INFORMATION THEORY INSPIRED VIDEO CODING METHODS : TRUTH IS SOMETIMES BETTER THAN FICTION

Nitin Khanna, Fengqing Zhu, Marc Bosch, Meilin Yang, Mary Comer and Edward J. Delp

Video and Image Processing Lab (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

ABSTRACT

In recent years there has been a growing interest in developing novel techniques for increasing coding efficiency and error resilience of video compression methods. These developments are often based on the concepts from information theory. In this paper, we describe an overview of some of these recent advances including distributed video coding, scalable coding, content-based methods, and multiple description coding.

1. INTRODUCTION

With the popularity of new digital video applications such real-time video streaming, IP-TV, and peer-to-peer (P2P) video, it has become important to improve coding efficiency (data rate), video quality, and robustness to errors. Shannon eloquently indicated that many of these issues can be addressed by the "use" of appropriate redundancies in the source [1]. However, any redundancy in the source will usually help if and only if it is utilized at the decoder. A typical video signal usually has a great deal of both spatial and temporal redundancy. Modern video codecs have been developed to remove the redundant information that the video decoder does not need to decode the sequence. Conventional hybrid video codecs have succeeded in increasing video quality while reducing data rate [2, 3]. With the commercial success of video coding and transmission, the video coding research community has been investigating next-generation services [4, 2, 5]. The new functionalities in video coding provide higher robustness, interactivity and scalability. In current video coding standards, spatial and temporal correlations are exploited at the encoder which leads to an encoder structure with high complexity. This satisfies the needs of applications such as video streaming, broadcast systems, and consumer applications such as DVDs. However, for applications such as wireless sensor networks, wireless PC cameras, and mobile video cameras for video surveillance, high complexity encoders are not feasible. Thus, a new approach known as distributed video coding [6] was proposed in recent years. This is based on the information theoretic work of coding with side information of Slepian and Wolf[7] and Wyner and Ziv[8]. In distributed coding, source statistics are exploited at the decoder so that it is feasible to design a simplified encoder.

When digital video is transmitted over variable bandwidth channels, stored on media of different capacity, and displayed on variety of devices, a method of adapting the digital video to the needs of the user becomes inevitable. Scalable video coding provides the flexible adaptation to accommodate such a goal [9]. The bit stream is divided into a base layer and one or more enhancement layers. The base layer provides fundamental information of the sequence and each enhancement layer can be added to improve the quality incrementally.

One interesting approach to reduce the data rate is to use methods based on scene content where a different coding method is used in various regions of the scene. In early video coding systems this was achieved by either reducing the size of the frame and/or a combination of frame skipping. Some content-based video coders introduced new coding techniques that focused on the semantic meanings of objects represented in the video sequence [10]. Using these semantic meanings average data rates lower than those achievable by modern codecs can be obtained by coding different regions at different rates. In particular, regions belonging to areas whose specific details are not perceived by the viewer, could be skipped or encoded at a much lower data rate.

In addition to reducing the data rate, robust error resilience is also important due to the high demand of applications in scalable, multicast and P2P environments [11]. In particular, Multiple Description Coding (MDC) has been proposed as an effective solution to combat error-prone channels, especially for real-time applications when retransmission is unacceptable [12, 13, 14]. MDC generates multiple, equally important descriptions so that each description can be decoded independently with an acceptable decoding quality. The decoding quality is improved when more descriptions are received. A variety of MDC algorithms have been proposed [12, 14].

This work was partially supported by the endowment of the Charles William Harrison distinguished professorship at Purdue University and by a grant from Cisco Systems, Inc.

