EXTENDED TEXTURE PREDICTION FOR H.264/AVC INTRA CODING

Johannes Ballé and Mathias Wien

RWTH Aachen University
Aachen, Germany
Email: {balle,wien}@ient.rwth-aachen.de

ABSTRACT

Efficient intra prediction is an important aspect of video coding with high compression efficiency. H.264/AVC applies directional prediction from neighboring pixels on an adjustable block size for local decorrelation. In this paper, we present an extended prediction scheme in the context of H.264/AVC that comprises two additional prediction methods exploiting self-similar properties of the encoded texture. A new macroblock type is implemented, allowing for flexible selection of the available prediction methods for sub-partitions of the macroblock. Depending on the content of the encoded video sequence, substantial gains in rate-distortion performance are achieved. The results may indicate directions towards an enhanced intra coding scheme with improved rate-distortion performance.

Index Terms— intra coding, markov models, H.264/AVC

1. INTRODUCTION

In H.264/AVC [1], intra prediction is used to locally decorrelate neighboring blocks. Key idea of the concept first proposed in [2] is extrapolation of the pixel row and column directly adjacent to the current block, as shown in Figure 1(a). Source pixel values are extrapolated to destination pixels along one of 8 spatial directions. Additionally, DC prediction is available: the average of all source pixel values is copied to all destination pixels. H.264/AVC intra coding is very efficient while being computationally inexpensive; however, it is not possible to predict more complex texture than uniform regions, or directional structures such as gradients or borders between uniform regions.

The rest of this paper is organized as follows. We present two additional prediction modes exploiting extended spatial correspondences beyond the scope of adjacent pixels in Sections 2 and 3. Section 4 describes a new macroblock type for H.264/AVC that allows flexible combination of all mentioned prediction techniques. The results are discussed in Section 5.

2. DISPLACEMENT INTRA PREDICTION

Considering that natural textures often contain recurring features, a simple approach to extend intra prediction is imitating motion compensation in intra coding. We simply employ block matching and copying *within* the encoded frame (Figure 1(b)). This technique has earlier been proposed and implemented in [3]. The proposal was limited to blocks of 16×16 pixels and re-used the H.264/AVC inter prediction framework for coding of displacement vectors. [4] extended the concept to variable block sizes and allowed a combination of the new prediction mode with the existing H.264/AVC Intra_4x4 prediction.

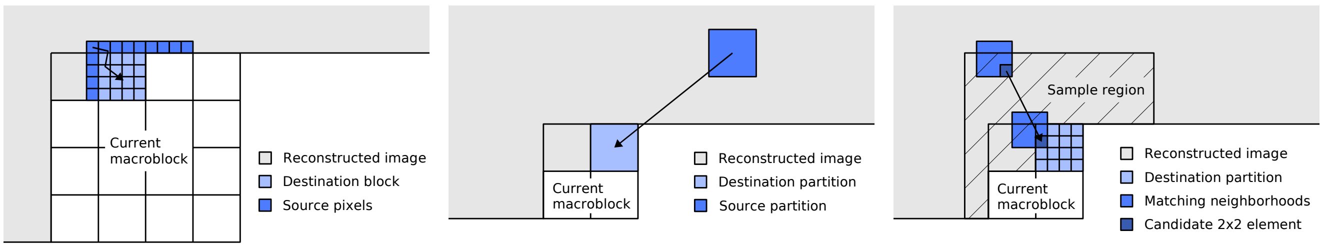
We enhanced the prediction mode using a nearest-neighbor extrapolation scheme for unreconstructed image regions, similar to *unrestricted motion vectors* introduced in H.263. Additionally, the displacement vectors are differentially encoded, analogously to the encoding of motion vectors. We tested some other vector prediction schemes; however, these resulted in no substantial gains in encoding performance while demanding a higher complexity. We re-use the 6-tap/bilinear filter combination known from H.264/AVC for quarter-pel accuracy. Finally, our modifications feature an optimized entropy coding scheme in order to better exploit the measured displacement vector characteristics.

