

EDGE-DIRECTED GEOMETRIC PARTITIONING FOR VERSATILE VIDEO CODING

Xuwei Meng*, Xinfeng Zhang[†], Chuanmin Jia*, Xia Li[‡], Shanshe Wang*, and Siwei Ma*

*Institute of Digital Media, Peking University, Beijing, China

[†]University of Chinese Academy of Sciences, Beijing, China

[‡]Shenzhen Graduate School, Peking University, Shenzhen, China

Email: {xwmeng, cmjia, ethanlee, sswang, swma}@pku.edu.cn, xfzhang@ucas.ac.cn

ABSTRACT

To improve the coding performance, geometric partition (GEO) was proposed for the upcoming VVC standard. GEO provides 140 partition candidates. The index of optimal GEO mode needs to be signaled explicitly. Considering different structural characteristics of different CUs and the correlation between spatial adjacent blocks and temporal collocated blocks, we propose a GEO mode prediction strategy by constructing a Most Probable Mode (MPM) list to reduce the overhead of GEO index and improve coding efficiency. Based on the observation of the high correlation between the partition mode and object boundaries, an edge-directed geometric partition scheme is proposed to construct the MPM list according to spatio-temporal edge information. The proposed method provides an objective BD-rate gain of 0.58% and 1.00% on average for RA and LDB configurations compared to VTM-6.0. Besides, it also promotes the visual quality of object boundaries.

Index Terms— Edge detection, MPM, geometric partition, inter prediction, video coding

1. INTRODUCTION

The block-based coding structure has been recognized as the core of the state-of-the-art video coding standards, such as H.264/AVC [1] and H.265/HEVC [2], because of its capability in achieving high compression efficiency. In AVC, a tree-structured variable block-size partition was employed for motion estimation (ME) and motion compensation (MC) to approximate the motion vector (MV) fields within each 16×16 macroblock (MB). To further improve the flexibility of block partition, quadtree-based (QT) block partition structure [3] and non-square QT structure [4] were introduced in HEVC. Then, the quadtree with a nested multi-type tree (QT-MTT) using binary tree and ternary tree partition structure [5] replaced the conventional QT partition in Versatile Video Coding (VVC), a new video coding standard, developed by Joint Video Experts Team (JVET) [6] since October 2017.

Although the partition structure becomes more flexible and efficient, the rectangular partition in the current video

coding standard is not able to represent the arbitrary shape of an actual object with a uniform MV in a rectangular coding block. This constraint limits the performance of inter prediction and may cause discontinuity on object edges. To address this issue, many non-rectangular partition methods were developed in the context of AVC and HEVC. To improve the performance of motion compensation, a wedge-based geometric partition (WBP) approach [7] was proposed to generate a better partition of the discontinuous MV fields by splitting a 16×16 MB into two wedge-shaped parts using a straight line. In which, the encoder performed ME on each part to find the best-matched corresponding reference block, and each of the two resulting wedge-shaped parts was predicted with an associated MV. This kind of non-rectangular partition method could achieve high coding efficiency, but also dramatically increase the encoder complexity since more recursive splitting iterations were required. Although many optimization approaches [8–10] were proposed to reduce the encoder complexity, there was still a long way for this technology to be adopted by video coding standards because of its lousy trade-off between the complexity and coding efficiency.

In VVC, these partition methods have also been investigated. In [11, 12], geometric partition based on the QT-MTT partition structure was proposed, which is shown in Fig. 1. Although the complexity still remained comparable with that in AVC and HEVC, the coding performance was less efficient because of the more flexible QT-MTT partition structure adopted in VVC. To achieve a better trade-off between the performance and complexity, JVET-L0124 [13], which has been adopted in VVC, proposed only to apply triangle partition (TPM) for merge mode. By doing so, the most time-consuming process, ME for geometric partition, was removed in the encoder. The adoption of TPM improved the representation capability of inter prediction, but the arbitrary object shape was still challenging to approximate with the pre-defined triangular and rectangular coding unit (CU). To further improve coding efficiency, a new geometric partition scheme (GEO) [14] for merge mode with up to 140 partition candidates was proposed. In GEO, one partition mode was selected for a CU from the pre-defined 140 candidates, and an

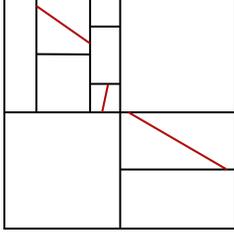


Fig. 1. An example of QT-MTT based geometric partition scheme. All leaf node coding units of QT-MTT can be further geometrically partitioned by a straight line.

index (0 ~ 139) was signaled to indicate the selected mode. Due to different structural characteristics of different CUs and the correlation between spatial adjacent blocks and temporal collocated blocks, it was unnecessary to use 140 partition modes for each CU. Moreover, the side information of GEO index (up to 8 bits) was also a significant burden of overhead which existed the potential for optimization.

