Joint Source-Channel Optimization for Layered Video Broadcasting to Heterogeneous Devices

Wen Ji, *Member, IEEE,* Zhu Li, *Senior Member, IEEE,* Yiqiang Chen, *Member, IEEE*

***Abstract*—Heterogeneous QoS video broadcast over wireless network is a challenging problem, where the demand for better video quality needs to be reconciled with different display size, variable channel condition requirements. In this paper, we present a framework for broadcasting scalable video to heterogeneous QoS mobile users with diverse display devices and different channel conditions. The framework includes joint video source-channel coding and optimization. First, we model the problem of broadcasting a layered video to heterogeneous devices as an aggregate utility achieving problem. Second, based on scalable video coding, we introduce the temporal-spatial content distortion metric to build adaptive layer structure, so as to serve mobile users with heterogeneous QoS requirements. Third, joint Fountain coding protection is introduced so as to provide flexible and reliable video stream. Finally, we use dynamic programming approach to obtain optimal layer broadcasting policy, so as to achieve maximum broadcasting utility. The objective is to achieve maximum overall receiving quality of the heterogeneous QoS receivers. Experimental results demonstrate the effectiveness of the solution.**

***Index Terms*—Heterogeneous QoS, video broadcasting, SVC, Fountain coding, elastic video, optimization.**

I. Introduction

A. Motivation

Joint source-channel coding (JSCC) is an effective approach in designing error-resilient wireless video broadcasting systems [1]- [9]. In recent years, JSCC attracts increasing interests in both research community and industry because it shows better results in robust layered video transmission over error-prone channels. In [10] and [11], good review of various techniques available during these years may be found. However, there are still many open problems in terms of how to serve heterogeneous users with diverse screen features and variable reception performances in wireless video broadcast system. One particular challenging problem of this heterogeneous QoS video provision is: the users would prefer flexible video with better quality to match their screens, at the same time, the video stream could be reliable received. The main technical difficulties are as follows:

- A distinctive characteristic in current wireless broadcast system is that the receivers are highly heterogeneous in terms of their terminal processing capabilities and available bandwidths. In source side, scalable video coding (SVC) has been proposed to provide an attractive solution to this problem. However, in order to support flexible video broadcasting, the scalable video sources need to provide adaptation ability through a variety of schemes, such as scalable video stream extraction (e.g. [12]- [16]), layer generation with different priority(e.g. [17]- [20]) and summarization (e.g. [21]), before they can be transmitted over the error-prone networks.
- Video content in different scalable domain have different rate-distortion (R-D) characteristics, e.g. 15fps@D1 and 30fps@CIF format video show different R-D results in temporal and spatial directions. Imperfectly, classical information theory in term of quality fidelity can not efficiently measure the temporal or spatial scalability. In wireless broadcasting system, the type of heterogeneous devices access in terms of different display terminal characteristic and different bandwidth requirement should be taken into consideration when optimize the broadcasting system.
- Since layered video data is very sensitive to transmission failures, the transmission must be more reliable, have low overhead and support large numbers of devices with heterogeneous characteristics [22]. In broadcast and multicast network, conventional schemes such as adaptive retransmission have their limitations, for example, retransmission may lead to implosion problem [23]. Forward error correction (FEC) (e.g. [24] [25]) and unequal error protection (UEP) are employed to provide the quality of service support for video transmission. However, in order to obtain as minimum investment as possible in broadcasting system deployment, server-side must be designed more scalable, reliable, independent, and support vast number of autonomous receivers. Suitable FEC approaches are expected such that can eliminate the retransmission and lower the unnecessary receptions overhead at each receiver-side.

Conventionally, the joint source and channel coding are designed with seldom consideration in heterogeneous characteristics, and most of the above challenges are ignored in practical video broadcasting system. This leads to the need for heterogeneous QoS video provision in broadcasting network. This paper presents the point of view to study the hybrid-scalable video from new quality metric so as to support users' heterogeneous requirements. In this paper, we propose a framework for joint scalable video and channel coding, adaptive layer generation, and joint optimization. In video generation phase, the scalable video with heterogeneous char-

This work was supported in part by the National Natural Science Foundation of China (61001194), and Hong Kong Research Grant Council (RGC) and Hong Kong Polytechnic University new faculty grant.

W. Ji, Y. Chen are with the Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academic of Sciences, P.R. China (e-mail: jiwen@ict.ac.cn; yqchen@ict.ac.cn).

Z. Li is with the Futurewei Technology, USA (e-mail: zhu.li@ieee.org).

acteristics support is encoded along the max users' satisfaction direction. Through suitable FEC approach, it allows any number of heterogeneous receivers to acquire video content with optimal efficiency. Based on components decomposition, the overall broadcasting efficiency is modeled by a generalized broadcasting utility maximization problem. Each video layer corresponds to a decomposed subproblem, and the interfaces among layers are quantified as functions of the optimization variables coordinating the subproblems. Finally, joint coding and optimization solution is proposed to satisfy broadcasting users' heterogeneous requirements.

B. Related work

Generally, the joint video source-channel coding problem includes joint coding and optimal rate allocation design between video coding and channel coding, which provides various protection level to the video data according to its level of importance and channel conditions. Most of related work in video transmission focus on: 1) finding an optimal bit allocation between video coding and channel coding, such as in [8] [9]; 2) designing the video coding to achieve the target source rate under given channel conditions, such as in [3]; 3) designing the channel coding to achieve the required robustness, such as using low-density parity check (LDPC) [4], Turbo [5], Reed-Solomon (RS) [6], and Fountain [7] codes; 4) designing joint optimization framework, including all available error control components together with error concealment and transmission control, to improve global system performance, such as in [2] [11].

Recent innovations in video coding offer new perspectives on variable resolution video provision [26]. The advances in SVC [27] provide great flexibility for scalable rate and quality video in joint coding applications especially in video broadcasting over wireless networks scenario [28]. SVC can generate multi-scalable video by utilizing manipulations at different domains (temporal, spatial and quality) in order to meet diverse user preferences. There has been a large amount of activities in research and standard development in this area. In [14], a low complexity joint SVC and LDPC coding methodology is proposed to minimize the end-to-end distortion over a lossy channel. This work demonstrates the effectiveness of such a JSCC approach for protecting SVC-coded video in wireless transmission. For the problem of scalable video streaming transmission over lossy channels, the common idea is still to explore the scalability provided by SVC, use unequal error/loss protection (UEP/ULP) and hybrid ARQ-FEC, to give important layer more protection [29] [30]. Furthermore, through cross layer optimization schemes [31] [32], layered coded video can provide better adaptive QoS in wireless local area networks. However, retransmission and redundant receptions problems in multi-receiver cases still limit the deployment in broadcasting scenario. Consequently, many researchers begin to find another way.

Since reliable video transmission in wireless network depends on its major technique, FEC, the performance of joint channel coding control is critical. From the view of channel coding, generally, LDPC, Turbo, RS, etc., belong to fixed-rate channel code constructions. When combined with variable-length video coding, the channel code rate needs to be carefully chosen to match video encoder as well as channel conditions [6]. Besides, since the capabilities of video terminals in terms of display size and access bitrates have achieved remarkable improvements, higher-level video quality and higher heterogeneous requirement become critical challenges in broadcasting scenario. Consequently, in multimedia broadcast/multicast services (MBMS) scenario, research community and industry began to recommend application layer FEC based on Fountain codes [22] [33] [38] [40]. The reasons lie in that 1) Fountain codes are rateless, this means the number of encoded packets that can be generated from server-side is potentially limitless [34], thus, this method can serve a wide range of heterogeneous receivers; 2) Fountain codes on lossy channels need not assume any knowledge of the channel, such that they are very suitable candidates for retransmission-limited applications such as transmitting data on multicast/broadcast circumstance [35]; 3) Fountain codes exhibit linear encoding and decoding complexity while still keep low coding overhead [36] [37], such that the deployments of both server-side and receiver-side become much easier. Therefore, Fountain codes [38] [39] have been employed for multimedia broadcast/multicast service in universal mobile telecommunication system [40]. Schierl *et al* [37] showed that using Fountain codes is particularly suitable to serve heterogeneous receivers with different display size, which is a typical case when SVC-coded video in mobile transmission circumstance. In unicast scenario, Ahmad *et al* [41] developed a flexible transmission framework with application-layer ACKs from receivers instead of data retransmission, and showed that using Fountain codes as the FEC method in video transmission can outperform previous FEC schemes over a wide range of transmission bit rates. Besides, Wagner *et al* [42] provided the solution when scalable video is from multiple servers cases. In multicast/broadcast scenario, Luby *et al* [38] provided basic end-to-end protocols design. Cataldi *et al* [43] and Ahmad *et al* [44] focused on how to design Fountain codes (Raptor, and LT) so as to provide unequal loss protection. [43] proposed slide-window method while [44] introduced appropriate degree distribution design. In a word, appropriate rate assignment of the Fountain codes to different video layers allows for scalable layer-specific video provision.

These show great advantages in using Fountain codes to provide flexible error protection in scalable video coding, especially in adapting the source coding at the video server side to receivers with heterogeneous characteristics. However, much work remains to be done. We need to further consider source adaptation and overall users' satisfaction maximum problem in broadcast network so as to provide the *flexible* and *reliable* video to heterogeneous receivers.

C. Summary of Contributions

In this paper, we focus on using joint scalable video and Fountain coding to solve the challenge in video broadcasting system: broadcasting to heterogeneous devices, with different display resolution requirements and working in variable channel conditions. We start by formulating the overall broadcasting quality achieving problem as a hiberarchy optimization

[45] problem. Through decompose utility composition, the deduction is carried out in a systematic way, which involves adaptive video layer structure based on hybrid temporal-spatial metric, together with embedded Fountain coding. The objective is to find an optimal solution in maximizing the overall users' utility so as to support many-effort heterogeneous devices with better-effort received quality. The main results and contributions of this paper include:

1) *Framework*: This paper presents a framework for broadcasting video content to heterogeneous users. Through joint temporal-spatial layer generation and rateless erasure protection, the framework can satisfy the requirements from multi-user's diverse display devices and the requirements from multi-user's variable channel characteristics. The framework includes hybrid temporal-spatial scalable video generation, joint Fountain coding and joint optimization for heterogeneous multi-user video broadcasting.

