Engineering Journal issn: 0125-8281

Review

Reducing Complexity on Coding Unit Partitioning in Video Coding: A Review

Ei Ei Tun and Supavadee Aramvith*

Multimedia Data Analytics and Processing Research Unit, Department of Electrical Engineering, Chulalongkorn University, Bangkok 10330, Thailand
*E-mail: supavadee.a@chula.ac.th (Corresponding author)

Abstract. In this article, we present a survey on the low complexity video coding on a coding unit (CU) partitioning with the aim for researchers to understand the foundation of video coding and fast CU partition algorithms. Firstly, we introduce video coding technologies by explaining the trending standards and reference models. They are High Efficiency Video Coding (HEVC), Joint Exploration Test Model (JEM), and VVC, which introduce novel quadtree (QT), quadtree plus binary tree (QTBT), quadtree plus multi-type tree (QTMT) block partitioning with expensive computation complexity, respectively. Secondly, we present a comprehensive explanation of the time-consuming CU partitioning, especially for researchers who are not familiar with CU partitioning. The newer the video coding standard, the more flexible partition structures and the higher the computational complexity. Then, we provide a deep and comprehensive survey of recent and state-of-the-art researches. Finally, we include a discussion section about the advantages and disadvantage of heuristic based and learning based approaches for the readers to explore quickly the performance of the existing algorithms and their limitations. To our knowledge, it is the first comprehensive survey to provide sufficient information about fast CU partitioning on HEVC, JEM, and VVC.

Keywords: Video coding, complexity reduction, coding unit partitioning, High Efficiency Video Coding (HEVC), Joint Exploration Test Model (JEM), Versatile Video Coding (VVC).

ENGINEERING JOURNAL Volume 26 Issue 11
Received 6 July 2022
Accepted 21 November 2022
Published 30 November 2022
Online at https://engj.org/
DOI:10.4186/ej.2022.26.11.43

DOI:10.4186/ej.2022.26.11.43

1. Introduction

The key purpose of video coding is to reduce the required amount of bits under a comparable visual quality. In literature, researchers from the video coding area follow one of two research directions: coding efficiency and computational complexity. To achieve better coding efficiency, the ongoing video coding standard introduces innovative and complicated coding tools, causing an extreme computational complexity over its predecessor. The better the coding efficiency, the higher the computational complexity. Consequently, it increases the power consumption and hardware cost, introducing difficulties in transporting HD/UHD, 3D, and VR videos in real-time video applications. During the past decade, several research ideas have contributed to reducing the computational complexity of video coding standards. However, there are still various research holes in complexity reduction direction to get a better trade-off between coding efficiency and computational complexity.

This article aims to provide a comprehensive review of complexity reduction on time-consuming CU partitioning for High Efficiency Video Coding (HEVC) [1], Joint Exploration Test Model (JEM) [2], and Versatile Video Coding (VVC) [3] since they have introduced expensive CU tree partition called quadtree (QT), quadtree plus binary tree (QTBT), quadtree plus multi-type tree (QTMT) block partitioning, respectively. Firstly, we introduce necessary information about HEVC, JEM, and VVC in Section 2. In Section 3, we conduct a comprehensive and beneficial explanation about three time-consuming and complicated CU partitioning especially for new video coding researchers. Consequently, these researchers can understand the whole CU partition process of QT, QTBT, and QTMT within a short period. We then mainly categorize the existing fast encoding research, especially for CU partitioning of HEVC, JEM, and VVC, into heuristic and learning-based approaches in Section 4. In the end, we conclude in Section 5.

2. Video Coding Technologies

ISO/IEC and ITU-T are international standard development organizations that have standardized several video compression standards for several years. Motion Picture Expert Group (MPEG) of ISO/IEC and Video Coding Expert Group (VCEG) of ITU-T are mainly involved in contributing to video coding standardization and innovative coding tools. H.261 [4], MPEG-4 [5], H.264/Advanced Video Coding (AVC) [6], HEVC, and VVC standards are the superior standards within the previous three decades.

2.1. HEVC

In 2013, MPEG and VCEG launched a new team, called Joint Collaborative Team on Video Coding (JCT-VC), to develop H.265/HEVC for Ultra High Definition (UHD) video applications since H.264/AVC is only applicable up to Standard Definition (SD)/High Definition (HD) video applications. Compared to H.254, HEVC can achieve a 50% bitrate reduction under the same video quality because of its innovative and time-consuming coding tools especially a flexible quadtree-based block partition. Consequently, the computational complexity of HEVC is significantly higher than that of its predecessor H.264/AVC.

2.2. JEM

Beyond HEVC, the research specialists from MPEG and VCEG were combined as Joint Video Exploration Team (JVET) in 2015 [7]. JVET contributed several advanced coding tools in the Joint Exploration Model (JEM) software to display the potential advantages of launching a future video coding standard called Versatile Video Coding (VVC). Among these coding tools, quadtree plus binary tree (QTBT) block partitioning is one of the leading tools. Even though the coding efficiency of JEM can achieve up to 40% over HEVC, its computational complexity may interfere with the installation of VVC on real-time video applications.

2.3. VVC

In 2018, JVET developed VVC targeting further quality improvement on high-resolution videos such as ultra-high definition (UHD) and virtual reality (VR) since HEVC cannot meet the growing demand in the future video market. To achieve a significant coding efficiency over HEVC, VVC adopted several innovative and time-consuming coding tools under a remarkable computational complexity. Among these advanced coding techniques, quadtree plus multi-type tree (QTMT) block partitioning has obtained impressive attention because of its outstanding coding efficiency gain.

