Graduate School of Fundamental Science and Engineering
Waseda University

博　士　論　文　概　要
Doctoral Dissertation Synopsis

論　文　題　目
Dissertation Title

Soft-decision-oriented Prediction for Video Compression Beyond VVC

VVCを超える映像圧縮向け軟判定指向予測

申　請　者
(Applicant Name)
Yoshitaka KIDANI
木谷　佳隆

Department of Computer Science and Communications Engineering, Research on Audiovisual Information Processing

May, 2024

Most of the video coding schemes in widespread use today are international standards developed by ITU-T and ISO/IEC. Among these standards, H.266 | Versatile Video Coding (VVC), which was standardized in 2020, currently has the highest coding performance and is expected to be used for video transmission over narrow-band networks such as terrestrial and mobile access lines. However, due to an accelerating increase in the amount of video data traffic, there is a need for coding performance beyond VVC. Therefore, new video compression technologies that can further improve coding performance have begun to be investigated.

Hybrid coding, which combines conventional temporal and spatial redundancy removal methods, is considered a coding method beyond VVC. In addition, new neural network-based coding is also being considered. Hybrid coding is a video coding scheme consisting of prediction, transformation, quantization, filter-in-the-loop, and entropy coding. The joint study group of ITU-T and ISO/IEC has developed a model named Enhanced Compression Model (ECM) based on hybrid coding to pursue coding performance beyond VVC. On the other hand, neural network-based coding has shown coding performance that exceeds VVC in some cases, but it has problems in terms of complexity, such as encoding and decoding runtime. Therefore, this dissertation focuses on achieving coding performance that exceeds VVC by extending the prediction method in hybrid coding.

Prediction consists of three processes: acquisition of reference samples, interpolation of reference samples at fractional-sample-precision positions, and generation of predicted samples. Each process has two or more different modes. Such a variety of predictions can ideally improve coding performance if the appropriate mode can be selected according to the image characteristics of the coded block. On the other hand, many modes do not necessarily improve coding performance because they increase the signaling overhead from the encoder to the decoder required for mode selection. Furthermore, diverse modes increase and encoding and decoding runtime for mode selection and raise the memory bandwidth of the decoder required to acquire reference samples.

This dissertation aims to achieve a highly efficient video coding method to address the increasing video data traffic. Therefore, this dissertation targets improving the efficiency of prediction methods to enhance the coding performance of hybrid coding. Specifically, this dissertation on extensions of prediction and study the methods that implement ideal soft decision criteria as real-world solutions considering actual operations rather than the conventional deterministic mode-switching decision (i.e., hard decision) criteria. In Chapters 2 to 5, this dissertation clarifies hard decision elements in the conventional prediction methods and proposes to soften the hard decision elements, considering the better trade-off between coding performance and complexity. This dissertation consists of the following six chapters.

Chapter 1 describes the background and outline of this dissertation.

Chapter 2 proposes an intra Switchable Interpolation Filter (SIF) with variable thresholds based on coding block sizes and quantization parameter values. To obtain high prediction accuracy over a wide range of video coding bitrates, the proposed method introduces a soft-decision-oriented threshold that switches between intra SIFs with different cutoff frequencies. Experimental results demonstrate that the proposed method provides better coding performance than the conventional method.

Chapter 3 proposes a memory bandwidth-constrained Overlapped Block Motion Compensation (OBMC) method. This method treats the number of motion vectors and the lengths of Interpolation Filter (IF) for neighboring blocks, which are determinants of memory bandwidth in OBMC, as variables dependent on the current coding block sizes. Furthermore, the proposed method generalizes the problem set as a constrained objective function that maximizes memory bandwidth for a predefined upper limit and derives soft-determined variable OBMC parameters. Experimental results show that the proposed method provides better coding performance than the conventional method with the same worst-case memory bandwidth as the conventional method.

Chapter 4 proposes a new mode of Geometric Partitioning Mode (GPM). Specifically, the proposed method introduces GPM with inter and intra predictions (GPM-Inter/Intra) in addition to the conventional GPM, i.e., GPM with different inter predictions (GPM-Inter/Inter). In other words, the proposed method can soften the application conditions of GPM with GPM-Inter/Intra. Furthermore, to suppress the signaling overhead of GPM-Inter/Intra, the proposed method restricts the types of selectable intra-prediction modes in GPM-inter/intra. Experimental results show that the proposed method has superior coding performance compared to the conventional method and qualitatively suppresses artifacts observed in the conventional method for low-delay video coding configurations.

Chapter 5 proposes a new mode of Intra Block Copy (IBC). Specifically, the proposed method introduces bi-predictive IBC using two block vectors in addition to the conventional uni-predictive IBC using a single block vector. In other words, the proposed method can soften the application condition of IBC with a bi-predictive IBC. Furthermore, to avoid increases in encoding runtime while maintaining the coding performance of the proposed IBC, the proposed method introduces an early termination method. This method determines cases where the application of the bi-predictive IBC is ineffective and suspends the proposed IBC's coding process. Experimental results show that the proposed method does not significantly increase the encoding runtime and achieves better coding performance than conventional methods.

Chapter 6 concludes this discussion with future directions.

List of research achievements for application of Doctor of Engineering, Waseda University

Full Name： 木谷　佳隆　　seal or signature

Date Submitted(yyyy/mm/dd): 2024/06/23

種類別 (By Type)	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む） (theme, journal name, date & year of publication, name of authors inc. yourself)
Academic paper	○Y. Kidani, H. Kato, K. Kawamura, H. Watanabe,"Geometric Partitioning Mode with Inter and Intra Prediction for Beyond Versatile Video Coding," IEICE, vol.E105, no.10, pp.1691-1703, Jun. 2022. ○Y. Kidani, K. Unno, K. Kawamura, H. Watanabe, "Memory Bandwidth Constrained Overlapped Block Motion Compensation for VideoCoding," The Institute of Image Information and Television Engineers Transactions on Media Technology and Applications (ITE-MTA), vol.11, no.1, pp.1-12, Jan. 2023. 木谷佳隆, 海野恭平, 河村圭, "H.266 | VVC 対応4K/8K リアルタイムコーデックの開発," 映像情報メディア学会和文論文誌, vol.78, no.1, pp.115-123, Jan. 2024. 鶴崎裕貴, 木谷佳隆, 柴田達雄, "ケーブルテレビの実環境に基づく放送・通信映像伝送動的切換システムの実現可能性検証,"映像情報メディア学会和文論文誌, vol.76, no.6, pp.747-756, Nov. 2022. 鶴崎裕貴, 木谷佳隆, 海野恭平, 河村圭, "VVC におけるマージモードの差分動きベクトル探索の高速化,"電子情報通信学会論文誌D vol. J105-D, no.3, pp.245-253, Mar. 2022. 木谷佳隆, 河村圭, "8K マルチアングル対応リアルタイムエンコーダの開発," 映像情報メディア学会誌, vol.74, no.4, pp.715-718, Dec. 2020.
International Conference Paper	○Y, Kidani, H. Kato, K. Kawamura, "Bi-predictive intra block copy for enhanced video coding beyond VVC," 2024 IEEE International Conference on Image Processing (ICIP), [Accepted on June 2024]. ○Y. Kidani, K. Unno, K. Kawamura, S. Naito, "Block-Size Dependent Overlapped Block Motion Compensation," 2020 IEEE International Conference on Image Processing (ICIP), pp.1191-1195, Virtual, Oct. 2020. ○Y. Kidani, K. Kawamura, K. Unno, S. Naito, "Blocksize-QP Dependent Intra Interpolation Filters," 2019 IEEE International Conference on Image Processing (ICIP), pp.4125-4129, Taipei, Sep. 2019. H. Kato, Y. Kidani, K. Kawamura, S. Naito,"Extended Intra Block Copy with Adaptive Filtering and Overlapped Block Averaging," 2023 IEEE Visual Communications and Image Processing (VCIP), Jeju, Dec. 2023. H. Kato, Y. Kidani, K. Kawamura, S. Naito, "Adaptive boundary width of Geometric Partitioning Mode for Beyond Versatile Video Coding," 2022 IEEE Visual Communications and Image Processing (VCIP), Suzhou, Dec. 2022. Y. Kidani, H. Tsurusaki, K. Unno, K. Kawamura, "Fast Decision Method for Adaptive Motion Vector Resolution in Versatile Video Coding," International Workshop on Advanced Image Technology 2022 (IWAIT 2022), HongKong, Jan. 2022. Y. Kidani, H. Tsurusaki, K. Unno, K. Kawamura, "Fast Decision Method for Merge with Motion Vector Difference in Versatile Video Coding," International Workshop on Advanced Image Technology 2022 (IWAIT 2022), HongKong, Jan. 2022. Y. Kidani, H. Yamashita, S. Matsumoto, "Proposal of RF/IP Adaptive Video Distribution Scheme over Cable Television Access Networks," 2020 SCTE (Society of Cable Telecommunications Engineers)-ISBE (International Society of Broadband Experts) Cable-Tec Expo 2020, Virtual, Oct. 2020.

List of research achievements for application of Doctor of Engineering, Waseda University

Full Name： 木谷　佳隆　　　seal or signature

Date Submitted(yyyy/mm/dd): 2024/06/23

種類別 (By Type)	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む） (theme, journal name, date & year of publication, name of authors inc. yourself)
Domestic Workshop	木谷佳隆, 加藤晴久, 河村圭, “Beyond VVC 向けのイントラブロックコピー拡張方式に関する一検討,” 2023年度画像符号化シンポジウム（PCSJ2023）, 2023年11月 木谷佳隆, 中條健, “招待講演: VVC の性能を超える映像符号化の技術探索活動,” 2022年度画像符号化シンポジウム（PCSJ2022）, 2022年11月 木谷佳隆, 海野恭平, 河村圭, “4K ライブ映像伝送のためのVVC 符号化特性評価,”映像情報メディア学会2021 年度冬季大会, 2021 年12 月 木谷佳隆, 加藤晴久, 河村圭, 渡辺裕, “VVC 拡張方式向けの幾何学分割モードに関する一検討,” 2021年度画像符号化シンポジウム（PCSJ2021）, 2021年11月 齋藤雄太, 木谷佳隆, 海野恭平, 河村圭,“VVC における幾何学的分割モードの符号化処理量削減に関する一検討,”第21 回情報科学技術フォーラム（FIT2022）, 2021 年9 月 木谷佳隆, 海野恭平, 河村圭, “VVC におけるLFSNT 制御の一検討,”第20 回情報科学技術フォーラム（FIT2021）, 2021 年8 月 木谷佳隆, 河村圭, 海野恭平, “オーバーラップブロック動き補償のVVCへの適用検討,” 映像情報メディア学会2020年度冬季大会, 2020年12月 木谷佳隆, 河村圭, “VVC 対応の4K リアルタイムソフトウェアデコーダの試作開発,”映像情報メディア学会メディア工学研究会, 2020 年9 月 木谷佳隆, 河村圭, 内藤整, “次世代動画像符号化向けサイズ・量子化パラメータ依存型イントラ補間フィルタの検討,” 2018年度画像符号化シンポジウム（PCSJ2018）, 2018年11月 木谷佳隆, 河村圭, 内藤整,“次世代動画像符号化方式向けモード・サイズ依存型係数走査方式の検討,”電子情報通信学会画像工学研究会, 2018 年5 月 木谷佳隆, 河村圭, 内藤整, “次世代動画像符号化方式向けイントラ参照画素強フィルタの主観評価, 映像情報メディア学会2017 年度冬季大会, 2017 年12 月
Award	公益財団法人通信文化協会, 第69回前島密賞, 2024年2月 一般社団法人映像情報メディ学会技術振興賞進歩開発賞（研究開発部門）, 2023年5月 一般社団法人映像情報メディア学会丹羽高柳賞業績賞, 2023年5月 一般社団法人映像情報メディア学会丹羽高柳賞論文賞, 2023年5月 International Workshop on Advanced Image Technology 2022 (IWAIT 2022) Best Paper Award, Jan. 2022 一般社団法人映像情報メディア学会2021年度冬季大会鈴木記念奨励賞, 2022年8月 一般社団法人映像情報メディア学会2021年度冬季大会学生優秀発表賞, 2022年8月 一般社団法人電子情報通信学会画像符号化シンポジウム(PCSJ2021) 学生論文賞, 2021年12月 一般社団法人電子情報通信学会第20 回情報科学技術フォーラム（FIT2021）FIT奨励賞, 2021 年8 月 一般社団法人映像情報メディア学会メディア工学研究会優秀発表賞, 2020 年9月 一般社団法人映像情報メディア学会映像情報メディア未来賞次世代テレビ技術賞, 2019 年5 月

Soft-decision-oriented Prediction for Video Compression Beyond VVC

VVCを超える映像圧縮向け軟判定指向予測

July, 2024

Yoshitaka KIDANI
木谷　佳隆

Soft-decision-oriented Prediction for Video Compression Beyond VVC

VVCを超える映像圧縮向け軟判定指向予測

July, 2024

Waseda University Graduate School of Fundamental Science and Engineering

Department of Computer Science and Communications Engineering, Research on Audiovisual Information Processing

Yoshitaka KIDANI
木谷　佳隆

Most of the video coding schemes in widespread use today are international standards developed by ITU-T and ISO/IEC. Among these standards, H.266 | Versatile Video Coding (VVC), which was standardized in 2020, currently has the highest coding performance and is expected to be used for video transmission over narrow-band networks such as terrestrial and mobile access lines. However, due to an accelerating increase in the amount of video data traffic, there is a need for coding performance beyond VVC. Therefore, new video compression technologies that can further improve coding performance have begun to be investigated.

Hybrid coding, which combines conventional temporal and spatial redundancy removal methods, is considered a coding method beyond VVC. In addition, new neural network-based coding is also being considered. Hybrid coding is a video coding scheme consisting of prediction, transformation, quantization, filter-in-the-loop, and entropy coding. The joint study group of ITU-T and ISO/IEC has developed a model named Enhanced Compression Model (ECM) based on hybrid coding to pursue coding performance beyond VVC. On the other hand, neural network-based coding has shown coding performance that exceeds VVC in some cases, but it has problems in terms of complexity, such as encoding and decoding runtime. Therefore, this dissertation focuses on achieving coding performance that exceeds VVC by extending the prediction method in hybrid coding.

Prediction consists of three processes: acquisition of reference samples, interpolation of reference samples at fractional-sample-precision positions, and generation of predicted samples. Each process has two or three different modes. Such a variety of predictions can ideally improve coding performance if the appropriate mode can be selected according to the image characteristics of the coded block. On the other hand, many modes do not necessarily improve coding performance because they increase the signaling overhead from the encoder to the decoder required for mode selection. Furthermore, diverse modes increase encoding and decoding runtime for mode selection and raise the memory bandwidth of the decoder required to acquire reference samples.

This dissertation aims to achieve a highly efficient video coding method to address the increasing video data traffic. Therefore, this dissertation targets improving the efficiency of prediction methods to enhance the coding performance of hybrid coding. Specifically, we focus on extensions of prediction and study the methods that implement ideal soft decision criteria as real-world solutions considering actual operations rather than the conventional deterministic mode-switching decision (i.e., hard decision) criteria. In Chapters 2 to 5, this dissertation clarifies hard decision elements in the conventional prediction methods and proposes to soften the hard decision elements, considering the better trade-off between coding performance and complexity. This dissertation consists of the following six chapters.

Chapter 1 describes the background and outline of this dissertation.

Chapter 2 proposes an intra Switchable Interpolation Filter (SIF) with variable thresholds based on coding block sizes and quantization parameter values. To obtain high prediction accuracy over a wide range of video coding bitrates, the proposed method introduces a soft-decision-oriented threshold that switches between intra SIFs with different cutoff frequencies. Experimental results demonstrate that the proposed method provides better coding performance than the conventional method.

Chapter 3 proposes a memory bandwidth-constrained Overlapped Block Motion Compensation (OBMC) method. This method treats the number of motion vectors and the lengths of Interpolation Filter (IF) for neighboring blocks, which are determinants of memory bandwidth in OBMC, as variables dependent on the current coding block sizes. Furthermore, the proposed method generalizes the problem set as a constrained objective function that maximizes memory bandwidth for a predefined upper limit and derives soft-determined variable OBMC parameters. Experimental results show that the proposed method provides better coding performance than the conventional method with the same worst-case memory bandwidth as the conventional method.

Chapter 4 proposes a new mode of Geometric Partitioning Mode (GPM). Specifically, the proposed method introduces GPM with inter and intra predictions (GPM-Inter/Intra) in addition to the conventional GPM, i.e., GPM with different inter predictions (GPM-Inter/Inter). In other words, the proposed method can soften the application conditions of GPM with GPM-Inter/Intra. Furthermore, to suppress the signaling overhead of GPM-Inter/Intra, the proposed method restricts the types of selectable intra-prediction modes in GPM-inter/intra. Experimental results show that the proposed method has superior coding performance compared to the conventional method and qualitatively suppresses artifacts observed in the conventional method for low-delay video coding configurations.

Chapter 5 proposes a new mode of Intra Block Copy (IBC). Specifically, the proposed method introduces bi-predictive IBC using two block vectors in addition to the conventional uni-predictive IBC using a single block vector. In other words, the proposed method can soften the application condition of IBC with a bi-predictive IBC. Furthermore, to avoid increases in encoding runtime while maintaining the coding performance of the proposed IBC, the proposed method introduces an early termination method. This method determines cases where the application of the bi-predictive IBC is ineffective and suspends the proposed IBC's coding process. Experimental results show that the proposed method does not significantly increase the encoding runtime and achieves better coding performance than conventional methods.

Chapter 6 concludes this discussion with future directions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Video compression technology, i.e., video coding, has evolved tremendously through the efforts of numerous researchers worldwide. Most of the video coding schemes in widespread use today are international standards developed by ITU-T and ISO/IEC. A new generation is developed every ten years, considering user demand for High-Definition (HD) and Ultra-HD (UHD) video services, hardware processing capability, and sufficiently superior coding performance compared to the previous standard. For example, H.265 | High Efficiency Video Coding (HEVC) [6] was standardized in 2013 by the Joint Video Experts Team (JVET), the collaboration team of ITU-T and ISO/IEC. HEVC is widely used for UHD broadcasting and internet video streaming over broadband networks such as satellite and static access lines (e.g., cable television and optical fiber). H.266 | Versatile Video Coding (VVC), developed in 2020 [7], currently has the highest coding performance and is expected to be used for video transmission over narrowband networks such as terrestrial [8] and mobile access lines. Specifically, VVC has twice the coding performance of HEVC, i.e., VVC can achieve around 50% bitrate saving compared to HEVC at the same subjective video qualities.

However, there is a need for coding performance beyond VVC due to an accelerating increase in video data traffic according to the analysis report on the internet data traffic [9, 10]. The projected increase in traffic volume from 2024 to 2029 is assumed to be caused by the introduction of Cross Reality (XR) services such as Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR) [9]. For example, UHD VR video with data more than UHD video is predicted to be widespread [10]. In addition, 3D volumetric video streaming, which leverages 2D video coding on the back end, is expected to further accelerate data traffic [11, 12]. Therefore, new video compression technologies that can further improve coding performance have begun to be investigated.

In the exploration of compression technologies beyond VVC, there are some studies of hybrid coding and neural network-based coding. Hybrid coding, which combines conventional temporal and spatial redundancy removal methods. Specifically, hybrid coding consists of

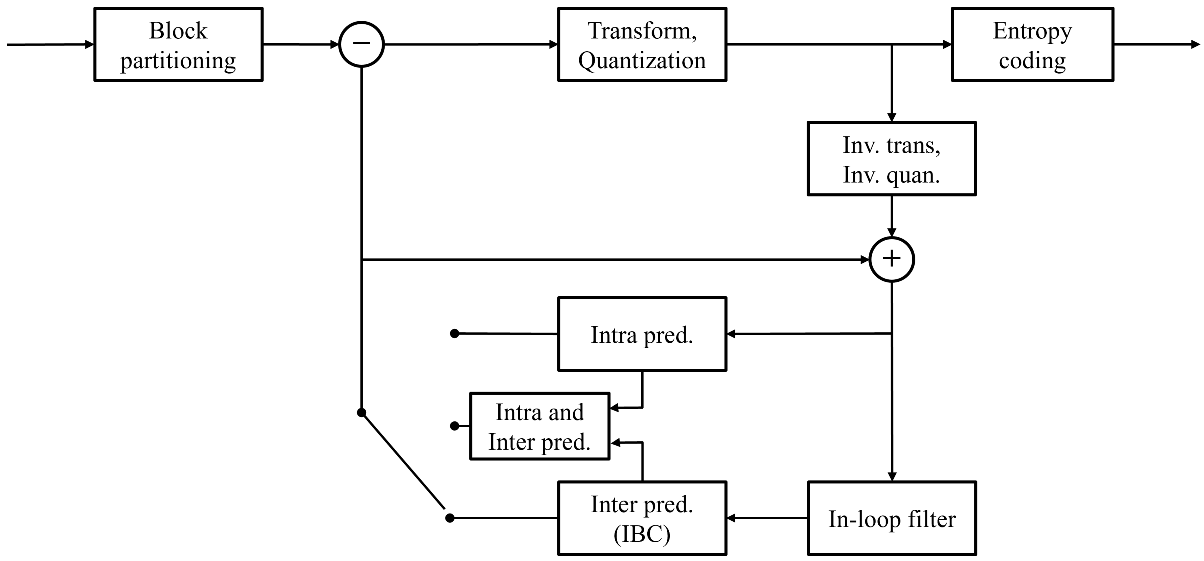


Figure 1.1: Function block diagram of VVC encoder.

prediction, transform, quantization, in-loop filtering, and entropy coding. JVET has developed a model named Enhanced Compression Model (ECM) based on a hybrid coding architecture to pursue coding performance beyond VVC [13]. On the other hand, neural network-based coding has been actively studied mainly in academia and has shown coding performance that exceeds VVC, but it has problems in terms of complexity, such as encoding and decoding runtime compared to hybrid coding [14, 15]. Therefore, this dissertation focuses on the study of hybrid coding with a view to future practical use. In addition, since many new prediction tools have been proposed for ECM and have attracted researchers' attention, this dissertation targets improved coding performance by extending prediction methods.

1.2 Prediction

VVC's high coding performance is due to several new video coding tools that were not in HEVC. VVC follows a hybrid coding architecture, the same as the pre-HEVC standards. Specifically, as shown in Fig. 1.1, VVC initially partitions video frames into coding blocks, generates the residual sample of coding block by block-wise prediction, then applies transform with quantization, and finally performs entropy coding.

Prediction comprises three processes: acquisition of the reference samples; interpolation of reference samples at fractional-sample-precision positions; and generation of prediction samples. Each process has two or three different modes. For instance, VVC has the following various modes in each process to improve coding performance. First, regarding acquisition schemes of reference samples, three schemes are in VVC: inter-prediction, intra-prediction, and Intra Block Copy (IBC), as shown in Fig. 1.2. Fig. 1.2(a) shows inter prediction, also called Motion Compensation (MC), that obtains reference samples from different coded frames using up to two Motion Vectors (MVs). Fig. 1.2(b) shows intra prediction that gets reference

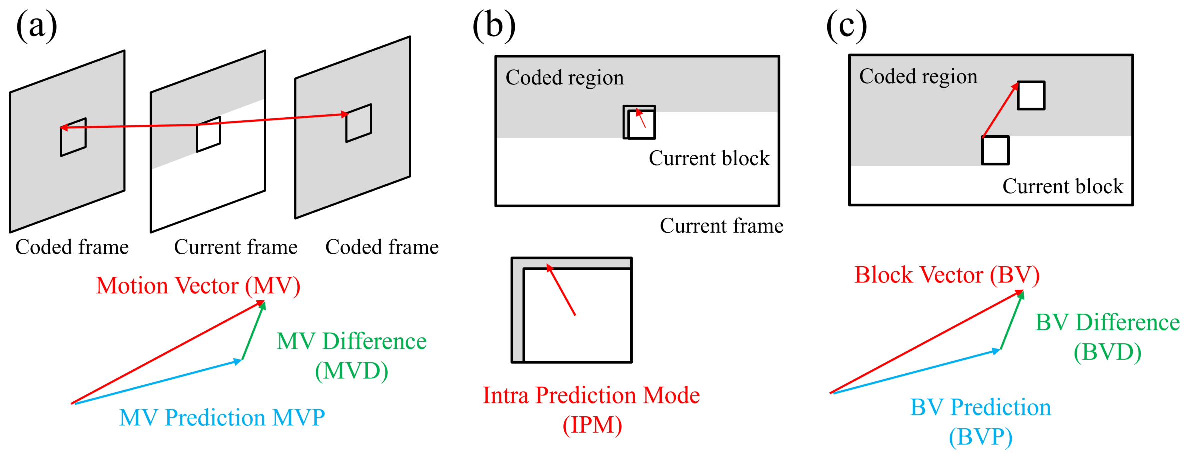


Figure 1.2: Overivew of basic three types of prediction methods in HEVC and VVC: (a) inter prediction; (b) intra prediction; (c) IBC.

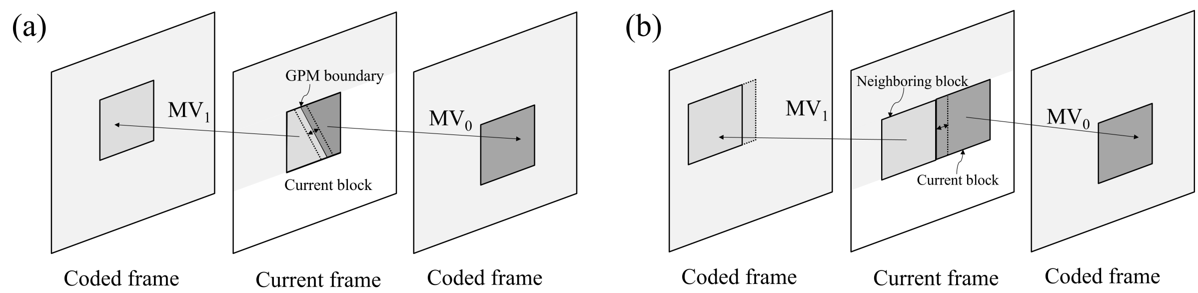


Figure 1.3: Comparison algorithm of blending inter prediction samples in GPM and OBMC: (a) GPM; (b) OBMC. Dashed line areas indicate blending areas.

samples from the coded region adjacent to the coding block within the same frame using an Intra Prediction Mode (IPM). Fig. 1.2(c) shows IBC that fetches reference samples from the broader coded region than that of intra prediction within the same frame using a single Block Vector (BV). Second, regarding the interpolation of reference samples, VVC extends the Interpolation Filter (IF) length more than that of HEVC and introduces Switchable Interpolation Filters (SIFs) with different cutoff frequencies that are applicable for inter- and intra-predictions. Third, regarding the generation of prediction samples, VVC adopts new prediction sample fusion tools such as Combined Inter and Intra Prediction (CIIP) and Geometric Partitioning Mode (GPM). CIIP fusions inter and intra prediction samples using a uniform weight across the entire block. GPM partitions a rectangular coding block into two regions by the pre-defined 64 types of straight lines (i.e., GPM boundary) and blends two different inter prediction samples using weights based on distance from the GPM boundary within the coding block, as shown in Fig. 1.3(a). The GPM fusion can accurately predict boundary areas between foreground and background with different motions.

Additionally, ECM adopts Overlapped Block Motion Compensation (OBMC), originally used in H.263. Comparing GPM, OBMC blends the current block and neighboring block inter prediction samples using weights based on distance from the coding block boundaries, as shown in Fig. 1.3(b). The OBMC fusion can avoid prediction errors around block boundaries in cases where current and neighboring blocks have different motions.

Such a variety of predictions can ideally improve coding performance if the appropriate mode can be selected according to the image characteristics of coding blocks. On the other hand, many modes do not necessarily improve coding performance because they increase the signaling overhead required for mode selection. In addition, diverse modes increase encoding and decoding runtime for mode selection and raise the decoder's memory bandwidth required to acquire reference samples. Those trade-offs in video coding are explained in the next section.

1.3 Trade-offs in Video Coding

There are two major trade-offs in video coding. The first is a trade-off between the code volume of the bitstream, i.e., bitrate, and distortion of coded pictures (i.e., reconstructed sample values) compared to original pictures (i.e., original sample values), which define coding performance. The bitstream includes signaling overhead specifying the mode in prediction and the level value of coefficients generated from the residual signal. Many modes reduce the distortions but also increase the signaling overhead. Therefore, searching for a better trade-off between the bitrate and distortion is indispensable to improve coding performance.

The second is a trade-off between coding performance and complexity. Here, the complexity of encoding and decoding runtime, the decoder's memory bandwidth, and the ease of parallel processing are commonly evaluated. As a well-known technique for improving coding performance, Rate-Distortion (RD) optimization [16] is generally used. The RD optimization is implemented in the reference software as an encoder functionality of minimizing a cost function J that is defined as the weighted sum of bitrate and distortion ($J = R + \lambda D$), where R, λ, and D denote bitrate, Lagrange multiplier, and distortion. The RD optimization requires bitrate so that it can be performed at the final encoding stage. This suggests that many prediction modes increase encoding runtime. In addition, complicated prediction processes raise decoding runtime and the decoder's memory bandwidth. Now that HD/UHD has become popular, it is common to fetch reference samples from the decoder's external memory. An increase in the number of reference samples leads to a rise in memory bandwidth.

For practical use of video coding, pursuing a better trade-off is essential. For example, decreasing the encoding and decoding runtime is required for real-time video transmission, and reducing memory bandwidth is required to save power consumption, which is a critical issue for battery-powered mobile devices.

1.4 Problem Setting

A highly efficient video coding method beyond VVC is required to overcome the increasing video data traffic. This dissertation focuses on improving the prediction efficiency, i.e.,

Table 1.1: Comparison of the approach and selectable mode m in conventional and proposed prediction methods. The number of possible states for m are assumed within the interval from 0 to 1.

Method	Conventional	Proposed
Approach	Hard decision (Deterministic / Fixed)	Soft decision (Variable)
Selectable mode m	$m \in \{0, 1\}$	$m \in M \subseteq [0, 1] \cap \mathbb{Q}$

prediction distortion versus signaling overhead for mode selection, to enhance coding performance.

However, the improvement of prediction efficiency is coming to saturation, with the conventional method that switches two or three modes in each prediction process described in Sec. 1.2. To break through this saturation, it is necessary to assume a larger number of states, ideally a continuous number of states, in each prediction process and to adaptively use them according to the image characteristics of coding blocks. On the other hand, with a view to actual operation, the number of states must be discretized, and improving prediction efficiency is a trade-off between the number of modes and signaling overhead and complexity. Despite this compromise, discretizing the number of modes may provide quantitatively or qualitatively better performance than conventional deterministic mode-switching methods. For the above reasons, the concept of a soft-decision-oriented approach, commonly used in demodulating techniques for error correction of information source coding [17, 18, 19], is important. Therefore, this dissertation treats this approach as the background of the philosophy.

1.5 Dissertation Overview

In this dissertation, we aim to achieve a highly efficient video coding method to address the increasing video data traffic. Therefore, we target improving the prediction efficiency to enhance the coding performance of hybrid coding. Specifically, we focus on extensions of prediction and study the methods that implement ideal soft decision criteria as real-world solutions considering actual operations rather than the conventional deterministic mode-switching decision (i.e., hard decision) criteria. Tab. 1.1 compares the approach and selectable mode in conventional and proposed prediction methods, assuming that the number of possible states for m is within the interval from 0 to 1. In Chapters 2 to 5, we clarify hard decision elements in the conventional prediction methods and propose to soften the hard decision elements, considering the better trade-off between coding performance and complexity.

Fig. 1.4 summarizes the prediction methods targeted in each chapter of this dissertation,

Chapter	Prediction	Decision matter	Hard decision	Soft decision
2 {14}	Intra/ Intra SIF	Apply cubic IF or Gaussian IF	Fixed threshold defined by one block size	<u>Variable threshold defined by QP dependent block sizes</u>
3 {2, 13}	Inter/ OBMC	Apply OBMC or not	Fixed memory bandwidth determinants of OBMC	<u>Variable memory bandwidth determinants of OBMC depending on block sizes</u>
4 {4}	Inter/ GPM	Apply GPM or not	Only GPM-Inter/Inter	GPM-Inter/Inter + <u>GPM-Inter/Intra</u>
5 {7}	IBC	Apply IBC or not	Only Uni-predictive IBC	Uni-predictive IBC + <u>Bi-predictive IBC</u>

Figure 1.4: Summary of the prediction methods targeted in this dissertation, their corresponding decision matters, hard decision elements, and proposed soft decision elements, respectively. The "Chapter" column's curly brackets correspond to the journals and conference papers described as the list of publications in this dissertation. Underlines in the "Soft decision" column are new elements introduced by the proposed method.

along with the corresponding decision matters, hard decision elements, and proposed soft decision elements, respectively. The rest of this dissertation is organized as follows.

Chapter 2 [Block-size and QP Dependent Intra Switchable Interpolation Filters] explores improving the prediction accuracy of an intra SIF. First, we present block-size dependent intra SIFs with two different cutoff frequency characteristics (i.e., cubic and Gaussian IFs), which were finally adopted in VVC. The cubic and Gaussian IFs can be switched by a fixed threshold defined by one block size of the current coding blocks. Specifically, the Gaussian IF has a higher denoising (i.e., smoothing) effect of prediction distortion than the cubic IF. Hence, the conventional method applies the Gaussian IF to large-size blocks with flat image characteristics. In contrast, it applies the cubic IF to small-size blocks with complex image characteristics such as edges. However, since the block sizes and prediction distortions depend on the Quantization Parameters (QPs), the conventional intra SIF is not always optimal for wide QP ranges. Therefore, we propose an intra SIF with variable thresholds defined by QP dependent block sizes. In other words, we introduce the soft-decision-oriented threshold for switching different cutoff frequencies to obtain high prediction accuracy over a wide bit rate range. Finally, we demonstrate the experimental results that the proposed method provides better coding performance than the conventional method in a high QP range.

Chapter 3 [Memory Bandwidth Constrained Overlapped Block Motion Compensation] studies the better trade-off between coding performance and memory bandwidth for OBMC. First, we explain the memory bandwidth issue of OBMC and illustrate memory bandwidth determinants for regular MC and OBMC. Next, we introduce the idea that the application condition of the conventional OBMC is designed not to exceed the maximum (i.e., worst-case) memory bandwidth required for regular MC without OBMC. Specifically, the application of OBMC is determined by a fixed current block size and a fixed number of MVs of the current block and neighboring blocks, which are the determinants of memory bandwidth. However, such OBMC application decisions by the fixed memory bandwidth determinants leave room for OBMC application for the worst-case memory bandwidth and do not maximize the potential coding performance of OBMC. Therefore, we propose a memory bandwidth-constrained OBMC method that treats the memory determinants of OBMC, such as the number of MVs and the IF length of neighboring blocks, as variables depending on the current coding block sizes. Furthermore, we generalize the problem set as a constrained objective function that maximizes memory bandwidth for a predefined upper limit and derives soft-determined variable OBMC parameters. Finally, we show experimental results that the proposed method provides better coding performance than the conventional method with the same worst-case memory bandwidth as the conventional method.

