Geometry-based Partitioning for Predictive Video Coding with Transform Adaptation

Max Bläser, Jens Schneider, Johannes Sauer and Mathias Wien
Institut für Nachrichtentechnik
RWTH Aachen University
blaeser@ient.rwth-aachen.de

Abstract— **Rectangular block partitioning as it is used in state of the art video codecs such as HEVC can produce visually displeasing artifacts at low bitrates. This effect is particularly noticeable at moving object boundaries. This contribution presents a comprehensive geometry-based block partitioning framework in a post-HEVC codec for motion compensated prediction, intra-prediction and transform coding as a solution. The method is evaluated on the set of sequences defined by the Joint Call for Proposals on Video Compression with Capabilities beyond HEVC [1]. Our contribution aims at visually improving the quality of object boundaries and provides an objective BD-rate gain of 0.82% on average compared to the reference Joint Video Exploration Team (JVET) test model (JEM 7.0).**

Index Terms—**video coding, motion compensation, geometric partitioning, intra prediction, tranform coding**

I. Introduction

Improvements in inter-picture prediction have been one of the driving factors for increased video compression performance in the most recent developments lead by JVET, which is set to start on the next phase of standardization for video compression by issuing the Joint Call for Proposals on Video Compression with Capabilities beyond HEVC [1]. By comparison of new inter-prediction features, which have been integrated into the latest JVET test model (JEM 7.0) [2], two distinctions can be made: on the one hand, better block-based coding tools are available, which rely on advanced algorithms such as non-translational motion modeling, decoder side motion vector derivation or bi-directional optical flow [3]. On the other hand, the overall block structure of JEM has been significantly changed in comparison with HEVC. While HEVC was built on a quadtree block structure, with the option of splitting a leaf node coding block further into two or four rectangular prediction blocks, this paradigm has been replaced in JEM by the combination of a quadtree with a binary tree, which allows for even higher flexibility in segmenting each picture [4].

Although the compression performance of JEM compared to HEVC has been significantly increased by around 29% in terms of BD-rate savings [5], JEM still suffers from drawbacks prone to all block-based video compression schemes. Blocking artifacts at low bit rates may become visible due to rectangular block-based prediction and transforms, which is especially noticeable in textured areas and regions around moving object boundaries [6], [7]. In our contribution, we leverage Geometric Motion Partitioning (GMP) to mitigate these effects and demonstrate how GMP leads to more sharpness and less blocking artifacts at object boundaries. Our contribution also shows how GMP can be combined with non-rectangular transform coding and intra-prediction. Further, the GMP approach is harmonized with most other improved inter-prediction tools available in the JEM software.

In Section II, we briefly review GMP and explain the key design aspects of our comprehensive coding framework, such as the representation and coding of GMP parameters, the newly added features of shape-adaptive transform coding and intra-prediction used upon individual partitions. In Section III comparative coding results as well visual test results are presented and discussed. Section IV concludes the paper.

II. Geometric Motion Partitioning with QTBT

GMP, also abbreviated as GEO, has been extensively studied during the standardization of HEVC [8], [9] and in the context of AVC [10]–[13], where it showed promising results by giving improvements of up to 5% in terms of BD-rate reductions. In our previous contribution [14], we revisited GMP and evaluated its performance in an earlier version of the JEM test model (JEM 2.0), which is more similar with the HEVC standard in terms of its prediction and transform block structure. Although BD-rate reductions of 1% were reported in lowdelay P configuration, they were lower than expected by comparison with previous GMP implementations. This behavior is partially explainable by the fact that inter-prediction tools aiming at better motion representation, such as sub-CU-based motion vector prediction, bi-directional optical flow and decoder-side motion vector refinement [4] already improve block-based motion compensation substantially.

