Performance Evaluation and Parameter Optimization of SoftCast Wireless Video Broadcast

Dongxue Yang
Key Lab of Universal Wireless Communications, Ministry of Education
Beijing University of Posts and Telecommunications, Beijing, China
dongxueyang01@bupt.edu.cn

Yunqiang Bi
Key Lab of Universal Wireless Communications, Ministry of Education
Beijing University of Posts and Telecommunications, Beijing, China
biyunqiang@bupt.edu.cn

Zhongwei Si
Key Lab of Universal Wireless Communications, Ministry of Education
Beijing University of Posts and Telecommunications, Beijing, China
sizhongwei@bupt.edu.cn

ZhiQiang He
Key Lab of Universal Wireless Communications, Ministry of Education
Beijing University of Posts and Telecommunications, Beijing, China
hezq@bupt.edu.cn

Kai Niu
Key Lab of Universal Wireless Communications, Ministry of Education
Beijing University of Posts and Telecommunications, Beijing, China
niukai@bupt.edu.cn

ABSTRACT

Wireless video broadcast plays an important role in multimedia communication with the emergence of mobile video applications. However, conventional video broadcast designs suffer from a cliff effect due to separated source and channel encoding. The newly proposed SoftCast scheme employs a cross-layer design, whose reconstructed video quality is proportional to the channel condition. In this paper, we provide the performance evaluation and the parameter optimization of the SoftCast system. Optimization principles on parameter selection are suggested to obtain a better video quality, occupy less bandwidth and/or utilize lower complexity. In addition, we compare SoftCast with H.264 in the LTE EPA scenario. The simulation results show that SoftCast provides a better performance in the scalability to channel conditions and the robustness to packet losses.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication

General Terms

Design, Measurement, Performance

.

Keywords

wireless video broadcast, SoftCast, H.264, scalable video communication, linear transmission

1. INTRODUCTION

Wireless video broadcast is becoming increasingly important driven by the shift of users' video viewing habits from traditional TV to mobile devices. The multiple users in the wireless broadcast system have different channel characteristics, e.g., the channel quality, the bandwidth allocation, the ratio of packet loss, etc. The main challenge in wireless video broadcast is then how to satisfy all the users when they are interested in the same video.

Current video broadcast schemes, such as H.264 [1], are based on separate source and channel coding. They exhibit a cliff effect [2, 3] in the peak signal to noise ratio (PSNR) which reflects the reconstructed video quality. When the channel quality is below a certain point, the PSNR drops dramatically since the pre-determined channel coding cannot protect the video from the errors. When the channel quality is above this critical point, the PSNR approximately keeps constant. This indicates that the distortion due to the lossy source coding cannot be recovered no matter how good the channel is. In a word, the users cannot obtain a video quality scalable to their channel conditions.

To combat the cliff effect, SoftCast was proposed in [4, 5, 6] as a cross-layer wireless video broadcast design. It compresses the video and protects it from errors and losses in a unified manner. One crucial property of SoftCast is that each transformation module is linear and a stream of real numbers instead of bits is transmitted. As a result of the linear encoding, the video quality of each receiver is proportional to its channel quality. Moreover, SoftCast uses two-dimension discrete cosine transform (2D-DCT) to realize intra frame compression and uses three-dimension (3D-

MOBIMEDIA 2015, May 25-27, Chengdu, People's Republic of China

DOI 10.4108/icst.mobimedia.2015.259038

DCT) to conduct inter frame compression. The DCT removes the spatial redundancy and the time redundancy to improve compression efficiency, and no motion estimation and motion compensation is needed. Based on these advantages, SoftCast attracts more and more attentions recently [7, 8, 9, 10].