3. MARKOVIAN TEXTURE PREDICTION

Markovian Texture Prediction is based on the assumption that Markov random fields (MRF) sufficiently model the properties of natural textures. This model has been used for texture synthesis. A particularly noteworthy algorithm is presented in [5]: It has comparably low complexity requirements compared to other algorithms of its purpose. The input consists merely of a texture sample and a randomized start configuration. From this, a texture patch with visually similar properties is generated. A well-known application of this is "filling up" the texture of an unknown region in an image.

In [6], a "template matching" prediction method was presented which uses a variation of this algorithm to "synthesize" a predictor for a partition from the surrounding reconstructed blocks. The authors embedded this mode into the H.264/AVC codec next to the Intra_4x4 and Intra_8x8 directional prediction modes specified in the standard. The concept is further evaluated and refined in [7].

The algorithm implemented in the context of our paper is based on the work presented in [5], [6], and [7]. It works as follows: The destination partition is divided in 2×2 pixel blocks which are processed in raster scan order (Figure 1(c)).



(a) H.264/AVC Intra Prediction (b) Displacement Intra Prediction (c) Markovian Texture Prediction

Fig. 1. Prediction modes. (a) shows the conventional Intra_4x4 prediction of H.264/AVC. (b) and (c) depict the application of the new prediction modes to a partition size of 8×8 pixels.

For each of these blocks, a neighborhood matching is performed in the rectangular image region consisting of the destination partition extended by 12 pixels in every direction, excluding the destination partition itself and any unreconstructed regions. The pixel values of the 5 candidate blocks with the most similar neighborhood in terms of SAD are combined using a weighted average. This average block is assigned to the 2×2 block in the destination region.

4. EXTENDED TEXTURE PREDICTION

In principle, the Markovian approach is suitable for prediction of a wide range of different textures. However, a large neighborhood size (or a multi-scale approach) is necessary for the algorithm to grasp distinguishable (large-scale) texture features. The neighborhood size again is heavily limited by complexity considerations, as the decoder needs to perform the same search as the encoder. In effect, MTP is limited to small-scale texture prediction (at least in the context of H.264/AVC).

DIP has no such limitations, but as it requires signalization of a displacement vector, it becomes more cost efficient as the block size increases. On the other hand, small block sizes are required for a good prediction quality. Hence, a trade-off between predictor block size and displacement vector quality has to be found.

One idea leading to the design described in the following is that both modes might compensate for each others weaknesses when both are available to the encoder at the same time. Interestingly, a similar concept was proposed for texture synthesis in [8]. The algorithm presented there uses block copying to preserve distinguishable features of the texture while minimizing complexity; Markovian synthesis is used to fill the “gaps.”

We extend H.264/AVC by an additional macroblock type, named ETP in the following, which denotes that the macroblock shall be predicted using DIP, MTP, a simplified subset of H.264/AVC intra prediction modes, or using a combination thereof. The subset of H.264/AVC modes comprises vertical (V), horizontal (H), and DC prediction. In order to introduce flexibility in the mode and block size decision, we apply a quad-tree partitioning scheme onto each macroblock,

