Joint Power and Channel Resource Optimization in Soft Multi-View Video Delivery

TICAO ZHANG AND SHIWEN MAO (Fellow, IEEE)
Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA
Corresponding author: Shiwen Mao (smao@ieee.org)
This work was supported in part by the U.S. NSF under Grant IIP-1822055, Grant ECCS-1923717, and Grant CNS-1702957, and in part by the Wireless Engineering Research and Education Center (WEREC), Auburn University, Auburn, AL, USA.

ABSTRACT Existing wireless multi-view video (MVV) transmission schemes use digital compression to achieve a better coding efficiency. However, the digital schemes suffer from the cliff effect, which refers to the phenomenon that the video quality is a step function of wireless channel quality. In this paper, we first consider a soft MVV transmission scheme where the correlations between the inter-view data and texture-depth data are exploited by a 5-dimensional discrete cosine transform (5D-DCT). The linearly transformed 5D-DCT signals are modulated in an analog manner so that the video quality gracefully improves when the channel quality becomes better. The cumbersome bit and rate controls in digital solutions are replaced by simple power controls. Second, as with the increase of the number of cameras and data depths, the data size of MVV increases linearly. To reduce the heavy data traffic in soft MVV transmission, we proposed efficient resource (bandwidth and power) allocation algorithms. Simulations results demonstrate that the proposed distortion-resource (DR) optimization algorithm can ensure a best viewing quality under a resource constraint and the proposed resource-distortion (RD) optimization algorithm can minimize the resource usage for a target video quality requirement. Third, the impact of power control across texture and depth frame and the impact of view positions on synthesized virtual view quality are investigated. The efficacy of the proposed algorithm on both the reference viewpoint as well as the virtual viewpoint is verified via simulations.

INDEX TERMS Soft video delivery, SoftCast, multi-view video, resource allocation.

I. INTRODUCTION
Multimedia has become the most popular application for emerging network paradigms [1]–[3]. Recently, Multi-view videos (MVV) are emerging in various application domains such as education, healthcare, and 3D-home entertainment. It is a fundamental technology in virtual-reality (VR), naked-eye 3D, and free-viewpoint video streaming [4], [5]. Fig. 1 shows an example of the MVV transmission system, where a number of cameras are deployed at different positions. Each camera captures both texture maps (images) and depth maps (distances from the objects). These texture and depth information is known as multi-view plus depth (MVD). The MVD information is encoded via texture and depth encoding and then transmitted to the receiver via, e.g., a wireless channel. After decoding, the receiver synthesizes intermediate virtual viewpoint using depth-image-based rendering (DIBR) from the received MVD frames [6]. As can be seen, virtual view 2 can be synthesized with the texture and depth data from its left view (i.e., view 1) and right view (i.e., view 3). The receiver can then select its favorite viewing angles and enjoy an enhanced viewing experience.

Compared with conventional single-view video streaming, an MVV streaming usually generates a large amount of video data. The video data grows linearly with the product of the number of cameras and the number of frame depths. When it comes to 4K (or 8K) videos, the heavy data traffic would consume a considerable amount of wireless resources and hence may become infeasible for realtime transmission. Currently, one solution is an independent coding on both texture and depth, which is a backward compatible extension of the H.264/AVC standard [7], [8]. Each viewpoint is encoded separately and only the view corresponding to the user’s current selected viewpoint is transmitted [9], [10]. To further decrease the redundancy, multi-view video coding (MVC) is proposed as an extension of the H.264/MPEG-
AVC standard. It introduces the concept of disparity-compensated prediction [11]. By exploiting the inter-view dependency, MVC enables a higher compression efficiency than separate view coding [12]. However, to arrive at a given level of video quality, the transmission rate still increases nearly linearly with the number of views. The large data overhead still poses a challenging problem.

Moreover, the digital video compression that current MVC adopts, relies on Shannon’s separate source and channel coding approach [13]. The video is encoded at the transmitter first at a specific coding rate, which is called source coding. Then adaptive modulation and channel coding is adopted to facilitate reliable bit transmission. Over the past decades, coded transmission has dominated the existing wireless video transmissions. However, there still exists several weaknesses in this framework. First of all, the quantization process involved is a lossy process and the encoded video quality depends on the coding rate at the source. Once the video is encoded, its quality will not improve any more even if the channel condition allows the transmission of a better quality video. When the channel quality degrades, the received video may be garbled due to error spread and packet loss. This is called the cliff effect [14]. Moreover, the efficiency of channel coding greatly depends on the timeliness and precision of the channel feedback. In practice, the encoder adjusts the source coding rate according to the buffer size and the transmitter chooses the most appropriate modulation rate to transmit the packet. This usually requires a quite complicated bit and power allocation solution. In MVV, where a smooth navigation with 3D scenes with a minimum delay is required, timely and accurate feedback and a large buffer size are more indispensable to process and transmit the huge video data traffic.

To overcome the problems in conventional digital video transmission, soft video delivery is proposed in [15], where joint channel and source coding is exploited. By skipping quantization and entropy coding, the video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.

In this paper, we will incorporate the soft video delivery technique for wireless MVV transmissions. We use a five-dimensional discrete cosine transformation (5D-DCT) to jointly process video texture and depth frames from different cameras. The output is scaled and modulated in an analog manner. The video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.

In this paper, we will incorporate the soft video delivery technique for wireless MVV transmissions. We use a five-dimensional discrete cosine transformation (5D-DCT) to jointly process video texture and depth frames from different cameras. The output is scaled and modulated in an analog manner. The video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.

In this paper, we will incorporate the soft video delivery technique for wireless MVV transmissions. We use a five-dimensional discrete cosine transformation (5D-DCT) to jointly process video texture and depth frames from different cameras. The output is scaled and modulated in an analog manner. The video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.

In this paper, we will incorporate the soft video delivery technique for wireless MVV transmissions. We use a five-dimensional discrete cosine transformation (5D-DCT) to jointly process video texture and depth frames from different cameras. The output is scaled and modulated in an analog manner. The video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.

In this paper, we will incorporate the soft video delivery technique for wireless MVV transmissions. We use a five-dimensional discrete cosine transformation (5D-DCT) to jointly process video texture and depth frames from different cameras. The output is scaled and modulated in an analog manner. The video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.

In this paper, we will incorporate the soft video delivery technique for wireless MVV transmissions. We use a five-dimensional discrete cosine transformation (5D-DCT) to jointly process video texture and depth frames from different cameras. The output is scaled and modulated in an analog manner. The video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.

