Motion Modeling for Motion Vector Coding in HEVC

Michael Tok, Volker Eiselein and Thomas Sikora
Communication Systems Group
Technische Universität Berlin
Berlin, Germany

***Abstract*—During the standardization of HEVC, new motion information coding and prediction schemes such as temporal motion vector prediction have been investigated to reduce the spatial redundancy of motion vector fields used for motion compensated inter prediction. In this paper a general motion model based vector coding scheme is introduced. This scheme includes estimation, coding and dynamic recombination of parametric motion models to generate vector predictors and merge candidates for all common HEVC inter coding settings. Bit rate reductions of up to 4.9% indicate that higher order motion models can increase the efficiency of motion information coding in modern hybrid video coding standards.**

I. Introduction

Most modern video coding standards apply block based hybrid video coding with intra prediction and motion compensated inter prediction. In such way the spatial and temporal redundancy of video sequences to be encoded is reduced drastically which leads to highly efficient video coding schemes. The prediction residual is transform coded, quantized and entropy coded. This general procedure has not changed since the standardization of H.261 [1] more than 20 years ago. Modern codecs such as H.264/AVC [2] or the newest video coding standard HECV, approved in April 2013 [3], still follow this scheme with slight changes. Amongst others, the main improvements include variable block sizes for motion compensation and for transform coding, higher motion compensation precision, more complex motion vector prediction, differing post filters like the sample adaptive offset filter in HEVC [4] and new forms of context adaptive entropy coding.

In the whole coding process, the motion compensated inter prediction reduces the highest amount of redundancy in a video sequence and thus contributes remarkably to the coding efficiency of modern video codecs. The motion information (called motion vectors) to be transmitted to the decoder is highly redundant and thus predicted from neighboring blocks. Many video sequences contain camera motion with a higher order motion behaviour such as zoom or rotation. Higher order motion models can be utilized to improve the encoding of such sequences. In [5] e.g. Springer et al. describe how to extend the whole inter prediction process by warping already encoded video frames with the help of higher order motion models. Sun et al. propose to encode motion vectors with global motion parameters [6]. However, in some cases the motion information to be encoded differs slightly from a rigid parametric motion model. In such cases it is better to use a higher order motion model for motion vector prediction rather than for motion vector coding.

This paper describes a way to utilize so called parametric motion models to improve the prediction and thus coding of motion vectors for inter prediction in HEVC. To reduce the amount of model parameters to be transmitted, a highly dynamic buffering and coding scheme is presented that enables the generation of parametric motion vector predictors (PMVPs) and parametric merge (PMERGE) candidates for all kinds of encoding scenarios. This allows to derive parametric vector predictors for more than one inter reference per slice in HEVC by only transmitting one compressed model per frame.

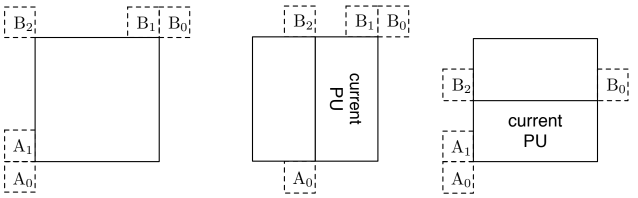
The remainder of this paper is organized as follows. The vector prediction and merge scheme based on parametric motion models is described in section II. The robust parametric motion estimation algorithm is shortly introduced in section III. A description of the dynamic motion model compression and combination scheme is given in section IV. Section V presents the experimental evaluation in terms of coding results for the HEVC test model HM 16 and finally, Section VI summarizes and closes this paper.

II. Parametric Motion Vector Prediction

In the beginning of the HEVC standardization, various motion vector prediction strategies have been proposed and evaluated by the JCT-VC partners [7] and the need of further investigation of MVP approaches and MERGE schemes became apparent [8]. The complex methods for AMVP and MERGE candidate generation as described in the final HEVC draft [3] generate precise motion vector predictors and motion representations for a wide variation of motion types ranging from spatial regular to temporal consistent motion. For spatially consistent motion, two candidates are chosen from neighboring PUs following the prediction schemes illustrated in figure 1. For spatial unsteady but temporal consistent motion an additional so called colocated predictor (merge candidate respectively) is derived from previously decoded video frames. In the MERGE candidate generation process, additional candidates for biprediction are generated. Further details of the AMVP and MERGE candidate derivation process are given by [3] and [9].

