Joint Source and Channel Analysis for Scalable Video Coding Using Vector Quantization over OFDM System

Farid Jafarian
Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.
farid.jafarian@birjand.ac.ir
Hassan Farsi*
Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran
hfarsi@birjand.ac.ir

Received: 08/Sep/2013 Revised: 08/Dec/2013 Accepted: 17/Feb/2014

Abstract
Conventional wireless video encoders employ variable-length entropy encoding and predictive coding to achieve high compression ratio but these techniques render the extremely sensitive encoded bit-stream to channel errors. To prevent error propagation, it is necessary to employ various additional error correction techniques. In contrast, alternative technique, vector quantization (VQ), which doesn't use variable-length entropy encoding, have the ability to impede such an error through the use of fix-length code-words. In this paper, we address the problem of analysis of joint source and channel for VQ based scalable video coding (VQ-SVC). We introduce intra-mode VQ-SVC and VQ-3D-DCT SVC, which offer similar compression performance to intra-mode H.264 and 3D-DCT respectively, while offering inherent error resilience. In intra-mode VQ-SVC, 2D-DCT and in VQ-3D-DCT SVC, 3D-DCT is applied on video frames to exploit DCT coefficients then VQ is employed to prepare the codebook of DCT coefficients. In this low bitrate video codecs, high level robustness is needed against the wireless channel fluctuations. To achieve such robustness, we propose and calculate optimal codebook of VQ-SVC and optimal channel code rate using joint source and channel coding (JSCC) technique. Next, the analysis is developed for transmission of video using an OFDM system over multipath Rayleigh fading and AWGN channel. Finally, we report the performance of these schemes to minimize end-to-end distortion over the wireless channel.

Keywords: Vector Quantization; Joint Source and Channel Coding; Video Coding; Wireless Channel.

1. Introduction

Source coding is an inseparable component in a video transmission system and is required to ensure manageable transmitted bitrates [1]. H.264/AVC is one of the most popular codecs in use today, and leans on de-correlating transforms, inter/intra prediction, and variable-length entropy coding. Another codec is 3D-DCT video encoder. Although H.264/AVC provides higher coding efficiency than 3D-DCT but 3D-DCT encoder has several advantages compared to H.264/AVC, such as reduced number of operations per pixel because of no motion estimation/compensation required and symmetric encoding-decoding [2].

With increasing interest on real-time video communications over wireless channels, the complex and challenging problem rise up due to the multi-path fading characteristics of the channel [3]. To prevent such error propagation, it is necessary to employ various additional error resilience and error correction techniques [1]. The VQ based encoders/decoders are an alternative which do not use variable-length entropy encoding and have the ability to prevent such an error, also it can be employed to very low bitrate video applications due to its compression performance rather than inter/intra predictive codecs [4]-[6].

The concept of Forward Error Correction (FEC) is to add redundancy bits to the block of video data for the purpose of error detection and correction. In common communication systems, source and channel coding have traditionally been designed independently of each other. This is justified by Shannon's principle which says that there is no performance loss if the two functions are separately treated [7]. However Shannon's theory is an asymptotic result that permits unlimited delay and complexity. In fact this principle relies on the crucial assumption that the source and channel codes can be of arbitrary long lengths. In practice, because of limitations on the computational power and processing delays this assumption does not hold. It is then of benefit to consider the problem of source and channel coding jointly. A review of JSCC can be found in [7]-[9].

In some streaming video applications like Unmanned Ground Vehicle (UGV) and Ground Control Station (GCS) communication scenario [10], due to power restriction, it is required to provide very low bitrate and a robust system to wireless video transmission. To this end, we introduce intra-mode VQ-SVC and VQ-3D-DCT SVC. In the intra-mode VQ-SVC encoder, 2D-DCT and in the VQ-3D-DCT SVC encoder, 3D-DCT is applied on video frames to exploit DCTs coefficients. Then VQ is employed to prepare the codebook of DCTs coefficients.

* Corresponding Author

In this very low bitrate video codecs, high level robustness is needed against the wireless channel fluctuations. To achieve such robustness, we propose and calculate optimal codebook of VQ-SVC and optimal channel code rate (CCR) using JSCC technique. Next, the analysis is developed for transmission of video using an OFDM system over multipath Rayleigh fading and AWGN channel.

In section 2, a brief of scalable video coding and VQ is introduced. In Section 3, the optimal codebook of VQ-SVC is presented. Joint source and channel rate optimization, the rate allocation algorithm, is presented in Section 4. Section 5 details simulation setting and in Section 6 simulation results are presented.

2. Summary of scalable video coding and vector quantization

2.1 Scalable Video Coding (SVC)

SVC selects a layered coding scheme to generate multi-layered bit-stream for heterogeneous environments. Scalable video coding is used for various transmission mediums which have various constraints. As today's wireless networks are not only widespread but they also represent a substantial part of larger heterogeneous networks, the support for low-bandwidth and robust transmission is highly desirable. As a general concept, a SVC stream has a base layer and several enhancement layers. As long as the base layer is received, the receiver can decode the video stream. As more enhancement layers are received, the decoded video quality is improved [11].

The SVC design which is an extension of the H.264/AVC [12-13] video coding standard can be classified as a layered video codec. SVC-based layered video coding is suitable for supporting heterogeneous devices with a scalable bit-stream. Such a stream allows for delivering a presentable quality (presentable quality refers to resolution, frame rate and bit rate of a decoded operation point of the scalable video bit stream) of the video depending on the device's capabilities. Paper [14] proposes two SVC structures for high efficiency video coding: one based on multi-loop decoding and the other based on single-loop decoding with High Efficiency Video Coding (HEVC). A detailed overview of SVC can be found in [15]. In [16], Scalable video streaming over fading wireless channels is presented. The authors use PSNR (Peak Signal to Noise Ratio) as a measure of video quality and develop a model to characterize the relationship between the rate and PSNR. The Discrete Cosine Transforms (DCT) based SVC is presented in [17]. The presented approach is applicable to different SNR scalable encoders.

DCT has a strong "energy compaction" feature which means the most signal information tends to be concentrated in a few low-frequency components of the DCT. By increasing the type of DCT, the signal information will be concentrated in fewer low-frequency components of the DCT, which results in more compression. The term "scalability" in this article refers to use DCTs (types: I, II, III, IV…) ([18-19]) for desired level of video compression and consequently accessing to quality and bit rate scalability. It should be mentioned that we use DCT transform in order to maintain acceptable visual quality of the very low bit-rate video transmission.

We use the 3D-DCT SVC in our analysis for a number of reasons. First, the generated embedded bit-stream is highly scalable. Depending on the estimated channel condition, the bit allocation scheme can easily extract varying subsets of the bit-stream represented the video of varying quality. Second, because the codec is DCT based, it can provide low bit rate video stream.

