New Compression Techniques for Robust and Scalable Media Communications

1. Research Team

Project Leader: Prof. Antonio Ortega, *Electrical Engineering*

Graduate Students: Hyukjune Chung, Huisheng Wang

2. Statement of Project Goals

As international standards (the latest examples being MPEG-4, H.264 and JPEG 2000) continue to be developed and provide widely adopted solutions for media communications, some of the most important research problems in multimedia compression are becoming those where pure compression performance is not the main requirement. Instead, in this project we are focusing on a number of system-level issues that will need to be resolved before a full deployment of media communications can take place.

The most popular and widely deployed media communication applications are likely to be those that are accessible over the Internet, both through the currently predominant wired access and through emerging wireless access channels. This requires the design of compression techniques that are robust to channel impairments (variations in bandwidth and delay, packet losses, bit errors, etc). Our work in this area concentrates on: (i) Long-term memory motion compensation, where the motion compensation gain can be significantly increased by extending the motion search range with multiple decoded frames as reference frames, instead of using just one decoded frame as in conventional motion compensation, (ii) Multiple Description Layered Coding (MDLC), where the advantages of layered coding and multiple description coding techniques are combined to provide robust video communication by adapting to the varying channel conditions on the fly, (iii) Wyner-Ziv Scalability (WGS), where based on the Wyner-Ziv framework, the proposed coder can support embedded representation and high coding efficiency by using the high quality version of the previous frame as side information in the enhancement-layer coding of the current frame.

3. Project Role in Support of IMSC Strategic Plan

While our work is mostly concerned with compression, we are addressing a series of issues that arise when compression technology gets integrated into real network systems. Approaches for robust communication and caching should become useful in applications where video transmission over existing networks, or future wireless links, is required. The scheduling for streaming applications considered should be integrated with other mechanisms for rate control that are embedded in video servers being studied within IMSC. Our research here complements that work in that we provide mechanisms for changing the transmission rate by modifying the quality of the video streams, while other techniques under consideration limit themselves to modifying the rate at which they transmit data, without actually modifying the data itself. The work presented here will become even more practical as progress is made in developing more efficient scalable video algorithms, e.g., those emerging from studies of Fine Grain Scalability (FGS) within MPEG-4.

Compression is a key component in providing media delivery to the broadest range of terminals and using all types of available transport mechanisms. The new compression technologies described in this project support the IMSC vision by allowing the integration of media delivery into current and future networks (e.g., the Internet) and physical channels (e.g., UWB).

4. Discussion of Methodology Used

Long-term memory motion compensation. We propose a novel adaptive motion search window location algorithm, which is the modified version of [18] to overcome the aliasing problem, which occurs when low resolution frames are used to estimate motion in high resolution ones. Our proposed fast search algorithm achieves low complexity by reducing the number of motion candidates based on the scene characteristics gathered from a low resolution version of the given sequence. However, for the case when the blocks being predicted and / or the candidate blocks contain strong high frequency components, it is more likely that the motion search window (MSW) location obtained by the proposed algorithm will be incorrect. This is because the low resolution may have lost some important information contained in the original resolution. To overcome this problem, in this project, we choose a motion vector predicted from neighboring blocks as additional information, which will provide a good location for MSW if the motion field is regular [20]. To locate the MSW we use both low resolution vectors and neighboring vectors. The proposed algorithm is implemented in reference software [19] for the H.264 video compression standard, and the parameters of the proposed algorithm are optimized for the implementation.

Multiple Description Layered Coding. By combining the advantages of Layered Coding (LC) and Multiple Description Coding (MDC), this approach can provide reliable video communication over a wider range of network scenarios and application requirements. MDLC improves LC in that it introduces redundancy in each layer so that the chance of receiving at least one description of base layer is greatly enhanced. Though LC and MDC are each good in limit cases (e.g., long end-to-end delay for LC vs. short delay for MDC), the proposed MDLC system can address intermediate cases as well. Same as a LC system with retransmission, the MDLC system can have a feedback channel to indicate which descriptions have been correctly received. Thus a low redundancy MDLC system can be implemented with our proposed runtime packet scheduling system based on the feedback information. The goal of our scheduling algorithm is to find a proper on-line packet scheduling policy to maximize the playback quality at the decoder. Previous work on scheduling algorithms has not considered multiple decoding choices due to the redundancy between data units, because of the increase in complexity involved in considering alternate decoding paths. In this work, we introduce a new model of Directed Acyclic HyperGraph (DAHG) to represent the data dependencies among frames and layers, as well as the data correlation between descriptions. The impact of each data unit to others is represented by messages passing along the graph with updates based on new information received.

Wyner-Ziv Scalability. This approach supports embedded representation and high coding efficiency by using the high quality version of the previous frame as side information in the enhancement-layer coding of the current frame. The traditional closed-loop prediction (CLP) approach requires the exact value of the predictor to create the residue. For a scalable coder either a single prediction is used, which leads to either drift or coding inefficiency, or a different prediction is obtained for each reconstructed version, which leads to added complexity. Moreover, it also suffers the inherent limitation that both encoder and decoder must be synchronized to avoid drift. On the other hand, Wyner-Ziv coding (WZC) only requires the correlation structure between the current signal and the predictor, so there is no need to incorporate a decoder at the encoder. The proposed coder uses nested lattice quantization followed by multi-layer Slepian-Wolf coders (SWC) with layered side information. Nested lattice quantization provides a straightforward mechanism for scalable source coding, where the coarse and fine quantizers generate the coarse and fine descriptions, respectively. The multi-layer SWC encoders can exploit the correlation between the quantized sample and the base and enhancement-layer reconstruction of the previous sample.

5. Short Description of Achievements in Previous Years

MDC coding. Previously proposed MDC schemes do not allow the level of redundancy to be easily changed. For example, this may require changing the transform at both encoder and decoder. Clearly, these approaches are limited in their ability to adapt to changing transmission conditions. In our work [10], we have proposed a simple approach for MDC that uses a polyphase transform and deterministic redundancy (e.g., each sample of input data is transmitted several times, with different coding rates). This approach is useful in that it greatly simplifies the design of a MDC scheme, since the rate allocation determines the amount of redundancy. Moreover, it provides a great deal of flexibility as it enables the choice of redundancy to be almost arbitrary. We have introduced an optimal bit allocation algorithm that allows us to select the amount of redundancy to be introduced in the signal that best matches a given target packet loss rate [11]. Our results show significant differences between optimal and suboptimal choices of redundancy for a given packet loss rate. Moreover, given that the decoder remains unchanged when the bit allocation changes, it is possible to adapt very simply to the changes in channel behavior without requiring a change in the packet sizes, or the structure of the decoder.

