A DYNAMIC MODEL BUFFER FOR PARAMETRIC MOTION VECTOR PREDICTION IN RANDOM-ACCESS CODING SCENARIOS

Michael Tok, Marko Esche, Thomas Sikora

Communication Systems Group
Technische Universität Berlin
Berlin, Germany

ABSTRACT

Motion compensated inter prediction is a powerful tool used in modern hybrid video codecs to reduce the temporal redundancy of video sequences. However, the motion information needed for motion compensation is highly redundant as well. Thus, motion vector prediction and difference coding is a common method in modern video codecs. During the standardization of HEVC, new methods for motion prediction such as temporal motion vector prediction have been analyzed. This paper presents a method for motion vector prediction from perspective motion models in random access scenarios with hierarchical group of picture structures. To enable this kind of prediction a dynamic buffer system for generating, compressing and transmitting the underlying motion models is introduced. Bit rate reductions of up to 5% underline the performance of the complete system.

Index Terms— Hybrid Video Coding, HEVC, Motion Compensation, Motion Vector Prediction, Motion Merge

1. INTRODUCTION

The ever increasing spatio-temporal resolution of video content has led to the development of more efficient video coding standards for the last decades. H.264/AVC [1] e.g. has a compression ratio twice as efficient as the preceding standard MPEG-4 Visual [2] at a comparable video quality level and standardization efforts of the joint colaborative team for video coding (JCTVC) strive for a new standard with even higher compression rates. The new standard under consideration, called high efficiency video coding (HEVC), basically uses the same concepts as H.264/AVC. The higher performance mainly results from allowing larger prediction block and transform sizes and utilizing better in-loop filters such as Sample Adaptive Offset (SAO) [3].

However, inter prediction through block-wise translational motion estimation and compensation still is the common technique for temporal redundancy reduction and compression in hybrid video codecs. Unfortunately, the motion vector field used for such redundancy reduction is highly redundant itself. Thus, techniques, such as spatio-temporal prediction for compressing this motion information, are discussed by JCTVC and utilized in HEVC [4]. For each block, called coding unit (CU), in HEVC and for each reference index, a first motion vector predictor candidate is taken from the left neighboring CU or the CU above. If it exists, a second predictor is taken from the collocated CU of an already decoded frame. Finally, the best matching combination of predictor and reference index in terms of rate distortion optimization is encoded and the difference to the block's estimated motion is transmitted. Candidates for the Merge mode are generated in a similar way [4]: Motion vector information and reference index of a neighboring or collocated unit are simply copied.

Nonetheless, all such predictor and merge candidates are imprecise when it comes to complex global motion such as zoom, rotation or perspective deformation. During the standardization of H.264/AVC Sun et al. proposed a technique for encoding motion vectors with global motion models [5]. To allow slight differences between a global motion model and the vectors to be encoded, Yuan et al. presented a motion vector predictor for zoom motion [6]. An enhancement to HEVC that predicts motion vectors from perspective motion models in low delay settings was introduced in [7].

This paper presents two extensions to that approach. A dynamic model compression and buffering system is introduced, that derives and encodes all needed models to generate motion vector predictors in random access settings with hierarchical group of picture (GOP) structures. In addition to motion vector predictors, merge candidates are generated to decrease the amount of bits needed for motion information coding. The remainder of this paper is organized as follows. Section 2 describes the generation and utilization of motion vector prediction and merge candidates derived from perspective parametric motion models. The parametric motion estimation technique used to obtain the underlying homographies for parametric motion vector prediction (PMVP) and parametric merge (PMERGE) is described in Section 3. The novel dynamic model compression and buffering system enabling PMVP and PMERGE in random access coding scenarios is introduced in Section 4. Section 5 presents the experimental evaluation in terms of coding results for the HEVC test model HM 9 and finally, Section 6 summarizes this paper.