In this paper, as in our previous contribution [14], the proposed GMP approach operates as a block-based coding tool at the coding unit (CU) level. Thus, GMP does not replace the existing partitioning structure but provides an alternative partitioning mode, which is explicitly signaled to the decoder. The introduction of QTBT partitioning in JEM 7.0 provides an additional challenge for the integration of GMP with respect to the transform coding of the prediction

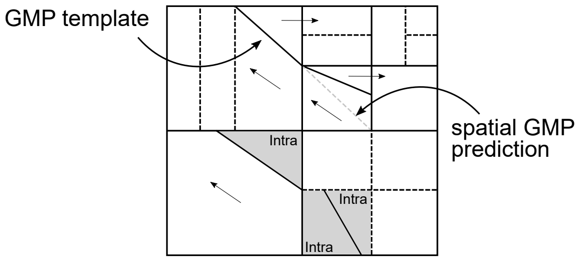
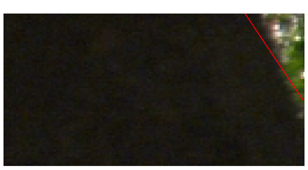


Fig. 1. Example of the proposed QTBT+GMP scheme. All leaf node coding units of the QTBT can be further geometrically partitioned by a straight line. Each partition can be inter- or intra-predicted.

error signal. An exemplified partitioning block partitioning, which is the combination of QTBT and GMP can be seen in Figure 1. In contrast to our previous contribution, GMP blocks can now be composed of inter-predicted and / or intra-predicted parts. While in HEVC, the residual signal of a coding unit can be further segmented using a separate residual quad-tree (e.g. transform tree), this feature has been waived in JEM versions $\geq$ 3.0 in favor of QTBT, where the size of the coding unit is identical to the size of the 2D block transform which may be applied to the residual. Although this has the advantage of a simpler design, it removes the ability to apply smaller subtransforms to those parts of the residual showing locally varying error distributions. The quantization of transform coefficients will therefore always affect the entire residual block. Figure 2 illustrates this issue for a large 128$\times$64 block, where the two motion compensated partitions show distinctly different distributions of the prediction error. In the example, only the smaller upper right partition contains significant energy, while the larger partition is well compensated. Applying a single Discrete-Cosine-Transform (DCT) of equivalent size and subsequent quantization would lead to an increased error in the reconstructed residual of the otherwise well compensated partition. In order to mitigate



(a) Original

(b) Residual

Fig. 2. Example of a large 128$\times$64 motion compensated GMP block (a), where the residual (b) shows a higher prediction error in one of the partitions.

this problem and locally isolate the residual for each segment, we propose to use the well known Shape-Adaptive-DCT (SADCT) introduced by Sikora and Makai [15] and refined for the application to video object coding in MPEG-4 as the so called Δ-SADCT by Kauff and Schüür [16]. We adapted the Δ-SADCT basis functions for usage with geometric partitions and subsequent HEVC transform coding within the JEM software.

Next to SADCT-based transform coding, we introduce usage of mixed intra- and inter-picture prediction within a GMP coded block as exemplified in Figure 1. Unlike [17], no averaging of inter-coded and intra-coded partitions is performed. Due to the potentially high encoder complexity needed to test all available intra-prediction options of JEM, only a simplified method of planar prediction is used for geometric partitions, further detailed down blow.

A. GMP Parameters

In the proposed approach, the partitioning line, splitting a block into two distinct segments, is parametrized by 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 two-point form of a straight line:

$$(y - y_0)(x_1 - x_0) = (y_1 - y_0)(x - x_0) \tag{1}$$

This two-point representation 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 S_0 or S_1, can be easily derived using the two following equations:

$$f(x, y) = \det \begin{pmatrix} x_0 - x & y_0 - y \\ x_1 - x & y_1 - y \end{pmatrix} \tag{2}$$

$$M(x, y) \begin{cases} 0, & \text{if } f(x, y) \geq 0 \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

As explicit coding of the coordinates $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ would consume too much rate, we use an approach to either signal spatial/temporal prediction of partitioning parameters or prediction from a fixed set of partitioning templates. Thus, there are two distinct GMP coding modes available:

1) Template-based coding
2) Temporal- or spatial prediction-based coding

Template-based coding may be used if no temporally or spatially neighboring blocks use GMP or if the block size is below a pre-defined threshold, e.g. if the width and height of the luminance component is smaller than 16 pixels. A list of 16 predefined templates is available, which are shown in Figure 3. Next to the list of templates, temporal and spatial prediction-based coding generates a second list of GMP candidates. These are derived by computing the intersection of neighboring partitioning lines with the current block boundaries (see Figure 1) or by temporal projection of GMP parameters from reference pictures. We refer to [14] for more details. For larger blocks,