3. Time-Consuming CU Partitioning

3.1. HEVC

As mentioned above, HEVC introduces the quadtree-based block partitioning with a new coding unit (CU) as one of the most significant inputs since its antecedent utilized only 16×16 macroblocks as the largest basic processing unit for Intra/inter prediction and transform coding. Since block partitioning is the first step in the video encoding process, the HEVC encoder initially splits each picture of a raw video sequence into small square blocks called CTU depending on the resolution of the raw video sequence. The size of a CTU can be 64×64 which is the default largest CTU size, 32×32, and 16×16 to get the best trade-off between coding efficiency and computational

DOI:10.4186/ej.2022.26.11.43

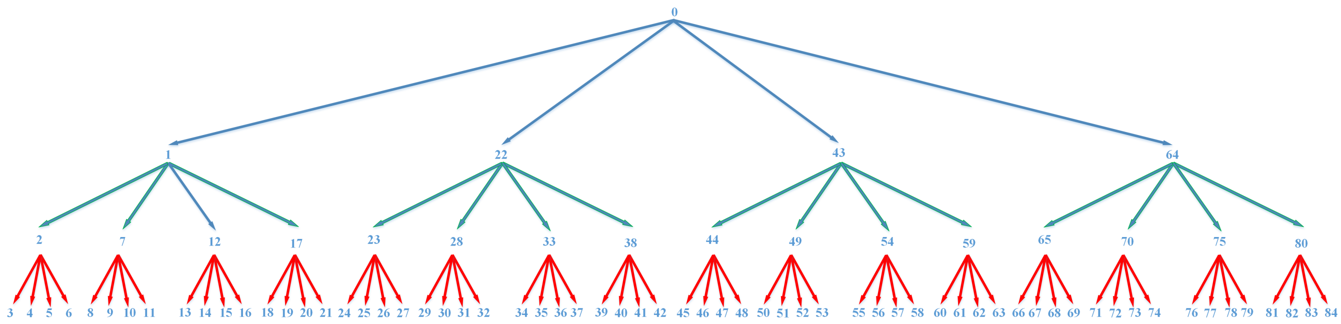

Fig. 1. The preorder quadtree traversal of RD cost computation for 85 CUs of a 64 × 64 CTU.

complexity. Please note that the largest CTU size can be extended more than 64 × 64, for example, 128 × 128. Then, each CTU can be recursively split into several smaller square CUs or can be a whole CU. For 64 × 64 CTU, the size of CU can be 64 × 64, 32 × 32, 16 × 16, and 8 × 8 which is the default smallest CU size. To examine the best quadtree partition structure for each CTU, HEVC introduces a preliminary encoding to check possible combinations of CU size, intra/inter prediction unit (PU) modes, and transform unit (TU) sizes. In preliminary encoding, there are two main processes: root-leaf traversal, which splits and calculates rough rate distortion (RD) cost, and leaf-root pruning, which compares the rough RD cost of the parent CU and its four sub-CUs and decides whether the parent CU split into four sub-CUs or not.

In root-leaf traversal as shown in Fig. 1, the rough RD cost for the CTU, called root CU, at position 0 (depth 0) is firstly calculated and split into four sub-CUs. Then, the rough RD cost for its first sub-CU at position 1 (depth 1) is calculated and divided into four sub-CUs. Then, the rough RD cost for the first sub-CU at position 2 (depth 2) is calculated and split into four sub-CUs called leaf CUs. After splitting into four leaf CUs, the rough RD costs for four leaf CUs at positions 3, 4, 5, and 6 (depth 3) are sequentially computed, and there is no further splitting from these leaf CUs.

Once the rough RD costs for four sub-CUs of each parent CU are calculated, the rough RD cost of the parent CU and the total rough RD cost for its sub-CUs are compared and there is no further splitting from the parent CU if the parent one is less than or equal the total one. This process is called leaf-root pruning. In detail, the rough RD cost of the parent CU at position 2 is less than or equal to the sum of four sub-CUs at positions 3, 4, 5, and 6, its four sub-CUs need to be pruned, and otherwise need to be split. Once leaf-root pruning for each parent CU finishes, it can be root-leaf traversal or leaf-root pruning based on the position of the parent CU. If the current parent CU is at position 2, it is required to do root-leaf traversal for the CU at position 7. If the current parent CU is at position 17, it is necessary to do leaf-root pruning again for the CU at position 1, which is the parent CU of four sub-CUs at position 2, 7, 12, and 17. After comparing a root CU at depth 0 with its four sub-CUs at depth 1, the best CU partition of a CTU with the smallest rough RD cost is one of 83,522 possible partitions. Fig. 2 shows one possible QT partition of HEVC.

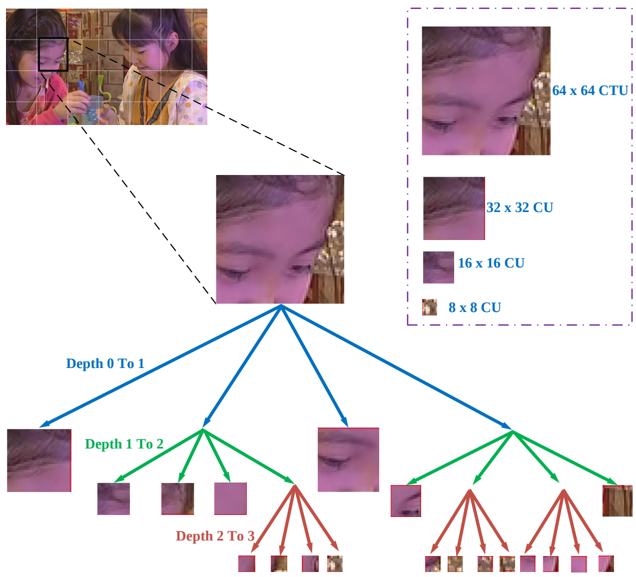

Fig. 2. Quadtree partitioning from CTU to CU.

The number of CUs for computing rough RD costs and comparisons depends on the size of the CTU and minimum CU size or maximum CU depth. In a 64 × 64 CTU with an 8 × 8 minimum CU size (maximum CU depth = 3), there is one root CUs and 84 sub-CUs to find time-consuming rough RD cost and 21 comparisons to find the best CU partition among 83,522 possible partitions. If CTU is 64 × 64 and the minimum CU size is 8 × 8 (maximum CU depth = 2), there are one root CUs and 20 sub-CUs for root-leaf traversal and 5 comparisons to find the best CU partition among 17 possible partitions as shown in Fig. 3. The total number of possible CU partitions $NoOfPart$ is calculated as in (1),

$$NoOfPart = (2^4 + 1)^{(D-1)^2} + (D\,mod\,2) \quad (1)$$

Where, maximum CU depth D can be 1, 2, 3 to represent minimum CU size 32×32 (depth 1), 16×16 (depth 2), and 8×8 (depth 3), respectively, and mod is the math function to find the remainder.

