A Novel JSCC Scheme for UEP-Based Scalable Video Transmission Over MIMO Systems

Chao Zhou, Chia-Wen Lin, *Senior Member, IEEE*, Xinggong Zhang, *Member, IEEE*, and Zongming Guo

Abstract— In this paper, we propose a novel joint source-channel coding (JSCC) scheme for scalable video transmission over multiple-input multiple-output (MIMO) systems. By exploiting the diversity of MIMO antennas and forward error correction (FEC)-based protection, our method aims to provide unequal error protection (UEP) for the video layers, which are mapped to appropriate antennas. Moreover, JSCC is also considered that we extract a proper subset of video layers and allocate suitable FEC redundancy to them. Jointly considering video layer extraction, FEC rate allocation, and video layer scheduling, we are able to achieve UEP so as to minimize end-to-end distortion. We formulate the scheme as a nonlinear integer optimization problem, which is known to be NP-hard. To find a near-optimal solution efficiently, we propose a low-complexity branch-and-bound algorithm, which partitions the original problem into a series of subproblems by a video layer branching technique. In each branch, the upper and lower distortion bounds are derived. In particular, we transform the video layer scheduling subproblem into a 0/1 multiple knapsack problem, which is NP-complete, and employ an evolutionary Lagrangian method to find a solution efficiently. For the FEC allocation subproblem, a Lagrange duality algorithm with fuzzy surrogate subgradient is proposed. The experimental results demonstrate that the proposed method has good efficiency while achieving close performance to the optimal results.

Index Terms— Branch-and-bound, joint source-channel coding (JSCC), scalable video transmission, unequal error protection (UEP).

I. INTRODUCTION

WITH the development of broadband wireless networks and mobile devices, delivering video over wireless networks has gained increasing attentions. Scalable Video Coding (SVC) [1] is the scalable extension of H.264/Advanced Video Coding (AVC) standard. It has drawn much attention due to its scalability in multiple dimensions (temporal, spatial, and quality scalability), which makes it flexible to various network conditions and terminal devices. To provide high-quality video transmission services in wireless links, joint source-channel coding (JSCC) has proven to be effective in improving video quality by jointly allocating bitrate between source coding and channel coding [2]–[5]. Besides, considering the unequal importance of the video layers, it is possible to achieve unequal error protection (UEP) and graceful quality gradation for scalable video transport by providing different levels of error protection to different video layers.

On the other hand, multiple-input multiple-output (MIMO) antenna systems have recently emerged to greatly enhance the performance of wireless systems in terms of capacity and reliability [6]–[8]. MIMO technology has been adopted in most of the next-generation communication protocols, such as IEEE 802.11n, WiMax, and 4G (LTE). An MIMO channel can be decomposed into a series of independent SISO subchannels with different bandwidth capacities and packet error rates (PERs). Thus, multiple data streams can be transmitted simultaneously using a spatial multiplexing technique [9], [10]. Considering the diversity of multiple antennas, we aim to provide UEP for video layers by both scheduling them to appropriate antennas and allocating proper forward error correction (FEC) redundancy.

In this paper, we propose a novel JSCC scheme for UEP-based scalable video transmission over MIMO systems, where a video is encoded with both temporal and quality scalability by SVC. Our method selectively extracts a suitable subset of video layers for transmission. As a result, UEP is achieved by both scheduling the extracted video layers to appropriate antennas and allocating FEC rate unequally, so as to minimize the end-to-end distortion. To accomplish these tasks, for each video layer, three factors need to be determined: 1) whether it would be selected for transmission; 2) which antenna it should be scheduled to; and 3) how much FEC redundancy it would be allocated with.

We formulate the JSCC scheme as a nonlinear integer optimization problem, which is, however, NP-hard. To find a practical near-optimal solution, we employ a branch-and-bound framework to derive a low-complexity branch-and-bound algorithm, which extracts the subset of video layers and partitions the original problem into subproblems by a branching technique. For each subset of video layers, we derive its distortion bounds by an appropriate video layer scheduling and FEC rate allocation. Finally, a

C. Zhou, X. Zhang, and Z. Guo are with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China.

C.-W. Lin is with the Department of Electrical Engineering, Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, and also with the Department of Computer Science and Information Engineering, Asia University, Taichung 413, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

pruning technique is utilized to remove unimportant branches, thus the complexity is reduced drastically. Our contribution is threefold.

1) To the best of our knowledge, we are among the first to jointly consider video layer extracting, video layer scheduling, and FEC rate allocation in formulating the problem of JSCC scheme for UEP-based scalable video transmission over MIMO systems. We formulate the problem into a nonlinear integer optimization problem and put the three aspects of video layer extracting, video layer scheduling, and FEC rate allocation into an efficient branch-and-bound framework to solve the optimization problem.

2) We transform the video layer scheduling problem into a 0/1 multiple knapsack problem (MKP) under the second Shannon theorem, and then propose an evolutionary Lagrangian method to find a suitable solution for the 0/1 MKP which is known to be NP-complete.

3) Using Lagrangian relaxation, we transform the FEC rate allocation problem into an unconstrained integer optimization problem. Then, based on Lagrange duality, we design a novel low-complexity fuzzy surrogate subgradient (FSSG) algorithm to solve the problem.

The rest of this paper is organized as follows. Section II provides a brief view of related work. In Section III, we formulate the proposed video transmission scheme into a nonlinear integer optimization problem. In Section IV, a branch-and-bound framework-based algorithm is proposed to solve the optimization problem. We give the details of video layer scheduling and FEC rate allocation in Sections V and VI, respectively. Section VII presents the experimental results. Finally, the conclusion is drawn in Section VIII.

II. RELATED WORK

Resource allocation for scalable video transmission over MIMO wireless networks has been studied to improve video transmission quality [11]–[15] for years. Li *et al.* [11] proposed a cross-layer optimization scheme, which includes packet prioritization and QoS mapping, for the delivery of SVC over the IEEE 802.11e wireless networks. Li *et al.* [12], [15] also studied resource allocation for scalable video transmission over MIMO, where bit loading and power allocation are jointly considered.

While considering the unequal importance of video layers, UEP has proven to be effective in achieving error-resilient video transport over error-prone networks. Existing UEP schemes for scalable video transmission can be mainly classified into two categories according to the video layers to be protected in a video bitstream: 1) UEP in different quality layers [16]–[20] and 2) UEP in different temporal layers [21]–[23]. In quality scalability of SVC, each video frame is encoded into multiple layers by multilevel quantization techniques. Since different layers of a scalable video bitstream are not equally important (i.e., the base layer is the most important, the first enhancement layer is the second, and so on), an intuitive way of protecting a scalable bitstream is to add more protection to the layers that impact the quality more. The UEP for quality layers has been addressed by

many researchers [16]–[20]. For instance, in [17], the problem of UEP between the base and enhancement layers of fine-granularity-scalability (FGS) coding is investigated and the concept of fine-grained loss protection is introduced. On the other hand, by adopting the hierarchical-B frame structure, temporal scalability is achieved in SVC. Considering the different importance of different types of frames, the anchor frames (i.e., I-frames and P-frames) should be given with more protection than the less-important nonanchor frames (i.e., B-frames). Researchers have tackled the problem of UEP for SVC by appropriately considering the various frame types [21]–[24]. For example, in [21]–[23], the sensitivity of succeeding frames is explored to minimize the mean distortion over the transmitted sequence. The bit error sensitivity of the three types of frames is considered so that the dependence among them is exploited in [24].

Nevertheless, joint consideration of quality scalability and temporal scalability of SVC for channel rate allocation was relatively rarely studied in [25] and [26]. Fang and Chau [25] jointly consider these two aspects and solve the problem using a genetic algorithm (GA) for MPEG-4 FGS video. However, GAs are considered to be slow and susceptible to premature convergence. In [26], a complex model was proposed to estimate the overall distortion of decoded frames, and the performance of FEC rate allocation is heavily dependent on the distortion model. However, these two works ignored the benefits of JSCC.

The JSCC has also proven to an effective means for error resilient video streaming. A JSCC framework for scalable video transmission over MIMO systems was proposed in [5]. In this method, the video layers are protected by multiplexing-diversity tradeoff of MIMO system and unequal FEC rate allocation. However, it only considered the quality scalability of SVC, but did not consider achieving UEP by video layer scheduling. In [27], a novel JSCC scheme for scalable video transmission over MIMO systems was proposed. In the scheme, we derived the FEC rate allocation under the assumption that FEC coding can achieve the theoretic upper bound of Shannon limits, and scheduled the video layers to multiple antenna elements by a heuristic scheduling method. However, the FEC rate may be allocated poorly since Shannon limits generally cannot be achieved in practice.

III. PROBLEM FORMULATION

The block diagram of our proposed JSCC scheme for UEP-based scalable video transmission is shown in Fig. 1. The input video sequence is first encoded into several layers with an SVC encoder, which is then sent over MIMO wireless channels. To achieve graceful quality degradation in lossy transmission, we allocate the bandwidth between the source and channel coding rates, and then perform UEP for the video layers. In particular, we adapt the source video rate by extracting a subset of video layers for transmission and dropping the remaining. For the extracted video layers, UEP is provided by both scheduling them to appropriate antennas and adding unequal FEC protections. Therefore, our scheme involves three key components: 1) video layer extraction; 2) video layer scheduling; and 3) FEC rate allocation.
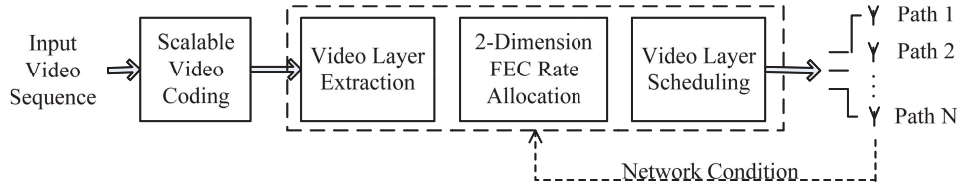
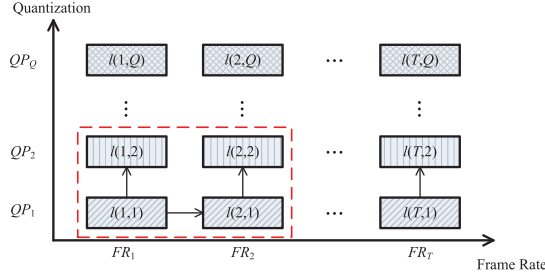Fig. 1.   System diagram of video transmission with JSCC focused on UEP.