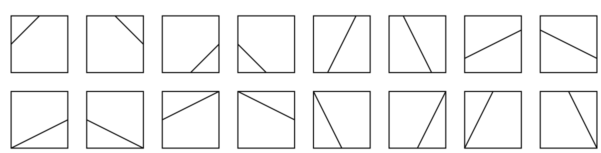


Fig. 3. The 16 predefined templates used for GMP prediction and coding. Note that the templates will be scaled accordingly for non-square blocks.

an additional refinement coding of the partitioning line may be applied to the GMP predictor. Table I summarizes all coding options for GMP parameters and also shows how parameters are binarized for the respective coding option.

TABLE I
CODING OPTIONS OF GMP PARAMETERS DEPENDING ON THE BLOCK SIZE.

Block size condition for width (w) and height (h)			
$w \leq 16$ & $h \leq 16$		$w > 16$ & $h > 16$	
template coding	FL(4)	template coding optional refinement	FL(4) EG(0)
		spatial / temporal prediction optional refinement	TU(8) EG(0)

FL(k) indicates a fixed-length code of k bits length, TU(k) a truncated-unary and EG(k) an Exp.-Golomb code of k^{th}-order respectively.

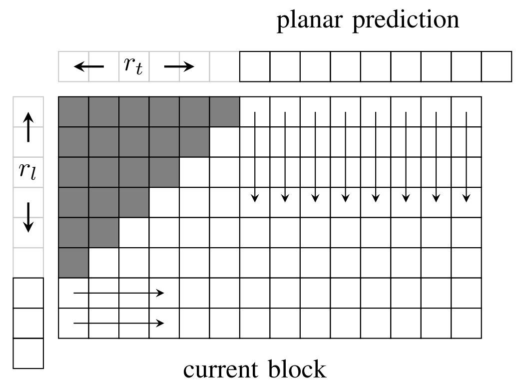


Fig. 4. Modified planar prediction with reference sample extension for GMP blocks.

B. Transform Coding using Shape-adaptive DCT

The SADCT operates on the basic principle of first applying a N-DCT on the columns of length N of the residual, followed by a horizontal M-DCT on the rows of the intermediate vertical transform. The column and row sizes N and M are inferred from the shape of the residual, which is given by the partition mask $M(x, y)$ in Equation 3. The SADCT transform coefficients are then quantized in the same manner as DCT coefficients and entropy coded. Here, the fact that the SADCT yields the same number of transform coefficients as pixels belonging to the segment (e.g. shape) gives the potential coding gain. The entropy coding of DCT coefficients in JEM has been adapted for SADCT coefficients accordingly. As each GMP block consists of two segments, two separate SADCT transforms are computed - one for each segment. It is noted that the encoder is tasked with determining which SADCT is retained based on a RD criterion. The option to apply the standard DCT-II on a GMP block is also available as this is beneficial for blocks with a uniform error distribution across both segments.

C. Intra-Prediction for Geometric Partitions

Intra-prediction tools have been extended in JEM by doubling the number of available prediction directions, adding new prediction filters and introducing a linear cross-component prediction mode. Additionally, intra-mode dependent non-separable secondary transforms can be applied to the intra residual [4]. For GMP partitions using intra-prediction however, planar prediction as used in HEVC and JEM is applied. The planar prediction has been modified by taking into account which reference boundary samples more likely belong to the same partition. This is exemplified in Figure 4. The reference samples denoted r_t and r_l likely belong to the shaded partition and thus shall not be used to perform planar prediction on the other segment. Otherwise, content associated with the shaded partition could propagate into the other segment. As with inter-predicted segments, the SADCT can be optionally applied to decorrelate the intra residual.

