

Article

A Trellis Based Temporal Rate Allocation and Virtual Reference Frames for High Efficiency Video Coding

Xiem Hoang Van ^{1,*}, Le Dao Thi Hue ¹ and Thuong Nguyen Canh ²¹ Faculty of Electronics and Telecommunications, University of Engineering and Technology, Vietnam National University, 144 Xuan Thuy Rd, Dich Vong Hau, Cau Giay Dist, Hanoi 10000, Vietnam; huele Dao94@gmail.com² Institute for Dataability Science, Osaka University, 2-8, Yamadaoka, Suita, Osaka 565-0871, Japan; Ngcthuong@gmail.com

* Correspondence: xiemhoang@vnu.edu.vn

Abstract: The High Efficiency Video Coding (HEVC) standard has now become the most popular video coding solution for video conferencing, broadcasting, and streaming. However, its compression performance is still a critical issue for adopting a large number of emerging video applications with higher spatial and temporal resolutions. To advance the current HEVC performance, we propose an efficient temporal rate allocation solution. The proposed method adaptively allocates the compression bitrate for each coded picture in a group of pictures by using a trellis-based dynamic programming approach. To achieve this task, we trained the trellis-based quantization parameter for each frame in a group of pictures considering the temporal layer position. We further improved coding efficiency by incorporating our proposed framework with other inter prediction methods such as a virtual reference frame. Experiments showed around 2% and 5% bitrate savings with our trellis-based rate allocation method with and without a virtual reference frame compared to the conventional HEVC standard, respectively.

Keywords: inter coding; HEVC standard; virtual reference; dynamic programming; adaptive quantization



Citation: Hoang Van, X.; Dao Thi Hue, L.; Nguyen Canh, T. A Trellis Based Temporal Rate Allocation and Virtual Reference Frames for High Efficiency Video Coding. *Electronics* **2021**, *10*, 1384. <https://doi.org/10.3390/electronics10121384>

Academic Editor: Otoniel Mario López Granado

Received: 10 May 2021

Accepted: 7 June 2021

Published: 9 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Context and Motivations

The wide growth of multimedia applications always demands more powerful video transmission over the internet with high compression performance as well as low complexity. Therefore, the Joint Collaborative Team on Video Coding (JCT-VC) announced the High Efficiency Video Coding (HEVC) [1] with 50% bitrate reduction at the same perceptual quality as the previous H.264/AVC standard [2]. Despite its success, there is still a demand for beyond HEVC coding to meet emerging requirements of higher resolutions (i.e., 4K, 8K) and diverse contents (screen contents, drone, etc.). In this context, in the newest video compression standard, H.266/Versatile Video Coding (VVC), many coding tools have been proposed, such as partitioning with quad-tree plus binary, adaptive transforms, new intra modes, affine motion estimation, dependent quantization, etc. [2]. Among all, inter coding tools are always bringing the highest bitrate reduction due to the high redundancy between consecutive temporal frames. Besides the affine motion estimation, virtual reference frame and temporal rate control are active research topics to improve the inter coding performance [3].

Rate control is one of the most powerful encoding tools to advance the rate-distortion (RD) performance of any video codec. In general, at a given quality, rates are allocated from coarse to fine in (i) group of pictures (GOP) or *temporal rate allocation*, then (ii) distributed among coding unit or spatial, or (iii) adaptive allocation to reduce the distortion propagation. Adaptive approaches are complicated and demand computational complexity and memory for either pre-analysis or online analysis. As a result, the common test condition of the most recent video coding standards, including H.266/VVC, still favors a

simple hierarchical rate allocation at the frame level [2]. For each GOP, HEVC and VVC classified frames into a temporal index (Tid), and depending on their hierarchical location and frame types (i.e., intra or inter), the quantized parameter (QP) is adjusted by a fixed value. This framework is very simple and easy to validate the performance of coding tools during the standardization process. However, with increasing competition from AOMedia Video Coding (AV1) [4] with an adaptive GOP, an effective temporal rate allocation for MPEG codecs is demanded.

HEVC adopts a simple temporal rate control via its hierarchical coding structure [1]. HEVC controls bits for each frame in a group of pictures (GOP) with a temporal index associated with its hierarchical reference model by adjusting the quantization parameter for each frame so that high Tid frames are encoded at a higher QP. However, the simple method in HEVC is not effective in preventing the propagation of distortion. Researchers have introduced adaptively allocate bitrates to each frame via a rate-distortion optimized scheme [5–7]. These methods, however, required modification in the decoder, which is non-compliant with the HEVC standard. Additionally, the adaptive quantization parameter at the frame level is an encoder-only tool, easy to modify in the configuration. In addition, adaptive quantization at the frame level would significantly improve the overall compression performance. Therefore, it is often left out from the MPEG (Moving Picture Experts Group) standardization process; thus, there are few works on this research topic.