2. DISTRIBUTED VIDEO CODING

Distributed video coding uses side information at the decoder to achieve lower complexity at the encoder [6, 15, 16, 17]. The theoretical basis for the problem dates back to work done in the 1970s. Slepian and Wolf addressed lossless compression of multiple correlated sources [7]. They described a lossless encoding scheme without side information at the encoder that performed as well as encoding with side information at the encoder. Wyner and Ziv extended this result to establish rate-distortion bounds for the lossy coding case [8]. Therefore, low complexity distributed video encoding methods are sometimes referred to as Wyner-Ziv video encoding. In a typiical Wyner-Ziv video codec the input video sequence is divided into two groups which are coded separately. Half of the frames could coded using a H.264 intraframe encoder, which are denoted as key frames. Between each key frame, one frame is independently encoded as a Wyner-Ziv frame (WZ frame). Use of "intra" key frames maintains low complexity at the encoder and also generates side information for the Wyner-Ziv frames at the decoder [6]. The side information can be extracted from the neighboring key frames by extrapolation or interpolation. Two channel coding methods have been used, turbo codes and low- density-parity-check (LDPC) codes. To reconstruct the Wyner-Ziv frames, the decoder first derives the side information from the previously decoded key frames. Side information is an initial estimate or noisy version of the current frame.

In [18] we introduced a rate distortion analysis of motion side estimation in Wyner-Ziv video coding as a model to examine its performance. Wyner-Ziv video coding is compared with two conventional Motion-Compensated Prediction (MCP) based video coding methods, i.e., DPCM-frame video coding and INTER-frame video coding. DPCM-frame coding subtracts the previous reconstructed frame from the current frame and codes the difference. INTER-frame coding performs motion search at the encoder and codes the residual frame. We have shown that Wyner-Ziv coding can achieve a gain up to 6 dB (for small motion vector variance) or 1-2 dB (for normal to large motion vector variance) over DPCM-frame video coding. On the other hand, INTER-frame coding outperforms Wyner-Ziv video coding by around 6 dB. Current Wyner-Ziv video coding schemes still fall far behind the state-of-the-art video codecs. A better motion estimator at the decoder is essential to improve the performance.

To address the coding efficiency of the Wyner-Ziv codec we proposed a backward-channel-aware motion estimation (BCAME) to code the key frames [15, 16] . The basic idea of BCAME is to perform motion estimation at the decoder and send the motion information back to the encoder through a backward channel. We refer to these backward predictively coded frames as BP frames. We observed that BP frames can significantly improve the coding efficiency with minimal usage of backward channel.

3. SCALABLE VIDEO CODING

When digital video is transmitted over variable bandwidth channels, stored on media of different capacity, and displayed on variety of devices, a method of adapting the digital video to the needs of the user becomes inevitable. Scalable video coding provides the flexible adaptation to accommodate such a goal [19]. The rate distortion analysis of scalable video coding is extensively studied in [9, 20, 21]. Because of the inherent scalability in wavelet transform, wavelet based video coding structure is used for scalability. A fully rate scalable video codec, referred to as SAMCoW (Scalable Adaptive Motion Compensated Wavelet) was proposed in [22]. The design based on hybrid video coding was standardized first in MPEG-2 [19]. Many other approaches have been proposed [23, 24] and new ideas are described below.

Temporal scalable video coding can be generated by a hierarchical structure, which allows the decoding at several frame rates for a bitstream. Therefore, it can significantly improve the coding efficiency over the single layer coding. The hierarchical structure shows an improved coding efficiency especially with cascading quantization parameters, where the base layer is encoded with the highest fidelity following lower quality coded enhancement layers (Figure 1).

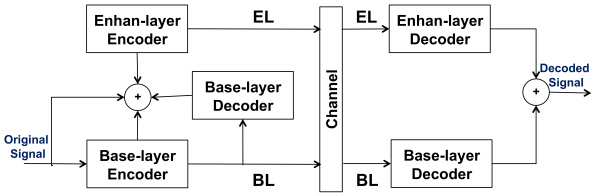

Figure 1. Layered Video Coding.