In this paper we provide the performance evaluation and the parameter optimization of the SoftCast wireless video broadcast system. In particular, we investigate the relationship between the video quality and three control parameters, chunk size, compression ratio and the group of pictures (GoP). We then analyze how to choose appropriate parameters to obtain a better video quality, occupy less bandwidth and/or utilize lower complexity. In the end, to provide a comprehensive performance evaluation, SoftCast is compared with H.264 in the LTE EPA scenario from different perspectives. The simulation results show that SoftCast provides better performance in the scalability to channel conditions, the robustness to packet losses, and so on.

The rest of the paper is organized as follows. Section 2 reviews the SoftCast scheme briefly. Section 3 analyzes the control parameters of the SoftCast and suggests the optimization principles. In Section 4, we evaluate the SoftCast's performance comparing with H.264 in the LTE EPA scenario. Section 5 concludes the paper.

2. REVIEW OF SOFTCAST

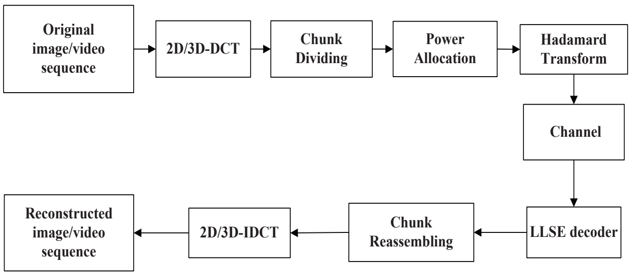
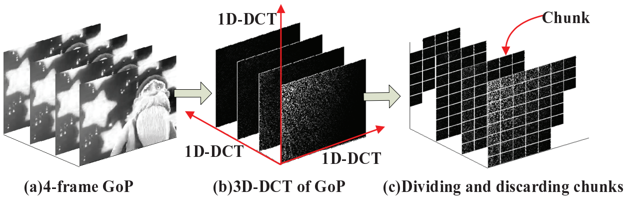


Figure 1: Framework of SoftCast.



(a)4-frame GoP (b)3D-DCT of GoP (c)Dividing and discarding chunks

Figure 2: 3D-DCT of a 4-frame GoP.

The framework of the SoftCast scheme is depicted in Fig.1. At the sender, SoftCast first performs 2D/3D-DCT on the original image/video. Most DCT components are zeroes, and the non-zero DCT components are spatially clustered as illustrated in Fig.2. In the chunk-dividing module, nearby DCT components are grouped into chunks. The energy is unequally distributed among the chunks, and the chunks with small energy can be discarded. The location information of the discarded chunks is broadcasted to the receivers. After chunk dividing, the DCT components of different chunks are scaled through a power allocation which computes the optimal power under a fixed power budget. The energy of the DCT components is redistricted in the Hadamard transformation which ensures that all the packets are equally important and the loss of each packet causes proportionate video distortion. The output of the Hadamard transformation is directly broadcasted into the channel. The above encoding process can be represented as

$$Y = HGX = CX, \tag{1}$$

where X is the matrix composed of the DCT components in a GoP and Y is the output of the encoder. The matrix G is a diagonal matrix with the scaling factors from power allocation as the entries, and H is the Hadamard transformation matrix. We write $C = HG$ as the encoding matrix for the ease of notation. Obviously, the output Y are real values instead of bits. SoftCast applies DCT instead of conventional quantization and entropy coding for the source coding, and abandons the channel coding for combating noises. In fact, the error protection of SoftCast is realized by the power allocation, and the resistence to packet loss is carried out by the Hadamard transformation.

The received signal $\hat{Y}$ is

$$\hat{Y} = CX + N, \tag{2}$$

where N denotes the white Gaussian noise. The linear least squared error (LLSE) decoder computes the estimation of the original DCT components as:

$$\hat{X}_{LLSE} = \Lambda_x C^{\mathrm{T}} (C\Lambda_x C^{\mathrm{T}} + \Sigma)^{-1} \hat{Y}, \tag{3}$$

where the diagonal matrix Λ_x consists of the variances of individual chunks, and the diagonal matrix Σ is composed of the noise power experienced by the corresponding packet. Note that Λ_x is transmitted as the metadata from the encoder and known by the receiver. The decoded DCT components are reassembled into chunks, and the discarded chunks are regarded as zero-valued. In the end, the DCT components are transformed back to the spatial domain by the inverse discrete cosine transform (IDCT) to reconstruct the image or video.

