BIT-DEPTH SCALABLE VIDEO CODING

Martin Winken, Detlev Marpe, Heiko Schwarz, and Thomas Wiegand

Fraunhofer Institute for Telecommunications – Heinrich Hertz Institute, Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany, [winken|marpe|hschwarz|wiegand]@hhi.fraunhofer.de

ABSTRACT

Scalable video coding (SVC) is an extension of H.264/AVC, which is currently being developed by the Joint Video Team. SVC supports any combination of temporal, spatial, and SNR scalability, but there is no special treatment for the case that two or more different bit-depths have to be supported by the scalable bit-stream. There is a number of applications (especially in the HD area) which would benefit from a design where an H.264/AVC conforming base layer with a sample bit-depth of typically 8 bit is extended in a backwards compatible way by an higher bit-depth enhancement layer. In this paper we present new techniques which can be used to achieve bit-depth scalability and show experimental results.

Index Terms— Video coding, Scalable video coding

1. INTRODUCTION

Scalable video coding (SVC) is a current standardization project of the Joint Video Team (JVT) of ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG). The current SVC draft [1] supports scalability in terms of SNR and/or spatial resolution, but up to now, there is no technique for processing two different bit-depths as part of a scalable bit-stream. One possible application for bit-depth scalability is coding of a video sequence in a way such that "legacy" H.264/AVC [2] decoders will be able to decode the 8 bit version of the video sequence, which is extended through a 10 bit enhancement signal for newer, bit-depth scalability-aware decoders. Since modern multimedia interfaces (HDMI 1.3) allow transmission of digital video data with up to 16 bits per sample and consumer displays already operate with a bit-depth of 10 bits or higher, it is desirable to provide a backwards-compatible way of coding high bit-depth video data. More detailed application examples regarding 10 bit DVD authoring and digital motion picture production are described in a JVT proposal from Thomson [3].

In this paper we present an approach how bit-depth scalability can be achieved as a simple enhancement to the current SVC draft with little increase in computational complexity at the decoder. Section 2 gives a description of the approach, Sec. 3 introduces the concept of inverse tone mapping, and Sec. 4 shows experimental results.

2. DESCRIPTION OF THE APPROACH

The current SVC draft allows scalability in terms of picture quality (SNR) and spatial resolution. The most obvious way to achieve scalability in terms of sample bit-depth within SVC would be to allow two different so-called coarse-grain SNR scalability (CGS) layers to support different bit-depths. Since using CGS generally requires only one motion-compensated prediction (MCP) loop for decoding, this would require performing MCP and especially the half-pel interpolation filtering on the high bit-depth data. In this paper we follow another approach, since from practical hardware implementation considerations, it is desirable to use instead only 8 bit data for MCP.

The basic architecture of our approach is illustrated in Fig. 1 (encoder parts for base and enhancement layer are separated by a dashed line). In our approach of bit-depth scalability, the high bit-depth input video signal is first down-converted to a low bit-depth version, which is encoded using single layer H.264/AVC. This low bit-depth video sequence builds the base layer. The tone mapping from high to low bit-depth image data can be of any kind and depends on the characteristics (e.g., dynamic range, pdf) of the high bit-depth data or on the specific target application area. The decoded low bit-depth representation is used to generate a prediction for the high bit-depth video signal. The way how this prediction process (so-called "inverse tone mapping") has to be performed depends on the chosen tone mapping scheme and is signaled as side information to the decoder. More details on this process will be given in the next section. The high bit-depth enhancement layer contains only a "texture refinement signal", which is the transform-coded difference signal between the original high bit-depth input data and its prediction from the base layer. Note, that using this approach MCP has to be performed only for the low bit-depth base layer. We therefore introduced a new flag in the slice header which signals whether the corresponding slice may contain motion data (as in CGS or spatial scalability) or if it is a "texture refinement only" slice for providing bit-depth scalability. In any

case, decoding is still possible with only one single prediction loop.

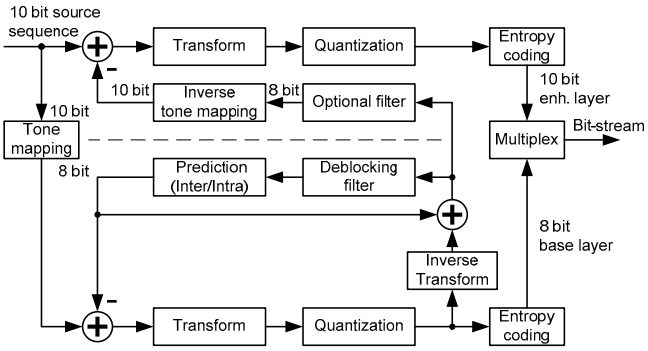

Fig. 1: Basic architecture of our approach.