D. Encoder Design

It is well known that GMP can significantly increase encoder complexity [10], [18], due to the need of performing motion estimation (ME) for each potential GMP configuration. Similar to other published approaches, an iterative search for determining GMP parameters was developed, basing the partitioning decision on the lowest achievable RD-costs per block. Resembling the processing used by rectangular inter-prediction, the steps of ME and coding mode decision (*MERGE* vs. *AMVP*) are decoupled from the core transform mode decision (*DCT-II* vs. *SADCT*). To summarize, the following steps are performed for every potential leaf CU of the QTBT at the encoder side:

1) Perform full-pel ME for all GMP templates and save best template based on RD-cost.
2) Generate list of temporal and spatial predictors.
3) Perform refinement by full-pel ME around GMP parameters given by best template and first temporal / spatial predictor. Determine best GMP prediction mode.
4) For refined GMP line parameters with lowest RD-cost, perform quarter-pel ME and test intra-prediction.
5) Estimate core transform mode for final GMP block.

Additionally, several techniques are employed to fast skip the GMP estimation for blocks that meet one or more of the following conditions:

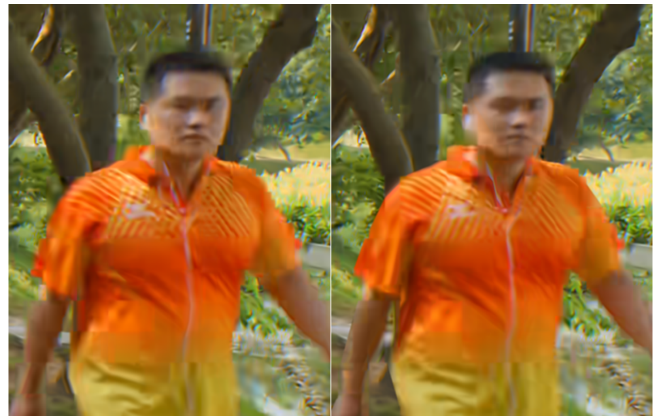
1) Skip GMP if RD-cost ratio for *SKIP* and *INTER* mode is lower than threshold, e.g $R_{\text{SKIP}}/R_{\text{INTER}} < \theta_{\text{th}}$ and $\theta_{\text{th}} < 0.85$.
2) Skip if block size is smaller than 8×8.
3) Skip uniform textured block measured by low pixel variance as in [19], e.g. $\sigma \leq 8 \cdot 2^{B-8}$, where σ is the weighted pixel variance and B the internal bitdepth.

The current prototype GMP implementation nevertheless increases encoder complexity of JEM 7.0 by a factor of 3.8. We acknowledge that this can be improved in the future.

III. EXPERIMENTAL RESULTS

Results are reported for an implementation of GMP in JEM 7.0. The proposed method was tested on HD and UHD sequences belonging to the SDR testset issued by the Joint Call for Proposals on Video Compression with Capability beyond HEVC.

(a) MarketPlace, RA, R1



(b) ParkRunning3, RA, R1

Fig. 5. Visual results comparing JEM 7.0 + GMP (left) against JEM 7.0 (right). Next to the sequence name, the frame number, encoder configuration and rate point is shown.