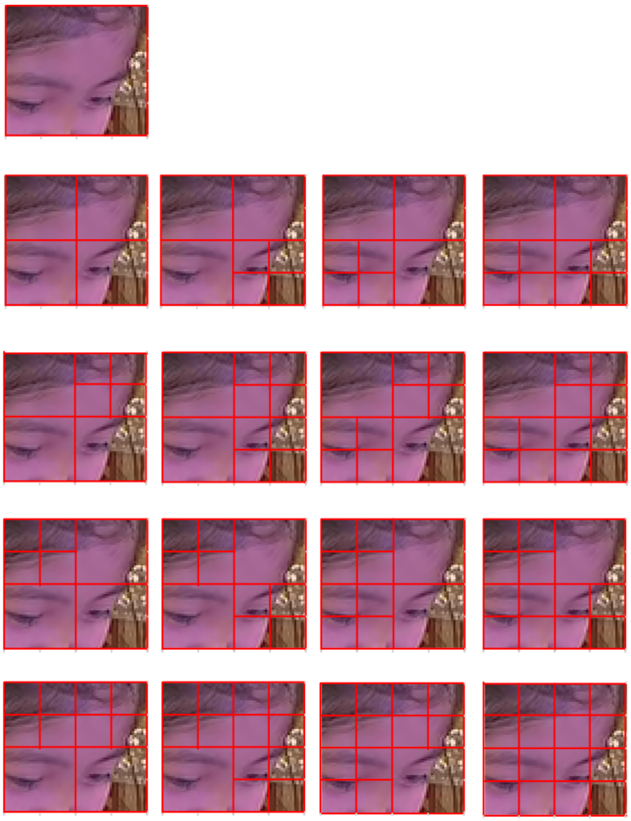

Fig. 3. Possible CU partitions of a 64 × 64 CTU with 16×16 minimum CU size (Maximum CU Depth = 2).

Since the quadtree-based CU partition in HEVC depends on the number of sub-CUs which need to spend the whole inter or intra mode prediction, it has been a research highlight to explore how to efficiently bring down the computational complexity of CU partition under a comparable video quality.

3.2. JEM

Even though HEVC outperforms its predecessor H.264/AVC due to its innovative quadtree-based block partition tools, there are still some crucial barriers mainly the shape of CU is the only square shape which may lead to a significant drawback to getting further possible developments on the coding flexibility. To fix the above main barrier of HEVC block partition, JVET has introduced a crucial quadtree plus binary tree-based block partition (QTBT) during the recent JVET development and released as JEM. As a consequence, the shape of CU can be a square or rectangle and CU itself can be used for prediction and transformation instead of PU and TU.

As shown in Fig. 4, an example of a QTBT partition, a 128×128 CTU (depth 0)is firstly split into four square sub-CUs (depth 1) partitioned by quadtree, similar to the quadtree block partition in HEVC, to get the smaller square CUs. Then, each 64×64 square CU can be split in three ways by quadtree or vertical or horizontal binary tree block partition to get four 32×32 square or two 64×32 or 32×64 rectangle sub-CUs (depth 2), respectively. It should be noted that once a binary tree splits a CU, it could no longer be divided by a quadtree. Each square sub-CU of quadtree can be split in three ways and binary tree sub-CUS only allow vertical binary tree (VBT) and horizontal binary tree (HBT). In QTBT, the minimum QT size and maximum BT size are the essential parameters for limiting the QT depth and BT depth.

To find a CU partition for reaching the best RD performance, JEM recursively checks all possible partition depths, called full Rate-Distortion Optimization (RDO) search. For each CU, JEM calculates an RD cost of the entire CU, then RD costs of 4 sub-CUs and 4 sub-CUs for QT and BT partition, respectively, are sequentially calculated. For example, in each 64×64 square CU, the number of sub-CUs for RDO search of JEM is $1 + 4 + 2 + 2 = 9$ while that of HEVC only has $1 + 4 = 5$. Similar to root-leaf traversal and leaf-root pruning of HEVC, JEM checks the RDO cost for all possible CU sizes and compares all parent CU and their sub-CUs. Then, JEM selects the best QTBT partition that minimizes the RD cost among several block partition combinations, bringing the impractical computational complexity for real-time video services.

3.3. VVC

In VVC, QTMT is the most effective and recently introduced coding tool. Different from the previous standard HEVC, which only provides quadtree and square block sizes, there are three partition trees: quadtree, binary, and ternary. Consequently, VVC can provide both square and rectangle CU sizes to achieve coding flexibility. In the QTMT block partition, CU can be split into five ways as shown in Fig. 6. Each 64×64 CU can be divided into four 32×32 CUs, two 64×32 CUs, two 32×64 CUs, one 64×32 CU and two 64×16 CUs, or one 32×64 CU and two 16×64 CUs if the partition tree is QT, VBT, HBT, VTT, or HTT, respectively. As a result, in 64×64 square CU at depth 1, the number of sub-CUs for RDO search of VVC is $1 + 4 + 2 + 2 + 3 + 3 = 15$ while that of JEM and HEVC only have $1 + 4 + 2 + 2 = 9$ and $1 + 4 = 5$, respectively. Like HEVC and JEM, VVC recursively does the exact RDO search to find the best CU partition under the best RD gain. Due to the several combinations of partition structure and complicated partition structure of QTMT shown in Fig. 5, the increased encoding complexity may block its practical installations in real-time application scenarios.