In this paper, we will incorporate the soft video delivery technique for wireless MVV transmissions. We use a five-dimensional discrete cosine transformation (5D-DCT) to jointly process video texture and depth frames from different cameras. The output is scaled and modulated in an analog manner. The video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.
transmission is presented. We consider practical DR optimization in Section IV and and RD optimization in Section V. Then, extensive simulations are performed to demonstrate the advantages of the proposed resource allocation algorithms in MVV transmission in Section VI. Finally, Section VII concludes the paper.

II. RELATED WORK

We divide our discussion on related work into two parts. In this section, we first provide a brief introduction of the recent work on MVV and MVD transmissions. We then review related works on soft video transmission.

A. MULTIVIEW VIDEO TRANSMISSION

MVV is a simple and effective extension of MV by providing camera-depth information. It enables efficient depth-image based rendering so that virtual views can be generated from a limited number of source views. The texture data is captured by multiple cameras. Meanwhile, every texture is accompanied with depth information. The 3D HEVC standard [12] exploits the dependencies between texture and depth information to remove the redundancy. The encoding process is realized by spatial prediction within each frame, temporal motion-compensation between different frames, transform coding of the prediction residual, and entropy coding [17]. To code the depth data, new intra coding modes, modified motion compensation, and motion vector coding are used.

Delivering MVV/MVD content over existing MVV streaming networks faces many challenges including network bandwidth variation, packet loss, delay and client view selection uncertainty [18]. So far, interactive MVV streaming, 3D video coding, and practical system implementation have been studied, e.g., see recent works [4], [5], [19]–[21]. However, very few works are focused on wireless MVV streaming. As introduced before, delivering 3D MVV video is a very challenging task over today’s wireless networks. It requires to carry a potentially much larger data traffic generated from large number of cameras, with a strict requirement on latency on the complex wireless environments. In [22], the authors incorporated MVV with the multiple input multiple output (MIMO) technique that employed precoding and spatial multiplexing for simultaneous transmissions. A resource control algorithm was proposed to achieve unequal error protection against channel errors. Reference [23] considered MVV transmission with multiple description coding. Multiple descriptions from texture and depth data of adjacent views were transmitted through separate wireless channels, so that the multi-path diversity could be exploited for improved reliability.

B. SOFT VIDEO TRANSMISSION

The interesting concept of soft video delivery (SoftCast) was first proposed in [15], [24]. Unlike traditional digital video transmission, SoftCast builds an analog code that achieves the compression-protection tradeoff with a suitable power allocation. Experiments demonstrate that the cliff effect in conventional digital video transmission can be avoided and users can enjoy a graceful video quality improvement according to the channel condition.

The SoftCast concept attracted considerable interest in the community. For example, Ref. [25] replaced the power allocation scheme in SoftCast with bit allocations, while Ref. [26], [27] combined the benefit of SoftCast and conventional digital video coding by considering an analog-digital hybrid coding scheme. In [28], the authors proposed an optimal channel and power allocation scheme under fast fading channels. The multiple antenna technique was exploited to improve the system performance. For example, Ref. [29] decomposed the MIMO channel into parallel sub-channels by MIMO precoding. By assigning high priority DCT coefficients to higher quality sub-channels, the reconstructed video quality could be optimized. Ref. [30] extended SoftCast to a wireless video multicast scenario with receiver antenna heterogeneity. The proposed scalable video multicast system allowed receivers to have a reconstructed video quality that was commensurable with the number of equipped antennas. In [31], a curve-fitting based source control algorithm was developed to find the cost distortion relationship, where cost consisted of bandwidth and transmit power, for soft video delivery.

To the best of our knowledge, Ref. [32] was the first work that investigated the soft video transmission for MVD. In this work, the metadata overhead could be greatly reduced with the proposed Gaussian Markov random filed (GMRF) model, and thus a better video quality could be achieved. However, compared with the overhead incurred by metadata, there exist a huge redundancy in the coded video data, especially in MVD transmission. How to jointly optimize the resources used, while achieving a satisfactory video quality, is still an open problem. In [33], we proposed a blind data detection method that recovered received video from the squared amplitude of received signals, which was almost metadata free. This work [33] was designed for a generic video and the AWGN channel.

III. SOFT VIDEO TRANSMISSION FOR MVV DELIVERY

In MVV delivery, the transmitter adopts multiple cameras to record the multi-color texture and depth frames. When the transmitter is notified via the feedback channel of the receiver preferred virtual viewpoint, it captures the data at several adjacent cameras near the requested virtual viewpoint. These captured data are then encoded and transmitted to the receiver. At the decoder side, the requested viewpoint is synthesized from the decoded texture and depth frames via DIBR.

In this section, we consider the case when there is plenty of bandwidth in the transmission channel and all the DCT coefficients will be transmitted to the receiver. Therefore we will focus on power allocation problem. At the encoder, a 5D-DCT is used for the entire texture and depth frames in one group of picture (GOP), which is a sequence of
successive MVD video frames. After power allocation for each DCT coefficient, the DCT coefficients are then mapped to I (in-phase) and Q (quadrature-phase) components for analog wireless transmission.

Specifically, the DCT coefficients are divided into $N$ rectangular chunks with size $h \times w$. Let $x_i[j]$ denote the $j$th DCT coefficient in the $i$th chunk. We scale all the DCT coefficients in the $i$th chunk by a common scaling factor $g_i$ for noise reduction. The scaled coefficient $s_i[j]$ is denoted as follows

$$s_i[j] = g_i \cdot x_i[j].$$

(1)

This analog-like scaling is also called power allocation. The optimal power scaling factor is obtained by minimizing the end-to-end distortion under a constrained power budget $P$ as follows [28].

**IV. DISTORTION RESOURCE (DR) OPTIMIZATION**

We next consider the more realistic case with limited channel resource (in the form of time slots or frequency bands). In soft video delivery, each scaled DCT chunk is transmitted in different time slots or frequency bands. To transmit a video of a large size, e.g., MVDs, a considerable amount of channel resources is required. This makes it hard for real time delivery. Fortunately, due to the compacting nature of DCT, most of the DCT components in high spatial frequency domain tend to have very small values. Therefore, we can discard a certain amount of high-frequency DCT chunks to satisfy the channel resource constraints, while not degrading the video quality too much. It is also worth noting that even if there are sufficient channel resources, it may still be helpful to drop some DCT chunks, since the saved power (by dropping some chunks) can be utilized more efficiently by re-allocating it to other more important chunks, especially when the power constraint is stringent. Then we have a joint chunk selection and power allocation problem.