3. TONE MAPPING AND ITS INVERSE

The term *tone mapping* specifies a method of mapping high bit-depth sample values to a lower bit-depth representation. There is a variety of ways how this mapping can be done. Generally, it depends on the characteristics of the source which mapping scheme is appropriate. In the simplest case, the least significant bits are just omitted, corresponding to a linear scaling operation. This would be appropriate if the high bit-depth signal covers the same dynamic range as the corresponding low bit-depth signal, but with a smaller size of the quantization interval. In other cases, for example, when the high bit-depth signal covers a larger dynamic range than the corresponding low bit-depth signal, non-linear mapping operations might be used. An example for the effect of using different tone mapping schemes for converting the luma component of the 10 bit Waves (Viper) sequence to an 8 bit range can be seen in Fig. 2. .For an overview of tone mapping schemes that are useful for video coding see [4].

In our approach to bit-depth scalability, the decoded low bit-depth signal is used to obtain a prediction for the high bit-depth signal. This prediction process has to be adapted to the tone mapping scheme that was used to generate the low bit-depth representation from the high bit-depth signal and is called *inverse tone mapping* in the following[1]. Since performing this process is required for decoding the high bit-depth signal, computational complexity is here a critical issue. We therefore have considered three simple variants for inverse tone mapping:

- **linear scaling** and clipping of the sample values x of the base-quality layer according to

$$\min(2^{M-N}x, 2^M - 1),$$

where the sample values x of the base-quality layer are represented with a bit-depth of N and the sample values of the high-quality enhancement layer are represented with a bit-depth of M with $M > N$.

- **linear interpolation** using an arbitrary number of interpolation points: For a low bit-depth sample with value x and two given interpolation points (x_n, y_n) and (x_{n+1}, y_{n+1}) the corresponding prediction sample y with a bit-depth of M is obtained according to the following formula for $x_n \le x \le x_{n+1}$:

$$y = \min\left(y_n + \frac{x - x_n}{x_{n+1} - x_n}(y_{n+1} - y_n), 2^M - 1 \right).$$

This linear interpolation can be performed with low computational complexity by using only bit shift instead of division operations if $x_{n+1} - x_n$ is restricted to be a power of two.

- **look-up table mapping:** for each possible low bit-depth sample value the corresponding high bit-depth sample value is specified.

Linear scaling is the default method to be used, if no other inverse tone mapping scheme has been specified. It requires very little computational complexity, since only bit shift and clipping operations are involved. Linear interpolation allows specification of a tone mapping scheme using a piece-wise linear approximation. Note that the value of $x_{n+1} - x_n$ is restricted to be a power of two in order to avoid division operations for obtaining interpolated sample values. Look-up table mapping allows the definition of arbitrary inverse tone mapping functions.

3.1. Signaling of the inverse tone mapping scheme

Since the statistical characteristics of the pictures may vary over the sequence, it is desirable to have the possibility to use a different tone mapping scheme for each picture. In our approach the inverse tone mapping parameters are first transmitted in the sequence parameter set (SPS) for the whole sequence and may be replaced by definition of a new inverse tone mapping scheme in the picture parameter set (PPS). This allows a high flexibility and requires only a small overhead of side information. Furthermore, it is possible to specify different inverse tone mapping schemes for the luma and the chroma components, if the sequence is coded using a corresponding color space.

[1] Note that this process cannot be an inverse in a rigorous mathematical sense since there is an obvious loss of information in the corresponding (forward) tone mapping process.