Chapter 4 [Geometric Partitioning Mode with Inter Prediction and Intra Prediction] researches improving the prediction accuracy of GPM. First, we summarize the algorithm of GPM with two different inter predictions (GPM-Inter/Inter), which is finally adopted in VVC. Whether GPM-Inter/Inter is applied is determined by signaling. However, GPM-Inter/Inter does not necessarily predict the boundary of the objects with high accuracy, especially for low-latency video coding configurations where we can fetch the reference samples only from past coded pictures. For example, GPM-Inter/Inter cannot accurately predict the boundary between the background and foreground, which appears after the intersection of two foreground objects, because the background region is not included in the past coded pictures. Therefore, we propose introducing GPM with inter and intra predictions (GPM-Inter/Intra) as a new selectable prediction mode of GPM in addition to GPM-Inter/Inter. In other words, the proposed method can soften the application conditions of GPM with GPM-Inter/Intra. Furthermore, to suppress the signaling overhead of GPM-Inter/Intra, we restrict the types of selectable intra-prediction modes in GPM-inter/intra. Finally, we demonstrate experimental results that the proposed method has superior coding performance compared to the conventional method and qualitatively suppresses artifacts observed in the conventional method for low-delay video coding configurations.

Chapter 5 [Bi-predictive Intra Block Copy] pursues a better trade-off between bitrates and distortion and encoding runtime of IBC. First, we describe conventional IBC methods, such as those used in VVC and ECM. All the conventional IBC methods use a single BV, i.e., uni-predictive IBC, and whether IBC is applied is determined by signaling. The uni-predictive IBC is classified into a BV-search-based IBC and a BV-search-free IBC, and we can select two IBC modes to obtain a better trade-off between bitrates and distortions. The

BV-search-based IBC method can decrease prediction distortion with high-accurate BV while increasing signaling overhead more than the BV-search-free IBC method. The characteristic of the BV-search-free IBC method is vice versa. However, there is room for further coding performance gain if we introduce a new IBC method that realizes an intermediate trade-off between BV-search-based and BV-search-free IBC methods. On the other hand, in that case, we must also consider how to realize a better trade-off between coding performance gains and encoding runtime of IBC since a new IBC method increases the selection of IBC modes. Therefore, we propose to introduce bi-predictive IBC as a new selectable prediction mode in IBC. In other words, the IBC application condition where only uni-predictive IBC can be selected is softened by introducing bi-predictive IBC. Furthermore, to avoid increases in encoding runtime while maintaining the coding performance of the proposed IBC, we introduce an early termination method. This method determines cases where the application of the bi-predictive IBC is ineffective and suspends the proposed IBC's coding process. Finally, we demonstrate experimental results that the proposed method does not significantly increase the encoding runtime and achieves better coding performance than conventional methods.

Chapter 6 concludes this dissertation with future directions.

Chapter 2

Block-size and QP Dependent Intra Switchable Interpolation Filters

2.1 Introduction

2.1.1 Background

[1] Before the development of the H.266 | Versatile Video Coding (VVC) standard, the Joint Exploration test Model (JEM) [20] was developed by the Joint Video Experts Team (JVET) to achieve compression capabilities beyond HEVC. JEM incorporates various new intra prediction tools that served as prototypes for VVC's intra prediction. For example, it includes intra mode coding with 67 intra prediction modes, Cross-Component Linear Model (CCLM) prediction, and a 4-tap intra Switchable Interpolation Filter (SIF). In particular, improving the Interpolation Filter (IF) required for generating reference samples at fractional-sample-precision positions in intra angular prediction is important for reducing prediction residuals.

JEM adaptively switches 4-tap cubic and 4-tap Gaussian IFs dependent on the prediction block sizes, whereas H.265 | High Efficiency Video Coding (HEVC) only uses a 2-tap bilinear interpolation filter. The purpose of switching cubic or Gaussian filters is to acquire the optimal reference samples by selecting cutoff frequency adaptively according to the picture characteristics. In practice, these filters are switched based on a fixed block size threshold, assuming two types of image characteristics: small-size and large-size prediction blocks. The SIF can reduce prediction errors and consequently improve coding performance. However, switching cutoff frequency based on the fixed threshold, i.e., fixed prediction block size, is not always optimal for various Quantization Parameter (QP) values due to the correlation between prediction block sizes and QP values. Specifically, prediction block size increases when QP values increase since QP is proportional to the logarithm of the quantization steps. In addition,

[1]This chapter is based on "Blocksize-QP Dependent Intra Interpolation Filters" [1], by the same author, which appeared in the Proceedings of IEEE International Conference on Image Processing (ICIP), Copyright(C)2019 IEEE.

improving coding performance in high QP conditions is important when transmitting video data over narrow bandwidth, such as mobile networks.

2.1.2 Proposal

This chapter proposes an intra SIF with variable thresholds based on prediction block sizes and QP values. In other words, we introduce the soft-decision-oriented threshold for switching different cutoff frequencies to obtain high prediction accuracy over a wide bit rate range.

2.1.3 Contribution

To evaluate the effectiveness of the proposed method, we implemented it on top of the VVC Test Model (VTM) version 2 (VTM-2) and performed the experiments according to the JVET Common Test Condition (CTC). The experimental results show that the proposed method provides a coding performance gain of 0.45% compared to the VTM-2 for the all intra configuration under JVET CTC. The gain is slightly larger than the conventional method using the fixed block-size criteria (0.41%). Furthermore, the proposed method achieves a coding performance gain of 0.43% more than the conventional method (0.20%) under a higher QP range than those used in JVET CTC.

2.1.4 Outline

The rest of this chapter is organized as follows. Related work and the corresponding problems are explained in Sec. 2.2. Sec. 2.3 presents the details of the proposed method. Sec. 2.4 describes the experimental results and discussion. Finally, we conclude this chapter in Sec. 2.5.

2.2 Related Work

Some related works propose a switchable interpolation filter using different filter taps or cutoff frequencies according to the criteria based on only prediction block size. Matsuo et al. proposed to apply a 4-tap or 6-tap DCT-based filter to a prediction block that is equal to or smaller than 8x8 samples while utilizing a 2-tap bilinear filter for prediction block that is larger than 8x8 samples [21, 22]. Wei et al. proposed to switch 4-tap DCT-based and 4-tap Gaussian filters according to the same criteria as Matsuo et al. [23]. Yoo et al. also proposed to switch 4-tap cubic and 4-tap Gaussian filters according to the prediction block-size-based criteria [24].

The design policy for conventional switchable filters based on prediction block size is as follows. Since large-size blocks generally have flat or simple textures, all the conventional methods proposed low-pass filters for large-size blocks. In addition, the effect of interpolation filters tends to be smaller for large-size blocks because the distance from the reference samples to the prediction samples near the lower right corner is larger. Therefore, a short-tap filter such as a 2-tap bilinear filter or a 4-tap Gaussian filter can provide sufficient smoothing effect. On the other hand, small-size blocks have complicated textures, and all samples in the block are close to the reference samples, so the effect of interpolation filters tends to be large. Therefore, a 4-tap DCT-based or cubic interpolation filter with a sharpness effect is applied to small-size blocks to reduce the prediction error.

2.3 Proposed Method

2.3.1 Analysis of Switchable Interpolation Filter

We performed a preliminary experiment to investigate whether the conventional switchable interpolation filter using a fixed block size threshold is optimal. Specifically, we analyzed the correlation between prediction block size and the selected filter between the 4-tap cubic and Gaussian filters by implementing them in the VTM-2 [25] instead of using the existing 2-tap bilinear filter for all-size blocks in VTM-2. We also replaced the prediction block-size-based selection criteria with an RDO-based type.

Figs. 2.1a and 2.1b show the 4-tap cubic and Gaussian filtered blocks for BasketballDrill (832×480) of the first frame with two QPs such as 22 and 37. BasketballDrill is a test sequence in the JVET Common Test Condition (CTC) [26]. Figs. 2.2a and 2.2a show the 4-tap cubic and Gaussian filters applied sample ratio normalized frame for each prediction block size. Here, non-filtered blocks appear when the horizontal, vertical, DC, or planar prediction is selected since the integer position is referenced.

From these figures, the three following trends can be revealed. First, cubic filtered blocks equal to or greater than 16×16 samples and Gaussian filtered blocks smaller than 8×8 samples appear in both QP values, where these filters are prohibited in conventional method's prediction block size-based criteria. Second, small-size prediction blocks coded by a smaller QP tend to select the cubic filter instead of the Gaussian filter and vice versa. On the other hand, large-size prediction blocks coded by a higher QP tend to select a cubic filter compared with the smaller QP case. Consequently, an optimal filter depends on prediction block size and QP values. Thus, it was confirmed that the prediction block size-dependent filters in the related works are insufficient for the Bjøntegaard Delta bitrate (BD-rate) aspect [27, 28]. The BD-rate is a well-known evaluation metric to quantify the difference in bitrate for equivalent levels of Peak Signal-to-Noise Ratio (PSNR), whose negative and positive values indicate coding performance gains and losses, respectively.

2.3.2 Block-size and QP Dependent Intra Switchable Interpolation Filters

This chapter proposes switching criteria based on the prediction block size and QP value for applying 4-tap cubic and Gaussian filters. Tab. 2.1 shows the proposed switching criteria. The column and row indicate prediction block sizes and QP values, respectively. The letters "C" and

Table 2.1: The portion of cubic filtered blocks (left) and Gaussian filtered blocks (right) at from the whole CTC test-set QP = 37.

QP value	prediction block size				
	4	8	16	32	64
18	C	G	G	G	G
19	C	C	G	G	G
···	C	C	G	G	G
27	C	C	G	G	G
28	C	C	C	G	G
···	C	C	C	G	G
35	C	C	C	G	G
36	C	C	C	C	G

“G” indicate the cubic and Gaussian filters, respectively. The prediction block size, i.e., whether to use the block’s height or width as the threshold, is determined based on the direction of intra angular modes because the non-square prediction blocks occur in VTM. When the angular is larger than 45°, the block size is defined as block width, and vice versa, as shown in Fig. 2.3. The proposed block-size definition, in contrast with conventional methods, allows the filters for non-squared blocks to select a suitable cut-off frequency.

The proposed method’s main difference is variable thresholds of prediction block sizes according to QP values. Here, the variable thresholds are derived by preliminary and exhaustive experiments, combining the QP value (12, 17, 22, 27, 32, 37, 42, 47, 52, and 57) and prediction block size thresholds (4, 8, 16, 32, and 64) for all test sequences in the JVET CTC. The derived thresholds endorse our assumption. The assumption is that the Gaussian filter is also likely to be applied for small-size prediction blocks in a low QP range, and the cubic filter is also likely to be applied for large-size prediction blocks in a high QP range.

2.4 Experimental Results and Discussion

2.4.1 Test Conditions

To evaluate the effectiveness of the proposed method, we implemented it on the VTM-2 and performed the experiments following JVET CTC [26]. Since the proposed method focuses on an intra-picture, the experiments were evaluated only for the all intra configuration of the JVET CTC. Coding performance is measured by the BD-rate of the luma (i.e., Y) component (BDY), which can be calculated by PSNR and bitrates of different coding methods at four different QP points. The complexity was assessed using the ratio of the proposed method’s encoder runtime (EncT) and decoder runtime (DecT) compared to VTM-2. As four QP points, we also used the

Table 2.2: Results in the CTC-QP range.

Class	Proposal			Yoo et al.		
	BDY [%]	EncT [%]	DecT [%]	BDY [%]	EncT [%]	DecT [%]
A1	-0.06	102	100	-0.18	103	100
A2	-0.17	102	100	-0.15	103	101
B	-0.43	101	101	-0.41	103	101
C	-0.76	100	101	-0.65	101	102
D	-0.55	100	100	-0.53	101	101
E	-0.73	101	100	-0.59	102	101
Overall (A∼E w/o D)	-0.45	101	101	-0.41	102	101

high-QP range (QP = 32, 37, 42, 47) in addition to the CTC-QP range (QP = 22, 27, 32, 37). In the JVET CTC's all-intra configuration, five classes were defined based on the use case with content resolution. The resolutions of test sequences are categorized by classes A1/A2, B, C, D, and E: 3840×2160, 1920×1080, 832×480, 416×240, and 1280×720. The frame rate ranges of the test sequences are from 24 to 60.

2.4.2 Rate-Distortion Curve Characteristics

Fig. 2.4 shows a Rate-Distortion (RD) curve of the sequence "BasketballDrill" (832×480, 50 FPS) for the all intra configuration in the high-QP range. Fig. 2.4 reveals that the coding performance gain comes not from the objective quality improvement but from the bitrate reduction by the proposed method. It can be considered that the bitrate reduction is realized as follows. The precision of intra-angular prediction is enhanced by high-precision reference samples generated using the optimal IF cutoff selected in the proposed method. This has led to a reduction in the prediction residual signal, resulting in minimal residual signal coding and a decrease in bitrate. Similar observations were made for the "Basketballdrill" sequence as with the other sequences.

2.4.3 Objective Evaluations

Tabs. 2.2 and 2.3 show the experimental results by the two QP range. The baseline of all results is VTM-2. In accordance with the JVET CTC, Class D's results are not included in the overall results.

Tab. 2.2 shows that proposed and conventional methods bring coding performance gains of 0.45% and 0.41% over VTM-2 under CTC-QP range. On the other hand, Tab. 2.3 shows that proposed and conventional methods provide coding performance gains of 0.43% and 0.20% over VTM-2 under high-QP range. A comparison of the coding performance gains of classes revealed that the gain with the low-resolution class is larger than that with the high-resolution

Table 2.3: Rsults in the high-QP range.

Class	Proposal			Yoo et al.		
	BDY [%]	EncT [%]	DecT [%]	BDY [%]	EncT [%]	DecT [%]
A1	-0.08	102	100	-0.18	103	100
A2	-0.10	102	100	-0.03	103	100
B	-0.52	101	101	-0.20	104	101
C	-0.72	100	101	-0.37	103	102
D	-0.50	100	100	-0.40	102	101
E	-0.57	101	100	-0.20	103	101
Overall (A∼E w/o D)	-0.43	101	101	-0.20	103	101

class. This is because the low-resolution test sequences have more small-size blocks with complex textures than the high-resolution test sequences. SIF is highly effective in these complex texture areas. Moreover, the proposed method can achieve double coding performance gain compared to the conventional method under high-QP ranges. This is because the cubic filter can be applied to large-size prediction blocks in the proposed method, while the cubic filter cannot be applied to these blocks in the conventional method. In addition, the proposed and conventional methods increase encoding and decoding runtime due to the SIF. The operation of the 4-tap IF in the proposed and conventional method is larger than the 2-tap bilinear filter in VTM-2.

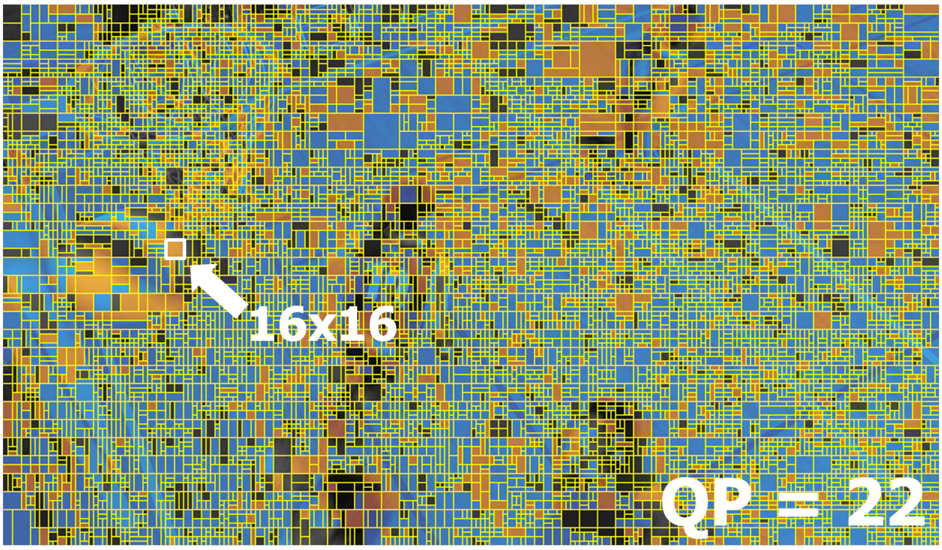
The experimental results demonstrated the effectiveness of the proposed method in terms of coding performance gains, especially for a wide QP range. On the other hand, the proposed method also provides a coding performance loss of 0.10% over the conventional method in some Class A1 sequences (FoodMarket and Tango). The coding performance losses are due to blurs of the two test sequences, which can spoil intra SIF. From the preliminary analysis, the optimal thresholds of the prediction block size are 4 for the two test sequences under the CTC and high-QP ranges. Hence, the coding performance of the conventional method, using a fixed threshold of prediction block (i.e., 8), is higher than the proposed method.

2.4.4 Subjective Evaluation

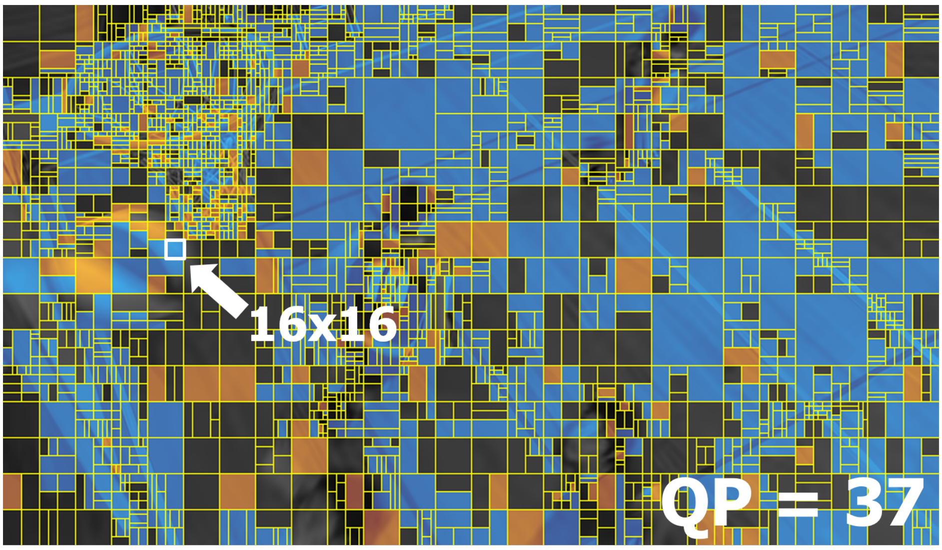
Fig. 2.5 compares subjective qualities and block map analyses of VTM-2 for BasketballDrill at QP = 37. Fig. 2.5 compares subjective qualities and block map analyses of VTM-2 for the “Basketballdrill” sequence at QP = 37. Firstly, the application of different filters does not significantly impact the subjective quality as long as the prediction block size remains constant. Secondly, the artifact within the red boundary is only observed in Yoo et al., where larger prediction blocks are present within the red boundary than VTM-2 and the proposed method. Therefore, it can be inferred that the variance in prediction block partitioning has a greater influence on subjective quality than the differences in filter application.

2.5 Conclusion

This chapter proposes intra SIFs with variable thresholds based on prediction block size and QP values. The proposed variable threshold can better switch the two IFs with different cutoff frequencies over a wide bit rate range than the conventional fixed threshold. The experimental results provided an average coding performance gain of 0.45% under all intra configurations of JVET CTC compared with the coding performance of VTM-2 using one intra IF. Furthermore, the proposed method provides twice the coding performance gain for a high QP range than the conventional method using intra SIFs based on the fixed block-size threshold. This demonstrates that the proposed method can maintain coding gains even in narrowband environments such as mobile networks.

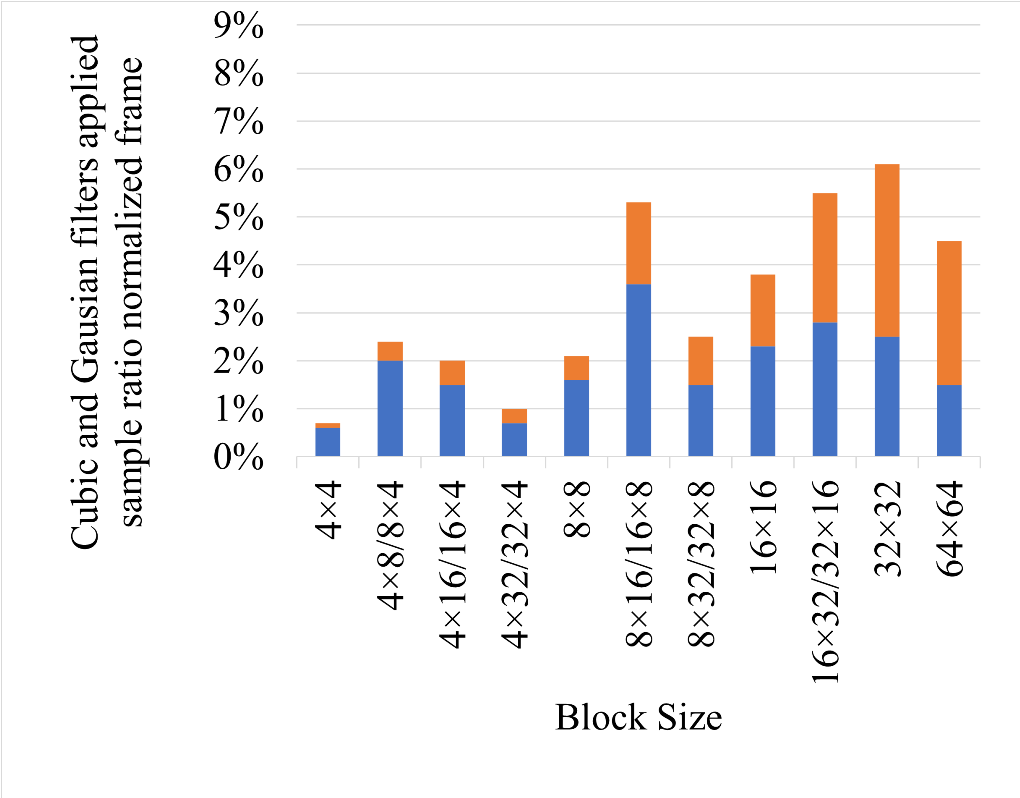


(a) QP = 22

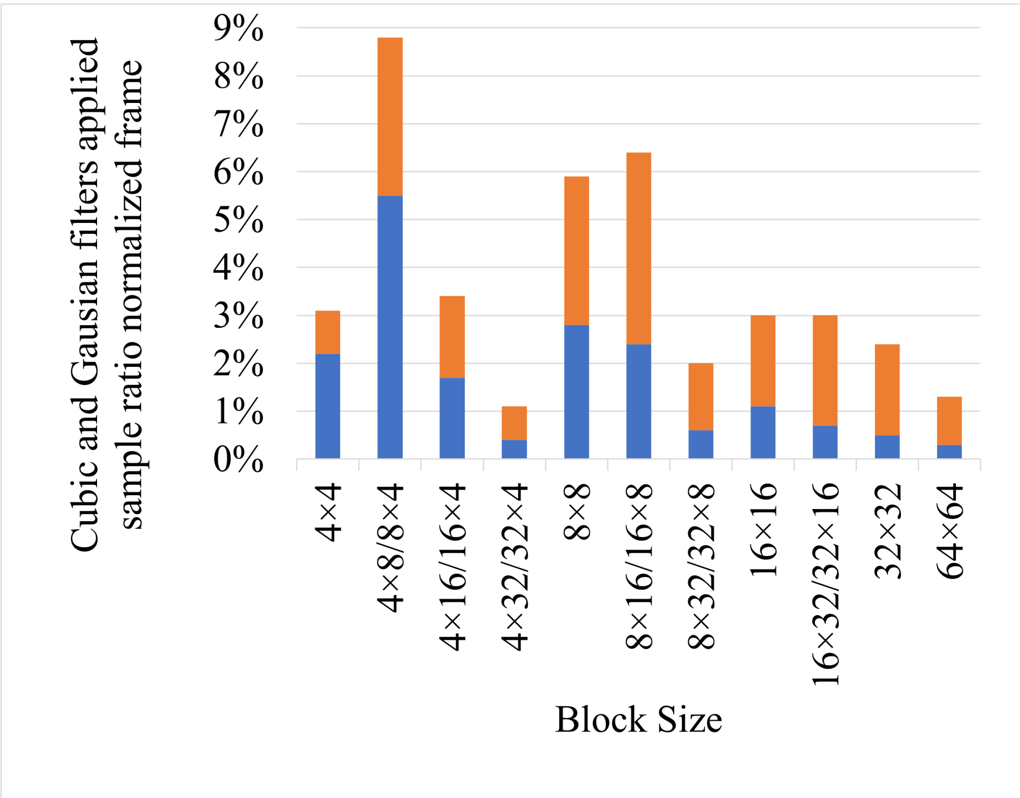


(b) QP = 37

Figure 2.1: 4-tap cubic and Gaussian filtered blocks for BasketballDrill (832×480) of the first frame. The yellow grids denote prediction block partitions. The blue, orange, and gray blocks indicate cubic, Gaussian, and non-filtered. (Copyright(C)2019 IEEE, [1] Fig. 1)



(a) QP = 22



(b) QP = 37

Figure 2.2: Cubic and Gaussian filters applied sample ratio normalized frame for each prediction block.

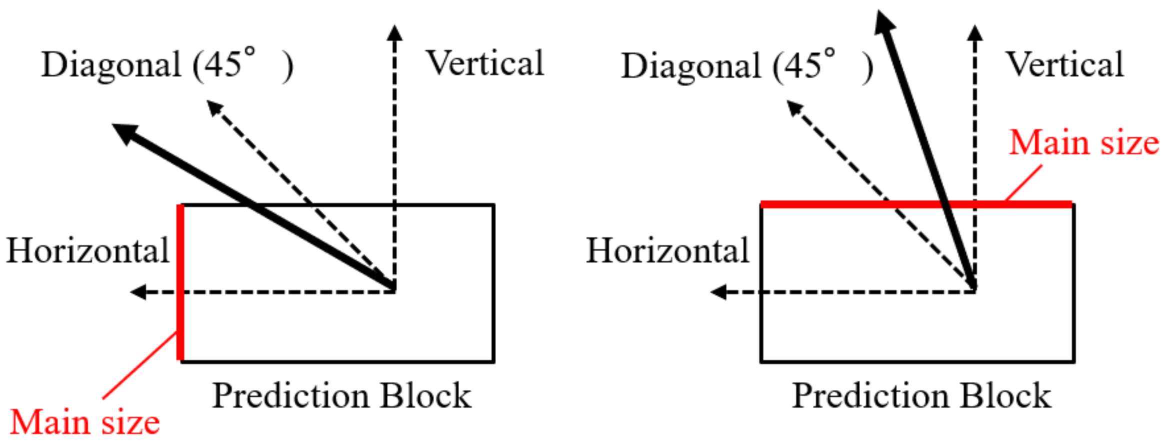


Figure 2.3: Main size of the prediction block for switching intra SIF is determined by the direction of intra angular modes. (Copyright(C)2019 IEEE, [1] Fig. 3)

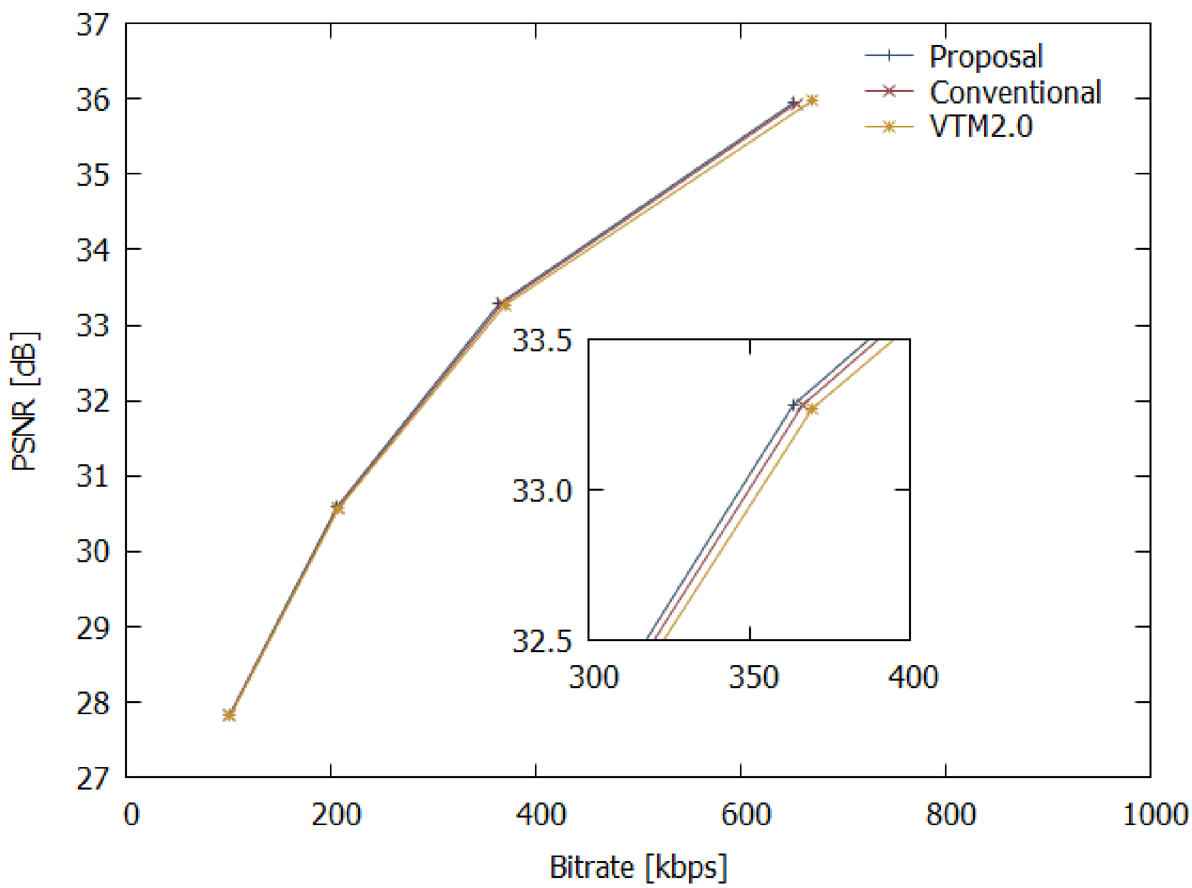


Figure 2.4: RD curve of "BasketballDrill" for all intra configuration in high-QP rage. (Copyright(C)2019 IEEE, [1] Fig. 4)

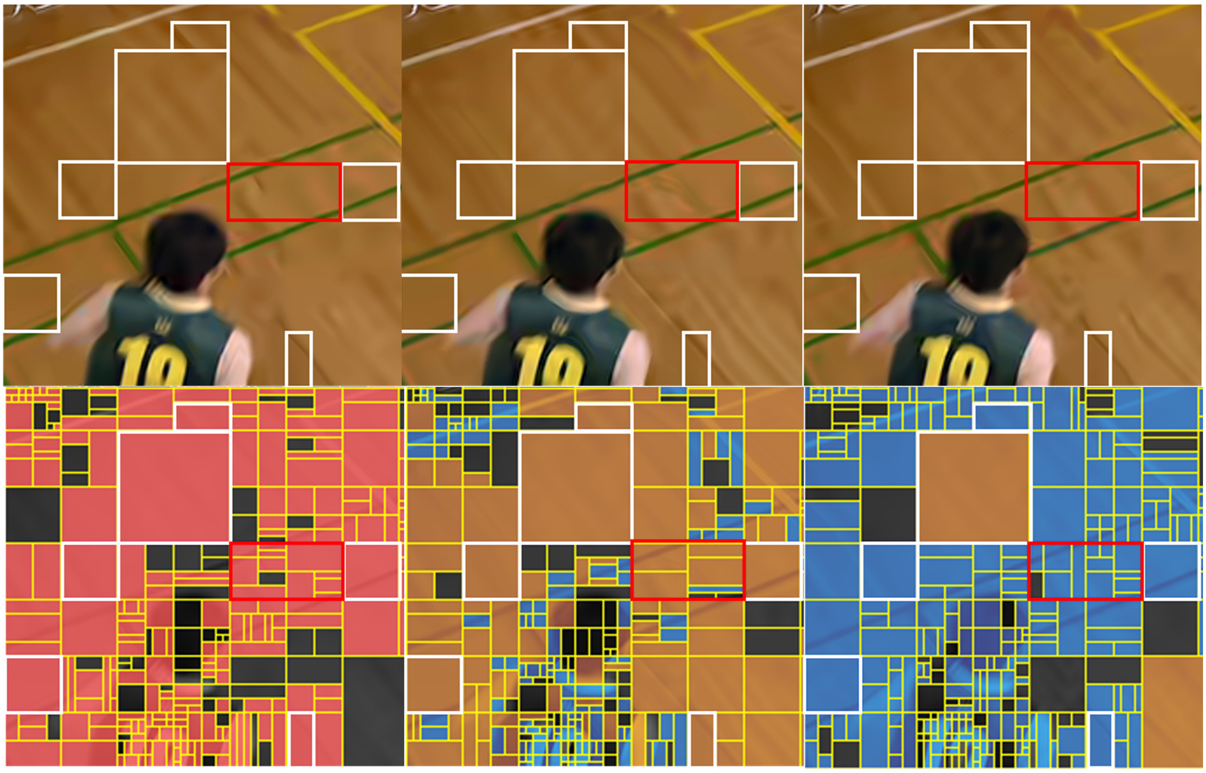


Figure 2.5: Comparison of subjective qualities and block map analyses of VTM-2 (left), conventional method (center), and the proposed method (right) at QP = 37. The red, blue, orange, and gray blocks indicate 2-tap bilinear filtered, 4-tap cubic filtered, 4-tap Gaussian filtered, and non-filtered blocks, respectively. The non-filtered blocks occur when the prediction block selects the horizontal, vertical, DC, and planar modes since the integer position is referenced. (Copyright(C)2019 IEEE, [1] Fig. 5)

Chapter 3

Memory Bandwidth Constrained Overlapped Block Motion Compensation

3.1 Introduction

3.1.1 Background

[1] In the block-wise Motion Compensation (MC) of H.265 | High Efficiency Video Coding (HEVC) and H.266 | Versatile Video Coding (VVC), called Regular MC (RMC), predicted sample values near the block boundary are discontinued when the motion vectors of current blocks and neighboring blocks (mv_{C} and mv_{N} hereinafter) differ. Overlapped Block Motion Compensation (OBMC) and De-Blocking Filters (DBF) are well-known solutions to this problem. As shown in Fig. 3.1, OBMC blends the predicted samples generated by mv_{N} across the block boundary into those generated by mv_{C} to reduce the block discontinuities [29, 30]. In contrast, DBF directly smooths block boundaries of the reconstructed blocks generated by the predicted blocks and residual blocks [6, 7]. OBMC and DBF provide additive improvements because of their different mechanisms, but OBMC has not yet been adopted in HEVC and VVC due to the issue of increasing the number of reference samples, which is approximated as memory bandwidth [31, 32]. We, therefore, focus on the memory bandwidth constrained OBMC method for further improvements of VVC. As that related work, uni-prediction based OBMC (i.e., OBMC using only one mv_{C} and one mv_{N}) [33] was proposed but not adopted in VVC. This is because the uni-prediction-based OBMC cannot retain the coding performance improvement with bitrate saving (coding performance gain hereinafter) of the bi-prediction-based OBMC (i.e., OBMC using two mv_{C} and two mv_{N}) [34].

