCT of 4 elements that is lesser as compared to ADE which requires one addition, two subtractions and six multiplication operations. So for a 4x4 block of data general DCT requires 128 (64x2) multiplication operations and 24 (4x4x3) addition operations while ADE requires 192 (96x2) multiplication operations and 96 (48x2) addition operations. So we modify the above scheme and propose an alternating transforms to reduce the computations and hence increase the speed.

This can be achieved by interchanging the stage 1 and stage 3 of the ADE scheme as illustrated in figure 4.

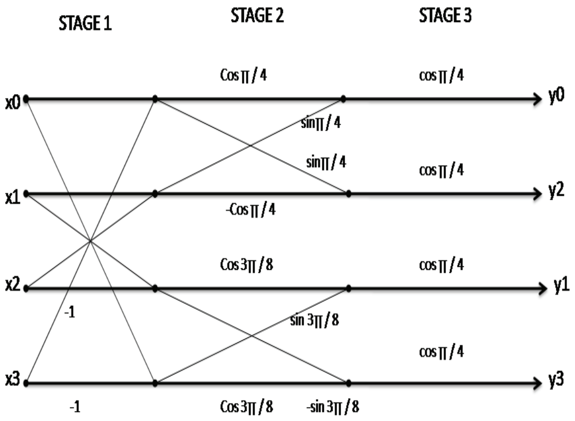


Figure 6: 4 PIONT DCT FOR MAFD

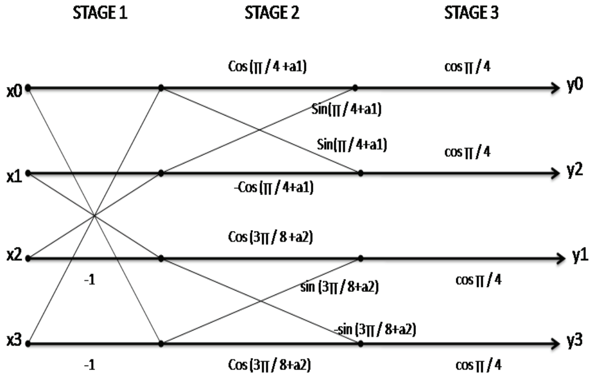


Figure 7: 4-POINT FAST DCT INCORPORATING ROTAION ANGLES

From the figure 7 it can be concluded that MAFD scheme require 96 (48x2) addition operations and 96 (48x2) multiplication operations to compute transform for a 4x4 block of data. Hence, resulting in a total reduction of 25% as compared to general DCT and 50% as compared to ADE scheme in computations.

V. RESULTS

In this section experimental results have been shown to demonstrate the effectiveness of the proposed scheme MAFD over ADE. For this purpose the four equi-space rotation angles are used $\frac{\pi}{4}, \frac{7\pi}{24}, \frac{8\pi}{24}, \frac{-\pi}{4}$ and four test video streams in grayscale mode are considered viz Miss America video consisting of 13 frames and Akiyo video with 30 frames both having resolution 176 x 144, 15 frames of Bear video of resolution 720 x 480 and 119 frames of Susie video with resolution 352 x 240 are chosen and both the procedures as explained in section 3 and 4 are implemented on blocks of 4 x 4 data of all the above video sequences using Image Processing Toolbox of MATLAB 7.0. Hence PSNR values, total number of bits required per pixel by each frame, time taken to compute DCT method and total execution time per frame, quality factor (PSNR / average bits per pixel) within the limitations of hardware and software are computed. Results are summarized as:

Table I: AVERAGE PSNR VALUES PER ENCRYPTED FRAME

VIDEO SEQUENCE	ADE	MAFD
MISS AMERICA (176X144) (13 FRAMES)	60.04938	60.01092308
AKIYO (176 X 144) (30 FRAMES)	55.984	55.92803333
SUSIE (352 X 240) (119 FRAMES)	55.20551	55.17861345
BEAR (720 X 480) (15 FRAMES)	55.872	55.8416

Table I above shows the comparison between average PSNR values for encrypted frames of different video sequences. It can be observed that both ADE and MAFD schemes results in approximately same average PSNR values.