2) *Utility-driven video broadcast*: This paper formulates overall users broadcasting quality achieving problem as a utility achieving model. The main idea is that for the broadcasting video, there is an associated utility function which can be a measurement of the user's heterogeneous-QoS performance. The utility is defined as the user's video quality and satisfaction level with respect to the allocated bandwidth. The utility function is built from hybrid temporal summarization and spatial preference metric, which is deduced from content rate-distortion model. Heterogeneous devices characteristics are embedded into utility parameters.

3) *Solution*: Based on the framework, we formulate the broadcasting policy through progressively generating layered scalable video and corresponding joint Fountain codes. We formulate the policy over a finite number of layer decision stages, and use dynamic programming (DP) to get the maximum overall utility of the whole broadcasting system.

4) *Performance Evaluation*: we take a progressive approach in simulations and carry out simulations over typical broadcast scenarios. Simulation results show that the proposed solution can maximize the total users' utility. By simulations, we show the effect of better-effort quality in many-effort users when video broadcast system serves to heterogeneous devices.

The rest of the paper is organized as follows. In section II, we describe the system model and general framework, analyze the broadcasting utility in video broadcasting among heterogeneous user groups and discuss how to deduce utility function. In section III, we present the solution of utility maximization in video broadcasting, which is based on DP. The experimental results and analysis are provided in section IV. Finally, the conclusion and future work are presented in section V.

II. System Model and Framework

A. Basic model

Consider a single cell broadcasting network, broadcast a layered video to a set $\mathcal{N}$ of heterogeneous users. We characterize the QoS of user $n, n \in \mathcal{N}$ by a utility function $U_n(r)$, which is a function of the quality of delivered video. Let N be the number of whole users, there is $N = |\mathcal{N}|$. In broadcasting system, $U_n(r)$ is often assumed to be increasing and dependent on the source rate r ($r \geq 0$). It can be defined in several forms, i.e. inverse to distortion [46], or positive to user satisfaction [47] [48] [49]. Associated with the layered video source, such as scalable video coding [50], or other methods which support layer transmission [51], it becomes the summation of all scalable layers $\sum_{l=1}^{L} U_{n,l}(r_l)$. Under wireless broadcast networks, the reconstructed video at the receiver usually differs from that at the encoder due to error-prone channel. Thus, the expected end-to-end utility (or distortion) is function of coded layer data as well as its probability of loss. In this work, we utilize joint source-channel coding for the effective selection of each layer coding rate and corresponding error protection rate that will allow for the most amount of overall broadcasting utility U_{system} for a given available bandwidth R. This problem is formulated as

$$\text{maximize} \quad U_{system} = \sum_{n=1}^{N} \sum_{l=1}^{L} \Big(\prod_{i=1}^{l} P\{r_i, \gamma_i, n, i\} \Big) U_{n,l}(r_l) \tag{1}$$

$$\text{subject to} \quad \sum_{i=1}^{L} (r_i + \gamma_i) \leq R \tag{2}$$

$$\boldsymbol{r} \geq 0, \boldsymbol{\gamma} \geq 0 \tag{3}$$

Here, $\boldsymbol{r} = [r_1, ..., r_L]$ and $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_L]$ are in R_+^L, which represent the video source rate and corresponding error protection rate in each layer. $U_{n,l}(r_l)$ denotes the corresponding lth layer utility for user n. $P\{r_i, \gamma_i, n, i\}$ represents the correct reception probability of user n when using γ_i to protect layered video r_i.

Remark: Since compressed video data has close correlation, video stream is usually transmitted layered and progressively. For the user n, $U_{n,l}(r_l)$ is achievable only when: 1) current layer can be received successfully when additional error protection part is no less than that of the user's channel condition requirement, that is $P\{r_i, \gamma_i, n, i\}$; 2) the anterior layers have been received successfully because the decoding and reconstruction of current layer depend on its previous layers, that is $\prod_{i=1}^{l} P\{\cdot\}$. For the whole broadcasting system, 3) the received video quality will be improved when more video layers are broadcasted reliably, that is $\sum_{l=1}^{L}(\cdot)U_{\cdot,l}(r_l)$; 4) the broadcasting utility is summarized as the aggregate utility of all the users, because the number of users aggregated and their experience of the video service reflect the performance of broadcasting system, that is $\sum_{n=1}^{N}(\cdot)U_{n,\cdot}(\cdot)$.

B. Providing reliability by joint Fountain coding

In this work, we use Fountain codes to provide flexible error protection in layered video data. We consider the downlink of a wireless broadcasting system with N users. The channel conditions are time-varying and modeled by a stochastic channel state vector $\mathbf{e} = (\varepsilon_1, ..., \varepsilon_N)$, where ε_n represents the stochastic result of nth user. Typically in practical broadcast system, the transmitter has no exact channel state information from receivers, however, some estimate of packet loss probability at each receiver is usually available, i.e., via empirical estimation, channel quality feedback [52], or application-layer ACKs periodically [53]. Then, when lth layer broadcasts to users

in rate r_l, the probability ε_n of loss this transmission to user n can be calculated through the model [54] [55]:

$$\varepsilon_n = \text{Prob}\Big[\frac{1}{2}(r_l+\gamma_l)\log(1+\frac{P}{N}) \le r_l(1+\delta)|e_n\Big] \quad (4)$$

$$= \text{Prob}\Big[(r_l+\gamma_l) \le \frac{2r_l(1+\delta)}{\log(1+\frac{P}{N})}\Big|e_n\Big] = F_{x|e_n}(\gamma_l|e_n) \quad (5)$$

where $F_{x|e_n}$ denotes the cumulative probability density function (CDF) of the channel state, conditioned on the estimated channel state e_n. $\frac{P}{N}$ is the transmission signal to noise ratio in transmitter. Based on this formulation [54], the conditional CDF can be empirically determined through channel measurements.

From the view of Fountain coding side, appropriate Fountain code can provide source data more reliable and matched protection level, in which the protection rate is a function of source rate and channel condition. The determinable expression of the function follows the Fountain coding principle: with an appropriate design [56], for user n with ε_n, which complies with $F_{x|e_n}$, the r_l message can be recovered at least when the erasure protection part γ_l satisfies

$$\gamma_l \ge r_l\Big(\frac{1+\delta}{1-\varepsilon_n} - 1\Big) \quad (6)$$

where δ is the overhead in Fountain code design. For the determinable number of input symbols k_c for a Fountain code unit, it can be measured by:

$$\delta = \begin{cases} O(\log(k_c)/k_c), & \text{for ideal distribution} \\ O(\log^2(k_c)/\sqrt{k_c}), & \text{for Soliton distribution} \end{cases} \quad (7)$$

Then, the number of correct reception users of layer r_l can be adjusted when allocate different γ_l.

C. From temporal-spatial content distortion to utility

In networks that provide heterogeneous QoS guarantees, utility is defined as the satisfaction level of a user with respect to heterogeneous characteristics. Since conventional quality scalable methods are not widely adopted in deployed systems and clients, in this work, we target hybrid temporal and spatial scalability of video stream. Hence, user satisfaction is measured in terms of the received spatial and temporal quality, corresponding to user terminal feature with different display resolution, and the other user terminal feature with different reception performance under variable channel condition. Since conventional rate-distortion in term of PSNR/MSE can not measure temporal and spatial scalable cases [57], we utilize temporal-spatial content rate-distortion metric [20] to measure the scalable layer structure. This content rate-distortion relation is built from the temporal scalability metric [21] in term of summarization, together with user preference in spatial domain, and shows effective especially when broadcasting system serves heterogeneous devices [58] [59].

Assume a GOP contains $V = \{f_1, ..., f_M\}$ frames, where M is the GOP size. These frames are in the highest temporal level $\eta_{\max}$ and the largest spatial level $\nu_{\max}$. Assume scalable video sub-stream is in temporal level ν and spatial level η, and contains $m, m \le M$ frames. In temporal domain, let Λ be the dropped frame set relative to V, there is $\Lambda_{\eta,\nu} = \{f_1^{Te}, ..., f_{M-m}^{Te}\}$. Notice, if $m = M, \Lambda = \varnothing$. From the view of user side, define the playback sequence as $V^* = \{f_1^*, ..., f_M^*\}$. If the user only received the frames in temporal level η, then the lost frames in $\Lambda_{\eta,\nu}$ will be replaced or concealed by the the nearest frames which are in temporal level η. A mapping relation $f_i^* = f_\theta$ when frame copy is used in decoder, where θ is the nearest frame to i, or other function relations between f_i^* and f_θ when error concealment techniques are used in the decoder. The resulting video rate reduction and distortion mainly rely on the drop set $\Lambda_{\eta,\nu}$. We use temporal rate-summarization distortion [21] to represent the temporal-level distortion.

Definition 1: (temporal content R-D [21]) Given the frame drop set $\Lambda_{\eta,\nu}$, for each spatial level ν, the rate and distortion in temporal domain are:

$$\begin{cases} R_{\eta,\nu}(\eta) &= \frac{1}{M}\sum_{i\in(V_\nu-\Lambda_{\eta,\nu})} r_\nu(f_i) \\ D_{\eta,\nu}(\eta) &= \frac{1}{M}\sum_{i\in\Lambda_{\eta,\nu}} d(f_i, f_i^*) \end{cases} \quad (8)$$

Herein, we use the video summary to represent the content information. The detailed computation is given by [21]. In a short, $d(\cdot)$ is computed as the principle component analysis (PCA) distance. $d(\cdot,\cdot)$ is the distortion between two frames' summary[1]. $r_\nu(f_i)$ is the rate of frame f_i in spatial level ν. For example, when user just received 15fps video while original video is 30fps video, the reconstructed frames lose consecutive summary, and might cause severe damage for user understanding the original video content. We summarize this as the temporal distortion, which can be measured by (8).

