Geometry-Adaptive Motion Partitioning Using Improved Temporal Prediction

Max Bläser [1], Cordula Heithausen and Mathias Wien

Institut für Nachrichtentechnik
RWTH Aachen University
Germany

[1] blaeser@ient.rwth-aachen.de

***Abstract*—Current state of the art video codecs such as HEVC are based on rectangular motion partitioning and compensation. To further enhance the compression performance, more flexible block partitioning strategies are needed. We present geometry based motion partitioning (GMP) in a post-HEVC framework with improved temporal prediction of GMP parameters. Our main contribution is a simple yet efficient temporal projection method for GMP parameters using available reference picture motion vectors, which allows the tracking of geometric partitioning lines along a motion trajectory. Average bit rate reductions by 1.35% are reported.**

***Index Terms*—video coding, motion compensation, block partitioning, geometric partitioning, inter prediction**

I. Introduction

Improved Inter-picture prediction is one of the driving factors for increased compression performance of today's video coding standards. While recent attention has been given to enhance motion compensation by using non-translational motion models [1] or optical-flow based optimizations [2], such methods still rely on rectangular block partitioning. It has been noted frequently however that such a method of block partitioning is sub-optimal for natural video as object borders rarely follow a purely rectangular pattern. Approximating motion boundaries by rectangular blocks therefore introduces a coding overhead.

A more flexible way of motion partitioning which allows arbitrarily shaped segments has been proposed Bläser et. al. in [3]. Although having the highest degree of freedom in terms of partitioning, the rate savings were comparably low at 0.6%. This was explained by the fact that the proposed scheme relied on simple, unsupervised spatio-temporal image segmentation, which is still an unsolved problem and thereby largely affecting the compression efficiency. In contrast to this work, we focus on a simplification of object-based block partitioning using a geometric model which allows the splitting of a motion compensated block by a straight line into two segments. This is denoted geometry-adaptive or geometric motion partitioning (GMP or GEO). Proposed in 2005 by Kondo et. al. [4], GMP has been researched in the context of AVC [5] and HEVC [6], [7]. Promising coding gains had been reported in the following investigations, leading to attempts to include GMP during the standardization phase of HEVC [8], [9].

Our GMP method differs from previous attempts in multiple aspects as it includes a flexible and efficient prediction of GMP parameters using a temporal projection. By scanning each reference picture motion vector field and translationally shifting the geometric partition boundaries with scaled motion vectors, projected locations of the partition boundaries can be derived. As only available reference pictures are needed for this process, it can be performed at the decoder. The paper is structured as follows: In Section II, we explain the proposed block partitioning method and specifically detail the process of temporal projection using motion vectors. We further present our approach on an implementation-driven representation of GMP parameters and give details about the coding aspects. In Section III comparative coding results and a statistical analysis of the GMP method are given and discussed. Section IV concludes the paper.

II. Geometric Motion Partitioning

Rectangular motion partitioning is the foundation of motion compensation in video codecs such as HEVC. While symmetric vertical and horizontal block partitioning has been available in AVC, HEVC introduced *Asymmetric Motion Partitioning* (AMP) [10], adding four more partitioning options. In GMP, such limitations are lifted and any partitioning by a straight line of a square block is allowed. However, significant coding cost has to be spent to signal the partitioning parameters.

A. GMP Line Derivation

Previous GMP methods mostly use a line partitioning described by two parameters, a radius and an angle, often denoted ρ and θ respectively, which model the partitioning by:

$$f(x,y) = x\ \cos\theta + y\ \sin\theta - \rho \tag{1}$$

In our novel approach however, we choose to represent

978-1-5386-0462-5/17/$31.00 ©2017 IEEE.