Since our previous work [11] demonstrated that bit allocation could be used to adjust the level of redundancy to match channel conditions we have studied this bit allocation problem in the context of a progressive wavelet coder such as SPIHT (Set Partitioning in Hierarchical Trees) [14.15]. Note that the bit allocation problem we have solved is useful not only for MDC but also other applications such as Region of Interest coding (ROI). In both cases, wavelet coefficients are divided by different factors before coding in order to enable different bit allocation to different regions. The basic idea in SPIHT (and in similar algorithms) is to perform bitplane encoding of the wavelet coefficients, so that larger coefficients start being transmitted first. By selecting appropriate weighting factors one can ensure that the coefficients in one region tend to be larger so that they are transmitted first and refined to higher resolution than those in other regions. Essentially this is equivalent to providing unequal bit allocation to each of the regions. While this technique has been used for ROI coding, we are the first to propose using it for MDC as well. In this work, we call ***priority-scaling factor*** (psf) the dividing factor. Our main contribution is to provide an analytical technique to determine what the psf value should be, given criteria such as relative importance of the regions or degree of redundancy in an MDC coder. Our approach is based on a novel Rate-Distortion (RD) model, which is an extension of Mallat's model to incorporate the specific characteristics of our proposed psf technique. In other words, our approach enables us to model analytically the RD values for any psf values, so that the optimal psf values can be calculated.

In addition we studied the use of MDC for correcting errors in DPCM and other memory based compression techniques [12.17]. For typical image/audio sources, which have memory, it is necessary to use these kinds of methods to achieve reasonable compression performance. However, approaches such as DPCM have the drawback of catastrophic error propagation, i.e., a 1 bit channel erasure in the prediction error will cause all samples after this erasure to be decoded incorrectly. In our work we propose to use a low rate description to correct erasures in the high rate description. This is a novel approach for erasure recovery in memory coders and results are very encouraging.

In the future, we propose to extend our work to motion prediction scheme in MPEG coders. Experimental results have shown that the proposed system can estimate data lost due to large burst of erasures.

We have also compared MDC to Layered Coding under various network conditions using a network simulator [13]. We find that for applications that require short latency, e.g. video conferencing, and for networks with large round-trip-time, MDC outperforms Layered Coding in case of even small packet losses over the network.

Standard-compatible MDC video coding: Another important contribution of our work is to demonstrate that it is possible to achieve some of the robustness of MDC while being fully compatible with existing standard video coders. We propose an unbalanced MDC system, where one of the descriptions has a high resolution and the other a low, but acceptable resolution (quality). The low-resolution description (LR) is explicitly used as redundancy, to be decoded in cases of losses in the high-resolution description (HR). Each of the descriptions are encoded and decoded by a standard compression system, in our case the H.263 video encoder. There has been limited work on MDC based video systems mainly because of the *prediction loop mismatch* problem. In a prediction-based decoder, if the prediction is based on information unavailable at the decoder, there is a prediction loop mismatch leading to poor MSE performance. Our system avoids the prediction loop mismatch problem because the central decoder is the same as the high-resolution decoder, with the LR description being used only when there are no losses in the HR description. Thus, the LR signal, unlike the LR signals in layered coding, is not meant to be usable by itself.

In our approach we obtain HR first and then for each block in each HR coded frame we decide to select only a certain number of those transform coefficients to be included in LR. This is done according to a criterion where given a total amount of allowable redundancy we choose the coefficients that minimize the distortion in case of loss. The motion vectors are duplicated in both the descriptions and it is assumed that there are no correlated losses, i.e., erasures do not occur at the same location in both the descriptions.

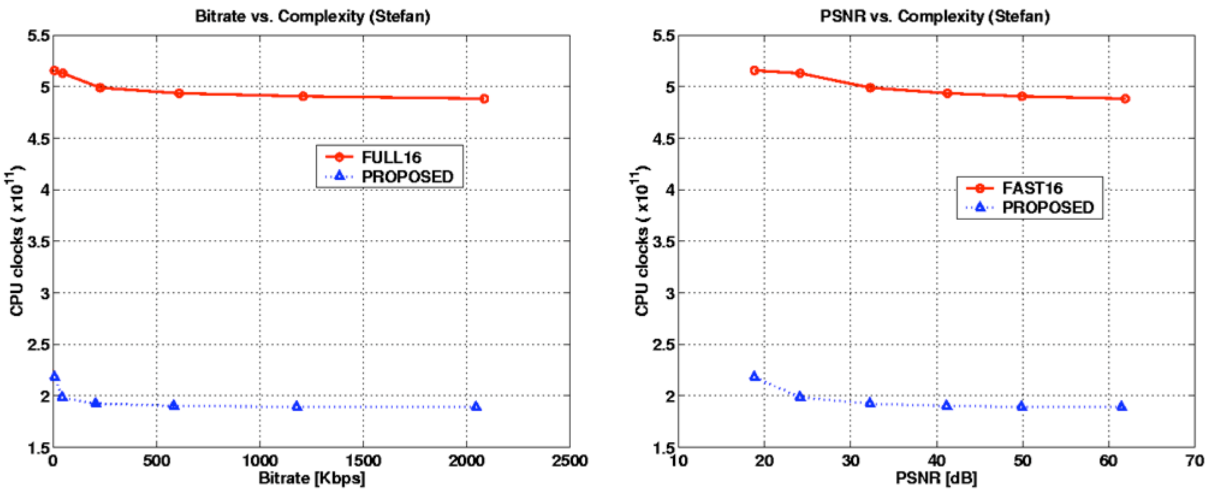
Streaming media delivery. We propose a framework for scalable streaming media delivery that involves a novel scheduling algorithm called *Expected runtime Distortion Based Scheduling*, EDBS. This algorithm decides the order in which packets should be transmitted in order to improve client playback quality in the presence of channel losses.

