JOINT SOURCE / CHANNEL CODING FOR WWAN MULTIVIEW VIDEO MULTICAST WITH COOPERATIVE PEER-TO-PEER REPAIR

Zhi Liu [+], Gene Cheung [#], Vladan Velisavljević [o], Erhan Ekmekcioglu [*], Yusheng Ji [#]

[+] The Graduate University for Advanced Studies, [#] National Institute of Informatics,
[o] Deutsche Telekom Laboratories, [*] University of Surrey

ABSTRACT

WWAN video multicast is challenging because of unavoidable packet losses, and inability to perform retransmission per packet for every user due to the NAK implosion problem. Previously proposed cooperative peer-to-peer repair (CPR) strategy, leveraging on the broadcast nature of wireless transmission and "uncorrelatedness" of receivers' channels, calls for peers with good WWAN channels to streaming server (rich peers) to locally relay packets lost to peers with bad WWAN channels (poor peers) over a secondary ad-hoc WLAN network. In the interactive multiview video streaming (IMVS) scenario, however, where users can each periodically select one out of many available views for decoding and display, a poor peer may not have a neighboring rich peer watching the same view for packet recovery via CPR.

In this paper, we propose a new CPR strategy for peers to repair lost packets to neighbors watching different views. The key idea is for server to transmit depth maps in addition to texture maps, so that lost frames in a different view can be reconstructed using depth-image-based rendering (DIBR). Like Forward Error Correction (FEC) packets, encoded depth maps incur an overhead in redundant information transmission to counter network losses, and we allocate optimal amount of bits for FEC packets (for same-view direct path protection via WWAN source) and depth map encoding (for different-view indirect path via CPR peers) to minimize expected distortion. Experimental results show our proposed CPR scheme offers a 3.4dB PSNR improvement over a non-CPR scheme that relies on FEC only.

***Index Terms*—** Multiview video, wireless multicast, cooperative networks, joint source/channel coding

1. INTRODUCTION

Video streaming over Wireless Wide Area Networks (WWAN) is challenging, given stringent playback deadlines of real-time video and unavoidable packet losses due to shadowing, channel fading and inter-symbol interference. Wireless video multicast is even more difficult, since the streaming server cannot tailor retransmission for every lost packet experienced by each client due to the well-known NAK implosion problem [1]. To overcome channel losses, typical WWAN video multicast schemes [2] inject a sufficiently large amount of Forward Error Correction (FEC) packets for channel coding, expending precious WWAN resource.

One recent approach [3, 4] to alleviate the wireless packet loss problem is to apply the *cooperative communication* principle for clients (peers) who are willing to help each other. In a nutshell, given the broadcast nature of wireless transmission (each transmission is heard by multiple receivers) and the "uncorrelatedness" of receivers' channels (hence unlikely for all receivers to undergo bad channel fades at the same time), peers that currently experience good channels (rich peers) can locally assist clients who currently experience losses (poor peers), *without* relying on the remote server for retransmission. In the case of WWAN multicast, if a secondary network such as an ad-hoc Wireless Local Area Network (WLAN) connects the peers locally [3], a WWAN multicast packet received by rich peers can be relayed to poor peers via WLAN without interrupting primary WWAN transmission.

The important observation in such *cooperative peer-to-peer repair* (CPR) scenario is that a packet can be delivered to a peer directly from server over WWAN (direct path), or indirectly via a CPR relay (indirect path). The net effect of such multi-path transmission is that a packet is lost to a peer *only if* both transmission paths fail. Given the more general packet loss condition (as opposed to non-CPR scenario), a smart joint source / channel coding (JSCC) scheme [5] can allocate more bits to source coding out of a fixed WWAN transmission bit budget to combat quantization errors without incurring more packet losses. We have previously shown [5] that such smart bit allocation scheme exploiting this *multi-path JSCC gain* can reap additional video quality (up to 6dB) over a similar CPR scheme but without smart bit allocation.

An orthogonal development over the past few years is multiview video technologies. Because of the continuing cost decrease of consumer-level cameras, a video sequence can now be recorded by a large array of cameras [6]; i.e., at each time instant, images of the same scene are simultaneously captured by multiple cameras from different viewpoints, creating a richer and higher-dimensional representation of the media content. Given encoded multiview content at the server, in an *interactive multiview video streaming* (IMVS) scenario [7], a viewer can freely select one out of many available views at the server periodically for real-time streaming, so that only frames of interested viewpoint are received. While IMVS offers viewers a new level of interaction via view switching, it complicates the packet recovery process; since viewers can now switch views during a video multicast (e.g., by subscribing to different multicast channels), a viewer may not have a neighboring peer watching the same view who can locally relay lost packets via CPR.