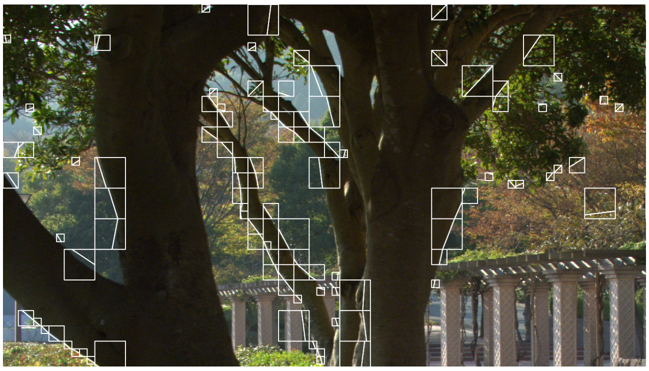

Fig. 1. Example of geometric motion partitioning as a block based coding tool for the *ParkScene* test sequence. In this sequence, the camera pans to the right.

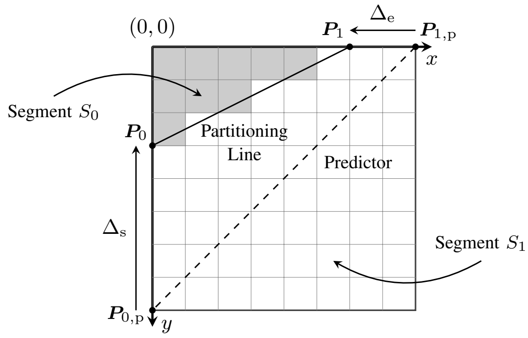

Fig. 2. Representation of geometric partitioning line using a predictor and two offset values.

the partitioning line, which splits a block into two distinct segments S_0 and S_1 (see Figure 2) using two points $\boldsymbol{P}_0 = [x_0, y_0]^{\mathrm{T}}$ and $\boldsymbol{P}_1 = [x_1, y_1]^{\mathrm{T}}$ located on the boundary of a given block, leading to the well known two-point form of a straight line:

$$(y - y_0)(x_1 - x_0) = (y_1 - y_0)(x - x_0) \tag{2}$$

This two-point representation has multiple benefits as it simplifies prediction, quantization and coding and allows for a direct integer-only implementation. A binary mask $M(x, y)$, which assigns each pixel of a given block to a specific segment, can now be easily derived using the two following equations:

$$f(x, y) = \begin{vmatrix} x_0 - x & y_0 - y \\ x_1 - x & y_1 - y \end{vmatrix} \tag{3}$$

$$M(x, y) \begin{cases} 0, & \text{if } f(x, y) \geq 0 \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

As direct coding of the two points $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ would consume too much rate, we use an approach combining both spatial and temporal prediction of the partitioning line. Thus, in the presented method, there are two distinct *GMP prediction modes* available:

- Spatial prediction
- Temporal prediction

For both cases – given a predictor line from a spatial context (e.g. neighborhood, rectangular partitioning modes) or from temporal projection – only integer start and end position offsets Δ_{s} and Δ_{e} from the line coordinates $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ and the predictor coordinates $\boldsymbol{P}_{0,\mathrm{p}}$ and $\boldsymbol{P}_{1,\mathrm{p}}$ need to be coded. This relationship is exemplified in Figure 2. A negative valued offset moves a point in mathematically positive sense along the block boundary and vice versa.

B. GMP Offset Quantization

In order to determine the optimal partitioning line at the encoder side, an iterative search is used. Given a set of initial partitioning parameters, in each step, the two offsets Δ_{s} and Δ_{e} are calculated as $\Delta_{\mathrm{s}} = \pm k_{\mathrm{step}} \cdot \Delta_{\mathrm{min}}$ and Δ_{e} accordingly, where k_{step} is a search range decreasing in each iteration (e.g. 8, 4, 2, 1) and Δ_{min} is the block-size dependent quantization step size of the minimum offset values. In our experiments presented in Section III, the quantization Δ_{min} was chosen as shown in Table I below. This quantization was optimized for an acceptable trade-off between encoder-complexity and coding performance.

TABLE I
QUANTIZATION OF GMP OFFSET VALUES.