[1] This chapter is based on “Memory Bandwidth Constrained Overlapped Block Motion Compensation for Video Coding” [2], by the same author, which appeared in the ITE Transactions on Media Technology and Applications,

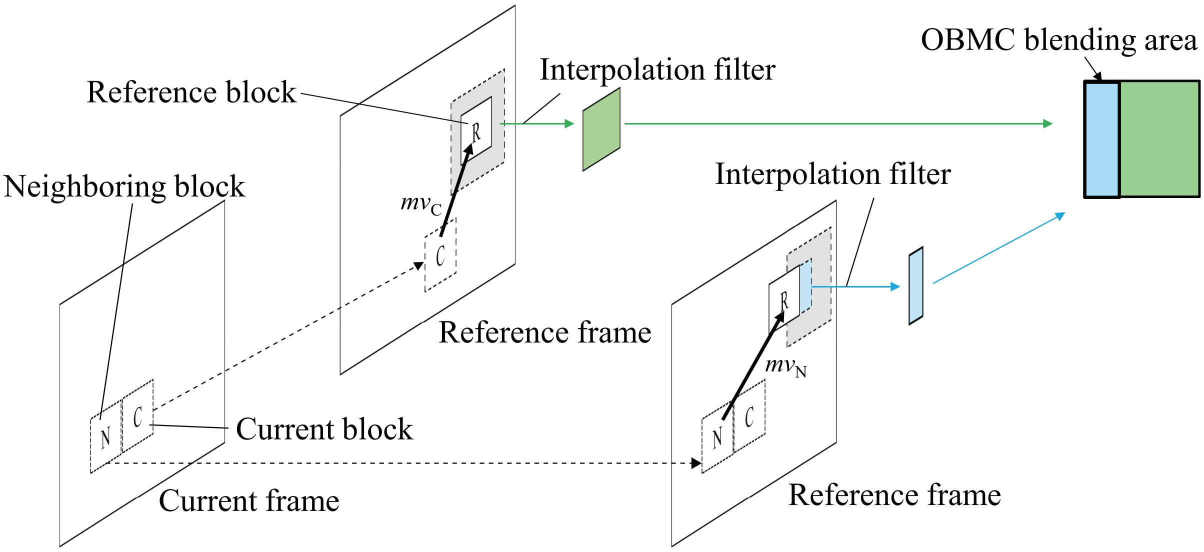


Figure 3.1: A mechanism of OBMC. C, N, and R represent the current, neighboring, and reference blocks, respectively. Shaded areas represent the extended reference sample areas against R when interpolation filters are required, that is when mv_{C} and mv_{N} indicate fractional-sample positions, respectively. (Copyright(C)2023 ITE, [2] Fig. 1)

3.1.2 Proposal

This chapter proposes a memory bandwidth-constrained OBMC method that treats the memory determinants of OBMC, such as the number of motion vectors and the interpolation filter length of neighboring blocks, as variables depending on the current coding block sizes. Furthermore, we generalize the proposed method as a constrained objective function that maximizes the memory bandwidth for an arbitrary preset upper limit to derive softened variable OBMC parameters.

3.1.3 Contribution

To evaluate the effectiveness of the proposed method, we implemented it on top of the VVC Test Model (VTM) version 10 (VTM-10) and performed the experiments according to the VTM Common Test Condition (CTC). Experimental results show that the proposed method provides an additional coding performance gain of 0.22% over VVC reference software. This gain is comparable to that of bi-prediction-based OBMC (0.33%), requiring 3.8 times the maximum memory bandwidth of VVC, but it is still greater than that of uni-prediction-based OBMC (0.12%).

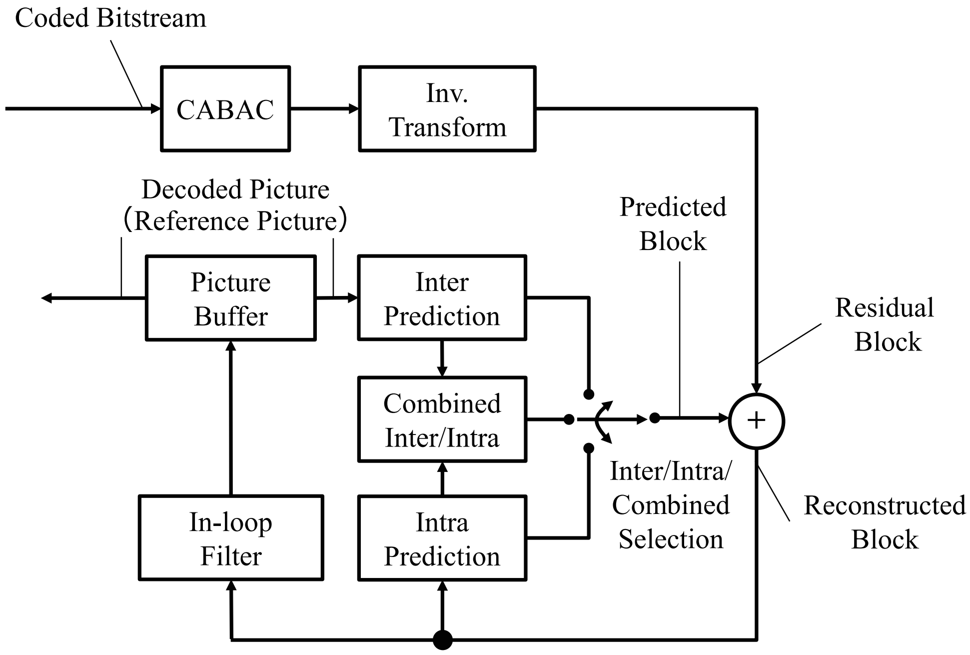


Figure 3.2: An example of the architecture for the general video decoder. (Copyright(C)2023 ITE, [2] Fig. 2)

3.1.4 Outline

The rest of this chapter is organized as follows. Memory bandwidth and related work are explained in Sec. 3.2. The problems are shown in Sec. 3.3. Sec. 3.4 presents the details of the proposed method. Sec. 3.5 describes the experimental results and discussion. Finally, we conclude this chapter in Sec. 3.6.

3.2 Memory Bandwidth and Related Work

3.2.1 Memory Bandwidth

Fig. 3.2 shows an architecture of the general video decoder, including the portions for the inter prediction, intra prediction, inverse transform, in-loop filter, and picture buffer. The function of RMC of HEVC and VVC is included in the inter prediction, and the reference samples for RMC are obtained from the picture buffer. As the video services achieve higher definition, external memory is now commonly used for the picture buffer to store the reference picture in the hardware video decoder [35]. This throughput of the access from the external memory to the main chip on the hardware video decoder is called the memory bandwidth.

The memory bandwidth for the hardware video decoder is designed by the M_{Wst} since it cannot be changed after manufacturing. In general, the M_{Wst} is approximated to be the maximum number of reference samples required for a predicted sample of RMC, as well as the related works [31, 32]. Hence, M_{Wst} is increased when more reference samples need to be

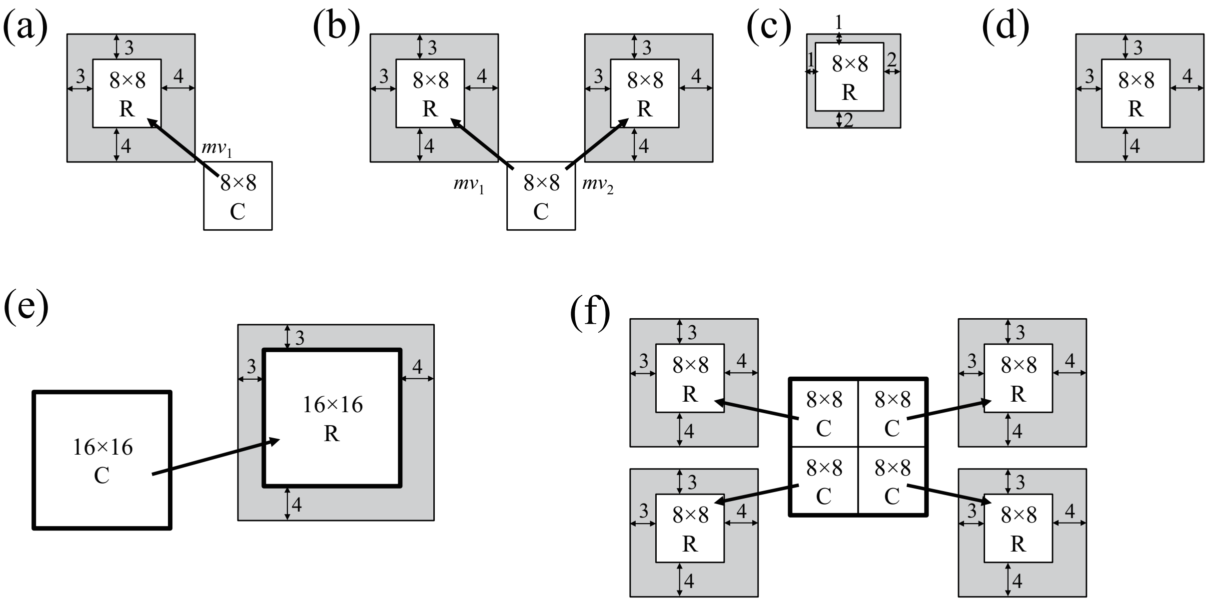


Figure 3.3: Examples of the memory bandwidth required for RMC of the current blocks increased by three determinants, i.e., the number of motion vectors, the number of interpolation filter taps, and the size of the current block against the size of pipeline processing for RMC. C, R, and shaded areas denote current blocks, reference blocks, and extended reference sample areas by interpolation filter against R, respectively. (a) Uni-prediction, (b) Bi-prediction, (c) 4-tap filter, (d) 8-tap filter, (e) an RMC pipeline organized by one 16×16 current block, and (f) an RMC pipeline organized by four 8×8 current blocks. (Copyright(C)2023 ITE, [2] Fig. 3)

fetched with the interpolation filter than the number of samples within the prediction block. In other words, the smaller the blocks are, the larger M_{Wst} becomes. The reduction of the M_{Wst} is critical, especially for mobile devices, since it saves on power consumption [36, 37, 35].

3.2.2 Memory Bandwidth Determinants

The M_{Wst} for RMC is varied depending on the number of motion vectors and the interpolation filter taps of the current blocks (n_{C} and t_{C}). Regarding n_{N}, the maximum number is two, i.e., bi-prediction, and it is utilized in HEVC and VVC, for example. When n_{N} is two, the required reference samples become double as shown in Fig. 3.3(a) and (b). Regarding t_{N}, 8-tap and 4-tap filters generate the predicted samples of the luma and chroma components, respectively. As t_{N} becomes longer, the required reference samples are increased as shown in Fig. 3.3(c) and (d).

The M_{Wst} for RMC also depends on the current block sizes. VVC diversifies the block partitioning including non-square shape, not in HEVC, so that selects the coding block size from the minimum 4×4 to the maximum 128×128 samples, for instance [7]. The extension of the block sizes increases the required internal memory, storing the reference samples from the external memory of the main chip on the hardware video decoder [38]. The current block's RMC

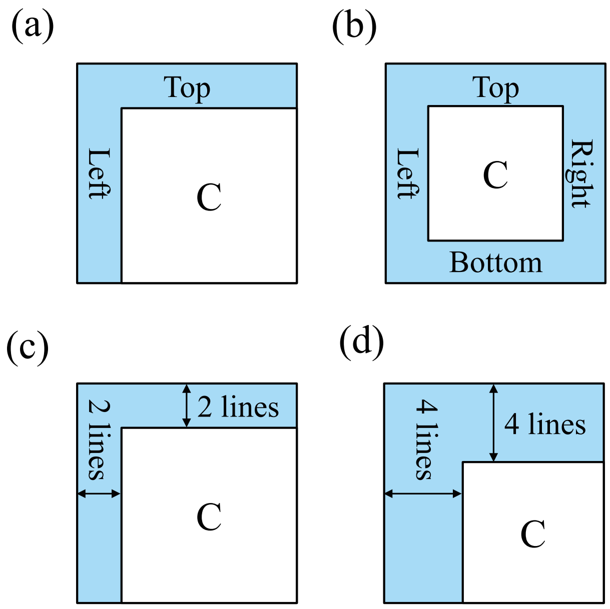


Figure 3.4: Examples of the memory bandwidth required for OBMC of the current blocks increased by two determinants, i.e., OBMC applicable locations and blending lines. C and sky-blue area denote current blocks and OBMC blending area. (a) Only top and left sides, (b) all sides, (c) 2-lines, (d) 4-lines. (Copyright(C)2023 ITE, [2] Fig. 4)

is generally conducted at the subblock level to reduce the required internal memory sizes. By Decoder-side Motion Vector Refinement (DMVR) in VVC, the maximum size of the subblock is defined as 16×16 samples. This means that each 16×16 predicted sample value in RMC is always the same as the non-subblock-wise RMC, assuring the 16×16 sample-wise pipeline processing within RMC.

The reference samples for the pipeline processing of RMC are fetched from the external memory all at once. Hence, the more the current block consists of the multiple smaller size blocks, the more reference samples are required, as shown in Fig. 3.3(e) and (f). To reduce the M_{Wst}, the minimum block sizes for the uni-prediction and bi-prediction of VVC are constrained by $4 \times 8/8 \times 4$ and 8×8, respectively [7].

3.2.3 Overlapped Block Motion Compensation

OBMC increases the memory bandwidth since the reference samples of neighboring blocks are required compared to those of only RMC. In addition to n_{N} and t_{N} for OBMC as in RMC, the reference samples required for OBMC are determined by the application locations and blending lines as shown in Fig. 3.4(a) and (b).

The original method of OBMC [29, 30] does not prohibit the application of OBMC to the current block boundaries on all four sides (e.g., top, left, right, and bottom), which is sometimes called non-causal OBMC [39]. The non-causal OBMC raises another implementation issue

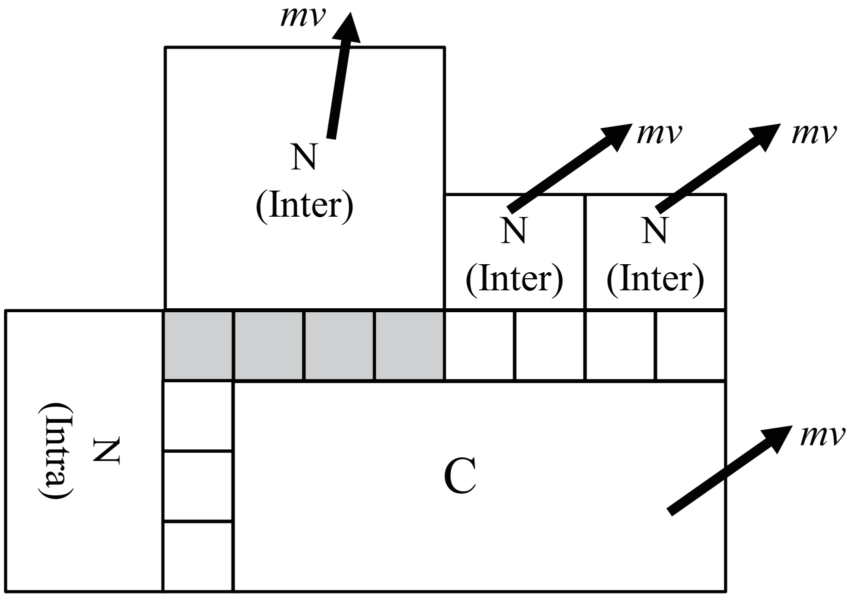


Figure 3.5: An example of the SbOBMC. C and N denote the current block and neighboring blocks. Shaded and non-shaded subblocks indicate SbOBMC applied and non-applied subblock. (Copyright(C)2023 ITE, [2] Fig. 5)

except for the memory bandwidth. Specifically, blending the right and bottom sides in the non-causal OBMC increases an encoding and decoding delay for block-wise processing in raster-scan order. In addition, for parallel processing, the non-causal OBMC must fetch ahead in the lower right blocks, which requires storage for many reference samples, increasing the internal memory size. To address the problem, the causal OBMC where OBMC can be applied only for the top and left sides is proposed [40]. This chapter follows the causal OBMC since we focus on the practical OBMC method.

Another method following the causal OBMC method but enabling the various size blocks, including non-square shaped types, is proposed as shown in Fig. 3.5 to realize the coding performance beyond HEVC (Chen2015 hereinafter) [41]. Chen2015 introduces the 4×4 subblock-wise OBMC to correspond to the neighboring blocks with various block sizes and prediction modes (i.e., inter or intra). The 4×4 Subblock-wise OBMC (SbOBMC) enables a detailed applicable determination depending on the prediction modes and similarity of the motion vector of the neighboring blocks as shown in Fig. 3.5 [41]. The SbOBMC applicable determination is conducted only when OBMC is determined to be applicable in a coding block, which is identified by a block-wise $obmc\_flag$ signaled from the encoder. For the signaling $obmc\_flag$, the encoder calculates the Rate-Distortion (RD) costs with and without OBMC applied after determining whether the current block is uni-prediction or bi-prediction.

Furthermore, Chen2015 proposes adjusting the OBMC blending lines depending on the current block sizes to reduce the M_{Wst}. Specifically, Chen2015 utilizes 4 lines when the width or height of the current block is larger than 8 samples, otherwise, it utilizes 2 lines. However, Chen2015 can apply OBMC even for bi-prediction current blocks having bi-prediction neighboring blocks, which increases the M_{Wst} against RMC in the small size current blocks.

To solve this problem, uni-prediction-based OBMC is proposed (Lin2019 hereinafter) [33]. Lin2019 enables OBMC only for uni-prediction current blocks having a uni-prediction neighboring block. To maintain the OBMC application rates, Lin2019 can apply OBMC for uni-prediction current blocks having bi-prediction neighboring blocks by converting neighboring blocks from bi-prediction to uni-prediction based on the distance between the current and reference pictures. Lin2019 further proposes prohibiting the application of OBMC for $4 \times 8/8 \times 4$ blocks to reduce the M_{Wst}.

3.3 Problem Statement

In this section, first, we derive the formula to calculate the M_{Wst} for RMC and OBMC and analyze the M_{Wst} of Chen2015 and Lin2019, when implemented in VVC as an example. Second, we ascertain the bottleneck of OBMC by changing t_{N} in these two methods.

The memory bandwidth of the final inter predicted $W \times H$ sample block $M_{\text{Inter}}^{W \times H}$ can be calculated depending on whether OBMC is applicable or not as

$$M_{\text{Inter}}^{W \times H} = \begin{cases} M_{\text{RMC}}^{W \times H} + M_{\text{OBMC}}^{W \times H} & \text{if applicable,} \\ M_{\text{RMC}}^{W \times H} & \text{otherwise,} \end{cases} \tag{3.1}$$

where W, H, $M_{\text{RMC}}^{W \times H}$, and $M_{\text{OBMC}}^{W \times H}$ denote the width of the current block, the height of the current block, the memory bandwidth of RMC, and the memory bandwidth of OBMC, respectively. Here, $M_{\text{RMC}}^{W \times H}$ can be calculated as

$$\begin{aligned} M_{\text{RMC}}^{W \times H} &= (W + t_{\text{C}} - 1) * (H + t_{\text{C}} - 1) \\ &\quad * n_{\text{C}} * \left(\frac{P}{W * H} \right), \end{aligned} \tag{3.2}$$

where P indicates the number of samples for the pipeline processing of RMC, which depends on the implementation, and 16×16 samples are provided in the VVC case as described in Sec. 3.2.2. $M_{\text{OBMC}}^{W \times H}$ can also be calculated as

$$\begin{aligned} M_{\text{OBMC}}^{W \times H} &= (M_{\text{OBMC}_{\text{T}}}^{W \times H} * \frac{W}{w} + M_{\text{OBMC}_{\text{L}}}^{W \times H} * \frac{H}{h}) \\ &\quad * n_{\text{N}} * \left(\frac{P}{W * H} \right), \end{aligned} \tag{3.3}$$

where $M_{\text{OBMC}_{\text{T}}}^{W \times H}$, $M_{\text{OBMC}_{\text{L}}}^{W \times H}$, w, and h represent the memory bandwidth of OBMC for the top-side block boundary, the memory bandwidth of OBMC for the left-side block boundary, the width of SbOBMC, and the height of SbOBMC, respectively. Finally, $M_{\text{OBMC}_{\text{T}}}^{W \times H}$ can be calculated as

$$\begin{aligned} M_{\text{OBMC}_{\text{T}}}^{W \times H} &= (w + t_{\text{N}} - 1) \\ &\quad * \left[\min\left(\frac{h}{l_1}, l_2 \right) + t_{\text{N}} - 1 \right]. \end{aligned} \tag{3.4}$$

Table 3.1: Combination of the determinants for OBMC in the conventional methods. $s_{\text{min/Uni}}$ and $s_{\text{min/Bi}}$ denote the minimum current block size for uni-prediction and bi-prediction of RMC.

Method	$s_{\text{min/Uni}}$	$s_{\text{min/Bi}}$	n_{C}	n_{N}
Chen2015	4×8	8×8	2	2
Chen2015/PhbtC	4×8	8×8	1	2
Chen2015/PhbtCN	4×8	8×8	1	1 ($2 \rightarrow 1$)
Lin2019	8×8	-	1	1 ($2 \rightarrow 1$)

Here, l_1 and l_2 indicate OBMC blending lines for smaller size blocks and larger size blocks, respectively. $M_{\text{OBMC}_\text{L}}^{W \times H}$ can also be calculated as $M_{\text{OBMC}_\text{T}}^{W \times H}$.

Tab. 3.1 and Fig. 3.6 show the M_{Wst} of Chen2015 and Lin2019 calculated with the derived formula, including non-proposed t_{N} and considering VVC. For the detailed analysis, M_{Wst} of two additional conditions based on Chen2015 is compared as described in "Chen2015/PhbtC" and "Chen2015/PhbtCN" in Fig. 3.6. Chen2015/PhbtC introduces OBMC prohibition for the bi-prediction current blocks, i.e., OBMC can be applied for uni-prediction current blocks having a bi-prediction or uni-prediction neighboring block in this method. In contrast, Chen2015/PhbtCN further introduces OBMC prohibition for uni-prediction current blocks having bi-prediction neighboring blocks, i.e., OBMC can be applied only for uni-prediction current blocks having uni-prediction neighboring blocks in this method. The difference of the OBMC applicable condition for each method is shown in Tab. 3.1. All M_{Wst} is calculated using only the luma component to simplify the comparison. The M_{WstVVC} is 2,024 $(= (16+7) * (4+7) * 2 * (16 * 16)/(4 * 16))$ reference samples, which are required in RMC for four sets of the bi-prediction $4 \times 16/16 \times 4$ current block with an 8-tap interpolation filter. Note that VVC has a 4×4 block-wise MC, i.e., affine MC, but its required memory bandwidth is constrained so as not to exceed the M_{WstVVC} [42]. Therefore, we focus on the RMC and OBMC.

From Fig. 3.6, it is clear that the bottle-neck is the application of OBMC for the bi-prediction current blocks having bi-prediction neighboring blocks since the M_{Wst} of Chen2015 exceeds the M_{WstVVC} by over 3.8 times ($\simeq$ 7,744/2,024) with the 8-tap filter and more than 1.7 times ($\simeq$ 3,520/2,024) even with the 2-tap filter. This is because $M_{\text{OBMC}}^{W \times H}$ is purely added into the M_{WstVVC} in Chen2015. Chen2015/PhbtC still exceeds the M_{WstVVC} even with the 2-tap filter, whereas Chen2015/PhbtCN and Lin2019 with the 2-tap filter become smaller than the M_{WstVVC}. This means that uni-prediction-based OBMC with the 2-tap filter can reduce the M_{Wst} lower than M_{WstVVC}. However, n_{N} and t_{N} have room to be adaptive in the larger size blocks since the required memory bandwidth becomes smaller in these blocks.

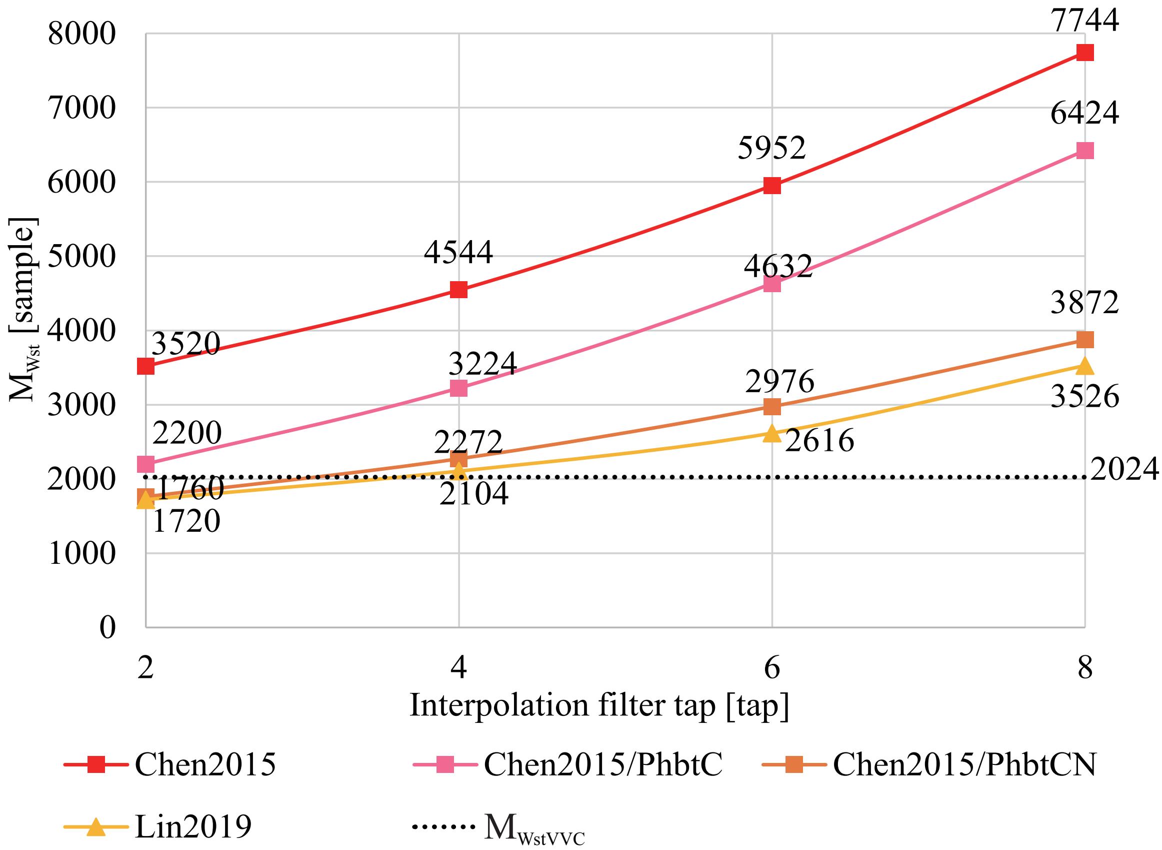


Figure 3.6: Analysis of the M_{Wst} for each OBMC method using 2, 4, 6, and 8-tap interpolation filters fixedly. (Copyright(C)2023 ITE, [2] Fig. 6)

3.4 Proposed Method

We propose a memory bandwidth constrained OBMC method with an adaptive number of motion vectors and interpolation filter taps of the neighboring blocks depending on the current block sizes (i.e., $n_{\mathrm{N}}^{W\times H} and t_{\mathrm{N}}^{W\times H}$), to retain the potential coding performance of OBMC while constraining $M_{\mathrm{OBMC}}^{W\times H}$. In this chapter, we tried to generalize the proposed method [43] as an objective function that maximizes $M_{\mathrm{Inter}}^{W\times H}$ with the $n_{\mathrm{N}}^{W\times H}$ and $t_{\mathrm{N}}^{W\times H}$ as variables such that the constraint, $M_{\mathrm{Inter}}^{W\times H}(n_N, t_N) \ < \ \mathrm{M}$, is satisfied. Here, M is an arbitrary memory bandwidth. This is because maximizing $n_{\mathrm{N}}^{W\times H}$ and $t_{\mathrm{N}}^{W\times H}$ contributes to increasing the coding performance gain of OBMC. The proposed method can be generalized using the following formula:

$$\hat{n}_{\mathrm{N}}^{W\times H}, \quad \hat{t}_{\mathrm{N}}^{W\times H} = \underset{n_{\mathrm{N}}, t_{\mathrm{N}}}{\mathrm{argmax}} M_{\mathrm{Inter}}^{W\times H}(n_{\mathrm{N}}^{W\times H}, t_{\mathrm{N}}^{W\times H})$$
$$\text{s.t. } M_{\mathrm{Inter}}^{W\times H}(n_{\mathrm{N}}^{W\times H}, t_{\mathrm{N}}^{W\times H}) < \mathrm{M}, \tag{3.5}$$

where $\hat{n}_{\mathrm{N}}^{W\times H}$ and $\hat{t}_{\mathrm{N}}^{W\times H}$ are the maximized combination of the interpolation filter taps and number of motion vectors of neighboring blocks for each current block size, $W \times H$, as an output of the formula.

The flowchart of the searching algorithm for the combination of $\hat{n}_{\mathrm{N}}^{W\times H}$ and $\hat{t}_{\mathrm{N}}^{W\times H}$ is shown in Fig. 3.7. The flowchart mainly consists of the following seven steps.

S1) Maximize the value of $n_{\mathrm{N}}^{W\times H}$ and proceed to S2

S2) Maximize the value of $t_{\mathrm{N}}^{W\times H}$ and proceed to S3

S3) Evaluate whether the constraint is satisfied with the current $n_{\mathrm{N}}^{W\times H}$ and $t_{\mathrm{N}}^{W\times H}$, and if so, determine them as $\hat{n}_{\mathrm{N}}^{W\times H}$ or $\hat{t}_{\mathrm{N}}^{W\times H}$, and proceed to S7. If not, proceed to S4

S4) Evaluate whether $t_{\mathrm{N}}^{W\times H}$ is the minimum value, and if so, proceed to S5. If not, reduce $t_{\mathrm{N}}^{W\times H}$, and return to S3

S5) Evaluate whether $n_{\mathrm{N}}^{W\times H}$ is the minimum value, and if so, proceed to S6. If not, reduce $n_{\mathrm{N}}^{W\times H}$, and return to S2

S6) Evaluate whether the constraint is satisfied with the current $n_{\mathrm{N}}^{W\times H}$ and $t_{\mathrm{N}}^{W\times H}$, and if so, determine them as $\hat{n}_{\mathrm{N}}^{W\times H}$ and $\hat{t}_{\mathrm{N}}^{W\times H}$, and proceed to S7. If not, determine that none of those that satisfy the constraints have been found, and proceed to S7 Here, the reason for prioritizing $n_{\mathrm{N}}^{W\times H}$ over $t_{\mathrm{N}}^{W\times H}$ in this search algorithm is that $n_{\mathrm{N}}^{W\times H}$ has a greater impact on coding performance. Note that whether OBMC is applied or not is finally determined by $obmc\_flag$ signaled by the encoder in the proposed method.

As an example, in this chapter, the OBMC applicable conditions without exceeding 1.0 and 1.5 times $\mathrm{M_{WstVVC}}$ (the “Proposal” and “Proposal/1.5×$\mathrm{M_{WstVVC}}$” hereinafter) are derived from the formula provided that these memory bandwidths are given as the arbitrary memory bandwidth M. Fig. 3.8 shows the derived OBMC applicable conditions without exceeding 1.0 and 1.5 times $\mathrm{M_{WstVVC}}$, i.e., $\hat{n}_{\mathrm{N}}^{W\times H}$ and $\hat{t}_{\mathrm{N}}^{W\times H}$ such that the constraint is satisfied in all sizes of the current block. In the proposed OBMC applicable conditions, 2, 4, 6, and 8-tap filters used in VVC can be selected for the $\hat{t}_{\mathrm{N}}^{W\times H}$ of the luma component. For the down-sampled chroma components, the half-tap filter for the luma component is utilized, but the 2-tap filter is used when the luma filter is 2-tap. In both conditions, OBMC can be applied for the $4 \times 8/8 \times 4$ uni-prediction current blocks having uni-prediction neighboring blocks, and for the other size current blocks having bi-prediction neighboring blocks, in which OBMC is prohibited in Lin2019.