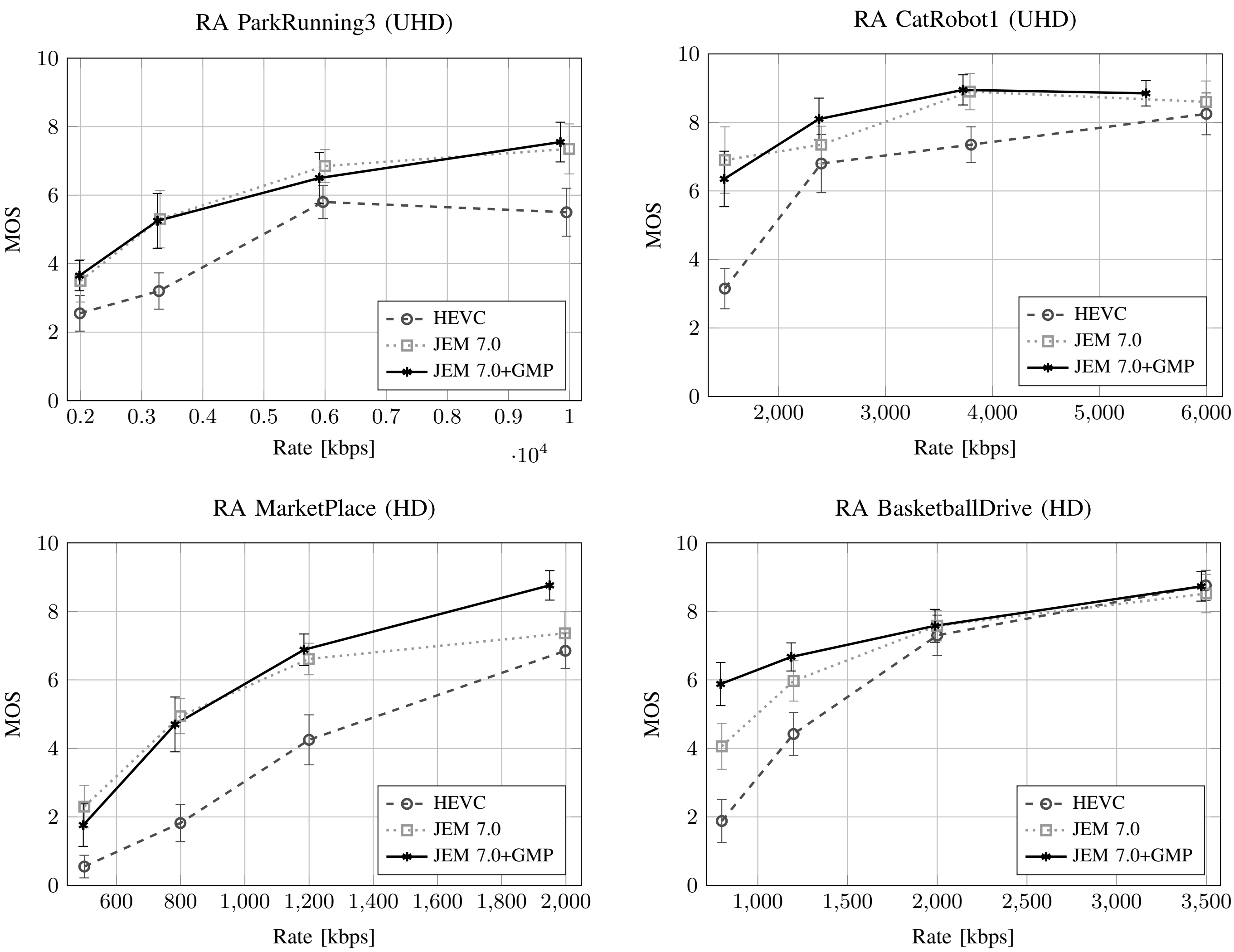


Fig. 6. Subjective test results showing rate vs. MOS plots including 2σ confidence intervals for selected sequences in the randomaccess category. ParkRunning3 and MarketPlace are typical examples of average performance, whereas CatRobot1 and BasketballDrive represent the best overall results for UHD and HD resolution. All subjective results can be found on: http://www.ient.rwth-aachen.de/cms/cfp-2018/

The coding performance is measured using the Bjøntegaard Delta rate change [20] against JEM 7.0 without GMP. Although the impact on the objective coding performance is small and sequence-dependent (shown in Table II), the method improves visual quality around moving object boundaries. Figure 5 shows two typical example patches for the lowest rate-point R1 considered by the Joint Call for Proposals (CfP) [1], where the impact of GMP on object boundaries can be assessed best. Particularly, natural objects with diagonal object boundaries (such as people) benefit from GMP, which is visible at the shoulder and head regions in *MarketPlace* and *ParkRunning3*. In the shown examples, object boundaries appear less blocky, sharper and more details of the background are revealed. Extensive visual tests were conducted as part of the CfP [21]. Test subjects were tasked to assess the visual quality of each sequence and rate-point of the proposal against two references (JEM 7.0 and HEVC) using an Absolute Category Rating from 0 (worst) to 10 (best). On average, the proposal performed very similar to JEM 7.0 with regard to the Mean Opinion Score (MOS) (JEM 7.0+GMP: 5.96, JEM 7.0: 6.01), with a slightly worse performance for UHD and a better performance for HD sequences compared to the JEM 7.0 anchor. Figure 6 shows selected rate vs. MOS plots: on the left, two examples which are considered of average performance are shown and two on the right which are the top performing results for UHD and HD resolution. For HD sequences in the RA category, GMP provides a measurable visual benefit.

