Joint Source and Channel Coding for MPEG-4 using Steganography of Channel Coding Information

Hassan Farsi[1], Pouriya Etezadifar[2]
1- Department of Electronics and Communications Engineering, University of Birjand, Birjand, Iran
Email: hfarsi@birjand.ac.ir
2- Department of Electronics and Communications Engineering, University of Birjand, Birjand, Iran
Email: p.etezadifar@birjand.ac.ir (Corresponding author)

Received: October 2012 Revised: April 2013 Accepted: May 2013

ABSTRACT:
The progress of the technology in the recent decades has caused the video transmission by communication channels to meet high demands. Therefore, many methods have been proposed to improve the quality of the video under channel errors. The aim of this paper is to increase the Peak Signal to the Noise Ratio (PSNR) for video reconstructed by the receiver .This is achieved by increasing channel encoder rate but in constant transmission rate. In the proposed method the video frames' parts containing more energy are coded by the local channel coding inside the source encoder. Then, the resulted data is embedded in the low power frames' parts. The proposed method is able to increase the channel coding rates without increasing the amount of the information for any frame. This method provides more robustness for video frames against channel errors. The proposed method is tested for different source coding rates and several Signals to Noise Ratio (SNRs) for channel and the obtained results are compared with a new method.

KEYWORDS: Video coding, Variable bit rate, Channel coding, F4 encoder, Steganography

1. INTRODUCTION

A few years ago, the concept of the video communication using cellular phones was considered highly impractical. But, today, increasing growth of the wireless channels and mobile telecommunication use and tend to serve the achievements of the modern telecommunications for the real-time multimedia data transmission such as video involves advance research in this area. Video recording and transmission features are very common in the most consumer cellular phones. So, the different methods of video coding with the lower bit rate and higher quality have been proposed so far. For instance: Cheung and Zakhor presented a bit allocation method for allocating source and channel bits between the subbands of a scalable video, such that the overall distortion was minimized given the channel conditions and a total bit budget [2]. Kondi proposed a JSCC scheme based on the universal rate-distortion curves for the motion compensated discrete cosine transform (DCT) based on the signal-to-noise-ratio (SNR) scalable video [3]. Bystorm and Modestino presented a JSCC method to optimally allocate source and channel coding bits with a fixed constraint on the transmission bandwidth for video transmission over Additive White Gaussian Noise (AWGN) channel [4]. Zhanget presented a distortion model for estimating the distortion due to quantization, error propagation and error concealment for video coding and transmission over the packet switched networks [5]. In the wireless and mobile telecommunication due to the nature of the wireless environment and channel errors the quality of the video is affected at the receiver. Several methods have been proposed to reduce the errors caused by the channel. Channel coding is an efficient method which is generally used to detect and correct the errors. However, channel coding was performed without considering the source coding rate. In the recent years different methods as a combination of channel and source coding have been introduced. These methods are known as the Joint Source and Channel Coding (JSCC). The JSCC is generally based on the channel estimation [2], [3], [4] and [5]. This means that in order to have the higher PSNR, the higher channel coding rate is required. In this paper we propose a new JSCC method which increases channel coding rate but the total transmission rate is maintained still constant. In fact by increasing the channel coding rate more robustness of the video frames can be provided. One of the most important problems in the general methods is that the channel information is required for the modification of the source and channel coding rates. However, in the proposed method the channel information is not needed. In other words, the proposed method is able to be applied on every source coding rate, independently and to improve the quality of the reconstructed video frames at the receiver. In this paper, the Peak Signal to the Noise Ratio (PSNR) and also the Bit Error Rate (BER) are used as measures to indicate the quality of the reconstructed video. This paper has been organized as follows:

In section 2, the method of encoding a video frame using the MPEG-4 part 2 is introduced for both Intra-picture (I-picture) and Predicted-Picture (P-picture). Then, in Section 3, the steganography method used in the proposed method is explained. In Section 4, the channel coding is briefly indicated. Next, in section 5, the proposed method is introduced and explained step by step. Finally, the performance of the proposed method is evaluated and compared with the method indicated in [1].

2. VIDEO FRAMES CODING

In this section, a group of pictures (GOP) is shortly introduced. Then video frames coding and decoding are explained using the MPEG-4. In the MPEG standard, three types of pictures are introduced by I, P and B briefly. The GOP is made by combining these pictures. As an example, a GOP Structure is shown in Figure 1.

I B B P B B B P B B B P B B B P I ......
I P P I P P I P P I P P I P P I ......
I I I I I I I I I I I I I I I I I I I I ....