Offset Quantization	**Block Size**			
	128×128	64×64	32×32	16×16
Δ_{min}	4	2	1	1

C. Spatial Prediction

Using spatial prediction of geometric partitioning predictors is needed for all cases where no geometric partitioning line can be found in the temporal neighborhood. In order to solve this problem, we use a similar approach as in [11]. In our method however, we derive predictor candidates from the six available HEVC rectangular partitioning modes and also add the two diagonals as a prediction option. As object boundaries are often smooth and continuous across block boundaries, there is also locally confined spatial correlation among the partitioning lines. This can be utilized for prediction by simply linearly extending the partitioning line of a neighboring block into the current block. As spatial prediction of GMP parameters is an established method, we refer to [12] for more details. For completeness, it is to be noted that also neighboring blocks using *Intra-prediction* may be used for prediction of GMP parameters as for example the directional Intra-modes contain information about the local texture gradients of the image content.

D. Temporal Prediction

It is to be expected that object boundaries often don't change their shape and are only subject to translational motion. Therefore, high temporal correlation between geometric partitioning patterns can be observed and exploited for prediction. Our method of temporal prediction is based on the simple notion that the block partitioning of an object is translationally shifted from picture to picture by the object motion. Thus, it is possible to motion-compensate the partitioning information and use it for prediction and coding of the current picture. In a practical application, this can be achieved by shifting the GMP parameters of reference pictures stored in the reference picture buffer using the motion vector field. As our GMP parameters are represented by two points, this can be achieved by a simple translation:

$$\boldsymbol{P}_{\mathrm{p,s}} = \boldsymbol{P}_{\mathrm{s}} + \boldsymbol{MV}_{\mathrm{p}}, \quad s = 0, 1 \tag{5}$$

$$\boldsymbol{MV}_{\mathrm{p}} = \begin{bmatrix} v_{\mathrm{x}} \\ v_{\mathrm{y}} \end{bmatrix} = -\frac{t_{\mathrm{b}}}{t_{\mathrm{d}}} \boldsymbol{MV} \tag{6}$$

Here, $\boldsymbol{P}_{\mathrm{p,s}}$ is the location of the projected GMP line parameters and $\boldsymbol{MV}_{\mathrm{p}}$ is the distance scaled motion vector. The scaling is performed according to the two parameters t_{d} and t_{b} which denote the time distance between the current picture and the first reference picture and the first and the second reference picture respectively. The method is exemplified in Figure 3. In our implementation, a search range centered around the current block of the reference picture motion vector field is scanned in a predefined order. For each motion vector which is encountered, the projection of collocated GMP parameters according to Eq. 5 is performed.

E. GMP Predictor List and Coding

When the projected GMP parameters have been determined, it is sufficient to test whether the projected line slices the current block at the collocated position and if so, add the GMP parameters to a candidate list of GMP predictors. As multiple lines may slice the current block, the candidate list will be expanded until a fixed limit of partitioning candidates is reached. At the encoder side, all available GMP predictors are subject to a rate-distortion optimization. The best available predictor from the predictor candidate list is signalled to the decoder by an index. At the decoder side, the entire projection process and list construction can be identically repeated as it is based on information contained in already decoded reference pictures. In summary, three syntax elements are needed to code the partitioning line:

1) GMP Prediction Mode Flag
2) GMP Predictor Index
3) GMP Offsets $(\Delta_{\mathrm{s}}, \Delta_{\mathrm{e}})$

The GMP Predictor Index is binarized using Truncated Unary code and each bin is coded by the *CABAC* engine. As the GMP Offsets are expected to follow distributions heavily centered at zero, coding them is performed using a combination of a Larger-Zero bin and Exponential Golomb Coding for the remainder.

III. Coding Results

We evaluated the proposed method using the *Joint Exploration Test Model* (JEM) of JVET [13], version 2.0, which is a test model for future video coding technology. This software already contains advanced methods of motion-compensation such as Overlapped Block Motion Compensation (OBMC), an affine motion model and bi-predictive optical flow (BIO) [2].

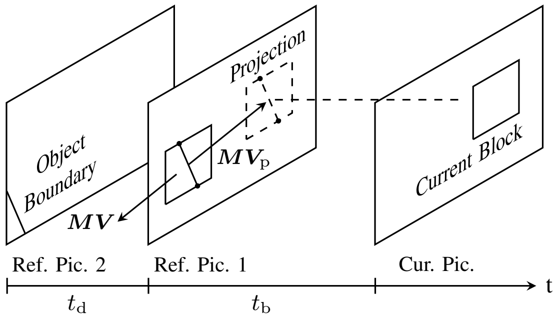

Fig. 3. Exemplified projection of a partitioning boundary from a reference picture using the motion vector field.