DOI:10.4186/ej.2022.26.11.43

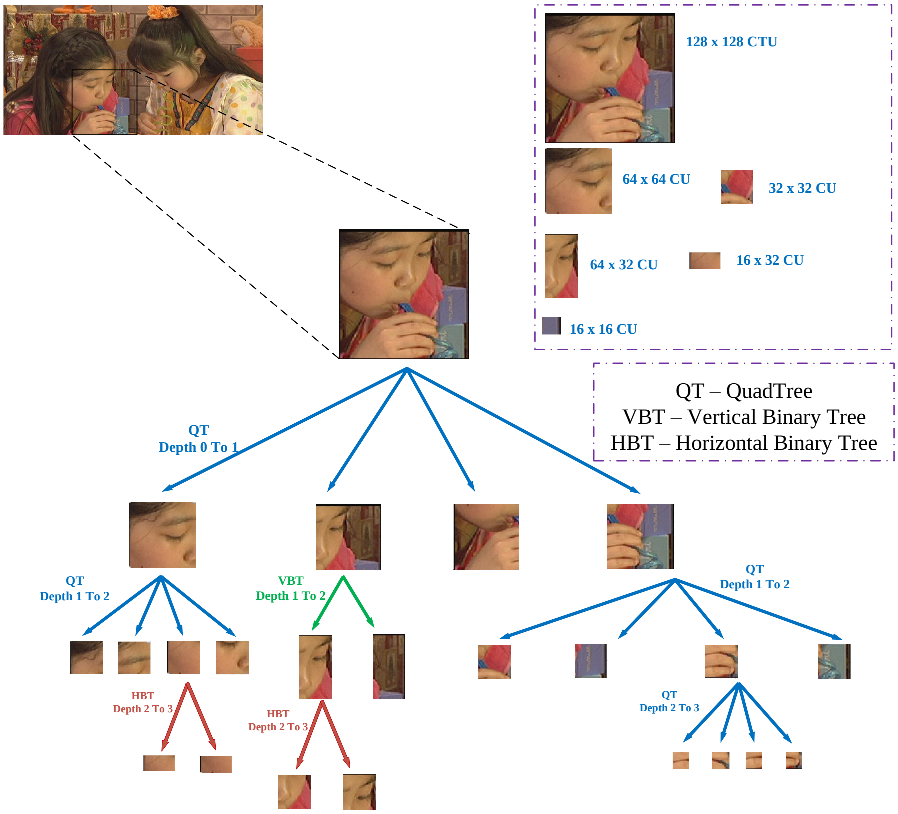

Fig. 4. Quadtree plus binary tree partition from CTU to CU in JEM.

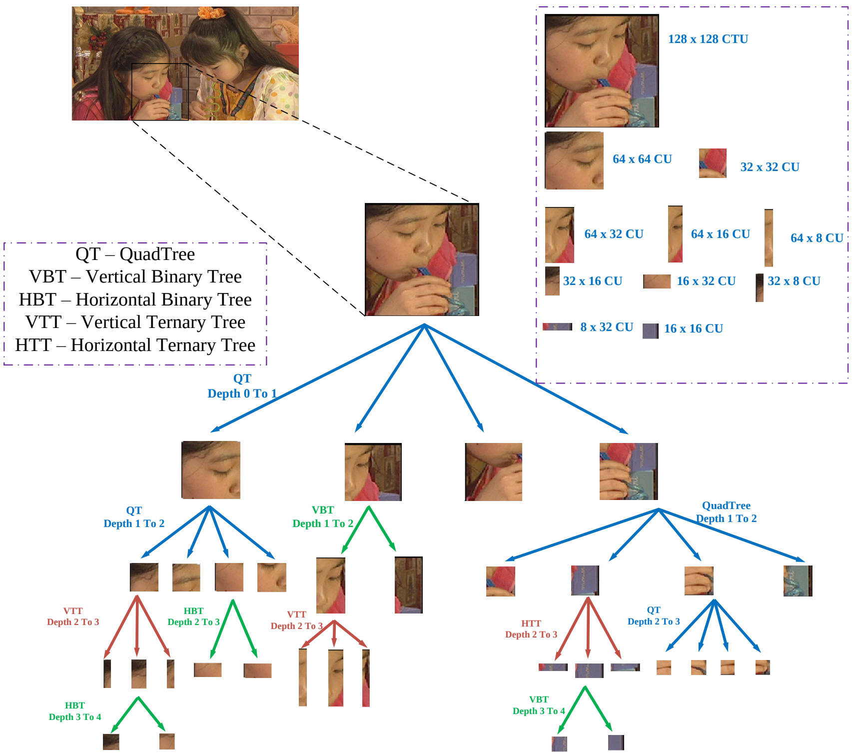

Fig. 5. QTMT partitioning from CTU to CU.

DOI:10.4186/ej.2022.26.11.43

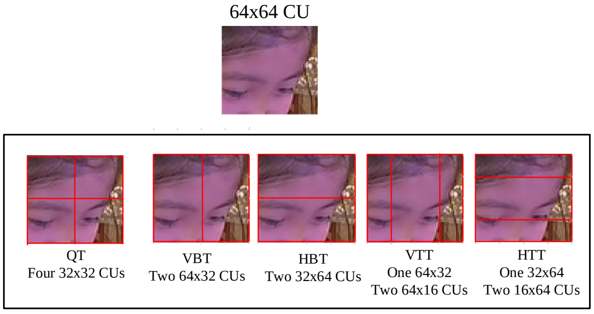

Fig. 6. Five partition structures of QTMT of 64 × 64 CU.

4. Fast CTU Partitioning Strategy

To adequately address the CTU depth decision problem in video coding, the existing works on Fast CTU Partitioning can be divided into two categories: heuristic-based and machine learning and deep learning (ML/DL) based approaches.