2.2 Vector quantization (VQ)

VQ has been widely used in very low bit rate video transmission because of its simple decoder structure and better rate distortion performance over traditional scalar quantization scheme such as Differential Pulse Code Modulation (DPCM) transform coding [20]. When a set of values is quantized jointly as a single vector, the process is known as VQ. Instead of transmitting a given data vector, a symbol which indicates the appropriate reproduction vector is used. This can result in considerable savings in transmission bandwidth, despite the expense of some distortion. A vector quantizer Q of dimension K and size N is a mapping from a point in k-dimensional Euclidean space into a finite set C containing output or reproduction points that exist in the same Euclidean space as the original point. These reproduction points are known as code-words and these set of code words are called a codebook C with N distinct code-words in the set [21]. Mapping function Q is defined as (1):

$$Q: R^k \rightarrow C \tag{1}$$

A codebook containing several code-vectors is used to encode the input vectors. The code-word closest to an input vector, i.e. one having the minimum Euclidean distance with respect to the input vector, is used to represent that input vector. Compression is achieved by transmitting the binary address of the selected code-word in the codebook. Guobin Shen in [22] proposed adaptive VQ with codebook updating based on locality and history for images. Its performance can be improved by exploiting inter index redundancies.

2.3 Intra-mode SVC and 3D-DCT SVC

The DCT has been accepted as an essential part of well-known transform block-based video compression standards, such as JPEG, MPEG1-2-4, and ITU-T H.263. Each basis vector in the DCT domain represents a spatial frequency component of the video frame. As mentioned before, the term *scalability* in this paper refers to quality and rate scalability features using DCT types for required level of video compression. The structure of intra-mode SVC is similar to intra-H.264/AVC with this difference that, in intra-mode SVC, DCT types [18-19] are employed as de-correlating transform, e.g., type-IV DCT is applied on video frames to exploit DCT coefficients.

Employing various types of DCT allows generating scalable video stream. Similarly, the 3D-DCT SVC employs DCT types as de-correlating transform.

Suppose each color component of video frame $\{f(i,t); t = 0,1,\dots; i = Y, C_r, C_b\}$, where t refers to the frame number and i is color component. Also, suppose L level of compression is desired. Each color component of the video sequence is partitioned into $8 \times 8 \times 8$ pixel blocks that are processed in a zigzag scan order. Each block is transformed by the 3D-DCT for L level, e.g., for L=1, type-I 3D-DCT and for L=2, type-II 3D-DCT are used. For L=1, the type-I 3D-DCT is mathematically as follow:

$$F(u,v,w,i) = c(u)c(v)c(w)\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{N} f(x,y,t,i) \times cos\left[\frac{(2x+1)u\pi}{2N}\right] \times cos\left[\frac{(2y+1)v\pi}{2N}\right] \times cos\left[\frac{(2t+1)w\pi}{2N}\right] \quad (2)$$

The inverse of 3D-DCT is:

$$f(x,y,t,i) = \sum_{u=1}^{N}\sum_{v=1}^{N}\sum_{w=1}^{N} c(u)c(v)c(w) \times F(u,v,w,i) \times cos\left[\frac{(2x+1)u\pi}{2N}\right] \times cos\left[\frac{(2y+1)v\pi}{2N}\right] \times cos\left[\frac{(2t+1)w\pi}{2N}\right]$$

$$c(w) = \begin{Bmatrix} \frac{1}{\sqrt{2}} & w = 0 \\ 1, & w > 0 \end{Bmatrix} \; c(u) = \begin{Bmatrix} \frac{1}{\sqrt{2}} & u = 0 \\ 1, & u > 0 \end{Bmatrix}$$

$$c(v) = \begin{Bmatrix} \frac{1}{\sqrt{2}} & v = 0 \\ 1, & v > 0 \end{Bmatrix} \quad (3)$$

2.4 Intra-mode VQ-SVC and VQ-3D-DCT SVC

Conventional wireless video encoders such as H.264/AVC and 3D-DCT employ variable-length entropy coding and predictive coding to achieve high compression ratios. But these techniques render the extremely sensitive encoded bit-stream to channel errors. To prevent error propagation, it is necessary to employ various additional error correction techniques [23]. In contrast, alternative technique, VQ which doesn't use variable-length entropy encoding, have the ability to impede such an error [23].

To demonstrate this ability, in this paper, we apply VQ on the coefficients resulting intra-mode SVC and 3D-DCT SVC and compare the performance (PSNR v.s. channel BER) with H.264/AVC and 3D-DCT. The channel conditions and the OFDM parameters are set according to the simulation setting presented in section 5. The results of our simulation are shown in Fig (2) parts a, b, c and d. From the simulation results it can be observed that VQ based encoders outperform both of the mentioned methods which don't employ VQ.

3. Joint source and channel codebook optimization

As mentioned in introduction section, to optimize the overall performance of video transmission in communication systems, it is necessary to consider the optimization of combination of source and channel characteristics. A traditional approach in combination of source and channel consists in first matching the source codebook to the source statistics in order to minimize the distortion of the source. Then, the labeling of the source dictionary to the channel code-words is optimized in order to minimize the distortion due to the channel errors. In this section, we analyze the codebook optimization for scalable video coding.

Let's suppose that the streamed coefficients of the video-source to be encoded are real-valued stationary and Ergodic process $\{X,; t = 0,1,\dots.; x \in X,\}$ with zero mean and variance σ_x^2 (DCT coefficients have Laplacian-like probability distributions). The source is needed to be encoded by means of VQ. K-dimensional, M-level VQ and a discrete memoryless channel (DMC) is considered. With input and output alphabet ψ={1, 2, …M}, we consider the source encoder as operation S (in fact is the mapping of k-th element X^k of the m-th VQ), channel coder as operation C, channel decoder as C_{de} and source decoder as S_{de}.

The source output is $X_s = \{X_s^1, X_s^2, X_s^3, \dots\}$ which we derive our distortion model expressions for a binary symmetric channel (BSC) with a given bit error probability. Given the bit error probabilities for any channel (AWGN, Rayleigh fading, etc.) and the fact that the probability of making an error from 0 to 1 is the same as that of 1 to 0, the channel can be represented as a BSC. Therefore, the distortion model presented in this paper can be used to find the distortion curves for any channel that can be represented as a BSC, given that the source coding rate and the bit error rate are known. Our distortion model is independent of modulation type.