The *Expected runtime Distortion* considers all the important elements that will affect the playback quality at client end, including, (i) the information contained in each data packet, which determines the reconstructed quality of the video, (ii) the playback end-to-end delay, which determines the duration of packet lifetime, (iii) the channel conditions, which affect the packet loss, and (iv) the packet dependencies, which produced by the scalable compression codec. All these factors are evaluated in our proposed transmission algorithm, which identifies the best schedule for packet delivery based on the result of this evaluation so that the playback quality is maximized. The detail of this algorithm and evaluation procedure can be found in [26]. We also provide a fast greedy search algorithm that achieves almost the same performance as an exhaustive search technique (98% of the time it results in the same schedule) with very low complexity and is applicable for real-time applications.

5a. Detail of Accomplishments During the Past Year

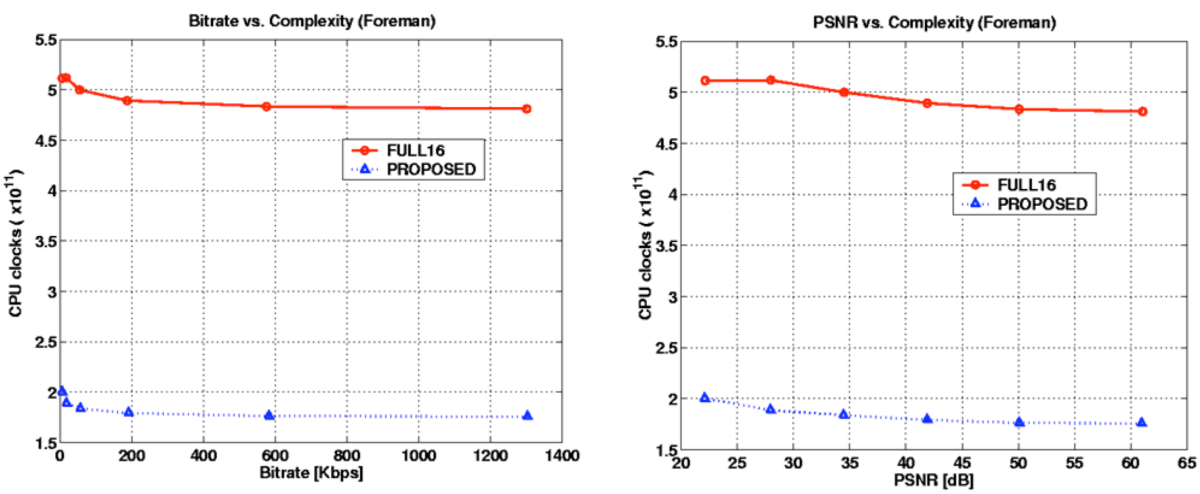
Long-term memory motion compensation. To demonstrate the performance of the proposed fast search algorithm, we implement our fast search algorithm in H.264 encoder reference software version JM 5.0c [19]. The proposed algorithm is implemented in the low-complexity mode of JM5.0c encoder. To locate MSWs, we employ the MRS-AWL algorithm proposed. For the fast search algorithm, we set the parameter values as follows. For the spatial reduction of the search range, T_S is set to 0.5, and the search range parameters are set to 6 or 10. For the temporal reduction of the search range, we set the threshold to 1. That is, if the normalized distortion Z is greater than 1, then we exclude the corresponding reference frame from the set of reference frames for a given macro-block. For JM 5.0c, the range of supported motion vectors is determined by the search range parameter. However, due to the nature of MRS-AWL algorithm, the motion vector can be much larger. Therefore, we also modified the reference software to support the larger motion vector range.

The simulation is done using an Intel Pentium IV 2.4 GHz PC with 512MB RAM. For experiments, qcif sequences *Stefan*, *Foreman*, and *Mother & Daughter* are used (100 frames for each with frame skip parameter 2). The computational complexity is measured by the number of CPU clock cycles for motion estimation. For the proposed algorithm, the overall complexity required for low resolution processing and low resolution motion estimation is also considered for the computational complexity estimate. For all the simulations, the first frame is coded as I-frame, and the other frames are encoded as P-frame, and for P-frames, all the inter block search modes are enabled. That is, in addition to 16x16 macro-blocks, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4 blocks are enabled. Also, UVLC is used as an entropy coding method. As a reference for comparison, we use a full-search method with low-complexity motion estimation and mode decision proposed in [20].



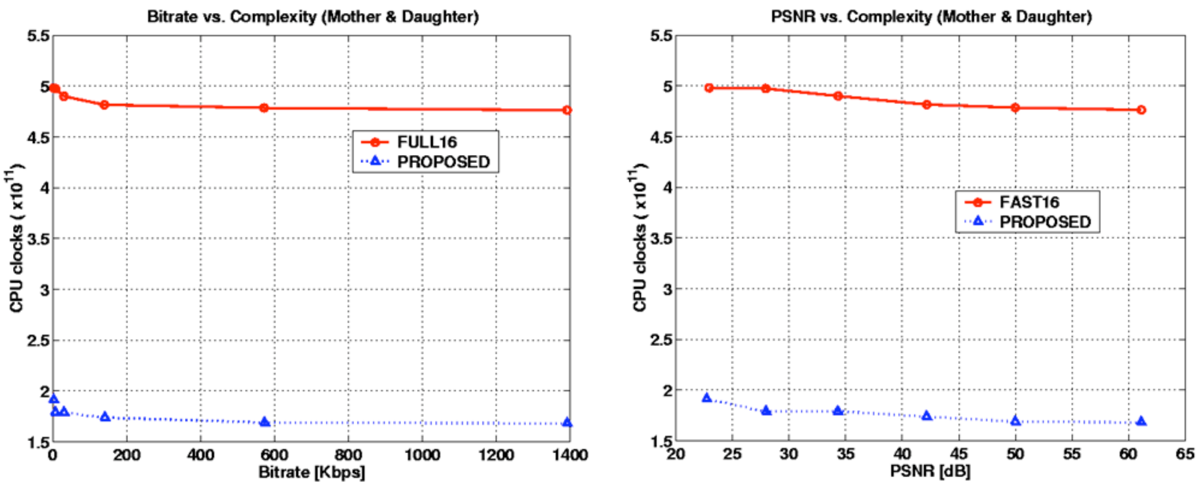
(a) (b)

Figure 1 Bitrate vs. Complexity (a) and PSNR vs. Complexity (b) graphs for H.264 encoder reference software with full-search with search parameter 16 (FULL16), and with proposed fast search (PROPOSED). Complexity is measured by the number of CPU clocks for motion estimation. The result is for *Stefan* sequence.