4.1. HEVC

4.1.1. Heuristic Approach

The primary strategy of heuristic approaches is to explore important information such as intermediate features that can early and roughly check the block partition without fully running the time-consuming RDO search for all possible QT block partitions. In [8], the frame-level or coding level-based CU size decision algorithm skipped uncommon CU sizes by exploring which CU sizes were not frequently utilized in the previous frames and referring to CU information in the neighbor and co-located CUs, respectively. According to the experiment results, this algorithm is more suitable for 720p video sequences. For the inter coding of HEVC, Xiong *et al.* [9] efficiently utilized the optical flow of the low-resolution frame (downscale 4x) to find a pyramid motion divergence that is very reasonable to decide CU splitting. For the intra-coding of HEVC, [10] introduced Bayes decision rules based fast CU splitting and pruning approach. To early determine the splitting and pruning of the current CU, the raw RD cost and time-consuming full RD cost are effectively utilized to skip full RD cost of the current CU and terminate the encoding process for its sub-CUs, respectively. As the best CU size is particularly content-dependent, the authors of [11] built a fast CU size decision approach by determining CU depth range and bypassing some particular depth levels which are not common depth levels of the adjacent and prior image frames. Also, they introduced motion homogeneity in order to early skip motion estimation for unimportant CU sizes. The simulation results show that the proposed approach can save more than 50% computation complexity for test sequences including slow moving objects such as Traffic and Vidyo1. Like the previous approach, [12] reduced the number of possible CU sizes which are needed to be processed in each treeblock by early deciding with texture homogeneity based adaptive threshold. Additionally, they skipped intra prediction part for large CU size jointly based on coding details and texture property of the adjacent CUs. According to the experimental results, it is especially leading for high resolution test sequences such as Class A and B. Also, the proposed system is effective for all video types with a significant performance among the latest intra codng approaches. To further reduce the computational complexity, Kim *et al.* [13] extended their previous keypoint-based CU size decision work by introducing a QP-based adaptive contrast threshold to rapidly detect keypoint after empirically observing the correlation between contrast threshold and QP. Since the rich texture area has a high potential to split into smaller CUs and go to higher CU depth, and several keypoins can be seen, their research can reasonably decide CU size by comparing the number of keypoints and split threshold. The authors of [14] found that the employment of temporal correlation at co-located CU of the previous frame can increase the compression gain bringing a more time for extracting temporal correlation than spatial correlation at adjacent CUs. Therefore, they explored the CU depth range by only utilizing spatial information. In [15], a fast intra encoding scheme was proposed to decide CU size by skipping uncommon CU sizes of the spatially neighboring CUs and selecting the common mode of the spatially neighboring CUs and parent CU as the best intra mode. In [16], the splitting and termination decisions was developed to remove the unnecessary calculations in large CUs and hinder the encoder from running time-consuming RDO search in small CUs, respectively, based on three efficient decision condition on the co-located CU and adjacent CUs. This algorithm can achieve more than 30% time saving especially for test video sequences including large smooth texture areas such as BasketballDrive and Kimono1. After conducting experiments for several selected test sequences with different QP values, the authors of [17] proposed a statistical-based CU partitioning to early stop the further partitioning of the current CU into sub-CUs by comparing the RD cost of the current CU with pre-defined threshold. In [18], we firstly represented the CU partitioning problem as a heuristic search and solved it with a simple optimizer, called Genetic Algorithm (GA), to quickly search for the optimal CU partitioning structure of each CTU. A reasonable chromosome and rough RD cost-based fitness function boosted the capability of GA. To additionally reduce the encoding time of HEVC inter coding, the temporal correlation was reasonably considered in our approach

4.1.2. ML/DL Approach

Most recently, most researchers have been interested in advanced and trending technologies such as machine learning (ML) and deep learning (DL) to employ in their research works. According to the learning ability of the ML/DL approach, these research works utilized it to derive from the extensive and complicated data into the best solution. Due to the nature of QT partitioning of HEVC, it can be designed as a binary classification work and solved by a logistic regression-based binary classifier in [19]. The selection of F-score-based feature sets was efficiently extracted for each CU size and quantization parameter (QPs). Particularly, the time saving of this algorithm is exceeding 60% for large resolution test sequences such as 720p, 1080p, and 1600p. The authors of [20] utilized a support vector machine (SVM) with image complexity based on three essential features for block partitioning such as global, directional, and local complexities to classify plain CU (Class 1), complex CU (Class 2), or uncertain decision (Class 3). Consequently, the computational complexity of HEVC intra coding can be significantly saved by early terminating the CU partition, skipping intra mode decision, or doing the same process of HM for classes 1, 2, or 3, respectively. This approach impressively achieve computational complexity reduction about 60% for low activity sequences such as Jonnny, KristenAndSara and Kimono. To early stop an expensive RDO search for the combination of CUs, PUs, and TUs, data mining based three decision trees with different attribute size was built in [21]. After analyzing QT block partitioning of HEVC, Zhang *et al.*[22] converted it into a hierarchical-based three-level binary classification model. To reduce uncertain decisions, a joint SVM classifier, including two binary classifiers and the voting module, is simply utilized to output three decisions such as split, non-split, and uncertain. In addition to binary classification for CU partition, [23] introduced binary plus multi-class classification based on SVM for PU mode classification.

The performance of the mentioned ML-based fast encoding algorithm depends on the manual or hand-crafted feature set while consuming additional processing time for feature selection. Therefore, in place of using hand-crafted features, it is desirable to intelligently and automatically extract QT partition-related features by DL to save the QT partition's encoding time further. For the VLSI-friendly purpose, the convolution neural network (CNN) architectures based on fast CU size and PU mode decision were developed for HEVC intra coding [24]. In 2018, the authors of [25] contributed DL on CU partitioning by introducing an extensive CU size database of intra and inter prediction to empower the DL approach with a significant time saving on CU partitioning.

4.2. JEM

4.2.1. Heuristic Approach

In [26], the authors defined the particular constraints. They bypassed binary tree(BT) at the second sub-CU of the current CU if the calculated rate-distortion costs of the current CU and its first sub-CU satisfy the pre-defined particular constraints. In [27], the partition and intra mode decision of formerly coded CU was reused if the same CU in the other partition selections has the same adjacent coded CU, globally depending on the heuristic information of the former coded CU. The QT plus binary tree partitioning was performed in [28] by adapting the maximum BT depth values of each image frame based on its temporal level. The authors of [29] presented a local constraint-based QT plus binary tree partitioning by dynamically getting the required parameters, which is vital for each CTU partitioning, from the last decoded frame without introducing further overhead. In 2018, they introduced a confidence interval for early terminating the QTBT splitting process with the usage of motion divergence field (MDF) in [30] to get a superior trade-off. Since the rich region including motion activities owns a hugh MDF energy, the time saving are remarkable on PeopleOnStreet, BasketballDrill, PartyScene and RaceHorses test sequences which all are including high movement activities.

After conducting several experiments with different splitting parameters to empirically analyze the impact of splitting parameters, the required splitting parameters of each CTU, such as minimum QT size, maximum BT size, and depth, were chosen and dynamically calculated based on the partitioning information of temporal co-located and spatial neighboring CTUs in [31]. The time saving of this algorithm is significant for test sequences including high activity such as PeopleOnStreet, RaceHorses, and BasketballPass like [30] and also obvious for low activity video sequences, e.g. BQSquare and KristenAndSara.