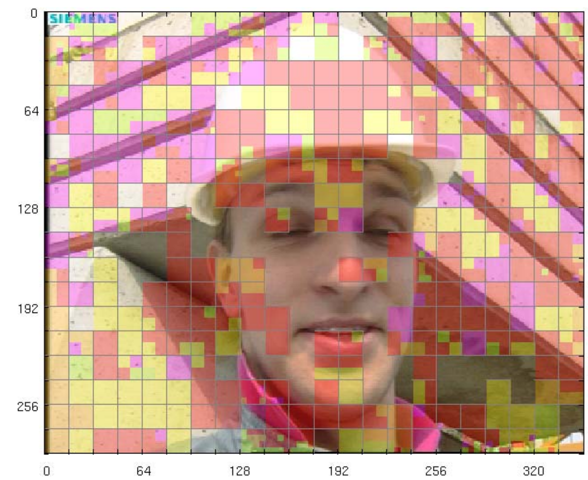


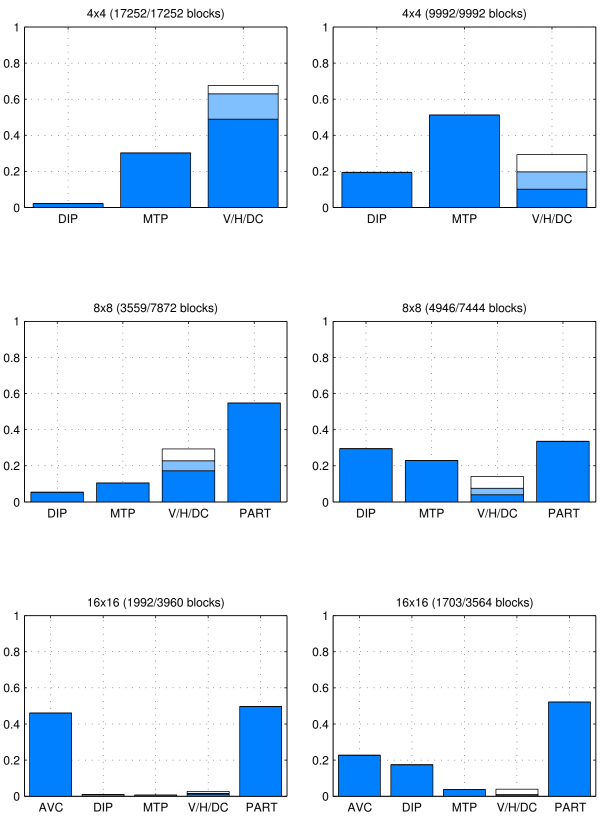
Fig. 2. Example for the applied prediction modes on the first frame of FOREMAN CIF at QP=32. Magenta: DIP; red: MTP; yellowish: V, H, DC; uncolored: conventional H.264/AVC intra macroblocks.

allowing block sizes from luma 16×16 down to 4×4. This allows mode decisions based on local signal statistics. In the context of the new mode, we aligned luma and chroma prediction as we expected the quality of the chroma signal to profit from DIP.

Since MTP and conventional prediction of submacroblock partitions relies on the final pixel values of the previously encoded partitions, the new macroblock type shares features of conventional intra prediction modes (immediate application of reconstructed residual) and conventional inter prediction modes (displacement vectors). The transform scheme used in H.264/AVC including Fidelity Range Extensions, however, was retained for the new macroblock type.

5. SETUP & RESULTS

We have integrated the presented extended prediction mode into I-slices and tested it in intra-only scenarios on progressive material. The coder was configured to use one slice per



(a) HARBOUR CIF (b) SPINCALENDAR CIF

Fig. 4. Mode distribution for selected sequences at QP=32.

Sequence	Res.	ΔPSNR [dB]	ΔRate [kbit/s]
BUS	CIF	0.347	–4.217
CITY	CIF	0.292	–4.358
CONCRETE	CIF	0.204	–1.575
CREW	CIF	0.217	–3.717
FOOTBALL	CIF	0.202	–2.835
FOREMAN	CIF	0.620	–9.946
HARBOUR	CIF	0.141	–1.707
MOBILE	CIF	0.342	–3.122
SHUTTLESTART	CIF	0.231	–4.382
SOCCER	CIF	0.246	–4.488
SPINCALENDAR	CIF	1.786	–22.606
TABLE	CIF	0.542	–8.107
Average		0.431	–5.922
CITY	4CIF	0.405	–6.096
SPINCALENDAR	4CIF	2.814	–33.931

Table 1. AVSNR results for the test set.