In spatial domain, define $\Psi_{\eta,\nu} = \{f_1^{'}, ..., f_M^{'}\}$ as the playback frames in spatial level ν and temporal level η. If the user only received the frames in spatial level ν, that is the reconstructed video will be stretched, from received frame size in term of height $f_h^{'}$ and width $f_w^{'}$, to original size f_h and f_w. The resulting video rate reduction mainly relies on the resolution downsampling set $\Psi_{\eta,\nu}$. We use expected spatial distortion to express the video spatial quality from the expectation of user preference.

Definition 2: (spatial content R-D [20]) Given the frame downsampling set $\Psi_{\eta,\nu}$, the rate and distortion in spatial domain are:

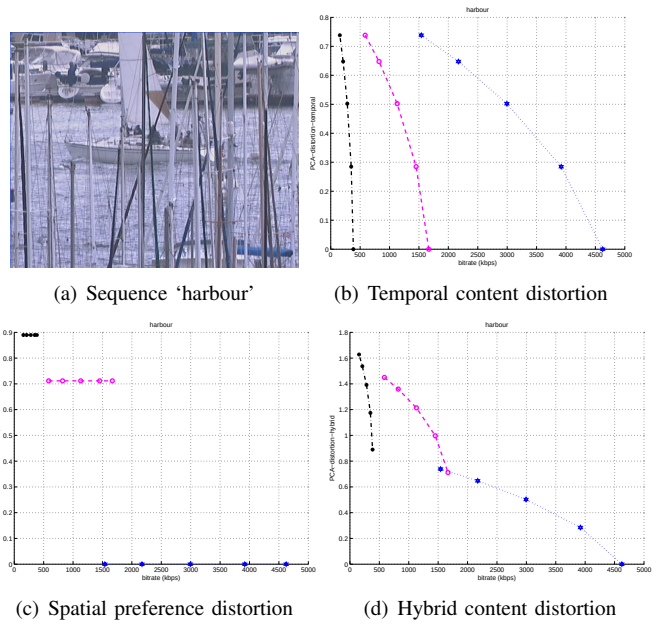
$$\begin{cases} R_{\eta,\nu}(\nu) &= \frac{1}{M}\sum_{i\in\Psi_{\eta,\nu}} r_\eta(f_i) \\ D_{\eta,\nu}(\nu) &= \frac{1}{M}\sum_{i\in\Psi_{\eta,\nu}}\Big(d(f_i) - \frac{f_w^{'}}{f_w}\cdot\frac{f_h^{'}}{f_h}\cdot d(f_i^{'})\Big) \end{cases} \quad (9)$$

where $d(\cdot)$ reflects the frame summary and is also computed as the PCA distance, which is consistent with definition 1.

We measure the utility with a function of received video quality, as measured by hybrid spatial-temporal content distortion model.

Definition 3: (Hybrid temporal-spatial utility [20]) Let (η, ν) be the maximum received temporal and spatial level of a user, and all previous levels in (i, j), $i \in [1, \eta]$, $j \in [1, \nu]$ are received correctly. Then, the utility is:

[1] In a brief, the algorithm is: first, project the video frames through PCA to a subspace that preserves most information while further reduces the correlation among temporal frames; second, select the desired number of dimensions with the largest eigenvalues for all projected data, then, compute each frame's summary through weight norm, that is $d(f_i)$; finally, compute the difference between temporal frames, that is $d(f_i, f_i^*)$.



(a) Sequence 'harbour' (b) Temporal content distortion

(c) Spatial preference distortion (d) Hybrid content distortion

Fig. 1. Content rate-distortion in term of summarization metric

$$\begin{cases} \text{Temporal:} & u_\eta = D^{max}_{\eta_{\max}} - D_{\eta,\nu}(\eta) \\ \text{Spatial:} & u_\nu = D^{max}_{\nu_{\max}} - D_{\eta,\nu}(\nu) \\ \text{Hybrid:} & \ddot{u} = \alpha_u(D^{max}_{\eta_{\max}} - D_{\eta,\nu}(\eta)) + \beta_u(D^{max}_{\nu_{\max}} - D_{\eta,\nu}(\nu)) \end{cases} \tag{10}$$

where $(\eta_{\max}, \nu_{\max})$ is the maximum temporal-spatial level in broadcast transmitter side, $\eta_{\max} \geq \eta, \nu_{\max} \geq \nu$. α_u, β_u are respective the influence parameters from temporal and spatial domains, also reflect the scalability metric in mean opinion score (MOS). Obviously, α_u and β_u respectively imply the influence of temporal scalability and spatial scalability in user's satisfaction measurement. However, in real-life videos, the value of α_u, β_u may depend on very complicated functions. To the best of our knowledge, no prior work has consider the joint impact of frame rate and image resolution on the bit rate. However, several works respectively provide the MOS results under different temporal or spatial resolution. For example, [60] gives the temporal frame rate impact on perceptual quality, and [61] presents the spatial resolution assessment on perceptual video quality. These work provide the cues for joint consideration. In this work, since we introduce hybrid temporal-spatial content metric, for the sake of fairly reflecting temporal and spatial influences of this two-dimension metric, we use $\alpha_u = \beta_u = 1$.

Fig.1 shows an example of normalized content rate-distortion, the testing video is standard sequence 'harbour'. Fig.1(b), (c) and (d) provide the results of temporal, spatial and hybrid content rate-distortion, respectively. In these three subfigures, the broken lines in black, pink, blue color, represent the results in spatial QCIF, CIF, D1 format. The five dots in each broken line orderly represent the results in temporal 1.875, 3.75, 7.5, 15, 30 fps. In Fig.1(b), the temporal distortion depends on the temporal frame rate, thus, QCIF, CIF, D1 format video sequences have same temporal distortion according to their frame rate. Fig.2(c) provide the spatial distortion. The

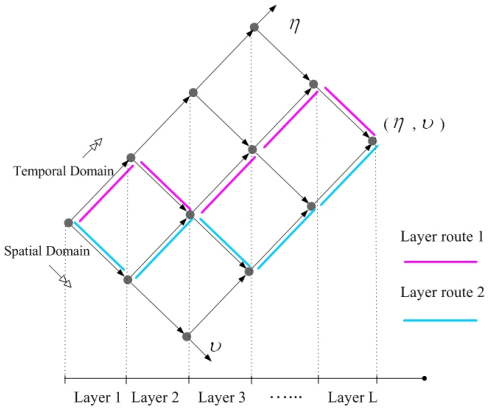


Fig. 2. Layer structure and layer route

results show that the spatial distortion depends on the spatial resolution because the same format video exhibits same spatial distortion. Although five dots have same spatial distortion due to their resolutions, the bitrates are still different because of their different frame rates. Thus, Fig.1(b) and (c) give the distortion in individual domain. Fig.2(d) provide hybrid temporal and spatial distortion, the weighted sum results. The temporal and spatial weights employ α_u and β_u in (10) so as to keep the same rule in utility measurement.

Since scalable video always support complex layer structure, to initiatively redesign the signal model for layered video coding can make the stream more compatible with the requirements from multiple heterogeneous devices. We measure the utility by a function of the received video quality, that is using the above hybrid temporal-spatial content distortion model to generate layer representation.

Definition 4: (Layer structure) A given video encoder includes interlaced $\nu_{\max}$ spatial scalable levels and $\eta_{\max}$ temporal scalable levels. (η, ν) is the l-th layer coordinate, where $\eta = 1, ..., \eta_{\max}$ and $\nu = 1, ..., \nu_{\max}$. A layer structure Ω is a contiguous set of coordinate elements that defines a layer path in temporal-spatial domain. The element in Ω is defined as $\omega_l = (\eta, \nu)_l$ so we have:

$$\Omega = \omega_1, ..., \omega_l, ...\omega_L, \qquad \max(\eta_{\max}, \nu_{\max}) \leq L \leq \eta_{\max} + \nu_{\max} - 1$$

Obviously, $\{\omega_1, ..., \omega_l\}$ reflects l layers structure.

Boundary conditions: Base layer is $\omega_1 = (1, 1)$. The highest enhancement layer is $\omega_L = (\eta_{\max}, \nu_{\max})$, in which $\eta_{\max}$ and $\nu_{\max}$ represent the largest frame rate and the highest spatial resolution, respectively.

Monotonicity: Given $\omega_l = (a, b)$ then $\omega_{l+1} = (a', b')$, where $a' - a \geq 0$, $b' - b \geq 0$ and $(a' - a) + (b' - b) = 1$. This forces the nodes in Ω to be monotonically scalable.

A layer structure example for hybrid temporal and spatial domain scalable video stream is illustrated in Fig.2. From base layer to top enhancement layer, video data is generated and will be broadcasted layered and progressively along these layers. The coordinate $(\eta, \nu)_l$ of lth layer implies different adaptive layer generation method. The reason lies in that 1) the layer route is non-unique. For example, suppose $(\eta_{\max}, \nu_{\max}) = (4, 3)$ which covers 3.75/7.5/15/30fps temporal scalability and QCIF/CIF/D1

spatial scalability. Then the red and blue lines in Fig.2 correspond to two layer route results. The red layer route means video layer will be generated or extracted along QCIF@3.75→QCIF@7.5→CIF@7.5→CIF@15→CIF@30→D1@30; while the blue layer route represents the generated path is QCIF@3.75→CIF@3.75→CIF@7.5→D1@7.5→D1@15→D1@30. 2) the resulting rate increment and content distortion of each layer is different. For example, suppose a user only receive first two layers reliably, then it means the final received video is in QCIF@7.5 under red layer route or in CIF@3.75 under blue layer route. Accordingly, the required available bandwidth and resulting video quality are different under these two adaptation operations.

Definition 5: (Layer utility) Under given layer structure Ω_L, the utility of l-th layer with coordinate $(\eta, \nu)_l$ is: $u(l) = \ddot{u}(l) - \ddot{u}(l-1)$. The corresponding bitrate increment of l-th layer is $x(l) = R(l) - R(l-1)$, where $R(\cdot)$ represents the total bitrate when video stream is extracted to corresponding ν-spatial and η-temporal levels.

