WAVELET VIDEO CODING

Jens-Rainer Ohm

Chair and Institute of Communications Engineering, RWTH Aachen University

ABSTRACT

Scalability at the bitstream level is an important feature for encoded video that is to be transmitted and stored with a variety of target rates or to be replayed on devices with different capabilities and resolutions. Conventional high-performance video compression schemes are based on the method of motion-compensated prediction, using a recursive loop in the prediction process. Due to this recursion and the inherent drift in cases of deviation between encoder and decoder states, scalability is difficult to realize and typically causes a penalty in compression performance. The method of interframe wavelet coding overcomes this limitation by replacing the prediction along the time axis by a wavelet filter, which can nevertheless be operated in combination with motion compensation. Recent advances in motion-compensated temporal filtering (MCTF) have proven that combination with arbitrary motion compensation methods is possible. Compression performance is achieved that is comparable with state of the art coders, or even better in some cases.

1. INTRODUCTION

Motion pictures must often be transmitted over variable bandwidth channels, both in wireless and cable networks. They have to be stored on media of different capacity, such as memory cards and high-capacity DVD; they have to be replayed on a variety of devices, ranging from small mobile terminals to high-resolution projection systems. Scalable video coding schemes are intended to encode the signal once at highest resolution, but enable decoding from partial streams depending on the specific rate and resolution required by a certain application. This enables a simple and flexible solution for transmission over heterogeneous networks, additionally providing adaptability for bandwidth variations and error control. It enables both multicast and unicast streaming applications with minimal processing at the server or in the network, and low-complexity decoding. It further allows simple adaptation for a variety of storage devices and terminals.

For video coding, a lack of efficiency can however be observed in combining scalable coding with the popular approach of hybrid motion-compensated prediction and block transform encoding, as it is implemented in most of today's standards. Therefore, research for more efficient scalable coding techniques is still a demanding area in video compression. Recent breakthroughs in motion-compensated temporal wavelet filtering have finally given a realistic perspective to implement highly efficient scalable video codecs.

This talk will highlight the principles of wavelet based video coding schemes. It presents a review of interframe wavelet video coding methods, including classification and detailed presentation of some of the motion-compensated wavelet coding schemes proposed recently.

2. THE PRINCIPLE OF MOTION-COMPENSATED TEMPORAL FILTERING

Application of motion compensation (MC) is often regarded to be implicitly coupled with frame prediction schemes. There is indeed no justification for this restriction, as MC can rather be interpreted as a method to align a filtering operation with a motion trajectory along the temporal axis. In the case of MC prediction, the filters are in principle LPC analysis and synthesis filters, while in cases of transform or wavelet coding, transform basis functions are subject to MC alignment. This is denoted as Motion-compensated Temporal Filtering (MCTF). If MCTF is used in combination with a 2D spatial wavelet transform, this is often denoted as 3D, or (depending on the sequence of the spatial and temporal processing) either as a 2D+t or t+2D wavelet transform.

The first approaches of MCTF date back to the early '90s and were mostly based on application of 2-tap (Haar) filters along the temporal axis. This could best be interpreted as generation of one average (Lowpass=L) and one difference (Highpass=H) frame from pairs of original frames. When the lowpass filtering is iterated in the fashion of a wavelet pyramid, the LLL.. frame at the end level would correspond to the (motion-compensated) average over a large number of frames. The real problem therein is to cope with spatially variant and inhomogeneous motion vector fields. Only recently, the combination of lifting filters with motion compensation has brought up a solution which allows perfect reconstruction in any case, even if arbitrary methods of MC are applied in MCTF. The lifting process consists of a sequence of so-called prediction and update steps, which are performed in reverse order during synthesis. As the prediction and update filters process samples from different frames, combination with motion compensation is easily done.