2. PARAMETRIC MOTION VECTOR PREDICTION

The spatial motion vector predictors and merge candidates commonly used in the current HEVC test model HM 9 [8] perform well for sequences with smooth translational motion. In addition, the collocated predictor candidate (and merge candidate respectively) allows to precisely predict temporally consistent motion with high precision. Figure 1(a) illustrates the position of these predictors. Nevertheless, as pointed out in [7], these candidates lack in precision when predicting higher order motion such as zoom, rotation, and perspective deformation. Figure 2 gives an example for such a case. To overcome this issue, additional parametric motion vector prediction and merge candidates can be added to the respective candidate lists.

For generating such candidates, parametric motion models (PMMs) have to be transmitted for each frame. With these models, vectors with endpoint $(x', y')^T$ can be derived at each block with center position $(x, y)^T$:

$$\begin{pmatrix} x' \cdot w' \\ y' \cdot w' \\ w' \end{pmatrix} = \mathbf{H} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1}$$

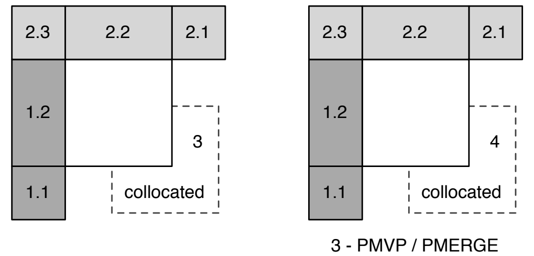
where $\mathbf{H}$ is the perspective parametric motion model containing the 8 perspective transformation parameters:

$$\mathbf{H} = \begin{pmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{pmatrix}. \tag{2}$$

The resulting vectors $\mathbf{v} = (x' - x, y' - y)^T$ are finally quantized to quarter pel precision and added to the candidate lists for PMVP and PMERGE. Figure 1(b) shows the position of the new set of predictors and merge candidates. Experiments have shown higher bit rate savings, when listing the parametric predictors and merge candidates before the collocated ones. Thus, bits for indexing the parametric (predictor) candidate selection are saved.

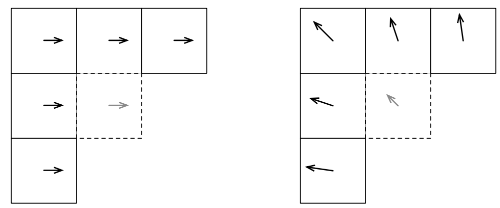
3. PARAMETRIC MODEL ESTIMATION

To estimate the needed PMMs, that describe the complex deformations resulting from camera motion, the motion estimation presented in [9] is used. This method is based on feature selection, tracking, and evaluation by a simplified RANSAC: For each frame 400 features are selected and tracked by KLT-feature-tracking. Subsequently, a modified RANSAC is applied on these features for robustly estimating an eight parameter perspective motion model only from background feature correspondences. To reduce the amount of iterations needed by RANSAC for finding a reliable subset, in each iteration



(a) in HM 9 reference (b) with parametric candidates

Fig. 1. Candidates for motion vector prediction and MERGE. For MVP the first two (existing) candidates are selected. For MERGE all existing candidates are used.



(a) For smooth translation. (b) For zoom.

Fig. 2. For complex global motion a problem with predicting (or merging) the gray dashed MVs from neighboring MBs occurs.

k only a four parameter model $\mathbf{H}_k$ is derived for correspondence classification

$$\mathbf{H}_k = \begin{pmatrix} \tilde{m}_{0,k} & \tilde{m}_{1,k} & \tilde{m}_{2,k} \\ -\tilde{m}_{1,k} & \tilde{m}_{0,k} & \tilde{m}_{3,k} \\ 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

This model is used to identify whether a feature correspondence is an inlier or not. Finally, the largest set of inliers is taken to calculate a final perspective motion model with the Minimum Least Squares Method, consisting of the eight parameters $m_0 \ldots m_7$.

4. DYNAMIC MODEL BUFFERING

The random access coding setting for HM 9 allows up to four reference frames per slice. Thus, for generating parametric vector predictors for each frame in the reference frame lists up, to four motion models per slice would have to be transmitted as well. Furthermore, each model consists of eight floating point parameters, each represented by 32 to 64 bit.