With this definition, a layer structure and its resulting rate & utility relation are identified. For the previous example in Fig.2, we encode the standard sequence 'harbour' under these two layer routes, as shown in Fig.3. Fig.3(a) gives the rate & utility relation from base layer to top enhancement layer; Fig.3(b) provides the rate and utility increment for each layer. The red and blue line correspond to the layer encoding results in Fig.2. Fig.3 shows that the layer route along red line is better than that along blue line in case of low available bandwidth because its utility ascends more quickly.

D. Modeling broadcast system through hybrid utility

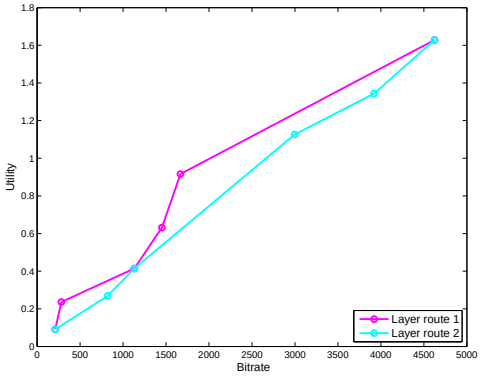
In this video broadcasting system, video content serves heterogeneous devices. The video server uses the spatial scalability to satisfy different device group users' subscription, and employs temporal scalability and corresponding joint Fountain coding to oppose variable channel condition. Herein, we use user groups to identify those users in different display size.

Assume there are $\mathcal{N} = \{n_1, ..., n_j, ..., n_{\nu_{\max}}\}$ users requiring the same content, where j is the index of heterogeneous group in term of display resolution, n_j is the number of users in group j who require same resolution video. n_1 and $n_{\nu_{\max}}$ represent the numbers of minimum and maximum resolution devices the video content supports, respectively. Obviously, there is $N = |\mathcal{N}| = \sum_{j=1}^{\nu_{\max}} n_j$. These users follow:

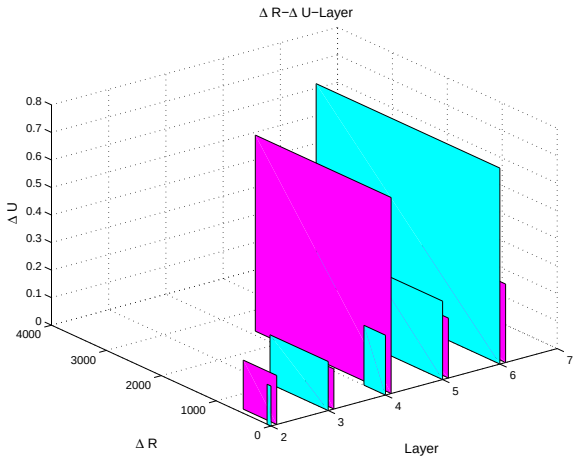
Feasibility Criterion: When current broadcasting layer is $(\eta, \nu)_l$, for the user $n \in \mathcal{N}$, who requires video in spatial level j, the spatial distortion and utility comply with definition 2 and 3 in case of $\nu \le j$; while the layer is infeasible for user n in case of $\nu > j$.

The reason lies in that high resolution device can decode and display low resolution video with declined user's satisfaction; while low resolution devices can not always decode and display high resolution video successfully.

To facilitate the discussion, we use $U(\Omega)$ to represent a content broadcasting with $\mathcal{N}$ heterogeneous users. We discuss the temporal, spatial and hybrid broadcasting utilities when system serves these $\mathcal{N}$ users, which are the components of system utility.



(a) Rate & Utility



(b) Rate-Utility-Layer

Fig. 3. Rate, utility, layer. Sequence: harbour

Theorem 1: Under layer structure Ω, for l-th layer, the total temporal utility is

$$u_{\eta l}(\Omega_l) = \frac{1}{M}\Big(\sum_{i\in\Lambda_{l-1}} d(f_i, f_i^*) - \sum_{i\in\Lambda_l} d(f_i, f_i^*)\Big) \cdot \sum_{\tau=\nu}^{\nu_{\max}} n_\tau \int_0^{\hat{\varepsilon}_l} \rho_\tau(\varepsilon)\mathrm{d}\varepsilon$$

where $\rho(\varepsilon)$ is the channel condition distribution of users, $\hat{\varepsilon}_l$ is the threshold variable.

Proof: Define Λ_{l-1} as the dropped frame set of $(l-1)$th layer. The corresponding spatial level is ν. From (8), the temporal distortion of $(l-1)$th layer is: $D_{\eta_{l-1}} = \frac{1}{M} \cdot \sum_{i\in\Lambda_{l-1}} d(f_i, f_i^*)$.

Then, for single user, according to definition 3, the utility of l-th layer due to temporal expanding is:

$$\begin{aligned} u_{\eta l}(\Omega_l) =& u_{\eta_l}(\Omega_l) - u_{\eta_{l-1}}(\Omega_{l-1}) \\ =& \Big(D^{max}_{\eta_{\max}} - D_{\eta_l,\nu}\Big) - \Big(D^{max}_{\eta_{\max}} - D_{\eta_{l-1},\nu}\Big) = D_{\eta_{l-1},\nu} - D_{\eta_l,\nu} \\ =& \frac{1}{M} \cdot \sum_{i\in\Lambda_{l-1}} d(f_i, f_i^*) - \frac{1}{M} \cdot \sum_{i\in\Lambda_l} d(f_i, f_i^*) \end{aligned}$$

Accordingly, for multiple users, the number of users in group τ who can receive the l-th layer correctly is $n_\tau \int_0^{\hat{\varepsilon}_l} \rho_\tau(\varepsilon)\mathrm{d}\varepsilon$, then, the number of users who can receive the l-th layer correctly in the whole heterogeneous devices users is $\sum_{\tau=\nu}^{\nu_{\max}} n_\tau \int_0^{\hat{\varepsilon}_l} \rho_\tau(\varepsilon)\mathrm{d}\varepsilon$.

Finally, combine these two results. Q.E.T.

Theorem 2: Under layer structure Ω, for l-th layer, the total spatial utility is

$$u_{\nu l}(\Omega_l) = \frac{1}{M} \cdot \sum_{\tau=\nu+1}^{\nu_{\max}} \theta \cdot n_\tau \int_0^{\hat{\varepsilon}_l} \rho_\tau(\varepsilon) \mathrm{d}\varepsilon$$

where $$\theta = \frac{f_w^{\nu+1}}{f_w^\tau}\frac{f_h^{\nu+1}}{f_h^\tau} \cdot \sum_{i\in\Psi_{\nu+1}} d(f_i^{\nu+1}) - \frac{f_w^\nu}{f_w^\tau}\frac{f_h^\nu}{f_h^\tau} \cdot \sum_{i\in\Psi_\nu} d(f_i^\nu)$$

The proof is similar to that of theorem 1.

Theorem 3: Under layer structure Ω, the hybrid utility of l-th layer is

$$u_{(\eta,\nu)_l}(\Omega_l) = \alpha_u u_{\eta l}(\Omega_l) + \beta_u u_{\nu l}(\Omega_l) \tag{11}$$

Proof: Add theorem 1 and 2 to the hybrid utility expression in definition 3. Q.E.T.

Remark: (Insight of system design) With theorem 1-3, each layer's utility increment is obtained when serves to heterogeneous devices users. In general, to maximize the whole broadcasting system utility depends on the adopted video adaptation scheme, which is also the solution to (1). Then the maximum utility achieving becomes a policy problem since it has close relation with layer structure Ω, which is also a tradeoff among temporal scalability, spatial scalability, and corresponding error protection.

III. Utility Maximization in Video Broadcasting

We formulate the broadcasting policy over a finite number of layer decision stages, and use dynamic programming (DP) to get the maximum overall utility of the heterogeneous devices.

A. Formulation as a DP process

We first define the sate as a combination of hybrid temporal-spatial video adaptation, then provide the control space and revenue function. Since the available bandwidth R is finite, in period τ, each decision state leads to different bandwidth consumption and the corresponding utility increment.

The state space: From (2), R consists of layered video data r and error protection part γ. In practical broadcasting system, broadcasting process begins from base layer, and video data is broadcasted progressively along video layers. Then the decision state space consists of three components, as shown in Table 1 (more details in subsection III-C):

TABLE I
State Space

State	Rate increment	Utility increment
$\gamma\uparrow$	boundary in (6)	$n_j\uparrow$ in Theorem 1,2
$\eta\uparrow$	$R_{\eta,\nu}(\eta)$ in Definition 1	$u\uparrow$ in Theorem 1
$\nu\uparrow$	$R_{\eta,\nu}(\nu)$ in Definition 2	$u\uparrow$ in Theorem 2

The revenue function: It is in fact utility function, which is additive in the sense of utility increment at period τ. Consequently, the total utility (revenue) coincides with formulation (1). We rewrite it as a DP form

$$U_{\text{system}} = u_1(r_1) + J(\tau, \eta, \nu, \varkappa_l, n_l) \tag{12}$$

where τ indexes period; η, ν represent the temporal and spatial level, respectively, and $l = \eta + \nu - 1$; $\varkappa_l$ represents the residual bandwidth for current layer l when broadcasting system servers current users, and $\varkappa_l = R - \sum_{l-1}(r(\eta,\nu)_{l-1} + \gamma_{l-1})$. n_l is the number of reliable receiving users for l layer, and for the first layer $l = 1$, initialize $n_l = N$. Obviously, the decision state for each period τ affects $(\eta,\nu)_l$, $\varkappa_l$, n_l and corresponding accumulated utility.

Therefore, the broadcasting problem can be formulated as an optimization of the accumulated utility, and the optimal hybrid temporal-spatial adaptive policy and corresponding error protection part would be equivalent to the solution of

$$\max U_{\text{system}} = u_1(r_1) + \arg\max_{(\eta,\nu)_l,\gamma_l} J^*(\tau, \eta, \nu, \varkappa_l, n_l) \tag{13}$$

B. Algorithm for utility Max

We design the DP algorithm in the following steps.

1) *Initialization*: For base layer $l = 1$, where $\eta = 1$, $\nu = 1$ and $l = 1$. Broadcasting system begins from to serve one user who has best channel condition, there is $\varkappa_1 = R - r_1$. Then, $U_{\text{system}} = u(r_1)$.