**A. PROBLEM STATEMENT**

Given a set of $N$ chunks of DCT coefficients with average energy denoted by $\lambda_1, \lambda_2, \ldots, \lambda_N$. Without loss of generality, we assume $\lambda_i \geq \lambda_j$, for all $i > j$. Let $M$ be the amount of available channel resources (e.g., time or frequency slots) and $P$ be the total power constraint for each chunk (and $P$ be the SNR budget for each chunk). We use a binary channel allocation vector $k = [k_1, k_2, \ldots, k_N]$ to denote the chunk selection of each GOP: $k_i = 0$ indicates that chunk $i$ is discarded, and $k_i = 1$ means that chunk $i$ is transmitted via a channel resource slot. We aim to find the optimal channel allocation $k^*$ and the optimal power allocation $\rho^* = [\rho_1^*, \rho_2^*, \ldots, \rho_N^*]$, so that the total video distortion is minimized. Mathematically, the problem can be formulated as follows.

**P1**

$$\min_{k, \rho} \text{MSE} = \frac{1}{N} \sum_{i=1}^{N} \frac{\lambda_i}{\tilde{\rho}_i + 1}$$

s.t. $\sum_{i=1}^{N} k_i \lambda_i \leq M$

$$\sum_{i=1}^{N} k_i \rho_i \leq \tilde{P},$$

(9)

(10)

Intuitively, since $k_i$ only takes binary values, $M$ should be in the range of $[0, N]$. If $M = N$, Problem (P1) will be exactly the same as Problem (P0). If $M < N$, to minimize distortion, we will retain the largest $M$ chunks, which we refer to as high-priority (HP) data, and discard the remaining smaller chunks, which we call low-priority (LP) data, i.e.,

$$k_i^* = \begin{cases} 1, & \text{if } i \leq M \\ 0, & \text{if } i > M. \end{cases}$$

(13)
The problem then becomes finding the optimal value $M^*$ and the optimal power allocation $\{\rho_i^*\}$. Similarly, by the Lagrange multiplier method, we can derive the optimal solution of $\rho_i$ as

$$\rho_i^* \approx \frac{N \sqrt{\lambda_i}}{\sum_{i=1}^{M} \sqrt{\lambda_i}} \cdot \tilde{P} \cdot k_i. \quad (14)$$

Then the MSE can be expressed as

$$\text{MSE} = \frac{1}{N} \left( \sum_{i=1}^{M} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=M+1}^{N} \lambda_i \right). \quad (15)$$

It can be seen that the total MSE in (15) can be expressed as a function of the power budget $P$ and the channel resource constraint $M$. Increasing the power budget will lead to an increase of $\rho_i$, which will help decrease the distortion. In other words, MSE is a monotone function in terms of $P$ (or $\tilde{P}$). However, it is still unclear how to find the optimal value of $M$ because of its discrete value and that it appears in the superscription of the summation term in (14). Such discrete nature makes it hard to obtain the optimal value $M^*$ in closed-form as what we did in the case of $\rho_i^*$.

**B. A GREEDY SEARCH APPROACH**

To find the optimal value $M^*$, we propose an exhaustive search based algorithm, as presented in Algorithm 1. The main idea is that, the transmitter has full knowledge of $\lambda_i$’s and the total power budget $P$, it can find the number of chunks that minimize the MSE by searching all the possible discrete channel resources in an exhaustive manner. With the optimal chunk selection, the video is actually compressed but without too much performance degradation, and users can enjoy a better experience since the transmission time is saved and the amount of video traffic is reduced. Meanwhile, the saved channel resources can be utilized by other users in the network.

**Algorithm 1 Distortion Resource Optimization Algorithm**

1. Initialize $\lambda_i = \mathbb{E}[|x_i|^2]$, for all $i$;
2. for $n = 1, 2, \ldots, N$ do
3. $M \leftarrow n$;
4. Calculate $\rho_{i,n}$ according to (13) and (14);
5. Calculate MSE$_n$ according to (15);
6. Calculate PSNR$_n$ as in (30) in Section VI-A;
7. end for
8. PSNR$_{\text{max}} = \max_{n \in \{1, 2, \ldots, N\}} \text{PSNR}_n$;
9. for $n = 1, 2, \ldots, N$ do
10. if PSNR$_n > \alpha \cdot \text{PSNR}_{\text{max}}$ then
11. $M^* \leftarrow n$;
12. break;
13. end if
14. end for
15. Calculate $\{k_i^*\}$ and $\{\rho_i^*\}$ using $M^*$ as in (13)-(14);
16. Output $M^*$, $\{k_i^*\}$, and $\{\rho_i^*\}$;

In Algorithm 1, we first compute the initial energy distribution of the chunks in each GOP. Based on the information, we find the optimal channel allocation and power allocation in an exhaustive manner. By Line 8, the algorithm can actually terminate and output the optimal channel number $M^*$, the optimal chunk selection $\{k_i^*\}$, and the optimal power control $\{\rho_i^*\}$. However, in Section VI, we observe from simulations that the Distortion-Resource (DR) curve tends to have a flat tail (e.g., see Figs. 3 and 4). This means that we are using a much larger number of channel resources to achieve only a slight improvement in PSNR. This is obviously inefficient. Hence, we introduce a control parameter $0 < \alpha \leq 1$ in Line 10 to search for a sub-optimal solution. By slightly sacrificing the PSNR performance, we can significantly reduce the video traffic and the channel resource usage.

**C. COMPLEXITY ANALYSIS**

In Algorithm 1, we traverse all the possible $n$ values in a greedy manner and for each $n$, the involved operations are all linear. Hence the complexity of Algorithm 1 is $O(N)$, which is negligible. For one GOP, the main complexity comes from the energy sorting process of the $N$ chunks, with complexity $O(N \log N)$. However, in practice, we do not necessarily need to sort all these chunks strictly according to their energy distribution. Instead, a more feasible way would be to sort these chunks in a zigzag scanning manner, which is used in the JPEG image compression. In this way, the sorting process can be avoided and the complexity can be greatly reduced. Moreover, in simulations, we find that for consecutive video frames, the optimal value $M^*$ does not change too much. This means that we may only need to find the corresponding $M^*$ for the first GOP, and then applies this $M^*$ to the remaining GOPs, which further reduces the complexity.

**V. RESOURCE DISTORTION (RD) OPTIMIZATION**

**A. PROBLEM STATEMENT**

Note that for a specific video sequence, different GOPs may have different levels of compressibility. In DR optimization, under a fixed power and channel budget, the distortion of consecutive GOPs may have large variations, which become quite annoying for viewers (although the overall PSNR could be maximized). In addition, human eyes are less sensitive to the differences of videos when the PSNR is very high. Based on these two observations, keeping the distortion relatively more stable may be a better choice. The saved power and channel resources this way can also be utilized by other users. Therefore, our problem becomes to find a good combination of chunk selection and power allocation for a target distortion. This problem is called resource distortion (RD) optimization.