In this paper, we propose a new CPR packet loss recovery strategy for peers to cooperatively repair packets of video stream *in different views* in a multiview video multicast scenario. In particular, server transmits depth maps in addition to texture maps over WWAN channels, so that a viewer A watching video in a different view than viewer B can nevertheless repair B's packet losses by synthesizing frame(s) in B's viewpoint via depth-image based rendering (DIBR) [8]. Essentially, transmitting additional depth maps constitutes a channel coding overhead for frame recovery of a different view via CPR, while FEC constitutes a channel coding overhead for packet recovery in the same view over WWAN. Hence the crux of the scheme lies in an optimal source / channel bit allocation scheme, so that in this multi-path transmission scenario, just the right amount of resource is devoted to channel coding in the direct WWAN path (FEC) and the indirect CPR path (depth maps), in order to minimize expected distortion. By enabling indirect CPR path recovery through transmission of depth maps, peers can exploit the aforementioned multi-path JSCC gain by allocating a larger portion of the WWAN bit budget to source coding even in a multiview streaming scenario. Experiments show that our proposed CPR scheme outperforms a non-CPR scheme that relies exclusively on WWAN FEC for direct path recovery by up to 3.4dB in video quality.

The outline of the paper is as follows. We first discuss related works in Section 2. We then outline our proposed multiview video multicast system in Section 3. The formulation of the joint source / channel bit allocation problem is formalized in Section 4. We discuss an optimization strategy to solve the bit allocation problem in Section 5. Results and conclusion are presented in Section 6 and 7, respectively.

2. RELATED WORK

Joint source / channel coding (JSCC) is a well-studied problem in signal processing [9], where the objective is to optimally allocate bits out of a fixed transmission budget to source coding (to lower quantization noise) and to channel coding (to contain reconstruction errors due to packet losses), such that the expected distortion at the receiver is minimized. Because the probability of packet loss in WWAN can be substantial, conventional WWAN video multicast schemes [2] optimizing JSCC involve a large overhead of FEC packets, leaving few bits for source coding to combat quantization noise. Our previous work on CPR [3, 5] differs from these traditional JSCC approaches by utilizing a secondary ad-hoc WLAN network for local recovery of packets lost in the primary network, so more bits can be allocated to source coding without inducing more packet losses. [5] studied video quality improvement possible by exploiting such multi-path JSCC gain when optimizing JSCC in a CPR scenario.

We note that our assumption of available multi-homing capable devices (each has multiple network interfaces to connect to orthogonal delivery networks simultaneously) is a common one in the literature [10, 11, 12], where different network optimizations are performed exploiting the multi-homing property. [10] shows that aggregation of an ad-hoc group's WWAN bandwidths can speed up individual peer's infrequent but bursty large content download like web access. [11] proposes an integrated cellular and ad-hoc multicast architecture where the cellular base station delivered packets to proxy devices with good channel conditions, and then proxy devices utilized local WLAN ad-hoc network to relay packets to other devices. [12] shows that smart striping of FEC-protected delay-constrained media packets across WWAN links can alleviate single-channel burst losses, while avoiding interleaving delay experienced in a typical single-channel FEC interleaver. Our current work extends this body of multi-homed literature by optimizing a different application: WWAN multiview video multicast.

In multiview video research, DIBR [8]—encoding of texture and depth maps for view synthesis of a different view—has been proposed for free viewpoint video [13], so that a new view (texture map) can be synthesized at the decoder using texture and depth maps of neighboring view(s) via DIBR techniques such as 3D warping [14]. To the best of the authors' knowledge, this work is the first attempt to cast encoded depth map as a *channel coding* tool (redundant information beyond texture bits) to combat packet losses in a video stream of a different view.

3. WWAN MULTIVIEW VIDEO MULTICAST SYSTEM

In this section, we first overview our WWAN multiview video multicast system. We then describe our source model for multiview video. Finally, we describe our network models for both WWAN used for source-to-peer transmission and ad-hoc WLAN used for peer-to-peer loss recovery.

3.1. System Overview

The components of our proposed WWAN multiview video multicast system, shown in Fig. 1, are the following. A WWAN server transmits different views of a multiview video content, synchronized in time, in different WWAN multicast channels such as Multimedia Broadcast/Multicast Service

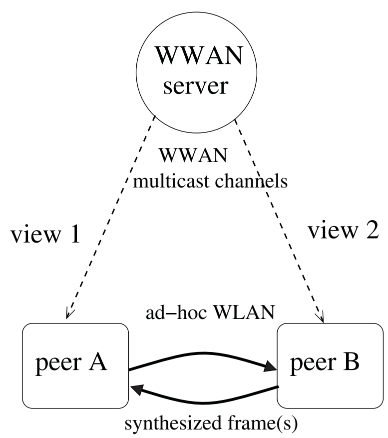

Fig. 1. System Model for Multiview Video Streaming

(MBMS) in 3GPP [15]. Peers interested in a particular view will subscribe to the corresponding multicast channel[1] and can switch views by switching multicast channels every T seconds (an *epoch*). Fig. 1 illustrates two peers, A and B, subscribing to two different views, 1 and 2. Peers are connected to their neighbors via ad-hoc WLAN, providing a secondary network for potential CPR frame recovery. For simplicity of derivation, in this paper we focus on the special case when only two peers watching different views of the same video are participating in CPR. Analysis of CPR network of more than two peers is left for future work.