3. OPTIMIZATION THROUGH THE CONTROL PARAMETERS

In this section, we optimize the performance of SoftCast through tuning parameters, aiming at reducing complexity and saving resource occupation but meanwhile providing better video quality. Three control parameters, chunk size, compression ratio (CR) and GoP, are considered in the optimization.

3.1 Chunk Size

SoftCast groups nearby spatial DCT components into chunks. When a chunk is discarded, the decoder estimates all DCT components in that chunk as zeros, and the distortion from discarding a chunk is merely the sum of the squares of the DCT components of that chunk. Hence, for a fixed compression ratio, improving the resolution for dividing and discarding the energy can minimize the distortion due to the discarded chunks. On the other hand, a smaller chunk size brings in a higher complexity. An optimal tradeoff between the performance and the complexity is highly useful for the system optimization. Since different images/videos have various characteristics in the DCT, it is hard to give a generalized theoretical analysis on the optimal tradeoff.

Therefore, in the following we try to find out this tradeoff through simulations.

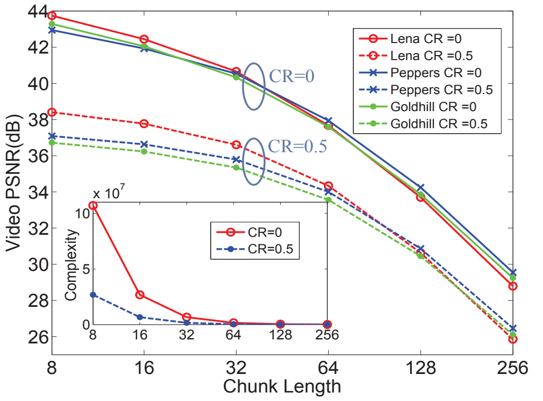


Figure 3: Performance evaluation for different chunk sizes.

To purely study the effects of chunk size selection, we use images (i.e., GoP=1) instead of videos in the following simulations. Three different types of pictures, *Lena*, *Peppers*, and *Goldhill* (512×512 pixel, gray) are chosen as the test images for the performance evaluation. The compression ratio is set to be 0 (no compression) and 0.5, respectively. Six choices of chunk size (8×8, 16×16,32×32, 64×64, 128×128, 256×256) are employed in the simulations.

The performance evaluation for different chunk sizes are shown in Fig.3. As expected, the performance degrades as the chunk size increases regardless of the compression ratio. The performance is improved significantly if the chunk size is changed from large to medium. However, if the chunk size is changed from medium to small, the improvement is not much. For example, for the case of CR=0.5, if the chunk size is changed from 256×256 to 32×32, the increase in PSNR is approximately 10dB for all the three images. When we further change the chunk size to 8×8, the increase in PSNR is less than 1dB. Similar results can be observed for the CR=0 case.

On the other hand, smaller chunk size brings in higher complexity due to the increase of dimension of the encoder matrix. A comparison of the computation complexity is included in Fig.3 as well. Although the increase in PSNR is trivial when the chunk size is changed form 32×32 to 8×8, the complexity increases dramatically for this change. By contrast, if the chunk size decreases from 256×256 to 32×32, the complexity remains almost the same but a significant gain in performance is observed.

From the above observations, we conclude that 32×32 is a cliff point of the chunk size which provides a good tradeoff between the performance and the complexity. To provide a subjective performance evaluation for different chunk sizes, *Lena*'s reconstructed images are shown in Fig.4. When the chunk size is 256×256, the recovered image has a certain degree of fuzzy. For the other two selections of chunk size, the subjective experience of image quality is similar, but the complexity of the size 32×32 is much smaller than that of the size 8×8. The subjective evaluation verifies the aforementioned conclusion on the optimal tradeoff between the performance and the complexity.