Since RD optimization involves distortion, channel, and power usage, it is a three-dimensional optimization problem, which is hard to solve. However, by fixing one factor, we can decompose the difficult problem into two sub-problems. For example, we can formulate a power distortion optimization problem that aims to minimize the power resources usage under a target distortion constraint MSE and a channel usage...
constraint $M$. The problem can be stated as follows.

$$(\text{P2}) \min\limits_{\rho} \tilde{P} = \frac{1}{N} \sum_{i=1}^{m} \rho_i$$

$$\text{s.t.} \frac{1}{N} \left( \sum_{i=1}^{m} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=m+1}^{N} \lambda_i \right) \leq \text{MSE}$$

Alternatively, we can formulate a channel distortion optimization problem that aims to minimize the channel resource usage under a distortion constraint $\text{MSE}$ and a power budget $\tilde{P}$ as follows.

$$(\text{P3}) \min\limits_{\rho} M$$

$$\text{s.t.} \frac{1}{N} \left( \sum_{i=1}^{m} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=m+1}^{N} \lambda_i \right) \leq \text{MSE}$$

Since the channel usage variable $M$ is discrete while the power scaling factor $g_i$ is continuous, these two problems belong to the class of mixed integer non-linear programming problems, which is generally NP-hard.

### B. POWER DISTORTION OPTIMIZATION

To find the minimal power use in Problem (P2), we search all the feasible $m \in [1, M]$ in an exhaustive manner. For each fixed channel resource usage $m$, we solve the following subproblem

$$(\text{P2a}) \min\limits_{\rho} \tilde{P} = \frac{1}{N} \sum_{i=1}^{m} \rho_i$$

$$\text{s.t.} \frac{1}{N} \left( \sum_{i=1}^{m} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=m+1}^{N} \lambda_i \right) \leq \text{MSE}$$

Define Lagrange multiplier $\mu > 0$, then the Lagrange function can be written as

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{m} \rho_i + \mu \left( \frac{1}{N} \left( \sum_{i=1}^{m} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=m+1}^{N} \lambda_i \right) - \text{MSE} \right)$$

Now we set $\frac{\partial \mathcal{L}}{\partial \rho_i} = 0, \text{ for } i = 1, 2, \ldots, m$, and $\frac{\partial \mathcal{L}}{\partial \mu} = 0$, to have

$$\rho_i^* = \begin{cases} \frac{1}{N} \cdot \sqrt{\lambda_i} \cdot \sum_{j=1}^{N} \sqrt{\lambda_j} - 1, & i = 1, 2, \ldots, m \\ \frac{1}{N} \sum_{i=m+1}^{N} \lambda_j \cdot \sum_{j=1}^{N} \sqrt{\lambda_j}, & i = m+1, \ldots, N. \end{cases}$$

The corresponding power distortion optimization algorithm is presented in Algorithm 2. Note that constraint (23) in Problem (P2a) implies that $\text{MSE} - \frac{1}{N} \sum_{i=m+1}^{N} \lambda_i > 0$ for a certain $m$. If this condition is violated, there will be no feasible solution and we will search for the next value of $m$. As shown in Algorithm 2, Line 5, the objective function of Problem (P2a) will have a closed form. By comparing all such $\tilde{P}$s, we choose the one that has the smallest value and adopt the corresponding power allocation $\{\rho_i^*\}$.

### C. CHANNEL DISTORTION OPTIMIZATION

For channel distortion optimization, we can search the value $M$ in a descending order. For each $M$, we solve the following subproblem.

$$(\text{P3a}) \min\limits_{\rho} \text{MSE} = \frac{1}{N} \left( \sum_{i=1}^{m} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=M+1}^{N} \lambda_i \right)$$

$$\text{s.t.} \frac{1}{N} \sum_{i=1}^{M} \rho_i \leq \tilde{P}. \quad (28)$$

If the objective value of Problem (P3a) is less than $\text{MSE}$, then corresponding $M$ is feasible. Then we decrease the value of $M$ by 1 and solve Problem (P3a) again, until we find an $M$ that is infeasible. For each fixed $M$, Problem (P3a) is a simple convex optimization problem. By the Lagrange multiplier method, we obtain the optimal solution as

$$\rho_i^* \approx \frac{N \cdot \sqrt{\lambda_i}}{\sum_{j=1}^{N} \sqrt{\lambda_j}} \cdot \tilde{P}. \quad (29)$$

We present the procedure in Algorithm 3. Note that Algorithms 1–3 are all greedy search algorithms and they share similar procedures. Thus the complexity analysis for Algorithm 1 also applies to Algorithm 2 and Algorithm 3.

### Algorithm 2 Power Distortion Optimization Algorithm

1. Initialize $\lambda_i = \mathbb{E}[|x_i|^2]$, for all $i$, and $\text{MSE}$;
2. for $m = 1, 2, \ldots, M$ do
3. $M \leftarrow m$;
4. Calculate $\{\rho_i, m\}$ according to (26);
5. Calculate $\tilde{P}_m$ according to (22);
6. end for
7. $\tilde{P}^* = \min_{m \in [1,2,...,M]} \tilde{P}_m$;
8. Output $\tilde{P}^*$ and the corresponding $\{\rho_i^*\}$;

### Algorithm 3 Channel Distortion Optimization Algorithm

1. Initialize $\lambda_i = \mathbb{E}[|x_i|^2]$, for all $i$, $\text{MSE}$, $m = N + 1$;
2. repeat
3. $m \leftarrow m - 1$;
4. Calculate $\rho_i, m$ according to (29);
5. Calculate $\text{MSE}_m$ as in (27);
6. until $\text{MSE}_m < \text{MSE}$
7. $M^* \leftarrow m$, $\rho_i^* \leftarrow \rho_i, m$, for all $i$;
8. Output $M^*$ and $\{\rho_i^*\}$;
D. POWER AND CHANNEL USAGE TRADEOFF
For a target video quality, we can either optimize the power consumption under a given channel usage constraint or optimize the channel usage under a constrained power budget. Therefore, we can obtain a power and channel usage tradeoff curve. On each point of the curve, the combination of the corresponding channel and power usage achieves the same target video quality. This tradeoff curve provides us a useful guideline for choosing a suitable power and channel usage pair based on practical resource constraints. In multi-user system, different viewers may have diverse power and channel resource budgets, where a joint optimization can be applied to save resource consumption.