Virtual reference frame (VRF) is another inter coding improvement approach by (i) generating virtual frames and (ii) utilizing virtual frames as an additional reference or for guided reconstructed frames [8]. Several works utilized the motion estimation or advanced deep learning framework to interpolate frames from decoded frames [9–13]. The deep learning approach shows higher gain but also introduces tremendous complexity in both encoder and decoder [11–13]. On the other hand, virtual frames are utilized as additional reference frames for motion estimation or reducing the syntax transmission overhead as a special merge mode. Although there are many works on VRF, there is lacking investigation on the impact of virtual frames on coding performance.

1.2. Contributions and Paper Organization

Although the virtual referencing-based method has provided important compression improvement for HEVC, there is still room for further improving HEVC performance. Previous works either researched the virtual reference frame or temporal rate control separately. Firstly, the prior VR creation mainly relied on the motion estimation and interpolation-based approach, which may be ineffective for fast motion content and compressed with high QPs or low rate. Secondly, the prior VR frames are used for all B-slices, which may always be effective in terms of the rate-distortion optimization (RDO) manner, especially for pictures at the low temporal layer positions. Finally, since the quality of the VR frames highly depends on the quality of existing decoded references, adaptive quantized frames at different layers would impact the performance of inter coding. In this context, this paper proposes:

- (i) a novel virtual reference frame creation where a multiple hypothesis motion estimation method is used for frame generation;
- (ii) an efficient rate allocation algorithm in which the trellis coding method is used to learn the temporal rate allocation.

We conducted a rich set of experiments with numerous train and test videos and showed around 5% BD-rate gains on top of the most recent HEVC reference software. The rest of this paper is organized as follows. Section 2 briefly describes the background work on HEVC inter coding and the related works on rate control. Section 3 presents the proposed VRF creation and temporal rate allocation, while Section 4 assesses the coding performance of the proposed methods. Finally, we summarize the contributions of this paper in Section 5.

2. Background and Related Works

HEVC adopted the inter-coding tools to take advantage of temporal correlation among frames in a GOP. This paper introduces novel temporal rate allocation and virtual reference frame exploitation for HEVC inter coding; hence, this section will briefly describe the background work on the HEVC inter coding and then analyzes the related works on HEVC rate control.

2.1. HEVC Inter Coding

Compared to the prior inter-coding tools of the H.264/AVC standard [2], the HEVC introduces several improvements in coding tools and coding structure, notably (i) block partitioning and (ii) motion estimation and reference picture management.

Block partition: HEVC inter coding allows compressing picture with more block partition shapes than intra coding, notably the partition modes of $2N \times 2N$, $2N \times N$, and $N \times 2N$ indicate the cases when coding block is not split, split into two equal-size prediction blocks horizontally and vertically, respectively [1]. While $N \times N$ refers to the case, the coding block is split into four equal-size blocks. In addition, there are four asymmetric motion partitions, $2N \times nU$, $2N \times nD$, $nL \times 2N$, and $nR \times 2N$ (see Figure 1). For each coding tree unit (CTU), the rate-distortion optimization (RDO) process is recursively computed to find the optimal coding and prediction unit, notably from 64×64 to 8×8 .

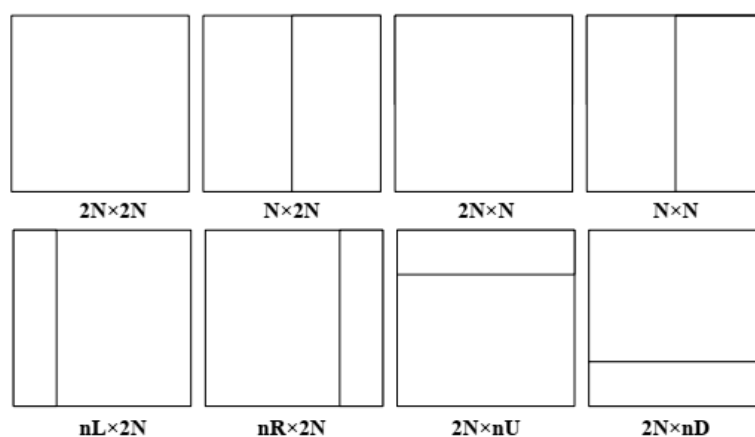


Figure 1. Illustration of block partitions in HEVC inter coding.

Motion estimation and reference picture management: Similar to the prior video coding standard, HEVC inter employs motion estimation to find the best-matched block in reference frames to reduce the redundancy between successive frames. For reference picture management, HEVC inter stores the previously decoded pictures into a decoded picture buffer (DPB). To identify these pictures, a list of picture order count (POC) identifiers is transmitted in each slice header, and the set of reference pictures is called the reference picture set (RPS), as shown in Figure 2.