Fig. 1. An example of the GOP

A picture which is coded by its information is called I-picture. Thus JPEG compression algorithm is used for coding the I-pictures [6]. The picture which is estimated by the nearest previous I or P picture is called P-picture. For the P-picture, the motion compensation predictor is also used. The picture which is estimated by the nearest previous and next, I or P picture is called B-picture. For the B-picture, the motion compensation predictor is also used, similar to the P-picture. We use two P-pictures and I-pictures through the GOP for the simulation, because using the B-pictures (Bidirectional-pictures) leads to a delay in the system. The delay is created in the B-pictures because the prediction is performed on the next and previous frames. Thus, the system is forced to wait until the next frames to be entered and buffered such that the estimation of the B-picture is possible to be performed.

2.1. An overview of the MPEG-4 encoding method

In this paper, we use the MPEG-4 part 2. The structure of the encoder and decoder of I and P, respectively, are shown in Figure 2. In the following section some of the operations used in the encoder and decoder shown in the Figure 2 are described.

Reorder: After quantization of the DCT coefficients, due to having many zero values, the obtained coefficients are arranged in zigzag form. Therefore the values related to the different frequencies are grouped [7]. Zigzag arrangement is shown in Figure 3.

Data Packetization: The occurring errors in the data transmission may cause the loss of synchronization between the transmitted video frames. To overcome this problem a resynchronized mechanism is required. One method for synchronization is to use the data Packetization. Thus, a number of markers are inserted in each packet for synchronization. The encoder divides each frame into packets and inserts related markers for synchronization at the beginning of each packet. The encoder and decoder perform resynchronization when the markers located in the beginning of each packet are received.

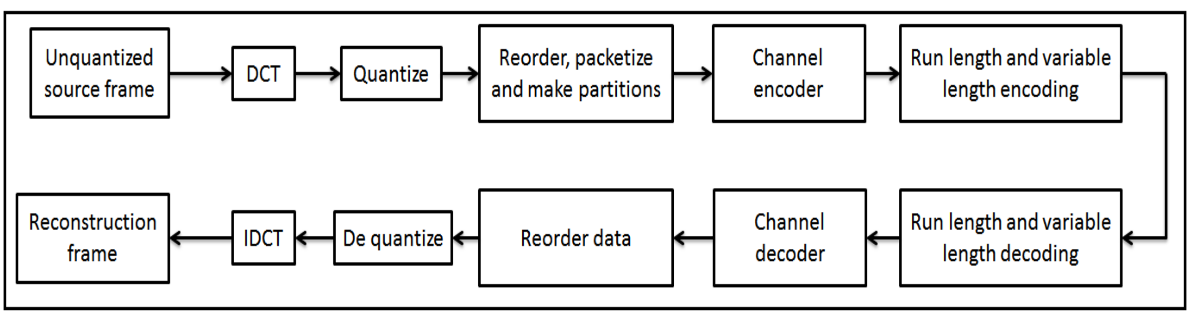

Fig. 2. MPEG-4 encoder and decoder for I-pictures

This procedure causes synchronization between the video frames and avoids the additive errors.

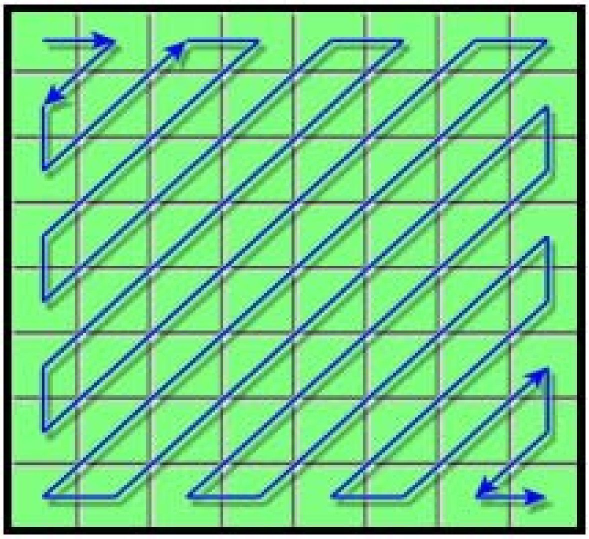

Fig. 3. Selection of the DCT coefficients for picking along the zigzag method

Data Partitioning: In this model, the data packet is divided into two areas. The main idea is to separate the data with the higher values (DC coefficients of DCT matrix, coding model information and motion vectors) from the data with the lower values (AC coefficients of DCT matrix and residual errors). For the I-picture transmission, the first area contains the source coding information and DC coefficients. However the second area containing lower values has AC coefficients. For the P-picture transmission, the first area contains the source coding information and motion vectors whereas the second area contains DCT data (texture, DC and AC coefficients). In the data partitioning the markers (which are unique) are used for synchronization. Figure 4 illustrates the packet structure for MPEG-4.

$$MSE\left(d_x, d_y\right) = \frac{1}{M_1 N_1} \sum_{(m,n)\in W} \left(b\,[m, n, k] - b\,[m - d_y, n - d_x, k-1]\right) \quad (1)$$