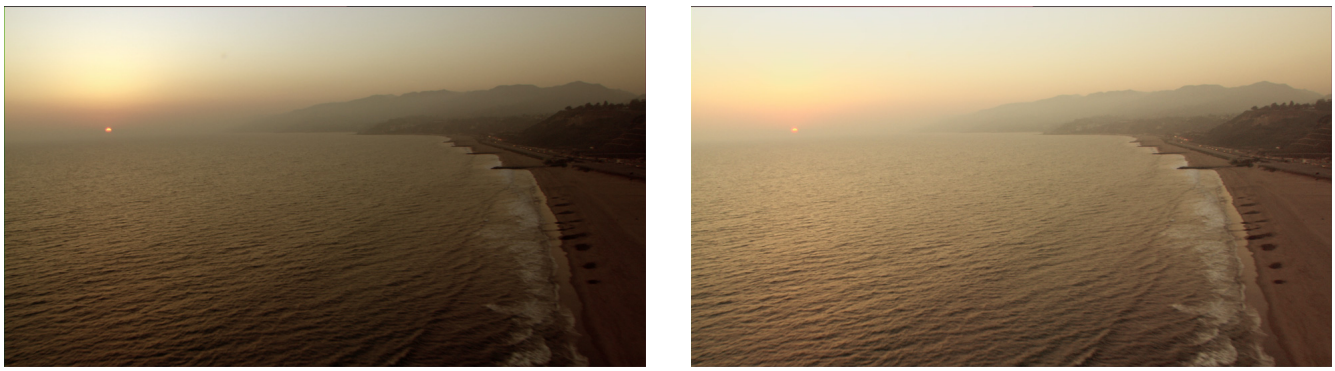

Fig. 2: Example for different luma tone mapping schemes (left: linear, right: according to [5]).

4. EXPERIMENTAL RESULTS

For our experiments, we extended the Joint Scalable Video Model (JSVM) reference encoder software to allow coding of video sequences with more than 8 bit per sample. The used sequences are part of the Viper test set distributed by FastVDO. Note, that in order to reduce memory requirements during the encoding process, the sequences have been cropped to a resolution of 704×576 samples. The tone mapping of the 10 bit source data to the 8 bit domain has been performed using the algorithm described in [5] for the CapitolRecords and Night sequences and by linear scaling, corresponding to omitting the two least-significant bits, for the Plane sequence. For the inverse tone mapping from the reconstructed 8 bit base layer, a simple look-up mapping has been employed where for luma and chroma two independent lookup-tables have been used. The used look-up tables are sequence-dependent and have been obtained by minimizing the mean squared error between the inverse tone mapped, non-coded 8 bit samples resulting from the tone mapping operator and the original 10 bit samples. With the vector x of the 10 bit source samples and the tone mapping operator f, the used inverse tone mapping scheme f^{-1} results from solving:

$$\min_{f^{-1}} \| x - f^{-1}(f(x)) \|_2^2$$

Note, that in our experiments for all pictures of the sequence, the same inverse tone mapping scheme has been used. The used inverse luma tone mapping curve for the CapitolRecords sequence is shown in Fig. 3. The "clipping" of 8 bit values larger than 206 is due to the fact that the tone mapping operator does not use the full dynamic range of the 8 bit domain and therefore these values do not occur.

For coding of base and enhancement layer, always the same quantization parameter (QP) settings have been used. A GOP size of 16 pictures was chosen and every 32nd picture has been intra-coded. As a reference, we show the performance of a 10 bit single layer H.264/AVC encoder as well as the resulting performance of coding the tone mapped 8 bit version of the sequence using a single-layer encoder where the reconstructed 8 bit signal has been inverse tone mapped to obtain a 10 bit signal using the same mapping scheme that has been employed for the scalable encoding process. The resulting plots are shown in Fig. 4. Note that the PSNR values are measured in the 10 bit sample domain. It can be seen that our scalable approach (red curves) clearly outperforms the variant using 8 bit data only (green curves) and nearly reaches the performance of a 10 bit single-layer coder (black curves) for the case using the rather complex tone mapping operator described in [5]. For the case of a linear tone mapping, as done for the Plane sequence, no gains to using the upscaled 8 bit base layer can be obtained by our scalable approach. For a further comparison, we also performed a simulation with a different inverse tone mapping scheme. Here, only a linear scaling of the reconstructed 8 bit base layer signal was allowed where the scaling factor was independently chosen for each 16×16 block of the sequence with an accuracy of one half. The resulting rate distortion performance is shown in the blue curves. It can be seen that using only a linear scaling operation is not sufficient in the case that a more complex tone mapping operator than linear scaling is used.

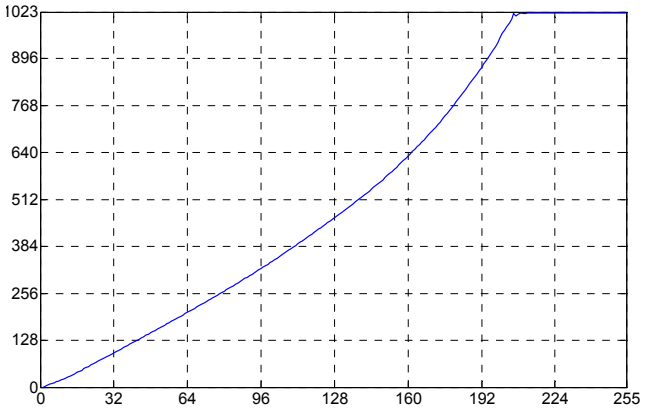

Fig. 3: Typical inverse luma tone mapping characteristic.