Fig. 5. Example for the prediction signal of H.264 (left) and H.264+ETP (right) for SPINCALENDAR 4CIF at QP=32.

frame. All sequences of the test set were temporally subsampled to 1 Hz. The rate-distortion performance was measured for QP = 22, 27, 32, 37 using constant QP encoding. Training of the CABAC contexts for the macroblock partitioning, the mode decision, and the displacement vectors was performed using a sequence compiled using a single frame of each of 51 different CIF sequences. In our current simulations, the deblocking filter is not enabled as adaptation to the new prediction modes was not trivial.

Our implementation is based on the open source encoder x264 [9] (SVN revision 553) and on FFmpeg [10] (SVN revision 6688) for decoding. x264 tends to perform slightly better than JM 11.0 on low QP settings while this is reversed for high QP settings. In the following RD plots, we include JM results for reference, while all other results are based on comparison to the original x264 revision.

Table 1 lists AVSNR results according to [11] for our test set. The average gain is in the range of 0.4 dB Bjøntegaard ΔPSNR or 5% Bjøntegaard rate savings, with a maximum for sequence FOREMAN with 0.62 dB and 9.9%. HARBOUR achieves the lowest gains. For SPINCALENDAR CIF, we observe remarkable gains of 1.79 dB ΔPSNR and 22.6% rate savings. For the 4CIF version, the gains further increase to 2.81 dB and 33.93%. RD plots for a selection of sequences are given in Figure 3.

The prediction mode selection is exemplary shown in Figure 2. Remarkably, macroblocks using simplified H.264/AVC intra prediction in the context of the new macroblock type are selected quite frequently. The reason for this might be reduced signaling overhead due to luma/chroma coupling combined with comparable prediction performance.

In Figure 4, the distribution of the selected modes for HARBOUR and SPINCALENDAR is provided. The figure is structured as follows. The sub-plot titles indicate the partition size and the number of encoded blocks of the respective size. On the y axis, the relative frequencies of the given modes are shown. In the lowest row, frequencies on macroblock level are given. AVC indicates the percentage of macroblocks encoded in conventional H.264/AVC modes. The frequencies for simplified H.264/AVC intra modes in the context of ETP are stacked for comparison. PART indicates the number of macroblocks that were partitioned further to 8×8 or 4×4 blocks. The same scheme applies to the 8×8 mode statistics in the middle row, where PART indicates the number of 8×8 blocks further partitioned to 4×4. Note that for all partition sizes, the achievable rate-distortion gains relate to the frequency the new modes are selected. Generally, our measurements indicate that on most sequences, MTP is selected more often than DIP, possibly due to its lower signalization overhead. However, in some cases DIP outperforms MTP.

To demonstrate how well DIP performs on certain tex-

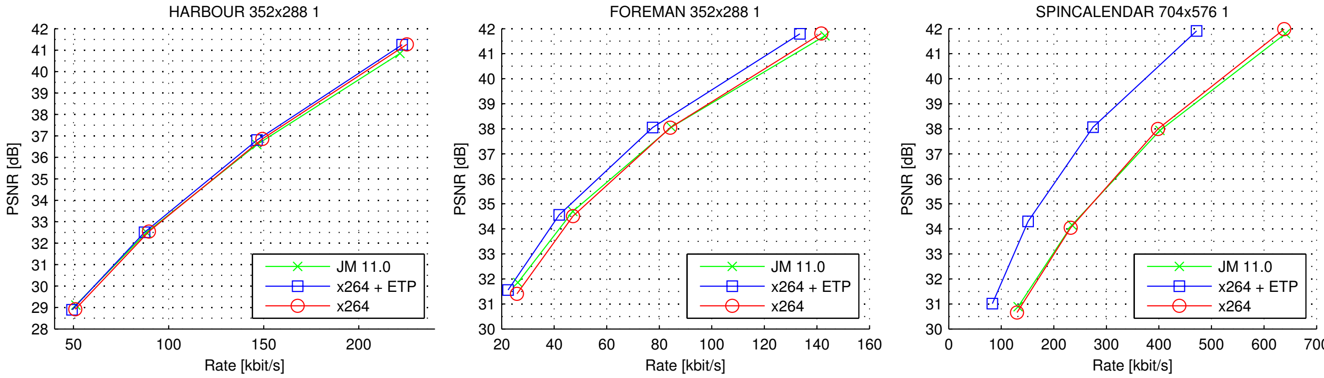


Fig. 3. Rate-distortion plots for selected test sequences.