4.2.2. ML/DL Approach

As mentioned above, the authors of [31] dynamically derived the splitting parameters of QTBT to adapt the local characteristics without introducing additional overhead at the CTU level. They designed a joint classifier-based decision tree approach for the CU level with information gain attribute evaluation (IGAE) to remove unnecessary RDO looping while controlling the misclassification rate. In [32], Random Forest (RF) based QTBT classification algorithm was designed to estimate whether QT splitting or BT splitting is suitable by utilizing off-line trained RF classifiers for each CU size. To reduce the false prediction risk, an uncertain region was defined and all possible modes for QT and BT splitting need to carry out in that region.

To reduce an enormous computation complexity of JVET QTBT block splitting, Jin *et al.* [33] firstly introduced an end-to-end DL model. They formulated QTBT splitting depth range (five classes) for each 32×32 CU as a multi-class classification since the earlier classification schemes classified splitting or non-splitting for each CU depth level. Due to their DL model's exact depth range classification, some RDO searches of the unnecessary depth can be early stopped for 32×32 CU when the depth level is not within five possible classes. In [34], the authors firstly did a statistical observation on the QTBT splitting to help implement the CNN architecture. Then, a novel CNN-based fast CU size algorithm for inter prediction of JVET was first developed. Since duplicated CU splitting structures may exist in each CU depth due to the splitting nature of QTBT, the above-mentioned DL-based approaches have some gaps in time saving performance matrix. Therefore, [35] measured intra 4×4 CU partition probabilities based on its content and partiality substituted the expensive RDO search by considering three partition types: QT and BT plus asymmetric binary tree (ABT).

4.3. VVC

4.3.1. Heuristic Approach

To reduce the most time-consuming QTMT part of VVC, a novel Bayesian decision rule-based CU splitting scheme was firstly presented in [36]. Since the horizontal binary tree (HBT) for each CU is firstly checked among the other new partitions such as vertical binary tree (VBT), horizontal ternary tree (HTT), and vertical ternary tree (VTT) after checking QT, the encoded information about HBT was extensively investigated to early stop the other partitions. It should be note that the overall algorithm achieves a consistent performance for ultra high resolution (UHD) test sequences such as Class A1 and A2. Yang *et al.* [37] proposed a novel fast intra encoding algorithm to decide both QTMT partition and the optimal prediction mode for intra prediction by connecting decision tree classifiers and utilizing gradient descent search with the optimized initial search, search direction, and step size function, respectively.

The authors of [38] accelerated the QT plus multi-type tree partitioning by considering the look-ahead estimation purpose which performs rough RDO search for pre-defined seven intra modes in advance. Based on the experimental results, their algorithm is highly applicable for UHD video applications and also produces a steady coding efficiency for all other test sequences.

In [39], the authors utilized the similarity between neighboring sub regions in both the horizontal and the vertical directions in order to detect an optimal multi-type-tree (MTT) partition.

4.3.2. ML/DL Approach

In [40], ML-based block partitioning was proposed to save the computational complexity of QTBT for JEM and QTMT for VVC by introducing the variable size for risk intervals for adjustable purposes and random forest (RF) classifiers for lightweight purposes. To save the encoding time for the time-consuming QTMT partition and intra mode selection for VVC intra coding, the authors of [41] proposed a fast encoding algorithm by efficiently utilizing RF classifier and texture characteristics, respectively. In [42], a CNN-based adaptive block partitioning for VVC intra encoding was designed by considering variable pooling size for each input CU shape.

To avoid the exhaustive RDO search for QTMT, a DL-based CU decision algorithm for VVC intra encoding was presented in [43] by building an extensive video dataset that includes enough CU splitting structure according to different video content.

In [44], a fast encoding approach for both intra and inter coding is implemented by utilizing Canny edge features in order to early skip partition modes in both directions and detecting a moving object of the current block with frame differentiating in order to early terminate, respectively. In [45], the authors defined a time-consuming CU partitioning as a classification problem and solved by utilizing a novel CNN architecture called multi-level tree CNN (MLT-CNN). In order to improve the classification accuracy, they additionally considered temporal features such as residual frame and picture order count (POC) in the training process. Mainly, MLT-CNN achieves a significant performance for Class A and B. In [46], in order to reduce the encoding time for VVC inter coding, each CTU is firstly split into 8×8 blocks and a CNN-based lightweight network is utilized to predict each block at the RDO stage for skipping QT partition search and uncommon partitions.

Since VVC is the newest coding standard and the CU partitioning of QTMT is very complicated, the solution for fast CU partitioning is very trending in the video coding research area. Therefore, we provide a reasonable and fair comparisons for only VTM version. In order to mention and fairly compare the the efficiency of these fast algorithms in reducing complexity of CU partitioning, we firstly categorize each algorithm by the same test model before starting any comparison and then calculate and present a reasonable comparison especially for a newest video coding standard VVC test model VTM 6 and VTM 11 in Table 1 and Table 2, respectively.

5. Discussion

Since the discussion about the advantages as well as the limitations are benefit for the beginners of video coding area to explore quickly the performance of the the existing algorithms, we generally discuss the advantages and limitation of heuristic and ML/DL approaches. Mainly, there are several heuristic approaches for reducing the computation complexity of CU partitioning. The general process

Table 1. Performance comparison of ML/DL based fast CTU partitioning in VTM 6.

Average on Class	Tang 2019 [44] on VTM 6		Yeo 2021 [45] on VTM 6		Liu 2022 [46] on VTM 6	
	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
A	3.35	30.17	3.85	36.95	2.12	35.88
B	3.78	34.73	3.76	30.93	1.59	31.79
C	1.82	16.20	2.56	25.98	0.39	22.94
D	5.49	30.11	2.35	22.05	0.19	12.34
Average	3.61	27.80	3.13	28.97	1.07	25.74

Table 2. Performance comparison of ML/DL based fast CTU partitioning in VTM 11.