5. CONCLUSION

We have presented a new technique for providing bit-depth scalability as an enhancement to the current SVC standard. An important feature of the presented approach is that motion-compensated prediction is performed using only the low bit-depth data (typically 8 bit per sample), so there is no increase in complexity for this component at the decoder.

Furthermore, it is possible to specify an inverse tone mapping scheme which can be adapted to the characteristics of the sequence. Our experiments show that this is necessary if a more complex tone mapping mechanism than simple linear scaling by omitting the least-significant bits is used. This approach provides a highly flexible, low-complexity way of incorporating bit-depth scalability into the current SVC standard.

6. REFERENCES

[1] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, M. Wien, "Joint Draft 8 of SVC Amendment," *Joint Video Team*, Doc. JVT-U201, Hangzhou, China, October 2006.

[2] *ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC*, "Advanced Video Coding for Generic Audiovisual Services," Version 3, 2005.

[3] Y. Gao, Y. Wu, "Applications and Requirement for Color Bit Depth Scalability," *Joint Video Team*, Doc. JVT-U049, Hangzhou, China, October 2006.

[4] A. Segall, L: Kerofsky, S. Lei, "New Results with the Tone Mapping SEI Message," *Joint Video Team*, Doc. JVT-U041, Hangzhou, China, October 2006.

[5] E. Reinhard, M. Stark, P. Shirley, J. Ferwerda, "Photographic Tone Reproduction for Digital Images." ACM Transactions on Graphics, vol. 21(3): 267–276

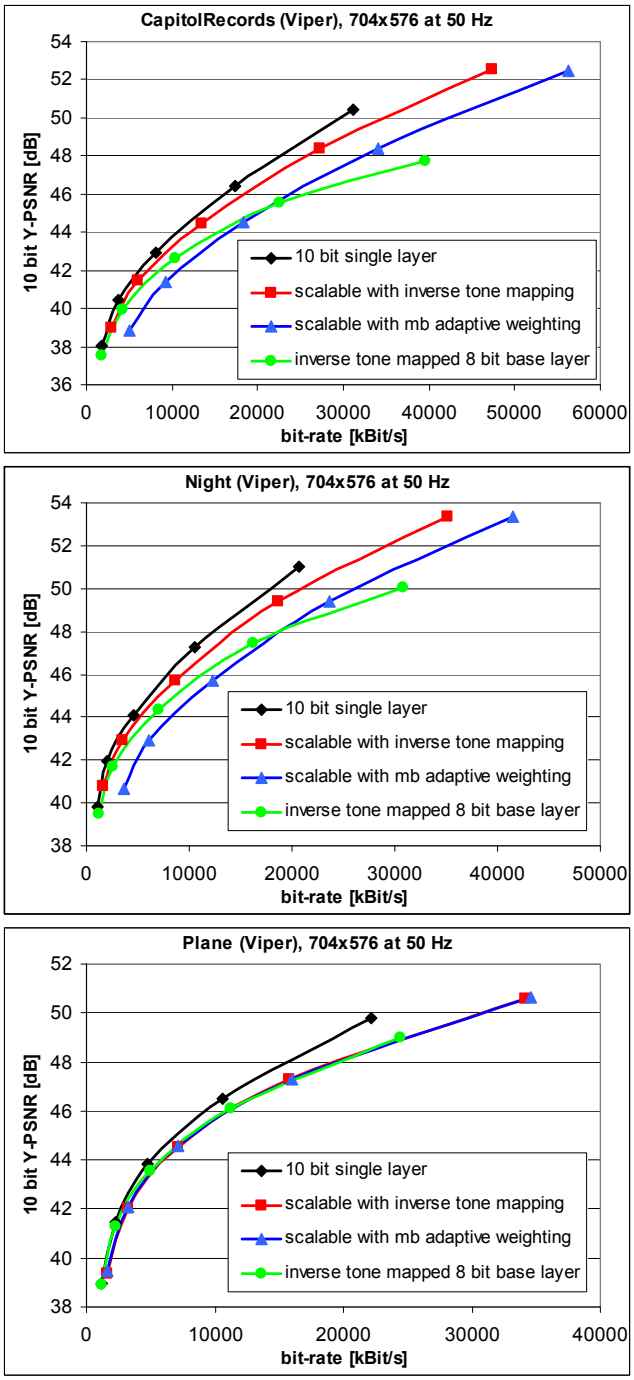

Fig. 4: Rate distortion performance of our approach using three different test sequences.