The comparison analyses of $M_{\mathrm{Inter}}^{W\times H}$ for each current block size regarding the $\mathrm{RMC_{Uni}}$, $\mathrm{RMC_{Bi}}$, Chen2015, the Lin2019/2-tap, Lin2019, Proposal, and Proposal/1.5×$\mathrm{M_{WstVVC}}$, are shown in Fig. 3.9. The Proposal can maximize $M_{\mathrm{Inter}}^{W\times H}$ (including M_{OBMC}) without exceeding the $\mathrm{M_{WstVVC}}$ through all size blocks as shown in Fig. 3.9. This is expected to maintain the potential coding performance gain of OBMC. The Proposal/1.5×$\mathrm{M_{WstVVC}}$ can also bring $M_{\mathrm{Inter}}^{W\times H}$ closer to 1.5 times $\mathrm{M_{WstVVC}}$ for small size blocks, but not for larger size blocks (e.g., 64×64, 64×128, and 128×128), which is the same level as the Proposal. This is expected to further preserve the potential coding performance gain in low-resolution sequences with many smaller-size blocks. These expectations will be clarified in Sec. 3.4

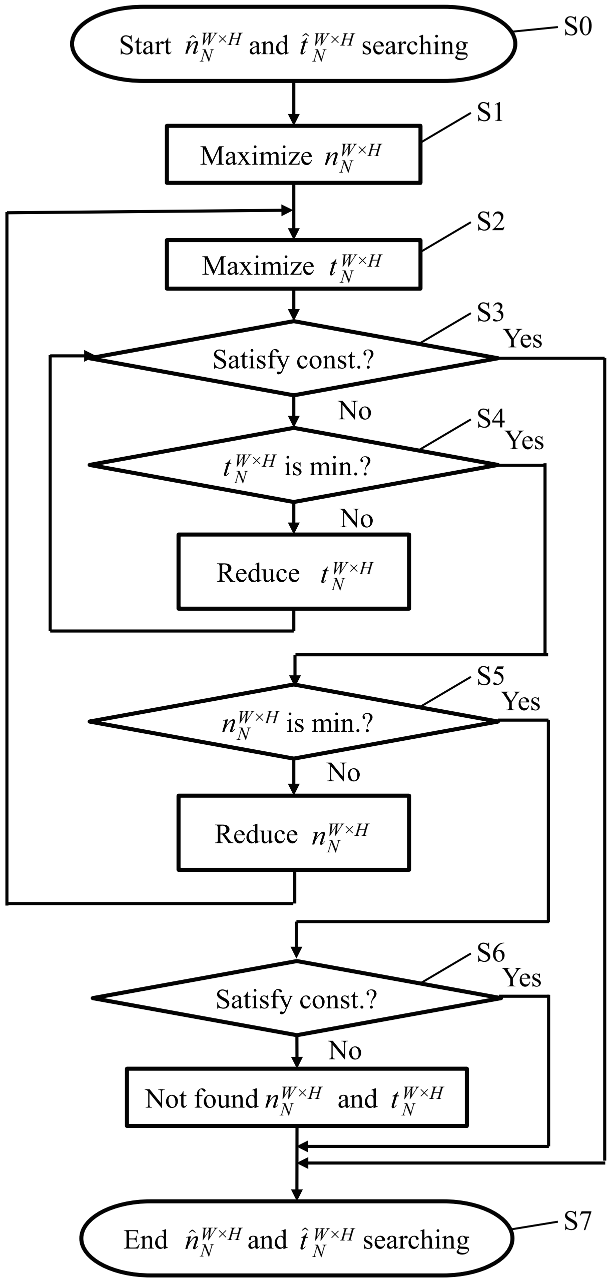


Figure 3.7: The flowchart of the searching algorithm for the combination of $\hat{n}_{\mathrm{N}}^{W\times H}$ and $\hat{t}_{\mathrm{N}}^{W\times H}$ in the proposed method. (Copyright(C)2023 ITE, [2] Fig. 7)

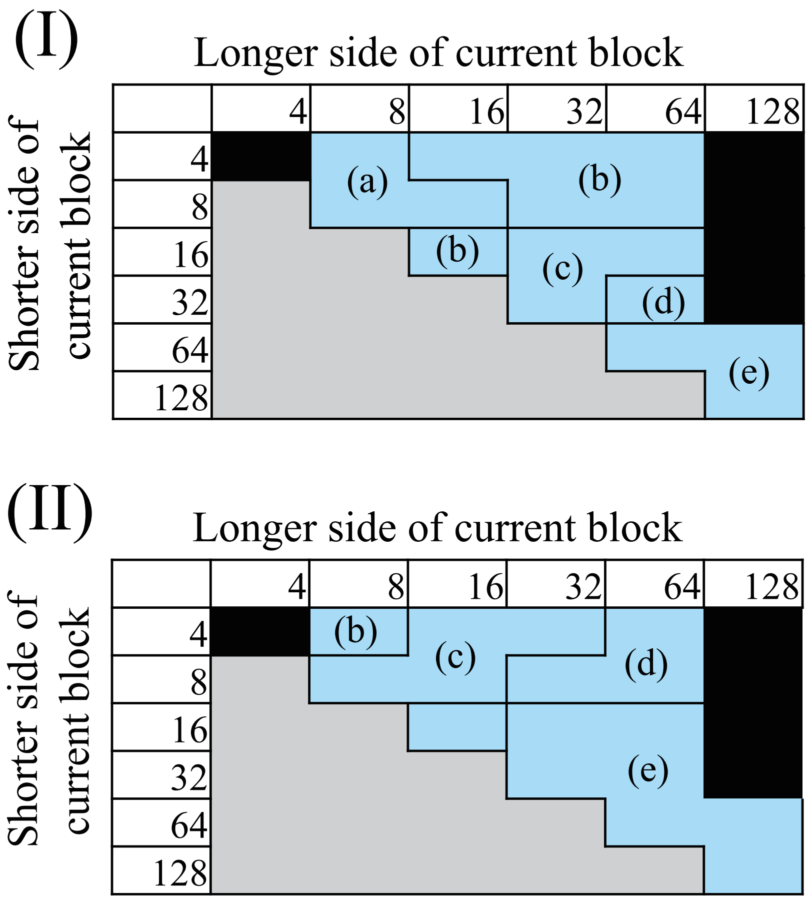


Figure 3.8: The proposed OBMC applicable conditions with 1.0 and 1.5 times $\mathrm{M_{WstVVC}}$: (I) = 1.0 times $\mathrm{M_{WstVVC}}$ and (II) = 1.5 times $\mathrm{M_{WstVVC}}$. (a)–(e) indicate the combination of $\hat{n}_{\mathrm{N}}^{W\times H}$ and $\hat{t}_{\mathrm{N}}^{W\times H}$: (a) = (1, 2), (b) = (2, 2), (c) = (2, 4), (d) = (2, 6), and (e) = (2, 8). Black areas denote OBMC's non-applicable areas. (Copyright(C)2023 ITE, [2] Fig. 8)

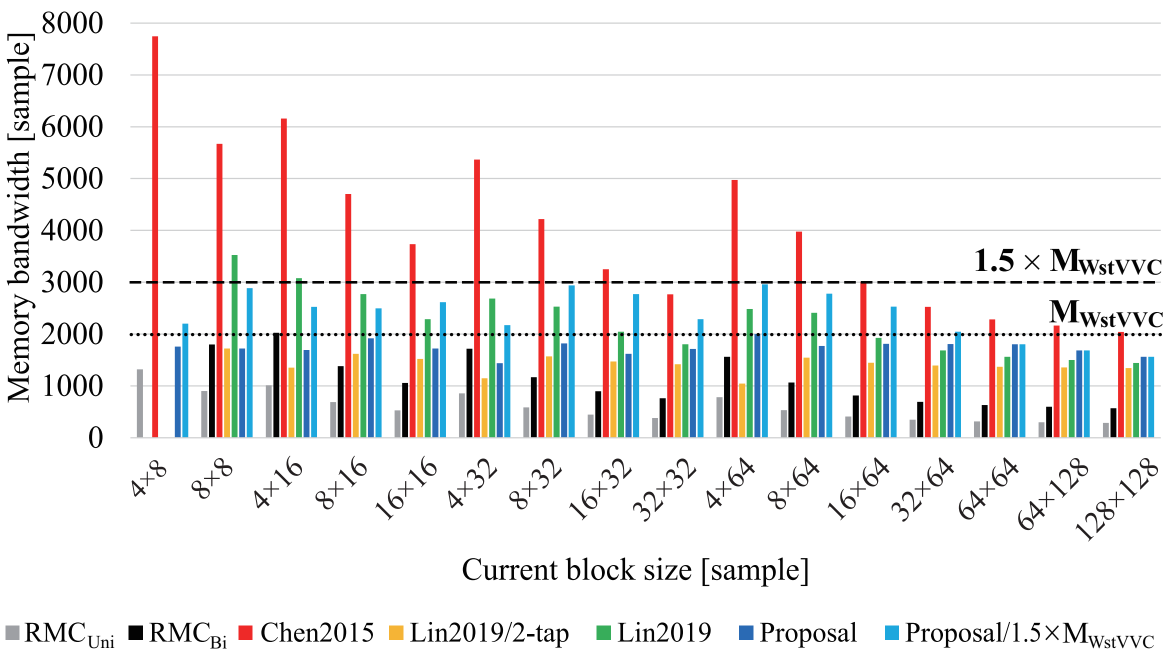


Figure 3.9: Comparison analysis of the memory bandwidth versus the current block size for each method based on VVC. RMC_{Uni} and RMC_{Bi} denote RMC for uni-prediction and bi-prediction. Copyright(C)2023 ITE, [2] Fig.9）

Table 3.2: Details of the VTM CTC test sequences from classes A to F categorized by resolutions, frame rates, and video content type, i.e., Camera-captured Content (CC), Screen Content (SC), and Mixed Content with CC and SC (MC).

Class	Resolutions [pixels $\times$ lines]	Frame rates	Video content
A1	3840×2160	30–60	CC
A2	3840×2160	50–60	CC
B	1920×1080	50–60	CC
C	832×480	30–60	CC
D	416×240	30–60	CC
F	$832 \times 480 \sim 1920 \times 1080$	20–60	SC and MC

3.5 Experimental Results and Discussion

3.5.1 Test Conditions

1) Software Settings: The VTM-10 [44] was used as the baseline software in our simulation experiments, and the proposed method was implemented in the VTM-10. To verify the trade-offs between the coding performance and memory bandwidth, the simulations for a total of nine different methods, i.e., Chen2015, Chen2015/PhbtC, Chen2015/PhbtCN, Lin2019, Lin2019 with 2–8tap filters, the Proposal, and Proposal/$1.5\times M_{\mathrm{WstVVC}}$ were conducted.

2) Encoder Configurations: The coding conditions were basically followed with the VTM Common Test Condition (CTC) [45]. The Random Access (RA) configuration, utilized for general video transmission, was used since OBMC is an inter prediction tool. The test sequences from classes A to F were used as listed in RA. They are categorized with different resolutions, frame rates, and video content as shown in Tab. 3.2. The partial test sequences, Class A(A1/A2) and B were encoded by only the first group of pictures in these sequences to reduce the encoding runtime. Four Quantization Parameter (QP) values, 22, 27, 32, and 37, were used for each test sequence to generate the different rate points.

3) Evaluation Metrics: The coding performance was evaluated by the Bjøntegaard Delta bitrate (BD-rate) of the luma and two chroma components, i.e., BD-rate Y, U, and V components (BDY, BDU, and BDV) [27, 28]. The BD-rate is the evaluation index used to quantify the difference of the generated bitrate for the identical level of Peak Signal-to-Noise Ratio (PSNR) between the two coding methods. The negative and positive values of BD-rate indicate the coding performance gains and losses. The complexity was evaluated by the two coding methods' relative encoding and decoding runtime (EncT and DecT) measured on a homogeneous cluster PC. In accordance with the VTM CTC, Classes D and F results are not included in the overall results.

3.5.2 Comparison of Overall Results

The overall results of each method compared to the VTM-10 in RA configuration, evaluated by BDY, BDU, BDV, and DecT, are listed in Tab. 3.3. In addition, the trade-off between BDY and the M_{Wst} of each method is shown in Fig. 3.10.

1) Coding performance: Tab. 3.3 shows that all the methods provide the coding performance gain against the VTM-10. These results prove the OBMC further improves the coding performance beyond VVC. Specifically, Chen2015 (0.33% gain) can be assumed to attain the full coding performance of OBMC. The Proposal achieves its comparable performance (0.22% gain) without exceeding the $\mathrm{M_{WstVVC}}$, which is still better than Lin2019/2-tap (0.12% gain). Proposal/$1.5\times\mathrm{M_{WstVVC}}$ brings the further coding performance (0.25% gain).

Fig. 3.10 shows that the Proposal achieves a better trade-off than those of the Chen2015 and Lin2019 series in terms of BDY and the M_{Wst}. The reason that the Proposal achieves the best trade-off is discussed as follows. First, the Proposal's coding performance gain is smaller than Chen2015/PhbtC but larger than Chen2015/PhbtCN. Comparing the three methods regarding the OBMC applicable condition, the common difference of the Proposal from the other two methods is that the shorter-tap filter is utilized for the smaller size current blocks. Here, it is clear that the shorter-tap filters do not affect the coding performance when comparing the Lin2019 series. As for the Chen2015/PhbtC, the other difference except for the filter is the OBMC prohibition for partial smaller size current blocks having bi-prediction neighboring blocks as in Fig. 3.8. On the other hand, as for the Chen2015/PhbtCN, the difference except for the filter is the OBMC application for partial larger size current blocks having bi-prediction neighboring blocks as in Fig. 3.8, which contributes to the coding performance gain of the Proposal against Chen2015/PhbtCN. In addition, a comparison of Chen2015/PhbtCN and Lin2019 shows that the OBMC application for the 4×8 size current block having uni-prediction neighboring blocks as in Fig. 3.8 also contributes to the coding performance gain of the Proposal against Lin2019.

2) Complexity: Tab. 3.3 shows that all methods increase EncT and DecT against the VTM-10. These results prove the OBMC increases the encoding and decoding runtime beyond VVC. The EncT increments of all the methods are not different, whereas the DecT increments of Chen2015 are larger than those of the other methods. This suggests that the OBMC application for the bi-prediction current block clearly further increases only DecT. Regarding EncT, this is because the number of RD cost calculations for OBMC signaling is the same between Chen2015 and Chen2015/PhbtC as described in Sec. 3.2.3. The reason that DecT of Chen2015/PhbtCN is the same level as Chen2015/PhbtC even with the additional constraint of OBMC is that Chen2015/PhbtCN maintains the OBMC application rates for the uni-prediction current blocks having bi-prediction neighboring blocks, with the conversion from bi-prediction to uni-prediction as described in Sec. 3.2.3.

Table 3.3: Overall results of the conventional and proposed methods over VTM-10 in RA configuration, which is evaluated by BDY [%], BDU [%], BDV [%] EncT [%], and DecT [%].

Method	BDY	BDU	BDV	EncT	DecT
Chen2015	-0.33	-1.04	-0.83	105	106
Chen2015/PhbtC	-0.27	-0.57	-0.57	104	103
Chen2015/PhbtCN	-0.14	-0.57	-0.28	105	104
Lin2019	-0.12	-0.43	-0.40	105	104
Lin2019/6-tap	-0.11	-0.24	-0.35	105	104
Lin2019/4-tap	-0.10	-0.43	-0.26	105	103
Lin2019/2-tap	-0.12	-0.43	-0.26	105	103
Proposal	-0.22	-0.54	-0.55	105	103
Proposal/1.5×M_{WstVVC}	-0.25	-0.64	-0.61	105	103

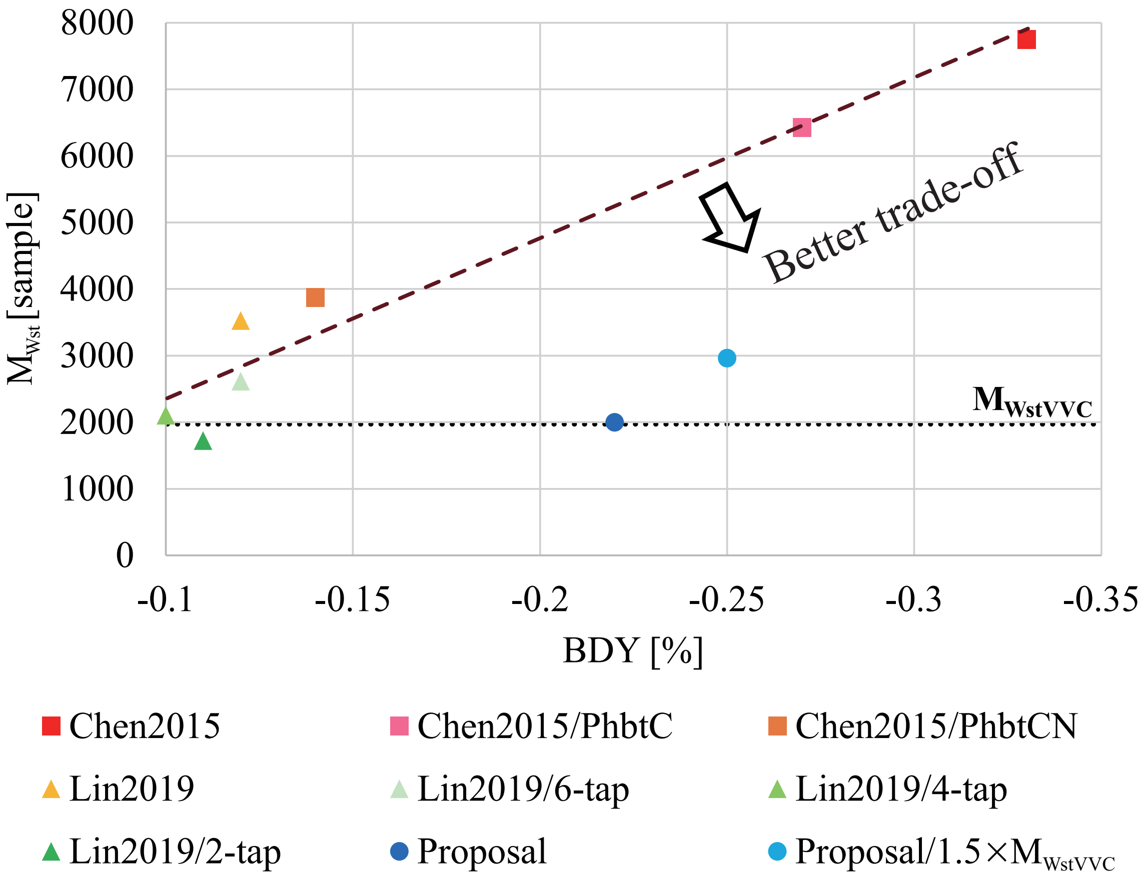


Figure 3.10: The analysis of the trade-off between BDY [%] and M_{Wst} [sample] for each method. Copyright(C)2023 ITE, [2] Fig.10）

Table 3.4: Sequence-level results of Chen2015, Lin2019, the Proposal, Proposal/1.5×M_{WstVVC} over VTM-10 in RA configuration, which is evaluated by BDY [%], BDU [%], BDV [%], EncT [%], and DecT [%].

	Chen2015			Chen2015/PhbtC			Chen2015/PhbtCN			Lin2019			Proposal			Proposal/1.5×M_{WstVVC}		
Sequence	BDY	EncT	DecT	BDY	EncT	DecT	BDY	EncT	DecT	BDY	EncT	DecT	BDY	EncT	DecT	BDY	EncT	DecT
Tango2	-0.31	105	104	-0.29	104	100	-0.10	105	101	-0.18	104	102	-0.27	104	101	-0.30	104	103
FoodMarket4	0.06	104	104	-0.06	103	102	0.05	104	102	0.03	104	102	-0.07	103	102	-0.01	103	101
CampfireParty2	-0.20	104	103	-0.19	104	102	-0.13	105	102	-0.15	104	102	-0.13	104	102	-0.18	104	101
Average Class A1	**-0.15**	**105**	**104**	**-0.18**	**104**	**102**	**-0.06**	**104**	**102**	**-0.10**	**104**	**102**	**-0.16**	**104**	**102**	**-0.16**	**104**	**102**
CatRobot1	-0.33	106	105	-0.35	104	102	-0.14	105	102	-0.23	105	102	-0.16	105	102	-0.41	104	102
DaylightRoad2	-0.09	106	106	-0.29	105	103	-0.05	105	104	-0.05	105	104	-0.23	105	103	-0.26	104	104
ParkRunning3	-0.51	106	107	-0.37	104	103	-0.27	105	104	-0.19	105	103	-0.35	104	103	-0.37	105	105
Average Class A2	**-0.31**	**106**	**106**	**-0.34**	**105**	**102**	**-0.15**	**105**	**103**	**-0.16**	**105**	**103**	**-0.25**	**105**	**103**	**-0.34**	**104**	**103**
MarketPlace	-0.47	106	108	-0.35	105	104	-0.18	106	106	-0.21	105	106	-0.20	105	105	-0.32	105	103
RitualDance	-0.18	105	105	-0.19	104	103	-0.17	105	103	0.04	105	107	-0.13	104	104	-0.15	104	103
Cactus	-0.04	106	105	-0.11	105	104	0.07	105	103	-0.10	105	103	-0.15	105	104	-0.14	105	103
BasketballDrive	-0.16	105	107	-0.03	104	102	0.00	105	103	0.02	105	104	-0.19	104	102	-0.02	104	99
BQTerrace	-0.43	106	105	-0.11	104	104	-0.03	106	103	0.10	105	102	0.01	105	103	-0.10	105	102
Average Class B	**-0.25**	**106**	**106**	**-0.16**	**105**	**103**	**-0.06**	**105**	**104**	**-0.03**	**105**	**105**	**-0.13**	**105**	**104**	**-0.14**	**105**	**104**
BasketballDrill	-0.65	105	106	-0.41	104	104	-0.27	106	104	-0.24	105	102	-0.35	105	103	-0.40	105	106
BQMall	-0.46	106	107	-0.36	105	105	-0.25	106	105	-0.22	105	105	-0.35	105	105	-0.36	106	105
PartyScene	-0.60	106	108	-0.45	105	106	-0.33	106	105	-0.20	105	104	-0.29	105	105	-0.37	106	106
RaceHorses	-0.55	106	110	-0.55	105	106	-0.31	105	107	-0.23	105	105	-0.42	105	106	-0.42	105	106
Average Class C	**-0.57**	**106**	**108**	**-0.44**	**105**	**105**	**-0.29**	**106**	**105**	**-0.22**	**105**	**104**	**-0.35**	**105**	**105**	**-0.39**	**105**	**104**
BasketballPass	-0.37	106	109	-0.31	105	106	-0.17	105	106	-0.03	105	105	-0.17	105	106	-0.26	105	108
BQSquare	-0.34	105	105	-0.19	105	103	-0.09	106	103	-0.03	104	101	0.06	105	103	-0.11	104	103
BlowingBubbles	-0.63	106	107	-0.57	105	105	-0.40	106	105	-0.25	105	104	-0.40	105	105	-0.44	105	106
RaceHorsesC	-0.67	106	111	-0.63	105	108	-0.55	106	107	-0.34	106	107	-0.53	106	108	-0.59	105	109
Average Class D	**-0.50**	**106**	**108**	**-0.42**	**105**	**105**	**-0.30**	**106**	**105**	**-0.16**	**105**	**104**	**-0.26**	**105**	**105**	**-0.35**	**105**	**104**
BasketballDrillText	-0.42	104	108	-0.36	103	106	-0.23	104	105	-0.15	104	105	-0.32	104	105	-0.36	104	108
ArenaOfValor	-0.15	104	108	-0.29	104	105	-0.28	105	104	-0.06	104	105	-0.27	104	105	-0.26	104	104
SlideEditing	0.18	101	101	0.13	101	102	0.13	102	102	0.10	102	102	0.13	101	101	0.15	101	101
SlideShow	0.71	103	102	0.51	103	102	0.54	103	103	0.54	103	102	0.53	103	102	0.55	103	103
Average Class F	**0.08**	**103**	**105**	**0.00**	**103**	**104**	**0.04**	**103**	**103**	**0.11**	**103**	**103**	**0.02**	**103**	**103**	**0.02**	**105**	**104**
Overall	**-0.33**	**105**	**106**	**-0.27**	**104**	**103**	**-0.14**	**105**	**104**	**-0.12**	**105**	**104**	**-0.22**	**105**	**103**	**-0.25**	**105**	**103**

3.5.3 Comparison of Sequence-level Results

To analyze the coding performance gain by OBMC observed in the overall results, the sequence-level results of each method are compared with the VTM-10 as shown in Tab. 3.4.

1) Coding performance: First, from comparing the average gain for each class, the coding performance gain is larger for low-resolution test sequences in common with all methods. This is clear by comparing RaceHorsesC in Class C and RaceHorses in Class D which differ only in resolution. Since the selection rate of smaller blocks is higher for low-resolution test sequences than for high-resolution test sequences, and the ratio of the OBMC applied area is higher for smaller blocks than for larger blocks, more coding performance gain by OBMC can be obtained. This characteristic matches the expectation described in Sec. 3.1 that Proposal/1.5x can provide more coding performance gain than the Proposal, especially in low-resolution sequences due to the extension of filter taps for smaller blocks.

Second, larger coding performance gain can be observed in the test sequences with various complicated motions in a picture such as Tango2, ParkRunning3, MarketPlace, and RaceHorses. Here, Tango2 and Racehorses have several moving objects with different motions, whereas ParkRunning3 and MarketPlace have camera shakes, which easily raise the difference in motion vectors between the blocks. In contrast, regarding the test sequences without these motions, such as BQTerrace and BQSquare, a significant small coding performance gain or even coding loss can be observed in all the methods except for Chen2015. These tendencies are consistent with the originally expected effects of OBMC as described in Sec. 3.1. Moreover, the coding performance gain of Chen 2015 in BQTerrace and BQSquare suggests that the OBMC application for the bi-prediction current block can further improve the coding performance of these sequences where the bi-prediction is originally effective. Finally, the coding losses of SC, such as SlideEditing and SlideShow, can be observed in common with all methods. This is because OBMC overshoots the block boundaries, including shaped edges in SC, degrading the objective qualities. For example, we can avoid it with the OBMC disabling flag for the overall sequence.

2) Complexity: The same tendency can be observed in DecT as the coding performance, i.e., the low-resolution test sequences have larger DecT. It demonstrates that these test sequences' high OBMC applied sample rates increase the DecT. In contrast, EncT is at the same level for high and low-resolution sequences.

3.5.4 Picture-level Analysis

1) Rate-Distortion curve characteristics: To analyze the coding performance gains and losses as discussed in Sec. 3.5.3, the Rate-Distortion (RD) curve of each method for RaceHorsesC and BQTerrace is compared as shown in Fig. 3.11. We selected RaceHorsesC and BQTerrace on behalf of the test sequences since they have the clearest tendencies as described in Sec. 3.5.3.

First, the comparison of each method for RaceHorsesC as shown in Fig. 3.11(a)–(c) shows that the coding performance gains of all methods against VTM-10 come from the bitrate savings,

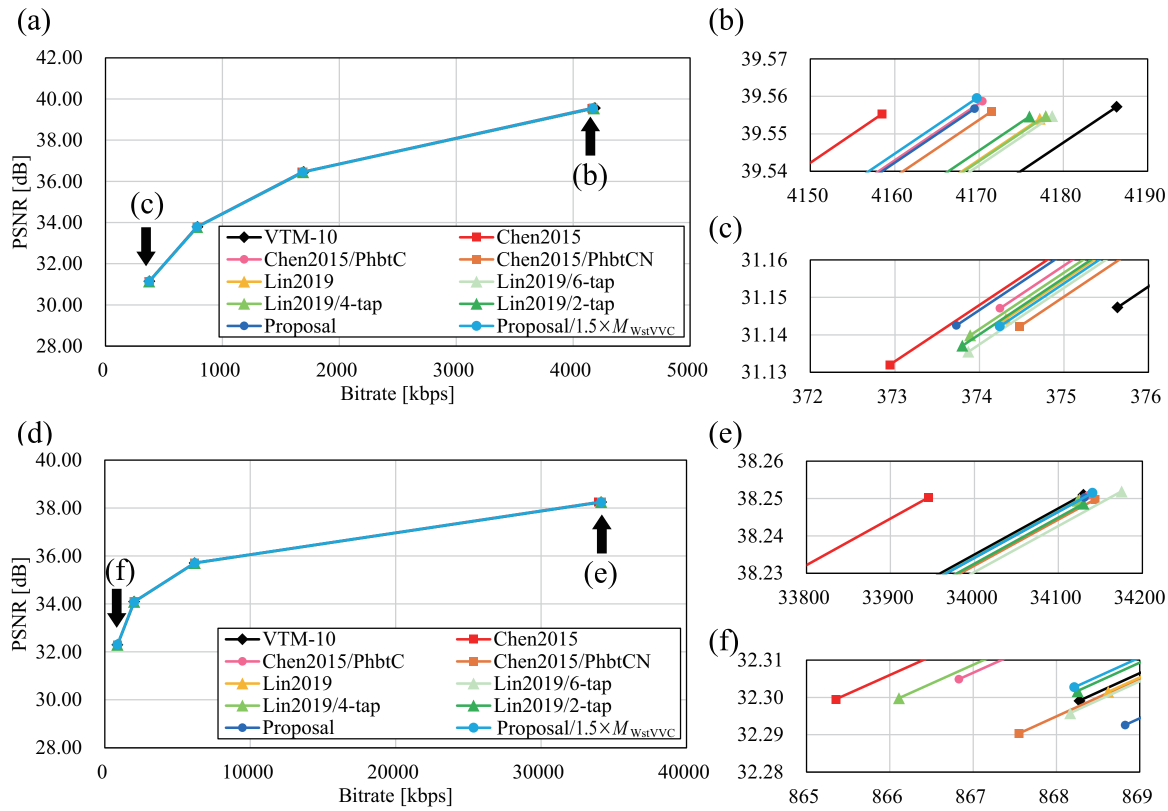


Figure 3.11: Rate distortion curves of the VTM-10 and each method for RaceHorsesC and BQTerrace. (a)-(c) RaceHorsesC, (d)-(f) BQTerrace. (Copyright(C)2023 ITE, [2] Fig. 11)

not from the improvement of the objective quality. The OBMC removes the discontinuity of the block boundary, decreases the residuals, and consequently reduces the bitrate. The larger bitrate savings can be observed at QP = 22 compared to those at QP = 37. This is because the small quantization step of the small QP raises the selection rates of smaller-size blocks and provides the coding performance gain and the effects seen in low-resolution test sequences. In addition, no difference among the Lin2019 series with different t_{N} corresponds to the discussion in Sec. 3.5.2.

Second, the comparison of each method for BQTerrace as shown in Fig. 3.11(d)–(e) shows that the coding performance gain of Chen2015 against VTM-10 also comes from the bitrate savings. In contrast, the bitrates for the other method are higher than VTM-10 at QP = 37, whereas the PSNR for some methods is smaller than VTM-10. The signaling overhead for OBMC and the degradation of the objective quality by OBMC prohibition for the bi-prediction current block seem to affect them, respectively.

2) OBMC applied sample rates: To reveal the discussion so far, we compare Chen2015, Lin2019, and the Proposal by two types of ratio of the OBMC applied samples for each current block size as shown in Fig. 3.12. One is the ratio of OBMC-applied samples to the inter frame samples $R_{\mathrm{OBMC/InterFrame}}$, which can compare the relative number of OBMC-applied samples

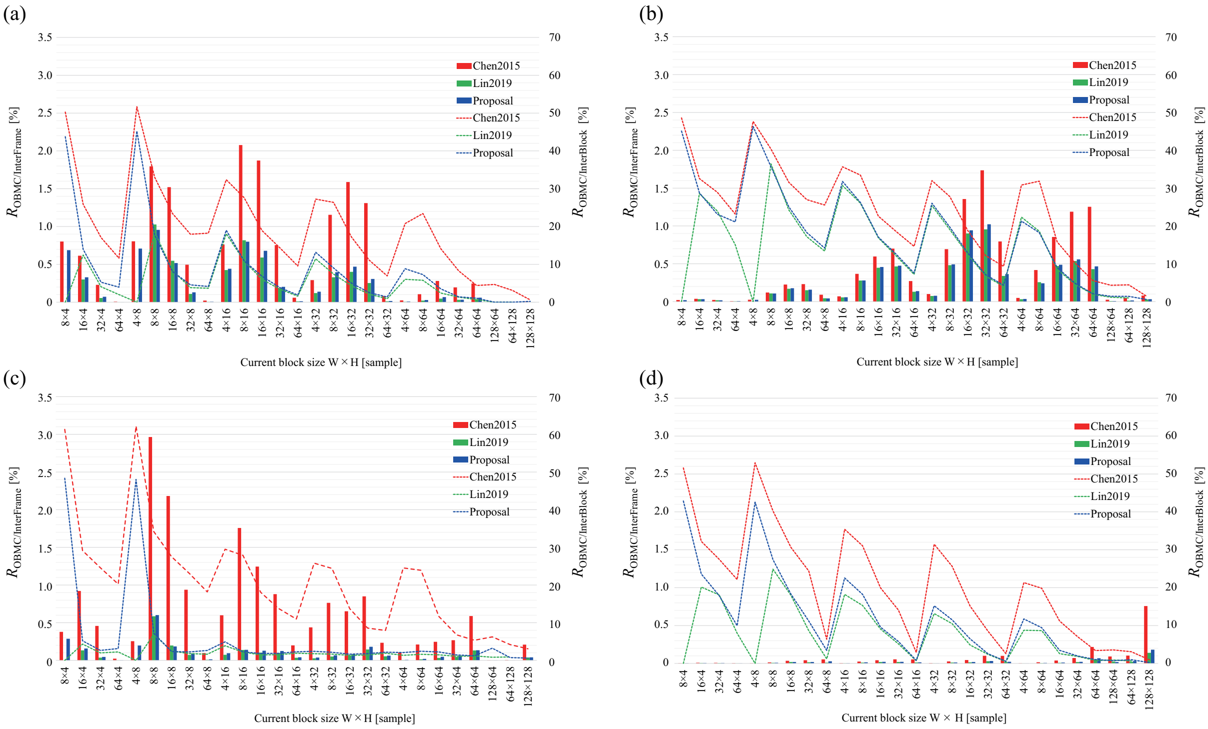


Figure 3.12: Comparison analysis of Chen2015, Lin2019, and the Proposal regarding the current block sizes versus a ratio of the OBMC applied samples to the inter frame samples $R_{\text{OBMC/InterFrame}}$ (shown as bar and left axis) and a ratio of those to the inter block samples $R_{\text{OBMC/InterBlock}}$ (shown as dotted plot and right axis). The total value generated by the RaceHorsesC/BQTerrace of QP = 22/37 is the number of those samples. (a) RaceHorsesC of QP = 22, (b) RaceHorsesC of QP = 37, (c) BQTerrace of QP = 22, and (d) BQTerrace of QP = 37. (Copyright(C)2023 ITE, [2] Fig. 12)

among the three methods and can evaluate their effects of OBMC. The other is the ratio of OBMC-applied samples to inter block samples $R_{\text{OBMC/InterBlock}}$, which can estimate the effect of OBMC depending on the current block sizes. We selected the same test sequences and QPs as Fig. 3.11.

Commonly in Fig. 3.12(a)–(d), the larger the current block size is, the higher $R_{\text{OBMC/InterBlock}}$ is, which is consistent with the discussion so far. Especially for RaceHorsesC, clear existences of OBMC applied samples can be observed in all methods at QP = 22 and QP = 37. In all methods, the peaks of $R_{\text{OBMC/InterBlock}}$ shift from smaller size blocks to larger size blocks at QP = 22 versus QP = 37, but maintain the total $R_{\text{OBMC/InterBlock}}$, which is the evidence of the coding performance gain by OBMC in RaceHorsesC. The number of Chen2015 is significantly higher than those of Lin2019 and the Proposal, while the other two methods are not so different except for $4 \times 8/8 \times 4$ current blocks. This corresponds to the reasons for the different coding performance gains for these three methods observed in RaceHorsesC of Tab. 3.4 and described in Sec. 3.5.3.

As for BQTerrace, the same level of $R_{\text{OBMC/InterFrame}}$ as that in RaceHorsesC can be observed in Chen2015 at QP = 22, while the smaller $R_{\text{OBMC/InterFrame}}$ as those in RaceHorsesC can be seen in the other two methods. This is consistent with the discussion regarding the OBMC effectivity for the bi-prediction current block in this test sequence as described in Sec. 3.5.3. At QP = 37, most of the OBMC applied samples except for 128×128 samples are dispensed, which causes no coding in the Proposal or the coding loss shown in Tab. 3.4 and described in Sec. 3.5.3.

3.6 Conclusion

This chapter proposed the memory bandwidth constrained OBMC. The proposed method is generalized as the objective function with the constraint that maximizes the coding performance with the number of motion vectors and interpolation filter taps of the neighboring blocks depending on the current block sizes. The constraint is not to exceed an arbitrary memory bandwidth and we set the worst-case upper limit of the memory bandwidth of VVC as an example. Experimental results showed that the proposed method achieves an additional coding performance gain of 0.22% over VVC reference software. This gain is comparable to the full performance of the conventional bi-prediction-based OBMC (0.33%) which requires 3.8 times the memory bandwidth of VVC, but it is still better than that of the conventional uni-prediction-based OBMC (0.12%).

Chapter 4

Geometric Partitioning Mode with Inter Prediction and Intra Prediction

4.1 Introduction

4.1.1 Background

[1] Geometric partitioning mode (GPM) is a newly adopted VVC inter prediction that contributes to the coding performance gain of H.266 | Versatile Video Coding (VVC). The overview of the GPM in VVC is shown in Fig. 4.1(a) [4, 46]. Different from the regular inter prediction performed on rectangular blocks in VVC, the GPM divides a coding block into two regions by the pre-defined 64 types of straight lines, generates inter prediction samples of luma and chroma component ("samples" hereinafter) for each GPM-separated region (i.e., P_0, P_1) with different Motion Vectors (MVs) (i.e., MV_0, MV_1), and then blends them to obtain the final inter prediction samples (i.e., P_{G}). With this feature, the GPM improves the prediction accuracy at the boundary between the foreground and background with different motions. The GPM especially contributes to the coding performance gain of the coding structure for the low-latency video transmission, called Low delay B (LB) configuration of Common Test Condition (CTC) defined by the Joint Video Experts Team (JVET). The coding performance gain is up to 1.54% bitrate savings compared to the whole bitrate savings (= 30%) in the LB configuration [46, 45].