Let us define input/output variables and parameters of the coding system. In the system, variables X and $\hat{X}$ represent the source and reconstructed signals respectively, also X_s and $\hat{X}_s$ represent the source encoder output and the source decoder input, X_c and $\hat{X}_c$ the channel encoder output and the channel decoder input respectively. For the source encoder, we have $s: X \to X_s^k, \varphi = \{S_1, S_2, \dots., S_M$ which $s(X) = X_s^k$ if $X \in S_i$ Also for the channel encoder we have $C: X_s^k \to X_c^k, \varsigma = \{C_1, C_2, \dots., C_M\}$ which $C(X_s^k) = X_c^k \;\; if\; X_s^k \in C_i$

At the decoder side, for the channel decoder we have $C_{de}: \psi \to \hat{X}_c^k, C_d = \{C_{d,1}, C_{d,2}, \dots, C_{d,M}\}$, which $C_{de}(\hat{X}_c^k) = \hat{X}_c^k \;\; if\; \hat{X}_c^k \in C_{d,i}$ and also for the source decoder we have $S_{de}: \hat{X}_s^k \to \hat{X}^k, S_d = \{S_{d,1}, S_{d,2}, \dots, S_{d,M}\}$, which $S_{de}(\hat{X}_c^k) = \hat{X}^k$ if $\hat{X}_s^k \in S_{d,i}$.

Let's suppose that the codebook (reproduction alphabet) be C_{book}={$C_{b,1}$, $C_{b,2}$, …., $C_{b,j}$, …, $C_{b,M}$} and the source encoder is described in term of a partition φ={S_1, S_2, …., S_M}. The total average distortion of the coding system is denoted by $D_{s+c}(\varphi, C_{book}, C)$ for L scalable layers. For a layer l∈L we have [24]:

$$D_{s+c,l}(\varphi, C_{book}, C) = \frac{1}{k}\sum_{i=1}^{M}\sum_{j=1}^{M} P\left(\hat{X}_{c,j}^k | X_{c,i}^k\right) \int P_l(x) d_l(x,\hat{x}) dx \quad (4)$$

We assume that the distortion caused by representing the source vector X by a reproduction vector (also called a code-vector) $\hat{X}$ is given by a non-negative distortion measure $d_l(x, \hat{x})$ and $P_l(x)$ is the k-fold probability density function [24], and the channel is DMC with probability of output $\hat{X}_{c,j}^k$ condition to given input $X_{c,i}^k, p(\hat{X}_{c,j}^k \mid X_{c,i}^k)$.

In paper [24], the optimum codebook is calculated for a non-scalable source. While in our paper we calculate the optimum codebook for a scalable source. For the L scalable layers the equation (4) becomes:

$$D_{s+c,l}(\varphi, C_{book}, C) = \frac{1}{k}\sum_{l=1}^{L}\sum_{i=1}^{M}\sum_{j=1}^{M} P(\hat{X}_{c,j}^k|X_{c,i}^k)\int P_l(x)d_l(x,\hat{x})dx \quad (5)$$

It is clear that for a fixed C_i the problem of minimizing the total average distortion is identical to the VQ design problem with a modified distortion measure [24]. Therefore for a fixed C_i and a fixed C_{book} the optimum partition $\varphi^* = \{S_1^*, S_2^*, \dots, S_M^*\}$ is such that:

$$S_i^* = \{x: \sum_{l=1}^{L}\sum_{j=1}^{M} p(\hat{X}_{c,j}^k \mid X_{c,i}^k)d_l(x,\hat{x}) \le \sum_{l=1}^{L}\sum_{j=1}^{M} p(\hat{X}_{c,j}^k \mid X_{c,h}^k)d_l(x,\hat{x}), \forall h \ne i\} \quad i \in \psi \qquad (6)$$

Where $d_l(x,\hat{x}) = \|x-\hat{x}\|^2$ is the squared error distortion. Similarly, to find the optimum codebook, the total average distortion should be determined for a fixed C_i and a fixed φ, therefore the equation (5) becomes to:

$$D_{s+c,l}(\varphi, C_{book}, C) = \frac{1}{k}\sum_{l=1}^{L}\sum_{j=1}^{M} P(\hat{X}_{c,j}^k|X_{c,i}^k)\int P_l(x)d_l(x,\hat{x})dx \quad (7)$$

The optimum codebook, $C_{book,j}^* = \{C_{b,1}^*, C_{b,2}^*, \dots, C_{b,j}^*, \dots C_{b,M}^*\}$, must satisfy $C_{book,j}^*$=arg min E$(d(x,\hat{x}) \mid V = \hat{X}_{c,j}^k), j \in \psi$ where $\hat{X}_j^*$ is equal to $C_{b,j}^*$, and V is used to denote the random variable at the channel output [24].

Let's attempt to solve the unconstrained minimization problem. The best set is determined by setting the partial derivatives of the equation (7) with respect to the $\hat{x}$'s equal to zero, i.e.:

$$\frac{\partial D_{s+c}(\varphi, C_{book}, C)}{\partial \hat{x}} = -\frac{2}{k}\sum_{l=1}^{L}\sum_{j=1}^{M} p(\hat{X}_{c,j}^k|X_{c,i}^k)\int(x-\hat{x})p_l(x)dx = 0 \quad (8)$$

Which, in turn, it implies that

$$\hat{x}_j^* = \frac{\sum_{l=1}^{L}\sum_{j=1}^{M} p(\hat{X}_{c,j}^k|X_{c,i}^k)\int x p_l(x)dx}{\sum_{l=1}^{L}\sum_{j=1}^{M} p(\hat{X}_{c,j}^k|X_{c,i}^k)\int p_l(x)dx}, j \in \psi \qquad (9)$$

Therefore the optimum codebook can be obtained by (9). We generally refer to the algorithm channel-optimized algorithm VQ (COVQ) which is detailed in [24] for optimizing the codebook.

4. Joint source and channel rate optimization

In this section, we focus on the application of JSCC in video communications, to find an optimal bit allocation between source coding and channel coding based on the Shannon's separation theory [8], [25] and [26].

4.1 Operational rate distortion (ORD) theory

Every lossy data compression scheme and every channel coding technique have only a finite set of admissible mode, suppose for the source mode we have M admissible modes $s = \{s_1, s_2, \dots, s_M\} \in S^M$ and for the channel coding modes, $c = \{c_1, c_2, \dots, c_M\} \in C^N$ Therefore there is only a finite number of possible rate distortion pairs for any given source and channel coding technique, so ORD is based on this fact [27]. For any given source mode and channel coding technique mode and given channel state information (CSI), we will have a set of Rate-Distortion points ($^{M\times N}$ points). For each modulation scheme, probability of error of the channel can be estimated using a given CSI [28-29]. Using this estimated probability of error, experimentally ORD points are calculated [8].