The HEVC DPB also contains two lists, list 0 and list 1, which are referred through the reference picture index. For biprediction, i.e., in random access coding configuration, two pictures are selected (one from each list).

2.2. HEVC Rate Control

Rate control has been an important research topic in video coding over the years. A temporal domain is an effective way to reduce the transition rate. HEVC uses a fixed hierarchical bit rate allocation. It controls bits for each frame in a group of pictures (GOP) with a temporal index (Tid) associated with its hierarchical reference model, depicted in Figure 2. By simply adjusting the QP parameters for each frame, high Tid frames are encoded at a higher quantization parameter (QP). However, this simple method in HEVC

is unable to prevent distortion propagation. Therefore, various rate control algorithms were developed for video coding standards. For HEVC, an early rate control algorithm was proposed using a pixel-wise unified rate-quantization (R-Q) model [14]. This R-Q model is almost the same as the conventional quadratic R-D model in [15]. Afterward, Li et al., in [16], proposed an R- λ model where λ is the Lagrange Multiplier based on a frame's complexity. This rate control algorithm allocates target bits to a GOP, a frame, or a CTU. Further extending this work, the authors in [17] proposed a gradient-based R- λ model and inter-frame rate control for HEVC. In order to enhance facial detail for video coding conferencing, the work in [18] developed a rate control algorithm based on a weighted R- λ model, which can allocate more target bits to important regions in a video frame. Considering the coding performance of intra coding, a structure similarity (SSIM) based game theory approach was introduced in [19] to optimize the CTU level bit allocation. Similarly, a ρ domain bit allocation and a rate control algorithm were proposed in [20] to optimize bit allocation for key frames. Recently, the authors in [21] presented a highly parallel hardware architecture for rate estimation in HEVC intra coding. Although the aforementioned rate control algorithms [14–20] show promising performance for HEVC, their high computational complexity may not be suitable for a number of emerging video applications.

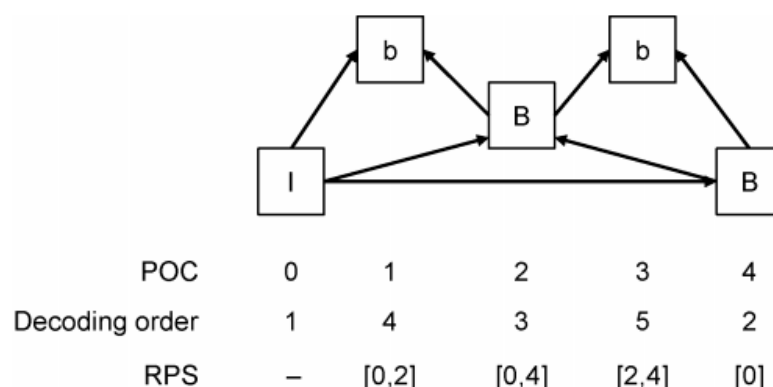


Figure 2. Example of a temporal prediction structure in HEVC inter coding.

In this work, we focused on developing a frame-level rate control algorithm. In order to do so, the learning process requires training data to have raw video contents at various sizes and resolutions. Therefore, using recently advanced learning frameworks such as deep learning [13] would require frame-level features for multiple frames in each GOP, thus demands a tremendous amount of computation, not only at training but also at deployment. Additionally, the learning framework often relies on a quality metric such as MSE, which does not well reflect the commonly used BDBR score in codec development. As a result, with our limited computing resources, we were not able to use a more expensive learning algorithm. To address this problem, and motivated by RDOQ, we proposed a simple but efficient temporal rate allocation solution for HEVC following a greedy method based on the trellis coding approach.

3. Proposed HEVC Improvement Tools

In this work, we aimed to develop a simple, non-normative MPEG video coding tool by integrating a VR creation solution into the HEVC and refining the existing temporal bitrate allocation scheme. We avoided the pre-analysis approach, which requires looking ahead several frames or GOPs, thus subsequently introduces significant complexity and is not suitable for relative applications. To achieve this target, we readjusted the predefined heuristic quantization level by a learned method. Unfortunately, the learning approach usually faces the challenge of the huge computational complexity of video coding, various datasets, as well as the nonlinearity of the encoding process [12]. We addressed these challenges by using a dynamic programming approach—trellis-based rate allocation (TRA).

The proposed TRA algorithm was integrated into both original HEVC and HEVC with virtual reference frames (VRF). Hence, this section introduces a novel HEVC with a virtual reference frame framework and is followed by the proposed TRA solution.

3.1. Proposed HEVC Architecture

The proposed HEVC encoding architecture is described in Figure 3 with highlighted modified modules. Virtual reference frames are generated from the reference decoded frames based on the multiple hypothesis motion estimation and then put to the DPB. Both the previously decoded references and the virtual references are used as the reference for motion estimation of the current frame.