The WWAN server first multicasts one epoch worth of video to peers. Then during WWAN transmission of the next video epoch, cooperative peers will repair lost packets / frames of the first video epoch. When the server multicasts the third video epoch, peers repair the second video epoch, and video in the first epoch is decoded and displayed. View-switching delay is hence two epochs. An epoch corresponds to the size of a *Group of Pictures* (GOP) which is on the order of 15 frames. For video frame rate of 30 fps, two epochs amounts to one second, which is tolerable.

Each epoch of video of a view j is composed of compressed texture and depth maps of a GOP in that view, and FEC packets to protect these texture and depth maps against WWAN packet losses. Using decoded texture map $F_{i,j}$ and depth map $X_{i,j}$ of instant i in view j, a peer can synthesize texture map $F'_{i,k}$ of the same instant i of a different view k via DIBR to a first-order approximation, so that if a neighboring peer experiences catastrophic packet losses and cannot recover all packets containing texture maps, a synthesized frame $F'_{i,k}$ can be sent from a cooperative peer to the poor peer in order to resume motion-compensation loop (albeit resulting in coding drift due to encoder / decoder mismatch).

Note that an alternative scheme is to differentially code frames in view k using frames in view j as predictors, so that by transmitting the coded residuals, it also enables frame recovery for a poor peer watching view k. In contrast, by transmitting depth maps, our proposed view synthesis procedure can be performed for a number of nearby views k's, each to a different degree of fidelity.

[1]We assume that due to power constraint, a peer will only subscribe to the one multicast channel he/she is interested in viewing at a time.

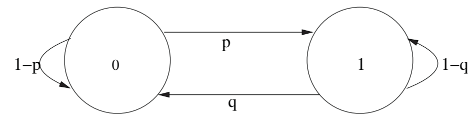

Fig. 2. 2-state Markov model for WWAN packet losses with parameters p and q. 1 (0) indicates a bad (good) state.

3.2. Source Model

We assume a video encoder at the server encodes a GOP of N frames of texture and depth maps of the same view at a time: a leading intra-coded I-frame followed by $N-1$ motion-compensated P-frames, each predicted using the previous frame as predictor. A frame, texture map F_i or depth map X_i, is correctly decoded if it is correctly received, *and* its predictor, F_{i-1} or X_{i-1}, (if any) is correctly decoded. A correctly decoded texture map F_i receives a benefit of $d_Z(i)$ distortion reduction, given texture map quantization level Z. Assuming further that quantization level Z' is used for depth map encoding, each texture F_i and depth map X_i will be mapped to $m_Z(i)$ or $n_{Z'}(i)$ packets respectively for transmission.

If a frame F_i cannot be correctly decoded by a peer, he may recover a synthesized frame F'_i from his neighbor (watching a different view) via CPR. The resulting distortion reduction of synthesized F'_i will be $d'_{Z,Z'}(i) < d_Z(i)$.

For subsequent P-frames F''_k's, $k > i$, after a synthesized frame F'_i received from neighboring peer, because predictor F_i used during encoding differs from received F'_i, a *coding drift* will ensue, resulting in distortion reduction $d''_{Z,Z'}(i,k) < d_Z(k)$, i.e., the distortion reduction of frame F''_k given coding drift in the motion-compensation loop has started at frame F'_i. We discuss how $d_Z(i)$, $d'_{Z,Z'}(i)$ and $d''_{Z,Z'}(i,k)$ in a GOP are generated for optimization and simulation in Section 6.

3.3. Network Models

To model WWAN packet losses, we assume a two-state Markov model with parameters p and q, where packet loss rate (PLR) is $p/(p+q)$ and average burst length is $1/q$. See Fig. 2 for an illustration of a two-state Markov model.

Given the described two-state Markov model, we first define the following terms, which will help ease formulation in Section 4. Let $p(i)$, $i \geq 0$, be the probability of having *exactly* i consecutive correctly delivered packets between two lost packets, following an observed lost packet, i.e., $p(i) = \Pr(0^i 1|1)$. Let $P(i)$ be the probability of having *at least* i consecutive correctly delivered packets, following an observed lost packet, i.e., $P(i) = \Pr(0^i|1)$. $p(i)$ and $P(i)$ can be mathematically written as follows:

$$p(i) = \begin{cases} 1-q & \text{if } i=0 \\ q(1-p)^{i-1}p & \text{o.w.} \end{cases} \tag{1}$$

$$P(i) = \begin{cases} 1 & \text{if } i=0 \\ q(1-p)^{i-1} & \text{o.w.} \end{cases} \tag{2}$$