However, as pointed out in [10], spatial and temporal MVP (and MERGE) candidates can lack precision if a sequences motion is neither spatially regular, nor temporally consistent,

978-1-4799-7783-3/15/$31.00 ©2015 IEEE



(a) 2Nx2N (b) lower Nx2N (c) right 2NxN

Fig. 1. Neighboring AMVP and MERGE candidate PUs for differing PU partitions in HEVC. One candidate is taken from the first existing A, the second one from B. The existence is checked in ascending index order ($A_0 \rightarrow A_1 \rightarrow$ and $B_0 \rightarrow B_1 \rightarrow B_2$).

leading to inefficient MV coding and poor MERGE candidates. In these cases the inter prediction efficiency is reduced which leads to a higher residual energy and thus to more transform coefficients to be transmitted. Such motion can be induced by arbitrarily moving objects and all kinds of camera position and zoom changes with varying velocity. For the latter so called parametric motion models (PMMs) can be estimated by various methods and be used to calculate parametric motion vector predictors and parametric merge candidates. Therefore, for each PU center $(x, y)^T$ a transformed position $(x', y')^T$ is obtained by

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

where $\mathbf{H}$ is a perspective parametric motion model, containing 8 perspective transform 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 final motion vector $\mathbf{v} = (x' - x, y' - y)^T$ is quantized to quarter pel precision and added to the AMVP and MERGE candidate list.

III. Parametric Model Estimation

To estimate the parametric motion models for PMVP and PMERGE, a parametric motion estimation aproach presented in [11] is utilized. This method is based on feature selection, tracking and robust regression by a simplified RANSAC: For each frame up to 400 features are selected and tracked by KLT-feature-tracking. Subsequently, a modified RANSAC is applied on these features to estimate an eight parameter perspective motion model robustly that describes the background deformation induced by camera motion. To reduce the amount of iterations in RANSAC for finding a reliable model, in each iteration k only a four parameter model $\tilde{\mathbf{H}}_k$ as shown in eq. (3) is derived for correspondence classification by two randomly selected features.

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

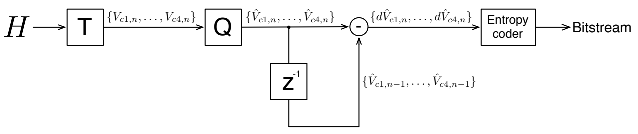

H
T
$\{V_{c1,n}, \ldots, V_{c4,n}\}$
Q
$\{\hat{V}_{c1,n}, \ldots, \hat{V}_{c4,n}\}$
-
$\{d\hat{V}_{c1,n}, \ldots, d\hat{V}_{c4,n}\}$
Entropy coder
Bitstream
z^{-1}
$\{\hat{V}_{c1,n-1}, \ldots, \hat{V}_{c4,n-1}\}$

Fig. 2. Coding scheme for the parametric motion models

This model is only used to identify whether a feature correspondence is an inlier, describing dominant parametric motion, or not. Over each iteration, the largest inlier set (consensus set) is kept. Finally, this consensus set is taken to calculate a final perspective motion model consisting of the eight parameters $m_0, \ldots, m_7$ by least mean of squares.

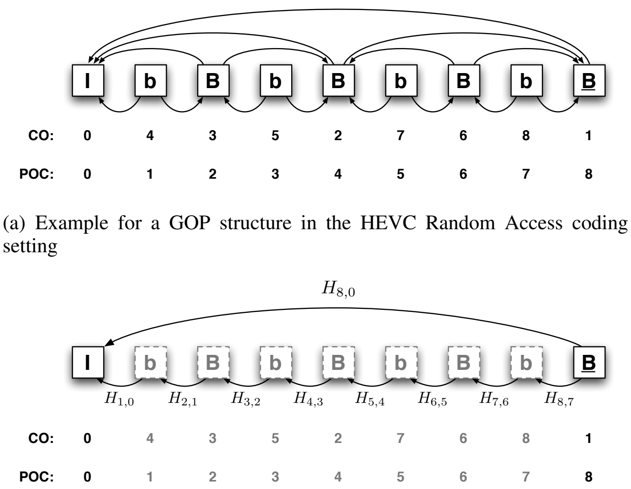
IV. Model Compression and Buffering

The coding settings of HM 16 allow up to four reference frames per slice. For generating parametric vector predictors, up to four PMMs per slice have to be transmitted as a consequence. As each model consists of eight real numbers represented as 32 or 64 bit floating point values, up to $4 \cdot 8 \cdot 64 = 2048$ additional bits per slice would be necessary for PMVP and PMERGE when transmitting the model parameters as raw data.

