Multiple Description Video Coding over Multiple Path Routing Networks

Shaoshuai Gao, and Hamid Gharavi
National Institute of Standards and Technology
100 Bureau Drive, Gaithersburg, MD 20899-8920
Email: sgao@nist.gov, gharavi@nist.gov

Abstract

This paper presents a multiple description (MD) video-coding scheme, which uses interlaced high signal to noise ratio (H-SNR) and low signal to noise ratio (L-SNR) coded frames to produce two bit-streams. At the decoder, when both bit-streams are received a high quality video will be reconstructed. If either one is received, a poorer but acceptable quality video can be reconstructed. In this approach, we also considered the mismatch between the encoding and decoding loops due to motion compensation if only one bit-stream is received. We first give the results of the rate distortion performance of our scheme assuming that only one description is received. We then present simulation results under packet loss circumstances, which are simulated by real-world IP networks and ad-hoc networks. It is shown that the proposed scheme has a superior performance when compared with the single description MPEG-4 video coder and MD coding using video redundancy coding.

1. Introduction

The main objective of Multiple Description Coding (MDC) [1]-[12] is to encode a source into two (or more) bit-streams such that a high-quality reconstruction is achieved when both (all) bit-streams are received successfully. On the other hand, a lower but still acceptable quality reconstruction can be accomplished in the presence of only one bit-stream. In a packet loss circumstance, as long as the packets of each description are not lost at the same time, a basic quality can always be obtained. As a result, it makes the compressed signal more error resilient over a packet-loss network. MDC in conjunction with multiple path transport (MPT), enables traffic dispersion and load balancing of the network, which improves the overall throughput of the network. In recent years MPT has been found to be particularly attractive for mobile ad-hoc networks [1], [19], which normally suffer from excessive bursts of packet drops due to their dynamically changing network topologies.

To satisfy the requirements of MDC for these applications, it is essential to introduce correlations between the two descriptions such that if one of the descriptions is lost, it can be estimated from the other. Introducing correlations would consequently result in expanding the source bandwidth. Optimizing tradeoffs between network resources, bandwidth expansions, and distortion are the most challenging aspects of MDC/multipath-routing, particularly for real-time signals.

In the case of video [10], it is more than just applying an MD image coder [6], [7] to the prediction error signal. Motion-compensated temporal prediction [11], for instance, discusses the use of multiple coding modes and redundancy allocation among them. Although in [11] a mismatch due to motion compensation has been taken into consideration, it is mainly based on the MDC on the spatial field. Wang et. al, proposed a scheme that encodes the GOBs to two thread, either of which is composed by interlaced H-SNR GOBs and L-SNR GOBs [12]. This MDC scheme is also based on the spatial field and doesn't consider the mismatch between the encoder and the decoder, which is a very important issue in terms of propagation distortion in the temporal direction.

To extend MDC on the temporal field, it is possible to modify the Video Redundancy Coding (VRC) approach, which was originally developed to improve the error resilience for a single description video transmission [13]. This approach uses more than one prediction thread to encode each video frame such that when one thread is lost, the frame rate will be reduced. Although VRC is a single description coder, it can be directly extended to an MD coder where each thread of P frame can be sent via a different route. In this case, each coded P-frame will depend entirely on the earlier P-frames of the same thread. This approach has been considered in [14], which uses a complicated multiple frame interpolation method, referred as Mcinterp, to conceal lost information. As our main objective in this paper is to enhance the error resilience of the encoder rather than error concealment, the VRC approach, referred to as MD-VRC, has been used for comparison purposes.

As far as our main contribution is concerned, we present a temporal-based MDC scheme that can completely solve the problem of mismatch. In the proposed scheme, video frames are encoded into two descriptions on the temporal field. Each description is composed of interlaced H-SNR and L-SNR frames. The order of H-SNR and L-SNR frames in the temporal direction changes alternatively between the two descriptions. We show that the proposed algorithm can significantly improve the performance compared with the single description (SD) MPEG-4 coder and the MDC-VRC coder under various IP-based test environments. This includes using our dual-path mobile ad-hoc routing testbed.