(a) (b)

Figure 2 Bitrate vs. Complexity (a) and PSNR vs. Complexity (b) graphs for H.264 encoder reference software with full-search with search parameter 16 (FULL16), and with proposed fast search (PROPOSED). Complexity is measured by the number of CPU clocks for motion estimation. The result is for *foreman* sequence.



(a)

(b)

Figure 3 Bitrate vs. Complexity (a) and PSNR vs. Complexity (b) graphs for H.264 encoder reference software with full-search with search parameter 16 (FULL16), and with proposed fast search (PROPOSED). Complexity is measured by the number of CPU clocks for motion estimation. The result is for *Mother & Daughter* sequence.

In Figure 1-3, we show bitrate vs. complexity and PSNR vs. complexity graphs for H.264 encoder reference software with full-search and with the proposed fast search. For this simulation, the quantization parameter (QP) for all the frames except for the first frame varies as 0,10,…,50, and QP for the first frame varies as 1,11,…,51. For the full-search, search range parameter is set to 16. As one can see from this figure, the proposed algorithm achieves similar bitrate and PSNR with significantly reduced computational complexity. The speed up factor of the proposed algorithm is around 2.5 for *Stefan*, 2.7 for *Foreman*, and 2.8 for *Mother & Daughter* as compared to the full-search.

The search range parameter for the proposed algorithm in level 2 (low resolution) is set to 8, this parameter corresponds to MSW of size 17x17 in level 2, and this MSW corresponds to MSW of size 68x68 in level 0 (integer resolution). Also, in the level 0 frame, search range parameter for the proposed algorithm is 6 or 10, which corresponds to MSW of size 13x13, and MSW of size 21x21 respectively. Therefore the proposed algorithm can use MSW of size 89x89, and this size of MSW corresponds to the search range parameter 44. Thus, in Table 1, we compare the bitrate and computational complexity of the proposed algorithm with the full-search with search range parameter 44 (FULL44). In this table, relative bitrate and speed up factor of our proposed algorithm are shown. Positive bitrate percentages represent relative increase in bitrate of the proposed algorithm, and vice versa. In this table QP is the quantization parameter for P-frames. Our proposed algorithm can detect motion vector as large as can be detected by FULL44. However, as one can see in this table, the speed up factor can be around 17 with small increase in bitrate as compared to FULL44.

Because in the original resolution, the proposed algorithm chooses search parameter 10 or 6 based on contents of scenes. Therefore, we compare the proposed algorithm with a full search with search parameter 10 (FULL10) in Table 1. As one can see from this table, our proposed algorithm achieves similar or lower bitrate with reduced computational complexity (speed up factor 1.6). This is because our proposed algorithm can locate MSW at a better position. Specifically for high motion scenes like *Stefan*, one can see that the proposed algorithm can achieve 21% bitrate saving as compared to FULL10.

Table 1 Relative bitrate and computational complexity of the proposed algorithm as compared to a full-search. Namely, FULL44 is a full-search with the search range parameter 44 and FULL10 is a full-search with the search range parameter 10. Bitrate entry in the table represents percentage increase (positive) or decrease (negative) of the proposed algorithm. SF represents speed up factor. QP is the quantization parameter for P-frames.

Stefan	FULL44		FULL10	
QP	bitrate	SF	bitrate	SF
0	+1.41%	15.47	-3.41%	1.55
10	+2.25%	15.50	-4.68%	1.55
20	+4.45%	15.44	-8.32%	1.55
30	+7.63%	15.40	-16.00%	1.54
40	+9.88%	15.36	-21.91%	1.52
50	-0.65%	14.41	+1.07%	1.35
Foreman	FULL44		FULL10	
QP	bitrate	SF	bitrate	SF
0	+0.58%	16.38	-0.45%	1.62
10	+1.67%	16.42	-0.09%	1.62
20	+3.52%	16.24	-0.26%	1.62
30	+4.08%	16.04	-0.77%	1.62
40	+4.57%	16.01	-3.66%	1.60
50	+0.48%	15.46	-3.66%	1.49
M&D	FULL44		FULL10	
QP	bitrate	SF	bitrate	SF
0	0.00%	17.10	+0.07%	1.66
10	+0.33%	17.13	+0.28%	1.66
20	+0.73%	16.69	+0.64%	1.62
30	+0.54%	16.43	+0.37%	1.60
40	+1.47%	16.77	+0.49%	1.61
50	-1.98%	16.10	-1.57%	1.47

Multiple description layered coding. Our general approach for multiple description layered coding (MDLC) [28] uses an MDC encoder to generate two base layer descriptions BL_1 and BL_2. For each base layer description BL_i, we create its corresponding enhancement layer EL_i. An extra enhancement layer description EL_0 is created as the difference between the original video signal and the combined signal of both base layer descriptions. The MDLC decoder selects which *EL* stream to decode given what base layer was received, and finally reproduces the signal by combining the base layer and enhancement layer information. The key advantage of our MDLC scheme is that it combines the hierarchical scalability provided by LC with the reliability introduced by adding redundancy into base layer with multiple descriptions. It provides the flexibility for the scheduling algorithm to choose the right base layer and enhancement layer descriptions based on the current network conditions and feedback information of previous transmission history.

For these experiments we have assumed that the transmitter would incorporate a run-time scheduler and a time window control. In the experiments shown here an observation of the channel and an *a priori* model, are used by the transmitter to select the most important packets. Our double time window control (DTWC) scheme, is different from the traditional method [29] in that it introduces separate windows of transmission opportunities (WoP) w_1 and w_2 for different descriptions. The DAG model used in previous research [29, 27] cannot be directly applied to a MDLC system because it can only consider one decoding choice. Here we introduce a new model called Directed Acyclic HyperGraph (DAHG). A DAHG is like a normal DAG, but each vertex is composed of a clique which contains a set of nodes and every pair of nodes are connected by an undirected edge to indicate their particular redundancy relationship. An *expected runtime packet distortion* is calculated for each candidate packet in the transmission buffer based on the channel conditions, data dependencies and correlation between packets, and the previous transmission history. At a given time, the packet with the greatest expected distortion is selected to be sent.