To compress affine motion models with six parameters Karczewicz et al. introduced a compression scheme based on parameter orthonormalization, quantization and entropy coding [12]. This scheme has also been used by Steinbach et al. in [13] to improve the performance of the H.263 inter prediction. Kordasiewicz et al. also introduced an extension to this method by applying context adaptive binary arithmetic coding on the orthonormalized and quantized model parameters [14]. Perspective motion models have the advantage of higher motion description precision when perspective deformations are part of a sequences motion. Unfortunately, the perspective parameters lead to nonlinear operations and thus cannot be orthonormalized like affine model parameters. Also the eight parameters (m_0 and m_7) differ in their order of magnitude (10^{-9} to 10^0 e.g.).

In [15], a lossy compression scheme for perspective motion model compression was presented. This scheme is described in figure 2. Each motion model is transformed (Block $\mathbf{T}$ in figure 2) to a set of four frame corner motion vectors $\{\mathbf{v}_{c1,n}, \ldots, \mathbf{v}_{c4,n}\}$ following eq. 1. These motion vectors have about the same order of magnitude and are more robust to quantization. Because the maximum precision of all motion vectors derived from the transmitted motion models is quarter pel. the four corner motion vectors representing a PMM can be quantized to quarter pel as well (Block $\mathbf{Q}$ in figure 2). Another advantage is the temporal redundancy of each corner motion vector allowing efficient difference coding (Blocks $\mathbf{z^{-1}}$ and $\ominus$ in figure 2). The final entropy coding is performed by Exponential Golomb coding.

With this scheme, one lossy compressed model per frame n for parametric motion vectors from frame n to $n-1$ is transmitted. To derive models for all reference frames, these



(a) Example for a GOP structure in the HEVC Random Access coding setting

(b) Long Term Motion model calculation through short term model concatenation

Fig. 3. Exemplary GOP structure for the Random Access setting (a) and corresponding concatenated Motion Model for PMVP in frame 8 with reference frame 0 (b)

models can be concatenated and, in the case of hierarchical GOP structures inverted as needed, to get models for all reference frames. For low delay, one model is transmitted with each frame. In the random access case (see figure 3(a)) a set of eight models is sent with the last frame **B** of each GOP. Figure 2 gives an example for the generation of additional models. To obtain a motion model $\mathbf{H}_{8,0}$ from POC (picture order) 8 to POC 0 e.g, the models $\mathbf{H}_{8,7}$ to $\mathbf{H}_{1,0}$ are simply multiplied (see figure 3(b)):

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

For models describing the motion to successive POCs, models are concatenated and inverted subsequently. A model $\mathbf{H}_{4,8}$ for instance is calculated by

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

That way, for each inter frame PMVP and PMERGE candidates for all reference frames can be generated although only one compressed model per frame on average is transmitted. This compression and buffering scheme leads to a bitrate of about 64 bit per frame in comparison to the 2048 for raw model transmission.

V. Exprimental Evaluation

The proposed scheme for PMVM and PMERGE candidate generation through dynamic model buffering has been incorporated in the HEVC reference software HM 16.0 for experimental evaluation. For performance verification, 12 test sequences with varying resolution, frame rates and video content were encoded with following three coding settings defined by [9]:

- Low Delay (B frames)
- Low Delay (P frames)
- Random Access with a GOP of 8 frames
  (B frames, 1 second intra period)

Table I depicts the test sequences' resolution as well as the encoding results in terms of BD rate and BD-PSNR [16] for the three encoder settings[1]. For sequences with complex higher order motion such as combinations of zoom, pan and perspective deformations like City, Room3D or Stefan gains of about 1.8% to 3.2% can be observed when using the low delay B frame setting. As pointed out in [10], most of the gain results from MVP candidates with much higher precision which leads to better inter prediction and thus less transform coefficients for residual transmission. Other sequences like BQMall or BQTerrace with very slow, very consistent motion do not benefit from PMVP or PMERGE. The additional signaling however leads to a higher bitrate and in this way to a loss of 0.2% to 0.4%.