2. Proposed coding scheme

The simplest way to produce two video descriptions is to duplicate the video bit-stream coded by single description coder (In this paper we call this encoder the duplicate coder). That is, either one of the two bit-streams can achieve the highest quality. However, if both are received, one of them will be totally useless. In this case, the redundancy will be 100%. Bear in mind that the basic concept behind MDC coding is to reduce the redundancy to a lower level (at a given distortion). Thus, our main objective is to find a method that can approach the highest quality with minimal added redundancy when only one description is received.

The basic idea is to encode the frames using two different quantization step-sizes in an interlaced manner. For one description the frames will be encoded using a lower quantization step-size (i.e., high-quality) and then switching to a higher quantization step-size in turn. For the other description, however, the order of the quantization step-size should be reversed in order to make the bitrate of the two descriptions nearly the same. That is, one description's step-size should be low, high, low, high, and so on. The other description's quantizer step-size should be high, low, high, low, and so on.

Figure 1 shows the proposed encoder, in which Q_1 and Q_2 are quantization step-sizes where $Q_1<Q_2$. There are two types of outputs for the encoder. One is High quality (H-SNR) coded frame (coded by a lower step-size Q_1) and the other is Low quality (L-SNR) coded frame (i.e., re-quantizing the samples by a higher step-size Q_2). As shown in Figure 1, it can be observed that the low quality reconstructed frame is used as a reference frame for both L-SNR and H-SNR encoders. (Justifications for this arrangement will be discussed in the next section).

Figure 2 shows the block diagram of the decoder, which consists of H-SNR and L-SNR decoders. The received bit-streams 1 and 2 from two channels will switch to the corresponding decoder on a frame-by-frame basis. It should be noted that there are two frame buffers in the H-SNR decoder: one for decoding the H-SNR frame and the other for reconstructing the L-SNR frame. The reconstructed L-SNR frame will then be used as a reference frame to decode the next frame. This is to make sure that both descriptions use the same reference frame to encode the next low or high quality frame. As far as the L-SNR decoder is concerned, it operates just like a SD decoder.

In error-free circumstances, the final decoded frame will always have the H-SNR quality whereas its L-SNR reconstructed version will always be used to predict future frames. However, in packet-based transmission environments and under noisy channel conditions, some MBs in the decoded H-SNR frame or in the L-SNR frame may be lost. Thus, to efficiently utilize both H-SNR and L-SNR frames to recover some or all of the missing data, we have designed an MB combiner scheme, which will be described later.

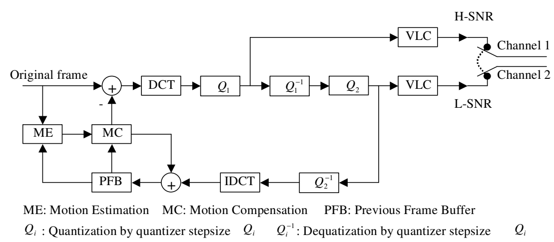

Figure 1. The proposed encoder

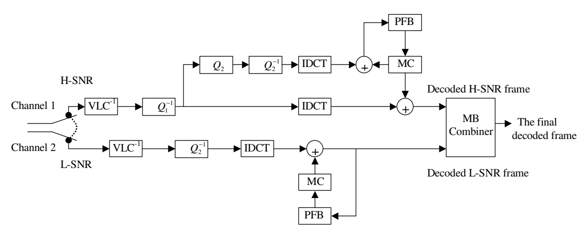

Figure 2. The proposed decoder

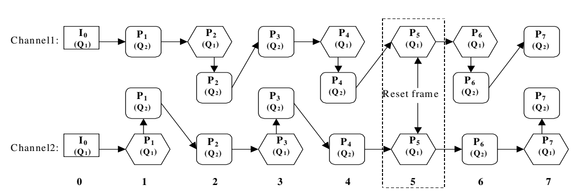

Figure 3. The proposed Multiple Description coding scheme using Interlaced High-SNR and Low-SNR frames