$q(i) = \Pr(1^i0|0)$ and $Q(i) = \Pr(1^i|0)$ are complementarily defined functions and can be similarly written:

$$q(i) = \begin{cases} 1-p & \text{if } i=0 \\ p(1-q)^{i-1}q & \text{0.w.} \end{cases} \tag{3}$$

$$Q(i) = \begin{cases} 1 & \text{if } i=0 \\ p(1-q)^{i-1} & \text{o.w.} \end{cases} \tag{4}$$

$R(m,n)$ is the probability that there are exactly m packets lost in n packets following an observed lost packet. $R(m,n)$ can be defined recursively as follows.

$$R(m,n) = \begin{cases} P(n) & \text{if } m=0 \;\&\; n \geq 0 \\ \sum_{i=0}^{n-m} p(i)R(m-1,n-i-1) & \text{if } 1 \leq m \leq n \end{cases} \tag{5}$$

$S(m,n)$ is the probability that there are exactly m packets correctly received following an observed received packet. It is a complementary function to $R(m,n)$ and can be similarly defined:

$$S(m,n) = \begin{cases} Q(n) & \text{if } m=0 \;\&\; n \geq 0 \\ \sum_{i=0}^{n-m} q(i)S(m-1,n-i-1) & \text{if} 1 \leq m \leq n \end{cases} \tag{6}$$

For ad-hoc WLAN, we make a simplifying assumption here that a neighboring peer watching a different view has CPR repair bandwidth of $500kbps$ to transmit synthesized frames F'_i's to a peer if the peer experiences irrecoverable WWAN packet losses. The assumption is a reasonable approximation of the CPR recovery capability, given that the WLAN bandwidth is in general larger than WWAN.

Note that to send multiple synthesized frames in sequence $F'_i, \ldots, F'_j$ to a neighboring peer via CPR, the first frame F'_i is an intra-coded I-frame while the subseuqent frames are differentially coded P-frames. Given available $500kbps$ CPR bandwidth, many frames can be repaired via CPR at reasonable texture map quantization level Z.

4. FORMULATION

In this section, we formalize our WWAN resource allocation problem: how to optimally allocate bits among texture maps (to minimize source quantization errors), depth maps (to enable CPR indirect path frame recovery via different-view synthesis) and FEC (to overcome WWAN direct path packet losses) such that the expected distortion at a peer is minimized? Essentially, we must choose the optimal quantization levels Z and Z' for texture and depth map encoding (with leftover WWAN transmission budget for FEC) to achieve minimum distortion.

4.1. WWAN Transmission Constraint

We first formally define the WWAN transmission constraint. Recall from Section 3.2 that each frame i of N total frames in a GOP in an epoch has $m_Z(i)$ and $n_{Z'}(i)$ packets for encoded texture and depth maps given texture and depth map quantization levels Z and Z', respectively. In addition, there are f FEC packets for direct-path (source to peer) WWAN loss protection on these texture and depth map packets. The transmission rate constraint is hence:

$$\sum_{i=1}^{N} m_Z(i) + \sum_{i=1}^{N} n_{Z'}(i) + f \leq C \tag{7}$$

where C is transmission budget for a GOP in packets for one epoch. There is no benefit leaving any transmission budget unused, so given texture and depth map quantization levels Z and Z', we will assume the leftover budget is used entirely for FEC $f_{Z,Z'}$; i.e.:

$$f_{Z,Z'} = C - \sum_{i=1}^{N} m_Z(i) - \sum_{i=1}^{N} n_{Z'}(i) \tag{8}$$

4.2. WWAN Transmission Failure Probability

Let $\alpha_{Z,Z'}$ be the *WWAN transmission failure probability* that $f_{Z,Z'}+1$ or more transmitted packets in a GOP are lost over WWAN to peer A (peer B)[2], so that there is at least one irrecoverable texture or depth map packet. Given the two-state Markov loss model discussed in Section 3.3, we can derive $\alpha_{Z,Z'}$ as follows. The probabilities of the WWAN channel being in bad and good state are $p/(p+q)$ and $q/(p+q)$, respectively. Given a GOP can begin transmission at either good or bad state, $\alpha_{Z,Z'}$ can be written as:

$$\alpha_{Z,Z'} = \frac{p}{p+q} \sum_{k=f_{Z,Z'}+1}^{C} R(k,C) + \frac{q}{p+q} \sum_{k=f_{Z,Z'}+1}^{C} S(C-k,C) \tag{9}$$

In words, (9) sums up the probabilities of losing $k \geq f_{Z,Z'}+1$ of C total packets in a GOP given transmission can start in either good or bad channel state.

4.3. GOP Packet Order

In this paper we assume a simple packet transmission order for a GOP as follows: texture map packets for frame 1 through

[2] We assume symmetry in the optimal solution, so that the WWAN transmission failure probability of peer A is the same as peer B.