3.2 Compression Ratio

SoftCast uses the same method to compress information across space and time, which employs a DCT to transform the data to its frequency representation. As mentioned earlier, the compaction of their frequency representation contributes to the realization of compression. When the chunk size is determined, the compression is affected by the ratio of chunk discarding. We define the compression ratio as the proportion of the discarded chunks to the total number of chunks. A higher compression ratio usually means a larger amount of energy abandon.

The compression should be done as much as possible to save bandwidth occupation and reduce complexity without degrading the reconstructed image quality. In the following we study how the compression ratio affects the performance in terms of PSNR. We use the same test images: *Lena*, *Peppers*, and *Goldhill*. Two chunk sizes are selected: 8×8 and 64×64. The performance evaluation for different compression ratios is depicted in Fig.5. We can see that the PSNR is linear to the compression ratio regardless of the chunk size. Smaller compression ratio gains better performance but requires more bandwidth occupation. In practice, the compression ratio can be tuned according to the bandwidth restriction. Once the compression ratio is determined, the smaller the chunk size is, the better the performance is, as shown in Fig.5.

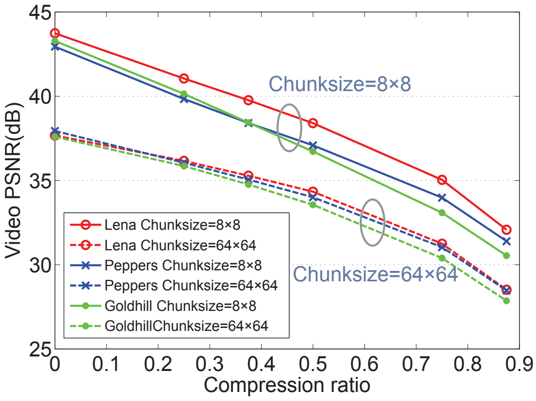


Figure 5: Performance evaluation for different compression ratios.

In the following, we investigate the performance as a function of the channel SNR in Fig.6. If there is no compression, the received PSNR is linear to the channel SNR. This feature benefits from the linear encoding of SoftCast. When the compression ratio is as high as 0.5 or even 0.75, i.e., only 1/2 or 1/4 chunks are reserved, the PSNR becomes slightly nonlinear in the high SNR region. For these cases, the energy abandoned in the chunk discarding is non-negligible, and it cannot be compensated by the channel.

3.3 Group of Pictures

The number of frames that the encoder compresses at one time is called GoP. In the following simulations, we take the first 128 frames of *foreman*, *paris* and *highway* as test videos. The chunk size and the compression ratio of the test videos are set to be 36×44 and 0.5 according to the above analysis and optimization. Fig.7 shows that the slope of the curve reduces as the GoP increases, which is due to the change of inter-frame correlation. When a sudden change occurs compared to the previous frames, PSNR may have a vibration.

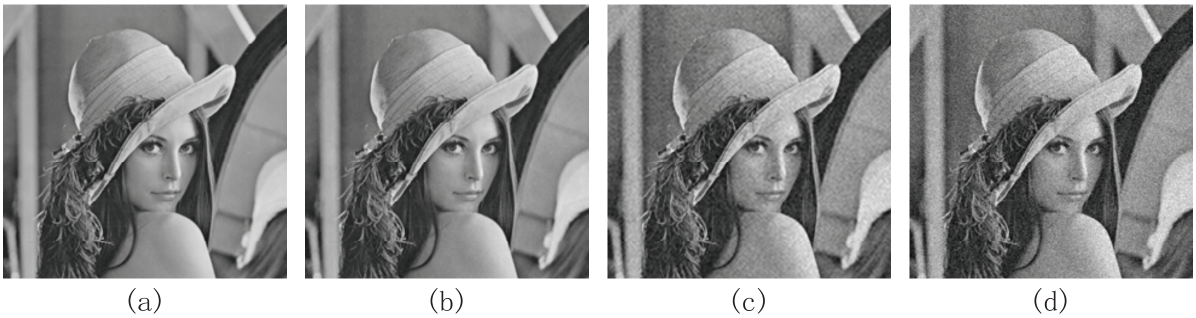

(a) (b) (c) (d)