Motion Compensation Block: The motion compensation algorithm is based on the Minimum Mean Square Matching Error (MMSE) between the two successive frames that are based on the following equations:

$$\begin{pmatrix} \hat{d}_x \\ \hat{d}_x \end{pmatrix} = \arg \underbrace{\min}_{(d_x, d_y)} MSE\left(d_x, d_y\right) \quad (2)$$

In (1), the third parameter (b) including (k) and (k - 1), calls the pixels of the two consecutive frames and the two (m) and (n) variables show the pixel location in a two dimensional space for each frame. The variable (W) represents the search window, and the motion prediction vector is obtained by searching at this range. W = 15 is considered in this paper. This means that the window size is 16 * 16 for searching. Also the prediction window size is considered to be 8*8.

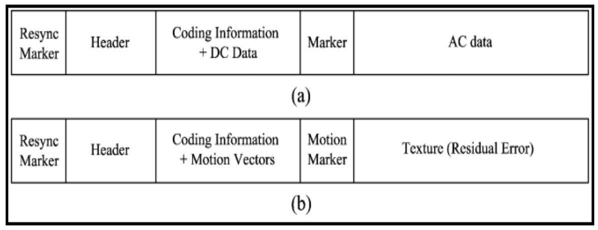

Fig. 4. MPEG-4 packet structure for (a) I frames and (b) P frames

In other words, the two consecutive frames within the search window are compared with each other, and the least value is stored as a member of the motion prediction matrix. Next, after the prediction of a frame completely and construction of the motion prediction matrix, the predicted value is subtracted from the previous frame. Finally, the obtained value is applied to the DCT block. After passing through the DCT block, the obtained value is quantized by 64 levels. At the receiver, as shown in Figure 2, firstly the received data is dequantized, and then the dequantized information is passed through IDCT block. Finally, using the motion prediction matrix, the input frame is added to the previous frame and therefore the transmitted frame is reconstructed.

3. STEGANOGRAPHY METHOD

The base of the steganography is to design a method or system in which the information is embedded to the original information such that the quality of the original information is maintained. Thus a picture or a message in other image (which is known as cover image) is hidden so that the resulting image would be visually similar to the cover image. Therefore the steganograph image contains a hidden message or image. There are many methods for steganography, such as: Hide & Seek, Jsteg, Outguess 0.1, Outguess 0.2, F3 and so on [9], [10], [11] and [12]. In this paper, we have used F4 steganography algorithm because it is able to provide a lot of information for embedding [11]. F3 method has two problems: first, it results in too many zeros and second, the odd numbers are more than even numbers which causes the statistical profile of pictures to be affected. These shortcomings are solved by using F4 method through mapping the coefficients (even negative coefficients indicate one, odd negative coefficients indicate zero, odd positive coefficients indicate zero and even positive coefficients indicate one.). The steganography process is performed by the comparison of the resulted bits using this mapping on DCT coefficients with the message bits.

4. CHANNEL CODING METHOD

Reed-solomon coding is used in this article which is a class of cyclic BCH code [16], [17], linear and non-binary and it is made by GF (q) field [18]. The reason for using this type of coding is non-binary characteristics. Therefore, the information of each block can be directly coded without converting to the binary. Thus the speed of the video frame coding increases.

5. THE PROPOSED ALGORITHM

In telecommunication system generally after source coding, channel coding is applied to compensate channel distortion. In the proposed method, the channel coding rates can be increased without increasing the transmission rate. The structure of the proposed encoder is shown in Figures 5 and 6 for I-pictures, but the same is also used for P-pictures. After calculating and quantifying of the DCT coefficients, the all DCT blocks energies in 8 * 8size are calculated by using (3).

$$Energy = \sum_{i=1}^{\substack{Number\ of\ all \\ DCT\ blocks}} \sum_{(m,n)\in W} (b_i[m,n])^2 \qquad (3)$$

Next steps are as follows:

- **Step 1:** In this stage, the threshold value is needed to be calculated. Threshold value for each frame is different from other frames. The threshold value is considered to be equal to 0.1 of the frame energy. Then, the number of the blocks having the more energy than this threshold value is calculated. If it is between35% to 45% of the total numbers of the blocks in each frame, the threshold value is considered as the main threshold. Note that this is obtained experimentally under the conditions that the minimum distortion is created in the reconstructed frame after steganography and the amount of the information for embedding in low-energy blocks is high. However, if the number of the blocks having the more energy than the threshold value is not in the determined range (35% to 45% of the total number of blocks) by giving an offset to the threshold value (note that if the number of blocks having the more energy than the predetermined threshold is less or more than the number of the blocks in determined range, the threshold should be increased or decreased, proportionally).we obtain the optimal threshold. This offset is created by adding or subtracting 1% of the total energy of the frame to the predetermined threshold.
- **Step 2:** The blocks containing the less energy than the threshold calculated in step 1, are selected.
- **Step 3:** Calculating the numbers of bits that can be embedded in the selected blocks in step 2 using F4 method.
- **Step 4:** The energy of each block is calculated and compared with the threshold level determined in Step 1.
- **Step 5:** If the block energy is greater than the threshold, this block is encoded using the local channel encoder. The code rate is calculated using (4).