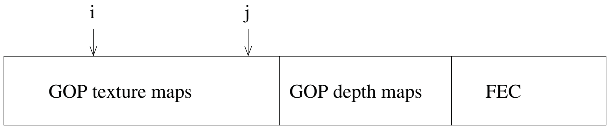

Fig. 3. Packet Order in a GOP

N, depth map packets for frame 1 through N, and FEC packets. This is illustrated in Fig. 3. Hence if a burst loss of length larger than $f_{Z,Z'}$ occurs in the depth and FEC portion of the packet sequence and a peer cannot recover all source packets, the peer can nonetheless use the received texture map packets for decoding and display of this view.

4.4. Distortion Objective

We now derive expression for the expected distortion given chosen quantization levels Z and Z' for texture and depth map encoding. Let $\bar{D}$ to be the initial distortion if no texture packet in a GOP is correctly received. Let i be the index of the first lost packet in a single burst loss in the GOP packet stream, and j be the index of the last lost packet in the same burst (we assume at most one burst loss takes place in a GOP in the analysis). Hence if burst length $j - i + 1 \leq f_{Z,Z'}$, then all texture and depth map packets can be recovered. In this case, we calculate the distortion reduction to be D_Z^0:

$$D_Z^0 = \sum_{k=1}^{N} d_Z(k) \tag{10}$$

If $j - i + 1 > f_{Z,Z'}$, we cannot recover all texture and depth map packets in the direct path given $f_{Z,Z'}$ FEC packets were used. In this case, we receive distortion reductions from frames F_i's constructed using texture map packets prior to the burst loss, *plus* distortion reductions due to synthesized frames F_i''s received from the peer's neighbor via CPR, *and* subsequent distortion reductions from correctly WWAN transmitted frames F_i'''s after peer's video decoder resumes the motion compensation loop.

First, let $\lambda(i)$ be index of the texture frame that contains packet i in the GOP packet stream. $\lambda(i)$ can be written as:

$$\lambda(i) = \min_{1 \leq \phi \leq N} \phi \quad \text{s.t.} \sum_{k=1}^{\phi} m_Z(i) \geq i \tag{11}$$

$\lambda(j)$ is hence the last texture frame affected by the single burst loss. If we assume CPR bandwidth is large enough for neighboring peer to transmit all lost frames, then the neighboring peer will then locally transmit synthesized frames $\lambda(i), \ldots, \lambda(j)$ for recovery via CPR, which has distortion reductions of $d'_{Z,Z'}(\lambda(i)), \ldots, d'_{Z,Z'}(\lambda(j))$. Subsequent frames s's have distortion reductions[3] $d''_{Z,Z'}(\lambda(j), s)$, $s > \lambda(j)$, thereafter, each $d''_{Z,Z'}(\lambda(j), s) < d_Z(s)$ due to coding drift.

[3]In general, synthesized frame distortion reduction $d'_{Z,Z'}(s)$ is smaller than drift-affected distortion reduction $d''_{Z,Z'}(i, s)$, $i < s$, since coding drift tends to decrease over time due to coded intra-blocks. Hence a synthesized frame F_i' should be used for repair only when the original frame F_i is lost in the direct path.

We divide the analysis of resulting distortion reduction due to single burst loss into two cases: GOP transmission starts when the channel is in bad (good) state. First, let $T_Z = \sum_{i=1}^{N} m_Z(i)$ be the total number of texture map packets in a GOP given texture map quantizer Z. If we begin at a bad state, expected distortion reduction $D_{Z,Z'}^B$ can be written as follows:

$$D_{Z,Z'}^B = \left(\frac{p}{p+q}\right) \left[q(1-p)^{T_Z - 1} D_Z^0 + L_{Z,Z'}^B\right] \tag{12}$$

where (12) states that if all T_Z texture map packets are received consecutively, then distortion reduction is D_Z^0. Otherwise, distortion reduction $L_{Z,Z'}^B$ depends both on the location i of the start of the burst loss, and length of the burst loss $j - i + 1$. Hence to find $L_{Z,Z'}^B$, one must consider all combinations of burst start location i and burst length $j - i + 1$:

$$\begin{aligned} L_{Z,Z'}^B \quad \approx \quad & \sum_{i=2}^{T_Z} \sum_{j=i+1}^{T_Z+f-1} q(1-p)^{i-2} p(1-q)^{j-i} q \, D_{Z,Z'}(i,j) + \\ & + \sum_{j=1}^{T_Z+f-1} (1-q)^j q \, D_{Z,Z'}(1,j) \end{aligned} \tag{13}$$

where each term in the first double summation in (13) corresponds to the probability that $i - 1$ packets (for $i \geq 1$) were first received, followed by $j - i + 1$ packets of single burst loss, and followed by a received packet. We denote the resulting distortion for given i and j as $D_{Z,Z'}(i,j)$. Each term in the summation in the second line of (13) corresponds to the probability that j packets are first lost, followed by a received packet.