2) *Recursion*: At period τ, let $J^*(\tau, \eta, \nu, \varkappa_l, n_l)$ be the maximum expected revenue to go associated with initial state $(\tau, \eta, \nu, \varkappa_l, n_l)$. Then $J^*(\tau, \eta, \nu, \varkappa_l, n_l)$ satisfies Bellman's equation which relates the optimal revenue in the current state to the expected future revenues:

$$\begin{aligned} J^*(\tau, \eta, \nu, \varkappa_l, n_l) = \max\{&u(\tau, \eta, \nu) + J^*(\tau+1, \eta, \nu, \varkappa_l, n_l + 1), \\ &u(\tau, \eta, \nu) + J^*(\tau+1, \eta+1, \nu, \varkappa_{l+1}, n_{l+1}), \\ &u(\tau, \eta, \nu) + J^*(\tau+1, \eta, \nu+1, \varkappa_{l+1}, n_{l+1})\} \end{aligned} \tag{14}$$

3) *Termination*: At period τ, take $\varkappa_1 \le 0$ or $J^*(\tau, \eta_{\max}, \nu_{\max} + 1, \varkappa_{L+1}) = 0$ or $J^*(\tau, \eta_{\max} + 1, \nu_{\max}, \varkappa_{L+1}) = 0$.

C. Solution and analysis

For the basic problem in (1), to increase the summation of utility U mainly depends on the following aspects: 1) increase error protection part γ_l, which is equivalent to increase the number of reliable received users: from $\hat{\varepsilon}_l = 1 - \frac{r_l(1+\delta)}{r_l+\gamma_l}$, threshold $\hat{\varepsilon}_l$ is in increasing proportion to γ_l. Thus, improving γ_l can make more users receive r_l reliably, such that increase the summation of utility U; 2) increase source layer l so as to add $r_{(l+1)}$, which is equivalent to improve the utility value and affects the number of reliable received users: from section II-D, this includes increasing η or ν so as to improve the temporal utility or spatial utility, respectively. All the error protection γ_l part and original video part r_l satisfy the bandwidth constraint (2): $\sum_l (r_l + \gamma_l) \le R$.

Consequently, each state $(\tau, \eta, \nu, \varkappa_l, n_l)$ results in a basic utility increment $u(\tau, \eta, \nu)$. The next decision can be:

① Increase γ_l, such that make n_l+1 user receive current layer data reliably. Then, the expected future revenue is: $\{u(\tau, \eta, \nu) + J^*(\tau+1, \eta, \nu, \varkappa_l', n_l + 1)\}$.

② Increase layer to $l+1$ through increase temporal level to $\eta+1$, so as to add the r_{l+1}. Then, new layer broadcast begins with utility $\{u(\tau, \eta, \nu) + J^*(\tau+1, \eta+1, \nu, \varkappa_{l+1}, n_{l+1})\}$. The utility computing method complies with theorem 1.

③ Increase layer to $l+1$ through increase spatial level to $\nu+1$, so as to add the r_{l+1}. The utility computation begins from a new layer, that is $\{u(\tau, \eta, \nu) + J^*(\tau+1, \eta, \nu+1, \varkappa_{l+1}, n_{l+1})\}$. The computing method complies with theorem 2.

The decisions follow two principles. First, the number of served users in layer $l+1$ is no more than that in layer l.

The reason lies in that the decoding and reconstruction of each layer depend on its previous layers. For the users who received posterior enhancement layer but lost base layer or anterior enhancement layers, the utilities in this layer become useless. Second, both r_l and γ_l share available bandwidth R, thus $\varkappa_l = R - \sum_{k=1}^{l-1}(r_k + \gamma_k)$.

Theorem 4: (Optimal solution) For initial state $\tau = 1$, the optimal revenue $J^*(\tau_1, \eta_1, \nu_1, \varkappa_1, n_1)$ of the basic problem is equal to $J_1(\tau_1)$, following the recursion algorithm (14), if $g_\tau^* = g^*(\tau, \eta, \nu)$ maximizes the right side of equation (14) for each τ and state (τ, η, ν), the policy $\pi^* = \{g_1^*, ..., g_N^*\}$ is optimal.

Proof: For any admissible policy $\pi^* = \{g_2, ..., g_N\}$ and each $\tau = 2, ..., N$, denote $\pi^\tau = \{g_2, ..., g_\tau\}$. For $\tau = 2, ..., N$, let $J_\tau^*(\tau, \eta, \nu)$ be the optimal revenue for the $1 \sim \tau$ stage problem that starts at state $(1, \eta_1, \nu_1)$ and period 1, and ends at period τ, there is

$$J_\tau^*(\tau, \eta, \nu) = \max_{\pi^\tau}\left\{u_1(1, \eta_1, \nu_1) + \sum_{k=2}^{\tau} u(k, \eta_k, \nu_k)\right\} \tag{15}$$

For $\tau = 1$, define $J_1^* = u_1(1, \eta_1, \nu_1, \varkappa_1, n_1)$. We have $J_1^* = J_1 = u_1$. Assume that for τ and all $u_{\tau-1}$ $(\tau = 2, ..., N)$, we have $J_{\tau-1}^* = J_{\tau-1}(\tau - 1, \eta_{\tau-1}, \nu_{\tau-1}, \varkappa_{\tau-1}, n_{\tau-1})$. Then, since $\pi^\tau = (u_\tau, \pi^{\tau-1})$, we have for all $(\tau, \eta_\tau, \nu_\tau)$

$$J_\tau^*(\tau, \eta, \nu) = \max_{(u_\tau, \pi^{\tau-1})}\left\{u_\tau(\tau, \eta_\tau, \nu_\tau) + u_1(1, \eta_1, \nu_1) + \sum_{k=2}^{\tau-1} u(k, \eta_k, \nu_k)\right\}$$

$$= \max_{u_\tau}\left\{u_\tau(\tau, \eta_\tau, \nu_\tau) + \max_{\pi^{\tau-1}}\left[u_1(1, \eta_1, \nu_1) + \sum_{k=2}^{\tau-1} u(k, \eta_k, \nu_k)\right]\right\} \tag{16}$$

$$= \max_{u_\tau}\left\{u_\tau(\tau, \eta_\tau, \nu_\tau) + J_{\tau-1}^*(\tau - 1, \eta_{\tau-1}, \nu_{\tau-1}, \varkappa_{\tau-1}, n_{\tau-1})\right\} \tag{17}$$

$$= J_\tau(\tau, \eta, \nu) \tag{18}$$

In equation (16), we move the maximum over $\pi^{\tau-1}$ inside the braced expression, using an optimality principle [62]: the tail portion of an optimal policy is also the optimal for the subproblem. In equation (17), we use the definition of $J_{\tau-1}^*$. In equation (18), we use the induction hypothesis $J_{\tau-1}^* = J_{\tau-1}(\cdot)$ because of the iterative operation. Therefore, an optimal policy exists, which is obtainable via value iteration.

IV. Experimental Results

A. Simulation setup

In the simulation, we investigate a wireless video broadcasting system in [59]. We use the H.264 extended SVC (JSVM [63]) video encoder to generate layered video stream with both temporal and spatial scalability support. In broadcasting scenario, as shown in Fig.2, scalable video covers three levels of spatial resolution: ranging among QCIF, CIF and D1 formats, which serve three groups users in corresponding display size; and covers five temporal levels: 1.875, 3.75, 7.5, 15, 30fps, which serve users with variable channel conditions. In order to evaluate the performance of proposed solution, we investigate a simple video broadcast network with thirty six users belong to three groups, as follow,

Group 1: ten users, requiring QCIF format video

Group 2: eighteen users, requiring CIF format video

Group 3: eight user, requiring D1 format video

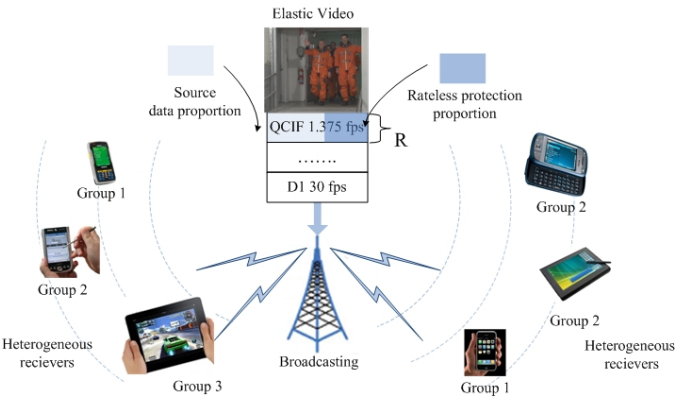
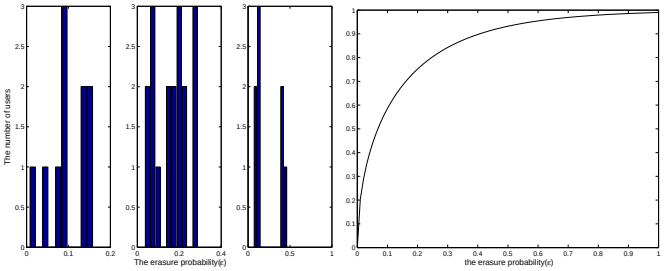


Fig. 4. Elastic rate video broadcasting to heterogeneous devices

These users are in wireless mobile error-prone circumstance and their channel states comply with the erasure rates distribution in Fig.5(a). The empirical CDF of users' channel states is shown in Fig.5(b). The following results are only meant to demonstrate the effectiveness of the proposed methods, not a full scale deployment in real network.



(a) Erasure rates distribution (b) Empirical CDF of users' state

Fig. 5. Mobile users erasure rate state example

B. Broadcasting utility and adaptive layer structure

In this subsection, we present the numerical results in modeling broadcasting utility through temporal-spatial content metric, we also show the role of layer structure.