Table II: AVERAGE BITS REQUIRED PER PIXEL PER FRAME

VIDEO SEQUENCE	ADE	MAFD
MISS AMERICA	1.230169	1.105838462
AKIYO	1.30983	1.14553
SUSIE	1.273624	1.068459664
BEAR	1.314787	1.212766667

Table II above shows the average bits required per pixel per frame values for different video sequences. It can be observed that MAFD scheme results in a decrease in number of bits per pixel requirement to 15% (approx) as compared to ADE scheme.

Table III: EXECUTION TIME TAKEN BY DCT METHOD

VIDEO SEQUENCE	ADE	MAFD
MISS AMERICA	1.450846	1.041923077
AKIYO	1.509333	1.072233333
SUSIE	4.931202	3.441420168
BEAR	21.57433	14.97227

Table III above shows the experiment results for execution time of DCT methods for different video sequences using both the schemes. It can be observed that MAFD scheme results in a decrease in execution time of DCT to 40% (approx) as compared to ADE scheme.

Table IV: TOTAL EXECUTION TIME PER FRAME

VIDEO SEQUENCE	ADE	MAFD
MISS AMERICA	2.241538	1.832615385
AKIYO	2.358	1.9209
SUSIE	12.27923	10.78944538
BEAR	137.9821	131.38

Table IV above shows the experiment results for average total execution time taken by different video sequences using both the schemes. It can be observed that MAFD scheme results in a decrease in total execution time to 22% (approx) as compared to ADE scheme.

Table V: AVERAGE QUALITY FACTOR PER FRAME

VIDEO SEQUENCE	ADE	MAFD
MISS AMERICA	49.06731	54.93738462
AKIYO	43.0224	49.4201
SUSIE	43.38891	51.78489916
BEAR	43.31273	47.31186667

Table V above shows the experiment results for average quality factor per frame for different video sequences using both the schemes. It can be observed that MAFD scheme results in an increase in quality to 12% (approx) as compared to ADE scheme.

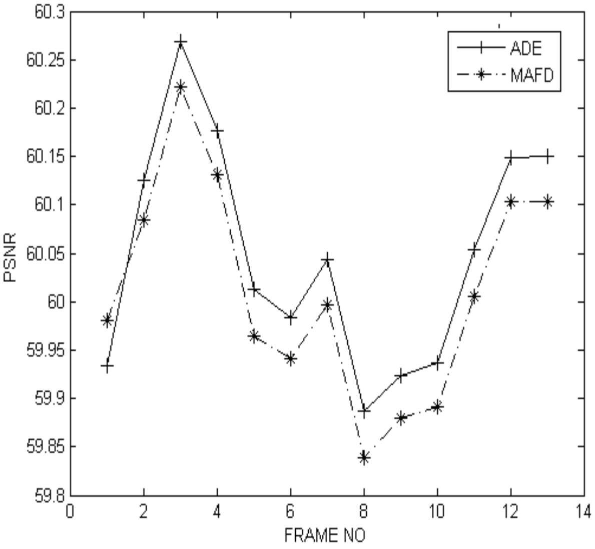


Figure8: COMPARISON OF PSNR VALUES OBTAINED BY APPLYING BOTH METHODS FOR ENCRYPTED MISS AMERICA VIDEO

Figure 8 above represents the comparison between PSNR values of encrypted frames of Miss America video obtained by both the schemes (ADE and MAFD). It can be observed that PSNR values of encrypted frames obtained by MAFD scheme is lower except for the first frame (I frame).