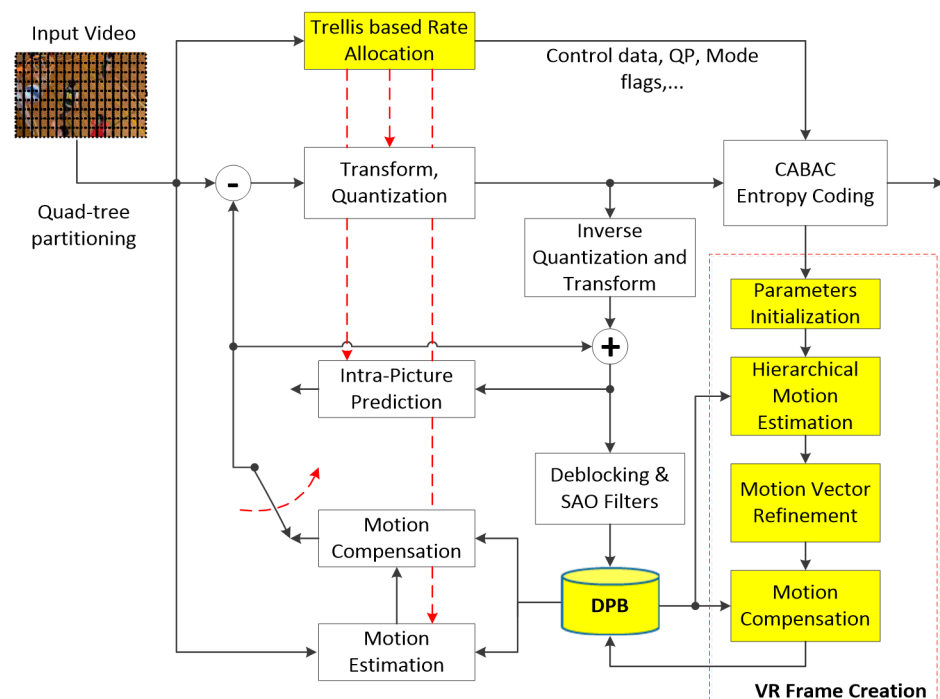


Figure 3. Proposed HEVC architecture.

3.1.1. Virtual Reference Frame Creation

To achieve high efficient encoder, we created a new virtual reference frame for each inter coding frame based on a hierarchical motion estimation (HME) and compensation technique. To fully exploit the statistic information from decoded data and the texture correlation between two consecutive references, we proposed an advanced motion-compensated temporal interpolation (MCTI) based VR frame creation, in which the motion vector field is adaptively generated using the hierarchical motion estimation (ME) solution with a coarse block size of 16×16 or 32×32 initialized in the backward ME. In this stage, the minimization of regularized mean absolute difference (RMAD) was adopted [22]. Figure 3 highlights the proposed MCTI structure. Since a large block size is hard to cover the motion information of tiny activity or video taken from a far camera, a finer block size may be chosen. Afterward, the motion vector refinement was adopted to refine the motion field. Finally, the motion compensation was used to create the VR frame. The creation of the VR frame can be performed as follows:

- **Hierarchical ME:** First, decoded frames obtained from two reference lists, list 0 and list 1, are low pass filtered and used as references in a motion estimation process. Our proposed VR frame creation uses both forward and backward ME to generate the forward interpolated frame and the backward interpolated frame. In these modules, a block matching algorithm is used to estimate the motion between the next and previous decoded frames.

To achieve more accurate motion vector information, an up-sample motion vector field (MVF) was performed. The motion vector value obtained from the previous step is an integer number; there will be jagged edges in the interpolated frame. To alleviate this problem, we up-sampled the reference frames by a factor of 2, both in horizontal and vertical directions, as shown in Figure 4. After up-sampling the MVF, coarsely refine the MVF considering its 8-neighboring MVs and choose that MVF has the smallest error cost.

- **MV refinement:** To achieve better motion field, a motion vector refinement (MVR) process is employed. In MVR, the temporal bidirectional ME (BiME) and the spatial weighted vector median filtering (WVMF) are chosen to refine the motion information derived from the hierarchical ME stage [22,23]. In BiME, the motion vectors of each interpolated block are refined in a small search area and following an assumption that the motion trajectory between consecutive frames is linear. While the spatial WVMF improves the motion field spatially coherent by looking, for each interpolated block, a candidate motion vector at neighboring blocks can better represent the motion trajectory. This filter is also adjustable by a set of weights, controlling the filter strength and depending on the block distortion for each candidate motion vector. Since the quality of decoded references and video content highly affect the final VR frame quality, we adopted a statistical learning-based parameter optimization solution to initialize the block size, search range, and search refinement areas for the proposed MCTI method [24].
- **Motion compensation:** Finally, the motion compensation process is applied to the decoded frames and obtained MV to achieve the VR frames.