The H frames resulting from a prediction filter output are at first sight very similar to standard MC prediction frames; in principle, both uni-directional and bi-directional frame prediction can be applied, corresponding to the concepts of P and B frames in well-known standard coders. The basic difference is however to generate these prediction frames not over a (possibly infinite) time series, but rather over a (finite) wavelet tree, which makes the whole concept non-recursive. This property is the main advantage of the application of MCTF instead of the well-known standard concepts, as the decoder becomes non-recursive, such that in fact no prediction loop exists.

3. PERSPECTIVES

New perspectives in video compression are enforced by these recent advances in MCTF. The non-recursive structure of MCTF based encoders provides high flexibility in bitstream scalability for different temporal, spatial and quality resolutions and better error resilience than conventional (prediction based) coders. In fact, due to the MCTF process the coded representation inherently provides new capabilities to better separate relevant and irrelevant parts of the information in a motion picture sequence. The lowpass frames highlight those information parts of a movie which are consistent over a large number of frames, establishing a means for powerful exploitation of multiple-frame redundancies as hardly achievable by conventional frame-to-frame or multi-frame prediction methods.

Due to the non-recursive structure, higher degrees of freedom are possible both for encoder and decoder optimization. In principle, a decoder can integrate additional signal synthesis elements whenever the received information is incomplete, such as frame-rate up-conversion, film-grain noise overlay or other elements of texture and motion synthesis. Such methods could easily be integrated as a part of the MCTF synthesis process; as no prediction loop exists, synchronization loss between encoder and decoder is not a severe problem. From this point of view, even though in the lifting interpretation many elements of MCTF can be regarded as extensions of proven techniques from MC prediction based coders, this framework exhibits and enables a number of radically new options in video encoding, which could finally initiate a breakthrough of analysis/synthesis based approaches.

On the other hand, when a wavelet transform is applied for the encoding of the L and H frames as resulting from the MCTF process, the commonalities with 2D wavelet coding methods are obvious. If the sequence of spatial and temporal filtering is exchanged (2D+t instead of t+2D wavelet transform), MCTF could also be interpreted as a framework for further interframe compression of (intra-frame restricted) 2D wavelet representations such as JPEG 2000. From this point of view, a link between the previously separate worlds of 2D wavelet coding with their excellent scalability properties and compression-efficient motion-compensated video coding schemes is established by MCTF. This shows the high potential for future developments in the area of motion picture compression, even allowing seamless transition between intraframe and interframe coding methods, depending on the application requirements for flexible random access, scalability, high compression and error resilience. Furthermore, scalable protection of content, allowing access management for different resolution qualities of video signals, is a natural companion of scalable compression methods.

In summary, these new wavelet codecs provide numerous advantages over non-scalable conventional techniques based on motion-compensated prediction:

- No recursive predictive loop as in the current standards (MPEG-x, H.26x), such that no drift occurs if decoding is performed at various bit-rates and resolutions;
- Separation of noise and sampling artefacts by usage of longer temporal filters;
- Flexible exploitation of long range as well as short range temporal redundancies;
- Adaptability in the spatial and temporal filtering methods, number of decomposition levels, and filter choices, which makes improvements possible that are not feasible in predictive coding.

As a consequence, wavelet video coding schemes can provide flexible spatial, temporal, SNR and complexity scalability with fine granularity over a wide range of bit rates, while maintaining a very high coding efficiency. Compared to state of the art single layer coders such as MPEG-4 Advanced Video Coding aka ITU-T H.264, methods of interframe wavelet coding have been shown to perform typically superior in sequences with quiet motion and high spatial detail, while for other types of sequences comparable performance is achieved.

MPEG's recent Call for Proposals for new highly-efficient scalable video coding technology, and the current plans to develop such a scalable video framework as part of the MPEG-21 standard, reflect the situation of new advances made in the field. Even though it is premature to predict the technical perspectives of such a new standardization effort which will run for more than two years from now, it is well possible that the interframe wavelet technologies described in this talk, or similar technology developed from this ground, could become one of the key players in future video standardization.