Figure 4 shows an experimental result on comparing the performance of LC and MDLC both of which use on-line scheduling. The results show that our proposed MDLC scheme provides more efficient and robust video communication in a wider range of network scenarios and application requirements.

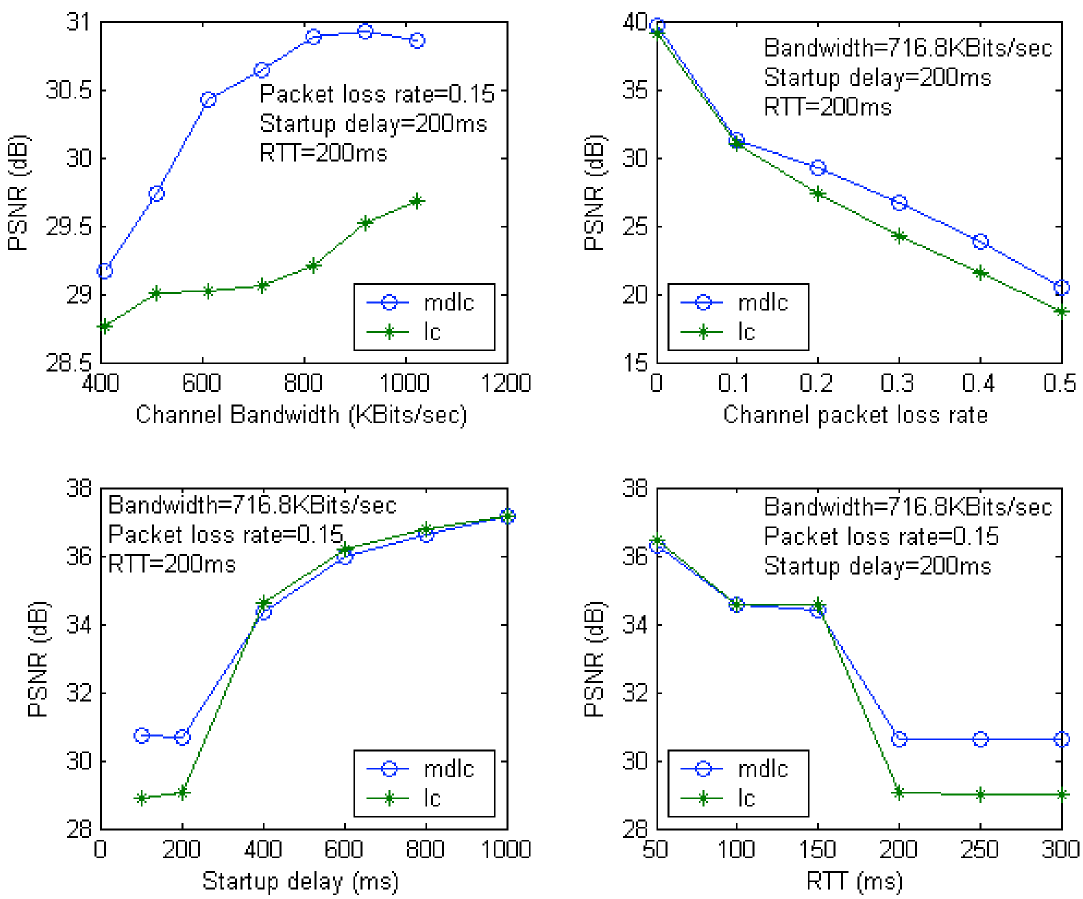


Figure 4. Comparison of PSNR for *Stefan* sequence between MDLC and LC with various parameters. The results show that by using the proposed MDLC system the playback quality improves about 2dB compared to the traditional LC system in the scenarios when a channel has large packet loss rate, long RTT or an application has stringent delay requirement.

Wyner-Ziv Scalability. The temporal evolution of DCT coefficients can be usually modeled by a first-order Markov process $x_k = \rho x_{k-1} + z_k$, where x_k is a DCT coefficient in the current frame and x_{k-1} is the corresponding DCT coefficient after motion compensation in the previous frame. Let x_k, $\hat{x}_k^b$ and $\hat{x}_k^e$ be the current sample, its base and enhancement-layer reconstruction, respectively. We assume that the motion vectors are transmitted separately and the same motion vector is shared for each macro block among all layers. Developed on a DPCM source model, the proposed encoder [33] consists of a pair of nested lattices (Λ_e, Λ_b) to support two-layer source quantization, and multi-layer SWC encoders to exploit the correlation between the quantized sample and the base and enhancement-layer reconstruction of the previous sample.

To demonstrate the performance of the proposed approach, we consider two-layer coding of first order Gauss-Markov sources of length 10^5 samples. The results are calculated as an average of these samples. We use nested scalar quantization and ideal SWC (i.e., the conditional entropy is calculated to approximate the rate after SWC) in our approach. The following methods are compared:

(1) the proposed coder.

(2) WZ-FGS, where the enhancement layer of x_k is coded only with the information from its base layer by nested quantization. The reconstruction at the decoder still takes into account $\hat{x}^e_{k-1}$.

(3) WZ-Simulcast, where the enhancement layer of x_k is coded directly with side information $\hat{x}^e_{k-1}$, i.e., each layer is coded separately or by simulcast.

Figure 5 shows the performance comparison between the above methods for various enhancement layer rates with ρ=0.99. The base layer is identical for all coders using Wyner-Ziv coding method. Also provided for reference are the Wyner-Ziv bound and the performance of nonscalable coder at the same total rate. The enhancement layer of the proposed coder performs better than WZ-FGS at high rates with gains over 3dB. The gain over WZ-Simulcast diminishes as the enhancement-layer rate increases. This is reasonable since when the enhancement-layer rate is high, $\hat{x}^b_k$ provides less useful information to code x_k as compared to $\hat{x}^e_{k-1}$ (when $\rho \to 1$), while the penalty due to the redundancy of simulcast is small when the base-layer rate is low.