E. VIEW SYNTHESIS
In MVV transmissions, the texture video data contains detailed video content information, while the depth data plays important roles in view synthesis. The quality of both the texture and depth frames determines the virtual view quality. In digital MVV transmissions, bit allocation and power assignment are performed to ensure a good virtual view synthesis performance, which is usually quite complicated [12], [34]. In [34], the view synthesis optimization algorithm is integrated into the encoding process to enable rate-distortion optimization. To achieve a bit rate distribution balance between the texture data and the depth data, a complex combinatorial optimization problem has to be solved. For a two-view scenario, the video/depth rate distribution can be 86%/14%. This way, the depth data is encoded at a low cost. Similar to the bit rate distribution in digital transmissions, in soft video transmissions, we will investigate the power allocation across the texture and depth video data.

Suppose there are N views in the system, as shown in Fig. 2. For simplicity, in this work, we consider the case where equal power control among different reference views is assumed. Power allocation between texture data and depth data is investigated. Specifically, before the 5D-DCT operation, each texture view is scaled by a common factor of $\beta/N$ and each depth view is scaled by a factor of $(1-\beta)/N$. After 5D-DCT, all the video data are linearly transformed and modulated in an analog manner. At the decoder, an inverse process is performed. Followed by a digital renderer to generate the user’s preferred virtual views. Compared with the complicated bit allocation in digital video transmissions, the proposed soft video framework simplifies the process into a power allocation problem. Our target becomes to investigate the impact of the scaling factor $\beta$ on the quality of the synthesized virtual view. We will provide our study and discussion on parameter $\beta$ in Section VI-D1.

VI. SIMULATION STUDY
A. PARAMETER SETTING
1) PERFORMANCE METRIC
In our simulation study, we use both the objective performance metric PSNR and the perceptual metric SSIM [16] for video quality assessment. PSNR is defined (in dB) as

$$\text{PSNR} = 10 \log_{10} \left( \frac{(2^B - 1)^2}{\text{MSE}} \right),$$

where $B$ is the number of bits used to encode pixel luminance (usually 8 bits) and $\text{MSE}$ is the mean squared error between all the pixels between the decoded and the original video. In soft video delivery, since DCT is a linear transformation, the MSE stays the same after the transformation. Hence, we substitute (15) into (30) and we get the...
corresponding distortion. Generally, improvements of PSNR of magnitude larger than 0.5dB are visually noticeable. A PSNR below 20dB is considered not acceptable.

We also use SSIM to measure the similarity of the original and reconstructed images to test the performance of the proposed method [16]. For two $N \times N$ images $x$ and $y$, the SSIM index is computed as [16]

$$\text{SSIM}(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

(31)

where $\mu_x$ and $\mu_y$ are the means of $x$ and $y$, respectively; $\sigma_x^2$ and $\sigma_y^2$ are the variances of $x$ and $y$, respectively; $\sigma_{xy}$ is the covariance of $x$ and $y$; $c_1 = (k_1 \cdot L)^2$ and $c_2 = (k_2 \cdot L)^2$; $L$ is the dynamic range of the pixel values; $k_1 = 0.01$, and $k_2 = 0.03$. An SSIM value closer to 1 suggests higher perceptual similarity between the original and the decoded image.

2) TEST VIDEO

We use two standard reference MVD videos, balloons and kendo, at 30 fps. We choose view points 1, 3, 5. Three cameras are used with a distance 10cm away. Their resolution is $1024 \times 768$ pixels for texture and depth frames at a frame rate of 20 fps. The video sequences are selected from the video database [35]. In addition, some standard single view monotone CIF video sequences from video database [36] are used in our simulations.

3) PARAMETER SETTING

For soft video delivery, we set the GOP size to 4. In existing chunk-based schemes of soft video delivery, we divide each frame into $8 \times 8 = 64$ chunks. For MVD video, we read both the texture and depth data from the three cameras. Thus one GOP consists of $3 \times 2 \times 4 \times 64 = 1536$ chunks. The camera configuration is summarized in Table 1. For Algorithm 1, we choose $\alpha = 0.98$. We use the 3D HEVC test model (HTM) v15.0 software [37] renderer to synthesize a virtual viewpoint from the received texture and depth frames. We assume the AWGN channel in the simulations.

| Table 1. Camera configuration for video sequences kendo.yuv and balloons.yuv. |
|---|---|---|---|---|---|
| View | Focal Length | Position | CShit | ZNear | ZFar |
|---|---|---|---|---|---|
| 1 | 2241.25607 | 5.0 | 701.5 | 448.251214 | 11206.280350 |
| 3 | 2241.25607 | 15.0 | 701.5 | 448.251214 | 11206.280350 |
| 5 | 2241.25607 | 25.0 | 701.5 | 448.251214 | 11206.280350 |

B. DR OPTIMIZATION PERFORMANCE

We first evaluate the DR optimization performance. We investigate the maximum PSNR that can be achieved under a given resource (i.e., channel and power) constraint. As mentioned before, there will be 1536 chunks in each GOP. Suppose in each channel slot (e.g., time or frequency), only one chunk can be transmitted and let the maximum number of available channel slots $M$ for each GOP be 1536. We fix the noise variance to be 1 and vary the total transmit power budget $\tilde{P}$ for each GOP. The DR optimization results for video sequences kendo.yuv and balloons.yuv are shown in Fig. 3 and Fig. 4, respectively.

It can be seen that for a given channel resource use $N$, the PSNR generally increases with the power budget $\tilde{P}$. This is also confirmed by our discussion of (15). However, for a given power budget $\tilde{P}$, the PSNR does not necessarily increase with the channel usage, especially when the power budget is low. For example, when the power budget $\tilde{P}$ is 5dB, the maximum PSNR point for the video sequences kendo.yuv and balloons.yuv are attained when $N = 250$ and $N = 196$, respectively. A maximum channel usage of 1536 does not always lead to the highest PSNR. The reason is in soft video delivery, different chunks are not of equal importance although each of them consumes one channel use.

1Equivalently, we are changing the value of the SINR $\tilde{P}$ for each GOP. Since we assume a fixed noise level, in the following, we do not distinguish $P$ and $\tilde{P}$ anymore.
Under a fixed power budget, allocating more power to HP chunks and allocating less power to LP chunks (or even discarding them) helps improve the PSNR performance. Finally, we note that these DR curves generally have a flat tail, which means when the PSNR is above a certain level, improving the channel use is not efficient as improving the power budget. For example, for video sequence *kendo*.yuv, when the power budget $\tilde{P} = 10$ dB and the channel use $N = 668$, the PSNR is 40.74 dB. Now improving channel usages doesn’t improve PSNR value any more. However, a power increase from 10 dB to 15 dB brings a PSNR improvement of 4.52 dB and a power increase from 10 dB to 15 dB brings a PSNR improvement of 8.28 dB.

