MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

A Study of Encoding and Decoding Techniques for Syndrome-Based Video Coding

Min Wu, Anthony Vetro, Jonathan Yedidia, Huifang Sun, Chang Wen Chen

TR2005-038 May 2005

Abstract

In conventional video coding, the complexity of an encoder is generally much higher than that of a decoder because of operations such as motion estimation consume significant computational resources. Such codec architecture is suitable for downlink transmission model of broadcast. However, in the contemporary applications of mobile wireless video uplink transmission, it is desirable to have low complexity video encoder to meet the resource limitations on the mobile devices. Recent advances in distributed video source coding provide potential reverse in computational complexity for encoder and decoder 1, 2. Various studies are done within a context of a scheme that is similar in spirit to that in 2. In this paper, we conduct a study of encoding and decoding techniques of syndrome-based video encoding scheme for mobile wireless applications. The innovation of this study is the adoption of low resolution low quality reference frames for motion estimation at the decoder. To accomplish the goal of reducing computational complexity while maintaining good reconstruction quality at the decoder, we investigate the following related strategies: 1. compression of low resolution and low quality sequence; 2. rate-distortion study with complexity constraints at the encoder; 3. building enhanced evidences for syndrome decoder. Extensive exerimental results have confirmed the effectiveness these techniques for syndrome based video coding.

IEEE International Symposium on Circuits and Systems (ISCAS)

A Study of Encoding and Decoding Techniques for Syndrome-Based Video Coding

Min Wu
Department of ECE
University of Missouri-Columbia
Columbia, MO 65211
Email: mw463@mizzou.edu

Anthony Vetro, Jonathan S. Yedidia, Huifang Sun
Mitsubishi Electric Research
Laboratories
Cambridge, MA 02139
Email: {avetro,yedidia,hsun}@merl.com

Chang Wen Chen
Department of ECE
Florida Institute of Technology
Melbourne, FL 32901
Email: cchen@fit.edu

***Abstract*— In conventional video coding, the complexity of an encoder is generally much higher than that of a decoder because of operations such as motion estimation consume significant computational resources. Such codec architecture is suitable for downlink transmission model of broadcast. However, in the contemporary applications of mobile wireless video uplink transmission, it is desirable to have low complexity video encoder to meet the resource limitations on the mobile devices. Recent advances in distributed video source coding provide potential reverse in computational complexity for encoder and decoder [1], [2]. Various studies are done within a context of a scheme that is similar in spirit to that in [2]. In this paper, we conduct a study of encoding and decoding techniques of syndrome-based video encoding scheme for mobile wireless applications. The innovation of this study is the adoption of low resolution low quality reference frames for motion estimation at the decoder. To accomplish the goal of reducing computational complexity while maintaining good reconstruction quality at the decoder, we investigate the following related strategies: (1) compression of low resolution and low quality sequence; (2) rate-distortion study with complexity constraints at the encoder; (3) building enhanced evidences for syndrome decoder. Extensive experimental results have confirmed the effectiveness these techniques for syndrome based video coding.**

I. Introduction

In current video coding standards, the complexity of an encoder is generally much higher than that of a decoder because some encoding specific components, such as motion estimation, have computationally intensive operations even when efficient fast motion search is used [3]. Such architecture is suitable for downlink transmission model of TV broadcast when the system has few encoders and numerous decoders. In contemporary media-rich uplink wireless video transmission, for instance, a video camera cell phone transmits video wirelessly to the base station, complexity is of paramount concern because battery-powered mobile handheld devices usually have limited processing power and memory. Therefore, it is desirable to have a low complexity video encoder to meet the resource limitations.

Recently, distributed video coding schemes have been proposed to provide potential reverse in computational complexity for decoder and encoder [1], [2]. The theoretical background of these schemes is based on Slepian-Wolf and Wyner-Ziv distributed source coding theories [4], [5]. The Slepian-Wolf theory states that two statistically dependent discrete signals, compressed by two independent encoders but decoded by a joint decoder, can achieve the same rate as they are compressed and decoded jointly even if the encoders are independent. The counterpart of this theorem for lossy source coding is Wyner-Ziv theory on source coding. Wyner-Ziv theory states that for a Gaussian random signal and its side information, the conditional Rate Mean Squared Error Distortion function for this signal is the same as the case when its side information is available only at the decoder.