4.2 Optimal bit allocation between source and channel coding using Lagrange multiplier

To solve a problem there are four distinctive steps. First, an appropriate system performance evaluation metric should be selected. Second, the constraints are needed to be specified. Third, a model of the relationship between the system performance metric and the set of adaptation parameters are needed to be established. Finally, the best combination of adaptation parameters that maximize the system performance while meeting the required constraints needs to be identified. Therefor we present the formal approach to formulate the joint source and channel coding problem and provide solution approaches to such a problem. Let S, be the set of source coding parameters and C the set of the channel coding parameter. The general formulation of the optimal bit allocation problem is to minimize the total expected distortion, i.e., provide the best video delivery quality, for the frame(s), given the corresponding bit rate constraint.

$$\min_{\substack{s\in S^M,\ c\in C^M \\ s.t.\ R(s,c)\le R_0}} E[D(s,c)]\ Min\ E[D(s,c)] \qquad (10)$$

Where $E[D(s,c)]$is total expected distortion, $R(s,c)$ the total number of bits used for both source and channel coding, and R_0 is the bit rate constraint for the frame(s). For solving constrained optimization problem Equation (11), which is a convex problem the Lagrange multiplier method is suitable. Lagrange method introduces a variable which is called a Lagrange multiplier that controls the weights of the constraint when added as a penalty to the objective function. In this method constrained problem is converted into an unconstrained problem [8], as:

$$\{S^*(\lambda), C^*(\lambda)\} = arg\ min\{E[D(s,c)] + \lambda R(s,c)\} \quad s \in S^M,\ c \in C^M \qquad (11)$$

Where λ≥ 0 is the Lagrange multiplier. The solution of Equation (10) can be obtained by solving Equation (11) with the appropriate choice of the Lagrange multiplier while bitrate constraint $R(S^*(\lambda), C^*(\lambda)) \le R_0$ is satisfied, [8]. In practice, due to the finite set of source and channel coding parameters, the objective function and the constraints are not continuous thus the constraint may not be met with equality. In this case, the solution obtained by solving Equation (11) will be the convex hull approximation solution to Equation (10). Let U denote the coding parameter set $S \times C$ and let $u \in U$ be an element of this set, a coding parameter including both source and channel coding parameters. Assuming that is $u^*(\lambda)$ an optimal solution to Equation (11), so we have:

$$D\big(u^*(\lambda)\big) + \lambda R\big(u^*(\lambda)\big) \leq D\big(u(\lambda)\big) + \lambda R\big(u(\lambda)\big), \forall u \in U \quad (12)$$

which can be rewritten as:

$$R(u) \geq -\frac{1}{\lambda} D(u) + \left\{R\big(u^*(\lambda)\big) + \frac{1}{\lambda} D\big(u^*(\lambda)\big)\right\} \quad (13)$$

This means that all rate-distortion pairs $\{R(u), D(u)\}$ must lie on the upper right side in the rate-distortion plane of the line defined by Equation (13) using an equal sign. $\{R\big(u^*(\lambda)\big), D\big(u^*(\lambda)\big)\}$ is on the line and the line has slope $-1/\lambda$. For a different value of λ, the line with slope of $-1/\lambda$ will meet at least one optimal solution $\{R\big(u^*(\lambda)\big), D\big(u^*(\lambda)\big)\}$.

In this paper we deal with a set of ORD points for each layer which is calculated through the experiment. The distortion for layer l is expressed as $D_{s+c,l}(R_{s+c,l})$ and the overall distortion as:

$$D_{s+c} = \sum_{l=1}^{L} D_{s+c,l}(R_{s+c,l}) \quad (14)$$

R_{s+c} is the bit rate used for source and channel coding for the scalable layer l, $R_{s+c,l}$ is equal to:

$$R_{s+c,l} = \frac{R_{s,l}}{R_{c,l}} \quad and \quad R_{s+c} = \sum_{l=1}^{L} R_{s+c,l} \quad (15)$$

Where $R_{s,l}$ and $R_{c,l}$ are the source and channel rates, respectively, for the scalable layer l. For L scalable layers, the total transmitted bitrate is equal to R_{s+c}. Now we expand discussion for the video, as mentioned the optimization problem in video transmission is:

$$\begin{aligned} &Min\, D_{s+c}(max\ \ PSNR_{s+c}) \\ &s.t. R_{s+c} \leq R_{budget} \end{aligned} \quad (16)$$

The solution, $\{D^*_{s+c}, R^*_{s+c}\}$ to the constrained optimization problem Equation (10) is obtained by converting it into an unconstrained problem using Lagrangian optimization function:

$$\begin{aligned} &J_l = D_{s+c}(R_{s+c,l}) + \lambda R_{s+c,l} \\ &l = 1,2,\ldots\ldots,L \end{aligned} \quad (17)$$

The unconstrained problem can now be solved using these Lagrangian cost functions, with the optimal solution being the argument of the following unconstrained minimization problem:

$$\min_{R_{s+c,l}(\lambda), D_{s+c,l}(\lambda)} (J_1 + J_2 + \cdots J_L) \quad (18)$$

To solve this problem, universal rate-distortion characteristic (URDC) is proposed [17]. In this method, universal rate-distortion curves are established using simulations at different source coding rates. These curves are used to establish operational rate distortion curves, which are then used to specify the value of distortion at different source coding and channel bit error rates. Using these operational rate-distortion curves, the optimal bit allocation and the minimum distortion are obtained, as follows:

Given a set of parameters for the channel (e.g., signal-to-noise ratio), channel coding, and the particular modulation scheme, the probability of bit error, P_b, are calculated for the set of channel coding rates, R_c, of interest. This can be performed using simulations or by theoretical means. It constructs a reference as a performance of the channel coding over the especial channel with the given parameters. Moreover, this channel performance analysis needs to be performed offline and only once [17].

Toward calculating the impact of the errors due to both source coding and channel transmission on a set of data, it is perceived that for a given set of layer source rates, the distortion for a particular layer, $D_{s+c,l}$, given a particular source coding rate, $R_{s,l}$, is a function of the bit error rate. Thus, the rate-distortion function of the layer for a fixed source rate, $R_{s,l}$, is a function of the bit error rate (after channel decoding), P_b. It is then possible to plot a family of $D_{s+c,l}$ versus 1/ P_b curves given a set of source coding rates of interest, so these are defined as the URDCs of the source [17].

5. Simulation setting

In this study, pixel Common Inter-mediate Format (CIF) resolution video sequences Football and Coastguard video (at 15 and 30 frames per second, respectively) [32] are employed as video sources. Next, the 3D-DCT is applied on $8\times8\times8$ blocks of video, and intra mode H.264/AVC is applied on frames of videos. Similarly, intra-mode VQ-SVC and VQ-3D-DCT SVC are applied on the video streams and based on the prepared optimum codebook calculated by equation (9) the vectors of data are converted to the symbols which indicate the reproduction vectors. Next, the symbols are coded using Rate-Compatible Punctured Convolutional (RCPC) codes [30]. Punctured convolutional codes are convolutional codes obtained by puncturing some outputs of the convolutional encoder.