In this paper, we propose a GEO mode prediction strategy by constructing a Most Probable Mode (MPM) list in the encoder and the decoder to reduce the overhead of GEO index and improve coding efficiency. Based on the observed high correlation between the partition mode and object boundaries, an edge-directed geometric partition scheme (E-GEO) is proposed to construct the MPM list according to spatio-temporal predicted edge information. The main contributions are summarized as follows,

- We propose a GEO mode prediction strategy by using an MPM list in encoder and decoder. To the best of our knowledge, it is the first MPM-based geometric partition framework for inter prediction.
- We propose a spatio-temporal edge detection method to construct the MPM list. Spatial and temporal pixels are jointly used to estimate the edge map of current CU.

2. EDGE-DIRECTED GEOMETRIC PARTITION

In this section, we firstly review the conventional GEO algorithm, then analyze the distribution of geometric partition in the decoder, which expertly guides us to design the proposed framework. Finally, the proposed edge-directed geometric partition method is elaborated.

2.1. Geometric partition framework

To achieve a better trade-off between coding efficiency and the computational complexity, the GEO partition scheme proposed in this paper is only applied for merge mode. The first step of GEO is the Merge Candidate List construction like other merge modes, such as regular merge mode and affine merge mode. Secondly, the best partition mode is selected

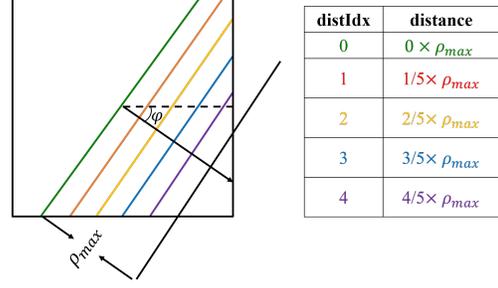


Fig. 2. Distance offset of GEO partition mode.

from the GEO mode set for each CU. And the best Motion Vectors corresponding to the two partitions of each CU are determined from Merge Candidate List. Then, motion compensation is conducted by using the best MVs derived in the previous step, and geometric partition edge blending is performed according to the selected partition mode to get the final reconstructed CU. Details will be shown in Section. 2.1.1 and Section. 2.1.2.

2.1.1. GEO mode set

The split boundary of geometric merge mode is described by the angle φ_i ($i = 0, 1, \dots, (I - 1)$) and the distance offset ρ_j ($j = 0, 1, \dots, (J - 1)$). Here, φ_i represents a quantized angle between 0 and 360 degrees with step 11.25 degree ($I = 32$). The distance offset ρ_j represents a quantized offset of the largest ρ_{max} with a fixed step ($J = 5$). In addition, split directions overlapped with QT-MTT and TPM splits are excluded. Hence, there are 140 modes in total.

For the case that φ is equal to 0, ρ_{max} is equal to $w/2$. If φ is equal to 90 degree, ρ_{max} is set to be $h/2$. The 1.0 sample backshift is to avoid that the split boundary is too close to the corner. When $0 < \varphi < \frac{\pi}{2}$, ρ_{max} can be formulated as [14],

$$\rho_{max}(\varphi, w, h) = \cos(\varphi) \times \left(\frac{h}{2 \tan(\frac{\pi}{2} - \varphi)} + \frac{w}{2} \right) - 1.0, \quad (1)$$

where h and w are height and width of the current CU, respectively. As shown in Fig. 2, for a given φ , ρ_{max} is calculated according to Eq. (1). Then, 5 distances corresponding to $0, 1/5 \times \rho_{max}, 2/5 \times \rho_{max}, 3/5 \times \rho_{max}, 4/5 \times \rho_{max}$ are derived.