The GPM in VVC organized by two different inter predictions ("GPM-Inter/Inter" hereinafter) has room to further improve the prediction accuracy if the final prediction samples can be generated using the intra prediction as well, which generates the prediction samples by fetching the reconstructed samples adjacent to a coding block in the same picture. Furthermore, if the application of intra-prediction to GPM is realized, GPM could be applied to the lower-latency coding structure, called Low delay P (LP) configuration, since the

[1] This chapter is based on "Geometric Partitioning Mode with Inter and Intra Prediction for Beyond Versatile Video Coding" [3], by the same author, which appeared in the IEICE Transactions on Information and Systems, Copyright(C)2022 IEICE.

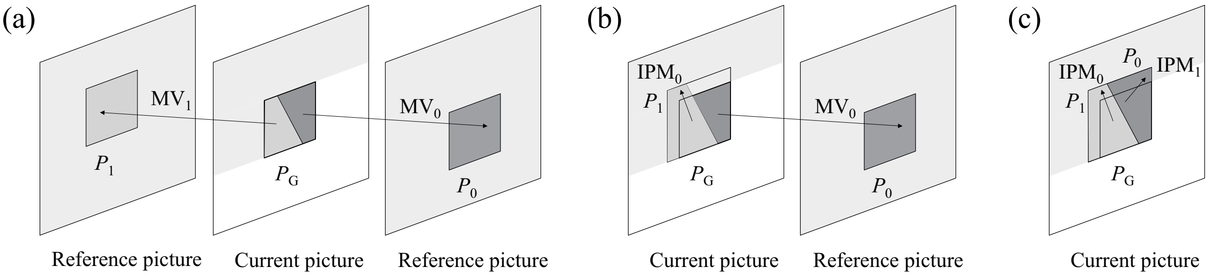


Figure 4.1: Overview of the GPM's generation process of prediction samples. (a) GPM-Inter/Inter in VVC, (b) GPM-Inter/Intra, and (c) GPM-Intra/Intra. The shaded regions of the current picture and the reference picture indicate the reconstructed sample areas available for inter and intra predictions. (Copyright(C)2022 IEICE, [3] Fig. 1)

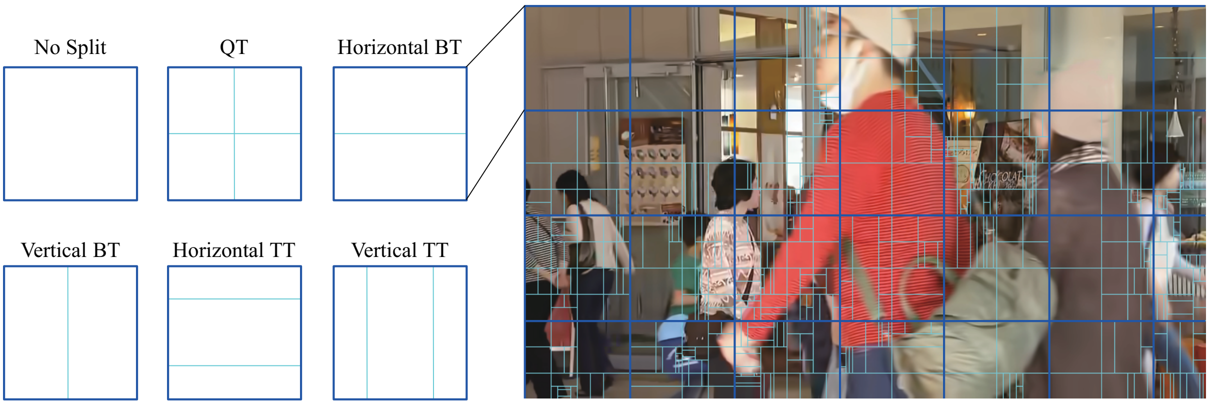


Figure 4.2: Quad-Tree (QT), Binary-Tree (BT), and Ternary-Tree (TT) block partitioning in VVC and an example of the recursive Quad-Tree plus Binary-Ternary Tree (QTBTT) block partitioning. Blue grids denote the coding tree blocks, and sky-blue lines indicate the QT, BT, or TT splitting lines. (Copyright(C)2022 IEICE, [3] Fig. 2)

inter prediction with two different MVs ("bi-prediction" hereinafter) is prohibited in LP configuration. However, efficient methods to apply the intra prediction to GPM to improve GPM have not been proposed.

Geometric partitioning of the coding blocks before the prediction stage ("GEO" hereinafter) [47]–[48] is a potential solution, while GEO significantly increases encoder and decoder complexities, such as the need for the additional transforms with adaptive shapes. Due to the high level of complexity, GEO is not adopted in VVC. Combined Inter and Intra Prediction (CIIP) with Triangular Partitions (CIIP-TP) [49] is another potential solution. Here, CIIP [46] is another new inter prediction in VVC, which generates the final prediction samples by blending inter and intra prediction samples without the additional partitioning at the prediction stage, such as GPM. CIIP-TP further extends CIIP such that rectangular coding blocks are diagonally split, the prediction samples in the two split regions are generated using inter or intra prediction, and finally combined. On the other hand, CIIP-TP is not adopted in VVC due to its small performance improvement over VVC reference software since CIIP-TP is restricted to two types of splitting shapes and only one intra-prediction mode (i.e., Planar mode).

4.1.2 Proposal

This chapter proposes introducing GPM with inter and intra predictions (GPM-Inter/Intra) as a new selectable prediction mode in GPM. In other words, introducing GPM-Inter/Intra softens the GPM application condition where only GPM-Inter/Inter can be selected. Furthermore, we propose restricting the number of Intra Prediction Modes (IPMs) applicable in GPM-Inter/Intra to suppress the signaling overhead of GPM-Inter/Intra.

4.1.3 Contribution

To evaluate the effectiveness of the proposed method, we implemented it on top of the VVC Test Model (VTM) version 11 (VTM-11) and performed the experiments according to the JVET CTC. The experimental results show that the proposed method provides additional coding performance gains of 0.17%, 0.45%, and 1.15% for the Random Access (RA), LB, and LP configurations. The proposed method improves coding gains for RA and LB by 1.3 times compared to conventional methods. In addition, subjective quality improvement by the proposed GPM-Inter/Intra can be observed in the LP configuration for test sequences where GPM is highly applied.

4.1.4 Outline

The rest of this chapter is organized as follows. Related work and the corresponding problems are explained in Sec. 4.2. Sec. 4.3 presents the details of the proposed method. Sec. 4.4 describes the experimental results and discussion. Finally, we conclude this chapter in Sec. 4.5.

4.2 Related Work

4.2.1 Block Partitioning in VVC

The maximum coding block size of VVC, i.e., coding tree block size, is extended from 64×64 samples of H.265 | High Efficiency Video Coding (HEVC) to 128×128 samples [50]. Furthermore, a recursive Quad-Tree plus Binary-Ternary Tree (QTBTT) block partitioning is introduced as shown in Fig. 4.2 [50]. The recursive QTBTT block partitioning generates various sizes and shapes of rectangular blocks including non-square blocks, making uniform the sizes of the coding blocks, prediction blocks, and transform blocks. It improves the coding performance while increasing the encoder and decoder complexity compared to those of HEVC [50].

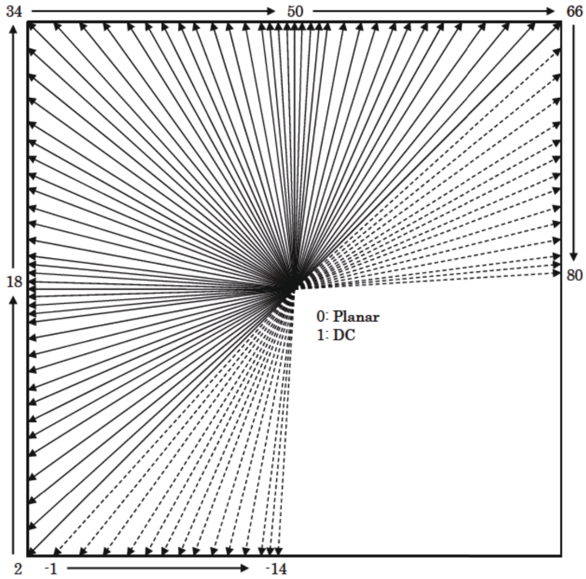


Figure 4.3: Intra prediction modes in VVC. (Copyright(C)2022 IEICE, [3] Fig. 3)

In addition, dual-tree, which allows the coding tree block of the luma and chroma components to separate coding tree structure, is newly introduced in intra slice (i.e., I slice) where only the intra prediction can be utilized, while not in inter slice (i.e., B or P slice) where the inter and intra prediction can be utilized [50]. The dual-tree improves the coding performance because the block partitioning size of the chroma components can be enlarged when the change of the sample values of the chroma components is relatively small compared to the luma component, for instance.

4.2.2 Intra Prediction in VVC

For the intra prediction for a luma component, Planar, Direct Current (DC), and angular modes similar to HEVC can be utilized in VVC. The Planar and DC modes effectively predict pictures with flat characteristics, whereas the angular mode effectively predicts the object edges. In particular, the number of angular modes is increased from 32 to 64, as shown in the solid arrows of Fig. 4.3 to improve the prediction accuracy for a coding block [51]. Moreover, wide angular modes which replace partial angular modes described as dashed arrows (= No. -1–No. -14 and No. 67–No. 80) in Fig. 4.3 are also newly adopted to improve the prediction accuracy for non-square blocks [51]. This means the total number of modes is the same as the regular angular modes. Whether the wide angular modes are applied or not is determined by the block sizes and shapes. For the intra prediction for chroma components, the direct mode, which applies the same IPM as the luma component to the chroma components to save the signaling overhead for the chroma component, is available in VVC, the same as HEVC.

In VVC, the IPM derivation is designed based on an IPM candidate list, similar to HEVC. The candidate list size (the maximum number of candidates) is extended from three in HEVC to six in VVC. Planar mode is registered preferentially, and the remaining candidates are

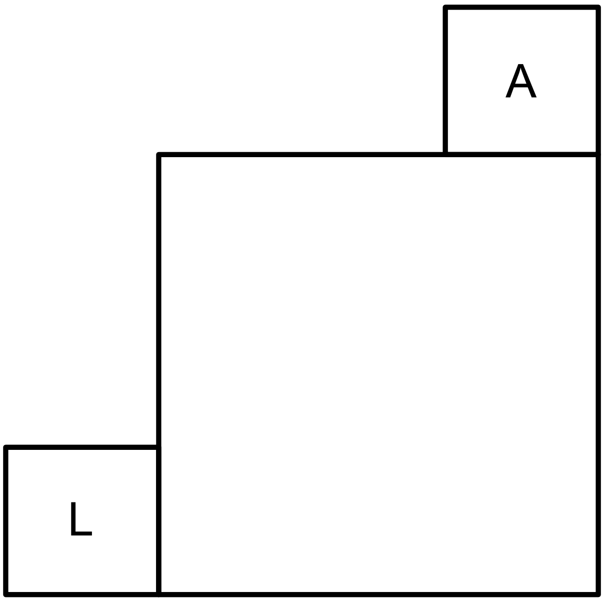


Figure 4.4: Available neighboring blocks for deriving IPM candidates in VVC. A and L denote the neighboring blocks on the above and left sides of a coding block. (Copyright(C)2022 IEICE, [3] Fig. 4)

determined depending on the IPMs of the two neighboring blocks, A and L, as shown in Fig. 4.3. Specifically, when these two modes are Planar or DC, the remaining candidates are registered as DC, angular No. 50 (i.e., Vertical mode), angular No. 18 (i.e., Horizontal mode), angular No. 46, and angular No. 54. Otherwise, A and L angular modes and also the three angular modes close to the direction of the A and L angular modes are registered. Moreover, the list already includes the target IPM candidate, it is not registered to reduce the signaling overhead for the duplicated IPM candidate. This is called the pruning process in VVC. After the list construction, the actual IPM can be uniquely identified by the index signaled by the encoder.

4.2.3 Inter Prediction in VVC

In VVC, the inter prediction is organized by an Adaptive Motion Vector Prediction (AMVP) mode and a merge mode, the same as HEVC [46]. AMVP mode derives the MV by adding the MV prediction from the decoded region and the MV difference signaled by the encoder. On the other hand, merge mode derives directly the MV from the merge candidate list, including a maximum of six sets of merge candidates (i.e., six sets of MV_0 and MV_1). These merge candidates are registered from the spatially and/or temporally neighboring inter prediction blocks. Because of these different MV derivation processes, AMVP and merge modes effectively predict the picture with non-uniform and uniform motions, respectively.

The maximum number of MVs applicable to a block is two, the same as HEVC. The minimum block sizes are defined as 4×8 or 8×4 for the inter prediction with only one MV ("uni-prediction", hereinafter), and 8×8 samples for the bi-prediction, respectively, due

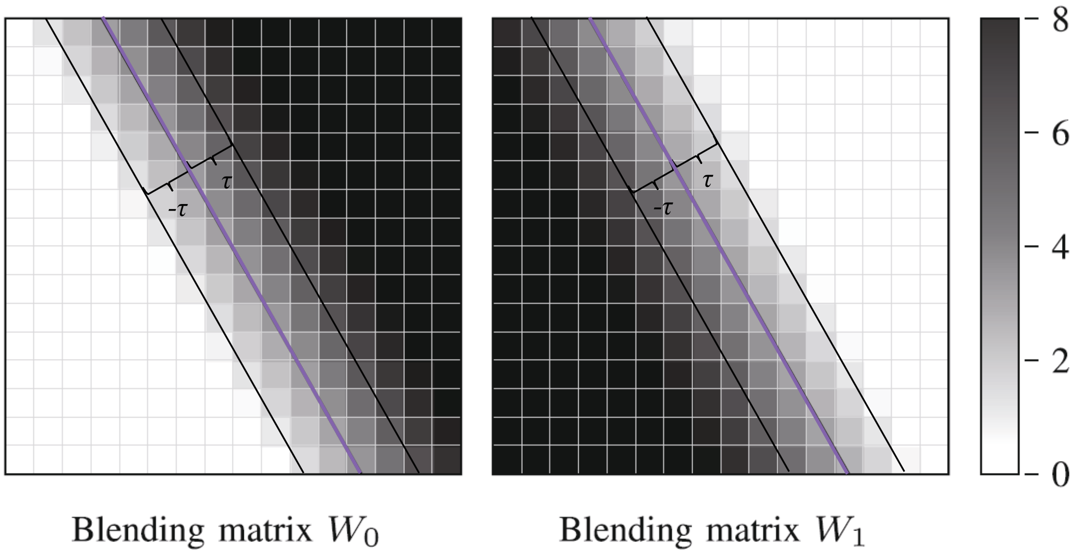


Figure 4.5: An example of the GPM blending matrix for each GPM-separated region based on Fig. 5 of Gao et al., 2021 [4]. The purple lines indicate the GPM block boundary.

to the worst-case memory bandwidth requirement.

Various new technologies have been developed to improve the prediction accuracy of the AMVP and merge modes in VVC [46]. In this chapter, we focus on the GPM among them since it has the potential for enhanced compression beyond VVC.

4.2.4 Geometric Partitioning Mode in VVC

As described in Sec. 4.1, P_{G} is derived by blending P_0 and P_1 with the integer blending matrices, W_0 and W_1, which contain weights in the value range of [0, 8]. The blending process is expressed by the following formula,

$$P_{\mathrm{G}} = (W_0 \circ P_0 + W_1 \circ P_1 + 4) \gg 3 \tag{4.1}$$

with

$$W_0 + W_1 = 8\ J_{w,h}, \tag{4.2}$$

where "$\circ$" in Eq. (2) represents the Hadamard product and $J_{w,h}$ is an all-ones matrix with the coding block size, $w \times h$. The weights of W_0 and W_1 depend on the displacement between the sample to be prediction and the GPM block boundary, $d(x_{\mathrm{c}}, y_{\mathrm{c}})$, where x_{c} and y_{c} denote the individual sample position within a coding block. In fact, the one of the W_0 and W_1 is given by a ramp function $\gamma_{x_c,\, y_c}$ as,

$$\gamma_{x_c,y_c} = \begin{cases} 0 & d(x_{\mathrm{c}}, y_{\mathrm{c}}) \leq -\tau, \\ \frac{8}{2\tau}(d(x_{\mathrm{c}}, y_{\mathrm{c}}) + \tau) & -\tau < d(x_{\mathrm{c}}, y_{\mathrm{c}}) < \tau, \\ 8 & d(x_{\mathrm{c}}, y_{\mathrm{c}}) \geq \tau, \end{cases} \tag{4.3}$$

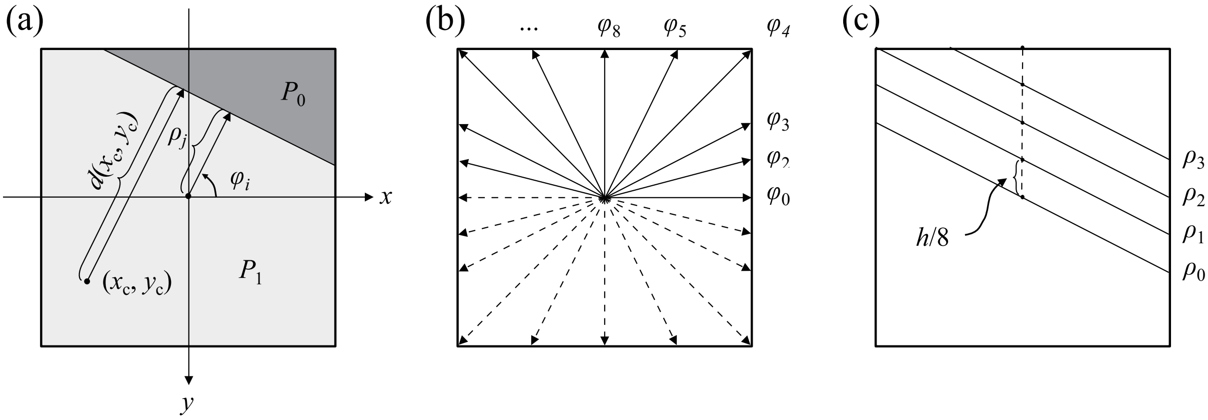


Figure 4.6: (a) Example of a Hessian normal form-based GPM block boundary; (b) quantized angle parameters φ; (c) quantized distance parameters ρ. (Copyright(C)2022 IEICE, [3] Fig. 6)

and the other blending matrix is derived from Eq. (2). Here, τ indicates the width of the soft blending area, which contains non-maximum or non-minimum weights within the matrices, and two samples are selected for τ of GPM-Inter/Inter in VVC. Examples for W_0 and W_1 are shown in Fig. 4.5. For the chroma components, the same matrices as that for luma are utilized, and the matrices are downsampled when the color format is 4:2:0, which has chroma planes with half the width and height of the luma plane.

Here, the shape of the GPM block boundary required for $d(x_c, y_c)$ is defined by the Hessian normal form with the combination of the angle φ and Euclidean distance ρ between the GPM block boundary and the center position of the coding block as shown in Fig. 4.6(a). The angle φ is quantized into 20 discrete angles φ_i shown in Fig. 4.6(b), with the range of $[0, 2\pi)$ symmetrically divided. The φ_i is designed with fixed tan(φ_i) values corresponding to the aspect ratio of the coding block, i.e., $\{0, \pm 1/4, \pm 1/2, \pm 1, \pm 2, \pm\infty\}$ and can be uniquely identified by the angleIdx signaled from the encoder. Similarly, the ρ is quantized into 4 discrete distance ρ_j shown in Fig. 4.6(c). Starting from ρ_0, which passes through the center of the block, the distance between ρ_0 and ρ_3 is calculated at equal intervals based on the height h, width w and angle φ of the block. In VVC, $d(x_c, y_c)$ is rounded into the integer precision sample position to avoid the additional interpolation in the GPM prediction process when calculating W_0 and W_1. The shapes of the GPM block boundary are restricted to a total of 64 combinations of 20 φ_i and 4 ρ_j, minus 16 redundant offsets. Here, the 16 redundant offsets are the 10 angles that overlap due to the 180-degree rotation and the 6 overlapping split lines due to the BT and TT splitting. An index to specify the shape of the GPM block boundary among the 64 candidates, $gpm\_partition\_idx$, is defined in VVC [7].

The MV_0 and MV_1 for the two GPM-separated regions are derived by GPM specific merge indices, gpm_merge_idx0 and gpm_merge_idx1, and the same merge candidate list as the regular merge mode [7]. gpm_merge_idx0 and gpm_merge_idx1 indicate the different merge candidates within the list. Hence, assuming the maximum merge candidates for GPM are six, the variation of gpm_merge_idx0 and gpm_merge_idx1 becomes 30 (= $gpm\_merge\_idx0 \times$

gpm_merge_idx1 = 6 $\times$ (6 - 1)) at most. It means that the encoder needs to select the best combination of the GPM block boundary shape and the MVs among the 1920 (= 64 $\times$ 6 $\times$ (6-1)) candidates.

To reduce the encoder complexity, the encoder of the VVC reference software, VTM version 11 [52], has an early termination method based on full Rate-Distortion (RD) Optimization (RDO) [16] such as the following three-step cost comparisons; In the first step, the Sum of Absolute Difference (SAD) costs for all 1920 candidates are compared and 60 candidates with smaller SAD costs are selected. In the second step, the Sum of Absolute Transformed Difference (SATD) costs for the 60 candidates are compared, and 8 with smaller SATD costs are selected. In the last step, the RD costs for the 8 candidates are compared, and the best combination with the smallest RD cost is determined.

Furthermore, to refrain from the decoder complexity, the range of GPM applicable block sizes is restricted from 8 $\times$ 8 to 64 $\times$ 64 (i.e., $8 \leq h \leq 64$ and $8 \leq w \leq 64$) in VVC. In addition, GPM is prohibited for blocks with an aspect ratio of 1 : 8 / 8 : 1 among the applicable block size range (i.e., 8 $\times$ 64 / 8 $\times$ 64) because it was confirmed that the improvements by GPM for these block sizes are relatively small in the VVC standardization process [53].

With the feature and algorithm, GPM-Inter/Inter can generate highly accurate prediction samples around the block boundaries between foregrounds and backgrounds with different motions. Especially, GPM contributes to increasing the coding performance of the coding structure without significant increments of the encoder complexity for the LB configuration as described in Sec. 4.1

However, GPM has room to improve the prediction accuracy further if the final prediction samples can be generated using intra prediction as well. Moreover, GPM-Inter/Inter cannot be utilized for the LP configuration where bi-prediction is prohibited.

4.3 Proposed Method

4.3.1 GPM with Inter and Intra Prediction

We propose a GPM with inter and intra prediction methods to improve the coding performance of GPM. Furthermore, we propose the following restriction of IPMs to maximize the coding performance of the proposed method. In this chapter, the detail of the proposed methods [54, 55] is described as follows.

The increment of the IPM number improves the intra prediction accuracy while increasing the signaling overhead and encoding time to determine the best IPM candidate. Hence, we propose to restrict the number of applicable IPM candidates to four at most: parallel angular mode to the GPM boundary as shown in Fig. 4.7(a) ("Parallel mode" hereinafter), perpendicular angular mode to the GPM boundary as shown in Fig. 4.7(b) ("Perpendicular mode" hereinafter), Planar mode, and IPM of the neighboring blocks ("Neighbor mode" hereinafter). The decoder can derive the actual IPM from the IPM candidate list similar to the merge candidate list and

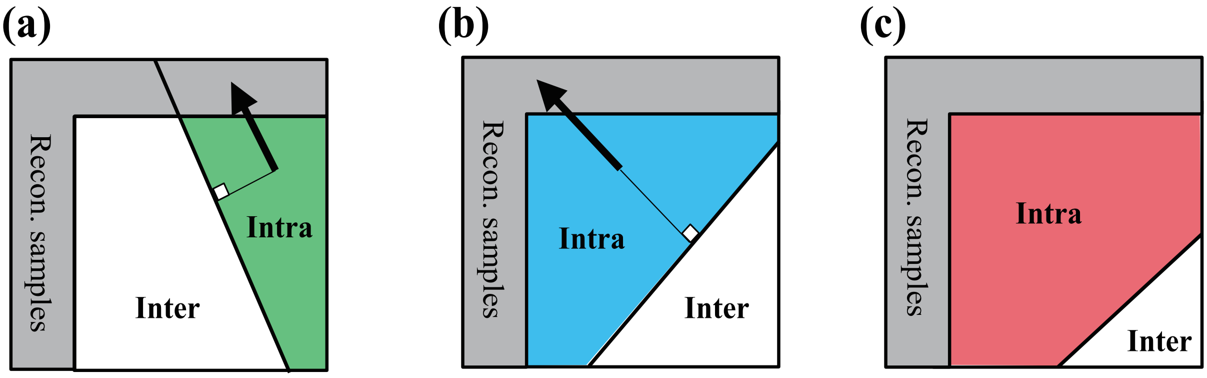


Figure 4.7: Examples of the GPM-Inter/Intra block applied by the proposed IPM candidates. (a) Parallel mode, (b) Perpendicular mode, and (c) Planar mode. Gray-shaded regions indicate the reconstructed sample areas. (Copyright(C)2022 IEICE, [3] Fig. 7)

Table 4.1: The restriction method of available neighboring blocks to derive the Neighbor mode for each GPM-separated region. GPM angleIdx corresponds to the angleIdx of GPM in VVC. A, L, and AL indicate the positions of the applicable neighbor blocks; A includes AL, A, and AR of Fig. 4.8; L includes AL, L, and BL of Fig. 4.8; L+A includes all the positions of Fig. 4.8.

GPM angleIdx	0	2	3	4	5	8	11	12	13	14
P_0	A	A	A	A	L+A	L+A	L+A	L+A	A	A
P_1	L+A	L+A	L+A	L	L	L	L	L+A	L+A	L+A
GPM angleIdx	16	18	19	20	21	24	27	28	29	30
P_0	A	A	A	A	L+A	L+A	L+A	L+A	A	A
P_1	L+A	L+A	L+A	L	L	L	L	L+A	L+A	L+A

the index for the IPM of each GPM-separated region signaled from the encoder.

The IPM number shown in Fig. 4.3 corresponding to the Parallel and Perpendicular modes can be identified with the index for specifying the shape of the GPM block boundary. This is because all the angles of the GPM block boundary are covered by the angles of the angular modes in VVC. When the angle of the GPM block boundary is above right or bottom left in the 45-degree direction, Parallel mode has two IPM candidates, No. 2 or No. 66, but No. 66 is always utilized in this proposed method.

For Neighbor mode, a maximum of two candidates can be derived from up to five neighboring blocks, as shown in Fig. 4.8. A new Neighbor mode is registered in the IPM candidate list only when the neighboring block is intra predicted, and the IPM is not yet registered. That is, a pruning process similar to that of the merge candidate list is also introduced in the IPM candidate list to avoid IPM duplication. To register more effective Neighbor modes, the maximum number of neighbor blocks that can be referenced is increased from 2 in HEVC

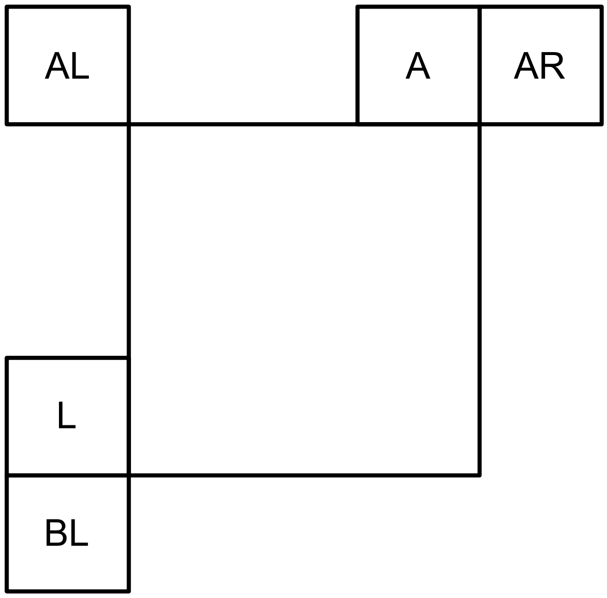


Figure 4.8: Available neighboring blocks for the Neighbor mode in the proposed method. AL, A, AR, L, and BL indicate the positions of the neighboring block: above left, above, above right, left, and bottom left, respectively. (Copyright(C)2022 IEICE, [3] Fig. 8)

to 5 in Fig. 4.8 while restricting the available neighboring blocks according to the shape of the GPM block boundary. The restriction method is represented as Tab. 4.1.

The registration order of IPM candidates is designed such that IPM candidates expected to be more effective against GPM are registered earlier. First, the Parallel mode, registered first in the proposed method, can be estimated to be effective when the GPM block boundary is extended over the reconstructed samples as shown in Fig. 4.7(a). Then, the Neighbor mode is registered next to the Parallel mode and is expected to be as effective as the Parallel mode since the IPM candidate of the neighboring blocks considered for the GPM block boundary can be applied. Next, the Perpendicular mode registered next to the Neighbor mode is considered to be effective when the total distances between the samples in the GPM-separated regions and the reconstructed samples are closer than that of the Parallel mode as shown in Fig. 4.7(b). After that, the Planar mode registered next to the Perpendicular mode is assumed to be effective when the GPM-separated region size is large, as shown in Fig. 4.7(c). Finally, the DC mode, which is usually applied preferentially in the regular intra prediction as well as Planar mode, can be registered as the final candidate in the case that the list is not filled due to the absence of the Neighbor mode that can be caused by the available check or pruning process for Neighbor mode.

In the proposed method, the same blending matrices as GPM-Inter/Inter are utilized for generating the final inter prediction samples (i.e., both GPM-Inter/Intra and GPM-Intra/Intra). In other words, the derivation method for the shape of the GPM block boundary is the same as GPM-Inter/Inter.

4.3.2 Prohibition of GPM-Intra/Intra

We also propose a prohibition of GPM-Intra/Intra to maximize the coding performance of the proposed method. The prohibition of GPM-Intra/Intra can be realized by the flag, $gpm\_intra\_enabled\_flag$, signaled from the encoder for specifying whether inter or intra prediction is applied for each GPM-separated region. Specifically, the signaling is designed so that when the intra prediction is applied to one region, it cannot be applied to the other region.

The prohibition of GPM-Intra/Intra improves the coding performance can be explained as follows. First, GPM is easily applied to the boundary between foreground and background with different motions as described in Sec. 4.1. This means that inter prediction is sufficient to predict the background areas with no motion or the foreground areas, including large flat regions. Second, the intra prediction accuracy for the bottom right region within the prediction block is lower than that of the above-left region. This is because the distance between the sample to be prediction and the reconstructed sample is large. With these features of the GPM and intra prediction, the application rate of GPM-Intra/Intra can be estimated to be lower than that of the other GPM. In other words, the prohibition of GPM-Intra/Intra can be expected to have no impact other than saving the signaling overhead, and the GPM-Inter/Intra will be effective even for the LP configuration.

The prohibition of GPM-Intra/Intra will not reduce the encoder complexity but will reduce the decoder complexity as follows. Regarding the encoder, this is because all SAD costs to apply each IPM candidate to the two GPM-separated regions in all GPM shapes need to be calculated for GPM-Inter/Intra at the first early termination stage as described in 4.2.4, even without prohibiting GPM-Intra/Intra. On the other hand, as for the decoder, the prohibition can avoid increasing the circuit size for the intra prediction, which is a critical problem for hardware-based video decoders.

4.3.3 The Other Specification of Signaling

In the proposed method, GPM-Inter/Intra is applied only for inter slices. This means the dual-tree in VVC does not need to be considered. Therefore, the direct mode is always utilized for deriving the chroma component IPMs to save the signaling overhead further.

The block size ranges to which the proposed method can be applied are not changed from those of GPM-Inter/Inter as described in Sec. 4.2.4. The reasons are as follows. For small-size blocks, applying intra prediction reduces the worst-case memory bandwidth requirement, making GPM applicable. On the other hand, GPM for small-size blocks increases the overhead more than improves the prediction accuracy. For larger size blocks, the intra prediction accuracy becomes lower, as described in Sec. 4.3.2. Furthermore, the calculation and storage of the blending mask for large-size blocks become a burden, especially for the decoder.

In addition to the original signals in GPM-Inter/Inter, two additional block-level signals, $gpm\_intra\_enabled\_flag$ as described in Sec. 4.3.2 and gpm_intra_idx, are introduced in the proposed method. gpm_intra_idx is signaled to identify the IPM candidate for GPM only when the number of IPM candidates is larger than one. $gpm\_intra\_enabled\_flag$ is coded with fixed-length code, whereas with a zero-order exponential Golomb, the same as that for signaling merge candidates.

Table 4.2: Details of the proposed methods categorized by IPM candidate list size, registrable IPM candidates and their registering order, and prohibition of GPM-Intra/Intra. The arrow in the third column indicates the list's registering order of registrable IPM candidates. The bracketed Planar and DC modes in Prop. 5 and Prop. 6 are registrable when Neighbor modes are not registered. "On" and "Off" within the fourth column denote the existence and absence of the prohibition of GPM-Intra/Intra in the proposed method.

Method	IPM candidate list size	Registrable IPM candidates and their registering order	Prohibition of GPM-Intra/Intra
Prop. 1	1	Parallel	On (Default)
Prop. 2 (Phbt.Off)	2	Parallel → Planar	Off
Prop. 2	2	Parallel → Planar	On
Prop. 3 (Phbt.Off)	2	Parallel → Perpendicular	Off
Prop. 3	2	Parallel → Perpendicular	On
Prop. 4 (Phbt.Off)	3	Parallel → Perpendicular → Planar	Off
Prop. 4	3	Parallel → Perpendicular → Planar	On
Prop. 5	3	Parallel → Neighbor → Perpendicular (→ Planar)	On
Prop. 6	4	Parallel → Neighbor → Perpendicular → Planar (→ DC)	On

Table 4.3: Details of the VTM CTC test sequences from classes A to F categorized by resolutions, frame rates, and video content type, i.e., Camera-captured Content (CC), Screen Content (SC), and Mixed Content with CC and SC (MC).

Class	Resolutions [pixels $\times$ lines]	Frame rates	Video content
A1	3840×2160	30–60	CC
A2	3840×2160	50–60	CC
B	1920×1080	50–60	CC
C	832×480	30–60	CC
D	416×240	30–60	CC
F	$832 \times 480 \sim 1920 \times 1080$	20–60	SC and MC

Table 4.4: Overall performance of each method, including the existing GPM-Inter/Inter compared to the anchor in RA, LB, and LP configurations, evaluated by BDY, BDY, BDV, EncT, and DecT.