Since our focus is on JSCC, we assume perfect channel estimation of each sub-carrier gain and perfect suppression of multipath by the guard interval. In this simulation we use the general OFDM system with FFT size 256, occupied subcarriers 151, Ratio of Cyclic prefix time (guard interval) set as 1/8, and modulation type set as QPSK. The parameters of the simulated channel are presented in Table (1). These parameters are according to the ITU Channel Model for Vehicular Test Environment. Motion causes Doppler shift in the received signal components, the Doppler frequency f_D equals f_0v/c in which v is the mobile speed, f_c is the carrier frequency and c is the light speed [31]. In our paper carrier frequency is equal f_c=1.9 GHz and the maximum Doppler frequency is equal f_D=90 Hz.

6. Simulation results

In this section, at first, the compression performance of the proposed intra-mode VQ-SVC and VQ-3D-DCT SVC are compared with that of intra-mode H.264/AVC and 3D-DCT. It is important to note that the results are not presented with a view to establishing which H.264/AVC provides better source compression. Rather, they are intended to show that the proposed VQ based codecs provide compression performance comparable to these other related codecs.

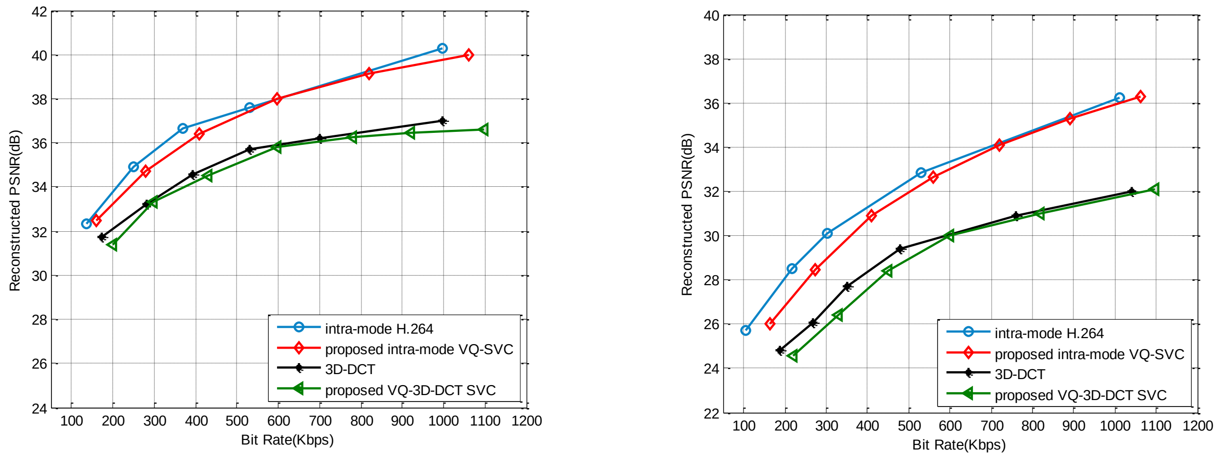
Fig. 1 illustrates the RD curves for two sequences. In Fig. 1(a), for Coastguard, it is observed that proposed intra-mode VQ-SVC has an RD performance similar to H.264/AVC. It is also obvious that at 540-820 Kbps, the proposed intra-mode VQ-SVC and H.264/AVC provide the same PSNR performance. Obviously, at 590-800 Kbps, VQ-3D-DCT SVC and 3D-DCT provide the same PSNR performance. As shown in Fig. 1(b), for Football, the proposed intra-mode VQ-SVC has an RD performance similar to H.264/AVC. In this figure, it can be observed that at 675-930 Kbps, intra-mode VQ-SVC and H.264/AVC result in the same PSNR performance. It is also clear that at 580-1050 Kbps, VQ-3D-DCT SVC and 3D-DCT provide the same PSNR. In Fig. 1(a) and (b), it is also obvious that intra-mode VQ-SVC and H.264/AVC outperform VQ-3D-DCT SVC and 3D-DCT.

As described in section 4, there is an optimal bit allocation in a communication scenario which is defined by the JSCC. As an example, Fig. 3 is presented to show the impact of the optimal bit allocation on the quality of video received by the end user. In the example, Football video is encoded by intra mode VQ-SVC and transmitted over wireless channel (at SNR=15, Mobile speed=50 km/h) in 5 different predefined modes; Mode-1{source code rate (SCR)=380 Kbps, channel code rate (CCR)=0.33}, Mode-2 {SCR=617 Kbps, CCR=0.5}, Mode-3 {SCR=807 Kbps, CCR=0.66}, Mode-4 {SCR=950 Kbps, CCR=0.75} and Mode-5 {SCR=998 Kbps, CCR=0.8}. From the Fig. 3 it can be observed that the received frame-74 in part *d* resulted by setting the Mode-3 as the bit allocation mode, have the highest quality.

Similarly, two simulations are performed for the Football video frames 1-100. The first simulation is performed in the different transmission environments SNR {SNR = 5, 10, and 15 dB} and the second, in the different mobile speeds {V = 30, 50 and 90 km/h}. Related results are presented in Table (2) and Table (3) respectively. The results obtained using repeated experiments and taking the average PSNR. As reported in Table (2), Average PSNR at SNR=15 at the Mode-3, compared with other modes, is maximum (PSNR=24.47 dB). It can be also observed that the mode is changed by reducing the SNR. For instance, at SNR=10 the optimum SCR is equal 617 Kbps and CCR is equal 0.5. Also, at SNR=5 the optimum SCR is equal 380 Kbps and CCR is equal 0.33. In the second simulation, quality of video received by the end user is evaluated at different mobile speeds at SNR=15 dB. As reported in Table (3), Average PSNR at mobile speed (V) 90 km/h at the Mode-1, compared with other modes, is maximum (PSNR=19.39 dB). It can be also observed that the mode is changed by reducing the mobile speed. For instance, at V=50, the optimum SCR is equal 617 Kbps and CCR is equal 0.5 and also at V=30, the optimum SCR is equal 807 Kbps and CCR is equal 0.66 and PSNR is equal to 27.76 dB.