To further clarify the channel usage saving, we plot the corresponding performance for video sequence *kendo*.yuv in Fig. 5. As can be seen, conventional Softcast uses all the channel and power resources to achieve a good PSNR. However, with the proposed algorithm, we can achieve a slightly higher PSNR value with a reduced channel usage. For example, in Fig. 5 (d) when the power budget $\tilde{P} = 5$ dB, in conventional SoftCast, the channel usage is 1536 and the achieved PSNR is 35.903 dB. However, with the proposed algorithm, we achieve the maximum PSNR of 36.236 dB with only 250/1536 = 16.2% of the original channel use. Moreover, due to the flat tail of the distortion resource curve, by slightly lowering the PSNR requirement (e.g., $\alpha = 98\%$ of the maximum PSNR), the channel usage can be further greatly reduced from 250 to 70, as shown in Fig. 5 (d).

Comparing these figures, we also note that the proposed algorithm saves more channel usage when $\tilde{P}$ is low. Hence the proposed method is more suitable for the case when channel condition is not good or the total power budget is limited.

Table 2 lists the channel usage comparison for different video sequences. The video file names in bold are MVD videos while others are standard monotone CIF video sequences used in SoftCast test, with a resolution of $352 \times 288$ at a frame rate of 20 fps. The proposed method significantly reduces the channel usage while still maintaining a satisfactory performance.

Note that the proposed method not only applies to MVV videos but also conventional standard videos. Considering that the channel usage grows linearly with the number of depth and camera data for MVV video transmissions, our proposed method is naturally more suitable to deal with the heavy data burden challenge caused by MVV videos.

### C. RD OPTIMIZATION PERFORMANCE

1) POWER DISTORTION OPTIMIZATION PERFORMANCE

Suppose the user requires a PSNR of 35 dB, under a constant channel usage of 1536 chunks per GOP. We plot the power allocation curve for consecutive MVV video sequences *kendo*.yuv in Fig. 6 and that for *balloons*.yuv in Fig. 7. Specifically, we plot the resource usage curve for the first 64 GOPs in the figures. It can be seen that the power allocation of consecutive GOPs have very small variations and the channel...
FIGURE 6. Power allocation for MVV video sequence *kendo.yuv*.

FIGURE 7. Power allocation for MVV video sequence *balloons.yuv*.

usage is maintained at a relatively low level. This intricate resource control helps the MVV video quality remain at the prescribed PSNR value of 35dB. Hence the viewer can enjoy a favorable viewing experience and the saved wireless resources can be utilized by other users.

2) CHANNEL DISTORTION OPTIMIZATION PERFORMANCE

Similarly, still suppose the user requires a PSNR of 35dB under a constant transmit power budget $\tilde{P} = 10$dB. We plot the channel allocation curve for different videos in Fig. 8 and Fig. 9. We note that the channel usage fluctuates at a relatively low and stable level compared with the large chunk number (1536) in one GOP. Hence, in practice, we can actually progressively allocate slightly more channel resources (say 100) without running algorithm 3 many times. The computational cost, therefore, can be further reduced.

3) CHANNEL POWER TRADEOFF

For a target video quality, there exists a tradeoff between power usage and channel usage. We plot the tradeoff curve
in Fig. 10 and Fig. 11. It can be observed that this kind of tradeoff varies with different target PSNR values. Under a lower target PSNR, both the power and channel usage required are quite low. Moreover, these curves tend to have a very sharp turning point when $M$ is relatively low. Hence, maintaining the power and channel usage pair near the turning point would be an efficient strategy. This observation has also been confirmed by the previous simulations.
FIGURE 14. Synthesize quality at virtual viewpoint 2 for different algorithms.

D. VIEW SYNTHESIS
1) IMPACT OF PARAMETER $\beta$
As mentioned before, the power scaling factor $\beta$ determines the power allocation ratio between the texture data and the depth data, which has a joint impact on the quality of synthesized virtual view. For each view, the power ratio between the texture data and the depth is $\beta/(1-\beta)$. $\beta = 0.5$ means an equal power control between texture data and depth data and $\beta > 0.5$ means more power is allocated to the texture data. We plot the impact of parameter $\beta$ on the quality of the synthesized virtual view in Fig. 12.

The test video sequence is kendo.yuv. We investigate the PSNR performance for the first frame in the first GOP of view 1, view 3, and the corresponding virtual views. The virtual view is synthesized based on the texture and depth data from view 1 and view 3. Recall that 5D-DCT is a linear transformation and it is performed on all the texture and depth data from each relevant views. Hence the MSE distortion of each component (texture data and depth data from each view) can be actually approximated by the average MSE of its corresponding GOP. This explains why the texture and depth data quality curve from view 1 crossed the curve from view 3 when $\beta = 0.5$. If we increase the value of $\beta$, we intend to allocate more power to texture data, hence the PSNR of the texture data increases and the PSNR of the depth data decreases. Remember that in this paper we assume power is equally allocated between different view points, hence view 1 and view 3 is of equal importance in viewing synthesis. That is why the PSNR quality curve of view 1 fits well with that of view 3. In real systems, there...
may be 50 or even more cameras, while a viewer may only be interested in one specific view. An equal power allocation for each view is obviously not optimal. Instead of performing an equal power control among different cameras, it may be a better strategy to allocate more power to the adjacent views of the user’s chosen one. This problem will be addressed in our future work. Finally, we note that the synthesized virtual view 2 reaches its best quality around $\beta = 0.5$. Moreover, we note that when $\beta$ is between 0.4 to 0.6, the virtual view quality remains at a relatively high level. As $\beta$ gets closer to either end of the interval [0, 1], the synthesized virtual view quality drops dramatically. This is because when $\beta$ is small (or large), there is a huge distortion in the texture (or the depth) data. This kind of imbalance between the texture data and depth data degrades the quality of the synthesized view. From this figure, we can see a good choice of $\beta$ is between 0.5 and 0.6 where the texture view qualities at view 1 and view 3 are slightly increased and the view quality at virtual view 2 almost remains at a constant high level.