$$R_c = \frac{N_s}{N_b} \qquad (4)$$

Where Rc is coding rate, Ns is the number of bits that can be embedded in the blocks with the lower energy than the threshold using the F4 and Nb is the number of the blocks with the higher energy than the threshold. In fact, the (4) indicates the coding rate of the channel encoder for the blocks with the higher energy than the threshold. The difference between the local and global channel encoders is that the obtained bits from the local channel encoder for each block are not embedded in the block itself. Because when the energy of a block is high, embedding in this block corresponds to the higher distortion and the lower PSNR in the relative frame. Instead, the data is embedded in the blocks containing the lower energy (these blocks have been identified previously in Step 4) than the threshold.

- **Step 6:** In this stage, after embedding the information related to the local channel encoder in the blocks with low energy, the blocks are rearranged according to the initial order. After reordering, packetizing and partitioning, the resulted blocks are then coded by the global channel encoder and using Huffman coding [14], [15].
- **Step 7:** After embedding the message bits, by considering one bit in Header, the blocks containing the embedded message can be specified by the other blocks. In addition, the number of bits embedded in each block is sent to the receiver as the Header. Therefore, the receiver has no problem to detect the embedded data and to assign them to the original blocks.

5.1. The proposed decoding structure

In the decoder, the information is reconstructed as follow:

- **Step 1:** After receiving the information, it is decoded by Huffman decoder followed by the global decoder and then reordering is performed. Next, the first step of the proposed method is applied. In this stage according to the Header, the input blocks are classified into two groups: High energy group and low energy group containing embedded information.

- **Step 2:** In this stage the data embedded in a block having low energy is decoded. Then this data with the blocks containing the high energy are decoded by the local channel decoder. Next, the data embedded in the low energy blocks is extracted. After decoding the blocks, they are rearranged and the transmitted frame is reconstructed.

In order to clarify the process of the proposed method, Figure 7 demonstrates the proposed technique as an example. As shown in Figure 7, each Foreman frame is divided into blocks in size 8 × 8, after computing DCT for each block. DCT coefficients are quantized. After quantization, the frame is converted to a set of quantized DCT blocks. In Figure 7, the two blocks one of which has the lower energy than the threshold and the other one has the higher energy than the threshold, have been shown. As shown the block having the high energy is coded by a local channel encoder and then the information related to local channel encoder is embedded in the block containing low energy. The new constructed block is replaced instead of the previous block. After reordering, packetizing and partitioning, the data are coded by the global channel encoder and then using variable length Huffman encoder is coded and transmitted. At the receiver (as shown in Figure 6), the embedded information in the block having lower energy than the threshold is extracted, and with the block having higher energy than the threshold are decoded by local channel encoder and the frame is reconstructed.

6. EVALUATION OF THE PROPOSED METHOD

First, we introduce several parameters used in the simulation and then we compare the results of the proposed method with the results of the new method indicated by Farooq in [1]. In this article the channel is Binary Symmetric Channel (BSC), the occurred error is Additive White Gaussian Noise (AWGN) and fading has Rayleigh PDF (Power Density Function). We have used two video files named Foreman and Walk [19]. The number of Foreman video frames is 123 that 3 frames are coded as the I-picture and the rest frames are divided into three parts. Each part has 40 frames which are coded and transmitted as P-pictures. It should be noted that 3 frames coded by the I-picture, at the beginning of each part, are transmitted. The numbers of the Walk video frames are 105 in which 5 frames are coded as the I-picture, and the rest frames are divided into five parts. Each part contains 20 frames which are coded as the P-picture. It should be noted that 5 frames coded by the I-picture, at the beginning of each part are transmitted. Frame size is equal to 288 * 352. Also frames transmission rate is 25 frames per second. This simulation has been performed at three transmission rates for source. Since the number of transmission frames is constant for a fixed source coding rate, the higher transmission rate corresponds to the higher channel coding rate and therefore results in more robustness against the noise for the video frames. The channel encoder rates for the two video files with different rates are shown in Table 1.Also, the channel encoding rates are calculated in (5) in terms of bits per second (bps).

$$R = \frac{k}{n} \quad (5)$$

In (5), (k)is the number of message bits that should be coded. (n) is the length of the sequence that includes the original message and the added information by the channel encoder. The Encoder is able to detect and correct the errors less than $\frac{n-k}{2}$. Thus if (n - k) (numbers of parity bits) is become higher, the more message bits can be corrected.