Figure 4: Lena's reconstructed images. (a) the original image; (b) Chunksize=8×8, PSNR=38.4dB; (c) Chunksize=32×32, PSNR=36.6dB; (d) Chunksize=256×256, PSNR=25.8dB.

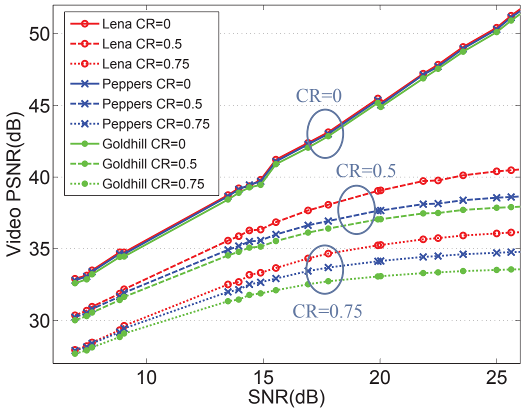


Figure 6: Performance evaluation for varying channel SNRs.

Within a certain range, the increase of GoP helps to further remove the redundancies between the frames and better video quality can be achieved. However, oversize GoP does not necessarily lead to better PSNR but definitely brings in higher complexity. The optimal selection of GoP is hard to decide in general because it closely depends on the characteristics of the video. We recommend using medium GoP, e.g., 10-20, in implementations. Performance optimization through tuning the GoP is unnecessary.

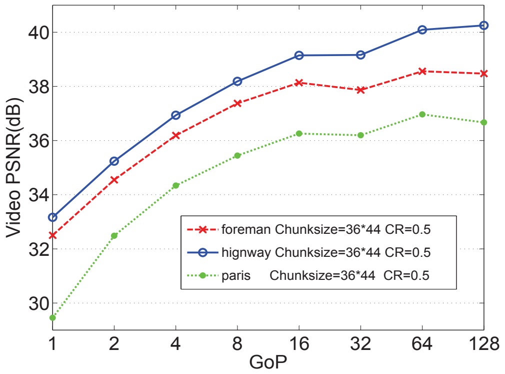


Figure 7: Performance evaluation for different GoPs.

3.4 Optimization Principles

Chunk size, compression ratio and GoP are the most important parameters of the SoftCast scheme. By choosing a proper chunk size, a tradeoff between the performance and the complexity can be achieved. Compression ratio can be used to balance the performance and the bandwidth occupation. A medium size of GoP is recommended, and it is not considered in the system optimization. The choice of the chunk size and the compression ratio is determined by the most crucial restriction in the system. If the complexity is the dominating factor, a relatively large chunk size should be selected, and a low compression ratio should be used to guarantee the performance. If the bandwidth restriction is dominating, the compression ratio should be tuned to satisfy the bandwidth requirement, and the performance can be improved by choosing small chunk size.

4. PERFORMANCE COMPARISON WITH H.264

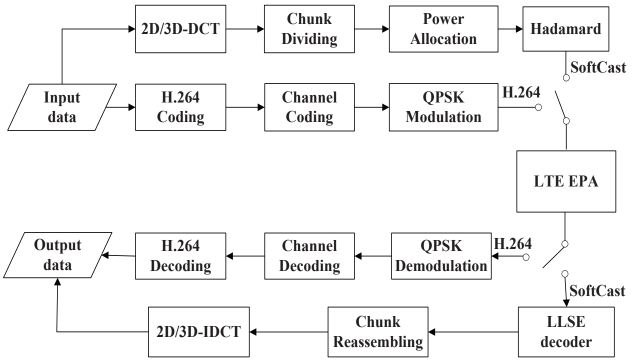


Figure 8: Famework of comparison between SoftCast and H.264.