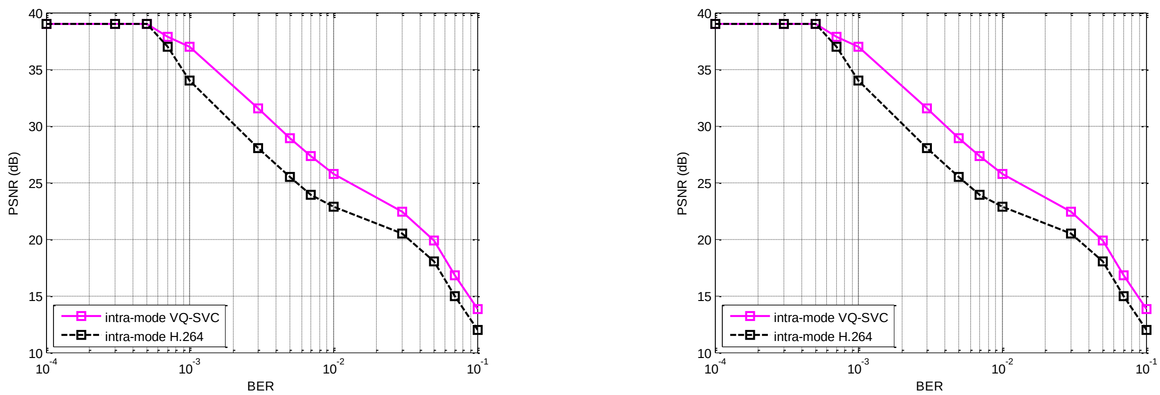
Next, to evaluate the performance of the proposed codecs and schemes over wireless channel, each video stream is compressed at the same bitrate and all utilize the same channel coding rate, i.e. 3/4 rate FEC. As observed in Fig.1, in 800 Kbps, for both sequences, the proposed intra-mode VQ-SVC and H.264/AVC provide the same PSNR (for Coastguard PSNR=39 dB and for Football PSNR=35 dB) and also VQ-3D-DCT SVC and 3D-DCT result in the same PSNR (for Coastguard PSNR=36.3 dB and for Football PSNR=31 dB). Therefore, in this section, Coastguard and Football are encoded at 800 Kbps.

In Fig. 2, parts *a* and *b*, the intra mode VQ-SVC is compared with the intra mode H.264/AVC over Coastguard and Football video respectively, and in Fig. 2, parts *c* and *d*, the VQ-3D-DCT SVC is compared with the 3D-DCT over Coastguard and Football video respectively. The simulation results show that VQ based encoders outperforms both of the mentioned methods which don't employ VQ. For example at 10^{-3}-10^{-2} BER range, intra mode VQ-SVC offer about 3.5 dB PSNR gain for Coastguard and 3.8 dB PSNR gain for Football compared to the best performing intra mode H.264/AVC stream. As well as, at 10^{-3}-10^{-2} BER range the VQ-3D-DCT SVC offer about 2.7 dB PSNR gain for Coastguard and 3 dB PSNR gain for Football compared to the best performing 3D-DCT stream.

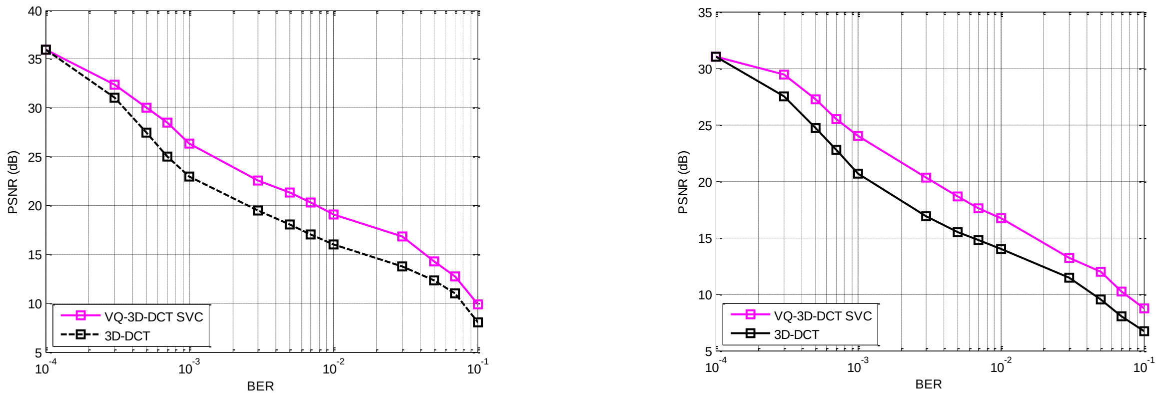


(a) (b)

Fig. 1. RD comparisons; intra-mode H.264, proposed intra-mode VQ-SVC, 3D-DCT and proposed VQ-3D-DCT SVC. (a) Coastguard sequence. (b) Football sequence.



(a) (b)



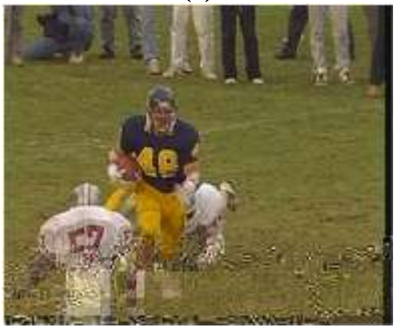
(c) (d)

Fig. 2. Simulation results of the proposed scheme, intra mode VQ-SVC, compared to intra mode H.264: (a) for Coastguard video (b) for Football video. Results of the proposed scheme, 3D-DCT VQ-SVC, compared to 3D-DCT: (c) for Coastguard video (d) for Football video.


(a)

(b)

(c)

(d)

(e)

(f)

Fig. 3. Simulation Results for channel condition including: SNR=15, Mobile speed= 50 km/h: (a) is the original frame of the Football video, frame number 74, (b) is the received frame in Mode 1, (c) is the received frame in Mode 2, (d) is the received frame in Mode 3, (e) is the received frame in Mode 4, (f) is the received frame in Mode 5.

Table 1. Channel Parameters and channel coding setting

Parameters	**Values**
Channel model	ITU channel model for vehicular test environment
Relative Tap delay (ns)	0, 310, 710, 1090, 1730 and 2510
Average power (dB)	0.0, -1.0, -9.0 -10.0, -15.0 and -20.0
Doppler frequency (Hz)	90
Channel coding	RCPC [a]; Rates: 1/3, 1/2, 2/3 and ¾

[a] RCPC: Rate-Compatible Punctured Convolutional

Table 2. JSCC simulation results of Football video frames 1-100 for different SNR = 15, 10 and 5 dB

Mode / Result	Mode 1: SCR [a] = 380 Kbps CCR [b] = 0.33	Mode 2: SCR = 617 Kbps CCR = 0.5	Mode 3: SCR = 807 Kbps CCR = 0.66	Mode 4: SCR = 950 Kbps CCR = 0.75	Mode 5: SCR = 998 Kbps CCR = 0.8
Average PSNR(dB) at SNR=15	20.39	21.58	24.47	20.81	20.15
Average PSNR(dB) at SNR=10	18.68	20.12	19.72	18.03	17.70
Average PSNR(dB) at SNR=5	17.11	15.89	14.08	12.75	12.03

[a] SCR: source code rate, b CCR: channel code rate

Table 3. JSCC simulation results of Football video frames 1-100 for different mobile speed (v): 90, 50 and 30 km/h