In general the low delay P frame setting is less efficient because the powerful biprediction is not used. Thus the achievable gains with PMVP and PMERGE are higher (and losses are lower respectively) compared to the low delay B setting. Videos encoded with hierarchical GOP structure and biprediction from more than one temporal direction as used in the random access setting can also be encoded with a lower bitrate when PMVP and PMERGE are used. It has to be mentioned that the random access coding setting encodes one intra frame per second. As these frames are only encoded with the use of intra prediction, PMVP and PMERGE cannot improve the coding efficiency for these frames. The Race sequence however can be encoded with about 4.9% less bits on average with the new MVP and merge candidate. Figure 4 shows exemplary rate distortion curves for four test sequences encoded with the three coding settings.

VI. Summary and Conclusion

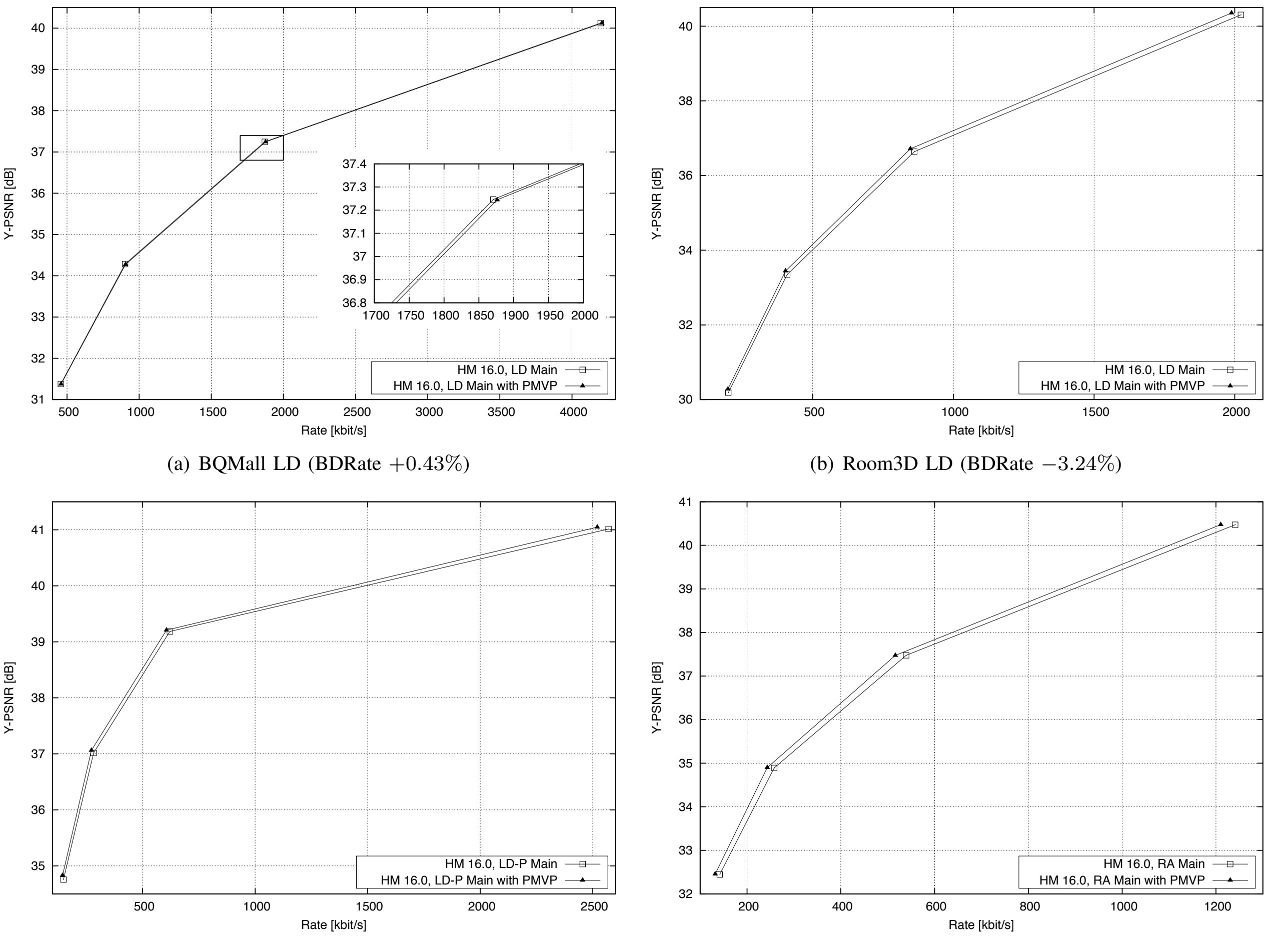
A new set of AMVP and MERGE candidates derived from parametric motion models has been presented. To obtain these candidates, a novel model compression and buffering scheme for higher order motion models has been introduced. This scheme is capable of generating needed additional motion models through model concatenation and inversion to enable the derivation of PMVP and PMERGE candidates in all standard GOP settings of HEVC. The benefit of PMVP and PMERGE has been evaluated by incorporating the model coding and buffering scheme and parametric motion vector derivation process in the HEVC Test model HM 16.0. Bit rate savings of up to 4.9% in terms of BD-rate indicate that the inter prediction process in HEVC can be improved by utilizing parametric motion models.