As shown in Table 1, if the transmission rate increases, the channel encoding rate will increase too. This means that the energy of the received frames has been increased.

Table1. Channel encoder rate for two video sequences Foreman and Walk with different transmission rates

Source transmission rate	Type of video sequence which transmit	n	k	R_c
384 Kbps	Foreman	15	7	0.466
768 Kbps	Foreman	15	3	0.2
2 Mbps	Foreman	31	2	0.064
256 Kbps	Walk	15	9	0.6
512 Kbps	Walk	15	5	0.33
1.5 Mbps	Walk	31	4	0.129

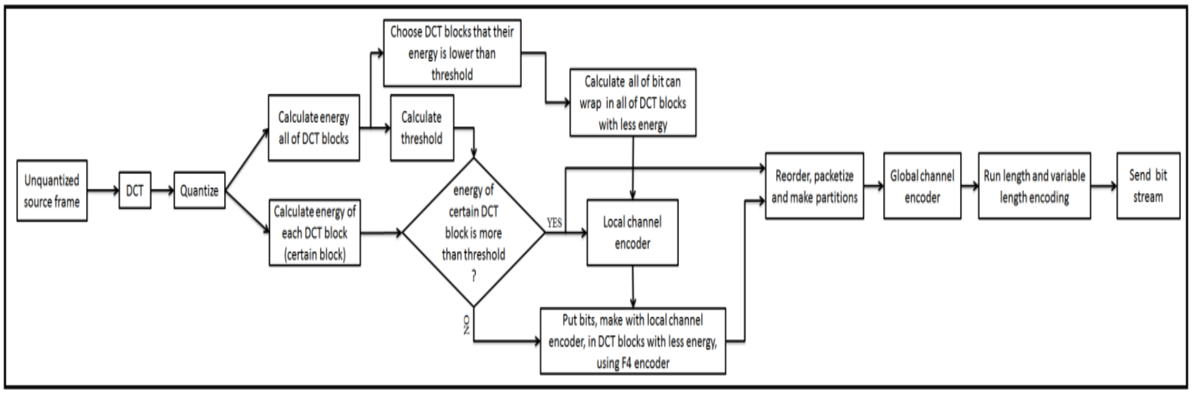

Fig. 5. Proposed encoder method for I-picture

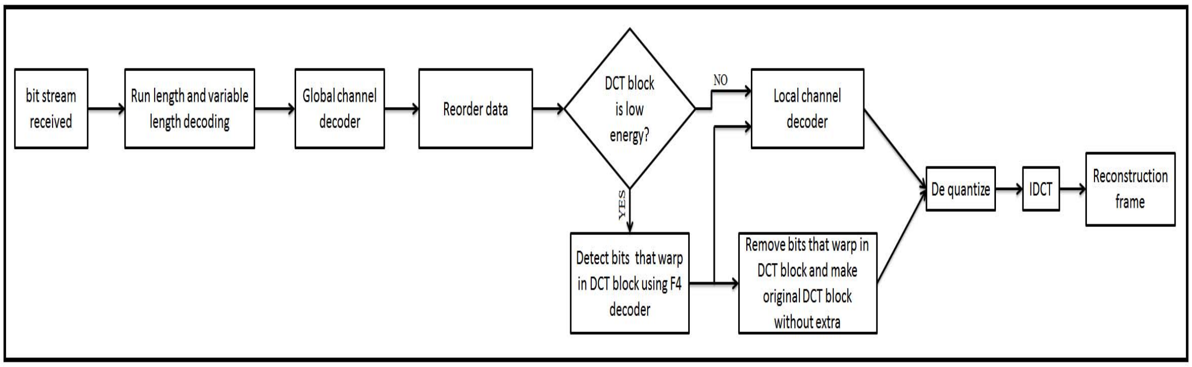

Fig. 6. Proposed decoder method for I-picture

After simulation of the proposed method, the obtained results have been compared with the results of Farooq method in [1]. Figures 8, 9, 10, show the obtained PSNR by applying the proposed method in comparison with [1] for all 123 frames of Foreman. The only difference in Figures 8, 9 and 10 is the source coding rate. This means that if the source coding rate increases the transmission of the frames will robust against the channel noise and distortion. Also Figures 11, 12 and 13 show the obtained PSNR by applying the proposed method in comparison with the method in [1] for all 105 frames of Walk. As shown the proposed method through embedding channel encoder bits in transmitted frames is able to increase the channel encoding rate without increasing the transmission bit rate. Therefore the quality of the reconstructed frames has considerably improved.