We also compare the proposed coder and the traditional closed-loop prediction (CLP) based methods, as shown in Figure 6. An optimal entropy-constrained uniform threshold quantizer (UTQ) is used in the CLP methods, and the enhancement layer is predicted only from its own base layer in the two-layer CLP scalable coder, as in MPEG-4 FGS. It is clear that the proposed coder consistently outperforms the CLP scalable coder. It is also noted that the nonscalable CLP and WZC coders perform quite closely in the middle and high rate range.

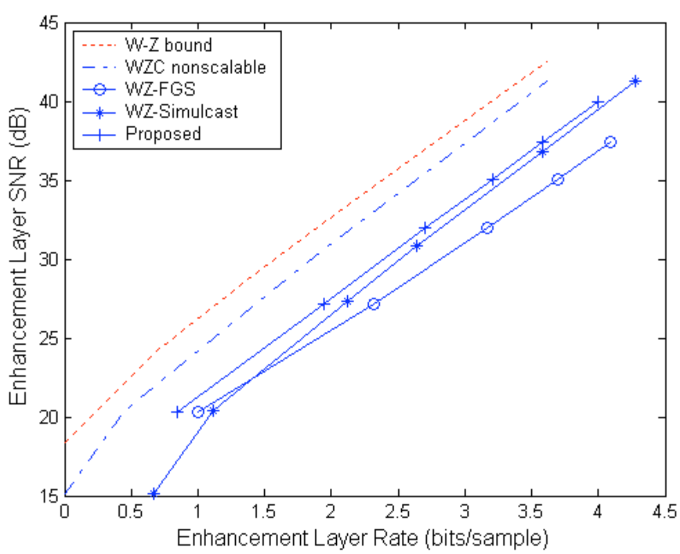


Figure 5. Comparison between the proposed two-layer scalable coder, WZ-FGS and WZ-Simulcast for Gauss-Markov source with ρ=0.99. Base-layer rate is 0.66 bits/sample.

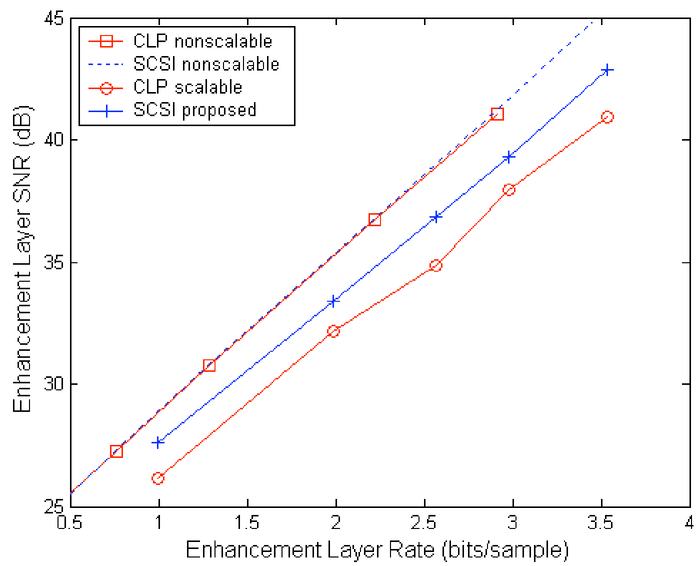


Figure 6. Comparison between the proposed two-layer scalable coder and the traditional CLP coders for Gauss-Markov source with ρ=0.99. Base-layer rate is 1.31 bits/sample.

6. Other Relevant Work Being Conducted and How this Project is Different

Long-term memory motion compensation. In this work, we propose a novel low complexity ME algorithm for long-term memory motion compensation (LTMC) based on a multiresolution search. The main idea of our approach is to use a lower resolution version of a given video sequence to obtain information about the motion estimation on the original resolution sequence. Multiresolution motion search has been well known and used for standard single-frame motion compensation as in [41.42.43.44.45.46.47.48,49.50]. These multiresolution motion searches use the hierarchy of lower resolution version of a frame to locate search positions only. However, our proposed multiresolution search gathers not only search position information also useful information such as search range (size of MSW), order of search, and stop criterion, etc.

Examples of our proposed framework can be found from our previous publications [18.51.52]. For the proposed work, we introduce several novel techniques to exploit multiresolution search efficiently in the context of LTMC.

Multiple Description Coding. Research work on robust communication using multiple description coding (MDC) techniques [1] has been actively conducted in recent years in both industry, e.g., Bell Labs [2.3] or AT&T Research [4.5.6], as well as in academia, e.g., the University of Illinois, at Urbana-Champaign [6.7], the University of California, Berkeley [8], the University of Washington [9] and Princeton University [5]. In an MDC scheme the source coder generates two or more compressed versions (or descriptions) of the signal, such that if *either* one is lost a reasonably good reproduction of the signal at the decoder can still be achieved. Conversely, high quality decoding is achieved when both descriptions are received. Obviously, the added robustness of sending two descriptions comes at the cost of redundancy and therefore worse coding performance. Some recent work has focused on transform-based MDC techniques [2], where the transform itself preserves some redundancy, while other researchers have proposed MDC approaches based on scalar quantization [5.6]. More recent approaches have demonstrated the benefits of combining MDC with forward error correction (FEC) to increase the robustness and enable unequal error protection to the various components of the signal [7.8.9].

The main novelty in our MDC approach is that we use explicit redundancy to generate the descriptions, that is, each sample of the signal can be coded and transmitted more than once. As will be discussed, this approach greatly reduces the complexity of designing the system. Previously proposed techniques tend to be optimized for an expected level of packet losses, and will not perform well in situations where there is a mismatch between expected and actual packet losses. In our proposed approach switching between different levels of redundancy is very straightforward so that a system can be made adaptive to changes in packet loss rates without requiring a complete redesign of encoder/decoder. We also introduce the model-based analytical bit allocation based on a novel rate-distortion model inside each description where we use the divide-and-multiply method to vary the quality of different parts of an image. Another major contribution of our MDC work is towards robust transmission of predictive encoded data. We have developed an algorithm for estimating data lost due to channel noise for a DPCM encoder and a MDC based video encoder that is compliant with the H.263 standard video codec. Our work on MDC has been published in [10.11.12.13.14.15.16.17].