Mode / Result	Mode 1: SCR [a] = 380 Kbps CCR [b] = 0.33	Mode 2: SCR = 617 Kbps CCR = 0.5	Mode 3: SCR = 807 Kbps CCR = 0.66	Mode 4: SCR = 950 Kbps CCR = 0.75	Mode 5: SCR = 998 Kbps CCR = 0.8
Average PSNR(dB) at v=90	19.39	17.97	16.38	14.56	14.01
Average PSNR(dB) at v=50	22.09	25.88	23.77	21.81	20.78
Average PSNR(dB) at v=30	23.87	26.20	27.76	26.01	24.43

[a] SCR: source code rate, b CCR: channel code rate

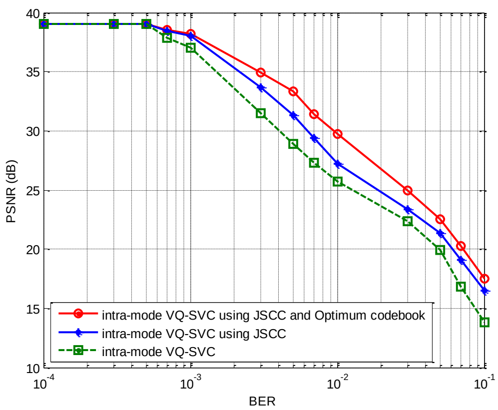
Next, we compare the performance of three proposed schemes; intra mode VQ-SVC with intra mode VQ-SVC using JSCC and with intra mode VQ-SVC using both JSCC and optimum codebook. We evaluate these schemes over Coastguard and Football video. The results are shown in Fig. 4, parts *a* and *b*. From our simulation it can be observed that intra mode VQ-SVC using both JSCC and optimum codebook outperforms the other two schemes; the intra mode VQ-SVC and the intra mode VQ-SVC using JSCC. For example as shown in Fig. 4 part *a*, at 10^{-3}-10^{-1} BER range, intra mode VQ-SVC using both JSCC and optimum codebook offers about 2.5-5dB PSNR gain for Coastguard, also as shown in Fig. 4 part *b*, it offers about 3-5.5 dB PSNR gain for the Football video compared to the intra mode VQ-SVC. Next, we compare the performance of three schemes; 3D-DCT VQ-SVC with 3D-DCT VQ-SVC using JSCC and with 3D-DCT VQ-SVC using both JSCC and optimum codebook. The results are shown in Fig. 4 *c* and *d*. From the simulation it can be observed that 3D-DCT VQ-SVC using both JSCC and optimum codebook outperforms the other two schemes; the 3D-DCT VQ-SVC and the 3D-DCT VQ-SVC using JSCC. For example, as shown in Fig. 4 part *c*, at 10^{-3}-10^{-1} BER range 3D-DCT VQ-SVC using both JSCC and optimum codebook offers about 3 dB PSNR gain for Coastguard and about 2-5 dB PSNR gain for Football, as shown in Fig. 4 part *d*, compared to the 3D-DCT VQ-SVC.

7. Conclusions

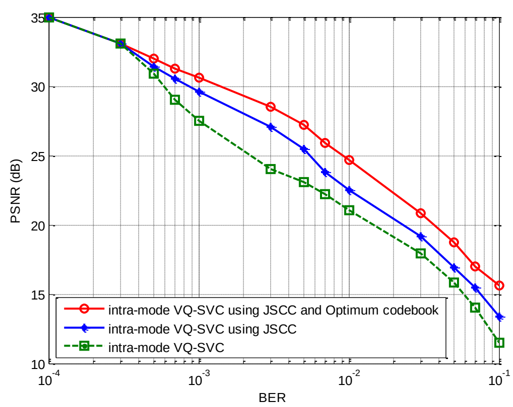
In some wireless streaming video applications, like UGV and GCS system, it is required to provide very low bit rate SVC. We demonstrate that intra-mode VQ-SVC and VQ-3D-DCT SVC outperform intra-mode H.264 and 3D-DCT SVC, respectively. To provide high quality and robust wireless video communication, it is required to employ JSCC technique. We use this technique in two objectives, first to find the optimum codebook for VQ and next to find the optimum channel and source code rates. Since the optimum codebook is not yet calculated for VQ-SVC, in this paper we analyze and apply it on the VQ-SVC. Then we develop the analysis for video transmission over multipath Rayleigh fading channel. The simulation results show that employing both optimal codebook and JSCC for VQ-SVC leads to higher-quality video delivery.

Acknowledgment

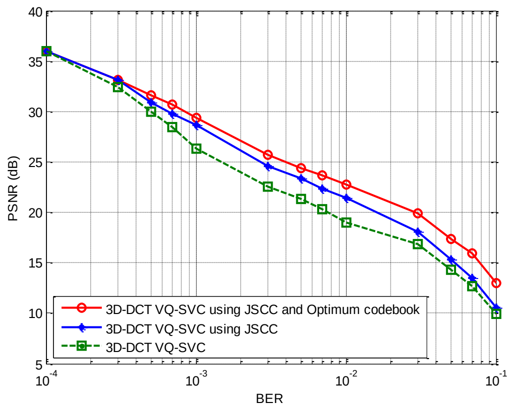
This work has been supported by the Research Institute for Information and Communication Technology (ITRC), Iran under Grant 19262/500/T



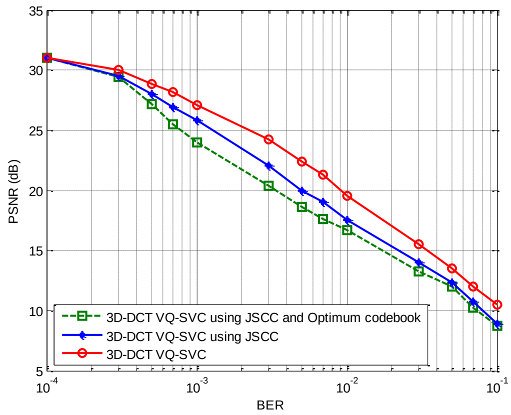
(a)



(b)



(c)



(d)

Fig. 4. Simulation Results of three proposed schemes; intra mode VQ-SVC; intra mode VQ-SVC using JSCC; intra mode VQ-SVC using both JSCC and optimum codebook: (a) for Coastguard video (b) for Football video. Results of three proposed schemes 3D-DCT VQ-SVC; 3D-DCT VQ-SVC using JSCC; 3D-DCT VQ-SVC using both JSCC and optimum codebook: (c) for Coastguard (d) for Football video.

References

[1] H.Wang, L.Kondi, A. Luthra and S.Ci, 4G wireless video communications, New York: Wiley 2009.

