A BLOCK-ADAPTIVE SKIP MODE FOR INTER PREDICTION BASED ON PARAMETRIC MOTION MODELS

Alexander Glantz, Michael Tok, Andreas Krutz, and Thomas Sikora

Communication Systems Group
Technische Universität Berlin
Berlin, Germany

ABSTRACT

Motion compensated prediction (MCP) in hybrid video coding estimates a translational motion vector for a given block which is then used for residual computation. However, when complex motion like zoom, rotation, and perspective transformation occur, the translational model assumption does not always hold. This may result in higher residual energy and splitting of blocks, respectively. This paper proposes a skip mode based on higher-order parametric motion models. Often, these models provide a better prediction quality resulting in lower residual energy and larger block sizes. The proposed technique estimates a higher-order motion model between two given pictures. The encoder decides in terms of rate-distortion optimization whether to use the new skip mode for a block and therefore not to transfer any additional information like coefficient data. Experimental evaluation shows that the proposed technique can improve the coding performance of next generation video coding standards significantly.

Index Terms— H.264/AVC, HEVC, video coding, parametric motion model, motion compensated prediction

1. INTRODUCTION

The emergence of ever increasing temporal and spatial resolution in video sequences and the need to transmit this content lead to current joint standardization activities between ISO/IEC MPEG and ITU in the Joint Collaborative Team on Video Coding (JCT-VC). The working title of the codec under development is HEVC (high efficiency video coding) and its test model [1] currently needs about 30 to 40% less bit rate than H.264/AVC [2] with comparable quality. Its main improvements include larger quadtree-based coding units, i.e. formerly macroblocks, larger transform sizes, better interpolation filters and an optional loop filter that is based on Wiener filtering. However, such as in state-of-the-art video coding, temporal redundancy is still removed assuming translational motion.

In MPEG-4 Part 2, global motion compensation (GMC) has been standardized to be used for video object planes (VOP) besides (local) block-based motion compensation [3]. The idea was to transmit a set of up to four motion vectors, i.e. indirectly defining a so-called higher-order or parametric motion model, for every VOP and let the encoder decide for a block whether to use GMC or translational motion compensation. However, GMC has not found broad acceptance due to several reasons. First, techniques for parametric (global) motion estimation were not very sophisticated and extremely time-consuming. Secondly, the used and standardized interpolation technique (bilinear) does not perform well enough.

Today, very advanced interpolation techniques can be used and more importantly, the requirements for video codecs have changed completely. Whereas at the time of MPEG-4 Part 2 standardization the Common Intermediate Format (CIF) with its 352×288 pixels was considered as state-of-the-art, today's standardization activities are dealing with video content of up to 8000×4000 (8k) pixels. That is a factor of 315 times the size of CIF. The additional transmission of a set of global motion parameters for every picture of a CIF sequence really is a big deal. For 8k, its negligible.

In this paper we propose a new SKIP mode that is based on a very sophisticated parametric motion estimation technique and compensation using cubic spline interpolation. Thereby, the advantage of the common SKIP mode is used while diminishing its drawback, i.e. the assumption of translational motion. The new mode is incorporated into the test model that is used in current HEVC standardization activities.

The remainder of this paper is organized as follows. Section 2 shortly introduces the technique for parametric motion estimation that is used for the prediction mode presented herein. In Section 3, the prediction mode itself is presented. Section 4 describes the experimental evaluation and the last section summarizes the paper.

2. PARAMETRIC MOTION ESTIMATION

The new mode that is proposed in this paper relies on the accurate description of camera motion. The model that is used in common hybrid video coding, i.e. a translational model, is not sufficient for complex motion like zoom, rotation, and perspective transformation. Therefore, the proposed mode

978-1-4577-1302-6/11/$26.00 ©2011 IEEE

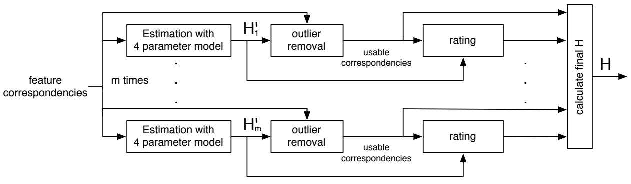


Fig. 1. Algorithm for parametric motion estimation using two different motion models and feature correspondencies, e.g. generated by block matching in common hybrid video coding or tracking as in KLT feature tracker.