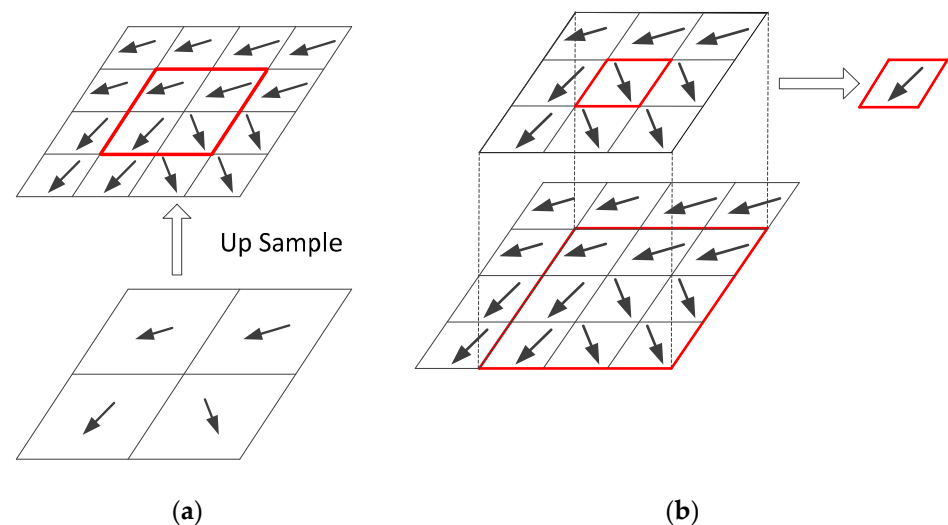


Figure 4. Hierarchical ME: (a) Upsample MVF; (b) MV selection.

3.1.2. Virtual Reference Frame Exploitation

To better use this new reference, we conducted exhaustive experiments to identify which frame position should be employed to generate VR frames. The generation of the VR frame benefits from the low complexity motion estimation and compensation approach combining with a data-driven optimal parameter configuration. The proposed VR is applied to the B-frame with bi-directional references as in [25].

3.2. Trellis-Based Rate Allocation (TRA)

Trellis search is a type of dynamic programming algorithm to find the optimal encoded sequence [26]. In video coding, trellis search is used for optimal quantization in rate-distortion optimized quantization (RDOQ) [27] of HEVC. Starting with initial scal5ar quantization (i.e., x), RDOQ recursively decides the optimal quantization level at each

coefficient location between candidates of $\{x, x - 1\}$ (additional candidate of 0 can be added) toward minimizing rate and distortion cost.

The overall step of our TRA is given in Algorithm 1. We modeled the bitrate allocation as an optimal search path of QP adjustment between temporal ID (Tid; illustrated in Figure 5). The details are explained as follows.

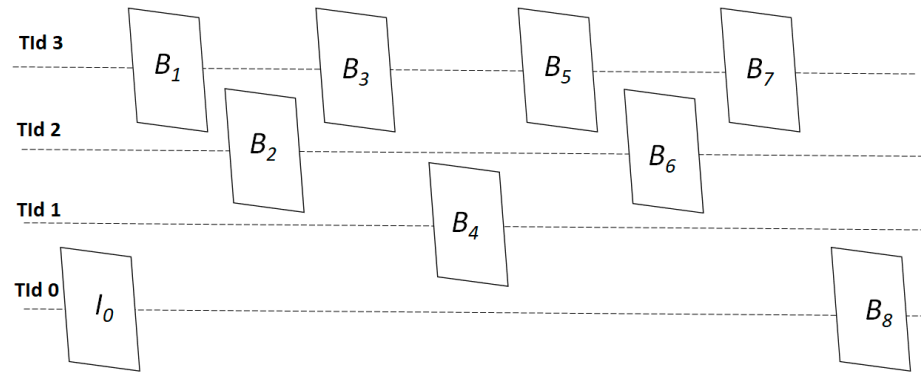


Figure 5. Tid illustration with HEVC Random Access configuration.

To avoid the worst case of the greedy algorithm, we have: (1) initialized the search with $\Delta QP^0 = [0, 0, 0, 0]$, which is the same as HEVC configuration; (2) using BDBR as the loss function to pick up the highest BDBR improvement path.