For lossy compression of affine parameters with six to twelve parameters Karczewicz et al. presented a compression scheme based on orthonormaliation, quantization and entropy coding [10]. This scheme is also used by Steinbach et al. in [11] to improve the performance of H.263 [12]. However, perspective models which are used for PMVP and PMERGE

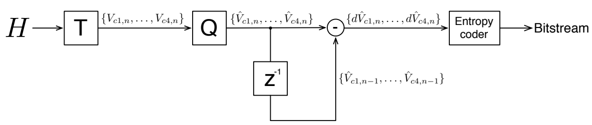


Fig. 3. Method for lossy perspective model compression.

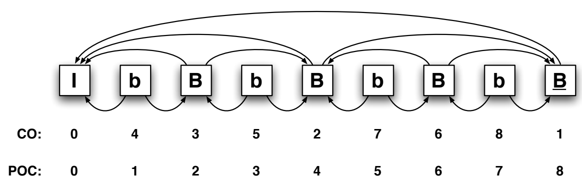


Fig. 4. Coding order (CO) and picture order count (POC) for HM 9 random access setting (RA), **b**-frames are not selected as reference for Inter-coding, one GOP ends after 8 frames (last frame of this GOP is marked as **B̲**)

because of their higher precision compared to affine models, cannot be orthonormalized in the same way.

In [13] a lossy compression scheme for perspective motion models was presented. Figure 3 explains this scheme. First, each model is transformed to a set of four frame-corner motion vectors following Eq.1 as these vectors are more robust to quantization than the perspective parameters themself. These vectors are quantized to quarter pel precision since the PMVP and PMERGE vectors have to have quarter pel precision as well. Finally, the vectors are encoded by temporal difference and subsequent Exponential Golomb coding.

In [7] this compression method is used to transmit one model per frame to generate PMVP candidates for the low-delay coding setting. To get candidates for each reference frame present in a reference frame list, previously transmitted models are concatenated in addition. However, the random access coding setting as defined in [14] has a more complex temporal prediction structure where frame coding order and display order differ. Figure 4 illustrates the (simplified) Inter prediction dependencies of the random access coding setting.

To enable PMVP and PMERGE candidates for this setting, a dynamic buffer system for transmitting and generating all needed models is required. This system transmits a set of eight short term motion models with every last frame of a GOP, marked as **B̲** in Figure 4. By concatenation and inversion of these models, perspective transformation parameters for all Inter prediction dependencies are calculated. Figure 5 gives an example of such model generation. To obtain a model $\mathbf{H}_{8,0}$ for instance, describing the transformation from POC (picture order count) 8 to POC 0, the models $\mathbf{H}_{8,7}$ to $\mathbf{H}_{1,0}$ are just multiplied:

$$\mathbf{H}_{8,0} = \prod_{i=0}^{7} H_{i+1,i}. \tag{4}$$

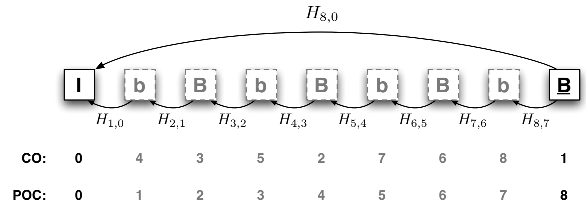


Fig. 5. With each **B̲**-Frame, a set of models for a whole GOP ($H_{8,7}, \ldots, H_{1,0}$) is transmitted. Thus, a model for describing the parametric motion between frame 8 and 0 ($H_{8,0}$) can be derived e.g.

HEVC test software	HM 9.0.1
Profile	Main
GOP settings	hierarchical B, random access
QP	$\{27, 32, 37, 42\}$
Largest CU size	64
Smallest CU size	8
Internal Bit-Depth	8
Motion search range	64×64

Table 1. HEVC coding settings for experimental evaluation

When a model for prediction from successive (in terms of POC) frames is needed, a homography is generated by concatenation and then inverted. A model $\mathbf{H}_{4,8}$ e.g. is simply derived by

$$\mathbf{H}_{4,8} \quad = \quad \left(\prod_{i=4}^{7} H_{i+1,i} \right)^{-1}. \tag{5}$$

5. EXPRIMENTAL EVALUATION

The proposed model compression and buffering scheme has been incorporated in the HEVC reference software HM 9 to enable the generation of PMVP and PMERGE candidates in the random access coder setting. In order to verify the performance of the presented techniques, eight test sequence with various resolutions and frame rates were encoded for experimental evaluation.