tures, the prediction signals (i.e., the signals before addition of the residual) of SPINCALENDAR with and without ETP are compared in Figure 5. While some of the stripes are predicted well by the directional prediction modes, especially the digits in the calendar are recovered almost entirely using DIP.

6. CONCLUSION

The presented extended prediction modes for intra coding provide interesting improvements of the rate-distortion performance compared to conventional H.264/AVC intra coding for specific sequences. Depending on the structural features of the encoded sequence, we observed Bjøntegaard savings of up to 2.8 dB or 33.9% rate savings for a specific 4CIF sequence. The average rate savings for the remaining test set are around 5%.

It should be noted that currently, the observed gains are realized at a potentially substantial increase of complexity, especially considering the search on the decoder side for Markovian prediction. For a standardized method, a good trade-off between decoder complexity and the achievable rate-distortion improvements would need to be defined. Future work might involve exploitation of the fact that different prediction modes such as MTP, DIP, and H.264/AVC Intra have different rate–distortion–complexity characteristics; for instance, introduction of a complexity penalty in the mode decision process.

As intra encoded regions tend to transport the highest portion of signal energy in the video stream, improved prediction and coding methods for these may be an important area for future improvements of the H.264/AVC coding scheme.

7. REFERENCES

[1] ITU-T, "Recommendation H.264: Advanced video coding for generic audiovisual services," Mar. 2005.

[2] Gisle Bjøntegaard, "Coding improvement by using 4x4 blocks for motion vectors and transform," in *SG16/Q15 VCEG, 3. Meeting, Document Q15-C23*, Eibsee, Germany, Dec. 1997, ITU-T.

[3] Siu-Leong Yu and Christos Chrysafis, "New intra prediction using intra-macroblock motion compensation," Doc. JVT-C151, Joint Video Team (JVT), 3rd Meeting, Fairfax, VA, USA, May 2002.

[4] Satoshi Kondo, Hisao Sasai, and Shinya Kadono, "Tree structured hybrid intra prediction," in *Proc. IEEE Int. Conference on Image Processing ICIP '04*, Oct. 2004, vol. 1, pp. 473–476.

[5] Li-Yi Wei and Marc Levoy, "Fast texture synthesis using tree-structured vector quantization," in *SIGGRAPH '00: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, July 2000, pp. 479–488, ACM Press/Addison-Wesley Publishing Co.

[6] Thiow Keng Tan, Choong Seng Boon, and Yoshinori Suzuki, "Intra prediction by template matching," in *Proc. IEEE Int. Conference on Image Processing ICIP '06*, Atlanta, GA, USA, Oct. 2006.

[7] Thiow Keng Tan, Choong Seng Boon, and Yoshinori Suzuki, "Intra prediction by averaged template matching predictors," in *Proc. IEEE Consumer Communications & Networking Conference '07*, Jan. 2007, pp. 405–409.

[8] Ying-Qing Xu, Baining Guo, and Harry Shum, "Chaos mosaic: Fast and memory efficient texture synthesis," Tech. Rep., Microsoft Research, Redmond, WA, USA, Apr. 2000.

[9] "x264—a free H.264/AVC encoder," http://www.videolan.org/developers/x264.html.

[10] "FFmpeg multimedia system," http://ffmpeg.mplayerhq.hu/.

[11] Gisle Bjøntegaard, "Calculation of average PSNR differences between RD curves," in *SG16/Q6 VCEG, 13. Meeting, Document VCEG-M33*, Austin, TX, USA, Apr. 2001, ITU-T VCEG.