Figure 3 shows the overall structure of the proposed scheme for transmission of coded bistreams via two channels. For the first frame, which has to be encoded in INTRA mode, low quantization step-size (i.e., H-SNR) is used for both bit-streams. The following frames are encoded in an interlaced manner using two different step-sizes Q_1 and Q_2, which results in H-SNR and L-SNR coded frames. As mentioned earlier, both H-SNR and L-SNR frames use the L-SNR reference frame for interframe prediction. This is to make sure that a loss of one frame (e.g., H-SNR or L-SNR frame) will not affect the decoding of the following frames. For example, if only the H-SNR frame is lost, the current frame can be decoded as the L-SNR frame without having any effect on decoding the following frames. On the other hand, if only an L-SNR frame is lost, the H-SNR frame can be decoded as an H-SNR frame. In the absence of the L-SNR frame, however, the H-SNR frame has to be re-quantized in order to recover an L-SNR reference frame for predicting its future frames.

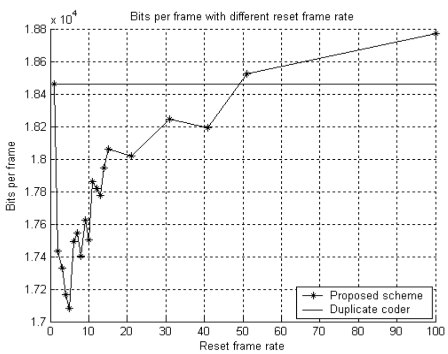

Figure 4. The effect of reset frame rate—"Foreman", QCIF, 7.5fps, 100 frames, packet size=200 bytes, Q_1=8, Q_2=16

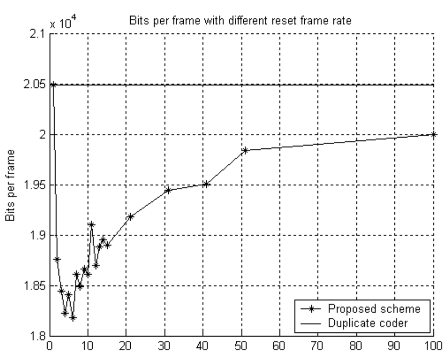

Figure 5. The effect of reset frame rate—"Coastguard", QCIF, 10fps, 100 frames, packet size=200 bytes, Q_1=8, Q_2=16

It is obvious that by selecting the reconstructed L-SNR frame as the reference frame for both channels, the prediction error between the high-SNR frame and the low-SNR reference frame will grow larger as the prediction progresses. Consequently, this could result in an increased bitrate, which could even surpass that of the duplicate coder. To overcome this situation, the L-SNR reference frame changes to H-SNR reference frame (reset reference frame) for both channels once every N frames.

To find the optimized reset reference frame rate, N, we evaluate the "Foreman" and "Coastguard" sequences (100 frames) with QCIF format for Q_1=8 (H-SNR), Q_2=16 (L-SNR). We change the reset reference frame rate from 1 (every frame is reset frame, just as the duplicate coder) to 100 (no reset frame). From Figure 4, we can see that for the "Foreman" sequence, when the reset rate is N = 5, the average bits per frame will be the lowest. That is, the reset rate of N = 5 can achieve the minimal redundancy rate. We can also observe that when the frequency of the reset is very low (e.g., always using the reconstructed L-SNR as the reference frame) the bit rate will exceed that of the duplicate coder. For the "Coastguard" sequence as depicted in Figure 5, we notice that N = 6 can achieve the minimal redundancy rate. However, if we set the reset rate to an even number, it will make the reset frames occur in one description and the average bits per frame of the two descriptions will consequently be very different due to the order of interlaced L-SNR and H-SNR frames in the proposed coding structure (Fig 3). For an odd number reset rate, the bitrates of the two descriptions are nearly the same. Therefore, in order to balance the bitrates, the reset reference frame rate, N, should be an odd number. Looking at both Figures we can see that N = 5 for both "Foreman" and "Coastguard" can provide the best results.