Fig. 2.   Structure of quality-temporal scalability of SVC bitstream.

A. Video Layer Extraction

As mentioned previously, we adapt the source rate of a scalable video by extracting a subset of video layers for transmission and dropping the remaining. When a video layer is extracted for transmission, it would be scheduled to a certain antenna; otherwise, it is dropped. The aim of this operation is to find the best subset of video layers under the bandwidth constraints. Considering the dependence between video layers, for a given quantization parameter and frame rate, we extract all video layers with larger quantization parameters and smaller frame rates. For example, assuming the given quantization parameter and frame rate are QP_2 and FR_2, respectively, the extracted video layers would include $l(1, 1)$, $l(1, 2)$, $l(2, 1)$, and $l(2, 2)$ as indicated in the red broken rectangle in Fig. 2, where $l(t, q)$ denotes the video layer at the tth temporal layer and the qth quality layer.

B. Video Layer Scheduling

Considering the diversity of channel qualities of multiple transmit antennas, UEP can be provided by scheduling the video layers to appropriate antennas so as to improve video quality. In this paper, if a video layer is extracted for transmission, all of its data will be scheduled to the same antenna. On the other hand, multiple extracted video layers are allowed to be scheduled to multiple antennas (but data from the same layer has to be scheduled to the same antenna) or to the same antenna. Thus, this is a many-to-one mapping problem.

C. FEC Rate Allocation

Application-level FEC has proven to be an efficient means for packet loss recovery. By unequal FEC rate allocation, video layers are protected unequally and graceful video quality is achieved. Considering both the temporal and quality scalability of SVC, we allocate redundancy unequally with the aim of minimizing the distortion.

In the application scenarios of SVC, the video is encoded into multiple layers with multidimensional video scalability

at the server side. Combining with the feedback network conditions (e.g., bandwidth constraint, PER), suitable video layers are extracted for transmission to maximize the video quality. Thus, estimating the video rate and weight for each layer is indispensable for layer extraction and our joint optimization problem.

Current H.264/SVC video coding standard supports scalability in the temporal, spatial, and quality dimensions. In this paper, we suppose a video sequence is encoded into $L = TQ$ layers with T layers in the temporal dimension and Q layers in the quality dimension. When a video layer belongs to the tth temporal layer and the qth quality layer, it is denoted by $l(t, q)$, as shown in Fig. 2. Some objective quality metrics, such as peak signal-to-noise ratio (PSNR), cannot represent quality changes in the temporal dimension. Thus, we adopt a temporal-quantization quality model [28] to measure the rate and weight of video layers. In particular, for the given frame rate FR_t and quantization parameter QP_q, the video rate and quality are written as

$$
\begin{cases}
R\left(FR_t, QR_q\right) = R_{max} \left(\dfrac{QR_q}{QR_{min}}\right)^{-a} \left(\dfrac{FR_t}{FR_{max}}\right)^{b} \\[3mm]
Q\left(FR_t, QR_q\right) = Q_{max} \dfrac{e^{-c \frac{QR_q}{QR_{min}}}}{e^{-c}} \dfrac{1 - e^{-d \frac{FR_t}{FR_{max}}}}{1 - e^{-d}}
\end{cases}
\tag{1}
$$

where FR_{max} and QP_{min} should be chosen based on the underlying application, R_{max} and Q_{max} are the actual rate and quality when coding a video at FR_{max} and QP_{min}, and a, b, c, and d are the model parameters. We define $R(FR_t, QP_q) = 0$, and $Q(FR_t, QP_q) = 0$ if $t = 0$ or $q = 0$. For each video layer $l(t, q)$, we denote its rate and weight by $r(t, q)$ and $w(t, q)$. Then, according to (1), we get $r(t, q)$ and $w(t, q)$ as

$$
\begin{cases}
r(t, q) = \ R\left(FR_t, QR_q\right) + R\left(FR_{t-1}, QR_{q-1}\right) \\
\qquad\quad - R\left(FR_t, QR_{q-1}\right) - R\left(FR_{t-1}, QR_q\right) \\
w(t, q) = \ Q\left(FR_t, QR_q\right) + Q\left(FR_{t-1}, QR_{q-1}\right) \\
\qquad\quad - Q\left(FR_t, QR_{q-1}\right) - Q\left(FR_{t-1}, QR_q\right).
\end{cases}
\tag{2}
$$

The intuition of deriving (2) from (1) is to generalize the case of SVC with single dimensional scalability proposed in [29]. For example, in SVC with single dimensional scalability, W_i denotes the PSNR when an SVC-coded bitstream is decoded with all i layers, then $w_i = W_i - W_{i-1}$ is the weight of layer i, which indicates the additional PSNR when the bitstream is decoded with all i layers compared with that decoded with only $i - 1$ layers. Similarly, in this paper, $w(t, q)$ denotes the additional visual quality gain when a video bitstream is decoded with all layers $l(i, j)(i \le t, j \le q)$ compared with that decoded without the $l(i, j)$th layer. The video rate $r(t, q)$ is derived in a similar way.

For more details about this model, one can refer to [28]. It is worth noting that although there is still no widely accepted model for combined temporal and quality scalability so far, there has been some promising results reported in [30]–[34]. This paper aims to propose an efficient optimization framework to incorporate such models, which can well characterize the rate $r(t, q)$ and weight $w(t, q)$ for each video layer. Besides the model in [28], in fact, other similar models, which can well characterize the rate $r(t, q)$ and weight $w(t, q)$ for each video layer, can also be plugged into our framework without changing the problem formulation and the solution algorithm. We adopted the model proposed in [28] under the consideration that, to the best of our knowledge, it and its extended versions should be among the most cited ones for combined temporal and quality scalability. Wang *et al.* [28] have some subsequent publications [35], [36] on the same issue but the models are generally similar. Promising results about the models and their successful applications to video transport were reported in [35] and [36]. While how to design a better model is beyond the scope of this paper, we believe that the experimental results based on the model in [28] can be used as a good example to demonstrate the effectiveness of our proposed scheme and solution algorithm.

Suppose there are N transmit antennas and each antenna $\{n | n \leq N\}$ has bandwidth capacity C_n (bps) and PER ρ_n, which depends on the channels characteristics. We define $\mathbf{X} = \{x_{tq}^n\}$ as a binary indicator matrix with size $L \times N$. Its entry $x_{tq}^n = 1$ if video layer $l(t, q)$ is scheduled to antenna n; otherwise $x_{tq}^n = 0$. Therefore, when video layer $l(t, q)$ is selected for transmission, it will be scheduled to one of the N transmit antennas, thus we have $\sum_{n=1}^{N} x_{tq}^n = 1$; otherwise $\sum_{n=1}^{N} x_{tq}^n = 0$. After scheduling the video layers to multiple antennas by matrix $\mathbf{X}$, video layer $l(t, q)$ is transmitted over the antenna whose PER is

$$p_{tq} = \begin{cases} \sum_{n=1}^{N} x_{tq}^n \rho_n & \text{if } \sum_{n=1}^{N} x_{tq}^n = 1 \\ 1 & \text{if } \sum_{n=1}^{N} x_{tq}^n = 0 \end{cases} \quad (3)$$

where ρ_n denotes the PER of antenna n.

To correct transmission errors, application-level FEC coding is applied. For video layer $l(t, q)$, suppose every k_{tq} video source packets are encoded with the packet-level FEC (M, k_{tq}) code to generate additional $M - k_{tq}$ redundant packets, where M is the length of FEC block. As long as a client receives at least k_{tq} out of the M packets, it can recover all the source packets. Then, the FEC error probability of video layer $l(t, q)$ can be calculated by

$$P_{tq}^{\text{FEC}}(\mathbf{X}, k_{tq}) = 1 - \sum_{i=0}^{M-k_{tq}} \binom{M}{i} p_{tq}^i (1 - p_{tq})^{M-i} \quad (4)$$

and p_{tq} is given in (3). In this paper, we adopt application-level FEC rather than physical layer FEC considering its flexibility and convenience for the experiments. In the physical layer, the number of available protection levels is equal to the number of available physical rates, which is usually rather limited and generally cannot be changed as it has already been specified

in the protocols. While in application layer, the available FEC rates is dependent on the length of an encoding block, e.g., if the encoding block length is M, the FEC coding rate can be $\{1/M, 2/M, \ldots, 1\}$. Note, application-layer FEC and physical rate selection can be jointly optimized to further improve the transmission quality and there is a fundamental tradeoff between these two aspects. For example, in the physical layer, different physical rates provide different transmission reliability and rate capacity, whereas in the application layer, different FEC rates provide different transmission reliability and consume different rate resource. Besides, it is worth noticing that there exists cliff phenomenon for FEC codes that for each given PHY transmission error probability, there exists a threshold that when the allocated FEC redundancy is higher than the threshold, the error probability becomes rapidly very low, otherwise it becomes rapidly very high. Moreover, the higher the PHY transmission error probability, the higher the threshold. Thus, when the average PER is becoming higher, the threshold is also becomes larger and more bandwidth resource is needed. Instead of adding insufficient FEC redundancy, which leads to high error probability, we discard some high video layers (unimportant video layers) so as to add enough FEC redundancy to the low video layers (important video layers) as the experiment results shows. In addition, the probability of packet error is a function of the packet length M as (4) shows. In general, by jointly optimizing the packet length with the other aspects, the video quality can be further improved with the price of increasing both the optimization complexity and the decoding complexity. In this paper, we consider the packet length as a constant and do not adjust the transmission process so as to reduce the complexity.