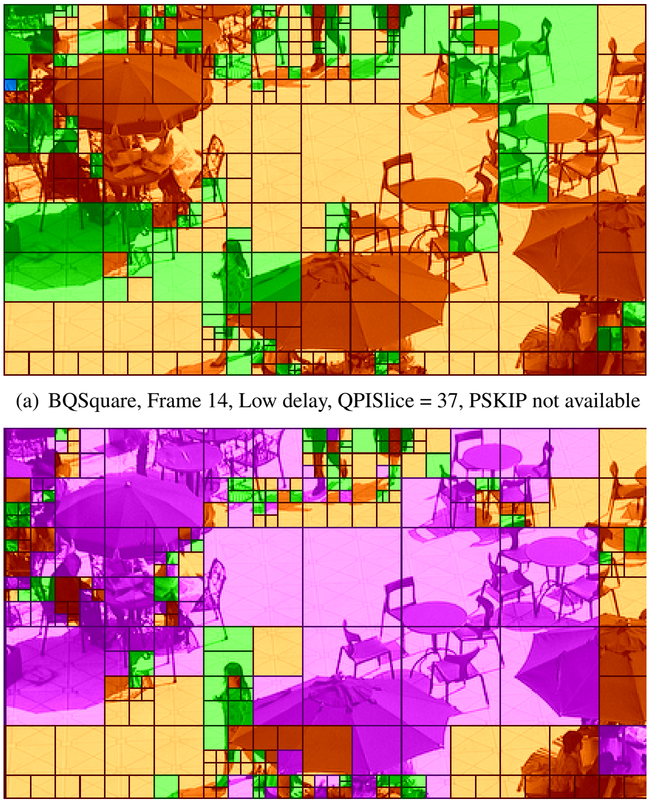
uses a higher-order motion model with eight parameters that includes all of the above.

However, only one set of parameters is used per picture correspondency at the moment. This means that, depending on the content of the sequence, the estimated motion is only describing a fraction of the picture's content, e.g. the background region. This is often also referred to as *global motion*.

The technique used for global motion estimation is an extension to the authors' previous work [4]. It is a two-step approach that first estimates local translational motion models for *good* features of a picture. It then estimates a parametric motion model using the tracked feature correspondencies between two given pictures of a video sequence with a modified version of a highly robust regression method based on the Helmholtz principle [5]. A block diagram of the technique is shown in Figure 1.

The feature correspondencies that are used for parametric motion model estimation are generated using the feature tracking method proposed by Kanade et al. [6], the so-called KLT tracker. The reasons for choosing the KLT tracker rather than correspondencies from common block matching are twofold. First, motion vectors from regular block structures, as is the case in hybrid video coding, are generated whether they contribute to the sought-after parametric model or not. This can severely affect the quality of the estimation. Additionally, motion vectors from hybrid video codecs are very likely to be quantized to quarter- or even half-pel accuracy. Results using the KLT tracker are subject to much higher motion information resolution.

The parametric motion estimation algorithm used herein derives a homography $\mathbf{H}$ for a pair of pictures from a field of feature correspondencies using the Helmholtz Tradeoff Estimator (HTE). The HTE is a robust estimator with the ability of detecting up to 80% of outliers in a given dataset for an underlying model using subsets. For every randomly chosen subset out of a given motion vector field, a 4-parameter model $\mathbf{H}_s$ is calculated. In the next step, every motion vector position is transformed using $\mathbf{H}_s$. The n-th percentile of the distances between estimated and true motion vector destinations computes a standard deviation that is used to part the subset into inliers and outliers. Here, n depends on the desired outlier tolerance. For the subset with the highest rating in terms of amount of inliers vs. inlier variance, a final homography $\mathbf{H}$ is calculated by least squares. For further information cf. [4].



(a) BQSquare, Frame 14, Low delay, QPISlice = 37, PSKIP not available

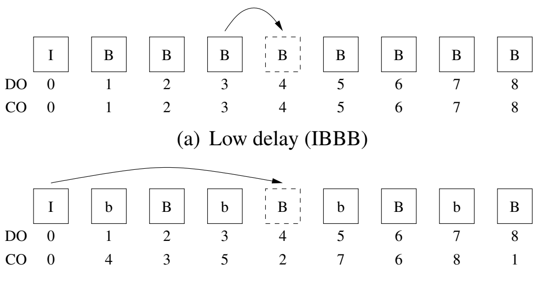
(b) BQSquare, Frame 14, Low delay, QPISlice = 37, PSKIP available

Fig. 2. Exemplary prediction structure using the HEVC test model HM 1.0 when Parametric Skip is either available besides common prediction modes or not (please refer to colored version; orange = SKIP, green = INTER, blue = INTRA, and magenta = PSKIP).