In [15] a video packet-recovery scheme using redundant packets is considered. The redundant packet only transports the most sensitive information in a coded video frame, such as the header information and the motion vectors (MVs). If there are some losses, the information in the redundancy packet can be used to conceal the lost MBs. Based on the same concept, we use the MB combiner in order to make the decoder get the best quality decoded frame by concealing the effect of the missing MBs.

The MB combiner works for two distinct cases. Case 1 assumes that some packets containing a number of MBs within the n^{th} H-SNR frame are lost. In this case, the corresponding decoded MBs in the n^{th} L-SNR frame are utilized to fill in the missing area. As a result, the modified decoded frame not only includes the decoded MBs from the H-SNR frame but also those were transferred from the L-SNR frame. Although there would be some degradation due to MBs replacement from the L-SNR frame, this cannot propagate to the following frames. Bear in mind that both channels use the L-SNR frame as the reference frame.

Case 2 is mainly concerned with losses in the first channel that could only affect the n^{th} L-SNR frame. In this situation, the missing MBs do not affect the final decoded frame but they can degrade the reference frame. To prevent this, it would be necessary to use the reference frame from the H-SNR decoder (via re-quantization) to fill in the lost area in the n^{th} L-SNR frame and thus, there would be no degradation at all.

In a more undesirable situation, packets from both channels that share some of the same coded MBs in the n^{th} frame may be lost. Under this condition, we can only estimate the MB by means of error concealment, which will be described in the results section.

3. The simulation results

We simulated the proposed coder using the MPEG-4 Verification Model (without B frame option and the video packet size = 200 bytes). To evaluate the performance of the proposed MD Video Coding using a different quantization step-size for each description, we use the same test sequences: "Foreman" (QCIF, 7.5fps, and 100 frames) and "Coastguard" (QCIF, 10fps, and 100 frames). In addition, for the sake of comparison, we also designed a MD version of video redundancy coding (MD-VRC). Note that MD-VRC uses the same INTRA coded frame (i.e., frame 0) for both channels. For the remaining frames, the even-indexed frames and the odd-indexed frames are independently encoded for transmission over channel 1 and channel 2, respectively. Obviously, each channel uses its own reference frame for interframe prediction. In addition to the MDC-VRC, in our evaluations we have also included the performance of the original MPEG-4 coder (single description coder).

In our first set of experiments, the same quantization step-size is used for Q_1, i.e., $Q_1 = 8$, whereas different step-sizes are selected for Q_2, i.e., $Q_2 = 8$ (i.e., duplicate encoder), $Q_2 = 16$, and $Q_2 = 31$ (note that 31 is the highest quantization step size that can be selected for the MPEG-4 encoder). Before evaluating each encoder under packet-loss transmission conditions, we first assess the rate distortion performance for one description in the absence of the other. Tables 1 and 2 show the rate distortion performance of different MD encoders for two sequences: "Foreman" and "Coastguard". For convenience, the proposed encoders are labeled as MDC. It can be seen that when both descriptions are received, the PSNR values are very close to each other. Note that for Q_2=8, the MDC behaves like a duplicate encoder. Its PSNR value is the same for one and two descriptions but with 100% added redundancy. The MDC with Q_2=16 has slightly lower PSNRs for both one and two-descriptions but with a lower redundancy. Increasing the step-size from Q_2 =16 to Q_2 = 31 does not appear to have any impact on lowering the redundancy rate but its PSNR quality drops by a wider margin with one description. Although MDC-VRC has the lowest redundancy with a good two-descriptions PSNR, its average PSNR drops sharply in the absence of one description. This is mainly because of the reduced frame rate where the average PSNR is measured by repeating the missing frames. From these initial results, we can observe that the MDC (Q_2=16) encoder can provide the best trade off between the redundancy rate and the quality. Therefore, this encoder will be used as our candidate MDC for further evaluations.