Considering the dependence of scalable video layers that a higher layer can be decoded if and only if all the lower layers have been decoded correctly. Thus, the expected distortion of the reconstructed video is expressed by

$$D(\mathbf{X}, \mathbf{K}) = \sum_{t \leq T, \, q \leq Q} w(t, q) \left(1 - \prod_{i \leq t, j \leq q} \left(1 - P_{ij}^{\text{FEC}}(\mathbf{X}, k_{ij}) \right) \right) \quad (5)$$

where $\mathbf{K} = [k_{11}, k_{21}, \ldots, k_{TQ}]$.

It is worth noticing that when video layer $l(t, q)$ is dropped, we have $k_{tq} = 0$. Thus, the total transmitted rate over antenna n is $\sum (M/k_{tq}) r(t, q) x_{tq}^n$ for all $k_{tq} \neq 0$, which must not exceed the antenna's bandwidth capacity. Now, we can formulate the system as

$$\{\mathbf{X}^*, \mathbf{K}^*\} = \underset{\{\mathbf{X}, \mathbf{K}\}}{\arg\min} \, D(\mathbf{X}, \mathbf{K})$$

$$\text{s.t.} \sum_{t \leq T, q \leq Q} \frac{M}{k_{tq}} r(t, q) x_{tq}^n \leq C_n \quad \forall n \leq N, k_{tq} \neq 0$$

$$\sum_{n \leq N} x_{tq}^n \in \{0, 1\} \quad \forall t \leq T, q \leq Q$$

$$x_{tq}^n \in \{0, 1\} \quad \forall t \leq T, q \leq Q, n \leq N$$

$$0 \leq k_{tq} \leq M, k_{tq} \in Z \quad \forall t \leq T, q \leq QC \quad (6)$$
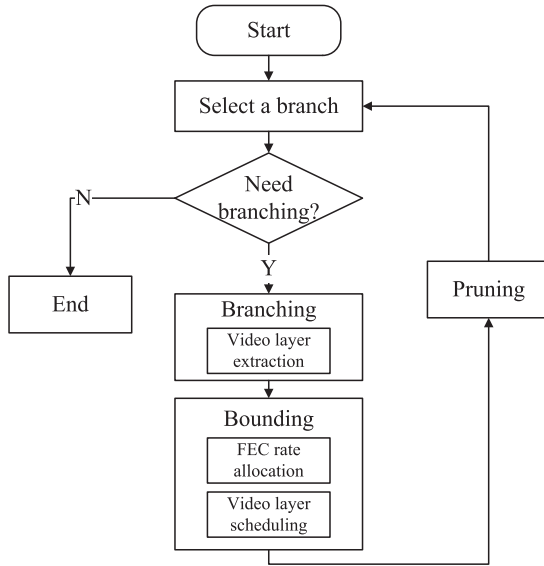
and Z denotes the set of integer.

Fig. 3. Flowchart of the proposed branch-and-bound algorithm.

The objective of this optimization problem is to minimize the total expected distortion by joint video source rate adaptation (layer extraction), video layer scheduling, and FEC rate allocation under the bandwidth constraints. It is a nonlinear integer optimization problem, and performing an exhaustive search (ES) over the whole solution space whose size is $M^L(N+1)^L$ would be computationally very expensive and, therefore, impractical.

IV. BRANCH-AND-BOUND ALGORITHM

To solve the optimization problem in (6), we propose a branch-and-bound algorithm. Branch-and-bound is an iterative method for solving optimization problems, especially for discrete and combinatorial problems. Fig. 3 shows the flowchart of the proposed branch-and-bound algorithm. Three important components are included in the branch-and-bound procedure. The first is called branching, which is used to partition the primal problem into some simpler subproblems (branches). The second is bounding, which is used to derive the bounds for each branch. It is worth noticing that the bounds of each branch are further affected by the FEC rate allocation scheme and video layer scheduling scheme. In Sections V and VI, we will give the details of how to schedule the video layers to appropriate channels and to allocate the FEC redundancy for the video layers. The last is pruning, which is used to delete some unimportant branches so as to reduce complexity.

A. Branching

We employ the branching technique to partition the original video source layers into mutually exclusive subsets of layers. Each branch, which is represented by a node, is a set of extracted layers for transmission. In this paper, the term branch and node are used interchangeably. Considering the dependence between video layers, we denote a branch by a two-tuple $b(t, q)$, and the branch should include all the video layers $l(i, j)$ with $1 \le i \le$ and $1 \le j \le q$.

Then, we branch the video layers as follows: starting from the branch with the maximal frame rate index T and

quantization-step index Q, i.e., branch $b(T, Q)$, we iterate every combination of frame rate and quantization step-size. Since branch $b(t, q)$ includes all video layers lower than video layer $l(t, q)$, including layer $l(t, q)$, it is branched into two child nodes, $b(t, q-1)$ and $b(t-1, q)$, and branch $b(t, q)$ is partitioned. By iterating this process, we obtain the branch-and-bound tree.

B. Bounding

For a certain branch, its distortion comes from two parts: 1) the distortion caused by dropping some video layers and 2) the distortion caused by transmission error for the extracted video layers, where the latter is further affected by both video layer scheduling and FEC rate allocation. In particular, since a branch $b(t_0, q_0)$ contains video layers $l(t_0, q_0)$ and all other video layers lower than $l(t_0, q_0)$, while dropping the remaining, the total excepted distortion for branch $b(t_0, q_0)$ can be computed

$$D(b(t_0, q_0), \mathbf{X}, \mathbf{K}) = \sum_{\substack{t \le t_0 \\ q \le q_0}} w(t, q) P_{tq}^V + \sum_{\substack{t_0 < t \le T \\ q_0 < q \le Q}} w(t, q) \quad (7)$$

where

$$P_{tq}^V = 1 - \prod_{\substack{i \le t \\ j \le q}} \left(1 - P_{ij}^{\text{FEC}}(\mathbf{X}, k_{ij})\right) \quad (8)$$

is the video error rate for layer $l(t, q)$. Then, we move to derive the distortion bounds of (7).

1) Lower Bound: The lower bound is generally derived by relaxing the constraints. From (4) and (7), we find that the distortion $D(b(t_0, q_0), \mathbf{X}, \mathbf{K})$ is an increasing function of both k_{ij} and $p_{ij}(i \le t_0, j \le q_0)$. A smaller k_{ij}, which implies more FEC protection, would lead to fewer distortion, but it comes with the risk of exceeding the bandwidth capacity. Under this observation, we bound the distortion by relaxing the bandwidth constraints that we set $k_{ij} = 1$ without considering the constraints of bandwidth capacity at first. Then, since p_{ij} is affected by video layer scheduling matrix $\mathbf{X}$, we further bound the distortion under the assumption that all video layers in branch $b(t_0, q_0)$ are scheduled to the antenna with the smallest PER. Therefore, we have

$$p_{ij} = \min(\rho_1, \rho_2, \ldots, \rho_N), \quad \text{if } i \le t_0, j \le q_0. \quad (9)$$

Thus, by substituting $k_{ij} = 1$ and (9) into (4), we obtain the lower bounds of the FEC error probability for video layer $l(i, j)$ in branch $b(t_0, q_0)$ as

$$P_{ij}^{\text{FEC}}(\mathbf{X}, k_{ij}) = P_{ij}^{\text{FEC}}(\mathbf{X}, 1) = p_{ij}^M \quad (10)$$

where p_{ij} is equal to the minimal PER given by (9). Then, the lower bound of the distortion for branch $b(t_0, q_0)$ is derived by substituting (10) into (7).

2) Upper Bound: The upper bound of the distortion in (7) is derived by finding a near optimal and feasible video scheduling matrix $\mathbf{X}$ and FEC rate allocation vector $\mathbf{K}$. However, these two aspects are dependent, since for a different video layer scheduling solution, the optimal FEC rate allocation is also different. Therefore, it is challenging to solve these

two problems jointly. Instead, we first schedule the video layers to multiple antenna elements based on the second Shannon theorem. Then, based on Lagrange duality, the FEC rate is allocated by our proposed FSSG algorithm. The details of video layer scheduling and FEC rate allocation are given in Sections V and VI, respectively.

C. Pruning

Pruning is implemented to delete unimportant branches so as to reduce the search space. The core of pruning is based on an observation that, for the above minimization problem, one branch and those branches rooted at it can be safely discarded if and only if the lower bound of distortion for this branch is higher than the upper bound for any other branches. Thus, the computational complexity can be reduced by pruning these branches.

In each iteration, the branch with the smallest lower bound of distortion is selected. When the lower bound is equal to the upper bound or the gap between them achieves a certain predefined threshold, the algorithm terminates; otherwise, the branch is further partitioned.

V. VIDEO LAYER SCHEDULING

According to the second Shannon theorem, by employing appropriate channel coding, the source bits can be transmitted with arbitrary transmission reliability when the source rate is not higher than the effective bandwidth capacity. Then, we are able to treat the nth ($1 \leq n \leq N$) antenna elements as a lossless channel with an effective bandwidth capacity of

$$\tilde{C}_n = C * (1 - \rho_n). \tag{11}$$

Then, we schedule the video layers to antennas under the effective bandwidth constraints. However, it is worth to noticing that it is only theoretically correct that there are no losses when the video rate is no more than $\tilde{C}_n$ according to Shannon limits. In practice, there is no such a channel coding scheme that can achieve this limit as Shannon limits are proved under the assumption that both the power and the encoding length are infinite. This is also the reason why we reallocate the FEC redundancy by our proposed FSSG method in Section VI.