1) Temporal-spatial metric and layer structure: Fig.6 illustrates the flexibility in layer structure. Every coordinate (η, ν) represents the temporal and spatial levels. The hollow dots represent the candidated operation points for adaptation layer route. The upper and lower remarks of the hollow dots includes rate increment & utility increment in spatial and temporal domains. For example, when 'harbour' sequence is used, for the point $(\eta, \nu) = (2, 1)$, the remarks $\frac{(56,0.09)}{(0,0.00)}$ mean if current position plays as the next enhancement layer, it relies on the temporal increment because spatial rate and utility are equal to zeros. And (56,0.09) shows that additional 56kbps rate can result in 0.09 utility increment.

Table II demonstrates the detailed coding results of the whole layer structure. The values in column 'Spatial-D', 'Temporal-D' and 'Utility' are spatial, temporal distortion and utility, respectively, and are computed as subsection II-C.

2) Utility and adaptive layer structure: Fig.6 provides two layer routes, distinguished by red and green lines, which represent two derived the optimal layer decisions through this work

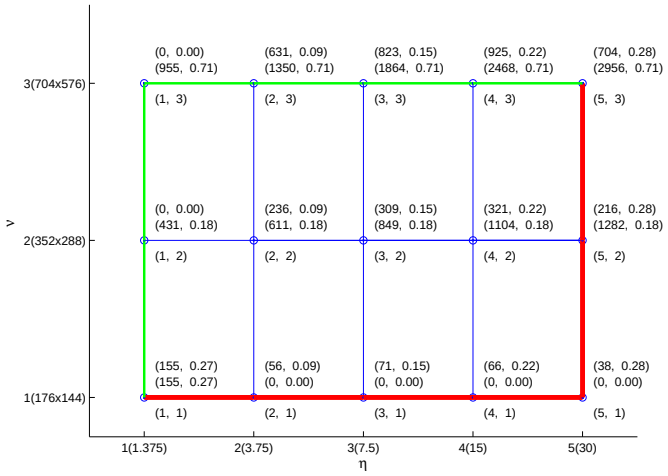
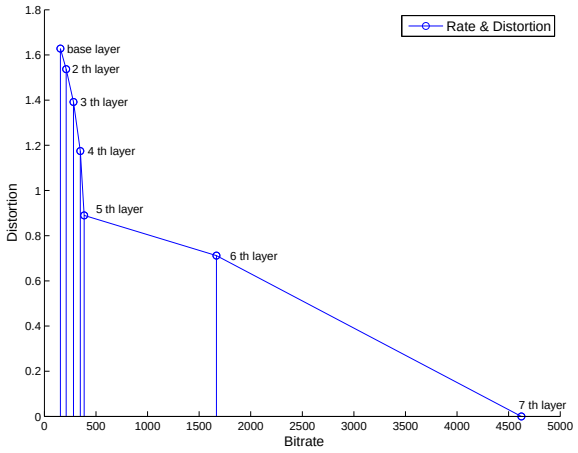
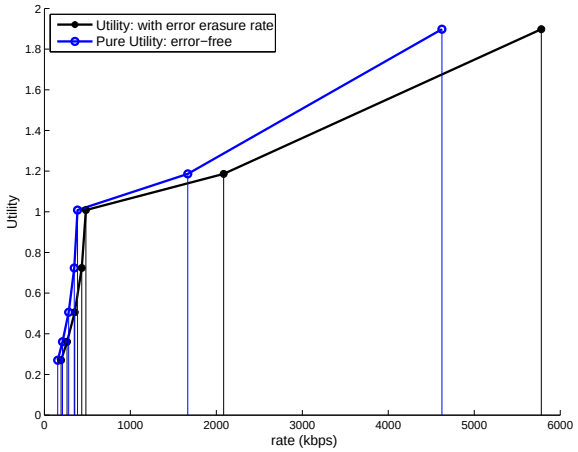


Fig. 6. Layer structure for sequence 'harbour'



(a) Rate & distortion



(b) Rate-Uti-Layer

Fig. 7. Rate, distortion, pure utility and utility. Sequence: harbour

and greedy generation algorithm (GGA) (more details in next subsection). For the route of red line, Fig.7(a) gives the rate & utility relation from base layer to top enhancement layer, video data is broadcasted progressively along these layers. Fig.7(b) gives the corresponding utilities. Fig.7(b) includes two types of utilities, one is the pure utility painted in blue line, which is computed according to subsection II-C. Pure utility means a user receive all the layered video data in error-free circumstance. Though this case is an ideal condition, it provides a rate comparison for error protection designs. In practical system, compressed layer data are encoded by channel coding so as to ensure reliable recovery by the receivers. Under Fountain code protection design, the practical utility painted in black line is also provided in Fig.7(b). The absolute values of both pure utility and utility are always in same, while the required bitrates to obtain the value are different. The reason lies in the channel protection part, that is γ in (2).

C. Heterogeneous QoS video broadcasting scenario

In this subsection, we present how the proposed scheme work and why it can maximize the whole broadcasting utility. Heterogeneous QoS video is generated (section II) so as to serve heterogeneous devices users. The layer structure is built from hybrid temporal summarization and spatial preference metric (subsection II-C). Through compute the broadcast utility (subsection II-D), derive the optimal layer decision through DP solution (section III).

1) Comparing algorithms: We further compare the proposed algorithm, noted as DP, with three typical algorithms, as follows,

(i) Greedy generation algorithm (GGA): under this algorithm, elastic rate video is generated along the largest utility route. This means, the enhancement layer decision relies on larger utility direction, as shown in the green line in Fig.6. The corresponding broadcasting situation is shown in the blue line in Fig.8.

(ii) Worst-user (WU) algorithm: it is the widely accepted approach in multicast/broadcast scenario and has been proposed for IEEE, (e.g., [64]). Under this design, video is generated and broadcasted based on the reception of worst user at each user group.

(iii) Opportunistic broadcasting (OPP) algorithm: since worst-user case may lead to available bandwidth waste when most users are in good channel conditions, this strategy can opportunistic video broadcasting over wireless networks that take into account users' variable reception, (e.g., [65]).

Fig.8 gives the aggregate utility results: the red line is the results of proposed DP scheme in which the layer route is also the red line in Fig.6; while the green line represents the results of GGA algorithm in which the corresponding layer route is the green line in Fig.6. Fig.8 also shows the aggregate utility results under the other two schemes: the blue line corresponds to the results under WU and the black line corresponds to the results under OPP. All these schemes are with consideration of users' heterogeneous reception. As shown in Fig.8, the results show three conclusions: (1) user performance consideration in unit of heterogenous group affects the rate allocation in joint layered video and error protection coding, because OPP shows better aggregate utility performance than WU. However, since OPP relies on the optimum selection of performance threshold, when the users in a user group are all in good reception, the required error protection rate decreases, WU shows better than

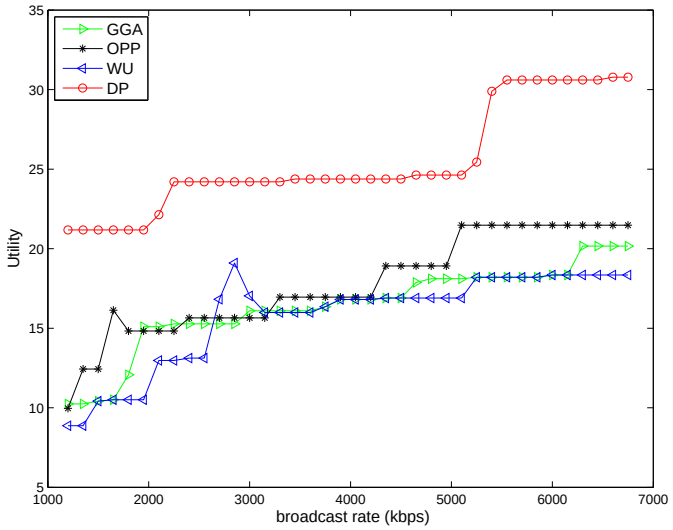


Fig. 8. Aggregate utility comparision: DP and GGA

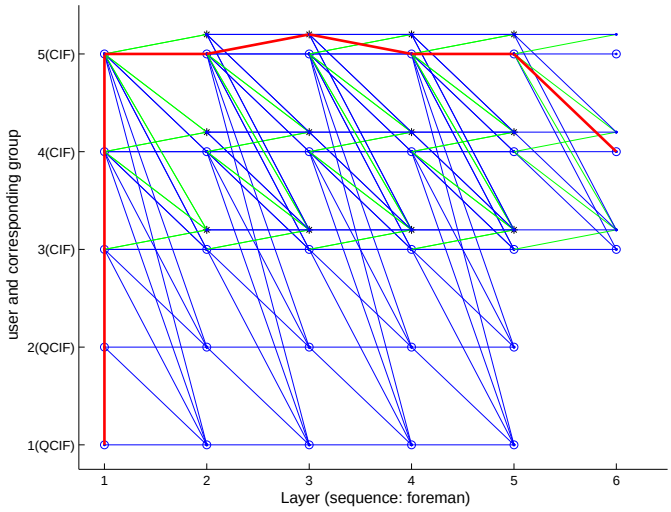


Fig. 9. Broadcast trellis based on DP solution.

OPP and GGA, which is reflected in the saltation in 2900kbps; (2) layer route directly affects the broadcasting utility, because GGA, WU and OPP show some fluctuations under different available bandwidth. The inflexions in 1600~2000kbps, 4200~5100kbps reflect the different layer routes results under these three schemes; (3) suitable optimal strategy benefits the whole broadcasting utility. The reason lies that the results show that the proposed DP scheme out-performs the others schemes in utility for all broadcasting rate range.

2) DP solution analysis: We use a very simple five users broadcasting system to vividly analysis why DP solution is optimal. 'Foreman' sequence is encoded in 6 layers. The bandwidth budget R = 450kbps, which is not enough for all the five users' correctly receiving:

Group 1: user 1 and user 2, requiring QCIF format video

Group 2: user 3, 4, 5, requiring CIF format video

The broadcasting process is illustrated as a trellis, as shown in Fig.9. The blue lines represent all the possible decision paths; green lines represent decision paths along spatial direction; red lines represent final optimal decision path. The trellis analysis shows that the proposed optimal broadcasting policy with DP solution can obtain maximum utility especially under bandwidth limited, the reasons lie in: 1) the max-utility achieving problem is constructed in piecemeal fashion [62], first transferring a broadcasting policy to sequential tail subproblem, then solving the tail subproblem through every stage (hollow dot in Fig.9) decision, and continuing in this manner until an optimal layer route policy for the entire problem is constructed. 2) during every stage, decision is made through computing broadcasting utility from each state and corresponding utility revenue, which can take advantage of the extra information (the value of the current state). This starts at the initial stage and ends at the stage within the terminal bandwidth budget, and has maximum sum of utility revenue. Therefore, through DP solution, elastic rate video broadcasting can achieve maximum utility.