3. PARAMETRIC SKIP

All hybrid video codecs, e.g. state-of-the-art H.264/AVC or HEVC, divide the picture which is subject to encoding into blocks of different size and assign prediction modes to them. A prediction mode defines a method for generating a signal from previously encoded data, i.e. either spatial or temporal, that minimizes the residual between prediction and original. Additionally, all codecs perform some kind of rate-distortion optimization (RDO) that weighs the acceptable energy of the residual against the amount of information needed for transmission.

Prediction modes that are currently used to remove temporal redundancy are so-called INTER and SKIP modes. Whereas the INTER mode transmits both residual data in form of quantized transform coefficients and motion vector



(a) Low delay (IBBB)

(b) Random access (hierarchical B with I pictures every sec.)

Fig. 3. Exemplary coding structures (DO = display order, CO = coding order). The dashed boxes represent pictures that are currently decoded with PSKIP using the source of the arrow.

information that points to previously encoded/decoded pictures, the SKIP mode normally transmits nothing at all. This is particularly efficient for sequences with no motion at all or motion that can exclusively be described by a translational model. On the other hand, the INTER mode is effective especially if motion occurs that differs from the translational model, e.g. complex motion like zoom or people moving. This can be seen in the prediction structure for an exemplary picture in Figure 2(a). Here, the encoder chooses to transmit residual data for large blocks of background regions and smaller blocks of moving people using the INTER mode (green).

The proposed prediction mode PSKIP (Parametric Skip) uses the advantage of the SKIP mode, i.e. no transmission of residual data, while diminishing its drawback, i.e. the sensitivity to complex motion. It is based on the sophisticated estimation of parametric motion models as described in Section 2. For a given coding structure (cf. Figure 3), the algorithm first estimates the parametric motion between the current picture and a previous one that is available in the decoded picture buffer. Especially for longer prediction distances as in hierarchical B picture coding this can significantly outperform SKIP and INTER modes that rely on the assumption of translational motion. For a given block, the mode then compensates the motion to generate the prediction which equals the decoded signal. This is done using cubic spline interpolation of degree three.

The anticipated success of this approach can exemplarily be seen in Figure 2(b). Here, the PSKIP mode is incorporated besides common prediction modes into the same coding environment that was used to generate Figure 2(a). Large areas of the picture are now predicted using the PSKIP mode (magenta) replacing both SKIP and INTER modes. As expected, solely the moving people regions are still predicted using mainly the INTER mode (green).

Figure 4 shows the PSKIP mode incorporated into a common hybrid video encoding environment besides INTER (including SKIP) and INTRA modes. Besides better prediction signal generation the drawback of adding the new mode is transmission of additional side information. In addition to extra bits spent on signaling the prediction mode for a block, the parametric motion models have to be transmitted to the receiver as well. At the moment this is done by sending 8 uncompressed parameters in floating point precision in the slice header. This results in 256 additional bits per picture.

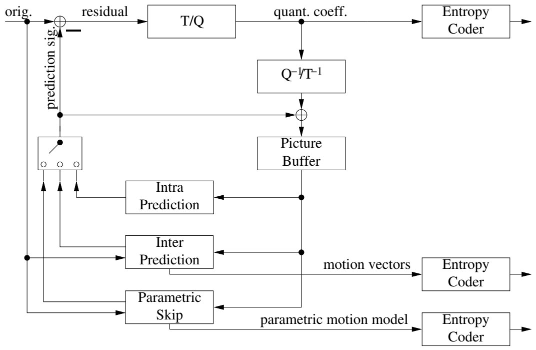


Fig. 4. Block-adaptive skip mode based on parametric motion models (Parametric Skip) within a hybrid video encoding environment.

Number of reference frames	2 (one per list)
RQT transf. size (min/max)	4/32 (RQT = residual quadtree)
Max RQT depth INTER	2
Max RQT depth INTRA	1
CU size / quadtree depth	64/4 (CU = coding unit)
Motion search range	64
Bit depth	8 (no internal bit depth increase)
Luma interpolation	Directional interpolation filter
Chroma interpolation	Bilinear interpolation
Entropy coder	Variable length coding (VLC)
Adaptive loop filter	off

Table 1. Low complexity coding settings.

4. EXPERIMENTAL EVALUATION

For experimental evaluation, the new mode has been incorporated into the HEVC test model HM 1.0 [1]. This software already saves between 30 to 40% bit rate compared to H.264/AVC and served as reference for measuring the coding performance. Table 1 shows the settings that have been used for the experimental evaluation. The main differences between HEVC and H.264/AVC are larger block sizes, larger transforms, better interpolation filters, and the use of an additional loop filter that is based on Wiener filtering (not used in low complexity coding setting).