Table 1 overviews the encoder settings used for experimental evaluation. Table 2 presents the test sequences' properties as well as the encoding results with PMVP only and PMVP in combination with PMERGE . The gains in terms of BD-rate and PD-PSNR [15] indicate, that sequences with complex and fast global motion such as Stefan or Race profit from the new predictor and merge candidates. For other sequences with higher order motion such as Stanford, City and BlueSky smaller gains can be observed.

Sequence	Resolution	frames	fps	PMVP BD-rate [%]	PMVP BD-PSNR [dB]	PMVP + PMERGE BD-rate [%]	PMVP + PMERGE BD-PSNR [dB]
Mountain	352 × 192	130	25	-1.46	0.07	-1.51	0.08
Stefan	352 × 240	300	30	-2.72	0.14	-2.71	0.14
Monaco	352 × 288	150	25	0.31	-0.01	0.30	-0.01
Race	544 × 336	100	25	-4.67	0.18	-5.01	0.19
Stanford	720 × 480	304	25	-1.07	0.03	-1.10	0.03
Palace	720 × 576	120	25	0.11	0.00	0.16	-0.01
City	1280 × 720	250	60	-0.88	0.03	-1.00	0.03
BlueSky	1920 × 1080	217	25	-0.30	0.01	-0.62	0.02
mean				-1.34	0.06	-1.44	0.04

Table 2. Sequence properties and encoding results for random access coding structure

Figure 6 shows exemplary rate distortion curves for the Stefan and Race test sequences. These curves show, that the bit rate savings through PMVP and PMERGE are obtained over a wide bit rate range.

Nevertheless, results for the Monaco and the Palace sequence indicate, that the additional candidates are impracticable for sequences with simple global motion such as pan. The increased predictor signalling costs induced by longer candidate lists lead to bit rate increases although the new predictors and merge candidates are not used.

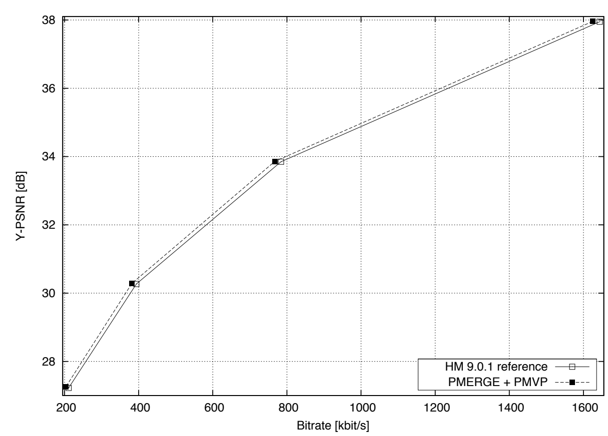
There is evidence, that a link between the performance of PMVP/PMERGE and the complexity of a sequence's motion exists. Thus, for further work, this dependency has to be analyzed in detail. A definition of 'motion complexity' in the context of motion prediction and compression has to be elaborated first.

6. SUMMARY AND CONCLUSION

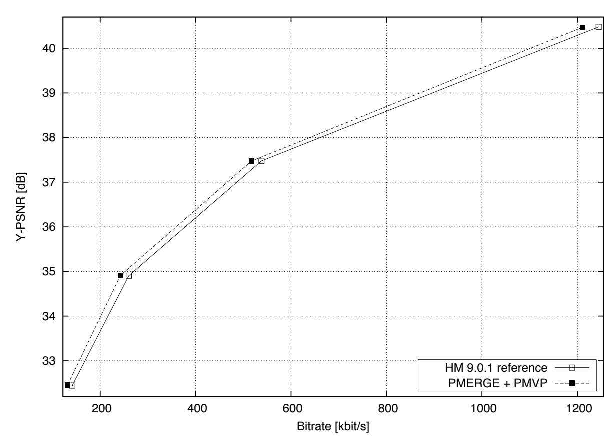
A model compression and buffering scheme for higher order motion model utilization in random access coding settings has been presented. The usability of this scheme for parametric motion vector prediction and parametric merge candidate generation in the HEVC test model HM 9 has been evaluated.