Method	RA [%]					LB [%]					LP [%]				
	BDY	BDU	BDV	EncT	DecT	BDY	BDU	BDV	EncT	DecT	BDY	BDU	BDV	EncT	DecT
GPM-Inter/Inter	-0.74	-1.08	-1.18	103	99	-1.54	-1.86	-1.80	105	98	0.00	0.00	0.00	100	100
Prop. 1	-0.88	-1.29	-1.38	104	100	-1.81	-2.58	-2.55	106	99	-0.78	-1.34	-1.24	107	101
Prop. 2 (Phbt.Off)	-0.89	-1.48	-1.47	104	100	-1.89	-2.75	-2.74	106	99	-0.87	-1.64	-1.54	109	102
Prop. 2	-0.94	-1.42	-1.49	104	100	-1.94	-2.86	-2.88	106	99	-0.95	-1.63	-1.54	108	101
Prop. 3 (Phbt.Off)	-0.89	-1.52	-1.49	104	100	-1.90	-2.70	-2.82	107	99	-0.91	-1.84	-1.76	108	101
Prop. 3	-0.93	-1.47	-1.48	104	100	-1.91	-2.62	-2.72	106	99	-0.95	-1.84	-1.59	108	101
Prop. 4 (Phbt.Off)	-0.91	-1.58	-1.54	105	100	-1.97	-2.84	-2.77	107	99	-1.01	-1.84	-1.81	109	102
Prop. 4	-0.96	-1.56	-1.55	105	100	-1.96	-2.94	-2.76	107	99	-1.09	-1.92	-1.71	110	102
Prop. 5	-0.97	-1.58	-1.62	107	100	-1.99	-3.16	-2.81	108	99	-1.15	-2.10	-1.96	111	102
Prop. 6	-0.96	-1.66	-1.67	107	100	-2.01	-3.10	-2.70	110	99	-1.12	-2.19	-1.91	113	101

Table 4.5: Sequence-level performance of GPM-Inter/Inter and Prop. 5 compared to the anchor (VTM-11 without GPM) in RA, LB, and LP configurations, which is evaluated by BDY, EncT, and DecT. Note that the performance of GPM-Inter/Inter in LP is not written since GPM-Inter/Inter is disabled in the LP configuration.

Sequense	GPM-Inter/Inter						Prop. 5								
	RA [%]			LB [%]			RA [%]			LB [%]			LP [%]		
	BDY	EncT	DecT	BDY	EncT	DecT	BDY	EncT	DecT	BDY	EncT	DecT	BDY	EncT	DecT
Tango2	-0.64	103	99	–	–	–	-0.75	108	100	–	–	–	–	–	–
FoodMarket4	-0.40	103	100	–	–	–	-0.54	108	101	–	–	–	–	–	–
Campfire	-0.20	103	99	–	–	–	-0.40	110	99	–	–	–	–	–	–
Average Class A1	**-0.41**	**103**	**99**	–	–	–	**-0.56**	**107**	**100**	–	–	–	–	–	–
CatRoad	-0.61	102	100	–	–	–	-0.80	107	99	–	–	–	–	–	–
DaylightRoad2	-0.46	102	100	–	–	–	-0.44	107	100	–	–	–	–	–	–
ParkRunning3	-0.56	104	100	–	–	–	-0.69	109	100	–	–	–	–	–	–
Average Class A2	**-0.55**	**103**	**100**	–	–	–	**-0.64**	**106**	**100**	–	–	–	–	–	–
MarketPlace	-0.44	103	100	-0.96	104	99	-0.60	108	100	-1.19	107	100	-0.67	110	101
RitualDance	-0.57	103	100	-1.04	105	98	-0.83	108	100	-1.87	110	99	-1.36	113	102
Cactus	-0.75	103	99	-1.40	105	95	-1.12	107	100	-2.00	108	96	-1.14	111	102
BasketballDrive	-0.32	103	100	-0.80	105	99	-0.49	108	101	-1.31	110	99	-0.93	112	101
BQTerrace	-0.38	103	100	-0.43	104	100	-0.41	106	101	-0.54	107	100	-0.49	109	102
Average Class B	**-0.49**	**103**	**100**	**-0.93**	**105**	**98**	**-0.69**	**106**	**100**	**-1.38**	**108**	**99**	**-0.92**	**111**	**102**
BasketballDrill	-1.20	103	98	-2.23	106	98	-1.85	108	101	-2.94	110	99	-1.92	115	102
BQMall	-2.30	102	98	-3.06	106	96	-2.61	107	101	-3.81	110	97	-2.28	111	101
PartyScene	-0.67	104	100	-1.08	107	98	-1.01	108	101	-1.79	111	100	-1.14	111	101
RaceHorses	-1.52	104	98	-1.83	107	98	-2.01	108	100	-2.47	111	99	-1.79	112	100
Average Class C	**-1.42**	**103**	**99**	**-2.05**	**106**	**98**	**-1.87**	**107**	**99**	**-2.75**	**111**	**99**	**-1.78**	**112**	**101**
BasketballPass	-0.79	103	99	-1.61	107	98	-1.20	108	101	-2.05	111	99	-1.29	110	102
BQSquare	-0.08	103	100	-0.96	105	99	-0.19	106	103	-0.87	107	101	-0.07	108	104
BlowingBubbles	-0.68	104	99	-1.78	107	97	-0.84	109	102	-2.01	111	98	-0.75	111	103
RaceHorses	-1.29	104	99	-2.17	106	96	-1.67	108	100	-3.03	110	98	-1.76	109	101
Average Class D	**-0.71**	**104**	**100**	**-1.63**	**106**	**98**	**-0.98**	**107**	**100**	**-1.99**	**110**	**99**	**-0.97**	**109**	**102**
FourPeople	–	–	–	-1.73	103	97	–	–	–	-2.05	107	99	-1.07	110	103
Johnny	–	–	–	-2.05	102	98	–	–	–	-2.01	104	100	-0.39	109	103
KristenAndSara	–	–	–	-1.92	102	97	–	–	–	-1.86	105	99	-0.66	108	102
Average Class E	–	–	–	**-1.90**	**102**	**97**	–	–	–	**-1.98**	**105**	**99**	**-0.71**	**109**	**102**
BasketballDrillText	-1.21	102	99	-1.88	104	98	-1.82	106	101	-2.66	108	98	-1.76	110	102
ArenaOfValor	-0.90	101	99	-1.56	103	97	-1.14	104	100	-2.13	107	98	-1.08	109	101
SlideEditing	-0.02	100	100	0.21	101	101	-0.04	102	102	-0.14	104	101	0.30	107	102
SlideShow	-0.12	101	101	0.04	103	100	0.16	103	103	0.05	106	101	-0.13	108	102
Average Class F	**-0.56**	**102**	**100**	**-0.80**	**103**	**94**	**-0.71**	**103**	**101**	**-1.22**	**106**	**99**	**-0.67**	**108**	**102**
Overall	**-0.74**	**101**	**99**	**-1.54**	**103**	**99**	**-0.97**	**107**	**100**	**-1.99**	**108**	**99**	**-1.15**	**111**	**102**

4.4 Experimental Results and Discussion

4.4.1 Test Conditions

1) Software Settings: The VTM-11 [52] was used for the simulation software, and the proposed method was implemented in the VTM-11. To evaluate the effect of GPM-Inter/Intra compared to that by GPM-Inter/Inter, the coding performance and the complexity of VTM-11, disabling GPM-Inter/Inter, were evaluated as the baseline (i.e., anchor). In addition, a total of nine different proposed methods, as shown in Tab. 4.2, were conducted to verify the effects of IPM candidate list sizes, IPM candidate variations, and the prohibition of GPM-Intra/Intra.

2) Encoder Configurations: The coding conditions were followed with the VTM CTC [45]. RA, LB, and LP configurations defined in the VTM CTC were used because GPM-Inter/Inter and the proposed methods can be applied only to B and P slices. RA is often used for general video transmission, while LB and LP are utilized for low-latency video transmission. Only the performance and complexity of the proposed methods in the LP configuration were evaluated since GPM-Inter/Inter is disabled by default in the LP configuration.

3) Test Sequences: The test sequences from classes A to F, as listed in VTM CTC, were used. They are categorized with different resolutions, frame rates, and video content, as shown in Tab.4.3. For each test sequence, four Quantization Parameter (QP) values, 22, 27, 32, and 37, defined in the VTM CTC, were used to generate the different rate points.

4) Evaluation Metrics: The coding performance was evaluated by the Bjøntegaard Delta bitrate (BD-rate) of the luma and two chroma components, i.e., BD-rate Y, U, and V components (BDY, BDU, and BDV) [27, 28]. The BD-rate is the evaluation index used to quantify the difference of the generated bitrate for the identical level of Peak Signal-to-Noise Ratio (PSNR) between the two coding methods. The negative and positive values of BD-rate indicate the coding performance gains and losses. The complexity was evaluated by the two coding methods' relative encoding and decoding runtime (EncT and DecT) measured on a homogeneous cluster PC. In accordance with the VTM CTC, Classes D and F results are not included in the overall results.

4.4.2 Comparison of Overall Results

The overall results of each method compared to the anchor in RA, LB, and LP configurations, evaluated by BDY, BDU, BDV, EncT, and DecT, are described in Tab. 4.4.

First, regarding the coding performance, the GPM-Inter/Inter and all the proposed methods provide the coding performance gain against the anchor in RA, LB, and LP as observed in Tab. 4.4. The values of LB is the largest among the three coding conditions. The proposed methods give additional coding performance gains compared to GPM-Inter/Inter. From the comparison of GPM-Inter/Inter and Prop. 5 or Prop. 6, the coding performance gains of RA and LB become 1.3 times that of the GPM-Inter/Inter at most, and the additional coding performance gain with a maximum of 1.15% BD-rate savings is newly produced in LP. They are derived from increased GPM applied samples, which will be clarified in Sec. 4.4.4.

Second, regarding the complexity, the GPM-Inter/Inter and all the proposed methods increase EncT compared to the anchor. Similarly, the proposed methods provide additional EncT increments against GPM-Inter/Inter. The increment of the proposed method in LP is significant since the entire functionality of GPM is introduced into the anchor in LP, whereas only the functionality regarding the intra prediction of the GPM-Inter/Intra is added in RA and LB. On the other hand, there is almost no DecT increment of GPM-Inter/Inter and all the proposed methods over the anchor. This is because the decoder can specify the GPM block boundary shape, MVs, and IPMs signaled by the encoder without burden.

Third, the coding performance and complexity of the proposed methods compared by different IPM candidates, corresponding to the results of Prop. 1–6 of Tab. 4.4, are described as follows.

Regarding the performance, more coding performance gains are achieved as the number of IPM candidates (i.e., IPM candidate list sizes) increases. The comparison of Prop. 1–3 proves that adding Planar or Perpendicular modes to Parallel mode yields additional coding performance gain. The further coding performance gain of Prop. 4 against Prop. 2 and 3 indicates an additive effect of Planar and Perpendicular modes. Compared with Prop. 4 and 5, more coding performance gain by introducing Neighbor mode can be confirmed. The same level of BD-rate between Prop. 5 and 6, which have different IPM candidate list sizes, implies the coding performance gain is saturated around three IPM candidates. This is because the PSNR improvements and bitrate increments by extending IPM candidate list sizes from 3 to 4 are counterbalanced. The above observation will be clarified with an RD curve and analysis of GPM-Inter/Intra prediction samples described in Sec. 4.4.4.

Regarding complexity, EncT similarly increases as the number of IPM candidates increases. This is because it raises the RDO processing of the encoder side to derive the minimum RD cost corresponding to the best combination of the GPM block boundary shape, MVs, and IPMs. In contrast, DecT is not increased since the decoder can identify the best combination by their indices signaled by the encoder.

Fourth, the coding performance and complexity of the proposed methods without and with the prohibition of GPM-Intra/Intra, i.e., Prop. 2–4 (Phbt.Off) vs. Prop. 2–4 in Tab. 4.4, are described as follows. Regarding the coding performance, the proposed method with the prohibition gives further coding performance gains compared to that without the prohibition, which suggests the bitrate reduction by the prohibition contributes to the gains. Regarding EncT, there is no difference between the proposed methods with and without the prohibition since only the addition processing of the two SAD costs for GPM-Intra/Intra (i.e., two different IPMs) is different, as described in Sec. 4.3.1. DecT is also not different because the reduction of the GPM-Intra/Intra process is minor due to the lower application rate of the GPM-Intra/Intra, which will be clarified in Sec. 4.4.4.

4.4.3 Comparison of Sequence-level Results

In this section, the coding performance and complexity of GPM-Inter/Inter and the proposed method are compared by sequence level with different resolutions and content. The sequence-level results of GPM-Inter/Inter and Prop. 5 to the anchor in RA, LB, and LP, evaluated by BDY, EncT, and DecT, are described in Tab. 4.5.

Regarding the coding performance, no tendency on the resolutions can be confirmed from the comparison of the results of classes A1/A2, B, C, and D. In contrast, the following tendency can be found with the content in both GPM-Inter/Inter and Prop. 5, which is common to the RA, LB, and LP. Specifically, the coding performance gains are relatively large for the sequence with differently moving foregrounds and backgrounds across classes, e.g., RitualDance, BQMall, and RaceHorses. GPM was originally easy to apply in these sequences, so Prop. 5 provides additional coding performance gains against GPM-Inter/Inter. On the other hand, the coding performance gains of the sequence without foregrounds and backgrounds, e.g., BQTerrace and BQSquare, are relatively small. GPM is seldom applied in these sequences so the bitrate increments for GPM-Inter/Intra raise the BD-rate of Prop. 5 compared to that of GPM-Inter/Inter. In the pure SC sequences, i.e., SlideEditing and SlideShow, very minor coding performance gains or coding performance losses are observed. This is because these sequences often include moving objects with sharp edges, and the graduated blending matrices described in Sec. 4.2.1 overly smooth the sharp edges and rather raise residuals.

In association with these characteristics, the EncT increments of the test sequences with minor coding performance gains or losses are smaller than those with larger coding performance gains since GPM is terminated early in the encoder's RDO processing. On the other hand, the DecT of the sequences with larger coding performance gains is smaller than those with minor coding performance gains or losses because GPM increases the number of large-size blocks and reduces the decoding processing on QTBTT partitioning compared to the anchor.

4.4.4 Picture-level Analysis

As a picture-level analysis, the RD curves of the anchor and all the proposed methods for BQMall and BQSquare in the LP configuration are shown in Fig. 4.9. To investigate the effect of the proposed method, we selected BQMall and BQSquare, which produce the largest and smallest coding performance gains among all the test sequences, respectively.

First, regarding BQMall, the comparison by the same QP value shows that coding performance gains of the proposed method mainly come from the bitrate reduction as shown in Fig. 4.9(b) and (c). The reason is that GPM-Inter/Intra improves the prediction accuracy, which reduces residuals around the boundaries of the foreground and background with different motions. Not only a bitrate reduction but also the PSNR improvement can be observed in the high QP range as shown in Fig. 4.9(c). This is because the effect of GPM on the qualities of the reconstructed samples, which determines the size of PSNR, became more apparent in the high QP range. The quantization step of the high QP range is larger than that of the low

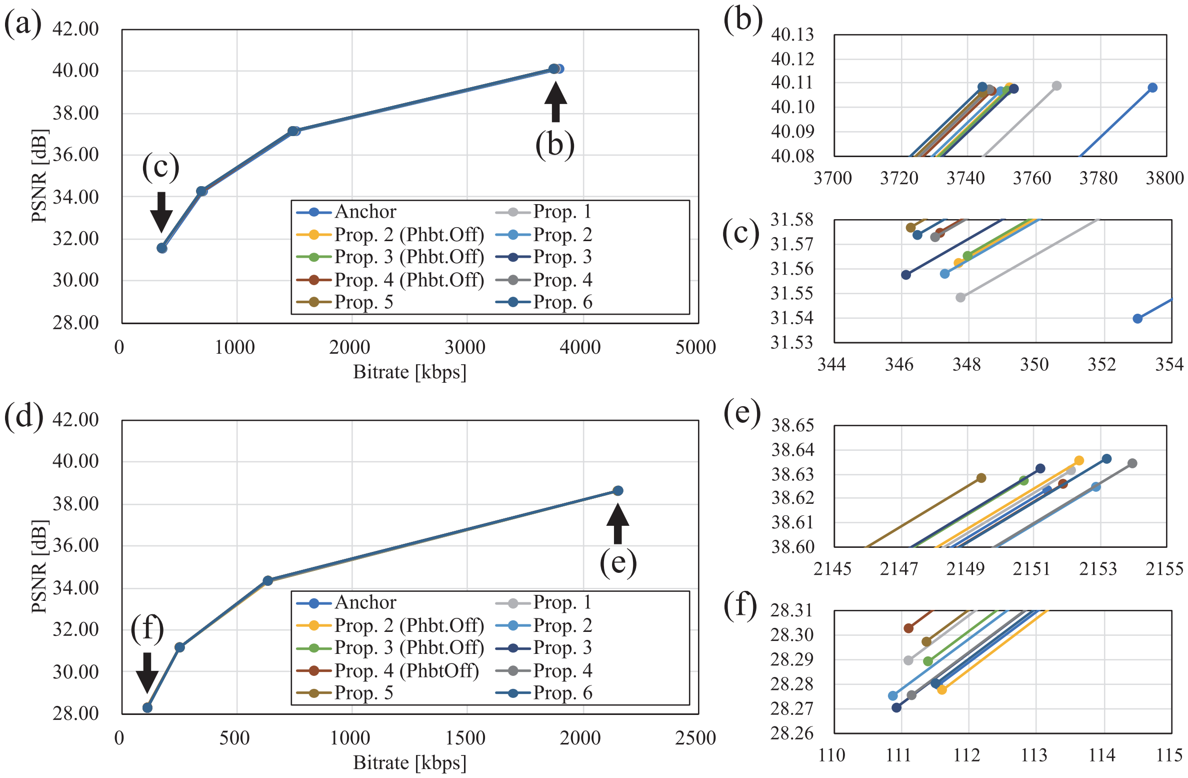


Figure 4.9: Rate distortion curves of the anchor and all the proposed methods for BQMall and BQSquare in the LP configuration. (a)–(c) BQMall, (d)–(f) BQSquare. (Copyright(C)2022 IEICE, [3] Fig. 9)

QP range so that most of the residual coefficients become zero. Therefore, the qualities of the reconstructed samples highly depend on those of the prediction samples. In this sense, the additional PSNR improvement can be observed as the number of IPMs increases in the high QP range. Moreover, the counterbalance between Prop. 5 and Prop. 6, and the additional bitrate reduction by the prohibition of GPM-Intra/Intra, i.e., Prop. 2–4 vs. Prop. 2–4 (Phbt.Off) can be confirmed, as described in Sec. 4.4.3.

On the other hand, regarding BQSquare, the bitrate reductions are very minor in both high and low QP ranges, while only small PSNR improvements are seen in Prop .4 (Phbt.Off) and Prop. 5 in the high QP range as shown in Figs. 4.9(e) and (f). The small coding performance gain of Prop. 5 in Tab. 4.5 comes from the small PSNR improvements. The same tendencies as BQMall and BQSquare can be observed in the other test sequences with larger and smaller coding performance gains, respectively.

To clarify the evidence of the difference in coding performance gains among RA, LB, and LP and the effect of the prohibition of GPM-Intra/Intra, an analysis of the total GPM applied samples organized by GPM-Inter/Inter, GPM-Inter/Intra, and GPM-Intra/Intra with Prop.4 (Phbt.Off) for BQMall and BQSquare is shown in Fig. 4.10. They are normalized by the total samples of the encoded test sequences. The difference in coding performance gains among RA, LB, and LP corresponds to the difference in their GPM applied samples.

Fig.4.10(a) shows that the total GPM applied samples in the RA and LB configurations become around 1.3 times that of only the GPM-Inter/Inter samples, which also matches the additional coding performance gain of Prop. 4 (Phbt.Off) as observed in BQMall. In contrast, Fig. 4.10(b) shows the total GPM applied samples in the RA and LB configurations have not increased much compared with only the GPM-Inter/Inter samples, which matches the coding performance loss of Prop. 4 (Phbt.Off) against the GPM-Inter/Inter as observed in BQSquare. In these two sequences, the total GPM-applied samples of RA and LP are at the same level, corresponding to the coding performance gains of the Prop. 4 (Phbt.Off) in the RA and LP configurations. The GPM-Intra/Intra samples are significantly smaller than the GPM-Inter/Inter or GPM-Inter/Intra samples. It means that the prohibition of the GPM-Intra/Intra does not affect the PSNR improvement and contributes to the bitrate reduction.

An analysis of the GPM-Inter/Intra prediction samples of Prop. 6 in each GPM applicable block size for two different test sequences and QPs in the LB configuration is shown in Fig. 4.11. A certain number of the GPM-Inter/Intra prediction samples in Figs. 4.11(a) and (b) is observed while not in Figs. 4.11(c) and (d), which corresponds to the coding performance gain of BQMall or the coding performance losses of BQSquare of the proposed method against the GPM-Inter/Inter in the LB configuration. The ratios of the GPM-Inter/Intra samples to all GPM applied samples (i.e., GPM-Inter/Intra and GPM-Inter/Inter prediction samples) in Figs. 4.11(a), (b), and (d) are decreased as the GPM applicable block size becomes larger. This is because the accuracy of the intra prediction degrades for the large block size where the prediction target samples are far from the reconstructed samples. Compared to the GPM-Inter/Intra prediction samples in Figs. 4.11(a) and (b), the majority of the samples shift from the smaller size blocks to the larger size blocks because the number of larger size blocks increases in the high QP range.

In addition, the ratio of GPM-Inter/Intra of each block size in the high QP range is larger than that in the small QP range from the comparison of Figs. 4.11(a) and (b). The reasons for this tendency are as follows. First, the inter prediction accuracy is relatively higher for low QP than for high QP. This is because the coding noise on the reference frame required for inter prediction is smaller for low QP than for high QP. On the other hand, the intra prediction accuracy is not always higher for low QP than for high QP. This is because intra prediction is highly dependent on the similarity of the reference samples themselves, as well as the presence or absence of coding noise on the reference samples adjacent to the current block. These characteristics based on the QPs of inter and intra predictions also correspond to GPM. Hence, at high QP, the application rate of GPM-Inter/Inter decreases, while those of GPM-Inter/Intra increases, which causes the increment in the ratio of GPM-Inter/Intra applied samples in Fig. 4.11.

Regarding IPMs, the number of GPM-Inter/Intra prediction samples applied by Parallel, Perpendicular, and Planar modes dominates. Especially, those of the Parallel mode are the largest, corresponding to our expectation described in Sec. 4.3.1. In contrast, those of Neighbor and DC mode are smaller. This is because the Neighbor mode is only registered in the IPM candidate list when the neighbor block is intra block, whereas the DC mode is registered in the IPM candidate list only when the list is not filled due to the absence of the Neighbor mode. In this sense, further coding performance gains for the Neighbor mode can be expected if the neighboring block is an inter prediction block but stores the IPM of the reference block, which has been studied in the exploration of JVET.

4.4.5 Subjective Evaluation

In addition to improving the coding performance, subjective quality improvements are provided by the proposed method. First, the significant improvements can be seen in GPM-Intra/Inter applied block areas of the LP configuration for BQMall from the comparison of Fig. 4.12, which are examples of the GPM applied block map of Prop. 5 and decoded image of the anchor and the Prop. 5. For example, the block noises around the man's shoulder in Fig. 4.12(b) are removed in Fig. 4.12(c). As another example, the blur of the woman's cap in Fig. 4.12(e) is sharpened in Fig. 4.12(d). These improvements correspond to the PSNR improvements in the high QP range, described in the RD curve of Sec. 4.4.4. In addition, Parallel, Perpendicular, and Planar modes were applied as expected in areas with the image characteristics described in Sec. 4.3.1. Note that the application of Neighbor mode was not observed in Fig. 4.12(a) due to its lower application rate, but it was confirmed in another frame.

Second, any of the subjective improvement and degradation by the proposed GPM cannot be seen in BQSquare as shown in Figs. 4.12(g) and (h) due to the lower application of the GPM-Inter/Intra blocks, which matches the objective evaluation results and analyses for BQSquare.

Third, an additional improvement was also confirmed in the LB configuration even where the conventional GPM (GPM-Inter/Inter) can be applied from the comparison of the block maps and decoded images of GPM-Inter/Inter and Prop. 5. An example is shown in Figs. 4.12(i)–(l). Specifically, the artifacts behind the boy and the blur near the boy's feet seen in Fig. 4.12(k) were removed in the GPM-Inter/Intra applied block area in Fig. 4.12(l). Other than this example, additional improvements have been identified, although not as much as in the LP configuration.

4.5 Conclusion

This chapter proposed a GPM with inter and intra prediction (GPM-Inter/Intra) to enhance coding performance beyond VVC. To maximize the coding performance of the proposed method, we propose restricting the number of intra prediction modes and prohibiting GPM with two different intra predictions. The experimental results show that the proposed method improved the coding performance gain by the existing GPM method in VVC, realized by two different inter prediction, by 1.3 times in RA and LB configuration. Additionally, the proposed method provided an additional coding performance gain with 1.15% bitrate savings in LP configuration. Moreover, a comparison of subjective qualities between the existing GPM and the proposed GPM revealed that the proposed GPM suppresses artifacts observed in the existing GPM.

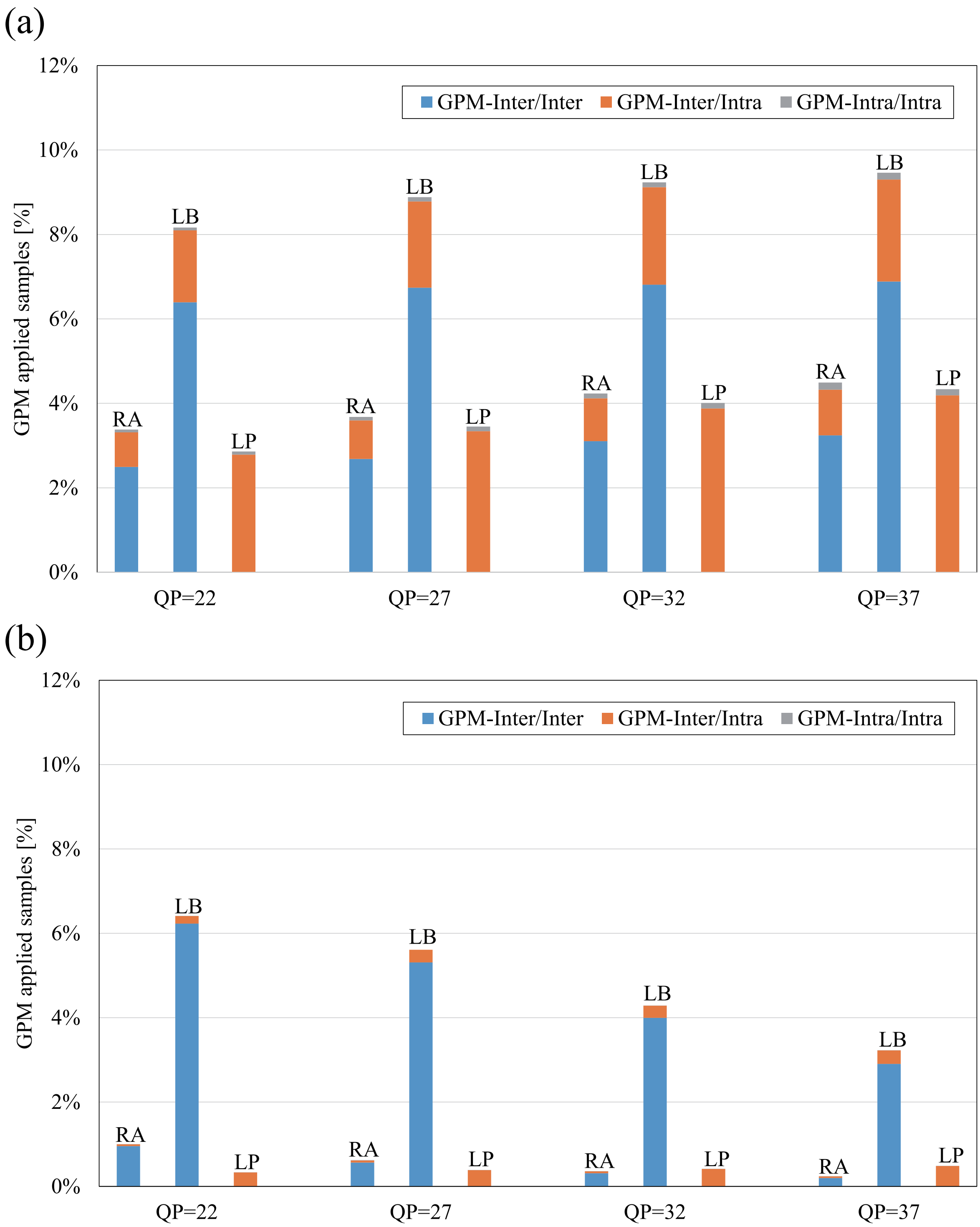


Figure 4.10: Analysis of the total GPM applied samples organized by GPM-Inter/Inter, GPM-Inter/Intra, and GPM-Intra/Intra with Prop. 4 (Phbt.Off) in RA, LB, and LP configurations. (a) BQMall, (b) BQSquare. (Copyright(C)2022 IEICE, [3] Fig. 10)

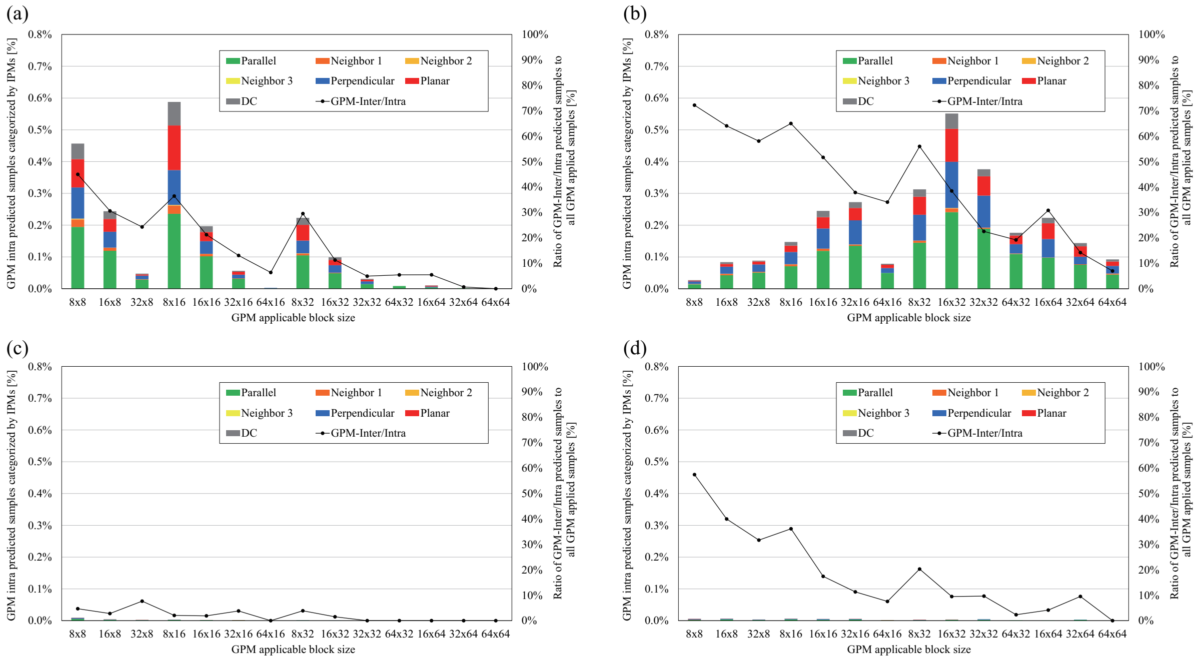


Figure 4.11: An analysis of the GPM-Inter/Intra prediction samples by Prop. 6 in each GPM applicable block size for two different test sequences and QPs in the LB configuration; (a) BQMall, QP = 22, (b) BQMall, QP = 37, (c) BQSquare, QP = 22, and (d) BQSquare, QP = 37. The left axis indicates GPM-Inter/Intra prediction samples categorized by IPMs. The right axis denotes the ratio of GPM-Inter/Intra prediction samples to all GPM applied samples (i.e., GPM-Inter/Intra and GPM-Inter/Inter prediction samples). Both values are normalized by the total encoded samples of each test sequence. (Copyright(C)2022 IEICE, [3] Fig. 11)

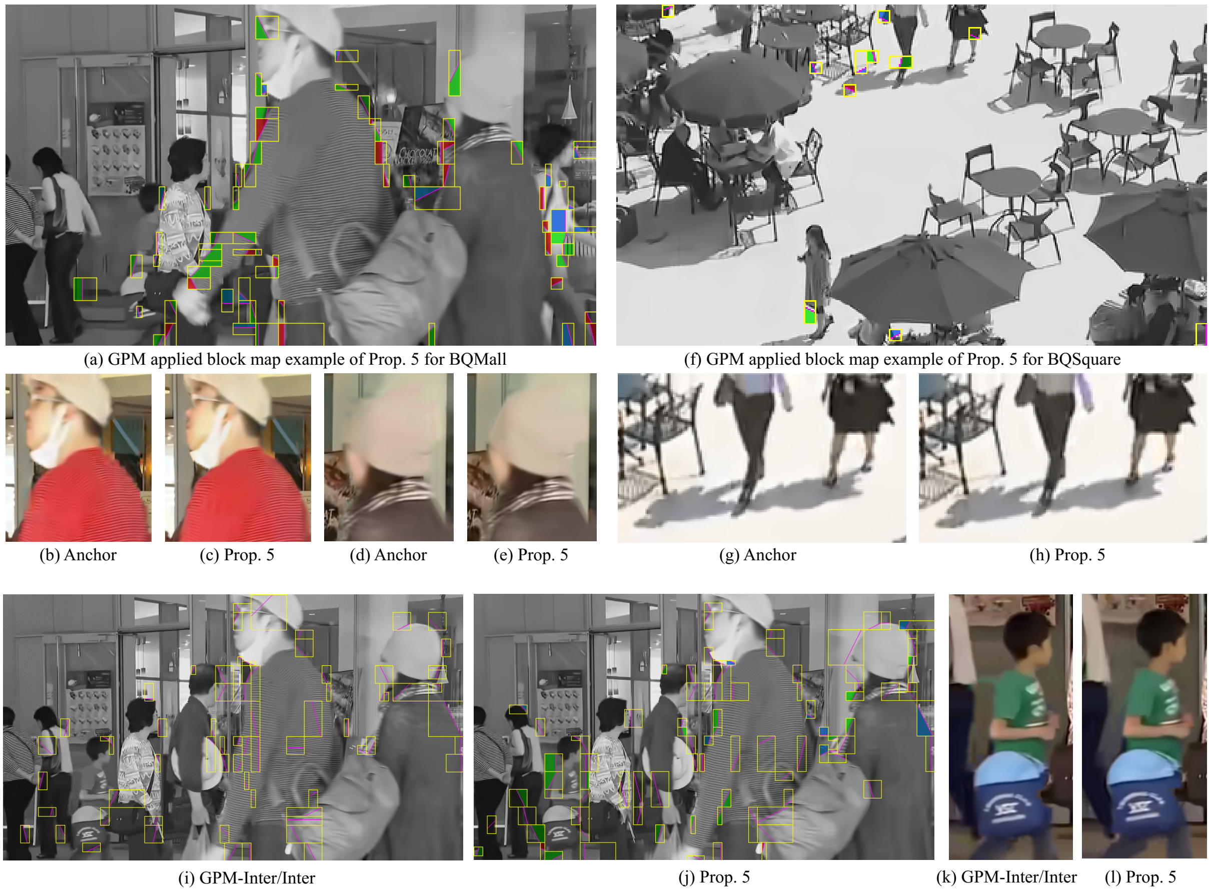


Figure 4.12: Subjective evaluation results of Anchor, GPM-Inter/Inter, Prop. 5 for BQMall and BQSquare in the LP and LB condition at QP = 37. (a), (f), (i), and (j) are examples of GPM-applied block maps. (a) Prop. 5 for the 77th frame of BQMall in LB; (f) Prop. 5 for the 32nd frame of BQSquare in LP; (i) and (j) GPM-Inter/Inter and Prop. 5 for the 68th frame of BQMall in LB. In these block maps, the area surrounded by the yellow grid is the GPM-applied area, while the other areas are the non-GPM areas. A purple line indicates the GPM block boundary. Gray, green, blue, and red areas within the yellow grid denote the areas where inter prediction, Parallel mode, Perpendicular mode, and Planar mode are applied, respectively. (b), (c), (d), (e), (k), and (l) are decoded images corresponding to the block maps for each method and each test condition. (Copyright(C)2022 IEICE, [3] Fig. 12)

Chapter 5

Bi-predictive Intra Block Copy

5.1 Introduction

5.1.1 Background

[1] Intra Block Copy (IBC) is a promising intra-coding tool for additional coding performance gains beyond VVC for Camera-captured Content (CC), while IBC is a specific coding tool for Screen Content (SC) in H.266 | Versatile Video Coding (VVC) and H.265 | High Efficiency Video Coding (HEVC) [56, 57]. The IBC generates prediction samples of coding blocks from the coded region within a picture using a single Block Vector (BV) that refers to the position of the reference block in the current picture. With the single BV, IBC can efficiently find reference blocks within SC's complex image regions, such as tiled and striped textures, even far from the coding block. To balance the prediction accuracy and signaling bitrates of BV, BV-search-based and BV-search-free methods, namely IBC Block Vector Prediction (BVP) and IBC Merge, are utilized in HEVC, VVC, and the reference software to explore the compression capability beyond VVC, i.e., Enhanced Compression Model (ECM) [57, 56, 58, 13]. As a result, IBC is highly applied to SC and thus achieves significant coding performance gains for SC compared to the conventional intra prediction.