Average on Class	Yeo [45] on VTM 11		Liu [46] on VTM 11	
	BDBR(%)	TS(%)	BDBR(%)	TS(%)
A	1.95	16.99	1.47	26.05
B	0.95	16.19	1.06	23.34
C	0.09	4.05	0.29	17.29
D	0.11	2.34	0.15	8.82
Average	0.78	9.89	0.74	18.88

of these approaches, to firstly extract manual or hand-craft features and to statistically define threshold based on empirical simulations for subjecting each hand-craft features to low computational complexity in the decision making. The main benefits of these heuristic approaches are very simple and straightforward to implement. However, they are not good enough to achieve a huge time saving compared to learning based approaches since the number of features is limited and the thresholds are statistically defined only on a small set of video frames. In contrast, ML/DL approaches design CU partition as classification problem such as hierarchical-based three-level binary classification model in [22] and binary plus multi-class classification in [23]. In learning based approaches, there are several significant advantages such as using joint features and high accuracy. In contrast, there are some limitations such as distinguishable feature selection can be time-consuming and needs to understand the problem domain well. Also, there is additional complexity overhead for features extraction, getting a optimal hyper-plane, and determining the best parameter.

6. Conclusion

This paper comprehensively analyzes fast CU partitioning schemes by categorizing two research groups: heuristic-based and ML/DL-based. Firstly, we give an interesting theory about the CU partitioning structure of QT, QTBT, and QTMT for HEVC, JEM, and VVC. We then categorize the start-of-the-art and existing methods and primarily focus on the fast CU partitioning algorithms for HEVC, JEM, and VVC. Finally, we include a discussion section about the advantages and limitations of heuristic based and DL/ML approaches.

Acknowledgment

We thank Multimedia Data Analytics and Processing Research Unit, Department of Electrical Engineering, Chulalongkorn University for the research support.

References

[1] 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, Dec 2012.

[2] J. Chen, E. Alshina, G. J. Sullivan, J. Ohm, and J. Boyce, "Algorithm description of Joint Exploration Test Model 7 (JEM 7)," *Document JVET-G1001-v1 of JVET, Torino, IT*, July 2017.

[3] B. Bross, J. Chen, and S. Liu, "Versatile video coding (draft 2)," *Document JVET-K1001, Ljubljana, SI*, July 2018.

[4] "Video codec for audiovisual services at px64 kbit/s," *ITU-T Rec. H.261*, version 1: Nov. 1990, version 2: Mar. 1993.

[5] "Coding of audio-visual objects—part 2: Visual," *ISO/IEC 14496-2 (MPEG-4 Visual version 1), ISO/IEC JTC 1*, Apr 1999.

[6] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[7] J. Chen, E. Alshina, and G. J. Sullivan, "Algorithm description of joint exploration test model 1," *JVET-A1001, JVET 1st Meeting, Geneva*, 2015.

[8] J. Leng, L. Sun, T. Ikenaga, and S. Sakaida, "Content based hierarchical fast coding unit decision algorithm for HEVC," in *2011 International Conference on Multimedia and Signal Processing*, vol. 1, 2011, pp. 56–59.

[9] J. Xiong, H. Li, Q. Wu, and F. Meng, "A fast HEVC inter CU selection method based on pyramid motion divergence," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 559–564, 2014.

[10] S. Cho and M. Kim, "Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 9, pp. 1555–1564, 2013.

[11] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective CU size decision method for hevc encoders," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 465–470, 2013.

[12] L. Shen, Z. Zhang, and Z. Liu, "Effective CU size decision for HEVC intracoding," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4232–4241, 2014.

[13] N. Kim, S. Jeon, H. J. Shim, B. Jeon, S.-C. Lim, and H. Ko, "Adaptive keypoint-based CU depth decision for HEVC intra coding," in *2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2016, pp. 1–3.

[14] D.-H. Kim, Y.-H. Kim, and W.-C. Park, "Selective CU depth range decision algorithm for HEVC encoder," in *The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014)*, 2014, pp. 1–2.

[15] L. Shen, Z. Zhang, and P. An, "Fast CU size decision and mode decision algorithm for HEVC intra coding," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 1, pp. 207–213, 2013.

[16] W.-J. Hsu and H.-M. Hang, "Fast coding unit decision algorithm for HEVC," in *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2013, pp. 1–5.

[17] J. Kim, Y. Choe, and Y.-G. Kim, "Fast coding unit size decision algorithm for intra coding in HEVC," in *2013 IEEE International Conference on Consumer Electronics (ICCE)*, 2013, pp. 637–638.

[18] E. E. Tun, S. Aramvith, and Y. Miyanaga, "Fast coding unit encoding scheme for HEVC using genetic algorithm," *IEEE Access*, vol. 7, pp. 68 010–68 021, 2019.

[19] Q. Hu, Z. Shi, X. Zhang, and Z. Gao, "Fast HEVC intra mode decision based on logistic regression classification," in *2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2016, pp. 1–4.

[20] D. Liu, X. Liu, and Y. Li, "Fast CU size decisions for HEVC intra frame coding based on support vector machines," in *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2016, pp. 594–597.

[21] G. Correa, P. A. Assuncao, L. V. Agostini, and L. A. da Silva Cruz, "Fast HEVC encoding decisions using data mining," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 4, pp. 660–673, 2015.

[22] Y. Zhang, S. Kwong, X. Wang, H. Yuan, Z. Pan, and L. Xu, "Machine learning-based coding unit depth decisions for flexible complexity allocation in high efficiency video coding," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2225–2238, 2015.

[23] L. Zhu, Y. Zhang, Z. Pan, R. Wang, S. Kwong, and Z. Peng, "Binary and multi-class learning based low complexity optimization for HEVC encoding," *IEEE Transactions on Broadcasting*, vol. 63, no. 3, pp. 547–561, 2017.

[24] Z. Liu, X. Yu, Y. Gao, S. Chen, X. Ji, and D. Wang, "CU partition mode decision for HEVC hardwired intra encoder using convolution neural network," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5088–5103, 2016.

[25] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan, "Reducing complexity of HEVC: A deep learning approach," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5044–5059, 2018.