Bit-rate-savings of up to 5% in terms of BD-rate are achievable with the presented techniques and show that additional motion vector prediction (and merge) candidates can improve the encoding efficiency of the HM dramatically.

However, for sequences with simple global motion such as pure translation, the additional parametric candidates lead to increased bit rates through higher predictor index costs. Thus, a deeper understanding of when PMVP and PMERGE candidates are needed, can lead to even higher coding efficiency.



(a) Stefan



(b) Race

Fig. 6. RD-curves for coding results with the HM 9 reference software (straight line) and HM 9 with PMVP + PMERGE (dashed line)

7. REFERENCES

[1] "Draft ITU-T recommendation and final draft inernational standard of joint video specification (ITU-T Tec. H.264/ICO/IEC 14496-10 AVC)," 2003.

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

[3] C.-M. Fu, E. Alshina, A. Alshin, Y.-W. Huang, C.-Y. Chen, C.-Y. Tsai, C.-W. Hsu, S.-M. Lei, J.-H. Park, and W.-J. Han, "Sample Adaptive Offset in the HEVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1755 –1764, Dec 2012.

[4] B. Bross, J. Jung, Y.-W. Huang, Y. H. Tan, I.-K. Kim, T. Sugio, M. Zhou, T. K. Tan, E. Francois K. Kazui, W.-J. Chien, S. Sekiguchi, S. Park, and W. Wan, "Bog report of ce9: Mv coding and skip/merge operations," Mar 2011.

[5] S. Sun and S. Lei, "Motion vector coding with global motion parameters," *ITU-T SG16/Q.6 VCEG document VCEG-N16*, Aug 2001.

[6] Hui Yuan, Yilin Chang, Zhaoyang Lu, and Yanzhuo Ma, "Model Based Motion Vector Predictor for Zoom Motion," *Signal Processing Letters, IEEE*, vol. 17, no. 9, pp. 787 –790, Sep 2010.

[7] M. Tok, A. Glantz, A. Krutz, and T. Sikora, "Parametric motion vector prediction for hybrid video coding," in *Picture Coding Symposium*, May 2012, pp. 381 –384.

[8] K. McCann, B. Bross, W.-J. Han, I. K. Kim, K. Sugimoto, and G. J. Sullivan, "High Efficiency Video Coding (HEVC) Test Model 9 (HM 9) Encoder Description," JCTVC, Dec 2012.

[9] M. Tok, A. Glantz, A. Krutz, and T. Sikora, "Monte-Carlo-based Parametric Motion Estimation using a Hybrid Model Approach," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1, 2012.

[10] M. Karczewicz, J. Nieweglowski, J. Lainema, and O. Kalevo, "Video coding using motion compensation with polynomial motion vector fields," in *First International Workshop on Wireless Image/Video Communications*, Sep 1996, pp. 26 –31.

[11] E. Steinbach, T. Wiegand, and B. Girod, "Using multiple global motion models for improved block-based video coding," in *Proceedings of the 6th IEEE International Conference on Image Processing*, Sep 1999, pp. 56–60 vol.2.

[12] ITU-T Recommendation H.263 Version 2 (H.263+), "Video coding for low bitrate communication," Jan 1998.

[13] M. Tok, A. Krutz, A. Glantz, and T. Sikora, "Lossy parametric motion model compression for global motion temporal filtering," in *Picture Coding Symposium*, May 2012, pp. 309 –312.

[14] F. Bossen, "Common HM test conditions and software reference configurations," JCTVC, Dec 2012.

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