Moreover, to achieve coding performance gains of IBC for CC, several extended methods have been studied, such as IBC with the BV of fractional-sample precision and interpolation filter [59]. They can achieve additional coding performance gains for CC by reducing IBC prediction distortions caused by sample fluctuations due to motion blur, common in CC. In addition, applying IBC to CC for ECM has been investigated, and only IBC BVP was adopted into the ECM version 9 (ECM-9) [58, 13], while IBC Merge was not adopted to avoid encoder runtime increases by IBC [60]. Despite these IBC extensions, there is room to achieve additional coding performance gains of IBC for CC and SC since all the conventional methods utilize only a single BV to generate the prediction samples. Furthermore, if we can minimize the increase

[1]This chapter is based on "Bi-predictive Intra Block Copy for Enhanced Video Coding Beyond VVC" [5], by the same author, which appeared in the Proceedings of IEEE International Conference on Image Processing (ICIP), Copyright(C)2024 IEEE.

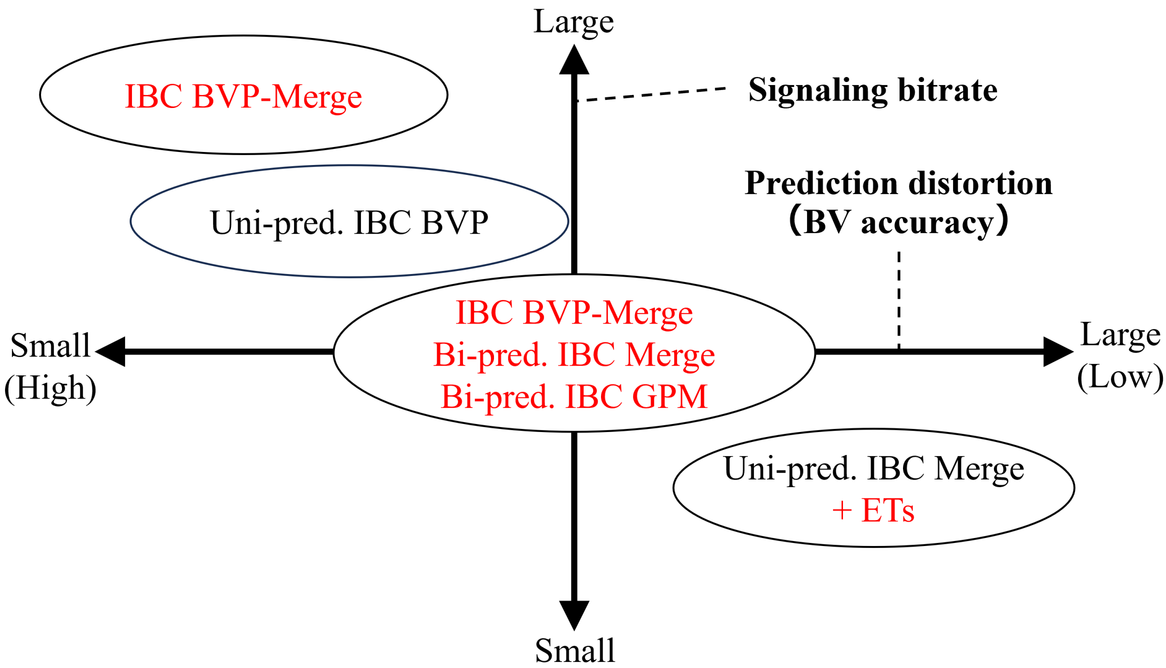


Figure 5.1: Trade-off between prediction distortions and signaling bitrates in IBC algorithms. The red character algorithms are new IBC algorithms proposed in chapter 5. (Copyright(C)2024 IEEE, [5] Fig. 1)

in encoder runtime of IBC Merge for CC while maintaining its coding performance gains, we can achieve reasonable additional coding performance gains for CC in ECM.

5.1.2 Proposal

This chapter proposes to introduce bi-predictive IBC as a new selectable prediction mode in IBC. In other words, the IBC application condition where only uni-predictive IBC can be selected is softened by introducing bi-predictive IBC. Furthermore, we propose an Early Termination (ET) method for applying search-free uni-predictive and bi-predictive IBC to avoid increased encoding runtime while maintaining the coding performance gains of the proposed IBC.

5.1.3 Contribution

To evaluate the effectiveness of the proposed method, we implemented it on top of the ECM-9 and performed the experiments according to the Common Test Condition defined by the Joint Video Experts Team (JVET). Experimental results show that the proposed method brings 0.15% and 0.30% coding performance gains for CC and SC compared to ECM-9 [58, 13] under all intra configuration, with negligible complexity increases [61]. The proposed method has been adopted in ECM version 10 [62, 13] because its effectiveness was recognized by JVET.

5.1.4 Outline

The rest of this chapter is organized as follows. Related work and the corresponding problems are explained in Sec. 5.2. Sec. 5.3 presents the details of the proposed method. Sec. 5.4 describes the experimental results and discussion. Finally, we conclude the chapter in Sec. 5.5.

5.2 Related Work

This section explains the mechanisms of IBC BVP and IBC Merge and then introduces various studies on IBC extension methods to achieve additional coding performance gains for CC and SC.

VVC and ECM can adaptively derive a single BV by IBC BVP and IBC Merge [7, 57, 58], depending on their trade-off between prediction distortions and signaling bitrates as described in Fig. **??**. Specifically, IBC BVP requires a signaled Block Vector Difference (BVD), a searched BV minus a Block Vector Prediction (BVP) commonly derived by the encoder and decoder. Therefore, IBC BVP increases signaling bitrates while reducing prediction distortion due to the highly accurate BV. In contrast, IBC Merge only signals a Merge index corresponding to the BV candidate within the Merge candidate list that registers multiple BV candidates from neighboring blocks. The neighboring block BVs are stored in the BV storage. Hence, IBC Merge decreases signaling bitrates while increasing prediction distortion compared to IBC BVP.

Multiple extension methods have been proposed based on the BV derivation schemes to achieve additional coding performance gains for CC and SC. Regarding CC, several methods of IBC using a fractional-sample-precision BV have been investigated [59, 63, 64, 65]. As another method to reduce prediction distortions, a method of IBC with adaptive interpolation filtering and overlapped-block-sample averaging has been proposed [66]. All the methods leverage the techniques developed for inter prediction and reveal the effectiveness of IBC to CC. Regarding SC, there are several efficient methods of applying IBC considering the image characteristics of SC: IBC with region-based BV searching [67, 68]; IBC with flipping modes [69]; IBC with affine deformation models [70, 71]; fusion of IBC and intra prediction [72]; and IBC with local illumination compensation [72].

Despite these multiple IBC extensions, there is room for additional coding performance gains in CC and SC. For instance, we can obtain more coding performance gain if we adaptively select new BV derivation schemes shown in Fig. 5.1 in addition to the conventional single-BV-based schemes, i.e., uni-predictive IBC BVP and IBC Merge, depending on these trade-offs. Furthermore, in ECM-9 [58], only IBC BVP is enabled for CC to avoid significant encoder runtime increases by IBC Merge [60]. This means that if we can minimize the increase in encoder runtime of IBC Merge for CC while preserving its coding performance gains, we can achieve reasonable additional coding performance gains for CC in ECM.

Table 5.1: Summary of the conventional and proposed IBC methods for CC and SC, denoting bold characters.

Content types	Method	
	Uni-predictive	**Bi-predictive**
CC	IBC BVP **IBC Merge + ETs**	**IBC BVP-Merge** **IBC Merge**
SC	IBC BVP IBC Merge	**IBC GPM**

5.3 Proposed Method

In this chapter, we propose a bi-predictive IBC using two BVs for CC and SC as a new IBC algorithm shown in Fig. 5.1 for reasonable additional coding performance gains. The derivation of two BVs in the bi-predictive IBC is based on the uni-predictive IBC Merge, increasing the encoder runtime for CC over ECM. To avoid the encoder runtime increases while maintaining coding performance gains, we propose encoder Early Terminations (ETs) of the IBC Merge for CC. The bi-predictive IBC has two content-dependent schemes for generating prediction samples and has two BV derivation modes for CC, as summarized in Tab. 5.1.

5.3.1 Bi-predictive IBC for CC and SC

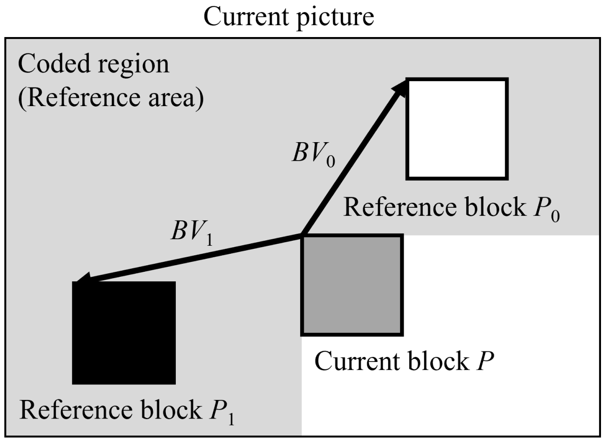
This section describes the proposed bi-predictive IBC from its four key features: BV storage, generation of prediction samples, BV derivation, and signaling. In particular, the BV derivation introduces three modes that achieve new trade-offs between prediction distortions and bitrates shown in Fig. 5.1.

First, regarding BV storage, the maximum number of BVs stored is extended from one to two to realize bi-predictive IBC. In contrast, the BV candidate list is maintained with a design based on uni-predictive IBC to ensure compatibility with uni-predictive IBC. Therefore, when two BVs are available in the BV storage for the following BV derivation process, each BV is added to the BV candidate list.

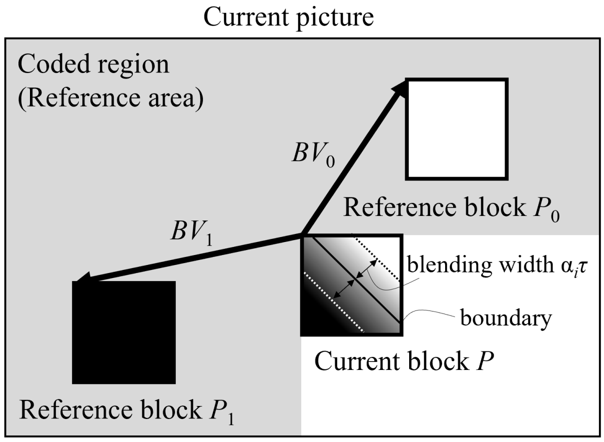
Next, regarding the generation of prediction samples, content-dependent sample blending schemes shown in Fig. 5.2 are adaptively applied based on sequence types. Specifically, a simple block-wise averaging is used to generate final prediction sample values P by two different IBC sample values P_0 and P_1 for CC (Fig. 5.2a), such as the following formula:

$$P(x, y) = (P_0(x, y) + P_1(x, y) + 1) \gg 1, \tag{5.1}$$

where (x, y) denotes the coordinates within the block. In contrast, the Geometric Partitioning Mode (GPM) adaptive blending schemes [58, 73, 74] are leveraged for SC (Fig. 5.2b) instead of the simple averaging. The reason for leveraging this GPM adaptive blending scheme is to



(a) Bi-predictive IBC Merge or IBC BVP-Merge for CC



(b) Bi-predictive IBC GPM for SC

Figure 5.2: Bi-predictive IBC algorithms. (Copyright(C)2024 IEEE, [5] Fig. 2)

preserve the edges often observed in the coding blocks of SC. In other words, simple averaging can lead to over-smoothing of the edges, resulting in increased prediction errors. The final prediction sample values P by the GPM scheme can be derived using the following formulae:

$$w(x,y) = \begin{cases} 0 & d(x,y) \leq -\alpha_i\tau \\ \frac{32}{2\alpha_i\tau}(d(x,y)+\alpha_i\tau) & -\alpha_i\tau < d(x,y) < \alpha_i\tau \\ 32 & d(x,y) \geq \alpha_i\tau, \end{cases} \tag{5.2}$$

$$\begin{aligned} P(x,y) =& (w(x,y) * P_0(x,y) \\ &+ (32 - w(x,y)) * P_1(x,y) + 16) \gg 5, \end{aligned} \tag{5.3}$$

where w, $\alpha_i\tau$, and d indicate the blending weight, width, and Euclidean distance of the prediction sample from the GPM boundary, respectively. Specifically, five different blending widths defined by the default width τ and five weight coefficients α_i = 1/4 1/2, 1, 2, and 4 can be adaptively selected by signaling, the same as GPM with adaptive blending for inter prediction in ECM-9. Formulae 5.2–5.3 indicate that P with $w = 32$ is derived from P_0, while P with $w = 0$ is generated from P_1, thereby conceptually partitioning the coding block into two regions.

Then, regarding BV derivation, the following three modes corresponding to the content-dependent sample blending schemes are introduced. The first mode is a bi-predictive IBC Merge that derives two different BVs from the BV candidate list using two Merge indices corresponding to the BVs. This BV derivation enables the bi-predictive IBC Merge to achieve a new trade-off shown in Fig. 5.1. Specifically, the reduction in prediction distortion achieved by bi-predictive IBC Merge comes at the expense of increased signaling bitrates, as compared to uni-predictive IBC Merge. Additionally, unlike uni-predictive IBC BVP, the bi-predictive IBC Merge does not signal BVD, resulting in smaller bitrates than that of the uni-predictive IBC BVP, while increasing the prediction distortion because of the search-free BV derivation. The second mode is an IBC BVP-Merge that derives two different BVs by leveraging BV derivation schemes of uni-predictive IBC BVP and uni-predictive IBC Merge. Here, BVD is dominant in IBC's signaling overhead. Hence, the IBC BVP-Merge can achieve two new trade-offs shown in Fig. 5.1, depending on the derived BVD. Specifically, when IBC BVP-Merge's BVD is smaller than the uni-predictive IBC BVP, IBC BVP-Merge can reduce prediction distortion more effectively than uni-predictive IBC Merge. When the IBC BVP-Merge's BVD is comparable to uni-predictive IBC BVP, IBC BVP-Merge can further reduce prediction distortion compared to uni-predictive IBC BVP. When IBC BVP-Merge's BVD is larger than uni-predictive IBC BVP, IBC BVP-Merge is not applied. The third mode is a bi-predictive IBC GPM for SC, where two different BVs corresponding to each GPM-separating region are derived from the BV candidate list using two Merge indices, similar to the bi-predictive IBC Merge mode. Depending on the existence of edges within a coding block, the bi-predictive IBC GPM can further reduce prediction distortions compared to the uni-predictive IBC methods.

Finally, regarding signaling, the encoder signals a sequence-level flag and block-level flag to specify the content-dependent blending schemes and adaptively apply them at a coding block. The encoder procedure to determine the IBC application will be explained in the next section.

5.3.2 Early Terminations of IBC Merge for CC

This section explains the proposed ETs of IBC Merge for CC to achieve additional coding performance gains without encoder runtime increases over ECM. Algo. 5.1 outlines the encoder procedure for each coding block on the intra prediction, IBC BVP, and IBC Merge with the proposed ETs. Intra prediction, IBC BVP and IBC Merge have processes to compute the Sum of Absolute Transformed Differences (SATDs) Δ and Rate-Distortion (RD) costs J, respectively. The runtime to compute J is much larger than Δ. The proposed encoder procedure has three-step ETs shown in ETs 1–3 of Algo. 5.1 to avoid the encoder runtime increases for CC caused by enabling bi-predictive IBC together with uni-predictive IBC Merge. The details of the ETs 1–3 are described below.

First, ET 1 is the early termination by counting non-zero coefficients of the best intra prediction mode NC_{Intra} after computing SATD Δ_{Intra} and RD costs J_{Intra} for all the intra

Algorithm 5.1: Proposed encoder procedure with ETs

Input: original CB and reconstructed neighboring CBs
Output: best prediction mode of CB: J_{Best}

for $i \in$ intra mode **do**
  Compute $\Delta_{\text{Intra}}[i]$
  Compute $J_{\text{Intra}}[i]$
  $J_{\text{Intra}}^{\text{min}} \leftarrow \min J_{\text{Intra}}[i]$, $J_{\text{Intra}}^{\text{max}} \leftarrow \max J_{\text{Intra}}[i]$
  $\Delta_{\text{Intra}}^{J\text{max}} \leftarrow \Delta_{\text{Intra}}$ corresponding to $J_{\text{Intra}}^{\text{max}}$
end for
$NC_{\text{Intra}} \leftarrow$ count non-zero coefficients in $J_{\text{Intra}}^{\text{min}}$
if $NC_{\text{Intra}} > 2$ **then** ▷ ET 1
  Compute Δ_{BVP}
  Compute J_{BVP}
  if $J_{\text{BVP}} < \beta_1 J_{\text{Intra}}^{\text{min}}$ **then** ▷ ET 2
    for $j \in$ BV candidates **do**
      Compute $\Delta_{\text{Merge}}[j]$
      if $\Delta_{\text{Merge}}[j] < \beta_2 max(\Delta_{\text{Intra}}^{J\text{max}}, \Delta_{\text{BVP}}$ **then)** ▷ ET 3
        Compute $J_{\text{Merge}}[j]$
        $J_{\text{Merge}}^{\text{min}} \leftarrow \min J_{\text{Merge}}[j]$
      end if
    end for
  end if
end if
$J_{\text{Best}} \leftarrow \min(J_{\text{Intra}}^{\text{min}}, J_{\text{BVP}}, J_{\text{Merge}}^{\text{min}})$
Return J_{Best}

prediction modes. Here, RD cost J is generally defined as the following formula:

$$J = D + \lambda R, \tag{5.4}$$

where D, λ, and R are distortions, a Lagrange multiplier, and bitrates, respectively. In ET 1, the subsequent IBC processing is terminated when the condition of ET 1 is not satisfied. Specifically, IBC is more likely to be applied to coding blocks with complex textures as described in Sec. 5.1.1; thus, IBC has non-zero coefficients for high-frequency components. Therefore, when NC_{Intra} is two or less, i.e., low-frequency components, it can be inferred that the application of IBC is unsuitable.

Next, ET 2 is the early termination by comparing RD costs of the best intra prediction mode and IBC BVP, i.e., $J_{\text{Intra}}^{\text{min}}$ and J_{BVP}. The constant value β_1 is 1.2 in this chapter, utilized in the similar early termination of IBC BVP for SC in ECM-9. In ET 2, the entire IBC Merge encoding process is terminated when the condition of ET 2 is not satisfied. This is because a coded block unsuitable for IBC BVP can be inferred to be unsuitable for IBC Merge.

Then, ET 3 is the early termination by comparing the SATDs of coding blocks in which the intra prediction, IBC BVP, and IBC Merge are applied (i.e., $\Delta_{\text{Intra}}^{J\text{max}}$, Δ_{BVP}, and Δ_{Merge}). Here, $\Delta_{\text{Intra}}^{J\text{max}}$ is the SATD corresponding to the intra prediction mode with the maximum RD cost, i.e., $J_{\text{Intra}}^{\text{max}}$. The constant value β_2 is 1.1 in this chapter, utilized in the similar ET of IBC BVP for SC in ECM-9. In ET 3, BV candidates for IBC Merge, not satisfying the condition of ET 3, are excluded from the targets of computing J_{Merge} because their J_{Merge} can be inferred to be larger than J_{Intra} and J_{BVP}.

5.4 Experimental Results and Discussion

The proposed method is implemented on top of ECM-9 [58, 13]. The experiments were conducted following the JVET's Common Test Condition (CTC) [75]. The coding performance was evaluated by the Bjøntegaard Delta bitrate (BD-rate) of the luma and two chroma components, i.e., BD-rate Y, U, and V components (BDY, BDU, and BDV) [27, 28]. The BD-rate is a well-known evaluation metric to quantify the difference in bitrate for equivalent levels of Peak Signal-to-Noise Ratio (PSNR), whose negative and positive values indicate coding performance gains and losses. Four Quantization Parameter (QP) values 22, 27, 32, and 37 were utilized to derive BD-rates. The complexity was assessed using the ratio of the proposed method's encoder runtime (EncT) and decoder runtime (DecT) compared to ECM-9. The CTC test sequences are classified Classes A through TGM by video resolutions and content types, as described in Tab. 5.2.

The experimental results provided in Tab. 5.2 demonstrate the advantage of the proposed method, which provides additional coding performance gains while maintaining a comparable complexity level to ECM-9. The detailed observations of the results for CC and SC are separately described below.

First, regarding CC, Tab. 5.2 presents the improved performance of the proposed method with a 0.15% coding performance gain on overall results under all intra configuration. Comparing the average performance of each class on average and each test sequence reveals the proposed method's sequence-dependent effectiveness. In particular, coding performance gains of more than 0.5% are observed in DaylightRoad2 and BQTerrace. DaylightRoad2 and BQTerrace contain image regions with tiles and stripes throughout the sequence. Moreover, DaylightRoad2 is a video from a moving vehicle window, and BQTerrace includes a camera pan scene. Such complex image regions with motion blur contribute to frequent applications of bi-predictive IBC and thus result in improved performance. In addition, the proposed method maintains DecT and increases EncT by only 1%, compared to ECM-9. The proposed uni-predictive IBC Merge and bi-predictive IBC do not increase DecT since they can derive BVs without BV searching in the decoder. Additionally, the minor increase in EncT is attributed to the proposed ETs, which will be further discussed in the ablation study.

Next, regarding SC, Tab. 5.2 presents the improved performance of the proposed method with a 0.30% coding performance gain on overall results without increasing EncT and DecT. Similar to the CC results, the sequence-dependent effectiveness of the proposed method is observed in SC. In particular, coding performance gains of more than 0.5% are observed in SlideEditing, FlyingGraphic, and Desktop. SlideEditing and Desktop contain scrolled consoles and windows, whereas FlygingGraphic includes randomly moving texts and shapes. Such complex image regions having edges within the blocks contribute to frequent applications of bi-predictive IBC GPM and thus result in improved performance for SC, too.

Tab. 5.3 shows the ablation study of the proposed method for CC with overall BD-rates Y, EncT, and DecT over ECM-9. Due to the greater impact of the luma component on perceptual quality than the chroma components, only BD-rates Y are compared in the evaluation. The uni-predictive IBC Merge brings a 0.15% coding performance gain while significantly increasing EncT by 7.8%. In addition, the bi-predictive IBC raises the coding performance gain up to 0.21% while further increasing EncT up to 9%. Meanwhile, the proposed ETs can reduce EncT from 9% to 1% while almost retaining the coding performance gain from the proposed method, revealing the proposed ET's effectiveness.

To investigate the differences in the effectiveness of the proposed method, Figs. 5.3 and 5.4 provide the application rate of each IBC method and the first decoded frames overlaying IBC-applied blocks for BQTerrace and BQSquare with two QPs (22 and 37), respectively. BQTerrace and BQSquare have similar content, but the resolutions and observed coding performance gains differ. Specifically, BQTerrace is a Class A sequence with the largest coding performance gain of 0.68%, whereas BQSquare is a Class D sequence with a small coding performance gain of 0.07%.

Comparing the application rates of Fig. 5.3 reveals that the proposed IBC has high application rates for the two test sequences under low-QP conditions, confirming the effectiveness of the proposed method as intended in Fig. 5.1. Specifically, at QP = 22, the proposed method increases the application rates of IBC in BQTerrace and BQSquare, with more than half of the total rates attributed to the proposed IBC. Meanwhile, in BQTerrace at QP = 37, the rate of the proposed method drops, though the total rates are almost the same as QP = 22. Moreover, in BQSquare at QP = 37, even the total rates decreased compared to those at QP = 22.

The different tendencies on the IBC application rates of BQTerrace and BQSquare under high QP conditions could lead to different coding performance gains of BQTerrace and BQSquare. Specifically, the drops in the rate of the proposed IBC in BQTerrace and BQSquare at QP = 37 come from coding degradation, which smooths or loses the edges of complex image regions. In particular, in BQTerrace, uni-predictive IBC is highly selected to avoid over-smoothing by the proposed bi-predictive IBC. Furthermore, in BQSquare, the total rates decrease due to the loss of edges and the increased signaling overhead for adaptation of the proposed method and uni-predictive IBC. Despite the increased signaling overhead, the total application rate of IBC is maintained in BQTerrace. This is because BQTerrace has a higher resolution than BQSquare and has many large blocks that are less affected by the signaling overhead.

Figure 5.4(a)–(d) demonstrates the application of the proposed IBC to complex image regions and the different tendencies on IBC application rates dependent on QP and sequence

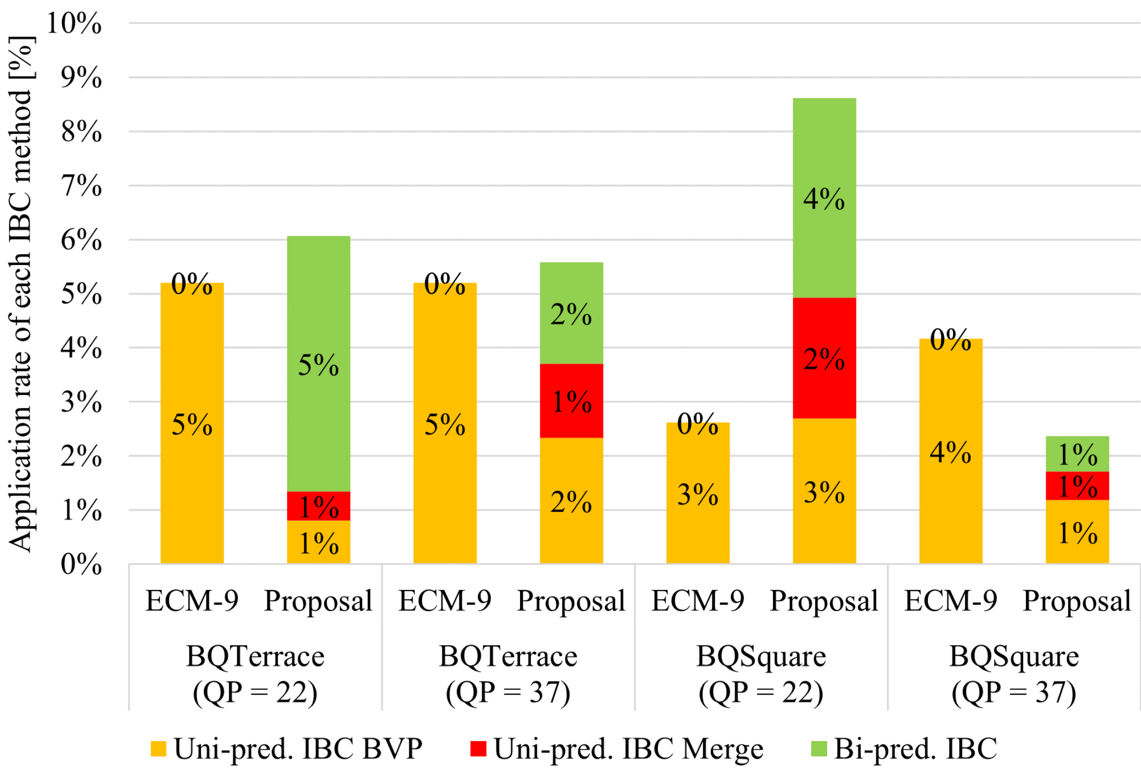


Figure 5.3: Application rates of each IBC method normalized by all samples in each test sequence. (Copyright(C)2024 IEEE, [5] Fig. 3)

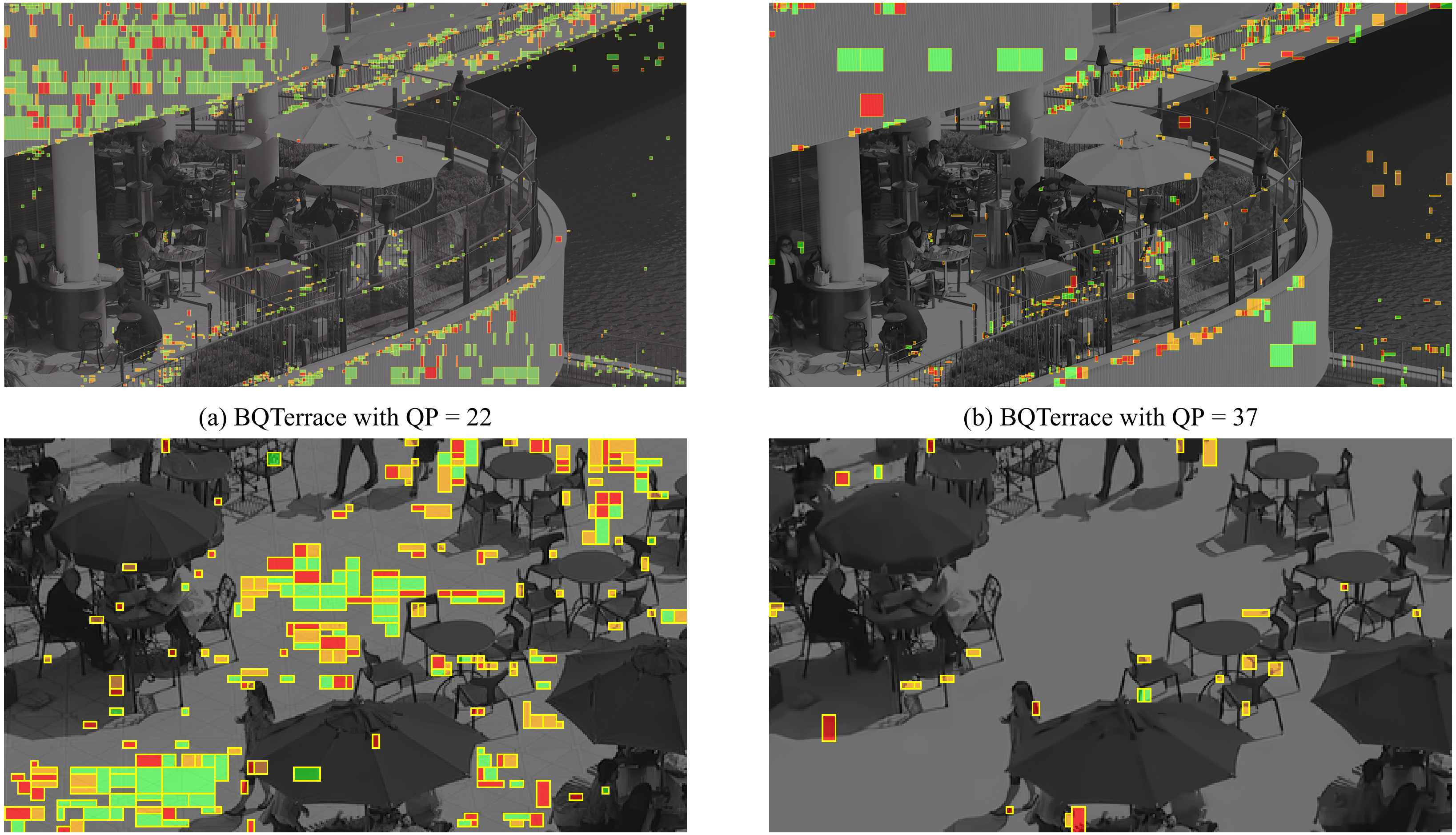
resolution, discussed in Fig. 5.1. Specifically, regarding BQTerrace at QP = 22 shown in Fig. 5.4a, the many small blocks that apply the proposed IBC are observed in the terrace's outer wall and the bridge's railing. In contrast, regarding BQTerrace at QP = 37 shown in Fig.5.4b, large blocks that apply the proposed IBC and small ones that apply uni-predictive IBC are observed in such regions. Regarding BQSquare at QP = 22 shown in Fig. 5.4c, the smaller blocks that apply the proposed method and uni-predictive IBC are observed around the terrace tile pattern. Meanwhile, regarding BQSquare at QP = 22 shown in Fig. 5.4d, IBC-applied blocks drastically decrease in such regions due to the loss of the tile pattern. Based on the discussion above, the proposed method has the potential to achieve additional coding performance gains or further reduce complexity by considering the adaptive application of bi-predictive IBC dependent on QP and block size.

Table 5.2: Results of the proposed method over ECM-9. The CC overall BD-rates with EncT and DecT exclude the results of Class D following CTC.

Cotenty type | Class	Sequence	BDY [%]	BDU [%]	BDV [%]
CC | A1	Tango2	-0.02	-0.15	-0.05
	FoodMarket4	-0.01	0.13	-0.07
	Campfire	-0.01	-0.04	0.06
CC | A2	CatRobot	-0.21	-0.20	-0.08
	DaylightRoad2	-0.53	-0.65	-0.53
	ParkRunning3	0.00	0.00	-0.04
CC | B	MarketPlace	0.01	0.04	0.16
	RitualDance	-0.01	-0.07	-0.08
	Cactus	-0.17	-0.26	-0.26
	BasketballDrive	-0.23	0.07	-0.40
	BQTerrace	-0.68	-0.73	-0.77
CC | C	BasketballDrill	-0.05	-0.10	-0.22
	BQMall	-0.11	0.11	0.05
	PartyScene	-0.18	-0.28	-0.27
	RaceHorses	0.03	0.05	-0.10
CC | D	BasketballPass	0.00	-0.34	-0.29
	BQSquare	-0.07	0.01	-0.03
	BlowingBubbles	-0.01	0.09	0.08
	RaceHorses	0.02	0.08	0.05
CC | E	FourPeople	-0.14	-0.14	-0.07
	Johnny	-0.20	-0.03	-0.48
	KristenAndSara	-0.15	-0.08	0.05
CC | Overall		**-0.15**	**-0.13**	**-0.17**
CC | EncT [%] / DecT [%]		**101.0 / 100.4**		
SC | F	BasketballDrillText	-0.03	-0.08	-0.08
	ArenaOfValor	-0.03	0.03	0.09
	SlideEditing	-0.51	-0.44	-0.48
	SlideShow	-0.12	-0.13	-0.26
SC | TGM	FlyingGraphic	-0.53	-0.47	-0.50
	Desktop	-0.69	-0.73	-0.78
	Console	-0.32	-0.19	-0.29
	ChineseEditing	-0.15	-0.09	-0.13
SC | Overall		**-0.30**	**-0.26**	**-0.30**
SC | EncT [%] / DecT [%]		**100.3 / 99.5**		

Table 5.3: Ablation study of the proposed method for CC with overall BD-rates Y, EncT, and DecT over ECM-9.