2) IMPACT OF VIRTUAL VIEW POSITIONS
In this experiment, view 3 and view 5 are reference views and the views between them are synthesized virtual views. Fig. 13 shows the impact of the virtual view position on the video perceptual quality. We consider the first frame of video sequence kendo.yuv. Video power is equally distributed between the texture and depth data ($\beta = 0.5$). We change the SNR value from 0dB to 20dB. Both the PSNR performance and the SSIM performance are presented. As can be seen, with the increase of SNR, both PSNR and SSIM increase. When SNR is 0dB and 10dB, the view performance at the mid-point (virtual view 4) tends to have a higher perceptual quality than the view quality at other positions. This is due to the equal power allocation between adjacent views. Hence, if a user is more interested in the virtual view that is closer to view 3 (e.g., virtual view at position 3.2), we may want to allocate more power to the view at position 3. When SNR is 20dB, there is generally no big difference between the video quality data at different positions.

3) PERFORMANCE THE PROPOSED ALGORITHM
In this subsection, we test the performance of our proposed DR algorithm on the quality of synthesized virtual views. Each simulation is performed 10 times and we present the averaged results.

Fig. 14 presents the synthesis quality at view point 2 for different algorithms, along with the synthesized frames. Under a channel constraint 1536, the conventional SoftCast scheme utilizes all the channel resources and allocate power on each chunk. In contrast, our proposed DR algorithm discard LP chunks and only retain the HP chunks. The channel usage is greatly saved and the PSNR performance is improved. For example, when power usage is 10dB, for kendo.yuv, both the PSNR and SSIM slightly increases and the channel usage is only $163/1536 = 10.6\%$ of the conventional SoftCast scheme. The proposed algorithm achieves an even better performance when SNR is low. Considering the huge data traffic in MVV, the saved wireless resources would be considerable. Finally, we have to mention that although the proposed algorithm is designed based on the PSNR metric, it still works well in terms of the SSIM metric. For video sequence balloons.yuv, the proposed algorithm improves the SSIM from 0.84184 to 0.91023 when SNR is 0dB. We also perform similar experiments on the RD algorithms, where similar observations are made. For space limitation, we omit the RD results here.

VII. CONCLUSION
In this paper, we integrated applying soft video delivery for MVV transmissions. Compared with the conventional digital based solutions, the proposed scheme improves video quality gracefully and the cliff effect can be avoided. Furthermore, complex bit allocation and rate control in digital systems can be replaced by a simple power allocation scheme. To handle the heavy data traffic caused by MVV, we proposed a resource control algorithm. The proposed DR optimization algorithm achieves the best viewing quality under a resource constraint. The proposed RD optimization algorithm minimizes the resource usage for a target video quality requirement. Hence the viewer can enjoy a stable viewing quality, which is favorable for MVV video streaming. With the proposed scheme, we also investigated the impact of the power control across the texture and depth data and view positions on synthesized virtual view quality. Simulation results demonstrated that the proposed algorithm worked well not only for referenced views but also for synthesized virtual views. Despite all these merits brought by soft video delivery, we find that soft video transmission may not be resilient to packet loss and is prone to channel variations. An extension toward hybrid digital and analog video coding would be a promising direction to fully exploit the benefit of both the digital and analog video coding.

REFERENCES
[1] S. Mao, Video Over Cognitive Radio Networks: When Quality of Service Meets Spectrum, 1st ed. New York, NY, USA: Springer, Jan. 2014.
[2] Y. Xu and S. Mao, “A survey of mobile cloud computing for rich media applications,” IEEE Wireless Commun., vol. 20, no. 3, pp. 46–53, Jun. 2013.
[3] M. Amjad, M. H. Rehmani, and S. Mao, “Wireless multimedia cognitive radio networks: A comprehensive survey,” IEEE Commun. Surveys Tuts., vol. 20, no. 2, pp. 1056–1103, 2nd Quart., 2018.
[4] Y. Chen, M. M. Hannuksela, T. Suzuki, and S. Hattori, “Overview of the MVC + D 3D video coding standard,” J. Vis. Commun. Image Represent., vol. 25, no. 4, pp. 679–688, May 2013.
[5] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, “Overview of the multiview and 3D extensions of high efficiency video coding,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 1, pp. 35–49, Jan. 2016.
[6] C. Fehn, “Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV,” Proc. SPIE, vol. 5291, May 2004, pp. 93–105.
[7] T. Zhang and S. Mao, “An overview of emerging video coding standards,” GetMobile, Mobile Comput. Commun., vol. 22, no. 4, pp. 13–20, May 2018.
[8] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, “Multi-view video plus depth representation and coding,” in Proc. IEEE Int. Conf. Image Process. (ICIP), San Antonio, TX, USA, Sep. 2007, pp. 201–204.
[9] Z. Liu, G. Cheung, and Y. Ji, “Optimizing distributed source coding for interactive multiview video streaming over lossy networks,” IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 10, pp. 1781–1794, Oct. 2013.

[10] L. Toni, G. Cheung, and P. Frossard, “In-network view synthesis for interactive multiview video systems,” IEEE Trans. Multimedia, vol. 18, no. 5, pp. 852–864, May 2016.

[11] Y. Chen, Y.-K. Wang, K. Ugr, M. H. Hannakshela, J. Lainema, and M. Gabbouj, “The emerging MVC standard for mobile services,” EURASIP J. Adv. Signal Process., vol. 2009, p. 8, Jan. 2009.

[12] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, L. Lakshman, P. Merkler, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, “3D high-efficiency video coding for multi-view video and depth data,” IEEE Trans. Image Process., vol. 22, no. 9, pp. 3366–3378, Sep. 2013.

[13] C. Berrou, A. Glavieux, and P. Thitmajishima, “Near Shannon limit error correcting coding: Turbo-codes,” in Proc. IEEE Int. Conf. Commun., vol. 2, May 1993, pp. 1064–1070.

[14] T. Kratochvíl and R. Štukavec, “DVB-T digital terrestrial television transmission over fading channels,” Radioengineering, vol. 17, no. 3, pp. 96–102, Sep. 2008.

[15] S. Jakubczak and D. Katabi, “SoftCast: One-size-fits-all wireless video,” ACM SIGCOMM Comput. Commun. Rev., vol. 41, no. 4, pp. 449–450, Oct. 2011.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[17] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[18] J. Chakareski, “Adaptive multiview video streaming: Challenges and opportunities,” IEEE Commun. Mag., vol. 51, no. 5, pp. 94–100, May 2013.

[19] A. D. Abreu, P. Frossard, and F. Pereira, “Optimizing multiview video plus depth prediction structures for interactive multiview video streaming,” IEEE J. Sel. Topics Signal Process., vol. 9, no. 3, pp. 487–500, Apr. 2015.