Algorithm 1 Proposed trellis-based bitrate allocation algorithm.

| # | Detail Descriptions | |
|---|---|--------------|
| 1 | Initialize offset $\Delta QP^0 = [0, 0, 0, 0]$, max range r , No. iteration K , excluding test $c_{ex}^{tid} = \{0\}$, $Tid = 0, 1, 2, 3$ | |
| 2 | For k in $\{1, \dots, K - 1\}$ | # Phase loop |
| 3 | For i in $\{0, 1, \dots\}$ | # Iter loop |
| 4 | For Tid in $\{0, 1, 2, 3\}$ | # Tid loop |
| 5 | Find exclude candidate $c_{ex}^{tid} \in \{c_{ex}^{tid}, \Delta QP_{tid}^k \pm k > 1\}$ | |
| 6 | Current candidate $c = \{\pm k\} \setminus \{c_{ex}\}$ | |
| 7 | Find the best $\Delta QP_{tid}^k \in c$ at tid w.r.t BDBR | |
| 8 | Add exclude candidates: $c_{ex}^{tid} \in \{c_{ex}^{tid}, c\}$ | |
| 9 | Output ΔQP^K | |

We then performed an iterative greedy search with the maximum K phases. In each phase, we searched the optimal QP offsets starting from Tid 0 to Tid 3 with respect to the best BDBR reduction. By starting from the most (Tid 0) to the least important picture (Tid 3), we avoided the great fluctuation in the algorithm (i.e., the BDBR gain tends to reduce after each step). An example of Tid structure in HEVC is shown in Figure 5. The lower Tid is, the more important it is, as being referred by more picture at lower Tid. In our practice, we used only two-phase, $K = 2$.

The best QP offset at the iteration k , and Tid i (i.e., denoted as ΔQP_i^k) is added to the optimal list ΔQP^K from the full candidate list of $\{0, \pm 1, \dots, \pm k\}$, which already removed candidates from the previous iteration (i.e., ΔQP_i^{k-1}). The maximum offset range was set to k in this work to limit the complexity. Based on the bitrate reduction rate (BDBR) of a given video dataset, we selected the best offset at high and low bitrate scenarios.

Given four Tids, at the maximum range of ± 2 (thus five available options), with six QPs for low and high rate, the number of tests for each sequence in the greedy search is

$$nQP \times r^4 = 6 \times 5^4 = 3750. \quad (1)$$

This still requires a tremendous amount of computation. Therefore, we used trellis-based dynamic programming to recursively search for the best offset of each TID. For one phase round of search (i.e., we named as a phase), it requires

$$nQP \times (5 + 4 + 4 + 4) = 6 \times 17 = 102, \quad (2)$$

for each sequence that significantly reduces simulations. Even though the proposed TRA could significantly reduce the searching space, it still demands a huge computation given multiple test/train sequences and several iterations. Therefore, we proposed a simplified version of TRA with a two-phase iterative approach based on the reference candidates. In the first phase, we limited the candidate list to $\{0, \pm 1\}$ which requires

$$nQP \times (3 + 2 + 2 + 2) = 6 \times 9 = 56, \quad (3)$$

half of experiments as in TRA. In the second phase, we only tested additional configuration, which is ± 1 difference from phase 1. If the optimal offset is +1, then only +2 is evaluated (i.e., 0 is evaluated in Phase 1). We skipped the case of the initial offset in Phase 1 being 0, as the difference to +2 is larger than 1. Therefore, the maximum additional experiments per iteration is

$$QP \times (1 + 1 + 1 + 1) = 6 \times 1 = 6. \quad (4)$$

The combination of experiments in (3) and (4) is significantly less than the original TRA in (2).

To train TRA, a decision was made to adjust the quantization level or ΔQP at each Tid related to the RD performance of the whole sequences. Fortunately, with many sequences, we adopted an offline simple training scheme, in which a decision is made based on the final BDBR performance. Then, we applied the learned TRA offset to test sequences without fine-tuning to avoid the complexity overhead. Therefore, we learned a data-dependent group of quantization ΔQP_i , $i = 0, 1, 2, 3$ while previous works independently learn to adjust ΔQP at a given QP setting.

4. Performance Evaluation

4.1. Training and Testing Conditions

To evaluate the coding performance of the proposed methods, Y-BDBR [28] is calculated under the Random Access (RA) configuration at the common test conditions [29]. We used the HM 16.20 [30] with GOP of 8, intra period of 32, and various quantization parameters for high rate (QP : 22, 27, 32, 37), which corresponds to the common test condition in HEVC and low rate (QP : 32, 37, 42, 45). For trellis-based rate allocation, we chose 16 sequences at various resolutions and frame rates while the test sequences are other common sequences in HEVC [31], and additional sequences can be found in [32,33]. Only the first 64 frames are used to reduce the training time but still produce high performance. The training and test sequences can be seen in Table 1, while Figures 6 and 7 illustrate the first frame of each video sequence.