In spatial scalability, each layer corresponds to some specific resolution. Besides the prediction for single-layer coding, spatial scalable video coding also uses the inter-layer correlation to achieve coding efficiency higher than simulcast coding. The inter-layer prediction can use the information from lower layers as the reference. This ensures that a set of layers can be decoded independent of all higher layers. In [25], an efficient inter-layer motion-compensation technique is proposed for enhancement layer in spatially scalable video coding. The lower bound on the rate distortion performance of subband motion compensation technique, which uses both the inter-layer spatial and intra-layer temporal redundancies for video spatial scalability, is studied in [26].

Coarse-granular scalability (CGS) and fine-granular scalability (FGS) are two popular concepts used in the design of SNR scalable coding. In CGS, SNR scalability is achieved by using similar inter-layer prediction techniques as spatial scalability [19]. FGS coding allows truncating and decoding a bitstream at any point with bit-plane coding. Progressive refinement (PR) slices are used in FGS to achieve full SNR scalability over a wide range

of rate-distortion points. A motion compensation (MC) scheme for FGS scalable video coding is proposed in [27], where two MC loops are used, one for the base layer and one for the enhancement layer. A technique called conditional replacement (CR), which adaptively selects between the base layer and enhancement layer prediction for each enhancement layer DCT coefficient, is used to simultaneously improve coding efficiency and reduce prediction drift.

4. CONTENT-BASED VIDEO CODING

The concept of dividing an image into textures and edges was introduced in 1959, known as Synthetic Highs [28]. The method described two approaches to encode each type of structure in an image. The goal is to determine where "insignificant" texture regions or "detail-irrelevant" regions in the frame are located and then use texture models for the pixels in these regions. "Insignificant" texture regions refers to those regions in a frame where changes are unnoticeable by the observer. An example of this approach is shown in Figure 2, where the black area in the frame on the right is not transmitted but reconstructed by the decoder from a texture region in a previous frame shown on the left.

Figure 2. General Approach: Texture and Motion Models Used to Construct a Frame.

Texture-based video coding method identifies homogeneous regions in a frame and labels them as textures [29]. Temporal consistency of the identified textures is ensured throughout the video sequence by using global motion models to warp texture regions from frame-to-frame. A set of motion parameters for each texture region is sent to the decoder as side information. The sequence is encoded using conventional video codec, e.g. H.264, with synthesizable regions labeled as skip macroblocks. At the decoder, frames are partially reconstructed except for the synthesizable parts which are inserted later using side information and key frames. We have examined spatial texture models that are based on features, such as color and edges, or based on statistical methods like Gray Level Co-occurrence Matrix (GLCM), and transform methods like Gabor filterbanks.After the feature extraction, segmentation such as split and merge or K-means clustering is performed to divide the frame into different regions based on their properties.

Similar to texture-based method, motion-based video coding uses the motion detection properties of Human Visual System [30]. Motion in a video sequence can be produced by static objects with a moving background due to camera motion, objects moving in different trajectories, or simply randomly moving objects. In these scenarios, a viewer gives priority to track fast and unpredictable motion objects. We call this type of motion "noticeable motion" which is different than slow or predictable motion which we call "non-noticeable motion". We developed our video codec assuming that for background or non-noticeable motion objects the viewer perceives just the semantic meaning of the objects, holding his/her attention to only noticeable objects. The implementation of this video codec is similar to the texture based approach, in the sense that we use a foreground-background extraction as the motion analyzer instead of the texture analyzer and we synthesize "non-noticeable motion" regions instead of textures.

We are able to show on average 15% improvement for data rate savings compare to sequences coded using H.264/AVC only. Because of the way these coders work, metrics such as PSNR are not suitable for measuring the visual quality. Perceptual quality user experiments were conducted to assess the impact the content-based methods on perceived quality [31], i.e. the trade-offs between visual quality and data rate savings. Results from these experiments indicate that the loss in quality using our video coding methods is acceptable.