Table 1. Redundancy rate distortion performance of different MD coders – "Foreman"

CODERS	MDC (Q_2=8)	MDC (Q_2=16)	MDC (Q_2=31)	MDC -VRC
Redundancy (%)	100.00	85.09	85.12	26.44
Two-description PSNR (dB)	33.00	32.73	32.72	32.74
Average one-description PSNR (dB)	33.00	31.21	30.69	22.45

Table 2. Redundancy rate distortion performance of different MD coders – "Coastguard"

CODERS	MDC (Q_2=8)	MDC (Q_2=16)	MDC (Q_2=31)	MDC-VRC
Redundancy (%)	100.00	79.69	80.33	33.03
Two-description PSNR (dB)	32.10	31.93	31.95	31.75
Average one-description PSNR (dB)	32.10	30.15	29.62	24.73

To evaluate the relative performance of the encoders in more realistic environments, we first ran a series of tests where packets in both bit-streams are discarded using the packet-loss statistics of [17], corresponding to 3%, 5%, 10%, and 20% packet-loss rates. These four error pattern files are used to simulate the Internet backbone (IP networks) performance for video coding experiments.

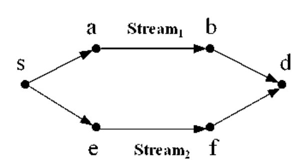

Figure 6. Test Scenario structure

Secondly, we evaluate the performance of these

MDC schemes using our mobile ad-hoc network testbed, which had been implemented utilizing the QualNet (Version 3.7) simulation tool. In this testbed, we considered the IEEE 802.11b standard and the DSR routing protocol [18]. However, since our objective is to evaluate the performance of MDC schemes, any discussion about multipath routing protocol aspects developed in our testbed is beyond the scope of this paper. Figure 6 shows our dual-path test scenario where nodes are not equally spaced and the communication is from node *s* to node *d*. In these experiments, the transmission power is 8.5 dBm, the receiver sensitivity is –93.0 dBm, the IEEE 802.11b data-rate is 2 Mb/s and the noise factor is 10.0. For simplicity, we assume that there’s no fading and the path loss model is free space. Also, by adjusting the distance between neighboring nodes, we are able to achieve packet loss rates of 3%, 5%, 10%, and 20%, respectively. Note that in a point-to-point CSMA/CA multihop communication, packets are continually competing to access the media in their shared collision domains and this can significantly affect the end-to-end packet loss performance [15], [16], [19].

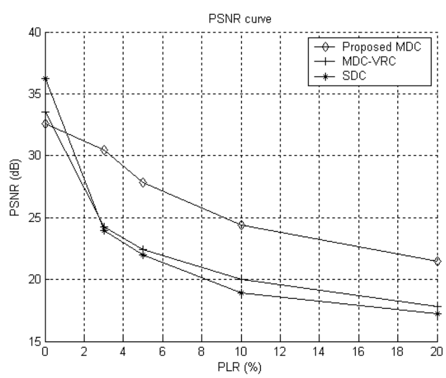

Figure 7. Objective quality comparison of different encoders over IP networks—“Foreman”, the bit rate is 144kbps.