For branch $b(t_0, q_0)$, we transform the video layer scheduling problem into a 0/1 MKP: given a set of $t_0 * q_0$ video layers with rate $r(t, q)$ and weight $w(t, q)(t \leq t_0, q \leq q_0)$, and the total N transmit antenna elements (knapsacks) with capacity $\tilde{C}_n(1 \leq n \leq N)$. The objective is to maximize the total weight under the effective bandwidth constraints. Then, the 0/1 MKP is formulated as

$$\text{MKP}: \ \max \sum_{t \leq t_0, \ q \leq q_0} \sum_{n \leq N} w(t, q)x_{tq}^n$$

$$\text{s.t.} \sum_{t \leq t_0, q \leq q_0} r(t, q)x_{tq}^n \leq \tilde{C}_n \quad \forall 1 \leq n \leq N$$

$$\sum_{n \leq N} x_{tq}^n \in \{0, 1\} \quad \forall t \leq t_0, \quad q \leq q_0$$

$$x_{tq}^n \in \{0, 1\} \quad \forall t \leq t_0, \quad q \leq q_0. \tag{12}$$

However, since the 0/1 MKP is a well-known NP-complete problem [37], we employ the evolutionary Lagrangian method proposed in [38] to find a practical near-optimal solution. For more details about this method, one can refer to [38] and the references therein. We solve the 0/1 MKP with the purpose of finding a feasible video layer scheduling solution. For branch $b(t_0, q_0)$, a feasible video layer scheduling solution means each video layer in the branch has been scheduled to a certain antenna and the other video layers have been dropped, so we have $\sum_{n=1}^N x_{tq}^n = 1(t \leq t_0, q \leq q_0)$. Then, combining with $x_{tq}^n = 0(t > t_0, q > q_0)$, we obtain a feasible video layer scheduling solution for the primal problem.

Thus, for this 0/1 MKP, if $\sum_{n=1}^N x_{tq}^n = 1$ for all $t \leq t_0$, $q \leq q_0$, the solution is a feasible video layer scheduling solution. Otherwise, it means too many video layers have been selected and the branch needs to be further branched.

Nevertheless, the second Shannon theorem does not give the method of how to allocate FEC redundancy so as to achieve the highest transmission reliability. Based on the above video layer scheduling solution, we allocate FEC redundancy using our proposed FSSG presented in Section VI.

VI. PROPOSED FSSG ALGORITHM FOR FEC RATE ALLOCATION

For branch $b(t_0, q_0)$, we have obtained the video layer scheduling matrix $\mathbf{X}_0$ by solving the 0/1 MKP in Section V. Now, given $\mathbf{X}_0$, we would like to allocate the redundancy among the video layers to minimize the overall distortion. It is worth noticing that for a given matrix $\mathbf{X}_0$, $p_{tq}(t \leq t_0, q \leq q_0)$ is no longer a variable in (4) according to (3). Thus, the problem can be rewritten as follows:

$$\text{IP}: \ \{\mathbf{K}^*\} = \underset{\{\mathbf{K}\}}{\text{argmin}} \ D(b(t_0, q_0), \mathbf{X}_0, \mathbf{K})$$

$$\text{s.t.} \sum_{t \leq t_0, q \leq q_0} \frac{M}{k_{tq}} r(t, q)x_{tq}^n - C_n \leq 0 \quad \forall n \leq N \tag{i}$$

$$k_{tq} - M \leq 0 \quad \forall t \leq t_0, \quad q \leq q_0 \tag{ii}$$

$$-k_{tq} + 1 \leq 0 \quad \forall t \leq t_0, \quad q \leq q_0 \tag{iii}$$

$$k_{tq} \leq 0 \quad \forall t > t_0, \quad q > q_0 \tag{iv}$$

$$-k_{tq} \leq 0 \quad \forall t > t_0, \quad q > q_0 \tag{v}$$

$$k_{tq} \in Z \quad \forall t \leq T, \quad q \leq Q. \tag{13}$$

This is an integer programming (IP) problem, and there are $N + 2L$ constraints for the above IP problem: 1) constraint (i) indicates the total rate should not exceed the bandwidth capacity for each antenna; 2) constraints (ii) and (iii) indicate that when a video layer is selected for transmission, there are at most M and at least one source packet in an FEC block; and 3) constraints (iv) and (v) indicate no packet is transmitted for the dropped layers, i.e., $k_{tq} = 0$. For the sake of clarity, we denote these constraints as $\mathbf{g}(\mathbf{K}) \leq 0$.

The $N + 2L$ constraints couple the decision variables k_{tq} thus making the above IP difficult to solve. Using the Lagrangian relaxation approach, we relax the constraints by introducing an $(N + 2L) \times 1$ Lagrangian multiplier vector $\boldsymbol{\lambda}$ and the following Lagrangian function:

$$J(\mathbf{K}, \boldsymbol{\lambda}) \equiv D(b(t_0, q_0), \mathbf{X}_0, \mathbf{K}) + \mathbf{g}^{\text{T}}(\mathbf{K})\boldsymbol{\lambda}. \tag{14}$$

The relaxed problem is to minimize $J(\mathbf{K}, \boldsymbol{\lambda})$ over $\mathbf{K}$, resulting in the concave non-smooth dual function as

$$J(\boldsymbol{\lambda}) \equiv \min_{1 \leq k_{tq} \leq M} \tilde{J}(\mathbf{K}, \boldsymbol{\lambda}). \tag{15}$$

The dual problem is to maximize the dual function

$$J^* = J(\boldsymbol{\lambda}^*) = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} J(\boldsymbol{\lambda}) \tag{16}$$

where $\boldsymbol{\lambda}^*$ is the optimal dual solution and $J^* = J(\boldsymbol{\lambda}^*)$ is the optimal dual value. In most cases, the optimization problem in (16) is performed iteratively. At the termination of such iterations, a simple heuristic is applied to adjust the relaxed problem solutions so as to obtain a feasible result satisfying all the constraints.

In general, a subgradient method is used to maximize the dual function, and at the uth iteration, the multipliers are updated by

$$\boldsymbol{\lambda}^{u+1} = \boldsymbol{\lambda}^u + s^u \mathbf{g}^u \tag{17}$$

where $\mathbf{g}^u = \mathbf{g}(\mathbf{K}^u)$ is the subgradient of $J(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda}^u$, here $\mathbf{K}^u = \mathbf{K}(\boldsymbol{\lambda}^u) = \arg\min_{1 \leq k_{tq} \leq M} \tilde{J}(\mathbf{K}, \boldsymbol{\lambda}^u)$, and s^u is the stepsize at the uth iteration. However, to find the search direction $\mathbf{g}^u$, this method requires the minimization of all the subproblems, which is very time consuming.

In this paper, we propose to fuzzily utilize the historic surrogate subgradient information to obtain the search direction. When deriving the subgradient, instead of solving the optimization subproblems, only an approximate minimization is performed where the new iterate $\mathbf{K}(\boldsymbol{\lambda}^u)$ satisfies the following condition for the given set of multipliers $\boldsymbol{\lambda}^u$:

$$\mathbf{K}^u = \mathbf{K}(\boldsymbol{\lambda}^u) \in \left\{ \mathbf{K} \left| \begin{array}{l} \tilde{J}(\mathbf{K}, \boldsymbol{\lambda}^u) < \tilde{J}(\mathbf{K}^{u-1}, \boldsymbol{\lambda}^u) \\ \forall 1 \leq k_{tq} \leq M - 1. \end{array} \right. \right\} \tag{18}$$

Furthermore, the linear combination of historic subgradient information is utilized to predict the current subgradient for the next iteration as

$$\tilde{\mathbf{g}}^u = \sum_{i=1}^{u} v_i^u \mathbf{g}^i \tag{19}$$

where v_i^u is the weight of the ith historic gradient at the uth iteration. Then, we plug $\mathbf{g}^u$ in (17) by (19), and the multipliers are updated as

$$\boldsymbol{\lambda}^{u+1} = \boldsymbol{\lambda}^u + s^u \tilde{\mathbf{g}}^u. \tag{20}$$

In general, when a historic surrogate subgradient is close to the current subgradient, it contains more information for the search direction decision and, therefore, a high value of weight is given. The weight of each historic gradient is calculated as

$$v_i^u = \frac{\bar{v}_i^u}{\sum_{j=1}^{u} \bar{v}_j^u} \tag{21}$$

where

$$\bar{v}_i^u = \begin{cases} [J(\boldsymbol{\lambda}^u) + \varepsilon - J(\mathbf{K}^i, \boldsymbol{\lambda}^u)]/\varepsilon \\ \quad \text{if} \quad J(\mathbf{K}^i, \boldsymbol{\lambda}^u) < J(\boldsymbol{\lambda}^u) + \varepsilon \\ 0, \quad \text{else.} \end{cases} \tag{22}$$

In the above, all near-optimal solutions (having a cost within ε of the minimum cost) are considered and assigned a weight. Since the near-optimal solutions with positive weights are usually byproducts when the relaxed problem is solved, their corresponding $\mathbf{g}^u$ and v_i^u can be easily obtained. The parameter ε is set to satisfy

$$0 \leq \varepsilon \leq \frac{J^* - J(\boldsymbol{\lambda}^u)}{\alpha}, \quad \alpha \geq 2 \tag{23}$$

and the step-size is set to satisfy the following constraint:

$$0 < s^u \leq \frac{2(J^* - J(\boldsymbol{\lambda}^u))}{\alpha \|\mathbf{g}^u\|^2}, \quad \alpha \geq 2 \tag{24}$$

where α is a parameter. As a result, the proposed FSSG algorithm has several nice properties.