TABLE II
Simulation results for JEM 2.0 LDP configuration with GMP temporal projection on (TP) and off (No TP).

Sequence	Resolution	BD-rate change [%]	
		TP	No TP
Foreman	352×288	-2.86	-2.07
BlowingBubbles	416×240	-0.97	-0.74
BasketballPass	416×240	-0.63	-0.55
RaceHorses	832×480	-1.40	-1.29
BQMall	832×480	-1.57	-1.16
BasketballDrill	832×480	-0.67	-0.72
PartyScene	832×480	-0.41	-0.35
ChinaSpeed	1024×768	-0.65	-0.51
Johnny	1280×720	-2.46	-1.61
KristenAndSara	1280×720	-2.18	-1.55
FourPeople	1280×720	-1.18	-0.76
Vidyo1	1280×720	-1.61	-0.94
Vidyo3	1280×720	-1.74	-1.14
Cactus	1920×1080	-0.72	-0.47
Kimono	1920×1080	-0.48	-0-38
ParkScene	1920×1080	-1.40	-1.01
PeopleOnStreet	2560×1600	-0.77	-0.64
Traffic	2560×1600	-1.68	-1.00
Drummer	3840×2160	-2.18	-1.64
Average		**-1.35**	**-0.98**

The proposed method of geometry-adaptive motion partitioning has been implemented as a block based coding tool and is therefore competing against the already established symmetric and asymmetric partitioning options in HEVC. To enable GMP on the block level, an additional flag is signalled.

Simulations were run on a variety of sequences with different resolutions and varying amount of motion. We chose to test the *Lowdelay-P* encoder configuration of the Common Test Conditions (CTC) of JVET. Our results in terms of BD-rate savings compared to JEM 2.0 without GMP are given in Table II. From the results, an average rate reduction of 1.35% in case of using temporal projection and 0.98% for pure spatial prediction can be observed. It is to be noted that for "unchallenging" sequences in terms of their content such as Johnny or KristenAndSara, among the highest rate-reductions by 2.46% and 2.18% and also the highest improvements over spatial prediction can be achieved.

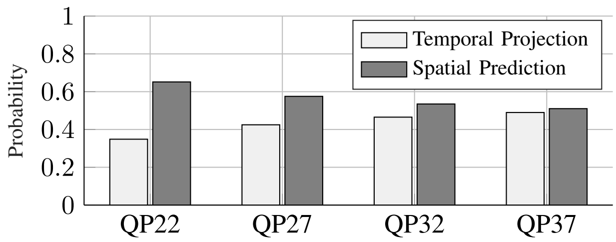

Fig. 4. Probabilities of the two GMP prediction mode usages measured over all tested sequences.

This can be explained by the fact that geometry-adaptive partitioning is better suited for video content containing objects with clearly defined edges such as heads and bodies and uniquely discernible motion (moving foreground, static background as in those test sequences). In Table III, we also provide an analysis how much area of the sequences were coded using geometric partitioning including temporal projection. Here, we see that on average roughly 5-8% of the tested sequences were coded using geometric partitioning. For the two best sequences Johnny and KristenAndSara, the mode usage was in a similar range with 6.6% and 6.8%. Further, it can be assumed that the projection process works better for smaller motions. This can be explained by analysis

TABLE III
GMP MODE USAGE OF ALL CODED PICTURES DEPENDING ON QP AND BLOCK SIZE FOR LDP USING TEMPORAL PROJECTION.

Block Size	**Coded in GMP mode [%]**			
	QP22	**QP27**	**QP32**	**QP37**
16×16	1.78	1.22	0.73	0.35
32×32	1.64	2.08	2.20	1.85
64×64	0.99	1.83	2.73	3.09
128×128	0.52	1.18	1.86	2.75
Total	**4.93**	**6.31**	**7.52**	**8.04**

of the GMP prediction mode distribution as seen in Figure 4. While the usage of spatial and temporal GMP prediction modes are quite balanced for the entire test set, they were more diverging for those sequences with the highest rate reductions: It was observed that temporal projection had a mode usage higher than on average, with 59% for Johnny and 51% for KristenAndSara.