However, the additional signaling costs for PMVP and PMERGE can lead to increased bit rates in sequences where

[1] For further results please visit www.nue.tu-berlin.de/research/modmvc

TABLE I
SEQUENCE RESOLUTIONS AND ENCODING RESULTS FOR THE HEVC CODING ORDERS

		Low delay (LD)		Low delay P (LD-P)		Random access (RA)	
Sequence	Resolution	BD [%]	BD-PSNR [dB]	BD [%]	BD-PSNR [dB]	BD [%]	BD-PSNR [dB]
BasketballDrive	1920 × 1080	-0.16	0.00	-0.54	0.01	-0.46	0.01
Bigships	1280 × 720	-1.77	0.05	-1.99	0.05	-0.81	0.02
BQMall	832 × 480	0.43	-0.02	0.33	-0.01	0.06	0.00
BQTerrace	1920 × 1080	0.29	0.00	-0.91	0.01	0.06	0.00
City	1280 × 720	-2.80	0.07	-4.01	0.10	-0.67	0.02
ParkJoy1	2560 × 1600	-0.71	0.02	-1.05	0.03	-0.57	0.02
Race	544 × 336	-0.69	0.02	-1.35	0.05	-4.91	0.19
Room3D	720 × 576	-3.24	0.14	-4.86	0.20	-0.86	0.04
Station2	1920 × 1080	-2.77	0.06	-4.48	0.09	-0.56	0.01
Stefan	352 × 240	-1.78	0.09	-2.28	0.12	-2.61	0.14
Tractor	1920 × 1080	-0.37	0.01	-1.20	0.04	-1.20	0.04
Waterfall	704 × 480	-2.95	0.08	-3.85	0.10	-0.13	0.00
Average		-1.38	0.03	-2.18	0.07	-1.06	0.04



(a) BQMall LD (BDRate $+0.43\%$)

(b) Room3D LD (BDRate -3.24%)

(c) Station2 LD-P (BDRate -4.48%)

(d) Race RA (BDRate -4.91%)

Fig. 4. Selected rate distortion curves for different coding settings and sequences

both techniques are not used. This is the case for sequences where too less or purely translational motion is dominant. For such cases adaptive switching schemes, deactivating PMVP and PMERGE automatically are needed.

REFERENCES

[1] T. Sikora, “Trends and perspectives in image and video coding,” *Proceedings of the IEEE*, vol. 93, no. 1, pp. 6–17, Jan 2005.

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

[3] “High efficiency video coding,” ITU-T, Recommendation, Apr 2013.

[4] 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.

[5] D. Springer, F. Simmet, D. Niederkorn, and A. Kaup, “Robust Rotational Motion Estimation for efficient HEVC compression of 2D and 3D navigation video sequences,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 1379–1383.

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

[7] G. Clare, J. Jung, and S. Pateux, “Preliminary results on motion vector prediction,” JCT-VC, Input document, Mar 2012.

[8] J. Jung, “Tool Experiment 11: Motion Vector Coding,” JCT-VC, Output document, Mar 2013.

[9] K. McCann, C. Rosewarne, B. Bross, M. Naccari, and G. J. S. K. Sharman, “High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description,” JCT-VC, Encoder Description, Jul 2014.

[10] 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.

[11] ——, “Monte-Carlo-based Parametric Motion Estimation using a Hybrid Model Approach,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, p. 1, 2012.

[12] 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.

[13] 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.

[14] R. Kordasiewicz, M. Gallant, and S. Shirani, “Encoding of Affine Motion Vectors,” *IEEE Transactions on Multimedia*, vol. 9, no. 7, pp. 1346 –1356, Nov 2007.

[15] 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.

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