2.1.2. Blending along the geometric partition edge

After predicting each non-rectangular partition using its own motion vector (MV_0 for P_0 and MV_1 for P_1), blending is applied to the two prediction blocks to derive samples around the partition edge, which is shown in Fig. 3. The blending process can be represented as follows,

$$P_B = (w \times P_0 + (8 - w) \times P_1 + 4) \gg 3, \quad (2)$$

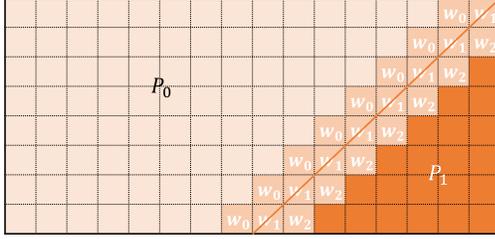


Fig. 3. An example of blending process for 16×8 CU.

where P_0 and P_1 are blocks reconstructed by using MV_0 and MV_1 , respectively. w corresponding to w_0 , w_1 and w_2 in Fig. 3 is the blending weight, derived by using look-up tables [14]. For the non-blending area, the samples in P_0 or P_1 are used as the final reconstructed samples directly. For the blending area, reconstructed samples are derived by Eq. (2).

2.2. GEO mode analysis

To analyze the influencing factors of GEO mode decision, we estimate the mode selection frequency in decoder, as shown in Fig. 4. GEO is selected mostly on the object edge, and the partition mode has a strong correlation with object boundaries. As we mentioned before, the large quantity of overhead caused by the GEO partition index is the primary reason for the limited coding efficiency. So, we propose to guide the MPM list construction using object edge information.



(a) *RacehorsesC* (QP=37, the 58th frame, RA)



(b) *BQMall* (QP=32, the 110th frame, LDB)

Fig. 4. Examples of GEO.

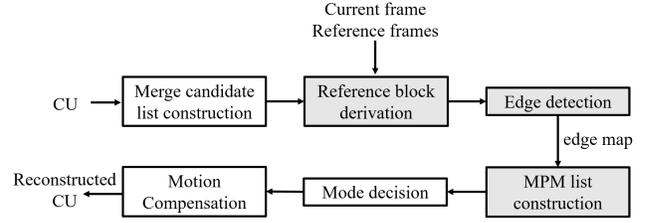


Fig. 5. The framework of the proposed E-GEO scheme in encoder, with the gray boxes highlighting our contributions.

2.3. Edge-directed geometric partition framework

To derive the MPM of current CU, edge estimation is needed in the proposed method. Moreover, as the original frame is unavailable at decoder, we can only use the reconstructed area of the current frame and reference frames to estimate the edge of current CU. As shown in Fig. 5, spatio-temporal based reference block derivation process is conducted firstly. Then guided filter [15] and canny detector [16] are used for edge detection. Finally, the MPM list is constructed according to the predicted edge map.

2.3.1. Reference block derivation

The reference block is derived according to spatio-temporal information. First, we derive the n -th MV (MV_n) from GEO merge candidate list, and round it to integer-pixel accuracy to avoid the interpolation process. Then, MSD (Mean of Squared Difference) between the template area of current CU (T) and that of its reference block (T^n) is used as the similarity evaluation criteria, as shown in Fig. 6. CU and T represent the current CU and its template area. CU^n and T^n are the reference block and its corresponding template area. Finally, the block with the smallest MSD is selected as the reference block of the current CU.

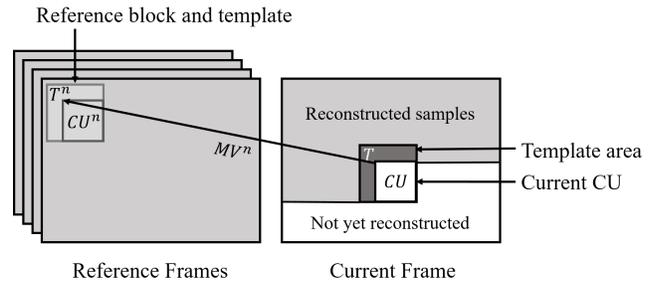


Fig. 6. Reference block derivation process.