We now derive the resulting distortion $D_{Z,Z'}(i,j)$ given burst starts at packet i and burst length is $j - i + 1$. If burst length $j - i + 1 \leq f_{Z,Z'}$, then peer can obviously recover all source packets and $D_{Z,Z'}(i,j) = D_Z^0$. Otherwise, the distortion reduction is $D_{Z,Z'}^1(i,j)$:

$$D_{Z,Z'}(i,j) = \begin{cases} D_Z^0 & \text{if } j - i + 1 \leq f_{Z,Z'} \\ D_{Z,Z'}^1(i,j) & \text{o.w.} \end{cases} \tag{14}$$

Given burst length $i - j + 1 > f_{Z,Z'}$, $D_{Z,Z'}^1(i,j)$ is the sum of the distortion reductions prior to the burst loss $d_Z(k)$'s, $k \leq \lambda(i) - 1$, distortion reductions $d'_{Z,Z'}(k)$'s of synthesized frames $\lambda(i) \leq k \leq \lambda(j)$, and subsequent distortion reductions $d''_{Z,Z'}(\lambda(j), k)$'s, $k > \lambda(j)$:

$$\begin{aligned} D_{Z,Z'}^1(i,j) \quad = \quad & \sum_{k=1}^{\lambda(i)-1} d_Z(k) + \\ & +(1 - \alpha_{Z,Z'}) \left[\sum_{k=\lambda(i)}^{\lambda(j)} d'_{Z,Z'}(k) + \sum_{k=\lambda(j)+1}^{N} d''_{Z,Z'}(\lambda(j), k) \right] \end{aligned} \tag{15}$$

Expected distortion reduction $D^G_{Z,Z'}$ when the channel begins at a good state can be similarly derived as $D^B_{Z,Z'}$, and hence the derivation is not presented for brevity. We simply present the derived equations as follows:

$$D^G_{Z,Z'} = \left(\frac{q}{p+q}\right)\left[(1-p)^{T_Z} D^0_Z + L^G_{Z,Z'}\right] \quad (16)$$

$$L^G_{Z,Z'} = \sum_{i=2}^{T_Z} \sum_{j=i+1}^{T_Z+f-1} (1-p)^{i-1} p (1-q)^{j-i} q \, D_{Z,Z'}(i,j) + + \sum_{j=1}^{T_Z+f-1} p(1-q)^{j-1} q \, D_{Z,Z'}(1,j) \quad (17)$$

To summarize, the expected distortion reduction of the entire GOP $D^o_{Z,Z'}$ given quantization levels Z and Z' for texture and depth maps is written as follows:

$$D^o_{Z,Z'} = \bar{D} - D^B_{Z,Z'} - D^G_{Z,Z'} \quad (18)$$

The goal is to minimize (18) with two degrees of freedom: texture and depth map quantization levels Z and Z'. Having formalized the bit allocation problem, we next describe an optimization strategy to solve this efficiently.

5. OPTIMIZATION

Given the definition of objective and constraints in Section 4, we now discuss our strategy on how to find the optimal texture and depth quantization levels Z and Z' (and FEC packet numbers f subsequently) given WWAN transmission budget C in packets. Suppose optimal channel coding resource can be optimally divided between FEC (for direct path WWAN loss protection) and encoding of depth maps (for indirect path CPR recovery). Intuitively then, the expected distortion $D^o_{Z,Z'}$ should have a unique minimum Z^* as a function of texture map quantization level Z, such that when $Z < Z^*$, insufficient number of bits are spent on channel coding and the channel noise overwhelms the distortion term, and when $Z > Z^*$, insufficient number of bits are spent on source coding and the source quantization noise overwhelms the distortion term. We can hence conclude that $D^o_{Z,Z'}$ should be unimodal as a function of Z with a unique local minimum.

Given texture map quantization level Z, the optimal division of channel resource between depth maps and FEC is also straightforward. It is due to the well known "cliff" effect of FEC: a sufficient level of FEC recovers most losses, above which the gain in recovery is minimum, and below which the deterioration of recovery is catastrophic. So a sensible strategy to optimally allocate resource between depth maps and FEC, is to first allocate all channel coding budget to FEC, then incrementally shift resource from FEC to depth maps, until an increase in expected distortion is observed.

From the above discussion, the optimal resource allocation problem between source and channel coding can be done iteratively as shown in Fig. 4.

1. Initialize texture map quantization level Z.
2. Given Z, find optimal channel coding resource allocation by first assigning all resource to FEC, then incrementally reallocate resource to depth mapss until an increase in distortion is observed.
3. If distortion $D^o_{Z,Z'}$ has reached local minimum, stop. Else, update Z via binary search and go to step 2.

Fig. 4. Iterative search algorithm to find optimal texture and depth map quantization levels Z and Z'.

6. EXPERIMENTATION

6.1. Experimental Setup

To test the effectiveness of our bit allocation scheme, we set up our simulation experiments as follows. For video source, we use view 3 and 5 of MPEG multiview test sequence `ballroom` [16] for video encoding, using H.264 JM version 16.2 [17]. GOP of 15 frames are encoded to a leading I-frame plus 14 P-frames using the same quantization level $Z \in \{35, 37, 39, 41, 43, 45\}$. Distortion reductions $d_Z(i)$'s are computed PSNRs given Z.