Multiple description layered coding. Our work is an extension of recent work at USC and elsewhere on scheduling algorithms for layered coded video. Chou and Miao [29] provide a rate-distortion optimized approach of packet scheduling over a lossy packet network. Based on a simple source model of a Directed Acyclic Graph (DAG), the proposed approach can allocate bandwidth among packets and schedule them in such a way that a Lagrangian cost function of expected distortion and rate is minimized. To reduce the scheduling complexity, Miao and Ortega [30.31] propose an Expected Run-time Distortion Based Scheduling (ERDBS) algorithm, which simply uses a greedy solution by explicitly considering the effects of data dependencies and delay constraints into a single importance metric. Recently the rate-distortion optimized approach is further extended to the scenario of proxy driven streaming by Chakareski, Chou and Girod [32].

Wyner-Ziv Scalability. Based on the Wyner-Ziv framework [34], several side information based video codecs have been proposed in the recent literature [35.36]. These can be thought of as an intermediate step between closing the prediction loop and coding independently. The closed-loop prediction (CLP) approach requires the exact value of the predictor to create the residue, whereas Wyner-Ziv coding (WZC) only requires the correlation structure between the current signal and the predictor, so there is no need to incorporate a decoder at the encoder. Some of the recent work addresses the problem of scalable coding in this setting. Sehgal *et al* [37] provided a theoretical approach by constructing several redundant Wyner-Ziv descriptions targeted at different fidelities. Here, based on the encoder's knowledge of the reconstruction status of previous samples at the decoder, a decision is made about which of those descriptions to send. Xu and Xiong [38] proposed an MPEG-4 FGS-like scheme by treating a standard coded video as a base layer, and building the bit-plane enhancement layers using Wyner-Ziv coding with current base and lower layers as side information. Steinberg and Merhav [39] formulated the theoretical problem of successive refinement of information, originally proposed by Equitz and Cover [40], in a Wyner-Ziv setting. Here we propose a practical scalable coder for the same setting as that of [39]. Our approach supports embedded representation and high coding efficiency by using the high quality version of the previous frame as side information in the enhancement-layer coding of the current frame.

7. Plan for the Next Year

As a future work, we are working on the technique, which selects motion search window size on the original resolution based on scene characteristics.

We will continue to study MDLC and extend it to consider more general MDC approaches. In particular we believe that there is a general class of systems where the data stored has significant redundancy, but where based on the state of the channel only a small fraction of the data is transmitted. Our current work on MDLC is just an initial example of these systems. Additional work will be devoted to exploring improvements to our media scheduling algorithms, where the target is to incorporate specific knowledge about congestion in the network, and aim to achieve "TCP friendly" scheduling.

We will adapt the proposed Wyner-Ziv scalable coder in the real video scenarios, and implement the practical Slepian-Wolf coder for this setting. It will be compared with the standard MPEG-4 FGS method.

8. Expected Milestones and Deliverables

2003-2004
- Improvements to scalable WZ coder
- Optimization of MDLC technique

2004-2006
- Implementation of Scalable WZ taking as a starting point a MPEG-4 FGS system
- Test of MDLC in real network environment

2006-2008
- Integrated robust media delivery platform, incorporating MDLC, FEC, adaptive scheduling and caching.

9. Member Company Benefits

N/A

10. References

[1] A. A. El-Gamal and T. M. Cover, "Achievable rates for multiple descriptions", IEEE Trans. Information Theory, vol. IT-28, no. 6, pp. 851-857, Nov. 1982.
[2] V. K. Goyal and J. Kovacevic, "Optimal multiple description transform coding of Gaussian vectors", in Proc. of IEEE Data Compression Conference, 1998.

[3] V. Goyal and J. Kovacevic, "Method description for transform coding using optimal transforms of arbitrary dimensions," US Patent, February 25, 1998. Filed.

[4] V. Vaishampayan, "Design of Multiple description Scalar Quantizers," IEEE Trans. Comm., vol. 39, no. 3, pp. 821-834, Nov. 1993.

[5] M. T. Orchard, Y. Wang, V. Vaishampayan, and A. R. Reibman, "Redundancy rate-distortion analysis of multiple description codingusing pairwise correlating transforms", in Proc. ICIP'97.

[6] S. D. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple descriptions wavelet based image coding", in Proc. ICIP'98.

[7] D.G. Sachs, R. Anand and K. Ramchandran, “Wireless image transmission using multiple-description based concatenated codes,” in Proc. IVCP 2000

[8] R. Puri and K. Ramchandran, “Multiple Description coding using forward error correction codes,” Proc. Asilomar Conf. on Signals and Sys., Pacific Grove, CA, Oct. 1999.

[9] A. Mohr, E. Riskin and R. Ladner, “Graceful degradation over packet erasure channels through forward error correction,” in Proc. DC 1999.

[10] W. Jiang and A. Ortega, “Multiple Description Coding via Polyphase Transform and Selective Quantization”, in Proc. VCIP’ 99.

[11] P. Sagetong and A. Ortega, “Optimal bit allocation for channel-adaptive multiple description coding”, in Proc. of IVCP 2000.

[12] R. Singh, and A. Ortega, “Multiple Description Coding Bases Erasure Recovery for State based Compression Scheme”, in Proc. MMSP’99.

[13] R. Singh, A. Ortega, L. Perret, and W. Jiang, “Comparison of Multiple Description Coding with Layered Coding using Network Simulator”, in Proc. IVCP 2000.

[14] P. Sagetong and A. Ortega, “Analytical model-based bit allocation for wavelet coding with applications to Multiple Description Coding and Region Of Interest coding,” IEEE International Conference on Multimedia and Expo (ICME 2001), Aug 2001.

[15] P. Sagetong and A. Ortega, “Rate-distortion model and analytical bit allocation for wavelet-based region of interest coding,” submitted to IEEE International Conference on Image Processing (ICIP 2002).

[16] D. Comas, R. Singh and A. Ortega “Rate-distortion optimization in a robust video transmission based on multiple description coding”, in Proc. MMSP 2001.

[17] R. Singh and A. Ortega, “A Robust DPCM scheme based on Multiple Description Coding”, accepted for publication in special issue of Journal of VLSI Signal Processing.

[18] H. Chung and A. Ortega, “Low Complexity Motion Estimation Algorithm by Multiresolution Search for Long-term Memory Motion Compensation”, in *Proc. IEEE Int. Conf. Image Processing*, Rochester, NY, Sep., 2002.