Video coding schemes built on the distributed source coding principles proposed recently have a general architecture. The encoder applies error correcting codes to each frame and generate syndrome bits [1], [2]. The decoder estimates each frame. Such estimate can be viewed as a noisy version of the original frame. The error correcting coder combines the syndrome bits and the noisy version to reconstruct each frame. The less the estimation error, the less the syndrome bits are needed.

The expected rate-distortion results are between the results obtained by running inter mode and intra mode of standard video compression, but still far from optimal. In [1], error correcting coding is applied to each block on a frame, in order to not exceed the correcting ability, the coefficients are very coarsely quantized, and refine bits are needed to be sent after distributed coding. In this case, only 20% bits are distributed coded, the rest of the bits are entropy coded. In [2], each frame estimated by the interpolation of previous frames is not very accurate, this will generate more syndrome bits to correct estimation errors. Therefore, building more accurate estimation for each frame will improve the overall coding efficiency.

There are various studies [6], [7] that have been conducted within the context of a reference scheme which is similar in spirit to that in [2]. In [6], Wyner-Ziv video coding is implemented in transform domain to improve the coding efficiency. In [7], hash codewords of the current frame are generated and sent to aid the decoder in accurately estimating the motion. The hash codewords are coarsely quantized version of a downsampled 8 by 8 image block. To save the bit rate, the distance between two hash codewords of co-located blocks on previous and current frame is calculated to decide whether or not to send the codewords. In our study, we use highly compressed version of each frame as reference to build

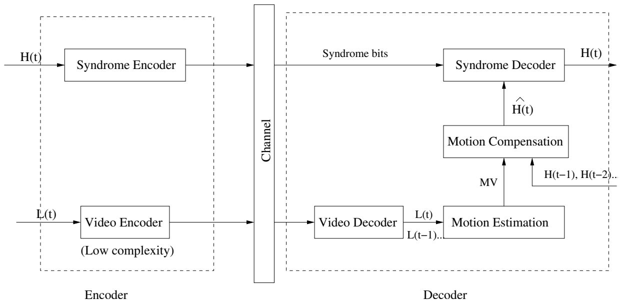

Fig. 1. Architecture of syndrome-based video source codec

more accurate estimation. Although we add some cost for compression and transmission of those low quality frames, the overall bit rate can be saved because more syndrome bits are saved by accurate estimation.

The rest of paper is organized as follows: In section II, we describe our system and investigate the following strategies: (1) compression of low resolution and low quality sequence; (2) rate-distortion study with complexity constraints at the encoder; (3) building enhanced evidences for syndrome decoder. In section III, extensive experimental results are given. In section IV, we provide a summary and some discussions.

II. Syndrome-based Video Source Coding

A. Syndrome-Based Coding Architecture

Figure 1 shows the architecture of this system. At the encoder, for each frame of the video, two sets of bits are transmitted to the decoder. The first set contains syndrome bits generated in syndrome encoding. The second set is highly compressed version of the current frame. At the decoder, we use the low quality frame as reference to perform motion search and compensation to predict the current high quality frame. The syndrome decoder combines the syndrome bits and the prediction to reconstruct the current high quality frame.

B. Encoder Side

1) Rate-distortion Study with Complexity Constraints at the Encoder: There are many choices to compress the low quality sequence. The state-of-art image/video coding standards include JPEG2000, MPEG2, H.263, and H.264. Since the encoder has limitation on processing power, the selection is based on the coding efficiency and computational complexity.

We select H.264 AVC to compress the low quality sequences because it has high coding efficiency and low complexity, comparing with other standards [8]. The new H.264 video coding standard provides a compression gain of 1.5 to 2 folds over previous standards, such as H.263+ and MPEG-4 Part 2. Unlike the popular 8×8 discrete cosine transform used in previous standards, the 4×4 transforms can be computed without multiplications, just additions and shifts, in 16-bit arithmetic, thus the computational complexity can be reduced, especially for the low-end processors.