Proposition 1: If $\forall i = 1, 2, \ldots, u$, the value of $J(\mathbf{K}^i, \boldsymbol{\lambda}^u)$ is less than the optimal dual cost J^*, i.e., $J(\mathbf{K}^i, \boldsymbol{\lambda}^u) < J^*$, the corresponding surrogate subgradient $\mathbf{g}^i$ is an acute angle with the optimal direction. That is

$$0 < J^* - J(\mathbf{K}^i, \boldsymbol{\lambda}^u) < (\mathbf{g}^i)^{\mathrm{T}}(\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u). \tag{25}$$

Proof: The proof is given in Appendix A.   ∎

This proposition indicates that in the FSSG algorithm, each of the selected historic surrogate subgradients is an acute angle with the optimal direction. Otherwise, they are not used since their weights are zero.

Proposition 2: When parameter $\alpha \geq 2$, the fuzzy subgradient direction is always in an acute angle with the optimal direction

$$0 \leq \frac{J^* - J(\boldsymbol{\lambda}^u)}{\alpha} < (\tilde{\mathbf{g}}^u)^{\mathrm{T}}(\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u). \tag{26}$$

Proof: The proof is given in Appendix B.   ∎

This proposition justifies that the combination of historic surrogate subgradients, i.e., the fuzzy subgradient, is always in an acute angle with the optimal direction.

Theorem 1: With the proposed FSSG algorithm, the multipliers move closer to an optimal $\boldsymbol{\lambda}^*$ step by step

$$\|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^{u+1}\| < \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u\| \quad \forall u \geq 1. \tag{27}$$

Proof: The proof is given in Appendix C.   ∎

Considering that it is very hard or even impossible to directly obtain the optimal value J^* during the iterations in practice, an approximate estimate based on the method proposed in [39] can be used. To simplify the implementation, during the iterations, we always set J^* is equal to the best result that has been already obtained in this paper. When the iterations terminate, the relaxed problem solutions are adjusted by applying some heuristics, such as the floor operation, to obtain a feasible result satisfying the constraints. Then, the distortion is obtained according to (7), which is also the upper bound for $D(b(t_0, q_0), \mathbf{X}, \mathbf{K}^*)$.

VII. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed algorithms. Three test video sequences, *Akiyo*, *City*, and *Crew*, with different motion characteristics and spatial details are used for the performance evaluation. All sequences have a spatial resolution of 352×288 pixels (CIF) are encoded into
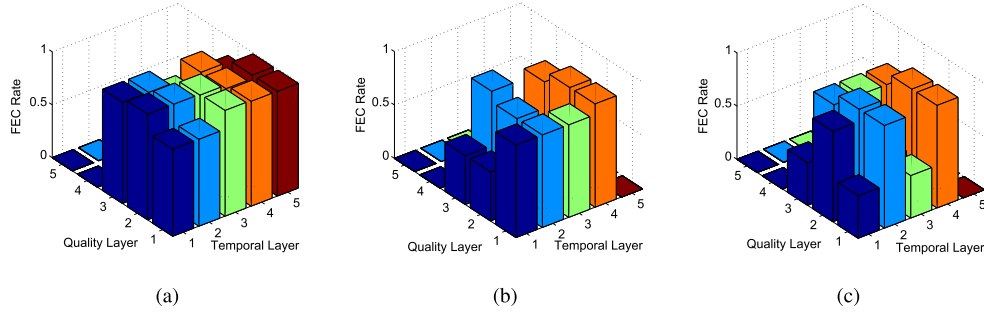
Fig. 4. FEC rate allocation for video sequence *City* with average bandwidth capacity 100 kb/s over various PER. (a) Average PER = 5%. (b) Average PER = 10%. (c) Average PER = 15%.

TABLE I

QR AND FRAME RATE FOR ALL VIDEO SEQUENCES

Layer Index	1	2	3	4	5
QP	44	40	36	32	28
Frame Rate	1.875	3.750	7.5	15	30

TABLE II

SCHEDULING RESULTS FOR *City* WITH BANDWIDTH
CONSTRAINT 200 kb/s AND AVERAGE PER = 10%

Antenna 1	$l(1,1), l(1,2), l(3,1)$
Antenna 2	$l(2,1), l(3,2), l(3,3), l(4,1)$
Antenna 3	$l(2,2), l(4,2), l(4,3)$
Antenna 4	$l(1,3), l(2,3)$

TABLE III

PARAMETERS FOR THE TESTED SEQUENCES

Video sequence / Parameters	Akyio	City	Crew
a	1.21	1.19	1.24
b	0.47	0.48	0.67
c	0.11	0.13	0.17
d	8.03	7.35	7.34
$Q_{\max}$	88.60	89.14	89.98
$R_{\max}$ (kbps)	163.23	658.25	1381.95

5 layers in both temporal and quality dimensions. In Table I, we give the average QR and frame rate for each layer, and in Table III, we show the details of model parameter for all sequences. For each video sequence, the details of bitrates and weights for all video layers are given in Tables IV–VI. Parameter α is set to two. Suppose there are four antennas with the same bandwidth capacity that $C_1 = C_2 = C_3 = C_4$, but having different PERs with $\rho_1 = 2\rho_2 = 4\rho_3 = 8\rho_4$. For the given PER, the GE packet loss model is used to generate packet losses. The experiments are implemented over a wide range of bandwidth capacities and average PERs. This experimental setup is considered sufficiently comprehensive to represent the diversity of different antenna elements and heterogeneous network conditions.

A. FEC Rate Allocation

We first evaluate the FEC rate allocation performance of our proposed scheme. Figs. 4–6 present the FEC rate allocation under various network conditions. Due to the space limit of this paper, we only present the results of sequence *City* because its $R_{\max}$ is medium among the selected three video sequences.

All the figures show that for the given bandwidth capacity constraints, some relatively higher layers are dropped as the average PER increases, so that the source rate is reduced and more redundancy bits can be added to protect the extracted video layers. For example, in Fig. 4(a), only the two highest quality layers are dropped and 15 video layers are transmitted when average PER is 5%. When average PER increases to

15%, the highest temporal layers are also dropped and only 12 video layers are transmitted.

However, although lower layers are more important than higher layers due to the inter-layer dependence, they may not always be given with stronger FEC protection. For instance, the FEC rate of the first quality layers is higher than the fourth quality layers in Fig. 5(b) (an FEC code with a higher code rate implies lower protection capability since the code rate is k/M for FEC(M, k)). This is mainly because multiple transmit antennas have different transmission reliability. We have given an example of the scheduling results in Table II. Combining with results in Fig. 5(b) and Table II, we can see that the more important video layers may not be given more FEC protections, this is because FEC rate allocation is not only dependent on video layers importance, but also affected by the video layer scheduling results, and vice versa. Under different bandwidth constraints and average PERs, it has similar characteristic and we omit the scheduling results of the other scenarios.

When comparing the results under different average PERs and bandwidth capacities, we observe that the FEC rate allocation and video layer extraction patterns are quite different. However, when some video layers need to be dropped due to either a high average PER or a low bandwidth capacity, the relatively higher layers are dropped preferentially to give more protections to the relatively lower layers.

B. Visual Quality Estimation

In this section, we compare the video quality of the proposed scheme against other schemes under various network conditions. The video quality is computed by subtracting the distortion in (5) from the maximal quality given in Table III.

1) Quality Evaluation Under the Single-Antenna Scenario: We first demonstrate the effectiveness of the branching

TABLE IV
VIDEO LAYERS' RATE (kb/s)/WEIGHT FOR *Akiyo*

Temporal layer / Quality layer	1	2	3	4	5
1	4.58/19.10	1.76/11.56	2.44/11.24	3.38/5.63	4.69/0.86
2	3.67/6.04	1.41/3.66	1.9591/3.56	2.71/1.78	3.76/0.27
3	6.34/4.51	2.44/2.73	3.38/2.65	4.69/1.33	6.49/0.20
4	1.00/3.00	3.86/1.81	5.34/1.76	7.40/0.88	10.25/0.13
5	19.74/2.32	7.60/1.41	10.53/1.37	14.58/0.68	20.20/0.10

TABLE V
VIDEO LAYERS' RATE (kb/s)/WEIGHT FOR *City*

Temporal layer / Quality layer	1	2	3	4	5
1	18.41/16.07	7.34/10.15	10.26/10.46	14.35/5.84	20.07/1.08
2	14.46/6.17	5.76/3.90	8.06/4.02	11.27/2.24	15.77/0.41
3	24.74/4.79	9.86/3.03	13.79/3.12	19.29/1.74	26.98/0.32
4	38.74/3.26	15.44/2.06	21.60/2.12	30.21/1.18	42.25/0.22
5	75.68/2.56	30.17/1.62	42.19/1.67	59.01/9.32	82.53/0.17

TABLE VI
VIDEO LAYERS' RATE (kb/s)/WEIGHT FOR *Crew*

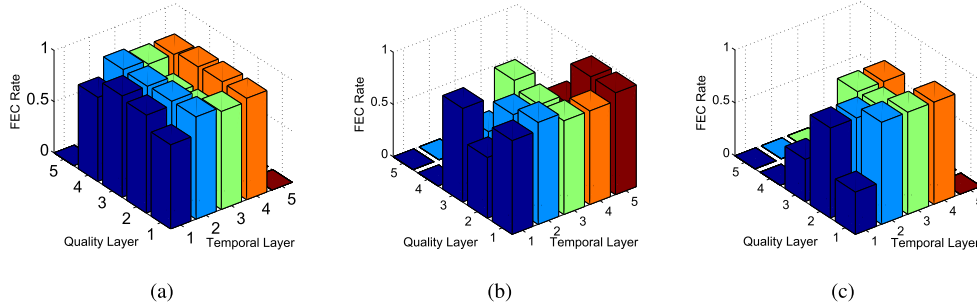| Temporal layer / Quality layer	1	2	3	4	5
1	20.99/13.01	12.43/8.22	19.79/8.48	31.51/4.74	50.18/0.88
2	17.39/6.89	10.30/4.35	16.40/4.49	26.11/2.51	41.57/0.46
3	30.46/5.78	18.04/3.65	28.72/3.77	45.73/2.11	72.81/0.39
4	48.76/4.12	28.87/2.60	45.97/2.68	73.20/1.50	116.54/0.28
5	97.44/3.34	57.70/2.11	91.87/2.18	146.27/1.22	232.89/0.23