[19] *JVT Reference Software, version JM50c*, http:// bs.hhi.de/ ~suehring/ tml/ download/ .

[20] T. Wiegand (ed.), “Joint Working Draft, version 2 (WD-2)”, Joint Video Team (JVT) of ISO/IECMPEG and ITU-T VCEG, JVT-B118r2, Mar., 2002.

[21] T. Wiegand and X. Zhang and B. Girod, “Long-term memory motion-compensated prediction”, *IEEE Trans. Circuits Syst. Video Technol.*, vol.9, no.1, pp. 70-84, Feb., 1999.

[22] G. J. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression”, *IEEE Signal Processing Magazine*, vol.15, pp. 74-90, Nov., 1998.

[23] M. Podolsky, S. McCanne and M. Vetterli, “Soft ARQ for Layered Streaming Media”, Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology. Special Issue on Multimedia Signal Processing, Kluwer Academic Publishers, 2001, to appear

[24] C. Papadopoulos and G. Parulkar, “Retransmission-Based Error Control for Continuous Media Applications”, in *Proc. NOSSDAV*, April, 1996.
[25] M. Lucas, B. Dempsey and A. Weaver, “MESH: distributed error recovery for multimedia streams in wide-area multicast networks”, in *Proc. IEEE Int. Conf. on Commun.*, vol. 2, pp 1127-32, June, 1997, Montreal, Que.
[26] H. Radha, Y. Chen, K. Parthasarathy and R. Cohen, “Scalable Internet video using {MPEG}-4”, Signal Processing: Image Communication, 15 p.p. 95-126, 1999.
[27] Z. Miao and A. Ortega “Optimal Scheduling for Streaming of Scalable Media”, in *34^{th} Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November, 2000.
[28] H. Wang and A. Ortega, “Robust video communication by combining scalability and multiple description coding techniques”, Image and Video Communications and Processing 2003 Conf., EI'2003, Jan. 2003.
[29] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," IEEE Transactions on Multimedia, Feb. 2001.Submitted.
[30] Z. Miao and A. Ortega, " Expected run-time distortion based scheduling for delivery of scalable media ", Packet Video Workshop 2002, Pittsburgh, PA, Jan. 2002.
[31] Z. Miao and A. Ortega," Fast adaptive media scheduling based on expected run-time distortion ", Proc. of Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA, Nov. 2002.
[32] J. Chakareski, P. A. Chou, and B. Girod, "Rate-distortion optimized streaming from the edge of the network," IEEE Workshop on Multimedia Signal Processing, St. Thomas, US Virgin Islands, Dec. 2002.
[33] H. Wang and A. Ortega, “Scalable predictive coding by nested quantization with layered side information”, submitted to IEEE International Conference on Image Processing, 2004.
[34] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder", IEEE Transactions on Information Theory, Vol. IT-22, No. 1, Jan. 1976.
[35] R. Puri and K. Ramchandran, “A video coding architecture based on distributed compression principles”, in UC Berkeley/ERL Technical Report, Mar. 2003.
[36] A. Aaron, R. Zhang, and B. Girod, “Wyner-Ziv coding of motion video”, in Proc. Asilomar Conf. Signals and Systems, Nov. 2002.
[37] A. Sehgal, A. Jagmohan, and N. Ahuja, “Scalable predictive coding as the Wyner-Ziv Problem”, in The 8th Int. Conf. Commun. Systems, Nov. 2002, Vol.1, pp. 101-106.
[38] Q. Xu and Z. Xiong, “Layered Wyner-Ziv video coding”, in Proc. VCIP’04, Jan. 2004.
[39] Y. Steinberg and N. Merhav, “On successive refinement for the Wyner-Ziv problem”, Submitted to IEEE Trans. Inform. Theory, Mar. 2003.
[40] W.H. Equitz and T.M. Cover, “Successive refinement of information”, IEEE Trans. Inform. Theory, vol. 37, pp. 269-275, Mar. 1991.
[41] G. Reynolds, F. Glazer, and P. Anandan, “Scene matching by hierarchical correlation”, in Proc. IEEE Computer Vision and Pattern Recognition Conf., pp. 432--441, Washington, DC, Jun. 1983.
[42] P. J. Burt, “Multiresolution techniques for image representation, analysis, and `smart' transmission”, in Proc. SPIE conf. Visual Commun. Image Processing}, pp. 2--15, Philadelphia, PA, Nov. 1989.
[43] M. Bierling and R. Thoma, “Motion compensating field interpolation using a hierarchically structured displacement estimator”, Signal Processing, pp.387--404, Dec. 1986.

[44] M. Bierling, "Displacement estimation by hierarchical blockmatching", in Proc. SPIE conf. Visual Commun. Image Processing, pp. 942--951, Boston, Nov. 1988.
[45] J. W. Woods and T. Naveen, "Subband encoding of video sequences", in Proc. SPIE conf. Visual Commun. Image Processing, pp. 724--732, Nov. 1989.
[46] K. M. Uz, M. Vetterli, and D. J. LeGall, "Interpolative multiresolution coding of advanced television with compatible subchannels", IEEE Trans. Circuits Syst. Video Technol., vol 1, no.1, pp. 86--99, Mar. 1991.\
[47] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression", IEEE Trans. Circuits Syst. Video Technol., vol. 2, no. 3, pp. 285--296, Sep. 1992.
[48] X. Lin J. Li and Y. Wu, "Multiresolution tree architecture with its application in video sequence coding: A new result", in Proc. SPIE conf. Visual Commun. Image Processing, pp. 730--741, 1993.
[49] T. Naveen and J. W. Woods, "Motion compensated multiresolution transmission of high definition video", IEEE Trans. Circuits Syst. Video Technol., vol. 4, no. 1, pp. 29--41, Feb. 1994.
[50] V. Bhaskaran and K. Konstantibides, Image and Video Compression Standards: Algorithm and Architectures, Kluwer Academic Publishers, Boston, MA, 1995.
[51] H. Chung , D. Romacho and A. Ortega, "Fast long-term motion estimation for h.264 using multiresolution search", in Proc. IEEE Int. Conf. Image Processing}, vol. 3, pp. 905--908, Barcelona, Spain, Sep. 2003.
[52] D. Romacho, Fast Search Algorithms for Long-Term Memory Motion Compensation: Master Thesis, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA, 2003.