Full motion search is a very efficiency way for video compression, since there is the processing power constraint at encoder, we do not perform full motion search in low quality sequence compression. However, Ting et al investigated the distribution characteristic of motion vectors [9]. They observed that motion vectors tends to concentrated in the previous frame, no matter the motion is slow or high. They also observed that most motion vectors are distributed around the search window, which means zero motion vector is a good estimate of full motion search. Thus, coding the difference of the current frame and previous frame (zero motion mode) is the best choice to compress low quality sequence under complexity constraint. In our scheme, we use H.264 IPPP zero motion mode to compress the low quality frames.

2) Compression of Low Resolution of Low Quality Frames: At low bit rate, few bits per pixel are allocated to each block of high resolution image. Such image has strong blocking artifacts. To further improve the compression at low bit rate, we compress the down-sampled version of the video sequence [10]. For those down-sampled images, more bits per pixel are allocated for each block. Such image displays more details, and has less blocking artifacts. Moreover, as we up-sample image at decoding side, since interpolation can further blur the blocking effect, we add another improvement to the process. Thus the down-sample approach has better performance both subjectively and objectively.

The down-sampling factors are not limited as integer. They could be fractional numbers, which make down/up-sampling much more complicated. An approach to determine the optimal down-sampling factor has been proposed [10]. However, due to the limitation on processing power of mobile devices, we constrain the down-sampling factor the integer power of 2.

C. Decoder Side

At the decoder side, we want to predict the current frame as accurate as possible to save the number of syndrome bits. Interpolation of previous frames does not provide very accurate prediction. We need to use some information from current low quality frame to improve the prediction. The low quality frames are used as reference to perform motion search to build motion compensation of the current frame. The syndrome decoder combines the prediction version with syndrome bits to reconstruct the current frame.

1) Building Enhanced Evidences for Syndrome Decoder: Motion search can be performed on current low quality frame and previous low quality frames (L-L), or on current low quality frame and previous reconstructed high quality frames (L-H). Those low quality frames at low bit rate have blocking artifacts. The blocking artifacts will disturb block matching when comparing with blocks in high quality previous frame. However, successive frames probably have the same blocking effect. The effect of blocking artifacts could be nullified by the similar blocking artifacts. Therefore, at very low bit rate, using L-L motion search has smaller prediction error than using L-H motion search.

When the difference between the low and high quality sequences is large, motion compensation is not accurate in prediction. Adding a constraint on motion search for very low

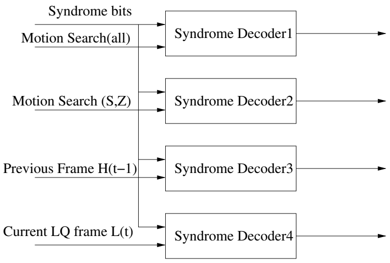

Fig. 2. Frame adaptive syndrome decoder

quality sequence will improve prediction. When the difference is small, motion compensation becomes accurate, and adding a constraint on motion search will increase the prediction error. Thus a weight factor is placed on the constraint. The cost function is then:

$$SSD + \lambda(Q, PSNR\_Diff)(MV)^2 \tag{1}$$

where, SSD is sum squared difference between the blocks in current frame and matching block in previous frame, and MV is motion vectors. λ is a empirical function of quantization parameter and PSNR difference between low quality and target high quality sequences.

2) Frame Adaptive Decoding for Syndrome Decoder: Motion compensation on each block may not be the optimal choice for prediction. For those high motion frames, the temporal correlation in successive frames might be low. Intra mode compression could be better than motion search exploiting temporal redundancy. Since the low quality frames are compressed by zero motion inter mode. We have the mode information for each block on inter frame. Each block in inter mode frame could possible be $Skip$, $Motion\ Compensation$, or $Intra$ mode. However, since the mode is chosen from zero motion mode and other modes. We still cannot guarantee that this mixed mode is optimal. If the complexity is not a major concern in decoder side, we could build parallel decoding procedures, and let syndrome decoder determine which one use the least number of syndrome bits.