Table 2 shows all test sequences that have been used as well as the results for low delay (IBBB) and random access (hierarchical B pictures with I pictures every second) coding. The rates have been evaluated for four QPISlice values each, i.e. $\{22, 27, 32, 37\}$. It shows that the performance of PSKIP

					Low delay		Random access	
Sequence	Source	Resolution	Frames	FPS	BD-rate	BD-PSNR	BD-rate	BD-PSNR
BlueSky	Taurus Media Technik	1920 × 1080	218	25	−8.2%	0.3 dB	−0.9%	0.0 dB
BQSquare	NTT DOCOMO Inc.	416 × 240	600	60	−2.8%	0.1 dB	−3.6%	0.1 dB
BQTerrace	NTT DOCOMO Inc.	1920 × 1080	600	60	−2.0%	0.0 dB	−1.4%	0.0 dB
Cactus	RAI	1920 × 1080	500	50	−1.2%	0.0 dB	0.1%	0.0 dB
City	ABC	1280 × 720	600	60	−3.6%	0.1 dB	−0.5%	0.0 dB
Desert	BBC	720 × 400	240	25	1.2%	0.0 dB	1.8%	−0.1 dB
Entertainment	RAI	720 × 576	250	25	0.0%	0.0 dB	0.0%	0.0 dB
PartyScene	NTT DOCOMO Inc.	832 × 480	500	50	−2.3%	0.1 dB	−2.8%	0.1 dB
Station2	Taurus Media Technik	1920 × 1080	250	25	−29.1%	0.9 dB	−9.7%	0.3 dB

Table 2. Test sequences and results of experimental evaluation in terms of BD-rate and BD-PSNR [7] for the low delay and random access coding structures. Negative BD-rates and positive BD-PSNR values indicate a gain in coding performance.

strongly depends on the content of the sequence that is subject to coding. Thus, performance varies between 29.1% gain (*Station2*, low delay) and 1.8% loss (*Desert*, random access).

The reason for the loss for some sequences is the missing RDO on a picture-level. Even if no block in a picture is encoded using PSKIP, the parametric motion model is sent to the receiver. Additionally, unnecessary overhead occurs since the parameters are sent uncompressed and block-based RDO has not been optimized to the new set of available modes. The occurrence of loss also seems to correlate with the resolution of the video sequence, which is obvious since the ratio between motion model and all other bits changes. For *Desert*, the reason for the loss is partly due to heat haze showing in the original. This causes extreme changes in the luminance values which negatively affects the performance of PSKIP.

However, the possibility to gain nearly 30% on a sequence that has already been compressed to about 30 to 40% relative to H.264/AVC shows the extreme potential of this technique. Scenes that do not comply with the assumption of translational motion only can benefit from the new mode. *Station2* and *BlueSky* for example show zoom respectively rotation that cannot be handled by common hybrid video codecs. Thus, further work should include a picture-based RDO to exclude single pictures from PSKIP usage and therefore increase coding performance even more.

5. SUMMARY

We presented a novel prediction mode PSKIP for hybrid video coding that is based on sophisticated parametric motion estimation. The PSKIP mode has been incoporated into the HEVC test model besides common prediction mode. For a given block, the encoder decides in terms of RDO whether to use the new mode. If so, all data is skipped and the signal is reconstructed using previously decoded pictures and a set of parametric motion parameters. Experimental evaluation shows that the new mode can significantly outperform next generation video coding standards.

6. REFERENCES

[1] K. McCann, B. Bross, and S. Sekiguchi, “High Efficiency Video Coding (HEVC) Test Model 1 (HM 1) Encoder Description,” Tech. Rep. JCTVC-C402, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, Guangzhou, China, Oct 2010.

[2] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, Jul 2003.

[3] T. Sikora, “The MPEG-4 video standard verification model,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 7, no. 1, pp. 19–31, Feb 1997.

[4] M. Tok, A. Glantz, A. Krutz, and T. Sikora, “Feature-based Global Motion Estimation using the Helmholtz Principle,” in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, Prague, Czech Republic, May 2011.

[5] R.L. Felip, X. Binefa, and J. Diaz-Caro, “A new parameter estimator based on the Helmholtz principle,” in *IEEE International Conference on Image Processing, 2005. ICIP 2005*, Sep 2005, vol. 2.

[6] J. Shi and C. Tomasi, “Good features to track,” in *Proc. CVPR ’94. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun 1994, pp. 593–600.

[7] G. Bjøntegaard, “Calculation of average PSNR differences between RD-curves,” *ITU-T SG16/Q.6 VCEG document VCEG-M33*, Mar 2001.