5. MULTIPLE DESCRIPTION CODING

One of the most difficult problems in video transmission is communication over error-prone channels, especially when retransmission is unacceptable. Multiple Description Coding (MDC) has been proposed as an effective solution to address this problem, due to its robust error resilience [12]. A variety of MDC algorithms have been proposed in recent years [12, 14]. In [14], three classes of MDC methods have been defined based on the predictor type. A four-description MDC which takes advantage of residual-pixel correlation in the spatial domain and coefficient correlation in the frequency domain is presented in [32]. However, this method often induces mismatch. Recently, we developed a new four-description MDC utilizing a hybrid architecture of both temporal and spatial correlations for error concealment [33]. It is optimal for high-motion sequences by using spatial concealment as the default method, and temporal concealment as a secondary method.

Due to the demand of applications in scalable, multicast and P2P environments, it is advantageous to use more than two descriptions. We developed a four-description MDC utilizing a hybrid architecture of both temporal and spatial correlations for error concealment [33]. Figure 3 shows the architecture of the method. The original video sequence is first split into two temporally correlated descriptions, which are called the Even and Odd sequences respectively. Each of the new sequences are then separated into two spatially correlated sequences through horizontal downsampling. The newly generated four descriptions are encoded independently. This is a Class A method [14], which has good mismatch control but loses certain coding efficiency. When both the spatially correlated descriptions

of odd frames or even frames are lost, temporal correlation is used for error concealment; otherwise, spatial correlation is always used to recover the lost data.

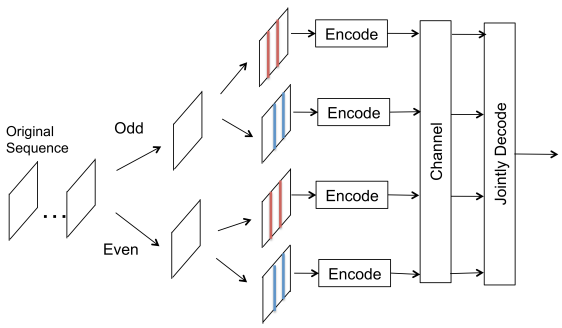

Figure 3. Architecture of Four-Description MDC.

The experimental results show that our method [33] provided a graceful degradation with the decrease in the number of descriptions received. Even when three descriptions were totally lost, our method could still achieve an acceptable quality. It was superior to two-description MDC, when clients could only receive as much as one fourth of the data due to bandwidth limitation or network congestion. For packet loss performance, experimental results showed that when packet loss rate (P_B) increased, PSNR of our method degraded more slowly than both types of two-description MDC.

6. CONCLUSIONS

This paper presented an overview of ongoing video coding research that is inspired by information theory. There are still a lot of challenges in video coding and many interesting approaches to be investigated [2]. The manta that "video coding is dead" is fiction. We need to be information theory archaeologist and combine methods that inspired by Shannon with the end user and application in mind.

7. REFERENCES

[1] C. E. Shannon, "A mathematical theory of communications," *The Bell System Technical Journal*, vol. 27:379-423, pp. 623–656, October 1948.

[2] G. J. Sullivan, J.-R. Ohm, A. Ortega, E. J. Delp, A. Vetro, and M. Barni, "dsp forum - future of video coding and transmission," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 76–82, Nov. 2006.

[3] G. J. Sullivan and T. Wiegend, "Video compression: From concepts to the H.264/AVC standard," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 18– 31, Jan. 2005.

[4] T. Sikora, "Trends and perspectives in image and video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 6– 17, Jan. 2005.

[5] L. Liu, F. Zhu, M. Bosch, and E. Delp, "Recent advances in video compression: What's next?" *Proceedings of the International Symposium on Signal Processing and its Applications*, Sharjah, United Arab Emirates, February 2007, pp. 1 –8.

[6] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71– 83, January 2005.

[7] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[8] 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. 22, no. 1, pp. 1–10, January 1976.

[9] G. Cook, J. Prades-Nebot, Y. Liu, and E. Delp, "Rate-distortion analysis of motion-compensated rate scalable video," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2170 – 2190, August 2006.

[10] M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second-generation image-coding techniques," *Proceedings of the IEEE*, vol. 73, no. 4, pp. 549–574, April 1985.