To comprehensively evaluate the performance, SoftCast is compared with H.264 in the LTE environment according to the framework shown in Fig.8. Since the reconstructed video quality changes with the diversity of video, we take combination of the first 32 frames of *akiyo*, *coast*, *flower*, *foreman* and *paris* to create a 160-frame test video for a fair comparison. For H.264, the GoP structure is IPPPPPPP, and the reference frame of P frame is the previous frame. The search range of motion estimation is 32×32. Seven block sizes (4×4, 4×8, 8×4, 8×8, 8×16, 16×8, 16×16) are used in motion estimation, and the motion precision is 1/4 pixel. We adjust SoftCast's parameters to ensure that the two schemes have the same resource occupation.

At the sender, data produced by H.264 codec is packed

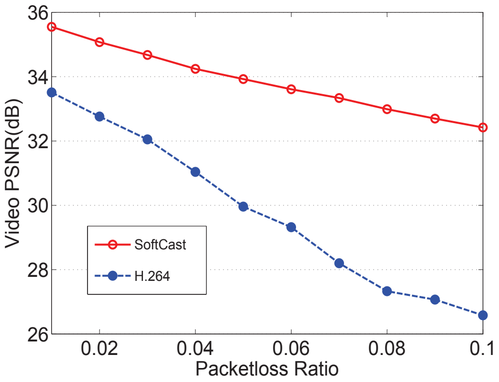


Figure 10: Comparison of resistence to packet loss between SoftCast and H.264.

into RTP packets of length 1200 bytes. Each RTP packet is added by a 24-bits cyclic redundancy check (CRC), encoded separately by FEC and QPSK modulation, and then transmitted over channels. In contrary, data produced by SoftCast coedc bypasses FEC and modulation. The packets of SoftCast and H.264 are transmitted separately over LTE EPA channels with the same statistical properties. The received H.264 packets are firstly demodulated and decoded. The decoder performs CRC for each RTP packet, and those error-free packets are forwarded to the H.264 decoder. The H.264 decoder is configured with the "motion copy" error concealment to tolerant a small percentage of erroneous RTP packets, which improves the quality of the reconstructed video. For SoftCast, the packets are directly forwarded to the SoftCast decoder.

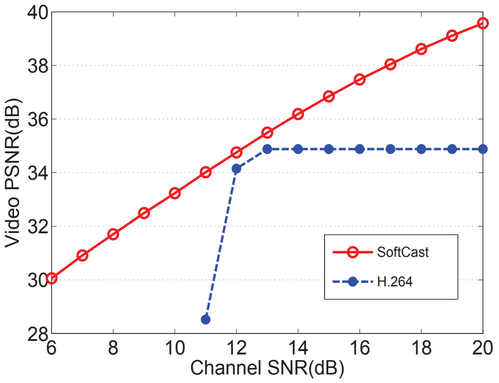


Figure 9: Performance comparison of scalable quality between SoftCast and H.264.

The simulation results are given in Fig.9. It is clear that H.264 suffers from a cliff effect. That is, when the channel SNR is below some critical point, the video quality degrades dramatically with the SNR since a large number of RTP packets are incorrect. In contrast, SoftCast's PSNR does not present a cliff but scales smoothly with the channel SNR. For all the SNR region, SoftCast clearly outperforms H.264 in terms of reconstructed video quality.