To illustrate the effectiveness of the TRA method, its performance on the training dataset in BDBR is shown in Figure 8. It is easy to observe that TRA already achieves good performance even with phase one, which includes ΔQP set of $\{\pm 1, 0\}$. Moreover, we observed that one iteration was enough for TRA to converge in both phases. For the 1080p training set, over Phase 1, an additional ΔQP set of $\{\pm 2, \pm 1, 0\}$ in Phase 2 showed no BDBR improvement for TRA. Interestingly, the TRA and VRF combination slightly (i.e., -0.28%) and greatly reduced (i.e., 1.54%) rates for TRA + VRF at a high and low rate, respectively. The reason for no improvement on TRA at Phase 2 is due to the change ΔQP at frame level, which will greatly impact overall performance. Therefore, ± 2 might depart too much from the conventional rate. However, while combining with the VRF, $\Delta QP \pm 2$ showed a significant gain of 1% BDBR in the training set on average. The loss in quality at low is compensated by additional virtual frames generated by VRF. Please refer to Table 2 for more results on the different training classes. We observed that using the

same ΔQP set for all training data does show a minor quality improvement at a high rate. This is due to the nature of the local optima of dynamic programming. At a low rate, an additional 0.31% Y-BDBR is achieved for low rate but almost identical results for high rate. Therefore, we adopted a simple resolution-based adaptive ΔQP selection. As the resolution is available as the input, the proposed method requires no additional complexity compared to other frameworks. For complex adaptive methods (i.e., fast/slow motion, simple/details scene, etc.), this could be an interesting topic for our future work.

Table 1. Training and testing sequences at various resolutions and frame rates.

| Resolutions | Training Sequences | Testing Sequences |
|-------------------|--|--|
| 1080p and Class A | ParkScene, BQTerrace, SnakeNDry, ReadySteadyGo | BasketballDrive, Kimono, Cactus, PeopleOnStreet, Traffic |
| 720p and Class F | Parkjoy, InToTree, DucksTakeOff, Johnny | FourPeople, KristenAndSara, Vidyo1, Vidyo3, Vidyo4, ChinaSpeed, SlideEditing |
| 480p | RaceHorses, BasketballDrill | BQMall, PartyScene, BasketballDrillText |
| 4SIF | Crew, Harbour | Ice |
| 240p | RaceHorses, BQSquare | BasketballPass, BlowingBubbles |
| CIF | Akiyo, City | Hall, Foreman, Football |

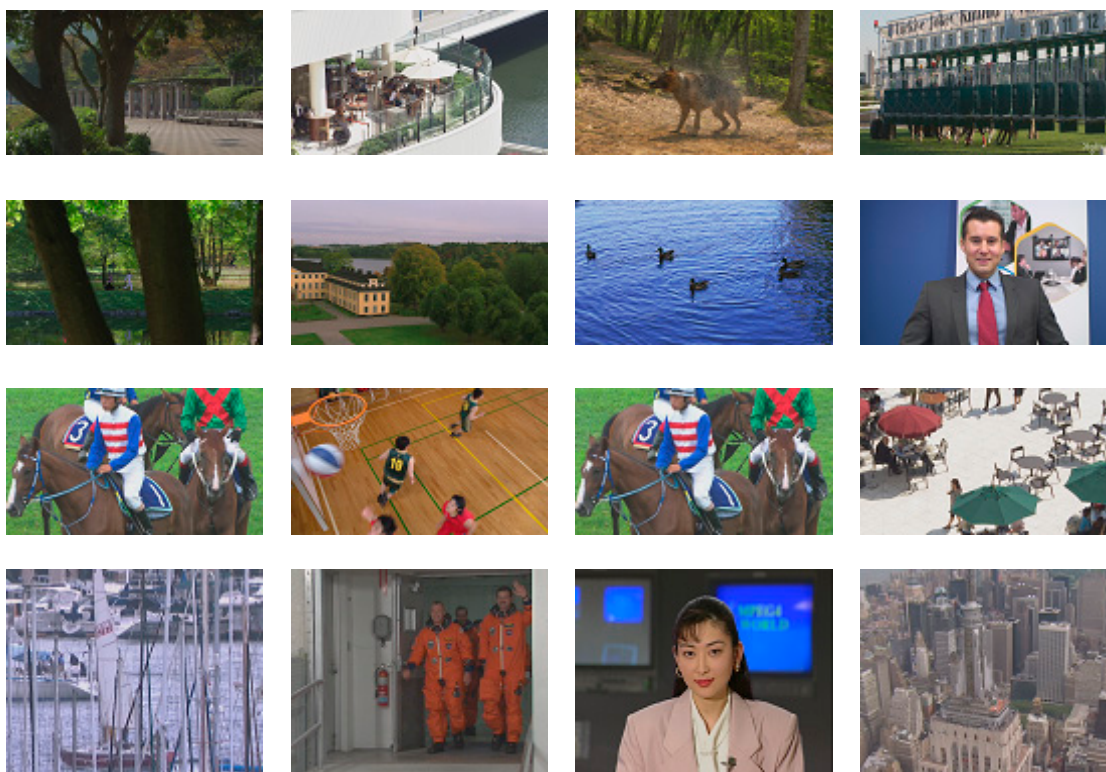


Figure 6. Illustration of the first frame of training videos. From left to right, top to bottom: ParkScene, BQTerrace, SnakeNDry, ReadySteadyGo, Parkjoy, InToTree, DucksTakeOff, Johnny, RaceHorses, BasketballDrill, Crew, Harbour, RaceHorses, BQSquare, Akiyo, City.