[26] P.-H. Lin, C.-L. Lin, and Y. Jen, "AHG5: Enhanced fast algorithm of JVET- E0078," *JVET-F0063, JVET 6th Meeting: Hobart, AUS*, 31 Mar. 2017.

[27] H. Huang, L. Shan, and Y.-W. Huang, "AHG5: Speedup for JEM-3.1," *JVET-D0077, JVET 4th Meeting, Chengdu*, 15–21 October 2016.

[28] Y. Yamamoto and T. Ikai, "AHG5: Fast QTBT encoding configuration," *document JVET-D0095*, Oct. 2016.

[29] Z. Wang, S. Wang, J. Zhang, and S. Ma, "Local-constrained quadtree plus binary tree block partition structure for enhanced video coding," in *2016 Visual Communications and Image Processing (VCIP)*, 2016, pp. 1–4.

[30] Z. Wang, S. Wang, J. Zhang, S. Wang, and S. Ma, "Probabilistic decision based block partitioning for future video coding," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1475–1486, 2018.

[31] ——, "Effective quadtree plus binary tree block partition decision for future video coding," in *2017 Data Compression Conference (DCC)*, 2017, pp. 23–32.

[32] T. Amestoy, A. Mercat, W. Hamidouche, C. Bergeron, and D. Menard, "Random forest oriented fast QTBT frame partitioning," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1837–1841.

[33] Z. Jin, P. An, L. Shen, and C. Yang, "CNN oriented fast QTBT partition algorithm for JVET intra cod-

ing," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, 2017, pp. 1–4.

[34] Z. Wang, S. Wang, X. Zhang, S. Wang, and S. Ma, "Fast QTBT partitioning decision for interframe coding with convolution neural network," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2550–2554.

[35] F. Galpin, F. Racapé, S. Jaiswal, P. Bordes, F. Le Léannec, and E. François, "CNN-based driving of block partitioning for intra slices encoding," in *2019 Data Compression Conference (DCC)*, 2019, pp. 162–171.

[36] T. Fu, H. Zhang, F. Mu, and H. Chen, "Fast CU partitioning algorithm for H.266/VVC intra-frame coding," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, 2019, pp. 55–60.

[37] H. Yang, L. Shen, X. Dong, Q. Ding, P. An, and G. Jiang, "Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, pp. 1668–1682, 2020.

[38] M. Lei, F. Luo, X. Zhang, S. Wang, and S. Ma, "Look-ahead prediction based coding unit size pruning for VVC intra coding," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4120–4124.

[39] S. Zhang, R. Zhang, and X. Jing, "A fast multi-type-tree split decision algorithm of intra coding unit in VVC," in *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2022, pp. 1–5.

[40] T. Amestoy, A. Mercat, W. Hamidouche, D. Menard, and C. Bergeron, "Tunable VVC frame partitioning based on lightweight machine learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 1313–1328, 2020.

[41] Q. Zhang, Y. Wang, L. Huang, and B. Jiang, "Fast CU partition and intra mode decision method for H.266/VVC," *IEEE Access*, vol. 8, pp. 117 539–117 550, 2020.

[42] G. Tang, M. Jing, X. Zeng, and Y. Fan, "Adaptive CU split decision with pooling-variable CNN for VVC intra encoding," in *2019 IEEE Visual Communications and Image Processing (VCIP)*, 2019, pp. 1–4.

[43] T. Li, M. Xu, R. Tang, Y. Chen, and Q. Xing, "DeepQTMT: A deep learning approach for fast QTMT-based cu partition of intra-mode VVC," *IEEE Transactions on Image Processing*, vol. 30, pp. 5377–5390, 2021.

[44] N. Tang, J. Cao, F. Liang, J. Wang, H. Liu, X. Wang, and X. Du, "Fast CTU partition decision algorithm for VVC intra and inter coding," in *2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2019, pp. 361–364.

[45] W.-H. Yeo and B.-G. Kim, "CNN-based fast split mode decision algorithm for versatile video coding (VVC) inter prediction," *Journal of Multimedia Information System*, vol. 8, no. 3, pp. 147–158, 2021.

[46] Y. Liu, M. Abdoli, T. Guionnet, C. Guillemot, and A. Roumy, "Light-weight CNN-based VVC inter partitioning acceleration," in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2022, pp. 1–5.

DOI:10.4186/ej.2022.26.11.43

Ei Ei Tun(S'18-M'20) received the B.C.Tech. and B.C.Tech.(Hons.) degree in computer technology from Computer University (Taungoo), Taungoo, Myanmar, in 2005 and 2006, and the M.C.Tech. (Distinction) degree in computer technology from the University of Computer Studies, Yangon, Myanmar, in 2010, and Ph.D. degree in electrical engineering from the Department of Electrical Engineering, Chulalongkorn University, Thailand in 2019. She was a lecturer at the University of Computer Studies, Yangon, Myanmar. She is currently a postdoctoral researcher with the Department of Electrical Engineering, Chulalongkorn University, Thailand. Her research interests include computer vision, and digital image processing, video coding, especially focusing on the complexity reduction of the newest video coding standards, VVC. She was one of the best paper candidates at the 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST) and the best poster candidates at the Asian Universities Alliance Postgraduate Academic Forum (AUA-PAF 2018), Tsinghua University, Beijing, China and the best presentation candidates at the 14th Regional Conference on Electrical and Electronics Engineering (RC-EEE 2021).

Supavadee Aramvith (S'95–M'01–SM'06) received the B.S. (first class honors) degree in Computer Science from Mahidol University in 1993. She received the M.S. and Ph.D. degrees in Electrical Engineering from the University of Washington, Seattle, USA, in 1996 and 2001, respectively. She joined Chulalongkorn University in June 2001. Currently, she is an Associate Professor at Department of Electrical Engineering with specialization Video Technology. She has successfully advised 9 Ph.D., 27 Master's, and 32 Bachelor's Graduates. She published over 130 papers in International Conference Proceedings and Journals with 4 International Book Chapters. She has rich project management experiences as the project leader and former technical committee chairs to the Thailand Government bodies in Telecommunications and ICT. She is very active in the international arena with the leadership positions in the international network such as JICA Project for AUN/SEED-Net, and the professional organizations such as IEEE, IEICE, APSIPA, and ITU.