[20] E. Ekmekcioglu, C. G. Gurler, A. Kondoz, and A. M. Tekalp, “Adaptive multiview video delivery using hybrid networking,” IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 6, pp. 1313–1325, Jun. 2017.

[21] R. Suenaga, K. Suzuki, T. Tezuka, M. P. Tehrani, K. Takahashi, and T. Fujii, “A practical implementation of free viewpoint video system for soccer games,” Proc. SPIE, vol. 9393, May 2015, Art. no. 93930G.

[22] Z. Chen, X. Zhang, Y. Xu, J. Xiong, Y. Zhu, and X. Wang, “MuVi: Multiview video aware transmission over MIMO wireless systems,” IEEE Trans. Multimedia, vol. 19, no. 12, pp. 2788–2803, Dec. 2017.

[23] Z. Liu, G. Cheung, J. Chakareski, and Y. Ji, “Multiple description coding and recovery of free viewpoint video for wireless multipath streaming,” IEEE J. Sel. Topics Signal Process., vol. 9, no. 1, pp. 151–164, Feb. 2015.

[24] S. Jakubczak and D. Katabi, “A cross-layer design for scalable mobile video,” in Proc. ACM MobiCom, Las Vegas, NV, USA, Sep. 2011, pp. 289–300.

[25] S. Aditya and S. Katti, “FlexCast: Graceful wireless video streaming,” in Proc. 17th Annua. Int. Conf. Mobile Comput. Netw., Las Vegas, NV, USA, 2011, pp. 277–288.

[26] Y. Li, Z. Li, Y. Liu, and Y. Wang, “SCAST: Wireless video multicast scheme based on segmentation and SoftCast,” in Proc. IEEE WCNC, San Francisco, CA, USA, Mar. 2017, pp. 1–6.

[27] B. Tan, J. Wu, Y. Li, H. Cui, W. Yu, and C. W. Chen, “Analog coded SoftCast: A network slice design for multimedia multicast,” IEEE Trans. Multimedia, vol. 19, no. 10, pp. 2293–2306, Oct. 2017.

[28] H. Cui, C. Luo, C. W. Chen, and F. Wu, “Robust uncoded video transmission over wireless fast fading channel,” in Proc. IEEE INFOCOM, Toronto, ON, Canada, Apr./May 2014, pp. 73–81.

[29] X. L. Liu, W. Hu, C. Luo, Q. Pu, F. Wu, and Y. Zhang, “ParCast+: Parallel videounicast in MIMO-OFDM WLANs,” IEEE Trans. Multimedia, vol. 16, no. 7, pp. 2038–2051, Nov. 2014.

[30] H. Cui, C. Luo, C. W. Chen, and F. Wu, “Scalable video multicast for MU-MIMO systems with antenna heterogeneity,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 5, pp. 992–1003, May 2016.

[31] D. Liu, J. Wu, H. Cui, D. Zhang, C. Luo, and F. Wu, “Cost-distortion optimization and resource control in pseudo-analog visual communications,” IEEE Trans. Multimedia, vol. 20, no. 11, pp. 3097–3110, Nov. 2018.

[32] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, “FreeCast: Graceful free-viewpoint video delivery,” IEEE Trans. Multimedia, vol. 21, no. 4, pp. 1000–1010, Apr. 2018.

[33] T. Zhang and S. Mao, “Metadata reduction for soft video delivery,” IEEE Netw. Lett., vol. 1, no. 2, pp. 84–88, Jun. 2019.

[34] W.-S. Kim, A. Ortega, P. Lai, and D. Tian, “Depth map coding optimization using rendered view distortion for 3D video coding,” IEEE Trans. Image Process., vol. 24, no. 11, pp. 3534–3545, Nov. 2015.

[35] Mobile 3D TV Content Delivery Optimization Over DVB-H System. Accessed: Oct. 12, 2019. [Online]. Available: http://sp.cs.ut.tu/dvb3dtv/stereo-video/

[36] Xiph.org Video Test Media [Def’s Collection]. Accessed: Oct. 12, 2019. [Online]. Available: https://media.xiph.org/video/derfl/erfl/

[37] Fraunhofer Heinrich Hertz Institute. 3D High Efficiency Video Coding (3D-HEVC)—JCT-VC. Accessed: Oct. 12, 2019. [Online]. Available: https://hevc.hhi.fraunhofer.de/3dhevc

TICAO ZHAO received the B.E. and M.S. degrees from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Auburn University. His research interests include video coding and communications, machine learning, and optimization and design of wireless multimedia networks.

SHIWEEN MAO (S’99–M’04–SM’09–F’19) received the Ph.D. degree in electrical and computer engineering from Polytechnic University (now New York University Tandon School of Engineering), Brooklyn, NY, USA. He joined Auburn University, Auburn, AL, USA, in 2006, as an Assistant Professor of electrical and computer engineering, where he held the McWane Associate Professorship, from 2012 to 2015. He is currently the Samuel Ginn Distinguished Professor and the Director of the Wireless Engineering Research and Education Center, Auburn University. His research interests include wireless networks, multimedia communications, and smart grid. He received the IEEE ComSoc TC-CSR Distinguished Technical Achievement Award, in 2019, the IEEE ComSoc MMTT Distinguished Service Award, in 2019, the Auburn University Creative Research and Scholarship Award, in 2018, the 2017 IEEE ComSoc ITC Outstanding Service Award, the 2015 IEEE ComSoc TC-CSR Distinguished Service Award, the 2013 IEEE ComSoc MMTC Outstanding Leadership Award, and the NSF CAREER Award, in 2010. He was a co-recipient of the IEEE ComSoc MMTT Best Journal Paper Award, in 2019, the IEEE ComSoc MMTT Best Conference Paper Award, in 2018, the Best Demo Award from IEEE SECON 2017, the Best Paper Awards from IEEE GLOBECOM 2016 and 2015, the IEEE WCNC 2015, and IEEE ICC 2013, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He was the Chair of the IEEE ComSoc Multimedia Communications Technical Committee (MMC), from 2016 to 2018. He is also an Area Editor of the IEEE Open Journal of the Communications Society and the IEEE Internet of Things Journal and an Associate Editor of the IEEE Transactions on Network Science and Engineering, the IEEE Transactions on Mobile Computing, the IEEE Transactions on Multimedia, the IEEE Multimedia, the IEEE Networking Letters, and GetMobile (ACM), among others. He is also a Distinguished Speaker from 2018 to 2021 and was a Distinguished Lecturer, from 2014 to 2018, of the IEEE Vehicular Technology Society.

* * *