We also compare the resilience to packet loss of the two schemes in the same scenario. We set the channel SNR to be 15dB and allocate the packet losses uniformly at random with increasing probability from 0.01 to 0.1. The simulation results are provided in Fig.10. In H.264, video quality drops sharply when the packet loss increases because the Huffman encoding and differential encoding bring dependencies between packets. In comparison, SoftCast's video quality degrades gently as the packet loss increases. The video quality is acceptable even at a packet loss rate as high as 10%. SoftCast's resilience to packet loss benefits from 3D-DCT, which makes the loss of a single packet distributed across the entire GoP. In addition, the Hadamard transformation makes the energy equally distributed among the packets so that the video quality drops linearly as the packet loss increases.

5. CONCLUSIONS

In this paper we have provided a comprehensive performance evaluation and parameter optimization on the SoftCast wireless video broadcast scheme. In particular, we have investigates how the performance is affected by the parameters: chunk size, compression ratio, and the group of pictures. Optimization principles on parameter selection have been provided to obtain a better video quality, occupy less bandwidth and/or utilize lower complexity. The optimized SoftCast is compared with H.264 in the LTE EPA scenario. The simulation results have shown that SoftCast clearly outperforms H.264 in terms of scalability to channel conditions and resistence to packet losses. However, there are some other questions we have to consider. H.264 and SoftCast both have DCT and IDCT, but the size of DCT in SoftCast is bigger than the DCT in H.264. The DCT of H.264 is used in block (block sizes are 4×4, 8×8, 16×16, i.e.), while the DCT of SoftCast is used in a frame or a GoP (frame sizes are 176×144, 352×288, 512×512, i.e.). SoftCast employs 3D-DCT to conduct inter frame compression, but its coding efficiency is relatively low. In addition, due to the lack of multiresolution characteristic for DCT operation, SoftCast has no spatial and temporal scalability.

6. ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China (No.61401037), the Fundamental Research Funds for the Central Universities (No. BUPT2014RC0102), and Qualcomm Corporation.

7. REFERENCES

[1] T. Wiegand, G. Sullivan, G. Bjintegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. In *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):560-576, July, 2003.

[2] S. Kokalj-Filipović and E. Soljanin. Suppressing the cliff effect in video reproduction quality. *Bell Labs Technical Journal.* 16(4):171-185, March, 2012.

[3] S. Kokalj-Filipović, E. Soljanin, and Y. Gao. Cliff effect suppression through multiple-descriptions with split personality. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 948-952, July, 2011.

[4] Szymon Jakubczak, Hariharan Rahul, and Dina Katabi. Softcast: One video to serve all wireless receivers. In *MIT Technical Report, MIT-CSAIL-TR-2009-005*, 2009.

[5] Szymon Jakubczak, Hariharan Rahul, and Dina Katabi. One size-fits-all wireless video. In *Proc. eight ACM SIGCOMM HotNets Workshop, New York*, NY, USA, October, 2009.

[6] Szymon Jakubczak and Dina Katabi. A cross-layer design for scalable mobile video. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 289-300, September, 2011.

[7] Xiaopeng Fan, Feng Wu, Debin Zhao, C. Oscar Au, and Wen Gao. Disributed soft video broadcast (dcast) with explicit motion. In *IEEE Data Compression Conference*, pages 199-208, April, 2012.

[8] Xiaolin Liu, Wenjun Hu, Qifan Pu, Feng Wu, and Yongguang Zhang. ParCast: soft video delivery in MIMO-OFDM WLANs. In *Internation Conference on Mobile Computing and Networking(Mobicom)*, pages 233-244, August, 2012.

[9] Ruiqin Xiong, Feng Wu, Jizheng Xu, and Wen Gao. Performance analysis of transform in uncoded wireless visual communication. In *IEEE International Symposium on Circuits and Systems(ISCAS)*, pages 1159-1162, May, 2013.

[10] Xiaopeng Fan, Feng Wu, Debin Zhao, and C. Oscar Au. Distributed wireless visual communication with power distortion optimization. In *IEEE Transactions on Circuits and Systems for Video Technology*, 23(6):1040-1053, June, 2013.