(a) (b) (c)

Fig. 5.   FEC rate allocation for video sequence *City* with average bandwidth capacity constraint 200 kb/s over various PER. (a) Average PER = 5%. (b) Average PER = 10%. (c) Average PER = 15%.
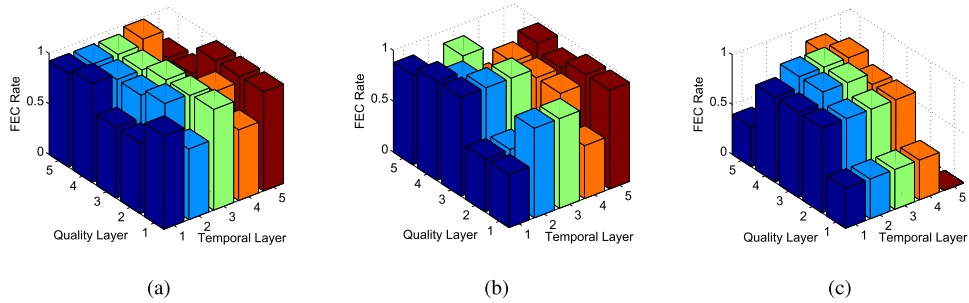

(a) (b) (c)

Fig. 6.   FEC rate allocation for video sequence *City* with average bandwidth capacity constraint 300 kb/s over various PER. (a) Average PER = 5%. (b) Average PER = 10%. (c) Average PER = 15%.

technique for layer extraction and that of the FSSG algorithm for FEC rate allocation. To eliminate the effect of video layer scheduling, we implement the algorithm under the single-antenna scenario. The results are compared with the performance of the following schemes.
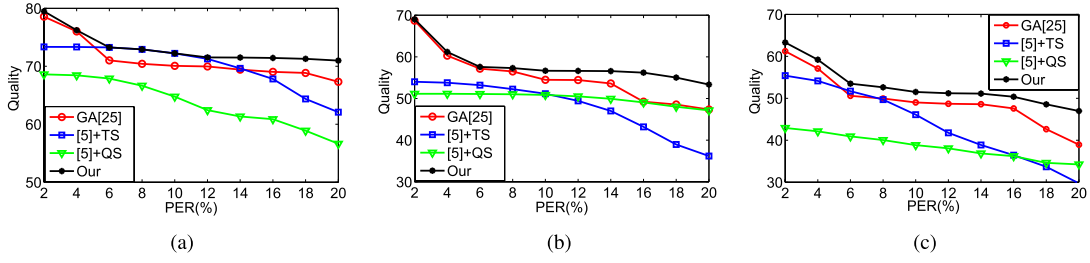
Fig. 7. Quality comparison of the proposed scheme against other schemes under the single-antenna scenario. Different video sequences are tested with bandwidth capacity 400 kb/s. (a) *Akiyo*. (b) *City*. (c) *Crew*.

Fig. 8. Quality comparison of the proposed scheme against other schemes under the single-antenna scenario. Different video sequences are tested with bandwidth capacity 800 kb/s. (a) *Akiyo*. (b) *City*. (c) *Crew*.
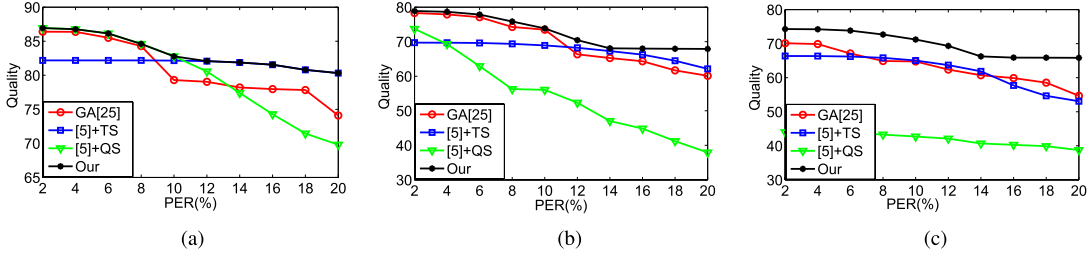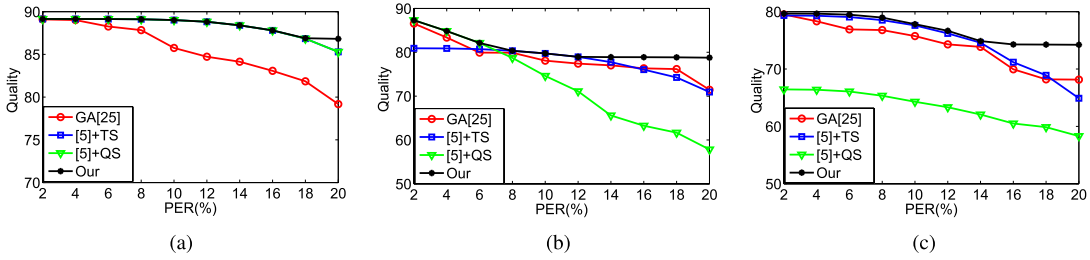

Fig. 9. Quality comparison of the proposed scheme against other schemes under the single-antenna scenario. Different video sequences are tested with bandwidth capacity 1200 kb/s. (a) *Akiyo*. (b) *City*. (c) *Crew*.

1) *GA [25]:* The GA is adopted in [25] to allocate FEC rate for the video layers. However, it does not consider joint rate allocation between the source coding and channel coding rates. When the bandwidth capacity is smaller than the source rate, this scheme becomes unusable. Thus, in our experiments, we modify this scheme so that video layer extraction is implemented by our proposed branch-and-bound algorithm, and then GA is used for FEC rate allocation to derive the bounds for each subproblem.

2) *JSCC [5]:* In the JSCC framework proposed in [5], by adopting the space time block code technique, the MIMO channel is treated as a virtual single channel (spatial diversity) with higher transmission reliability. Besides, it only considers the quality scalability of SVC, and controls the source rate by truncating higher video layers. When considering both quality and temporal scalability, however, this method is not suitable. In our experiments, we extend this JSCC method into two variants: JSCC [5] + TS and JSCC [5] + QS, that is, the layers are truncated in temporal dimension (JSCC [5] + TS) and in quality dimension (JSCC [5] + QS), respectively.

After video layer truncation, we implement our proposed FSSG algorithm for FEC rate allocation.

The results presented in Figs. 7–9 show that our proposed scheme always performs the best under different network conditions for all video sequences. The superior performance mainly comes from two aspects: 1) branch-and-bound based layer extraction and 2) FSSG-based FEC rate allocation. First, since both JSCC [5] + TS and JSCC [5] + QS implement the same FEC rate allocation algorithm with our proposed scheme, the significant performance improvement of our proposed scheme over JSCC [5] + TS and JSCC [5] + QS demonstrates that the proposed branch-and-bound based layer extraction strategy is significantly more effective than the methods by truncating video layers in a certain single dimension. On the other hand, the quality gap between our proposed scheme and GA [25] demonstrates the effectiveness of FSSG algorithm over GA algorithm, as both implement the same layer extraction scheme.

2) Quality Evaluation Under the Multiantenna Scenario: We then evaluate the quality performance of our proposed algorithms under multiantenna scenario. The ES method and our previous JSCC scheme [27] are implemented for
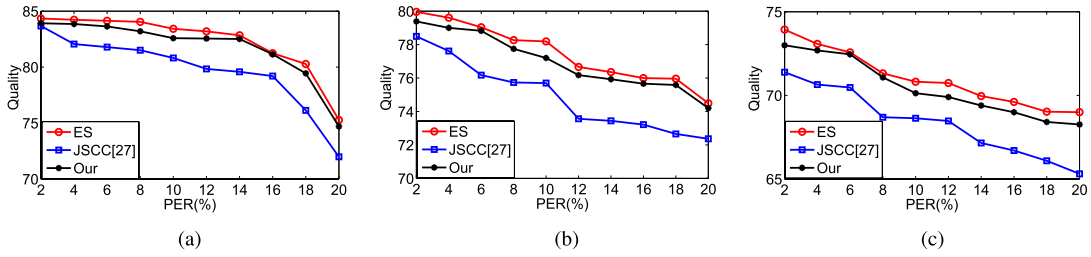
Fig. 10. Quality comparison of the proposed scheme against other schemes under the multiantenna scenario. Different video sequences are tested with average bandwidth capacity 300 kb/s. (a) *Akiyo*. (b) *City*. (c) *Crew*.