- Motion search for all blocks in inter frame. Use motion vectors and previous high reconstructed high quality frames to build motion compensated prediction.
- Motion search for $Skip$, and $Zero\ Motion$ blocks. Use motion vectors and previous reconstructed high quality frames to build those blocks. For those block with $Intra$ mode, use co-located block in the current low quality frame as prediction.
- Input previous reconstructed high frame as prediction.
- Input current low quality frame as prediction.

The architecture of decoding procedure is shown in figure 2. We feed above four choices of evidence into difference syndrome decoders. The four syndrome decoders are running simultaneously. If all syndrome decoders need more syndrome bits to reconstruct current frame, feedback information will be sent to syndrome encoder to ask for more bits. Whichever syndrome decoder successfully reconstructs the current frame, the decoding of current frame stops and all syndrome decoders continue to decoder the next frame.

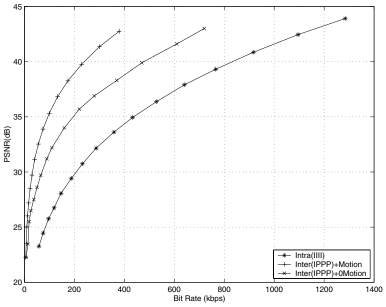

Fig. 3. Zero motion performance on Carphone QCIF

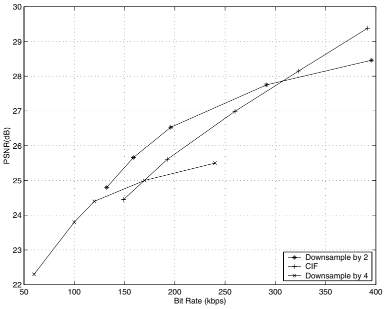

Fig. 4. Performance of downsampling on Foreman CIF

III. Experimental Results

In this section, we present the experimental result to confirm the effectiveness of the proposed techniques in syndrome-based encoding/decoding presented in Section II.

A. Encoder Side

We use Carphone and Foreman sequences in our test. The frame rate is 30 fps. First we test the performance of IPPP zero motion mode. Figure 3 shows the result on Carphone sequence. The result on Foreman is similar to that on Carphone. We observe that at low PSNR range (less than 30), zero motion mode is closer to IPPP full motion search mode. We also observe that zero motion has higher coding efficiency on slow motion sequence than that on high motion sequence. This phenomenon is reasonable. In slow motion sequence, the successive pictures are highly correlated in temporal domain. The motion vectors are more condensed on 0. In contrast, there is much less temporal correlation in fast motion sequences. The motion vectors are less condensed on 0.

Then we test the down-sampling approach in low quality sequence compression. Figure 4 show the result on Foreman sequence. The image is compressed with intra (IIII) mode. We test the image down-sampled by factor of 2, and 4. From the result, we observe that when we compress image lower than 24 dB, 4 is the best choice for down-sampling factor; when we compress image between 24 and 28 dB, 2 is the best choice; when we compress image higher than 28 dB, down-sampling provide no benefit in compression.

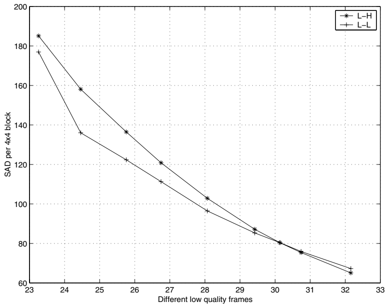

Fig. 5. Prediction error of L-L & L-H motion search for Carphone QCIF

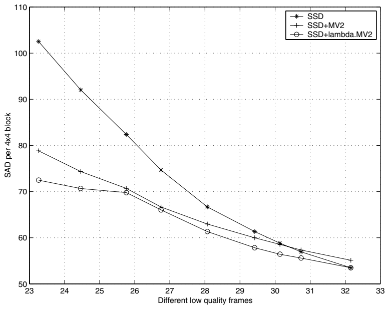

Fig. 6. Impact of adaptive constrained motion search on prediction

B. Decoder Side

First we select motion search mode. We perform L-L and L-H motion search, then we use the estimated motion vectors and high quality previous frame to build motion compensated prediction. Figure 5 shows the prediction error of Carphone sequence. We observed that when the PSNR of low quality sequences are less than 30 dB, using L-L motion search has smaller prediction error than that of L-H motion search. Since our application is mobile video transmission, the target PSNR is about 30 dB, the low quality frame should be lower than 30 dB, thus we choose L-L motion search in our system.