TABLE II
CODING RESULTS FOR RANDOM ACCESS (RA) AND LOWDELAY (LD) CONFIGURATION.

Sequence	Class	Δ-BD-Rate	
		RA	LD
BQTerrace	HD	−0.21 %	−0.78 %
BasketballDrive	HD	−0.09 %	+0.04 %
Cactus	HD	+0.07 %	−0.47 %
MarketPlace	HD	−1.47 %	−2.40 %
RitualDance	HD	−0.62 %	−0.58 %
Campfire	UHD	−1.06 %	
CatRobot1	UHD	−1.21 %	
DaylightRoad2	UHD	−1.48 %	
FoodMarket4	UHD	−0.74 %	
ParkRunning3	UHD	−1.14 %	
Average		−0.79%	−0.84 %

IV. CONCLUSION

In this paper an alternative method of block partitioning is presented, which comprises geometry-based prediction and transform coding. The method can reduce coding artifacts and improve visual quality, especially at low bitrates. Boundaries of moving objects often appear sharper and more natural. Further, GMP provides a slight improvement in coding efficiency. The method can be improved by lowering encoder complexity and better harmonization with existing coding tools such as affine motion compensation. Further, improvements with regard to the coding of SADCT transform coefficients and GMP parameters may boost the performance of GMP.

REFERENCES

[1] A. Segall, V. Baroncini, J. Boyce, J. Chen, and T. Suzuki, "Joint call for proposals on video compression with capability beyond HEVC," Macao, CN, 8th meeting, Tech. Rep. JVET-H1002, Oct. 2017.

[2] "Joint Exploration Test Model, version 7.0," https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/tags/HM-16.6-JEM-7.0, 2017.

[3] E. Alshina and K. Choi, "Samsung's response to Joint CfE on Video Compression with Capability beyond HEVC (SDR category)," Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, Torino, Italy, 7th meeting, Tech. Rep. JVET-G0029, Jul. 2017.

[4] J. Chen, E. Alshina, G. Sullivan, J.-R. Ohm, and J. Boyce, "Algorithm Description of Joint Exploration Test Model 7 (JEM 7)," Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, Torino, Italy, 7th meeting, Tech. Rep. JVET-G1001, Jul. 2017.

[5] M. Karczewicz and E. Alshina, "JVET AHG report: Tool evaluation (AHG1)," Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, Torino, Italy, 7th meeting, Tech. Rep. JVET-G0001, Jul. 2017.

[6] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 576–584.

[7] R. Lin, Y. Zhang, H. Wang, X. Wang, and Q. Dai, "Deep convolutional neural network for decompressed video enhancement," in *2016 Data Compression Conference (DCC)*, March 2016, pp. 617–617.

[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] 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.

[11] 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.

[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] C. Dai, O. D. Escoda, P. Yin, X. Li, and C. Gomila, "Geometry-adaptive block partitioning for intra prediction in image video coding," in *2007 IEEE International Conference on Image Processing*, vol. 6, Sept 2007, pp. VI – 85–VI – 88.

[14] M. Bläser, C. Heithausen, and M. Wien, "Geometry-Adaptive Motion Partitioning Using Improved Temporal Prediction," in *2017 Visual Communications And Image Processing*, St Petersburg, USA, Dec. 2017.

[15] T. Sikora and B. Makai, "Shape-adaptive dct for generic coding of video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 1, pp. 59–62, Feb 1995.

[16] P. Kauff and K. Schuur, "Shape-adaptive dct with block-based dc separation and delta dc correction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 3, pp. 237–242, Jun 1998.

[17] 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.

[18] 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.

[19] 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.

[20] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16/Q6 VCEG, Austin, USA, Tech. Rep. VCEG-M33, 2001.

[21] V. Baroncini, "Results of Subjective Testing of Responses to the Joint CfP on Video Compression Technology with Capability beyond HEVC," Joint Video Experts Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, San Diego, USA, 10th meeting, Tech. Rep. JVET-J0080, Jul. 2018.