No.	Method	BDY [%]	EncT [%]	DecT [%]
1	Uni-pred. IBC Merge	-0.15	107.8	100.5
2	No.1 + Bi-pred. IBC	-0.21	109.3	100.6
3	No.2 + Proposed ETs	-0.15	101.0	100.4



(a) BQTerrace with QP = 22

(b) BQTerrace with QP = 37

(c) BQSquare with QP = 22

(d) BQSquare with QP = 37

Figure 5.4: The first decoded frames overlaying IBC-applied blocks for BQTerrace and BQSquare with two QPs (22 and 37). The yellow grid indicates the boundary of the blocks, and the colors of blocks represent each IBC method {Orange: uni-predictive IBC BVP; Red: uni-predictive IBC Merge; Green: bi-predictive IBC; and Transparent: intra prediction}.(Copyright(C)2024 IEEE, [5] Fig. 4)

5.5 Conclusion

This chapter proposes a bi-predictive IBC using two block vectors as a new IBC. In addition, this paper also proposes early terminations of search-free IBC, which is utilized for deriving block vectors in bi-predictive IBC. The experimental results show the proposed method provides 0.15% and 0.30% coding performance gains with negligible increases in the encoder and decoder runtime for camera-captured and screen contents compared to ECM-9.

Chapter 6

Conclusion

In this dissertation, we aimed to achieve a highly efficient video coding method to address the increasing video data traffic. For this goal, we targeted improving the prediction efficiency to enhance the coding performance of hybrid coding architecture used in international video coding standards. Specifically, we focused on extensions of prediction and studied the methods that implement ideal soft decision criteria as real-world solutions considering actual operations rather than the conventional deterministic mode-switching decision (i.e., hard decision) criteria. In Chapters 2 to 5, we clarified hard decision elements in the conventional prediction methods and propose to soften the hard decision elements, considering the better trade-off between coding performance and complexity.

Chapter 2 [Block-size and QP Dependent Intra Switchable Interpolation Filters] explored improving the prediction accuracy of an intra Switchable Interpolation Filter (SIF). First, we presented block-size dependent intra SIFs with two different cutoff frequency characteristics, i.e., cubic and Gaussian Interpolation Filters (IFs), which were finally adopted in H.266 | Versatile Video Coding (VVC). The cubic and Gaussian IFs can be switched by a fixed threshold defined by one block size of the current coding blocks. Specifically, the Gaussian IF has a higher denoising (i.e., smoothing) effect of prediction distortion than the cubic IF. Hence, the conventional method applies the Gaussian IF to large-size blocks with flat image characteristics. In contrast, it applies the cubic IF to small-size blocks with complex image characteristics such as edges. However, since the block sizes and prediction distortions depend on the Quantization Parameters (QPs), the conventional intra SIF is not always optimal for wide QP ranges. Therefore, we proposed an intra SIF with variable thresholds defined by QP dependent block sizes. In other words, we introduce the soft-decision-oriented threshold for switching different cutoff frequencies to obtain high prediction accuracy over a wide bit rate range. Finally, we demonstrated the experimental results that the proposed method provides better coding performance than the conventional method in a high QP range.

Chapter 3 [Memory Bandwidth Constrained Overlapped Block Motion Compensation] studied the better trade-off between coding performance and memory bandwidth for Overlapped Block Motion Compensation (OBMC). First, we explained the OBMC memory bandwidth issue and illustrated memory bandwidth determinants for regular

Motion Compensation (MC) and OBMC. Next, we introduced the idea that the application condition of the conventional OBMC is designed not to exceed the maximum (i.e., worst-case) memory bandwidth required for regular MC without OBMC. Specifically, the application of OBMC is determined by a fixed current block size and a fixed number of Motion Vectors (MVs) of the current block and neighboring blocks, which are the determinants of memory bandwidth. However, such OBMC application decisions by the fixed memory bandwidth determinants leave room for OBMC application for the worst-case memory bandwidth and do not maximize the potential coding performance of OBMC. Therefore, we proposed a memory bandwidth-constrained OBMC method that treats the memory determinants of OBMC, such as the number of MVs and the IF length of neighboring blocks, as variables depending on the current coding block sizes. Furthermore, we generalized the problem set as a constrained objective function that maximizes memory bandwidth for a predefined upper limit and derives soft-determined variable OBMC parameters. Finally, we showed experimental results that the proposed method provides better coding performance than the conventional method with the same worst-case memory bandwidth as the conventional method.

Chapter 4 [Geometric Partitioning Mode with Inter Prediction and Intra Prediction] researched improving the prediction accuracy of Geometric Partitioning Mode (GPM). First, we summarized the GPM algorithm with two different inter predictions (GPM-Inter/Inter), which was finally adopted in VVC. Whether GPM-Inter/Inter is applied is determined by signaling. However, GPM-Inter/Inter does not necessarily predict the boundary of the objects with high accuracy, especially for low-latency video coding configurations where we can fetch the reference samples only from past coded pictures. For example, GPM-Inter/Inter cannot accurately predict the boundary between the background and foreground, which appears after the intersection of two foreground objects, because the background region is not included in the past coded pictures. Therefore, we proposed introducing GPM with inter and intra predictions (GPM-Inter/Intra) as a new selectable prediction mode of GPM in addition to GPM-Inter/Inter. In other words, the proposed method can soften the application conditions of GPM with GPM-Inter/Intra. Furthermore, to suppress the signaling overhead of GPM-Inter/Intra, we restricted the types of selectable intra-prediction modes in GPM-inter/intra. Finally, we demonstrated experimental results that the proposed method has superior coding performance compared to the conventional method and qualitatively suppresses artifacts observed in the conventional method for low-delay video coding configurations.

Chapter 5 [Bi-predictive Intra Block Copy] pursued a better trade-off between bitrates and distortion and encoding runtime of Intra Block Copy (IBC). First, we described conventional IBC methods, such as those used in VVC and Enhanced Compression Model (ECM). All the conventional IBC methods use a single Block Vector (BV), i.e., uni-predictive IBC, and whether IBC is applied is determined by signaling. The uni-predictive IBC is classified into a BV-search-based IBC and a BV-search-free IBC, and we can select two IBC modes to obtain a better trade-off between bitrates and distortions. The BV-search-based IBC method can decrease prediction distortion with high-accurate BV while increasing signaling overhead more than the BV-search-free IBC method. The characteristic of the BV-search-free IBC method is vice versa. However, there is room for further coding performance gain if we introduce a new IBC method that realizes an intermediate trade-off between BV-search-based and BV-search-free IBC methods. On the other hand, in that case, we also have to consider how to realize a better trade-off between coding performance gains and encoding runtime of IBC since a new IBC method increases the selection of IBC modes. Therefore, we proposed introducing bi-predictive IBC as a new selectable prediction mode in IBC. In other words, the IBC application condition where only uni-predictive IBC can be selected is softened by introducing bi-predictive IBC. Furthermore, to avoid increases in encoding runtime while maintaining the coding performance of the proposed IBC, we introduced an early termination method. This method determines cases where the application of the bi-predictive IBC is ineffective and suspends the proposed IBC's coding process. Finally, we demonstrated experimental results that the proposed method does not significantly increase the encoding runtime and achieves better coding performance than conventional methods.

In conclusion, all the proposed soft-decision-oriented predictions achieved better coding performance beyond VVC. In particular, the proposed method in Chapters 4 and 5 have been adopted in ECM, suggesting that they have been highly evaluated in the industry. The soft decision-oriented approach can be applied to functions other than prediction in hybrid video coding, which is the subject of this dissertation, such as block segmentation, transformation with quantization, filtering in the loop, and entropy coding. The specific procedure of the soft-decision-oriented approach is as follows: Identify hard decision elements of each process of the functions; Try to soften the hard decision elements to unlock the function's potential; Generalize soft-decision-oriented methods. Here, soft decisions can improve coding performance while increasing complexity. Therefore, we need to estimate an acceptable level of complexity that considers the evolution of hardware capabilities. Then, we are required to implement a generalized soft decision-oriented approach as a realistic solution that takes into account actual operations according to that acceptable level. This approach is essential to achieve practical video coding that can overcome the continuously increasing video data traffic in the future.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to Prof. Hiroshi Watanabe (渡辺裕教授) at the Fundamental Science and Engineering of Waseda University for the continuous support of my research work over three years.

Second, I appreciate the strong support and valuable advice from Prof. Jiro Katto (甲藤二郎教授), Prof. Wataru Kameyama (亀山渉教授) and Prof. Hiroyuki Kasai (笠井裕之教授) at the Fundamental Science and Engineering of Waseda University.

I am deeply thankful to Takaaki Ishikawa (石川孝明氏), invited researcher of Watanabe Laboratory for multiple helpful technical advice.

I would like to offer my special thanks to my colleagues in KDDI Rsearch, Inc., especially to Dr. Sei Naito (内藤整氏), Dr. Kei Kawamura (河村圭氏), Dr. Haruhisa Kato (加藤晴久氏), Dr. Kyohei Unno (海野恭平氏).

I also thank all the students in the Watanabe Laboratory.

Outside the laboratory and KDDI Research, Inc., my family always encouraged me. I take this opportunity to thank everyone involved.

Reference

[1] Y. Kidani, K. Kawamura, K. Unno, and S. Naito, “Blocksize-QP Dependent Intra Interpolation Filters,” in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4125–4129, 2019. [Online]. Available: https://doi.org/10.1109/ICIP.2019.8803456

[2] Y. Kidani, K. Unno, K. Kawamura, and H. Watanabe, “Memory bandwidth constrained overlapped block motion compensation for video coding,” *The Institute of Image Information and Television Engineers Transactions on Media Technology and Applications (ITE-MTA)*, vol. 11, no. 1, pp. 1–12, 2023. [Online]. Available: https://doi.org/10.3169/mta.11.1

[3] Y. Kidani, H. Kato, K. Kawamura, and H. Watanabe, “Geometric partitioning mode with inter and intra prediction for beyond versatile video coding,” *IEICE Transactions on Information and Systems*, vol. E105.D, no. 10, pp. 1691–1703, 2022. [Online]. Available: https://doi.org/10.1587/transinf.2022PCP0005

[4] H. Gao, S. Esenlik, E. Alshina, and E. Steinbach, “Geometric Partitioning Mode in Versatile Video Coding: Algorithm Review and Analysis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 9, pp. 3603–3617, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2020.3040291

[5] Y. Kidani, H. Kato, and K. Kawamura, “Bi-predictive intra block copy for enhanced video coding beyond vvc,” in *2024 IEEE International Conference on Image Processing (ICIP)*, 2024, [accepted on June 2024].

[6] “*Recommendation* H.265: High Efficiency Video Coding,” ITU-T, 2013. [Online]. Available: https://www.itu.int/rec/T-REC-H.265

[7] “*Recommendation* H.266: Versatile Video Coding,” ITU-T, 2020. [Online]. Available: https://www.itu.int/rec/T-REC-H.266

[8] “放送システムに関する技術的条件－情報通信審議会からの答申－”，総務省，2023．[Online]. Available: https://www.soumu.go.jp/menu_news/s-news/01ryutsu08_02000287.html

[9] “Ericsson Mobility Reprot, November 2023,” Ericsson, 2023. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2023

[10] “Cisco Annual Internet Report (2018-2023) White Paper,” Cisco Systems, 2020. [Online]. Available: https://www.cisco.com/c/ja_jp/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[11] “Transmission experiment using real-time codec compliant with the latest international standard of point cloud compression,” KDDI Research, Inc., 2023. [Online]. Available: https://www.kddi-research.jp/english/newsrelease/2023/012401.html

[12] 篠田 貴之，渡邊 勇二，佐々木 聡史，神谷 泰次，佐藤 肇，伊達 厚，“ボリュメトリックビデオ技術を用いたプロ野球中継”，映像情報メディア学会誌，vol. 78，no. 2，pp. 247–251，2024．[Online]. Available: https://doi.org/10.3169/itej.78.247

[13] Enhanced Compression Model (ECM). [Online]. Available: https://vcgit.hhi.fraunhofer.de/ecm/ECM

[14] 甲藤二郎，“【映像符号化】映像符号化・配信における深層学習の広がり”，電子情報通信学会誌，vol. 105，no. 5，pp. 383–386，2022．

[15] 猪飼智宏，“AI の映像符号化への応用”，映像情報メディア学会誌，vol. 78，no. 1，pp. 50–55，2024．[Online]. Available: https://www.ite.or.jp/contents/tech_guide/tech_guide202302_202401.pdf

[16] G. J. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE signal processing magazine*, vol. 15, no. 6, pp. 74–90, 1998.

[17] 安田豊，“ヴィタビ復号による誤り訂正方式の研究”，京都大学，1983．

[18] 日下卓也，“1 群 2 編 2 章 2-6 軟判定復号法”，電子情報通信学会知識の森，pp. 15–17，2012．[Online]. Available: https://www.ieice-hbkb.org/portal/01-2/01_02/

[19] 西村芳一，“[改定新版] データの符号化技術と誤り訂正の基礎”，CQ 出版社，2018．

[20] J. Chen, E. Alshina, G. J. Sullivan, J.-R. Ohm, and J. Boyce, “Algorithm description of Joint Exploration Test Model 7 (JEM7),” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-G1001. 2017.

[21] S. Matsuo, S. Takamura, and A. Shimizu, “Modification of intra angular prediction in HEVC,” in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1–4, 2012.

[22] S. Matsuo, S. Takamura, and H. Jozawa, “Improved intra angular prediction by DCT-based interpolation filter,” in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 1568–1572, 2012.

[23] R. Wei, R. Xie, L. Song, L. Zhang, and W. Zhang, “Improved intra angular prediction with novel interpolation filter and boundary filter,” in *2016 Picture Coding Symposium (PCS)*, pp. 1–5, 2016. [Online]. Available: https://doi.org/10.1109/PCS.2016.7906392

[24] S. Yoo, J. Heo, J. Choi, L. Li, and J. Lim, “CE3-3.1.1: Interpolation filter selection regarding intra mode and block size,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-L0130. 2018.

[25] J. Chen, Y. Ye, and S. Kim, “Algorithm description for Versatile Video Coding and Test Model 2 (VTM 2),” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-K1002. 2018.

[26] F. Bossen, J. Boyce, X. Li, V. Seregin, and K. Sühring, “JVET common test conditions and software reference configurations for SDR video,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-K1010. 2018.

[27] G. Bjøntegaard, “Calculation of average PSNR differences between RD-curves,” 2001.

[28] K. Andersson, F. Bossen, J.-R. Ohm, A. Segall, R. Sjöberg, J. Ström, and G. J. Sullivan, “Summary information on BD-rate experiment evaluation practices,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-Q2016. 2020.

[29] H. Watanabe and S. Singhal, “Windowed motion compensation,” in *Visual Communications and Image Processing ’91: Visual Communication*, K.-H. Tzou and T. Koga, Eds., vol. 1605, pp. 582 – 589, 1991. [Online]. Available: https://doi.org/10.1117/12.50301

[30] S. Nogaki and M. Ohta, “An overlapped block motion compensation for high quality motion picture coding,” in *[Proceedings] 1992 IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 184–187 vol.1, 1992. [Online]. Available: https://doi.org/10.1109/ISCAS.1992.229983

[31] P. Lai, S. Liu, and S. Lei, “Combined temporal and inter-layer prediction for scalable video coding using HEVC,” in *2013 Picture Coding Symposium (PCS)*, pp. 117–120, 2013. [Online]. Available: https://doi.org/10.1109/PCS.2013.6737697

[32] Y.-W. Chen and X. Wang, “AHG5: Reducing VVC Worst-case Memory Bandwidth by Restricting Bi-directional 4x4 Inter CUs/Sub-block,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-L0104. 2018.

[33] Z.-Y. Lin, T.-D. Chuang, C.-Y. Chen, C.-W. Hsu, Z.-Y. Lin, Y.-C. Lin, Y.-W. Huang, S.-M. Lei, X. Xiu, and Y. He, “CE10.2.1: Uni-prediction-based CU-boundary-only OBMC,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-M0178. 2019.

[34] G. Sullivan and J.-R. Ohm, “Meeting Report of the 13th Meeting of the Joint Video Experts Team (JVET),” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-M1000. 2019.

[35] R. Hashimoto and S. Mochizuki, "AHG5: How to Use the Software to Evaluate Memory Bandwidth," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-J0090. 2018.

[36] M. Budagavi and M. Zhou, "Video coding using compressed reference frames," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1165–1168, 2008. [Online]. Available: https://doi.org/10.1109/ICASSP.2008.4517822

[37] M. E. Sinangil, A. P. Chandrakasan, V. Sze, and M. Zhou, "Memory cost vs. coding efficiency trade-offs for HEVC motion estimation engine," in *2012 IEEE International Conference on Image Processing (ICIP)*, pp. 1533–1536, 2012. [Online]. Available: https://doi.org/10.1109/ICIP.2012.6467164

[38] H. Gao, X. Chen, S. Esenlik, J. Chen, and E. Steinbach, "Decoder-Side Motion Vector Refinement in VVC: Algorithm and Hardware Implementation Considerations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3197–3211, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2020.3037024

[39] Y. Chen and D. Mukherjee, "Variable block-size overlapped block motion compensation in the next generation open-source video codec," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 938–942, 2017. [Online]. Available: https://doi.org/10.1109/ICIP.2017.8296419

[40] Y. Liao, A. Leontaris, and A. M. Tourapis, "A low complexity architecture for video coding with overlapped block motion compensation," in *2010 IEEE International Conference on Image Processing (ICIP)*, pp. 2041–2044, 2010. [Online]. Available: https://doi.org/10.1109/ICIP.2010.5649822

[41] J. Chen, Y. Chen, M. Karczewicz, X. Li, H. Liu, L. Zhang, and X. Zhao, "Coding tools investigation for next generation video coding based on HEVC," in *Applications of Digital Image Processing XXXVIII*, A. G. Tescher, Ed., vol. 9599, p. 95991B, 2015. [Online]. Available: https://doi.org/10.1117/12.2193681

[42] H. Yang, H. Chen, J. Chen, S. Esenlik, S. Sethuraman, X. Xiu, E. Alshina, and J. Luo, "Subblock-Based Motion Derivation and Inter Prediction Refinement in the Versatile Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3862–3877, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3100744

[43] Y. Kidani, K. Kawamura, K. Unno, and S. Naito, "Block-Size Dependent Overlapped Block Motion Compensation," in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 1191–1195, 2020. [Online]. Available: https://doi.org/10.1109/ICIP40778.2020.9191338

[44] J. Chen, Y. Ye, S. Kim *et al.*, "Algorithm Description for Versatile Video Coding and Test Model 10 (VTM 10)," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-S2002. 2020.

[45] F. Bossen, J. Boyce, X. Li, V. Seregin, and K. Sühring, "VTM common test conditions and software reference configurations for SDR video," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-T2010. 2020.

[46] W.-J. Chien, L. Zhang, M. Winken, X. Li, R.-L. Liao, H. Gao, C.-W. Hsu, H. Liu, and C.-C. Chen, "Motion Vector Coding and Block Merging in the Versatile Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3848–3861, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3101212

[47] M. Bläser, J. Sauer, and M. Wien, "Description of SDR and 360° video coding technology proposal by RWTH Aachen University," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-J0023. 2018.

[48] B. Bross, K. Andersson, M. BlÃ¤ser, V. Drugeon, S.-H. Kim, J. Lainema, J. Li, S. Liu, J.-R. Ohm, G. J. Sullivan, and R. Yu, "General Video Coding Technology in Responses to the Joint Call for Proposals on Video Compression With Capability Beyond HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1226–1240, 2020. [Online]. Available: https://doi.org/10.1109/TCSVT.2019.2949619

[49] S. Blasi, A. S. Dias, and G. Kulupana, "Non-CE4: CIIP using triangular partitions," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-O0522. 2019.

[50] Y.-W. Huang, J. An, H. Huang, X. Li, S.-T. Hsiang, K. Zhang, H. Gao, J. Ma, and O. Chubach, "Block Partitioning Structure in the VVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3818–3833, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3088134

[51] J. Pfaff, A. Filippov, S. Liu, X. Zhao, J. Chen, S. De-Luxán-Hernández, T. Wiegand, V. Rufitskiy, A. K. Ramasubramonian, and G. Van der Auwera, "Intra Prediction and Mode Coding in VVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3834–3847, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3072430

[52] S. Blasi, A. S. Dias, and G. Kulupana, "Non-CE4: CIIP using triangular partitions," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-O0522. 2019.

[53] Z. Deng, L. Zhang, H. Liu, K. Zhang, and Y. Wang, "CE4-related: further constraints on block shapes for GEO," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-Q0309. 2020.

[54] Y. Kidani, H. Kato, and K. Kawamura, "AHG12: GPM with inter and intra prediction," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-W0110. 2021.

[55] ——, "EE2-related: Modified GPM with inter and intra prediction," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-X0078. 2021.

[56] T. Nguyen, X. Xu, F. Henry, R.-L. Liao, M. G. Sarwer, M. Karczewicz, Y.-H. Chao, J. Xu, S. Liu, D. Marpe, and G. J. Sullivan, "Overview of the screen content support in vvc: Applications, coding tools, and performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3801–3817, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3074312

[57] X. Xu, S. Liu, T.-D. Chuang, Y.-W. Huang, S.-M. Lei, K. Rapaka, C. Pang, V. Seregin, Y.-K. Wang, and M. Karczewicz, "Intra block copy in hevc screen content coding extensions," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 409–419, 2016. [Online]. Available: https://doi.org/10.1109/JETCAS.2016.2597645

[58] M. Coban, R.-L. Liao, K. Naser, J. Strom, and L. Zhang, "Algorithm description of Enhanced Compression Model 9 (ECM 9)," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AD2025. 2023.

[59] H. Chen, Y.-S. Chen, M.-T. Sun, A. Saxena, and M. Budagavi, "Improvements on intra block copy in natural content video coding," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2772–2775, 2015. [Online]. Available: https://doi.org/10.1109/ISCAS.2015.7169261

[60] C.-C. Chen, B. Ray, M. Coban, H. Wang, V. Seregin, M. Karczewicz, W. Chen, X. Xiu, C. Ma, H.-J. Jhu, C.-W. Kuo, N. Yan, and X. Wang, "EE2-1.8/1.9: IBC adaptation for coding of natural content," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AD0208. 2023.

[61] Y. Kidani, H. Kato, K. Kawamura *et al.*, "EE2-2.2/EE2-2.3/EE2-2.4: Bi-predictive IBC GPM, IBC BVP-merge, bi-predictive IBC merge, and IBC MBVD list derivation for camera captured and screen contents," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AE0169. 2023.

[62] M. Coban, R.-L. Liao, K. Naser, J. Strom, and L. Zhang, "Algorithm description of Enhanced Compression Model 10 (ECM 10)," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AE2025. 2023.

[63] J. Xu, "EE2-3.2-related: IBC adaptation for camera-captured contents," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-Z0157. 2022.

[64] B. Ray, H. Wang, C.-C. Chen, V. Seregin, and M. Karczewicz, "Non-EE2: IBC adaptation for coding of natural content," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AC0161. 2023.

[65] W. Chen, X. Xiu, C. Ma, H.-J. Jhu, C.-W. Kuo, and X. W. N. Yan, "Non-EE2: IBC with fractional block vectors," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AC0172. 2023.

[66] H. Kato, Y. Kidani, K. Kawamura, and S. Naito, “Extended Intra Block Copy with Adaptive Filtering and Overlapped Block Averaging,” in *2023 IEEE International Conference on Visual Communication and Image Processing (VCIP)*, pp. 1–5, 2023. [Online]. Available: https://doi.org/10.1109/VCIP59821.2023.10402667

[67] G. Venugopal, S. De-Luxan-Hernandez, K. Muller, H. Schwarz, D. Marpe, and T. Wiegand, “Region-based predictors for intra block copy,” in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 1166–1170, 2020. [Online]. Available: https://doi.org/10.1109/ICIP40778.2020.9191005

[68] D. Feng, C. Zhu, G. Lu, and L. Song, “Position-based motion vector prediction for textual image coding,” in *2022 Picture Coding Symposium (PCS)*, pp. 55–59, 2022. [Online]. Available: https://doi.org/10.1109/PCS56426.2022.10018060

[69] J. Cao, F. Liang, and J. Wang, “Intra Block Copy Mirror Mode for Screen Content Coding in Versatile Video Coding,” *IEEE Access*, vol. 9, pp. 31 390–31 400, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3060448

[70] D. Li, Z. Zhang, K. Qiu, Y. Pan, Y. Li, H. R. Hu, and L. Yu, “Affine Deformation Model Based Intra Block Copy for Intra Frame Coding,” in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2020. [Online]. Available: https://doi.org/10.1109/ISCAS45731.2020.9180453

[71] J. Adhuran, A. Fernando, G. Kulupana, and S. Blasi, “Affine Intra-prediction for Versatile Video Coding,” in *2020 European Signal Processing Conference (EUSIPCO)*, pp. 545–549, 2021. [Online]. Available: https://doi.org/10.23919/Eusipco47968.2020.9287579

[72] Y. Wang, K. Zhang, and L. Zhang, “Local-Aware Intra Block Copy for Video Coding,” in *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2023. [Online]. Available: https://doi.org/10.1109/ISCAS46773.2023.10181637

[73] Y. Kidani, H. Kato, K. Unno, K. Kawamura, N. Yan, X. Xiu, W. Chen, H.-J. Jhu, C.-W. Kuo, and X. Wang, “GPM adaptive blending,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AA0058. 2022.

[74] H. Kato, Y. Kidani, K. Kawamura, and S. Naito, “Adaptive boundary width of Geometric Partitioning Mode for Beyond Versatile Video Coding,” in *2022 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pp. 1–5, 2022. [Online]. Available: https://doi.org/10.1109/VCIP56404.2022.10008865

[75] M. Karczewicz and Y. Ye, “Common test conditions and evaluation procedures for enhanced compression tool testing,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AD2017. 2023.

[76] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. 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. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3101953

[77] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, “Overview of the Versatile Video Coding (VVC) Standard and its Applications,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3101953

[78] V. Seregin, J. Chen, F. L. Léannec, and K. Zhang, “JVET AHG report: ECM software development (AHG6),” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AE0006. 2023.

[79] V. Seregin *et al.*, “Exploration experiment on enhanced compression beyond VVC capability (EE2),” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AD2024. 2023.

[80] W.-J. Chien, J. Boyce, Y.-W. Chen, R. Chernyak, K. Choi, R. Hashimoto, Y.-W. Huang, H. Jang, R.-L. Liao, and S. Liu, “JVET AHG report: tool reporting procedure and testing (AHG13),” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-T2013. 2020.

[81] M. Bläser, J. Schneider, J. Sauer, and M. Wien, “Geometry-based Partitioning for Predictive Video Coding with Transform Adaptation,” in *2018 Picture Coding Symposium (PCS)*, pp. 134–138, 2018. [Online]. Available: https://doi.org/10.1109/PCS.2018.8456238

[82] Y. Kidani, H. Kato, and K. Kawamura, “Non-EE2: Bi-predictive IBC for natural and screen content,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-AD0134. 2023.

List of Publications

Peer Review Journal Papers

1. 木谷佳隆，海野恭平，河村圭，“H.266 | VVC 対応 4K/8K リアルタイムコーデックの開発,” 映像情報メディア学会和文論文誌, vol.78, no.1, pp.115-123, Jan. 2024

2. ○Y. Kidani, K. Unno, K. Kawamura, and H. Watanabe, “Memory Bandwidth Constrained Overlapped Block Motion Compensation for Video Coding,” *The Institute of Image Information and Television Engineers Transactions on Media Technology and Applications (ITE-MTA)*, vol.11, no.1, pp.1-12, 2023.

3. 鶴崎裕貴, 木谷佳隆, 柴田達雄, “ケーブルテレビの実環境に基づく放送・通信映像伝送動的切換システムの実現可能性検証,” 映像情報メディア学会和文論文誌, vol.76, no.6, pp.747-756, Nov. 2022.

4. ○Y. Kidani, H. Kato, K. Kawamura, and H. Watanabe, “Geometric Partitioning Mode with Inter and Intra Prediction for Beyond Versatile Video Coding,” IEICE Transactions on Information and Systems, vol.E105.D, no.10, pp.1691-1703, Jun. 2022.

5. 鶴崎裕貴, 木谷佳隆, 海野恭平, 河村圭, “VVC におけるマージモードの差分動きベクトル探索の高速化,” 電子情報通信学会論文誌 D vol. J105-D, no.3, pp.245-253, Mar. 2022.

6. 木谷佳隆, 河村圭, “8K マルチアングル対応リアルタイムエンコーダの開発,” 映像情報メディア学会誌, vol.74, no.4, pp.715-718, Dec. 2020.

Peer Review International Conference Papers

7. ○Y. Kidani, H. Kato, and K. Kawamura, “Bi-predictive Intra Block Copy for Enhanced Video Coding Beyond VVC,” in *2024 IEEE IEEE International Conference on Image Processing (ICIP)*, Oct. 2024. [Accepted on June 2024].

8. H. Kato, Y. Kidani, K. Kawamura, and S. Naito, “Extended Intra Block Copy with Adaptive Filtering and Overlapped Block Averaging,” in *2023 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1-5, Dec. 2023.

9. H. Kato, Y. Kidani, K. Kawamura, and S. Naito, “Adaptive Boundary Width of Geometric Partitioning Mode for Beyond Versatile Video Coding,” in *2022 IEEE Visual Communications and Image Processing (VCIP)*, pp.1-5, Dec. 2022.

10. Y. Kidani, H. Tsurusaki, K. Unno, and K. Kawamura, “Fast Decision Method for Adaptive Motion Vector Resolution in Versatile Video Coding,” in *International Workshop on Advanced Image Technology (IWAIT) 2022*, Jan. 2022.

11. Y. Kidani, H. Tsurusaki, K. Unno, and K. Kawamura, “Fast Decision Method for Merge with Motion Vector Difference in Versatile Video Coding,” in *International Workshop on Advanced Image Technology (IWAIT) 2022*, Jan. 2022.

12. Y. Kidani, H. Yamashita, and S. Matsumoto, “Proposal of RF/IP Adaptive Video Distribution Scheme over Cable Television Access Networks,” in *2020 Society of Cable Telecommunications Engineers (SCTE) - International Society of Broadband Experts (ISBE) Cable-Tec Expo*, Oct. 2020.

13. ○Y. Kidani, K. Unno, K. Kawamura, and S. Naito, “Block-Size Dependent Overlapped Block Motion Compensation,” in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 1191-1195, Oct. 2020.

14. ○Y. Kidani, K. Kawamura, K. Unno, and S. Naito, “Blocksize-QP Dependent Intra Interpolation Filters,” in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4125-4129, Sep. 2019.

Domestic Conference Papers

15. 木谷佳隆，加藤晴久，河村圭, “Beyond VVC 向けのイントラブロックコピー拡張方式に関する一検討,” 2023 年度画像符号化シンポジウム（PCSJ2023）, 2023 年 11 月

16. 木谷佳隆, 中條健, “招待講演: VVC の性能を超える映像符号化の技術探索活動,” 2022 年度画像符号化シンポジウム（PCSJ2022）, 2022 年 11 月

17. 木谷佳隆, 海野恭平, 河村圭, “4K ライブ映像伝送のための VVC 符号化特性評価,” 映像情報メディア学会 2021 年度冬季大会, 2021 年 12 月

18. 木谷佳隆, 加藤晴久, 河村圭, 渡辺裕, “VVC 拡張方式向けの幾何学分割モードに関する一検討,” 2021 年度画像符号化シンポジウム（PCSJ2021）, 2021 年 11 月

19. 齋藤雄太，木谷佳隆，海野恭平，河村圭, “VVC における幾何学的分割モードの符号化処理量削減に関する一検討,” 第 21 回情報科学技術フォーラム（FIT2022）, 2021 年 9 月

20. 木谷佳隆, 海野恭平, 河村圭, “VVC における LFSNT 制御の一検討,” 第 20 回情報科学技術フォーラム（FIT2021）, 2021 年 8 月

21. 木谷佳隆, 河村圭, 海野恭平, “オーバーラップブロック動き補償の VVC への適用検討,” 映像情報メディア学会 2020 年度冬季大会, 2020 年 12 月

22. 木谷佳隆, 河村圭, “VVC 対応の 4K リアルタイムソフトウェアデコーダの試作開発,” 映像情報メディア学会メディア工学研究会, 2020 年 9 月

23. 木谷佳隆, 河村圭, 内藤整, “次世代動画像符号化向けサイズ・量子化パラメータ依存型イントラ補間フィルタの検討,” 2018 年度画像符号化シンポジウム（PCSJ2018）, 2018 年 11 月

24. 木谷佳隆, 河村圭, 内藤整, “次世代動画像符号化方式向けモード・サイズ依存型係数走査方式の検討,” 電子情報通信学会画像工学研究会, 2018 年 5 月

25. 木谷佳隆, 河村圭, 内藤整, “次世代動画像符号化方式向けイントラ参照画素強フィルタの主観評価, 映像情報メディア学会 2017 年度冬季大会, 2017 年 12 月

Awards

26. 公益財団法人通信文化協会 第 69 回前島密賞, “H.266 | Versatile Video Coding の国際標準化と実用化への多大な貢献,” 2024 年 2 月

27. 一般社団法人映像情報メディア学会 第 50 回技術振興賞進歩開発賞（研究開発部門）, “4K/8K 解像度 H.266 | VVC 対応リアルタイムコーデックの開発,” 2023 年 5 月

28. 映像情報メディア学会 第 62 回丹羽高柳賞業績賞, “H.266 | Versatile Video Coding (VVC) の国際標準化と実用化への貢献,” 2023 年 5 月

29. 一般社団法人映像情報メディア学会 第 62 回丹羽高柳賞論文賞, “ケーブルテレビの実環境に基づく放送・通信映像伝送動的切換システムの実現可能性検証,” 2023 年 5 月

30. International Workshop on Advanced Image Technology 2022 (IWAIT 2022) Best Paper Award, Jan. 2022

31. 一般社団法人映像情報メディア学会 2021 年度冬季大会第 55 回鈴木記念奨励賞, 2022 年 8 月

32. 一般社団法人映像情報メディア学会 2021 年度冬季大会学生優秀発表賞, 2022 年 8 月

33. 一般社団法人電子情報通信学会 画像符号化シンポジウム (PCSJ2021) 学生論文賞, 2021 年 12 月

34. 一般社団法人電子情報通信学会 第 20 回情報科学技術フォーラム（FIT2021）FIT 奨励賞, 2021 年 8 月

35. 一般社団法人映像情報メディア学会メディア工学研究会 優秀発表賞, 2020 年 9 月

36. 一般社団法人映像情報メディア学会 映像情報メディア未来賞次世代テレビ技術賞, 2019 年 5 月