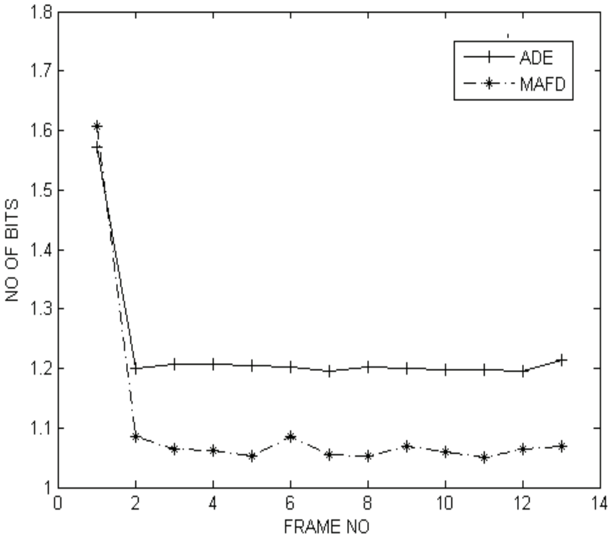


Figure 9: COMPARISON OF NO. OF BITS REQUIRED PER PIXEL OBTAINED BY BOTH METHODS FOR MISS AMERICA VIDEO

Figure 9 above represents the comparison between number of bits required after entropy encoding (Huffman Encoding) by encrypted frames of Miss America video obtained by both the schemes (ADE and MAFD). It can be observed that no of bits required by encrypted frames obtained by MAFD scheme is lower in case of P and B frames while its approximately similar in case of first frame(I frame).

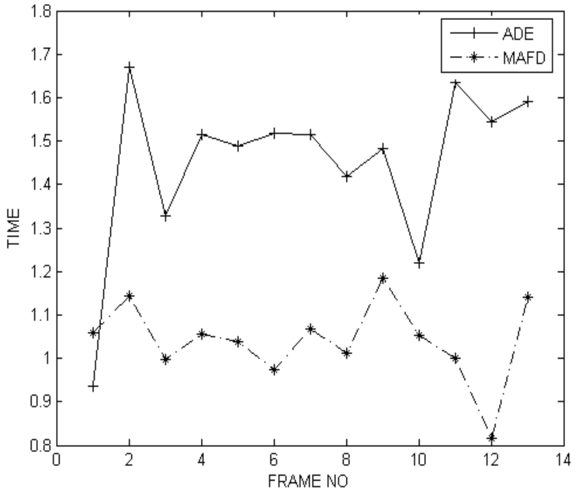


Figure10: COMPARISON OF EXECUTION TIME TAKEN BY DCT METHOD IN BOTH SCHEMES FOR MISS AMERICA VIDEO

Figure 10 above represents the comparison of execution time of DCT method taken for Miss America video by both the schemes (ADE and MAFD). It can be observed that time taken by DCT method as obtained by MAFD scheme is lower. This is due to reduction in computations in modified schemes as compared to the ADE scheme.

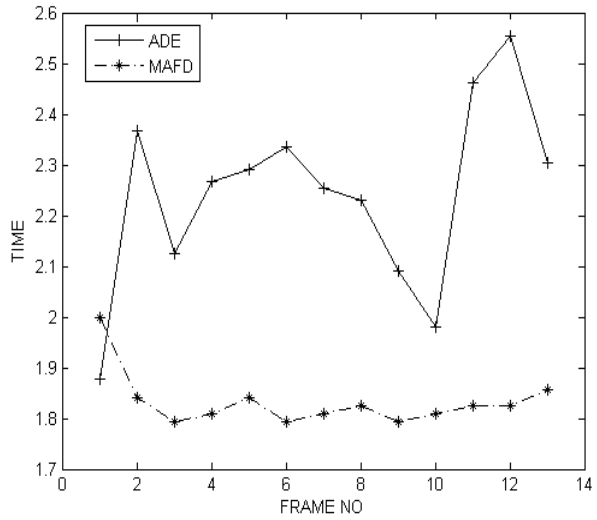


Figure11: COMPARISON OF TOTAL EXECUTION TIME PER FRAME FOR MISS AMERICA VIDEO

Figure 11 above represents the comparison of total execution time per frame taken for Miss America video by both the schemes (ADE and MAFD). It can be observed that time taken by DCT method as obtained by MAFD scheme is lower.