7. CONCLUSION

In this paper, we proposed a novel method in which the received video frames contain higher PSNR in comparison with the new method indicated by Farooq. Therefore, the reconstructed frames have a higher quality. In the proposed method, due to the use of the steganography for the embedding local channel coding information in low energy video frames, the numbers of bits embedded in the video frames are low in comparison with the used bits for the global channel coding. Therefore, when the channel errors increase, the proposed method considers more bits for channel coding but transmission rate is still maintained constant whereas in the new methods such as Farooq method for a constant transmission rate, it is impossible to provide the higher channel coding rate. In the other words, in the new JSCC methods, the higher channel coding rate results in a lower source coding rate and vice versa for the constant transmission rate. However, in the proposed method a local channel coding is applied on the high energy video frames and the relative local channel information is embedded in the low energy video frames, such that the transmission rate is still constant. Thus, the proposed method is able to provide the higher PSNR in comparison with Farooq method especially when the channel errors or SNRs increase and therefore the proposed method results in a higher quality for the reconstructed video frames.

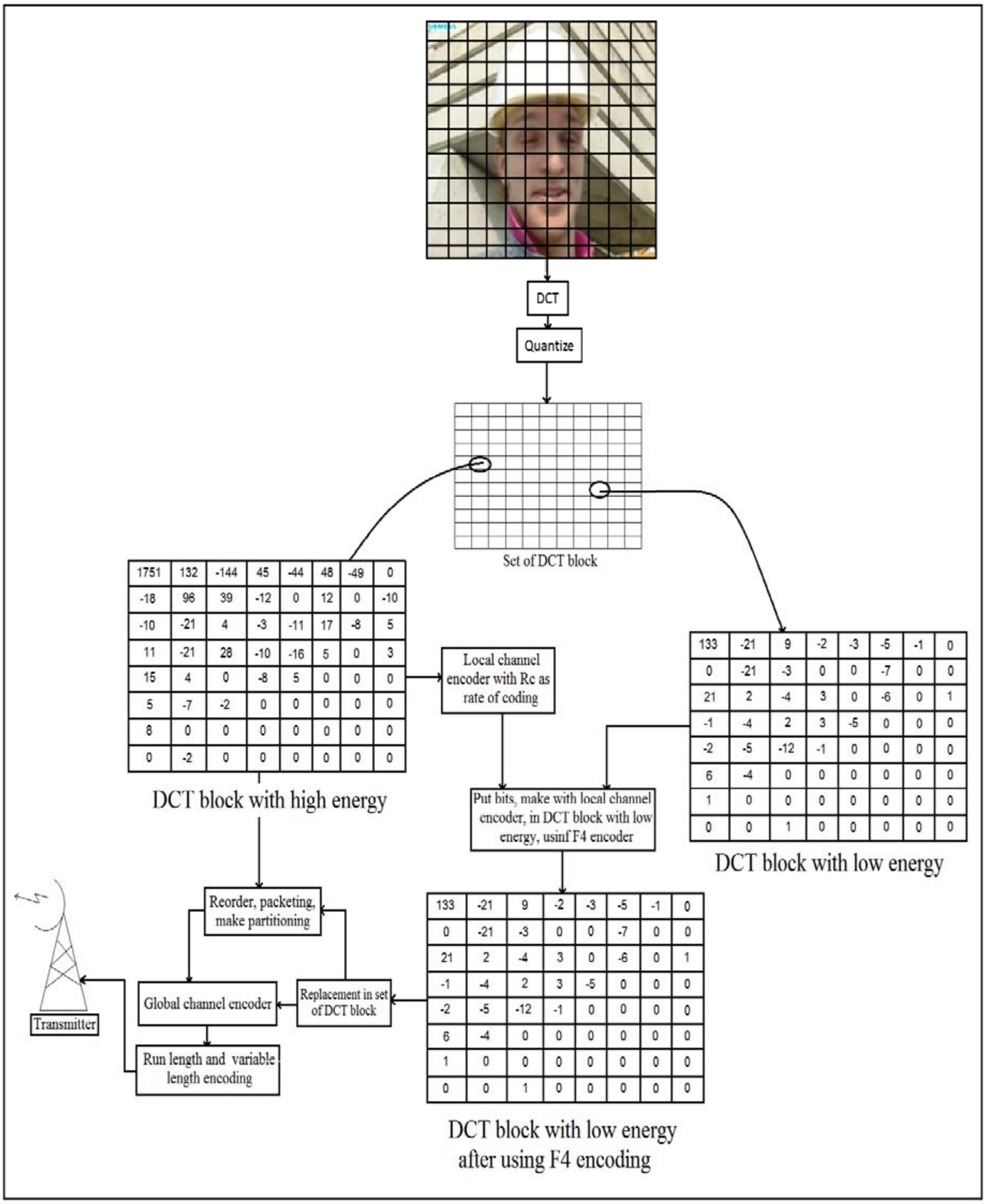

Fig. 7. An example of the performance of the proposed algorithm for I-picture

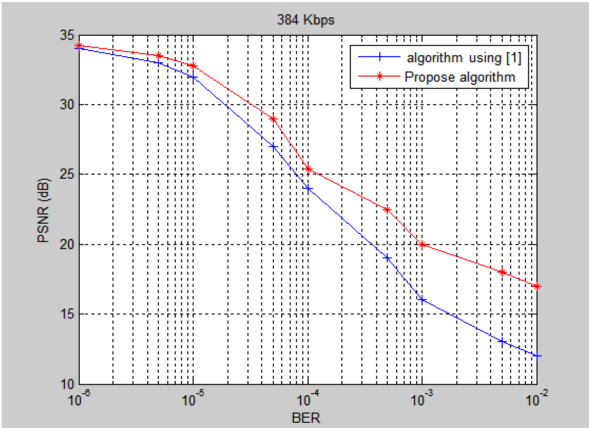

Fig. 8. Comparison between the PSNR for 123 frames of Foreman and the proposed method and algorithm used in [1], with source coding rate = 384 Kbps

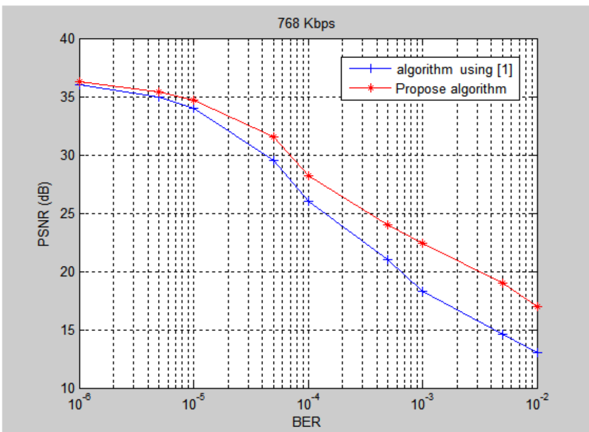

Fig. 9. Comparison between the PSNR for 123 frames of Foreman and the proposed method and algorithm used in [1], with source coding rate = 768 Kbps

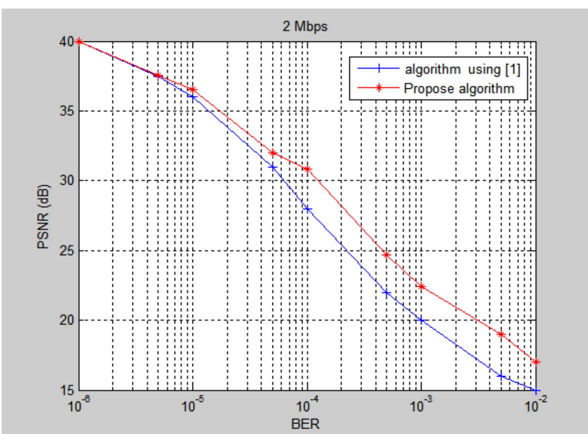

Fig. 10. Comparison between the PSNR for 123 frames of Foreman and the proposed method and algorithm used in [1], with source coding rate = 2 Mbps

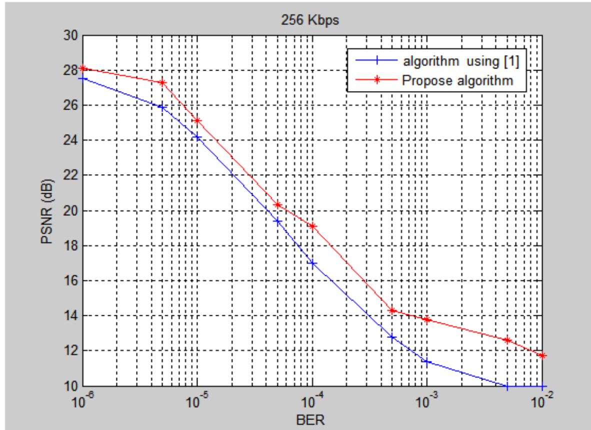

Fig. 11. Comparison between the PSNR for 105 frames of Walk and the proposed method and algorithm used in [1], with source coding rate = 256 Kbps

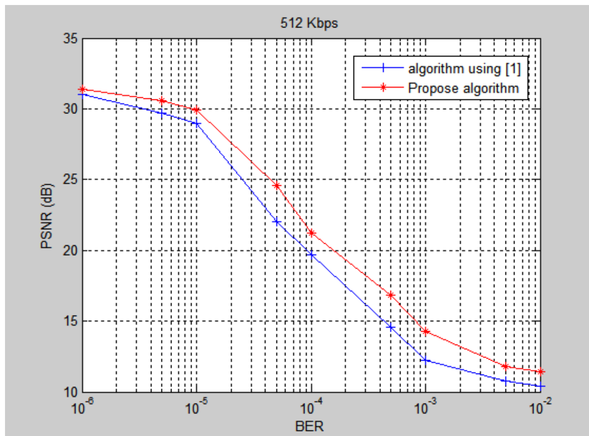

Fig. 12. Comparison between the PSNR for 105 frames of Walk and the proposed method and algorithm used in [1], with source coding rate = 512 Kbps