[2] S.Saponara, "Real-time and low-power processing of 3D direct/inverse discrete cosine transform for low-complexity video codec", Journal of Real-Time Image Processing Vol.7, No.1, pp.43-53, 2012.

[3] E.Dahlman, S.Parkvall, J.Skold, 4G: LTE/LTE-advanced for mobile broadband, Academic Press, 2011.

[4] A.S.Akbari, P.Bagheri Zadeh, T.Buggy, J.Soraghan, "Multiresolution perceptual and vector quantization based video codec", Multimedia Tools and Applications Vol.58, No.3, pp.569-583, 2012.

[5] R. Krishnamoorthy and R. Punidha, “Low bit-rate multi stage vector quantization based on energy clustered training set”, Multimeda Tools and Applications, pp.1-16, 2012.

[6] N.Adami, A.Boschetti, R.Leonardi and P. Migliorati “Embedded indexing in scalable video coding”, Multimeda Tools and Applications, pp.105-121, 2010.

[7] A. E.Gamal, Y. H. Kim, Network information theory, Cambridge University Press, pp.336-360, 2012.

[8] F.Zhai, A.Katsaggelos, "Joint source-channel video transmission", Synthesis Lectures on Image, Video, and Multimedia Processing, Vol.3, No. 1, pp.1-136, 2007.

[9] D.Ersson, J.Kron, M.Skoglund, E.G.Larsson, “Joint source-channel coding for the MIMO broadcast channel” IEEE Trans. on Signal Processing, Vol. 60, No.4, pp.2085-2090, 2012.

[10] C.Blumm, C.Heller, R.Weigel, “SDR OFDM waveform design for a UGV/UAV communication scenario” Journal of Signal Processing Systems, pp.11-21, 2012.

[11] H.Sun, A.Vetro, J.Xin, "An overview of scalable video streaming", Wireless Communications and Mobile Computing Vol.7, No.2, pp.172-159, 2007.

[12] ITU-T Recommendation, H.264 advanced video coding for generic audiovisual services, ISO/IEC 14496, 2003.

[13] T.Wiegand, G.J.Sullivan, G.Bjontegaard, A. Luthra, “Overview of the H.264/AVC video coding standard”, IEEE Trans. Circuits Syst. Video Technol. , vol. 13, no. 7, pp. 560576, Jul. 2003.

[14] H.Choi, J.Nam, D.Sim, and I.V.Bajic, "Scalable video coding based on high efficiency video coding (HEVC)", IEEE Conf. Communications Computers and Signal Processing, pp. 346-351, 2011.

[15] H.Schwarz, D.Marpe, T.Wiegand, "Overview of the scalable video coding extension of the H. 264/AVC standard", IEEE Trans. Circuits and Systems for Video Technology, Vol.17, No.9, pp.1103-1120, 2007.

[16] H.Zhang, Y.Zheng, M.A.Khojastepour, S.Rangarajan, "Scalable video streaming over fading wireless channels", IEEE Conf. Wireless Communications and Networking, pp.1-6, 2009.

[17] L.P. Kondi, F.Ishtiaq, A.K.Katsaggelos, "Joint source-channel coding for motion compensated DCT based SNR scalable video", IEEE Trans. image processing, Vol.11, No.9, pp.1043-1052, 2002.

[18] [Online] Available: http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Discrete_cosine_transform

[19] F.Matteo, S.G.Johnson, "FFTW library", 2009, [Online] Available: http://www.fftw.org, a free (GPL) C library that can compute fast DCTs (types I-IV) in one or more dimensions, of arbitrary size.

[20] J.E.Fowler, K.C.Adkins, S.B.Bibyk, S.C.Ahalt, “Real-Time video compression using differential vector quantization,” IEEE trans. of circuits and systems for video technology, vol.5, No.1, pp.14-24, 1995.

[21] R. Gray, Vector quantization, IEEE ASSP Magazine, Vol.1, No.2, pp.4-29, 1984.

[22] G.Shen, B.Z.Ming, L.Liou, “Adaptive vector quantization with codebook updating based on locality and history,” IEEE Trans. on image processing, Vol. 12, No. 3, pp. 283-295, 2003.

[23] S.M.M.Bokhari, A.R.Nix, D.R.Bull, “Rate distortion optimized video transmission using pyramid vector quantization” IEEE Trans. Image Processing, Vol.21, No.8, pp.3560-3572, 2012.
[24] N.Farvardin, V.Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers", IEEE Trans. Information Theory, Vol.37, No.1, pp.155-159, 1991.
[25] D.Persson, J.Kron, M.Skoglund, E.G. Larsson, "Joint source-channel coding for the MIMO broadcast channel", IEEE Trans. Signal Processing, Vol.60, No.4, pp.2085-2090, 2012.
[26] C.Guillemot, P.Christ, "Joint source-channel coding as an element of a QoS framework for ‘4G’wireless multimedia", Computer Communications Vol.27, No.8, pp.762-779, 2004.
[27] M.Bystrom, J.W.Modestino, "Combined source-channel coding schemes for video transmission over an additive white Gaussian noise channel", IEEE Journal on Selected Areas in Communications, Vol.18, No.6, pp. 880–890, 2000.
[28] L.Rugini, P.Banelli, "BER of OFDM systems impaired by carrier frequency offset in multipath fading channels", IEEE Trans. Wireless Communications, Vol.4, No.5, pp.2288-2279, 2005.
[29] S.T.Chung, A.J.Goldsmith, "Degrees of freedom in adaptive modulation: a unified view", IEEE Trans. Communications, Vol.49, No.9, pp.1561-1571, 2001.
[30] J.C.Moreira, P.G.Farrell, Essential of error control coding, Wiley: 2006.
[31] T.David, P.Viswanath, Fundamentals of wireless communication, Cambridge university press: 2005.
[32] Available http://media.xiph.org/video/derf/

Farid Jafarian received his B.S. (with highest honors) degree from Isfahan’s Sepahan Institute of Science and Technology, Isfahan, Iran and M.S. degree from University of Birjand, Birjand, Iran, in 2008 and 2012 respectively, all in Electrical Engineering. He was with the Research Institute for Information and Communication Technology (ICT), Tehran, Iran. He is currently senior lecturer in Bonyan University. His research interest and contributions are in the areas of communication theory, information theory and their applications to wireless communications (e.g., multiuser MIMO, cognitive radio and cooperative communication), joint source and channel coding (JSCC) and communications signal processing with a focus on video and speech coding. He is also interested in applications of information theory in Bioinformatics with a focus on genome modeling

Hasan Farsi received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D in the Centre of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D degree in 2004. He is interested in speech, image and video processing on wireless communications. Now, he works as associate professor in communication engineering in department of Electrical and Computer Eng., university of Birjand, Birjand, IRAN. His Email is: hfarsi@birjand.ac.ir