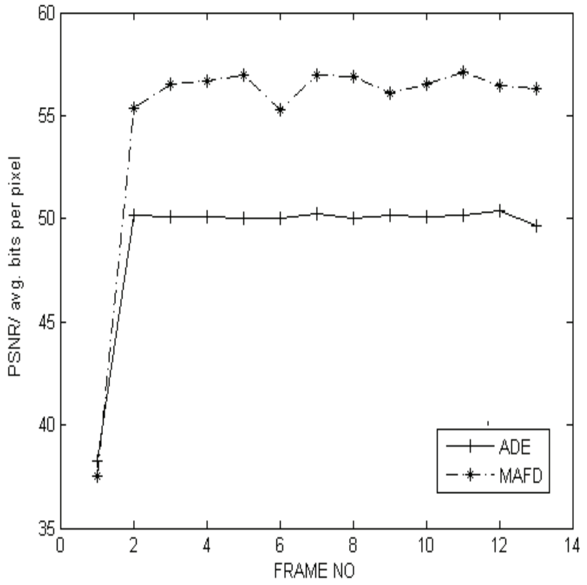
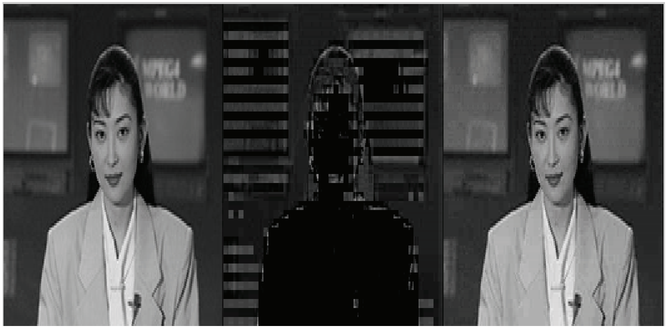


Figure 12: COMPARISON OF QUALITY FACTOR VALUES FOR MISS AMERICA VIDEO

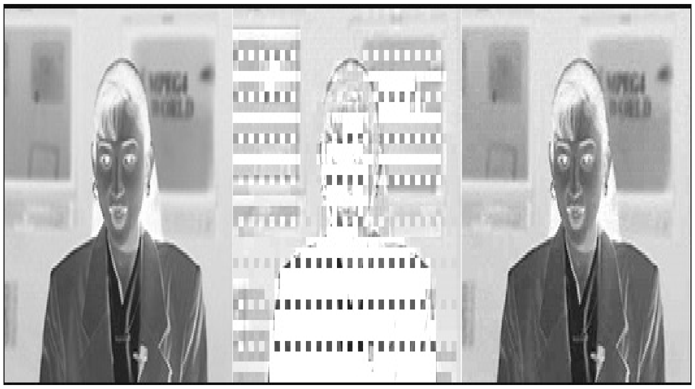
Figure 12 above represents the comparison of quality factor i.e. ratio of PSNR and average bit required per pixel per frame for Miss America video by both the schemes (ADE and MAFD). It can be observed that quality factor is higher in case of modified scheme (MAFD).

Figures 13 to 16 below displays the screenshots of original frame, reconstructed encrypted frames and the predicted frames of Akiyo, Miss America and Susie video sequences under both the methods i.e. ADE and MAFD.



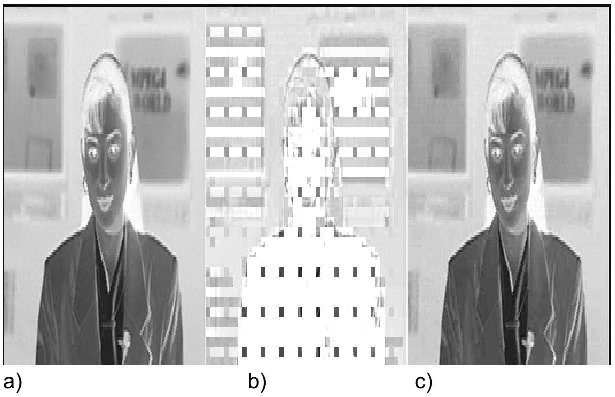
a) b) c)

Figure 13 : AKIYO VIDEO (FRAME 3 ADE scheme) a) ORIGINAL FRAME b) ENCRYPTED RECONSTRUCTED FRAME c) PREDICTED FRAME



a) b) c)

Figure 14 : AKIYO VIDEO (FRAME 29 ADE scheme) a) ORIGINAL FRAME b) ENCRYPTED RECONSTRUCTED FRAME c) PREDICTED FRAME



a) b) c)

Figure 15 : AKIYO VIDEO MODIFIED (FRAME 3 MAFD method) a) ORIGINAL FRAME b) ENCRYPTED RECONSTRUCTED FRAME c) PREDICTED FRAME