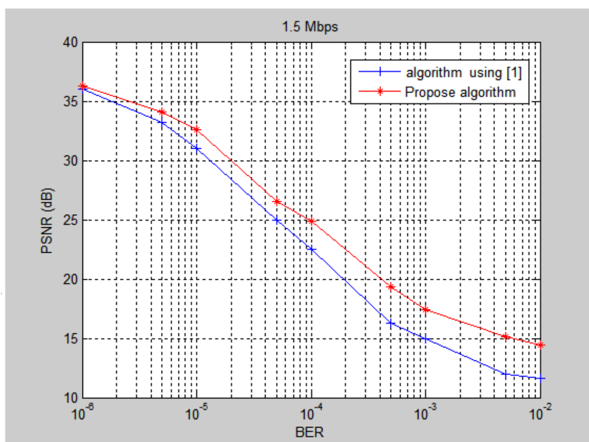

Fig. 13. Comparison between the PSNR for 105 frames of Walk and the proposed method and algorithm used in [1], with source coding rate = 1.5 Mbps

REFERENCES

[1] M. Farooq Sabir, R. W. Heath, and A. Cornard Bovik, **"Joint Source-Channel Distortion Modeling for MPEG-4 Video,"** *IEEE Trans. Image Processing.* , Vol. 18, No. 1, January. 2009.

[2] G. Cheung and A. Zakhor, **"Bit allocation for joint source/channel coding of scalable video,"** *IEEE Trans. Image Process.*, Vol. 9, No. 3, pp. 340–356, Mar. 2000.

[3] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, **"Joint source-channel coding for motion-compensated dct-based snr scalable video,"** *IEEE Trans. Image Process*., Vol. 11, No. 11, pp. 1043–1052, Sep. 2002.

[4] M. Bystrom and J. W. Modestino, **"Combined source-channel coding schemes for video transmission over an additive white Gaussian noise channel,"** *IEEE J. Sel. Areas Commun*., Vol. 18, No. 6, pp. 880–890, Jun. 2000.

[5] R. Zhang, S. Regunathan, and K. Rose, **"Video coding with optimal inter/intra-mode switching for packet loss resilience,"** *IEEE J. Sel. Areas Common*., Vol. 18, No. 6, pp. 966–976, Jun. 2000.

[6] D. Salomon, **"Data Compression,"** Third Edition, Springer, 2004.

[7] I. E. G. Richardson, **"H.264 and MPEG-4 Video Compression, Video Coding for Next-generation Multimedia,"** New York: Wiley, 2003.

[8] Y. Q. shi and H. Sun, **"Image and Video Compression for Multimedia Engineering, Fundamentals, Algorithms, and Standards,"** *ISBN 0-8493-3491-8*, 1999.

[9] N. Provos and P. Honeyman, **"Hide and Seek: An Introduction to Steganography,"** *IEEE: Security & Privacy*, Vol. 1, pp. 32-44, 2003.

[10] K. Solanki, N. Jacobsen, U. Madhow, Manjunath, B. S. Chandrasekaran, **"Robust image-adaptive data hiding based on erasure and error correction,"** *IEEE Trans. On Image Processing*, Vol. 13, No. 12, pp. 1627–1639, 2004.

[11] A. Westfeld, **"F5-A Steganographic Algorithm: High Capacity Despite Better Steganalysis,"** *Lecture Notes in Computer Science*, Vol. 2137, pp. 289-302, 2001.

[12] J. Portilla and E. P. Simoncelli, **"A parametric texture model based on joint statistics of complex wavelet coefficients"** *International Journal of Computer Vision*, Vol. 40, No. 1, pp. 49-71, 2000.

[13] J. C. Moreira and P. G. Farrell, **"Essentials of Error-Control Coding,"** *John Wiley & Sons*, Ltd, 2006.

[14] D. A. Huffman, **"A method for the construction of minimum redundancy codes,"** *Proc. IRE*, Vol. 40, pp.1098-l 101, 1952.

[15] J. G. Proakis, **"Digital Communications,"** Mcgraw Hill, Hardcover, 1995.

[16] Shannon, C. E., **"Communications in the presence of noise,"** *Proc. IEEE*, Vol. 86, No. 2pp. 447–458, February 1998.

[17] Carlson, B., **"Communication Systems. An Introduction to Signals and Noise in Electrical Communication,"** 3rd Edition, McGraw-Hill, New York, 1986.

[18] C. E. Shannon, **"A mathematical theory of communication,"** *Bell Syst. Tech. J*., Vol. 27, pp. 379-423, 623-656, July and October 1948.

[19] Video Test Media [derfs collection]; in https://media.xiph.org/video/derf/.