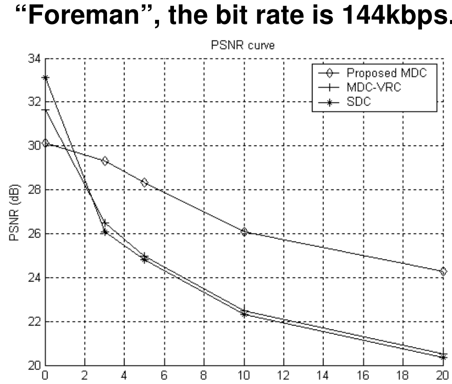

Figure 8. Objective quality comparison of different encoders over IP networks—“Coastguard”, the bit rate is 144kbps.

For fair comparison, we made sure that all the schemes are compared at the same average bitrate, which is achieved by adjusting quantization step-sizes. For the proposed MDC, this is done by changing the quantization step-size, Q_1, in such a way that the overall bitrate for both descriptions will remain the same when $Q_2 = 16$ is selected. In addition, the MB combiner that plays a crucial role in data recovery has been deployed. Bear in mind that if losses from both channels include the same data the decoder uses a simple error concealment algorithm. For the first frame (I-frame) a simple pixel interpolation algorithm is used to conceal the effect of missing MBs, whereas the following frames use the corresponding MBs from the previous frame. The same error concealment has also been used for the SDC encoder. For the MDC-VRC coder, we use the most adjacent frame to conceal the effect of missing packets (note that the first frame also uses pixel interpolation error concealment). We should point out that all the encoders use a random INTRA block refresh at the rate of 10%. In addition, the average PSNR values are measured by running each experiment 25 times. That is, a total of 2500 frames are tested for the final results.

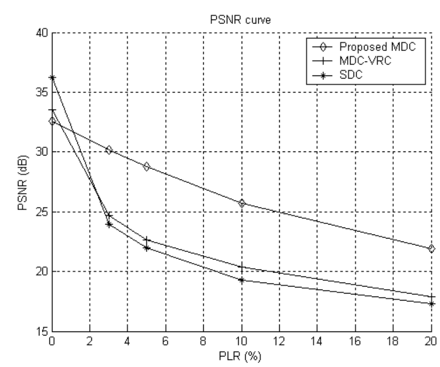

Figure 9. Objective quality comparison of different encoders over ad-hoc wireless networks—“Foreman”, the bit rate is 144kbps

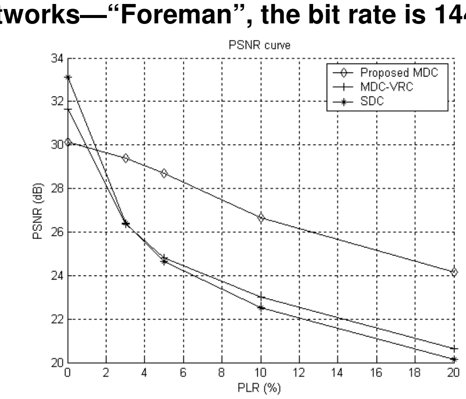

Figure 10. Objective quality comparison of different encoders over ad-hoc wireless networks—“Coastguard”, the bit rate is 144kbps.

Figure 7 and Figure 8 show the objective quality comparison of the three encoders over IP networks for "Foreman" and "Coastguard", respectively. As shown in these figures although the SDC encoder can achieve a much higher PSNR in an error-free case, its PSNR quality degrades rapidly with the increasing packet-loss rates. MDC-VRC can obtain a little improvement over the SDC coder in packet-loss circumstances. For "Foreman", only a 0.3-1.2 dB improvement can be observed. For "Coastguard" the improvement is even more minimal. For our proposed MDC coder, a much more graceful PSNR degradation can be observed compared with the other two coders and can get 3.5-6 dB improvement in packet-loss circumstances. Figure 9 and Figure 10 show the objective quality comparison of the three encoders over ad-hoc wireless networks for "Foreman" and "Coastguard," respectively. The similar results can be observed. However, a more graceful degradation can be seen for the proposed coder due to the higher burst packet loss in ad-hoc networks, which can be processed easier by the proposed MDC coder.