3) Computational complexity analysis: The complexity of above solution depends on the sum of the number of states in state space, and the number of arithmetic operations performed by the algorithm. For the above problem:

The complexity of DP solution: At each node, there are three candidated decision directions, as shown in Fig.9. Since progressive layer broadcasting is employed, for each epoch l, there can be at most incoming L arcs. From definition 4, it is $\eta_{\max} + \nu_{\max} - 1$. Since heterogeneous device characteristic is introduced, spatial layer increment will improve a user group's satisfaction, which accordingly leads to additional spatial dimension computation, as shown in the green line in Fig.9. Besides, in order to provide N heterogeneous users reliable transmission, and the users may belong to different user groups, the computational complexity can be upper-bounded by

$$C = (\eta_{\max} + \nu_{\max} - 1) \cdot \nu_{\max} \cdot \sum_{i=1}^{N} i$$

The complexity of GGA solution: Similar to the analysis in DP, for the GGA solution, since for each epoch l, the utility max rule determines the layer progressive direction, the number of arcs in the DP trellis can be reduced. Accordingly, the computational complexity is upper-bounded by

$$C = (\eta_{\max} + \nu_{\max} - 1) \cdot \sum_{i=1}^{N} i$$

Thus, the complexity difference between DP and GGA is $\nu_{\max}$ times. In practical system, user groups in term of display size as well as corresponding maximum spatial scalable level is a limited integer (e.g. no more than 30). However, GGA may make some spatial rate not feasible because some potentially useful arcs are removed. Consequently, if we have a larger L, more arcs need to be evaluated in the DP trellis, which could potentially improve the available bandwidth utilization.

V. Conclusion and Future Work

In this paper, we propose a framework of broadcasting flexible rate and reliable video stream to heterogeneous devices. Our objective is to maximize the total reception quality of

TABLE II
CODING RESULTS IN LAYER STRUCTURE (*harbour*)

v,η	Coding info.			PSNR metric	Distortion & utility in proposed metric		
	Res.	fps	kbps	Y-PSNR	Spatial-D	temporal-D	Utility
(1,1)	176×144	1.875	153.3134	43.3410	0.8897	0.7384	0.2699
(1,2)	176×144	3.75	208.5600	41.0628	0.8897	0.6475	0.3608
(1,3)	176×144	7.5	281.6472	39.4910	0.8897	0.5023	0.5061
(1,4)	176×144	15	348.4808	38.2045	0.8897	0.2848	0.7235
(1,5)	176×144	30	385.9312	37.3971	0.8897	0.0000	1.0083
(2,1)	352×288	1.875	577.8995	41.3016	0.7118	0.7384	0.4478
(2,2)	352×288	3.75	810.9797	39.2928	0.7118	0.6475	0.5387
(2,3)	352×288	7.5	1130.7328	37.8869	0.7118	0.5023	0.6840
(2,4)	352×288	15	1452.2896	36.6672	0.7118	0.2848	0.9015
(2,5)	352×288	30	1667.9576	35.7260	0.7118	0.0000	1.1863
(3,1)	704×576	1.875	1521.0505	40.3062	0.0000	0.7384	1.1596
(3,2)	704×576	3.75	2143.1361	38.4815	0.0000	0.6475	1.2505
(3,3)	704×576	7.5	2994.6176	37.2002	0.0000	0.5023	1.3958
(3,4)	704×576	15	3919.5728	36.1254	0.0000	0.2848	1.6132
(3,5)	704×576	30	4624.0016	35.2166	0.0000	0.0000	1.8980

heterogeneous QoS users, and the solution is based on joint temporal-spatial scalable video and Fountain coding optimization. We aim at heterogeneous devices characteristics including diverse display resolution and variable channel conditions. We introduce a hybrid temporal and spatial rate-distortion metric based on video summarization and user preference. Based on this hybrid metric, we model the total reception quality provision problem as a broadcasting utility achieving problem. We use adaptive layer structure embedded with hybrid temporal and spatial scalability to generate flexible rate video, while use rateless erasure protection to provide elastic and reliable transmission. Joint coding between layered video and rateless codes is embedded into the broadcasting utility solving process. Simulation results show that the proposed framework can maximize the total users' utility. In the future, we will expand and practice this method into more complex broadcasting system with consideration of more users heterogeneous devices characteristics.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (61001194), and Hong Kong Research Grant Council (RGC) and Hong Kong Polytechnic University new faculty grant.

The authors would like to thank the reviewers and editors for their insightful comments and suggestions.

REFERENCES

[1] H. Y. Shutoy, D. Gündüz, E. Erkip, and Y. Wang, "Cooperative Source and Channel Coding for Wireless Multimedia Communications", *IEEE J. Select. Topics Signal Process.*, vol. 1, no. 2, pp. 295-307, Aug. 2007.

[2] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-Distortion Optimized Hybrid Error Control for Real-Time Packetized Video Transmission", *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 40-53, Jan. 2006.

[3] Z. He, J. Cai, and C. W. Chen, "Joint Source Channel Rate-Distortion Analysis for Adaptive Mode Selection and Rate Control in Wireless Video Coding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511-523, Jun. 2002.

[4] N. Raja, Z. Xiong, and M. Fossorier, "Combined Source-Channel Coding of Images under Power and Bandwidth Constraints", *EURASIP J. Advances Signal Process.*, vol. 2007.

[5] X. Jaspar, C. Guillemot, and L. Vandendorpe, "Joint Source-Channel Turbo Techniques for Discrete-Valued Sources: From Theory to Practice", *Proc. IEEE*, vol. 95, no. 6, pp. 1345-1361, Jun. 2007.

[6] L. Qian, D.L. Jones, K. Ramchandran, and S. Appadwedula, "A General Joint Source-Channel Matching Method for Wireless Video Transmission", *in IEEE Proc. DCC*, pp.414-424, 1999.

[7] Q. Xu, V. Stankovic, and Z. Xiong, "Distributed Joint Source-Channel Coding of Video Using Raptor Codes", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 4, pp. 851-861, May. 2007.

[8] M. Bystrom and J. W. Modestino, "Combined Source-Channel Coding Schemes for Video Transmission over an Additive White Gaussian Noise Channel", *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 880-890, Jun. 2000.

[9] D. C. Cernea, A. Munteanu, A. Alecu, J. Cornelis, and P. Schelkens, "Scalable Joint Source and Channel Coding of Meshes", *IEEE Trans. Multimedia*, vol. 10, no. 3, pp. 503-513, Apr. 2008.

[10] K. Ramchandran, A. Ortega, K. M. Uz, and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source/channel coding", *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 6-23, Jan. 1993.

[11] F. Zhai, Y. Eisenberg, and A. K. Katsaggelos, "Joint Source-Channel Video Transmission", *Synthesis Lectures on Image, Video, and Multimedia Process.*, doi:10.2200/S00061ED1V01Y200707IVM010, 2007.

[12] M. Wien, H. Schwarz, and T. Oelbaum, "Performance Analysis of SVC",*IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1194-1203, Sept. 2007.

[13] Y. Wang, L. P. Chau, and K. H. Yap, "Spatial resolution decision in scalable bitstream extraction for network and receiver aware adaptation", *in Proc. ICME*, pp. 577-580, Jun. 2008.

[14] M. Stoufs, A. Munteanu, J. Cornelis, and P. Schelkens, "Scalable Joint Source-Channel Coding for the Scalable Extension of H.264/AVC",*IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 12, pp. 1657-1670, Dec. 2008.

[15] T. B. Abanoz and A. M. Tekalp, "Optimization of encoding configuration in scalable multiple description coding for rate-adaptive P2P video multicasting", *in Proc. ICIP*, pp. 3741-3744, Nov. 2009.

[16] E. Maani and A. K. Katsaggelos, "Optimized Bit Extraction Using Distortion Modeling in the Scalable Extension of H.264/AVC",*IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2022-2029, Sept. 2009.

[17] Q. Zhang, Q. Guo, Q. Ni, W. Zhu, and Y. Q. Zhang, "Sender-Adaptive and Receiver-Driven Layered Multicast for Scalable Video Over the Internet",*IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 4, pp. 482-495, Apr. 2005.

[18] D. Taubman and J. Thie, "Optimal Erasure Protection for Scalably Compressed Video Streams With Limited Retransmission", *IEEE Trans. Image Process.*, vol . 14, no. 8, pp. 1006-1019, Aug. 2005.

[19] J. Liu, B. Li, Y. T. Hou, and I. Chlamtac, "On optimal layering and bandwidth allocation for multisession video broadcasting",*IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 656-667, Mar. 2004.

[20] W. Ji and Z. Li, "Heterogeneous QoS Video Broadcasting with Optimal Joint Layered Video and Digital Fountain Coding", *in Proc. ICC*, Jun. 2011.

[21] Z. Li, G. M. Schuster, A. K. Katsaggelos, and B. Gandhi, "Rate-distortion Optimal Video Summary Generation", *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1550-1560, Oct. 2005.