Further, it is evident that the maximum number of iterations performed at the encoder for determining the RD-optimal partitioning line affects the coding performance. As motion estimation is required for each geometric block partitioning configuration, encoder complexity quickly becomes unmanageable when using too many iterations. Compared to JEM 2.0 without GMP, we can report a roughly 3× encoder complexity increase. However, our approach does not employ fast-decision strategies, such as a pre-analysis step of the texture or the option to skip the GMP mode estimation altogether e.g. for cases where it unlikely to outperform the *Skip mode* [10].

IV. Conclusion

This paper presents a geometry-adaptive motion partitioning method, competing against established rectangular block partitioning methods in a post-HEVC video coding framework. Our main contribution is a simple and implementation-friendly GMP parameter representation in combination with improved temporal prediction of the partitioning line. The temporal prediction is achieved by a projection process using the reference picture motion vector field. Average BD-rate saving on the used test set of 1.35% can be reported.

As Quad-Tree-Binary-Tree (QTBT) block partitioning has been integrated into version 4.0 of JEM, we wish to compare GMP to QTBT in the future.

In future work, temporal projection can be improved by also considering motion trajectories over longer temporal distances, involving multiple reference pictures. The rate distortion optimization of the partitioning line at the encoder can be improved by applying image segmentation methods in a pre-processing step.

References

[1] C. Heithausen, M. Bläser, M. Wien, and J. R. Ohm, "Improved higher order motion compensation in hevc with block-to-block translational shift compensation," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 2008–2012.

[2] A. Alshin and E. Alshina, "Bi-directional pptical flow for future video codec," in *2016 Data Compression Conference (DCC)*, March 2016, pp. 83–90.

[3] M. Bläser, C. Heithausen, and M. Wien, "Segmentation-based Partitioning for Motion Compensated Prediction in Video Coding," in *2016 Picture Coding Symposium*, Nuremberg, Germany, Dec. 2016.

[4] S. Kondo and H. Sasai, "A motion compensation technique using sliced blocks in hybrid video coding," in *IEEE International Conference on Image Processing 2005*, vol. 2, Sept 2005, pp. II–305–8.

[5] A. A. Muhit, M. R. Pickering, and M. R. Frater, "A fast approach for geometry-adaptive block partitioning," in *2009 Picture Coding Symposium*, May 2009, pp. 1–4.

[6] P. Bordes, E. Francois, and D. Thoreau, "Fast encoding algorithms for geometry-adaptive block partitioning," in *2011 18th IEEE International Conference on Image Processing*, Sept 2011, pp. 1205–1208.

[7] Q. Wang, X. Ji, M. T. Sun, G. J. Sullivan, J. Li, and Q. Dai, "Complexity reduction and performance improvement for geometry partitioning in video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 338–352, Feb 2013.

[8] P. Bordes, P. Chen, I.-K. Kim, L. Guo, H. Yu, and X. Zheng, "CE2: unified solution of flexible motion partitioning," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, Geneva, Switzerland, 5th meeting, Tech. Rep. JCTVC-E374, Mar. 2011.

[9] X. Zheng, H. Yu, S. Li, Y. He, and P. Bordes, "CE2: non-rectangular motion partitioning," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, Torino, Itlay, 6th meeting, Tech. Rep. JCTVC-F415, Jul. 2011.

[10] M. Wien, *High Efficiency Video Coding – Coding Tools and Specification*. Berlin, Heidelberg: Springer, Sep. 2014.

[11] Y. Chen, D. Mukherjee, J. an, and K. Rose, "Joint inter-intra prediction based on mode-variant and edge-directed weighting approaches in video coding," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 7372–7376.

[12] O. D. Escoda, P. Yin, C. Dai, and X. Li, "Geometry-adaptive block partitioning for video coding," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 1, April 2007, pp. I–657–I–660.

[13] Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, "Joint Exploration Test Model (JEM) 2.0." [Online]. Available: https://jvet.hhi.fraunhofer.de/