Then we test the impact of adaptive constrained motion search on prediction. Figure 6 shows the result on Carphone QCIF sequence. Experimental result shows that the adaptive result is always better than the motion constraint without adaptation. We also test the prediction error with these four different evidences in frame adaptive syndrome decoding. Figure 7 shows the experimental results on Foreman CIF sequence. Frame adaptive is always better than others.

IV. Summary and Discussion

In this paper we conduct a study of encoding and decoding techniques for syndrome-based video coding scheme for mobile wireless applications. We use very low quality frames as reference to build more accurate estimate of each frame. In the compression of low resolution and low quality sequence, we conclude that mixed zero motion compression mode is optimal trade off with processing power constraint and that down-sampling can further improve rate distortion when compression at very low bit rate.

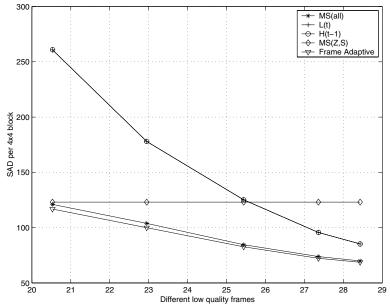

Fig. 7. Prediction error of four difference kind of evidences on Foreman CIF

We also investigated issues in building enhanced evidences for syndrome decoder. At the decoder, for those low quality frames, low to low motion search can achieve better prediction for current high quality frame. Adaptive constrained motion search tends to improve prediction. Frame adaptive reconstruction has the best result with high complexity. Extensive simulations show that the proposed syndrome based video coding techniques are suitable for mobile wireless applications.

It should be clear that better coding efficiency can be obtained using the techniques proposed in this paper. The proposed encoding techniques provide the decoder with the best reference to perform motion estimation given the low bit-rate and low complexity constraints. Furthermore, the proposed decoding techniques allow for reconstruction of the video at the target quality using fewer syndrome bits. Complete coding results will be reported in a future paper.

References

[1] R. Puri and K. Ramchandran, "PRISM: a new robust video coding based on distributed compression principles," in *Proc. of Conference on Communication, Control, and Computing*, Allenton, IL, October 2002.

[2] A. Aaron and B. Girod, "Towards practical wyner ziv coding of video," in *Proc. of IEEE International Conference on Image Processing*, Barcelona, Spain, September 2003.

[3] Y. Liu and S. Oraintara, "Complexity comparison of fast block-matching motion estimation algorithms," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2004, pp. 17–21.

[4] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transaction on Information Theory*, vol. 19, pp. 471–490, July 1973.

[5] D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transaction on Information Theory*, vol. 22, pp. 1–10, January 1976.

[6] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform domain wyner-ziv codec for video," in *in Proc. SPIE Visual Communications and Image Processing*, San Jose, CA, January 2004.

[7] A. Aaron, S. Rane, and B. Girod, "Wyner-ziv video coding with hash-based motion compensation at the receiver," in *in Proc. IEEE International Conference on Image Processing*, Singapore, October 2004.

[8] H. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in h.264/AVC," *IEEE Transaction on Circuits and System for Video Technology*, vol. 13, pp. 598–603, July 2003.

[9] C. W. Ting, L. M. Po, and C. H. Cheung, "Center-biased frame selection algorithms for fast multi-frame motion estimation in h.264," in *Proc. of IEEE International Conference on Neural Networks and Signal Processing*, Nanjing, China, December 2003, pp. 1258–1261.

[10] A. M. Bruckstein, M. Elad, and R. Kimmel, "Down-scaling for better transform compression," *IEEE Transactions on Image Processing*, vol. 12, pp. 1132–1144, September 2003.