[11] L. Liang, S. Paul, and E. J. Delp, "Unequal error protection techniques based on wyner-ziv coding," *EURASIP Journal on Image and Video Processing*, vol. 2009, 2009.

[12] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, September 2001.

[13] C. Zhu and M. Liu, "Multiple description video coding based on hierarchical b pictures," *IEEE Transactions on Circuits Systems Video Technology*, vol. 19, no. 4, pp. 511–521, April 2009.

[14] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceeding of the IEEE*, vol. 93, no. 1, pp. 57–70, January 2005.

[15] L. Liu, Z. Li, and E. Delp, "Backward channel aware wyner-ziv video coding: A study of complexity, rate, and distortion tradeoff," *Signal Processing: Image Communication*, vol. 23, no. 5, pp. 353–368, 2008.

[16] ——, "Efficient and low-complexity surveillance video compression using backward-channel aware wyner–ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 4, pp. 453–465, 2009.

[17] Z. Li, L. Liu, and E. J. Delp, "Wyner-Ziv video coding with universal prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 11, pp. 1430–1436, November 2006.

[18] Z. Li, L. Liu, and E. Delp, "Rate distortion analysis of motion side estimation in wyner–ziv video coding," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 98–113, January 2007.

[19] J.-R. Ohm, "Advances in scalable video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 42– 56, Jan. 2005.

[20] J. Prades-Nebot, G. W. Cook, and E. J. Delp, "An analysis of the efficiency of different SNR-scalable strategies for video coders," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 848– 864, April 2006.

[21] Y. Liu, P. Salama, Z. Li, and E. Delp, "An enhancement of leaky prediction layered video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 11, pp. 1317– 1331, November 2005.

[22] K. Shen and E. J. Delp, "Wavelet based rate scalable video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 109 – 122, February 1999.

[23] L. Overturf, M. Comer, and E. Delp, "Color image coding using morphological pyramid decomposition," *IEEE Transactions on Image Processing*, vol. 4, no. 2, pp. 177 –185, February 1995.

[24] K.-L. Hua, I. Pollak, and M. Comer, "Optimal tilings for image and video compression," *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, October 2006, pp. 391 –395.

[25] R. Zhang and M. Comer, "Efficient inter-layer motion compensation for spatially scalable video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 10, pp. 1325 –1334, oct. 2008.

[26] Z. Rong and M. Comer, "Rate distortion analysis of subband motion compensation for video spatial scalability," *Proceedings of the Picture Coding Symposium (PCS)*, Chicago, Illinois, 6-8 2009, pp. 1 –4.

[27] M. Comer, "Conditional replacement for improved coding efficiency in fine-grain scalable video coding," *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, Rochester, New York, September 2002, pp. II–57 – II–60.

[28] W. F. Schreiber, C. F. Knapp, and N. D. Kay, "Synthetic highs, an experimental tv bandwidth reduction system," *Journal of Society of Motion Picture and Television Engineers*, vol. 68, pp. 525–537, August 1959.

[29] M. Bosch, F. Zhu, and E. Delp, "Spatial texture models for video compression," *Proceedings of the IEEE International Conference on Image Processing*, San Antonio, Texas, September 2007, pp. 93–96.

[30] ——, "Video coding using motion classification," *Proceedings of the IEEE International Conference on Image Processing*, San Diego, California, October 2008, pp. 1588–1591.

[31] ——, "Perceptual quality evaluation for texture and motion based video coding," *Proceedings of the IEEE International Conference on Image Processing*, Cairo, Egypt, November 2009, pp. 2261–2264.

[32] C. Hsiao and W. Tsai, "Hybrid multiple description coding based on h.264," *IEEE Transactions on Circuits Systems Video Technology*, vol. 20, no. 1, pp. 76–87, January 2010.

[33] M. Yang, M. Comer, and E. Delp, "A four-description mdc for high loss-rate channels," *submitted to the Proceedings of the 28th Picture Coding Symposium*, Nagoya, Japan, December 7-10 2010.