[22] J. W. Byers, M. Luby, and M. Mitzenmacher, "A Digital Fountain

Approach to Asynchronous Reliable Multicast", *IEEE J. Select. Areas Commun.*, vol. 20, no. 8, pp. 1528-1540, Oct. 2002.
[23] J. Crowcroft and K.Paliwoda, "A Multicast Transport Protocol", *in Proc. SIGCOMM*, vol. 18, no. 4, pp. 247-256, Aug. 1998.
[24] D. Jurca, P. Frossard, and A. Jovanovic, "Forward Error Correction for Multipath Media Streaming", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1315-1326, Sept. 2009.
[25] W. Xiang, C. Zhu, C. K. Siew, Y. Xu, and M. Liu, "Forward Error Correction-Based 2-D Layered Multiple Description Coding for Error-Resilient H.264 SVC Video Transmission",*IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 12, pp. 1730-1738, Dec. 2009.
[26] C. A. Segall and G. J. Sullivan, "Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, No. 9, pp. 1121-1135, Sept. 2007.
[27] H. Schwarz, D. Marpel, 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, Sept. 2007.
[28] T. Schierl, K. Grüneberg, and T. Wiegand, "Scalable video coding over RTP and MPEG-2 transport stream in broadcast and IPTV channels", *IEEE Wireless Commun.*, vol . 16, no. 5, pp. 64-71, Oct. 2009.
[29] A. Majumda, D. G. Sachs, I. V. Kozintsev, K. Ramchandran, and M. M. Yeung, "Multicast and unicast real-time video streaming over wireless LANs", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 524-534, Jun. 2002.
[30] Z. Liu, Z. Wu, P. Liu, H. Liu, and Y. Wang, "Layer bargaining: multicast layered video over wireless networks", *IEEE J. Select. Areas Commun.*, vol . 28, no. 3, pp. 445-455, Apr. 2010.
[31] M. van der Schaar, Y. Andreopoulos, and Z. Hu, "Optimized Scalable Video Streaming over IEEE 802.11a/e HCCA Wireless Networks under Delay Constraints", *IEEE Trans. Mobile Computing*, vol. 5, no. 6, pp. 755-768, Jun. 2006.
[32] 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.
[33] M. Luby, "LT-codes," in *Proc. IEEE Symp. FOCS*, pp. 271-280, 2002.
[34] D. J. C. MacKay, "Fountain codes", *IEE Proc. Commum.*, vol. 152, no. 6, pp. 1062-1068, Dec. 2005.
[35] N. Rahnavard, B. N. Vellambi, and F. Fekri, "Rateless Codes With Unequal Error Protection Property", *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1521-1532, Apr. 2007.
[36] A. Shokrollahi, "Raptor codes",*IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551-2567, Jun. 2006.
[37] T. Schierl, S. Johansen, C. Hellge, T. Stockhammer, and T. Wiegand, "Distributed Rate-Distortion Optimization for Rateless Coded Scalable Video in Mobile Ad Hoc Networks", *in Proc. ICIP*, pp. VI-497-VI-500, Sept. 2007.
[38] M. Luby, T. Gasiba, T. Stockhammer, and M. Stockhammer, "Reliable Multimedia Download Delivery in Cellular Broadcast Networks", *IEEE Trans. Broadcast.*, vol . 53, no. 1, pp. 235-246, Mar. 2007.
[39] D. Vukobratović, V. Stanković, D. Sejdinović, L. Stanković, and Z. Xiong, "Scalable Video Multicast Using Expanding Window Fountain Codes", *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1094-1104, Oct 2009.
[40] "Multimedia Broadcast/Multicast Service (MBMS): Protocols and codecs", 3GPP TS 26.346 V6.1.0.
[41] S. Ahmad, R. Hamzaoui, and M. Al-Akaidi, "Adaptive Unicast Video Streaming With Rateless Codes and Feedback",*IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 2, pp. 275-285, Feb. 2010.
[42] J.-P. Wagner, J. Chakareski, and P. Frossard, "Streaming of Scalable Video from Multiple Servers using Rateless Codes", *in Proc. ICME*, pp. 1501-1504, Jul. 2006.
[43] P. Cataldi, M. Grangetto, T. Tillo, E. Magli, and G. Olmo, "Sliding-Window Raptor Codes for Efficient Scalable Wireless Video Broadcasting With Unequal Loss Protection", *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1491-1503, Jun. 2010.
[44] S. Ahmad, R. Hamzaoui, and M. M. Al-Akaidi, "Unequal Error Protection Using Fountain Codes With Applications to Video Communication", *IEEE Trans. Multimedia*, vol. 13, no. 1, pp. 92-101, Feb. 2011.
[45] M. Chiang, S. H. Low, and A. R. Calderbank, "Layering as optimization decomposition: A mathematical theory of network architectures", *Proc. IEEE*, vol. 95, no. 1, pp. 255-312, Jan. 2007.
[46] C. An and T. Q. Nguyen, "Analysis of Utility Functions for Video", *in Proc. ICIP*, pp. V-89-V-92, Jul. 2007.
[47] N. Shutto, Y. Handa, T. Higashino, K. Tsukamoto, and S. Komaki, "Measurements of a Utility Function for Video Download Service and its Application to Service Management", *in Proc. WCNC*, pp. 2894-2898, Mar. 2007.
[48] W.-H. Kuo, W. Liao, and T. Liu, "Adaptive Resource Allocation for Layer-Encoded IPTV Multicasting in IEEE 802.16 WiMAX Wireless Networks", *IEEE Trans. Multimedia*, vol. 13, no. 1, pp. 116-124, Feb. 2011.
[49] Y.-J. Yu, A.-C. Pang, Y.-C. Fang, and P.-F. Liu, "Utility-Based Resource Allocation for Layer-Encoded Multimedia Multicasting over Wireless Relay Networks", *in Proc. GLOBECOM*, pp. 1-6, Dec. 2009.
[50] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile Video Transmission Using Scalable Video Coding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, No. 9, pp. 1204-1217, Sept. 2007.
[51] Y. Li, Z. Li, M. Chiang, and A. R. Calderbank, "Content-Aware Distortion-Fair Video Streaming in Congested Networks", *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1182-1193, Oct. 2009.
[52] V. G. Subramanian, R. A. Berry, and R. Agrawal, "Joint Scheduling and Resource Allocation in CDMA Systems",*IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2416-2432, May. 2010.
[53] J. Huang, Z. Li, M. Chiang, and A. K. Katsaggelos, "Joint source adaptation and resource allocation for multi-user wireless video streaming", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 582-595, May 2008.
[54] E. Maani, P. V. Pahalawatta, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Resource Allocation for Downlink Multiuser Video Transmission Over Wireless Lossy Networks", *IEEE Trans. Image Process.*, vol. 17, No. 9, pp. 1663-1671, Sept. 2008.
[55] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio", *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359-378, May. 1994.
[56] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551-2567, June 2006.
[57] 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. PV*, pp. 1-9, May. 2009.
[58] W. Ji and Z. Li, "Joint Layered Video and Digital Fountain Coding for Multi-Channel Video Broadcasting", *in Conf. ACM Multimedia*, pp. 1223-1226, Oct. 2010.
[59] Z. Li, Y. Li, M. Chiang, R. Calderbank, and Y. C. Chen, "Optimal Transmission Scheduling For Scalable Wireless Video Broadcast with Rateless Erasure Correction Code", *in IEEE Conf. CCNC*, vol. 10, no. 13, pp. 1-5, Jan. 2009.
[60] Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual Quality Assessment of Video Considering Both Frame Rate and Quantization Artifacts", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 286-298, Mar. 2011.
[61] Y. Xue, Y.-F. Ou, Z. Ma, and Y. Wang, "Perceptual video quality assessment on a mobile platform considering both spatial resolution and quantization artifacts", *in Proc. PV*, pp. 201-208, Dec. 2010.
[62] D. P. Bertsekas, "Dynamic Programming and Optimal Control, Volume I", Press: Athena Scientific, 3rd, May. 2005.
[63] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model JSVM-6", Joint Video Team, Doc. JVT-S202, Apr. 2006.
[64] C. Eklund, R. B. Marks, S. Ponnuswamy, K. L. Stanwood, and N.J.M.V. Waes, "WirelessMAN: Inside the IEEE802.16 Standard for Wireless Metropolitan Networks", Press: IEEE Standards Information Network, May. 2006.
[65] Q. Le-Dang, T. Le-Ngoc, and Q.-D. Ho, "Opportunistic Multicast Scheduling with Erasure-Correction Coding over Wireless Channels", *in Proc. ICC*, May. 2010.

Wen Ji received the M.S. and Ph.D. degrees in communication and information systems from Northwestern Polytechnical University, China, in 2003 and 2006, respectively.

She is an Associate Professor in Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS). She was a Post Doctoral Fellow from 2007 to 2009, and was an Assistant Professor from 2009 to 2010 at ICT, CAS. Her research areas include video communication & networking, video coding, channel coding, information theory, and optimization.

Zhu Li (SM'07) received his Ph.D. degree in electrical and computer engineering from Northwestern University, Evanston, IL, in 2004.

He was an Assistant Professor with the Department of Computing, The Hong Kong Polytechnic University, from 2008 to 2010, and a Principal Staff Research Engineer with the Multimedia Research Lab (MRL), Motorola Labs, from 2000 to 2008. He is currently a Senior Staff Researcher with the Core Networks Research, Huawei Technology USA, Bridgewater, NJ. His research interests include audio-visual analytics and machine learning with its application in large scale video repositories annotation, mining, and recommendation, as well as video adaptation, source-channel coding and distributed optimization issues of the wireless video networks. He has 12 issued or pending patents and 60+ publications in book chapters, journals, and conference proceedings in these areas.

Dr. Li was elected Vice Chair of the IEEE Multimedia Communication Technical Committee (MMTC) 2008-2010. He received the Best Poster Paper Award at IEEE International Conference on Multimedia & Expo (ICME), Toronto, ON, Canada, 2006, and the Best Paper (DoCoMo Labs Innovative Paper) Award at the IEEE International Conference on Image Processing (ICIP), San Antonio, TX, 2007.

Yiqiang Chen received the B.A.Sc. and M.A. degrees from the University of Xiangtan, Xiangtan City, China, in 1996 and 1999, respectively, and the Ph.D. degree from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2002.

In 2004, he was a Visiting Scholar Researcher at the Department of Computer Science, Hong Kong University of Science and Technology (HKUST), Hong Kong. Currently, he is an Associate Professor and Vice Director of the pervasive computing research center at ICT, CAS. His research interests include artificial intelligence, pervasive computing, and human computer interface.