Fig. 11. Quality comparison of the proposed scheme against other schemes under the multiantenna scenario. Different video sequences are tested with average bandwidth capacity 400 kb/s. (a) *Akiyo*. (b) *City*. (c) *Crew*.
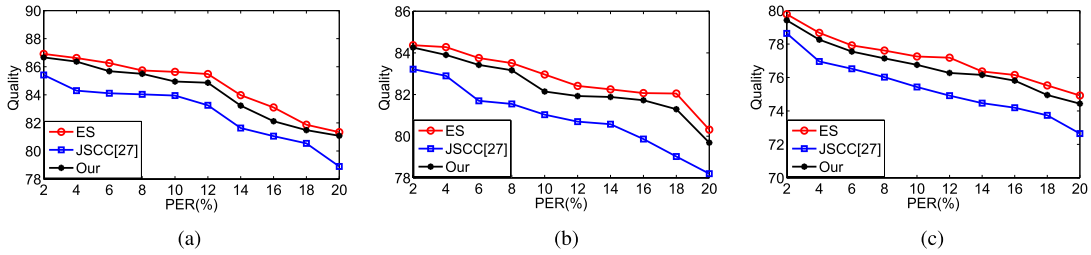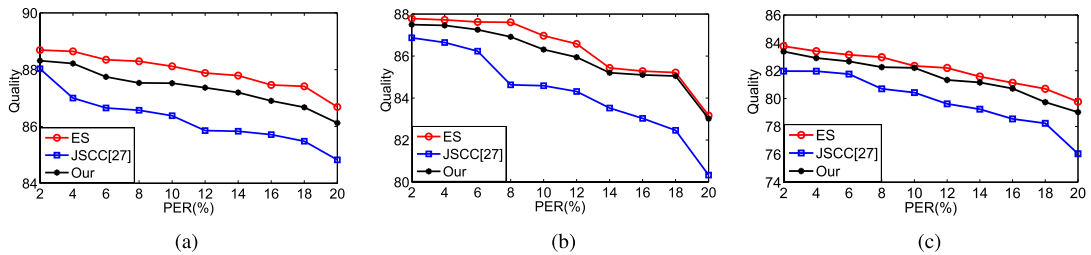

Fig. 12. Quality comparison of the proposed scheme against other schemes under the multiantenna scenario. Different video sequences are tested with average bandwidth capacity 500 kb/s. (a) *Akiyo*. (b) *City*. (c) *Crew*.

comparison. Considering the high computational complexity of ES, it is only implemented on the video layer scheduling, whereas the branching technique is implemented for layer extraction and the FSSG algorithm is implemented for FEC rate allocation. The results in Figs. 10–12 show that our proposed scheme achieves very close quality performance with the ES method and performs better than JSCC [27]. This performance loss of JSCC [27] is mainly because that FEC rate allocation is implemented under the Shannon limits, which generally cannot be achieved in practice. On the other hand, we also show the iteration results in Table VII. In ES, each video layer can be scheduled to one of the four antennas and the optimal scheduling solution is found by searching the whole solution space whose size is 4^{25}, the complexity is awesome. While both our proposed scheme and JSCC [27] reduce the computational complexity drastically and have acceptable iterations for practical implementation as Table VII shows.

On the other hand, for the same average PER, when the bandwidth capacity of single-antenna scenario is set equal to the sum of bandwidth capacity of multiple antennas for multiantenna scenario, the performance under multiantenna scenario is worse than that under single-antenna scenario as Figs. 9 and 10 show. This is because video layers scheduling is

TABLE VII

AVERAGE NUMBER OF ITERATIONS

Methods \ Video Sequences	Akiyo	City	Crew
Exhaustive Search	4^{25}	4^{25}	4^{25}
Our	157	203	249
JSCC[27]	100	132	160

a discrete problem, and some scrap bandwidth of each antenna cannot be effectively utilized for multiantenna scenario. However, from Figs. 9 and 12, we find that though the average bandwidth capacity of multiantenna scenario is much smaller than that of single-antenna scenario, it has better, at least not worse video quality, this also demonstrates the advantage of employing MIMO system for video transmission.

C. Subjective Quality Estimation

Since there is no widely accepted quality metric that assesses quality-temporal tradeoffs, we have done some viewing tests to obtain subjective MOS results so as to demonstrate the effectiveness of our proposed video transmission scheme and solution algorithm. The results are shown in Tables VIII–X. We have tested all the three video

TABLE VIII

MOS RESULTS FOR VIDEO *Akiyo* WITH AVERAGE PER = 10%

Methods \ Average Bandwidth	300kbps	400kbps	500kbps
Original	3.92	3.92	3.92
Our	2.25	2.92	3.12
JSCC[27]	1.92	2.58	2.75

TABLE IX

MOS RESULTS FOR VIDEO *City* WITH AVERAGE PER = 10%

Methods \ Average Bandwidth	300kbps	400kbps	500kbps
Original	3.12	3.12	3.12
Our	1.59	2.67	3.09
JSCC[27]	1.17	1.67	2.6

TABLE X

MOS RESULTS FOR VIDEO *Crew* WITH AVERAGE PER = 10%

Methods \ Average Bandwidth	300kbps	400kbps	500kbps
Original	3.11	3.11	3.11
Our	1.33	2.41	3.10
JSCC[27]	0.83	1.92	2.25

sequences under different bandwidth constraints, and the PER is set to 10%. All the video sequences, including the original video sequences, are tested by 15 viewers. The subjective score is given by all viewers independently and the average results are showed in Tables VIII–X. From the results we can find that our proposed scheme performs much better than the scheme proposed in JSCC [27]. Besides, when the average bandwidth is increasing, the MOS of our proposed scheme is very close to the original video sequences. At last, compared with results in Figs. 7–9 (see the revised paper), we can find that the MOS results have similar characteristics with the results obtained by the model. This also demonstrates the effectiveness of our proposed video transmission scheme and solution algorithm.

VIII. CONCLUSION

In this paper, we have proposed a novel JSCC scheme for UEP-based video transport over MIMO wireless networks. By jointly considering video layer extraction, video layer scheduling, and FEC rate allocation, the end-to-end video distortion can be minimized. We have formulated the problem as a nonlinear integer optimization problem. To reduce the computational complexities of video layer extraction, 2-D FEC rate allocation, and video layer scheduling, we have proposed an efficient branch-and-bound algorithm to solve the optimization problem. Experimental results show that our proposed schemes significantly improve the video quality compared with existing schemes.

APPENDIX A

PROOF OF PROPOSITION 1

Proof: The left inequality is obvious based on the assumption that $J(\mathbf{K}^i, \boldsymbol{\lambda}^u) < J^*$. From (15), we have

$$J(\boldsymbol{\lambda}^u) = \min_{1 \le k_{tq} \le M} J(\mathbf{K}, \boldsymbol{\lambda}^u) \le J(\mathbf{K}^i, \boldsymbol{\lambda}^u). \tag{28}$$

The above is also true at $\boldsymbol{\lambda}^*$

$$J^* = J(\boldsymbol{\lambda}^*) \le J(\mathbf{K}^i, \boldsymbol{\lambda}^*). \tag{29}$$

Therefore, we have

$$
\begin{aligned}
J^* - J(\mathbf{K}^i, \boldsymbol{\lambda}^u) &\le J(\mathbf{K}^i, \boldsymbol{\lambda}^*) - J(\mathbf{K}^i, \boldsymbol{\lambda}^u) \\
&= D(b(t_0, q_0), \mathbf{X}_0, \mathbf{K}^i) + \mathbf{g}^T(\mathbf{K}^i)\boldsymbol{\lambda}^* \\
&\quad - D(b(t_0, q_0), \mathbf{X}_0, \mathbf{K}^i) + \mathbf{g}^T(\mathbf{K}^i)\boldsymbol{\lambda}^u \\
&= (\mathbf{g}^i)^T(\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u) \tag{30}
\end{aligned}
$$

thus the right inequality of inequality (28) is also correct, and the proof is complete. ∎

APPENDIX B

PROOF OF PROPOSITION 2

Proof: The left inequality is obvious according to (16). We now prove the right inequality as follows. From the definition of v_i^u, if $v_i^u > 0$, then

$$J(\mathbf{K}^i, \boldsymbol{\lambda}^u) < J(\boldsymbol{\lambda}^u) + \varepsilon \forall i = 1, 2, \ldots, u. \tag{31}$$

Given the inequality (23), it implies that

$$
\begin{aligned}
J^* - J(\mathbf{K}^i, \boldsymbol{\lambda}^u) &> J^* - J(\boldsymbol{\lambda}^u) - \frac{J^* - J(\boldsymbol{\lambda}^u)}{\alpha} \\
&= \frac{(\alpha - 1)(J^* - J(\boldsymbol{\lambda}^u))}{\alpha}. \tag{32}
\end{aligned}
$$

From Proposition 1 and the condition that $\alpha \ge 2$, for any $i = 1, 2, \ldots, k$, we have

$$\frac{J^* - J(\boldsymbol{\lambda}^k)}{\alpha} < \frac{(\mathbf{g}^i)^T(\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^k)}{(\alpha - 1)} \le (\mathbf{g}^i)^T(\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^k). \tag{33}$$

Finally, considering that

$$v_i^u \ge 0, \text{ and } \sum_{i=1}^u v_i^u = 1 \tag{34}$$

we obtain

$$\frac{J^* - J(\boldsymbol{\lambda}^u)}{\alpha} < \sum_{i=1}^u v_i^u(\mathbf{g}^i)^T(\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u) = (\tilde{\mathbf{g}}^u)^T(\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u). \tag{35}$$

Thus, Proposition 2 is proved. ∎

APPENDIX C

PROOF OF THEOREM 1

Proof: From (20), we have

$$
\begin{aligned}
\|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^{u+1}\|^2 &= \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u - s^u\tilde{\mathbf{g}}^u\|^2 \\
&= \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u\|^2 - 2s^u(\tilde{\mathbf{g}}^u)^T(\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u) + (s^u)^2\|\tilde{\mathbf{g}}^u\|^2. \tag{36}
\end{aligned}
$$

Then, combined with Proposition 2 to be consistent, the above yields

$$
\begin{aligned}
\|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^{u+1}\|^2 &< \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u\|^2 - 2s^u \frac{J^* - J(\boldsymbol{\lambda}^u)}{\alpha} + (s^u)^2\|\tilde{\mathbf{g}}^u\|^2 \\
&= \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u\|^2 - s^u\left(2\frac{J^* - J(\boldsymbol{\lambda}^u)}{\alpha} - s^u\|\tilde{\mathbf{g}}^u\|^2\right) \\
&\le \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^u\|^2. \tag{37}
\end{aligned}
$$

The last term of the above inequality comes from that $2J^* - J(\boldsymbol{\lambda}^u)/\alpha - s^u\|\tilde{\mathbf{g}}^u\|^2 \ge 0$ according to the range of stepsizes in (24). Thus, Theorem 1 is proved. ∎

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[2] M. Srinivasan and R. Chellappa, "Adaptive source-channel subband video coding for wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 9, pp. 1830–1839, Dec. 1998.