Each depth map $X_{i,j}$ associated with a texture map $F_{i,j}$ is estimated as follows. $X_{i,j}$ is first estimated using the quick color block matching-based method from [18], then smoothed [19] enforcing the coherence in the depth maps across the spatial and temporal dimensions and views. Depth maps are encoded using quantization level $Z' \in \{41, 43, 45\}$. For given texture and depth map quantization level Z and Z', distortion reductions for synthesized frames $d'_{Z,Z'}(i)$'s and coding-drift affected frames $d''_{Z,Z'}(k,i)$'s, $k < i$, can be calculated. Assuming a maximum transmission unit (MTU) of 1500 bytes, quantization levels Z and Z' also induce the number of packets for each frame F_i in the GOP, $m_Z(i)$ and $n_{Z'}(i)$ for texture and depth map, respectively. We assume WWAN packet loss rate ranging from 0.1 to 0.3, and average burst length of 20 and 25 packets. As mentioned, WLAN bandwidth available for CPR is 500kbps.

6.2. Experimental Results

We first show the computed distortion reductions (in Peak Signal-to-Noise Ratio PSNR) of our proposed bit allocation scheme using derived equations in Section 4 in Fig. 5. We can see that for all three combinations of burst length and packet loss rate, there is a unique maximum at which the performance of our proposed scheme is at its best. This agrees with our intuition where below the theoretical maximum, there is not sufficient channel protection deployed and channel noise dominates the distortion term, and where above the maximum, there are not sufficient source coding bits and source quantization noise dominates. For the current setup, our scheme achieved this maximum at texture quantization

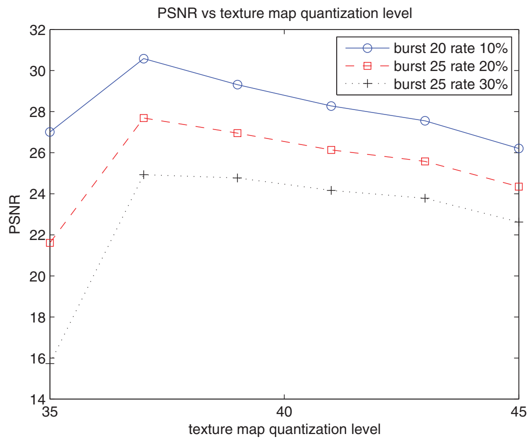

Fig. 5. Using `ballroom` sequence, showing derived expected PSNR Performance of proposed CPR scheme at different texture map QP, for burst length $20, 25, 25$ and loss rate $0.1, 0.2, 0.3$.

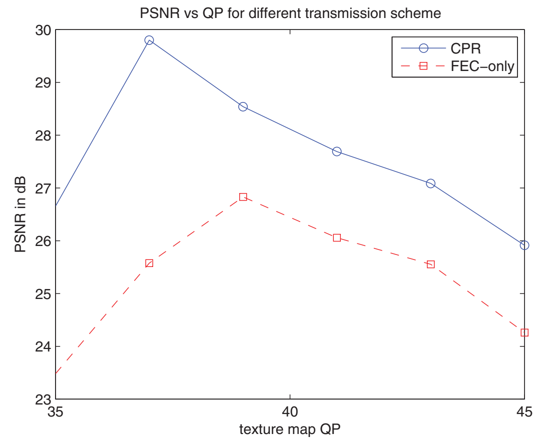

Fig. 6. Using `ballroom` sequence, comparing PSNR performance of proposed CPR scheme versus FEC-only scheme at different texture map QP, for burst length 25 and loss rate 0.2.

level $Z = 37$ for all three combinations of burst length and packet loss rate.

We have also conducted simulation experiments where the two-state Markov channel model for WWAN was simulated, virtual packets in WWAN were transmitted, ad-hoc CPR packets were exchanged, and resulting distortion for each epoch was recorded. Fig. 6 and Fig. 7 show the resulting PSNR versus texture map quantization level Z, for combination of burst length and packet loss rate $(25, 0.2)$ and $(25, 0.3)$, for our proposed scheme and a competing scheme that uses FEC only for direct path loss protection. We first see that the maximum performance is achieved at $Z = 37$ as predicted in our computed plot in Fig. 5 while the FEC-only case get to the peak at the $Z = 39$. Further, we see that our proposed scheme outperformed the FEC-only scheme by 2.9dB and 3.4dB for the two combinations, showing the merit of our proposal.