2.3.2. Edge detection

To reduce quantization noise and preserve object boundaries, we use the guided filter to replace the widely-used Gaussian filter in the canny detector, firstly. Then we conduct the remaining steps in the canny detector (Sobel operator, non-maximum suppression and double threshold) to derive the edge map. The guided filter process is as follows,

$$q = a_k I + b_k, \quad (3)$$

$$a_k = \frac{\sigma_k^2}{\sigma_k^2 + \epsilon}, \quad (4)$$

$$b_k = (1 - a_k)u_k, \quad (5)$$

$$\epsilon = \epsilon \times (1 \ll \text{bitDepth})^2, \quad (6)$$

where I is the input reconstructed image, q is the output of the guided filter. σ_k is the variance of image I . u_k is the mean of I . bitDepth is the internal bit-depth, which is 10 in VVC. ϵ represents the filter strength, and ϵ ($\epsilon \in [0, 1]$) is the parameter needed to be trained. Other than ϵ , two thresholds, thre_{high} and thre_{low} , used in the double-threshold process of the canny detector also need to be trained offline. The actual high threshold (THRE_{HIGH}) used in the canny detector is the k_{th} value in the list which is constructed by sorting the block amplitude map after non-maximum process according to the amplitude value from small to large, where k is $w \times h \times \text{thre}_{high}$. The actual low threshold is calculated by $\text{THRE}_{HIGH} / \text{thre}_{low}$.

2.3.3. MPM list construction

After the reference block derivation and edge detection, a predicted edge map corresponding to the amplitude of Sobel gradient and an angle map corresponding to the direction of Sobel gradient are derived for the MPM list construction. The

Table 1. MPM list construction.

MPMIdx	AngleIdx	DistIdx
0	bestAngIdx	bestDistIdx
1	bestAngIdx	bestDistIdx - 1
2	bestAngIdx	bestDistIdx + 1
3	bestAngIdx - 1	bestDistIdx
4	bestAngIdx - 1	bestDistIdx - 1
5	bestAngIdx - 1	bestDistIdx + 1
6	bestAngIdx + 1	bestDistIdx
7	bestAngIdx + 1	bestDistIdx - 1
8	bestAngIdx + 1	bestDistIdx + 1
9	default	default
...
M	default	default

first step in MPM list construction process is to derive the best partition mode, corresponding to angle (bestAngIdx) and distance (bestDistIdx), with the largest average amplitude. Assuming the length of MPM list is M , the MPM list construction process is shown in Table. 1. If the selected candidates do not fulfill the MPM list, a default value will be added in it. In this paper, we set the default value to be 0.

Table 2. Variance of edge detection parameters.

	ϵ	thre_{high}	thre_{low}
PSNR-var	0.000529	0.000571	0.000359
bitRate-var	112.95	119.75	108.91

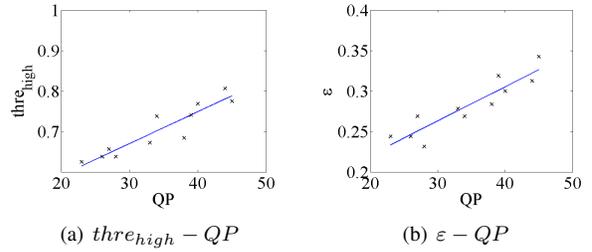


Fig. 7. Relationship between parameters and QP.

2.3.4. Parameter training

The parameter ϵ used in the guided filter, thre_{high} and thre_{low} applied in the canny detector are trained as follows,

- 1) In this paper, we choose four non-test sequences (*BasketballPass*, *BQSquare*, *BlowingBubbles* and *RaceHorses*), four typical QPs (22, 27, 32, 37) and two configurations (RA and LDB) for parameter training. The coding performance is tested with different parameter sets.
- 2) According to the coding performance derived in previous step, we calculate the PSNR-variance and bitRate-variance by changing one parameter with the other two parameters fixed. As shown in Table. 2, the parameter with larger variance represents that it has larger influence on coding efficiency. Based on this rule, we find that the most important parameter is thre_{high} , the second one is ϵ and the third one is thre_{low} .
- 3) For each B-frame, we experimentally select the optimal thre_{high} corresponding to the highest PSNR of Y component, which is plotted in Fig. 7(a). Because of the QPOffset for different B-frame, there may be several QPs for one test configuration. Hence, the number of points in Fig. 7 (a) is more than 4. It is clear that there

Table 3. Experimental results of the proposed E-GEO, Anchor: VTM-6.0.