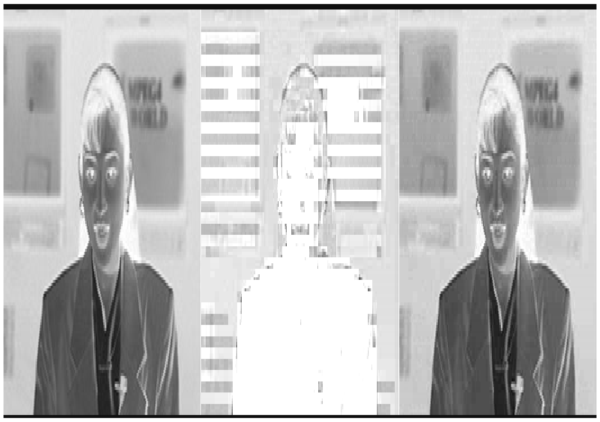


Figure 16 : AKIYO VIDEO (FRAME 29 MAFD method) a) ORIGINAL FRAME b) ENCRYPTED RECONSTRUCTED FRAME c) PREDICTED FRAME

VI. CONCLUSION AND FUTURE SCOPE

The procedures as explained in section 3 and 4 are implemented in MATLAB and compared practically on the basis of parameters PSNR, number of bits per pixel required, execution time taken by each frame to evaluate DCT method, total execution time taken by and quality factor (ratio of PSNR and number of bits per pixel required). It is shown both theoretically and practically that MAFD scheme requires less computational time as compared to that of ADE. And it can be concluded from the results that both the procedures are providing approximately same average PSNR values that is approximately 56 db. A higher PSNR value represents less error.

The average bits required per pixel per frame values for different video sequences in case of MAFD scheme results in a decrease in no of bits per pixel requirement to 15% (approx) as compared to ADE scheme.

Execution time of DCT methods for different video sequences using both the schemes in case of MAFD scheme results in a decrease in execution time of DCT to 40% (approx) as compared to ADE scheme. The reason for this can be explained theoretically by examining figure 4,figure 5, figure 6 and figure 7, for the evaluation of 1 -D 4 sample will require evaluating 24 multiplications and 12 additions. Hence it results in an increased overhead of computations. With the modified scheme, that is interchanging stage 1 and stage 3 will require 12 multiplications and 12 additions for a 1-D sequence of 4 sample values. Hence results in reduction in number of multiplications to half. So this results in a reduction in time taken to evaluate DCT function and hence the overall time for execution.

The experimental results for average total execution time taken by different video sequences using both the schemes shows that MAFD scheme results in a decrease in total execution time to 22% (approx.) as compared to ADE scheme.

Average quality factor per frame for different video sequences using both the schemes in case of MAFD scheme results in an increase in quality to 12% (approx) as compared to ADE scheme.

As the selective sign encryption of DC coefficients is used for the encryption purpose so the overhead due to encryption process is very less.

This work is carried out for 4x4 input blocks size due to this the reconstructed frames will be more accurate but on the other side as we decrease the block size the energy of the block decreases but results in an increase in computations and complexity as compared to the 8x8 blocks. Also both the schemes (ADE and MAFD) are compared on the basis of the parameters like PSNR, average bits per pixel, execution time etc. the analysis can be done on the basis of various attacks for which the system can be vulnerable.

REFERENCES

1. I. Agi and L. Gong, “An Empirical Study of Secure MPEG Video Transmission,” Proceedings of the Symposium on Network and Distributed Systems Security, pp 137-144, IEEE, 1996.
2. L. Qiao and K. Nahrstedt, “Comparison of MPEG Encryption Algorithms,” International Journal on Computer and Graphics, Special Issue on Data Security in Image Communication and Network, 22(3), pp 437-438, 1998.
3. Siu-Kei Au Yeung, Shuyuan Zhu and Bing Zeng,”Partial Video Encryption Based on Alternating Transforms”, IEEE Signal Processing Letters, Vol. 16, No. 10, pp 893-896, October 2009
4. I. Richardson, H.264 and MPEG-4 Video Compression. Hoboken, NJ: Wiley, 2003.
5. Jian Zhao, ”Applying Digital Watermarking Techniques to Online Multimedia Commerce”, In: Proc. of the International Conference on Imaging Science, Systems, and Applications (CISSA97), June 30-July 3, 1997, Las Vegas, USA.
6. http://trace.eas.asu.edu/