7. CONCLUSION

In the paper, we present a new Cooperative Peer-to-peer Repair (CPR) strategy for WWAN multiview video multicast. The key idea is to transmit depth maps in additional to texture maps, so that lost frames of a peer watching a different view can be partially repaired by neighbor via depth-image based rendering (DIBR). Since transmitting depth maps constitutes an overhead like forward-error correction (FEC), we formulate our problem as a bit allocation and find the optimal division of transmission budget among texture and depth map coding and FEC. Results show our scheme can be up to $3.4dB$ improvement over a scheme that relies on FEC only under burst loss condition.

8. REFERENCES

[1] J. Crowcroft and K. Paliwoda, "A multicast transport protocol," in *ACM SIGCOMM*, New York, NY, August 1988.

[2] J. Zfzal, T. Stockhammer, T. Gasiba, and W. Xu, "Video streaming over MBMS: A system design approach," in *Journal of Multimedia*, August 2006, vol. 1, no.5, pp. 25–35.

[3] X. Liu, G. Cheung, and C.-N. Chuah, "Structured network coding and cooperative wireless ad-hoc peer-to-peer repair for WWAN video broadcast," in *IEEE Transactions on Multimedia*, 2009, vol. 11, no.4.

[4] O. Alay, P. Liu, Y. Wang, E. Erkip, and S. Panwar, "Error resilient video multicast using randomized distributed space-time codes," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, March 2010.

[5] X. Liu, G. Cheung, and C.-N. Chuah, "Joint source/channel coding of wwan multicast video for a cooperative peer-to-peer collective using structured network coding," in *IEEE International Workshop on Multimedia Signal Processing*, Rio de Janeiro, Brazil, October 2009.

[6] T. Fujii, K. Mori, K. Takeda, K. Mase, M. Tanimoto, and Y. Suenaga, "Multipoint measuring system for video and sound—100 camera and microphone system," in *IEEE International Conference on Multimedia and Expo*, Toronto, Canada, July 2006.

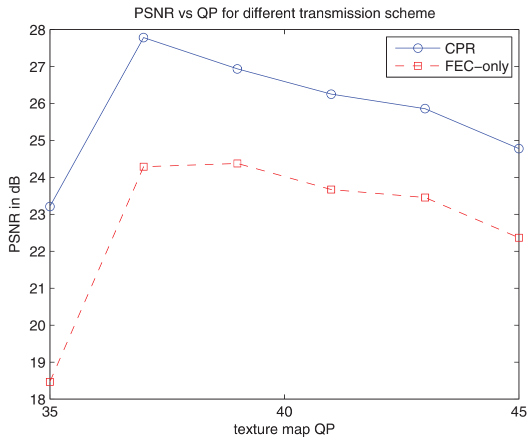

Fig. 7. Using `ballroom` sequence, comparing PSNR performance of proposed CPR scheme versus FEC-only scheme at different texture map QP, for burst length 25 and loss rate 0.3.

[7] G. Cheung, A. Ortega, and N.-M. Cheung, "Generation of redundant coding structure for interactive multiview streaming," in *Seventeenth International Packet Video Workshop*, Seattle, WA, May 2009.

[8] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *IEEE International Conference on Image Processing*, San Antonio, TX, October 2007.

[9] Gene Cheung and Avideh Zakhor, "Bit allocation for joint source/channel coding of scalable video," in *IEEE Transactions on Image Processing*, March 2000, vol. 9, no.3.

[10] P. Sharma, S.-J. Lee, J. Brassil, and K. Shin, "Distributed communication paradigm for wireless community networks," in *IEEE International Conference on Communications*, Seoul, Korea, May 2005.

[11] Randeep Bhatia, Li (Erran) Li, Haiyun Luo, and Ram Ramjee, "ICAM: Integrated cellular and ad hoc multicast," in *IEEE Transactions on Mobile Computing*, August 2006, vol. 5, no. 8, pp. 1004–1015.

[12] G. Cheung, P. Sharma, and S.J. Lee, "Smart media striping over multiple burst-loss channels," in *IEEE Transactions on Selected Topics in Signal Processing*, August 2007, vol. 1, no.2, pp. 319–333.

[13] T. Fujii and M. Tanimoto, "Free viewpoint TV system based on ray-space representation," in *Proceedings of SPIE*, 2002, vol. 4864, p. 175.

[14] H.-Y. Shum, S.-C. Chan, and S. B. Kang, *Image-Based Rendering*, Springer, 2007.

[15] *Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service (MBMS) user services; stage 1 (Release 6) (3GPP TS.26.246 version 6.3.0*, March 2006.

[16] "Test sequences at MERL," ftp://ftp.merl.com/pub/avetro/mvc-testseq.

[17] "The TML project web-page and archive," http://kbc.cs.tu-berlin.de/ stewe/vceg/.

[18] M. Tanimoto, T. Fujii, and K. Suzuki, "Multi-view depth map of Rena and Akko & Kayo," October 2007, MPEG Doc. M14888.

[19] E. Ekmekcioglu, V. Velisavljević, and S. T. Worrall, "Content adaptive enhancement of multi-view depth maps for free viewpoint video," *IEEE Selected Topics in Signal Processing*, to appear, April 2011.