Class	Sequence	GEO [14]		Proposed E-GEO	
		RA	LDB	RA	LDB
Class A1	<i>Tango2</i>	-0.23%	-	-0.30%	-
	<i>FoodMarket4</i>	-0.11%	-	-0.16%	-
	<i>Campfire</i>	-0.00%	-	-0.08%	-
Class A2	<i>Catrobot</i>	-0.31%	-	-0.41%	-
	<i>DaylightRoad2</i>	-0.19%	-	-0.32%	-
	<i>ParkRunning3</i>	-0.13%	-	-0.26%	-
Class B	<i>MarketPlace</i>	-0.18%	-0.51%	-0.37%	-0.74%
	<i>RitualDance</i>	-0.53%	-0.47%	-0.60%	-0.73%
	<i>Cactus</i>	-0.25%	-0.39%	-0.47%	-0.64%
	<i>BasketballDrive</i>	-0.22%	-0.30%	-0.36%	-0.46%
	<i>BQTerrace</i>	-0.15%	-0.02%	-0.24%	-0.24%
Class C	<i>BasketballDrill</i>	-0.46%	-1.18%	-0.76%	-1.39%
	<i>BQMall</i>	-1.79%	-1.46%	-2.13%	-1.75%
	<i>Partyscene</i>	-0.32%	-0.46%	-0.52%	-0.70%
	<i>RaceHorsesC</i>	-1.34%	-0.92%	-1.70%	-1.34%
Class E	<i>FourPeople</i>	-	-0.70%	-	-1.04%
	<i>BQJohny</i>	-	-1.48%	-	-1.86%
	<i>KristenAndSara</i>	-	-0.48%	-	-1.07%
Average	Class A1	-0.11%	-	-0.18%	-
	Class A2	-0.21%	-	-0.33%	-
	Class B	-0.27%	-0.34%	-0.41%	-0.56%
	Class C	-0.98%	-1.00%	-1.28%	-1.30%
	Class E	-	-0.89%	-	-1.32%
	Overall	-0.41%	-0.70%	-0.58%	-1.00%

is a positive relation between QP and the corresponding best $thre_{high}$. So we fit a linear model to represent their correlation. The fitted line, as shown in Fig. 7(a), can be formulated as,

$$thre_{high} = 0.0079 \times QP + 0.4344. \quad (7)$$

- 4) We train the parameter ε as step 3) with $thre_{high}$ fixed using Eq. (7). As shown in Fig. 7(b), the linear model for $\varepsilon - QP$ can be formulated as,

$$\varepsilon = 0.0042 \times QP + 0.1363. \quad (8)$$

- 5) Finally, parameter $thre_{low}$ is trained with $thre_{high}$ and ε fixed using Eq. (7) and Eq. (8). As there is no obvious rule for this parameter, we set it to be 3.0 in this paper.

3. EXPERIMENTAL RESULTS

In this section, we test the performance of the proposed E-GEO algorithm in VTM-6.0 [17]. The videos in the common test condition [18] are utilized as test sequences in our experiment under two configurations, Random Access Main10 (RA) and Low-Delay B Main10 (LDB). The first two seconds



Fig. 8. Subjective results of the proposed E-GEO (*PartyScene* and *RaceHorsesC*, QP=37, LDB)



Fig. 9. An example of the proposed E-GEO. (*RaceHorsesC*, QP=37, the 58th frame, RA)

of these sequences are encoded for performance evaluation. And the coding performance is measured by Bjontegaard's method [19] (Y-component) in terms of BD-rate.

Table. 3 summarizes the results of the proposed method. Two tests are conducted, GEO (CE4-1.1) [14] and the proposed E-GEO. It can be seen that the proposed method can achieve 0.58% and 1.00% bit-rate reduction for RA and LDB, which improves the coding gain of GEO by 41% and 43%. Note that the bit-rate reduction of class E in LDB configuration is higher. The reason is that the sequences of class E are conference videos with stable camera. In that case, the edge detection method is more accurate. E-GEO can also improve the subjective performance, as shown in Fig. 8. Compared to VTM-6.0, E-GEO can improve subjective quality obviously, especially on the boundaries of moving objects. We also

analyze the partition mode in the decoder, which is shown in Fig. 9. In which, red squares represent that the selected GEO mode is in the MPM list and white squares represent the contrary. It can be seen that the proposed method can predict the GEO partition mode effectively. Moreover, there is still space for optimization. And other GEO mode prediction methods will be investigated in our future work.