[3] 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.

[4] L. Kondi, F. Ishtiaq, and A. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1043–1052, Sep. 2002.

[5] H. Xiao, Q. Dai, X. Ji, and W. Zhu, "A novel JSCC framework with diversity-multiplexing-coding gain tradeoff for scalable video transmission over cooperative MIMO," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 7, pp. 994–1006, Jul. 2010.

[6] N. Seshadri, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[7] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[8] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[9] M. K. Jubran, M. Bansal, and L. P. Kondi, "Low-delay low-complexity bandwidth-constrained wireless video transmission using SVC over MIMO systems," *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 1698–1707, Dec. 2008.

[10] D. Song and C. W. Chen, "Scalable H.264/AVC video transmission over MIMO wireless systems with adaptive channel selection based on partial channel information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1218–1226, Sep. 2007.

[11] M. Li, Z. Chen, and Y. Tan, "Cross-layer optimization for SVC video delivery over the IEEE 802.11e wireless networks," *J. Vis. Commun. Image Represent.*, vol. 22, no. 3, pp. 284–296, Apr. 2011.

[12] M. Li, Z. Chen, and Y. Tan, "Scalable resource allocation for SVC video streaming over multiuser MIMO-OFDM networks," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1519–1531, Nov. 2013.

[13] C.-F. Tsai, C.-J. Chang, F.-C. Ren, and C.-M. Yen, "Adaptive radio resource allocation for downlink OFDMA/SDMA systems with multimedia traffic," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1734–1743, May 2008.

[14] M. Li, Z. Chen, and Y. Tan, "QoE-aware resource allocation for scalable video transmission over multiuser MIMO-OFDM systems," in *Proc. IEEE Vis. Commun. Image Process.*, Oct. 2012, pp. 1–6.

[15] M. Li, Z. Chen, and Y.-P. Tan, "Joint power allocation and bit loading for enhanced SVC video downlink transmissions over SDMA/OFDMA networks," in *Proc. IEEE Conf. Multimedia Expo*, Jul. 2011, pp. 1–6.

[16] H. Cai, B. Zeng, G. Shen, and S. Li, "Error-resilient unequal protection of fine granularity scalable video bitstreams," in *Proc. IEEE Int. Conf. Commun.*, vol. 3. Jun. 2004, pp. 1303–1307.

[17] M. van der Schaar and H. Radha, "Unequal packet loss resilience for fine-granular-scalability video," *IEEE Trans. Multimedia*, vol. 3, no. 4, pp. 381–394, Dec. 2001.

[18] C. E. Costa, Y. Eisenberg, F. Zhai, and A. K. Katsaggelos, "Energy efficient wireless transmission of MPEG-4 fine granular scalable video," in *Proc. IEEE Int. Conf. Commun.*, vol. 5. Jun. 2004, pp. 3096–3100.

[19] J. Kim, R. M. Mersereau, and Y. Altunbasak, "A multiple-substream unequal error-protection and error-concealment algorithm for SPIHT-coded video bitstreams," *IEEE Trans. Image Process.*, vol. 13, no. 12, pp. 1547–1553, Dec. 2004.

[20] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 819–828, Jun. 2000.

[21] T. Fang and L.-P. Chau, "A novel unequal error protection approach for error resilient video transmission," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2005, pp. 4022–4025.

[22] X. K. Ang *et al.*, "Unequal loss protection for robust transmission of motion compensated video over the internet," *Signal Process., Image Commun.*, vol. 18, pp. 157–167, Mar. 2003.

[23] F. Marx and J. Farah, "A novel approach to achieve unequal error protection for video transmission over 3G wireless networks," *Signal Process., Image Commun.*, vol. 19, no. 4, pp. 313–323, Apr. 2004.

[24] C.-L. Huang and S. Liang, "Unequal error protection for MPEG-2 video transmission over wireless channels," *Signal Process., Image Commun.*, vol. 19, no. 1, pp. 67–79, Jan. 2004.

[25] T. Fang and L.-P. Chau, "GOP-based channel rate allocation using genetic algorithm for scalable video streaming over error-prone networks," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1323–1330, Jun. 2006.

[26] E. Maani and A. K. Katsaggelos, "Unequal error protection for robust streaming of scalable video over packet lossy networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 3, pp. 407–416, Mar. 2010.

[27] X. Zhang, C. Zhou, and Z. Guo, "A novel JSCC scheme for scalable video transmission over MIMO systems," in *Proc. 19th IEEE Int. Conf. Image Process.*, Oct. 2012, pp. 2277–2280.

[28] Y. Wang, Z. Ma, and Y.-F. Ou, "Modeling rate and perceptual quality of scalable video as functions of quantization and frame rate and its application in scalable video adaptation," in *Proc. 17th Int. Packet Video Workshop*, May 2009, pp. 1–9.

[29] H. Mansour, Y. P. Fallah, P. Nasiopoulos, and V. Krishnamurthy, "Dynamic resource allocation for MGS H.264/AVC video transmission over link-adaptive networks," *IEEE Trans. Multimedia*, vol. 11, no. 18, pp. 1478–1491, Dec. 2009.

[30] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.

[31] R. Feghali, F. Speranza, D. Wang, and A. Vincent, "Video quality metric for bit rate control via joint adjustment of quantization and frame rate," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 441–446, Mar. 2007.

[32] J.-S. Lee, F. De Simone, T. Ebrahimi, N. Ramzan, and E. Izquierdo, "Quality assessment of multidimensional video scalability," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 38–46, Apr. 2012.

[33] H. Sohn, H. Yoo, W. De Neve, C. S. Kim, and Y.-M. Ro, "Full-reference video quality metric for fully scalable and mobile SVC content," *IEEE Trans. Broadcast.*, vol. 56, no. 3, pp. 269–280, Sep. 2010.

[34] R. Stapenhurst, J. Lu, and D. Agrafiotis, "Performance evaluation of objective video quality metrics on mixed spatiotemporal resolution content," in *Proc. 20th IEEE Conf. Image Process.*, Sep. 2013, pp. 64–68.

[35] Z. Ma, M. Xu, Y.-F. Ou, and Y. Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 5, pp. 671–682, May 2012.

[36] H. Hu, X. Zhu, Y. Wang, R. Pan, J. Zhu, and F. Bonomi, "QoE-based multi-stream scalable video adaptation over wireless networks with proxy," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2012, pp. 7088–7092.

[37] M. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA, USA: Freeman, 1979.

[38] Y. Yoon, Y.-H. Kim, and B.-R. Moon, "An evolutionary Lagrangian method for the 0/1 multiple knapsack problem," in *Proc. GECCO*, Jun. 2005, pp. 629–635.

[39] K. C. Kiwiel, "The efficiency of subgradient projection methods for convex optimization, part II: Implementations and extensions," *J. Control Optim.*, vol. 34, no. 2, pp. 660–676, 1996.
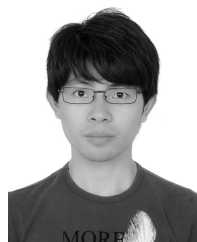
Chao Zhou received the Ph.D. degree from the Institute of Computer Science and Technology, Peking University, Beijing, China, in 2014.

He is a Senior Research Engineer with the Media Technology Laboratory, Central Research Institute, Huawei Technologies Company, Ltd., Beijing, China. His research interests include video streaming, joint source channel coding, rate adaptation, HTTP streaming, and multimedia communication and processing.

Dr. Zhou received the Best Student Paper Awards from the IEEE Visual Communications and Image Processing Conference in 2012.

Chia-Wen Lin (S'94–M'00–SM'04) received the Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000.

He was with Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, from 1992 to 2000, and with the Department of Computer Science and Information Engineering, National Chung Cheng University, Minhsiung, Taiwan, from 2000 to 2007. He is currently a Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU. He is also an Adjunct Professor with the Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan. His research interests include image and video processing, and video networking.

Dr. Lin's paper received the Top 10% Paper Award by the IEEE International Workshop on Multimedia Signal Processing in 2013, and he was a recipient of the Young Investigator Award by the Visual Communications and Image Processing in 2005, the Young Faculty Awards by CCU in 2005, and the Young Investigator Awards by the National Science Council, Taiwan, in 2006. He has served as an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, *IEEE Multimedia Magazine*, and *Journal of Visual Communication and Image Representation*. He is also an Area Editor of the *EURASIP Signal Processing: Image Communication*. He is currently the Chair of the Multimedia Systems and Applications Technical Committee of the IEEE Circuits and Systems Society. He served as the Technical Program Co-Chair of the IEEE International Conference on Multimedia and Expo (ICME) in 2010, and the Special Session Co-Chair of the IEEE ICME in 2009.

Xinggong Zhang (M'11) received the B.S. degree from Harbin Institute of Technology, Harbin, China, in 1995; the M.S. degree from Zhejiang University, Hangzhou, China, in 1998; and the Ph.D. degree from Peking University, Beijing, China, in 2011.

He was a Senior Research Staff Member with Founder Research China. He is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. His general research interests lie in multimedia networking and video communications. His current research interests include video conferencing, dynamic HTTP streaming, and content-centric networking.

Dr. Zhang received the Ministry of Education Science and Technology Progress Award in 2006, and the National Science and Technology Award in 2007.

Zongming Guo was born in 1966. He received the B.Sc. degree in mathematics and the M.Sc. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively.

He is the Deputy Dean of the Institute of Computer Science and Technology with Peking University. He has authored over 80 technical articles in refereed journals and proceedings in multimedia, image and video compression, image and video retrieval, and watermarking. His research interests include network video technology, including streaming media technology, IPTV, and mobile multimedia.

Dr. Guo led the Research and Developing Team, which received the State Administration of Radio Film and Television Award, the Ministry of Education Science and Technology Progress Award, and the National Science and Technology Award in 2004, 2006, and 2007, respectively, and the Government Allowance granted by the State Council in 2009.