4. Conclusion

In this paper, we proposed a multiple description video-coding scheme, which uses interlaced H-SNR and L-SNR frames to produce two bit-streams (descriptions). In contrast with other MD encoders, this scheme is based on a temporal information division. In order to minimize the effect of error propagation, both channels always use the L-SNR frames as the reference frame (except the following frame of the reset reference frame, for which both channels use the H-SNR frames). This is to ensure that if only one description is lost, at any given time, it will not affect the decoding of the following frames. The decoder also takes advantage of the MB combiner scheme to improve the data recovery process. The proposed MDC scheme is then compared with the modified VRC coder, which is referred to as MDC-VRC, and the SDC MPEG-4 encoder in two test scenarios. One is assuming that only one description is received and the other one assumes both descriptions suffer from packet losses. Under both test scenarios it has been shown that the proposed MDC can greatly outperform the MDC-VRC and SDC. Moreover, it should be noted that although we realized our MDC scheme based on the MPEG-4 coder, it can also be used on other coders, such as H.263, MPEG-2, and so on.

5. Acknowledgment

The authors would like to thank Dr. Bo Yan for his help in developing mobile ad-hoc network testbed for generating error pattern files.

6. References

[1] S. Mao, S. Lin, S. S. Panwar, Y. Wang, and E. Celebi, "Video transport over ad hoc networks: multistream coding with multipath transport," *IEEE J. Select. Areas Commun.*,vol. 21, pp. 1721-1737, Dec. 2003.

[2] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.

[3] V. A. Vaishampayan, "Design of multiple description scalar quantizer," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.

[4] Y.Wang, M. Orchard, and A. R. Reibman, "Optimal pairwise correlating transforms for multiple description coding," in *Proc. ICIP98*, Oct. 1998.

[5] H. Jafarkhani and V. Tarokh, "Multiple description trellis coded quantization," *IEEE Trans. Commun.*, vol. 47, pp. 799–803, June 1999.

[6] Y. Wang, M. Orchard, V. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Processing*, vol. 10, pp. 351–366, Mar. 2001.

[7] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in *Proc. ICIP98*, Oct. 1998.

[8] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Proc. VCIP 99*, Feb. 1999.

[9] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Mag.*, vol. 18, pp.74-93, Sep. 2001.

[10] Y. Wang, A. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, pp. 57-70, Jan. 2005.

[11] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description coding for video using motion compensated temporal prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 193–204, Mar. 2002.

[12] Y. Wang, C. Wu, "Low Complexity Multiple Description Coding Method for Wireless Video," in *Proc. AINA05*, vol. 02, pp. 95-98, Mar. 2005.

[13] S.Wenger, "Video redundancy coding in H.263+," in *Proc. AVSPN*, Aberdeen, U.K., Sept. 1997.

[14] J. G. Apostolopoulos, "Error-resilient video compression through the use of multiple states," in *ICIP00*, vol. 3, pp. 352-355, 2000.

[15] H. Gharavi and K. Ban, "Dynamic Adjustment Packet Control for Video Communications over Ad-hoc Networks," in *Proc. ICC04*, Paris, France, June 2004.

[16] H. Gharavi, "Control Based Mobile Ad-hoc Networks For Video Communications," IEEE Trans. Consumer Electronics, vol. 52, no. 2, May 2006.

[17] S. Wenger, "Error patterns for Internet experiments," ITU Telecommunications Standardization Sector, Red Bank, NJ, Doc. Q15-16r1, Oct. 1999.

[18] D. B. Johnson and D. A. Maltz, "Dynamic Source Routing in Ad-Hoc Wireless Networks", In *Mobile Computing*, edited by T. Imielinski and H. Korth, Chapter 5, pp. 153-181, Kluwer Academic Publishers, 1996.

[19] B. Yan, and H. Gharavi, "Multi-path Multi-Channel Routing Protocol," in *Proc. IEEE NCA06*, July 2006.