4. CONCLUSION

In this paper, an MPM-based GEO scheme was proposed to reduce the overhead of GEO index and improve coding efficiency. Based on the observation of the high correlation between partition mode and object boundaries, an edge-directed geometric partition scheme was adopted to construct the MPM list according to the spatio-temporal predicted edge information. Experimental results demonstrated that our proposed algorithm can predict the GEO partition mode efficiently and achieve 0.58% and 1.00% bit-rate reduction on average for RA and LDB configurations, respectively.

5. ACKNOWLEDGEMENT

This work was supported by National Key Research and Development Project (2019YFF0302703) Guangdong Key Research and Development Project (2019B010133001) and High-performance Computing Platform of Peking University, which are gratefully acknowledged.

6. REFERENCES

- [1] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [2] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [3] Il-Koo Kim, Junghye Min, Tammy Lee, Woo-Jin Han, and JeongHoon Park, "Block partitioning structure in the HEVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1697–1706, 2012.
- [4] Yuan Yuan, Il-Koo Kim, Xiaozhen Zheng, Lingzhi Liu, Xiaoran Cao, Sunil Lee, Min-Su Cheon, Tammy Lee, Yun He, and Jeong-Hoon Park, "Quadtree based non-square block structure for inter frame coding in high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1707–1719, 2012.
- [5] YW Chen, WJ Chien, and HC Chuang, "Description of SDR, HDR and 360 video coding technology proposal by Qualcomm and Technicolor—low and high complexity versions," *JVET document, JVET-J0021*, 2018.
- [6] Andrew Segall, Vittorio Baroncini, Jill Boyce, Jianle Chen, and Teruhiko Suzuki, "Joint call for proposals on video compression with capability beyond HEVC," *JVET document, JVET-H1002*, 2017.
- [7] Satoshi Kondo and Hisao Sasai, "A motion compensation technique using sliced blocks in hybrid video coding," in *ICIP. IEEE*, 2005, pp. II–305.
- [8] Renan U Ferreira, Edson M Hung, Ricardo L De Queiroz, and Debargha Mukherjee, "Efficiency improvements for a geometric-partition-based video coder," in *ICIP. IEEE*, 2009, pp. 1009–1012.
- [9] Abdullah A Muhit, Mark R Pickering, and Michael R Frater, "A fast approach for geometry-adaptive block partitioning," in *PCS. IEEE*, 2009, pp. 1–4.
- [10] Liwei Guo, Peng Yin, Yunfei Zheng, Xiaoan Lu, Qian Xu, and Joel Sole, "Simplified geometry-adaptive block partitioning for video coding," in *ICIP. IEEE*, 2010, pp. 965–968.
- [11] Yongjo Ahn and Donggyu Sim, "CE10: Diagonal motion partitions with uni-prediction constraint (test 10.3.3)," *JVET document, JVET-L0125*, 2018.
- [12] Max Blser and Johannes Sauer, "CE10: Results on geometric partitioning," *JVET document, JVET-L0417*, 2018.
- [13] Ru-Ling Liao and Chong Soon Lim, "CE10.3.1.b: Triangular prediction unit mode," *JVET document, JVET-L0124*, 2018.
- [14] Han Gao, Semih Esenlik, Elena Alshina, Anand Meher Kotra, Biao Wang, Max Blser, and Johannes Sauer, "CE4: CE4-1.1, CE4-1.2 and CE4-1.14: Geometric merge mode (GEO)," *JVET document, JVET-P0068*, 2019.
- [15] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2012.
- [16] John Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 6, pp. 679–698, 1986.
- [17] Jianle Chen, Yan Ye, and Seung Hwan Kim, "Algorithm description for Versatile Video Coding and Test Model 6 (VTM 6)," *JVET document, JVET-O1002*, 2019.
- [18] Frank Bossen, Jill Boyce, Karsten Suehring, Xiang Li, and Vadim Seregin, "JVET common test conditions and software reference configurations for SDR video," *JVET document, JVET-N1010*, 2019.
- [19] Gisle Bjontegarrd, "Calculation of average PSNR differences between RD-curves," *VCEG-M33, 13th VCEG Meeting, Austin, TX, USA*, 2001.