The learned ΔQP set is given in Table 2. Phase 2 only changes the Tid 0 in class 1080p, 720p and 240p. The main reason is related to the GOP 8, which has more Tid 0 frames and thus consumes a high rate. We also observed very similar results of the learned ΔQP with and without VRF with the exception of 1080p. A more detailed rate–distribution can be found in Section 4.3.

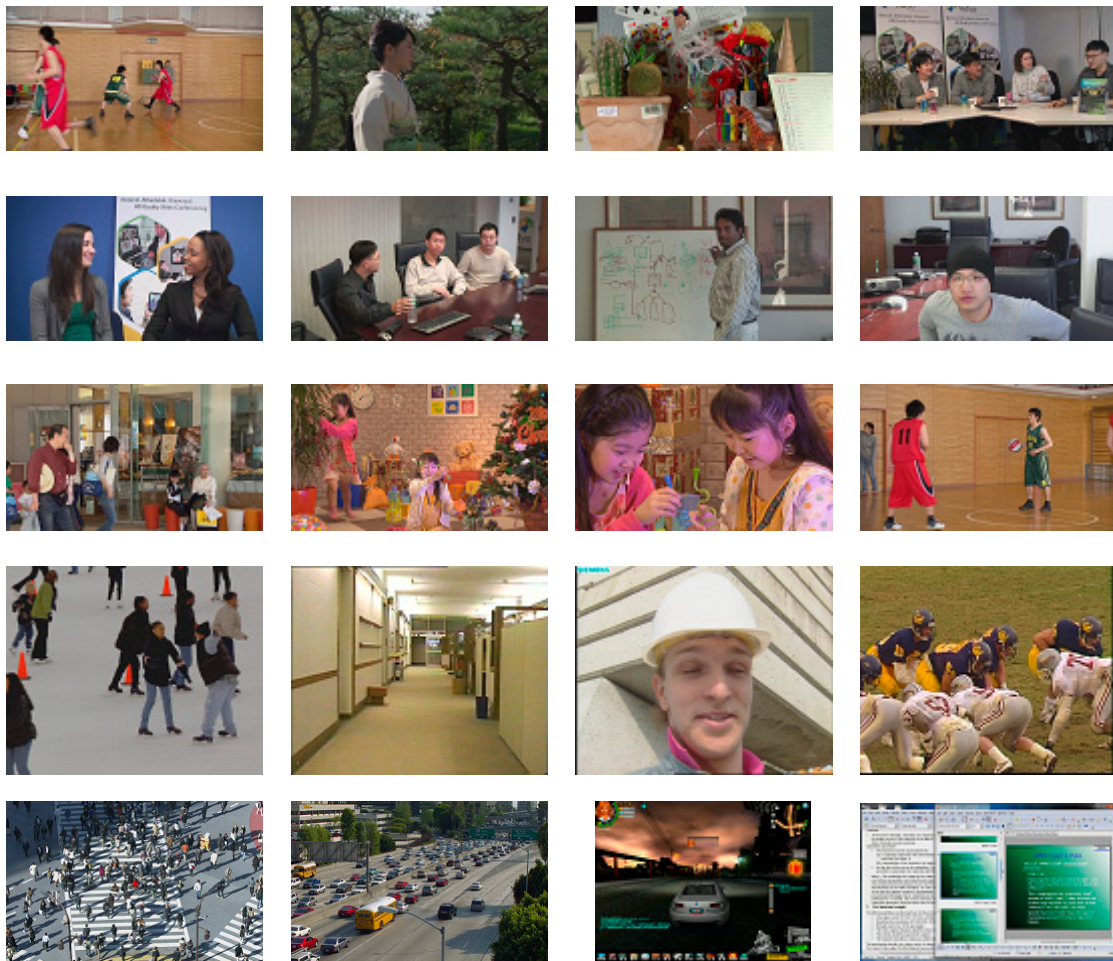


Figure 7. Illustration of the first frame of testing videos. From left to right, top to bottom: BasketballDrive, Kimono, Cactus, FourPeople, KristenAndSara, Vidyo1, Vidyo3, Vidyo4, BQMall, PartyScene, Ice, BasketballPass, BlowingBubbles, Hall, Foreman, Football, PeopleOnStreet, Traffic, ChinaSpeed, SlideEditing.

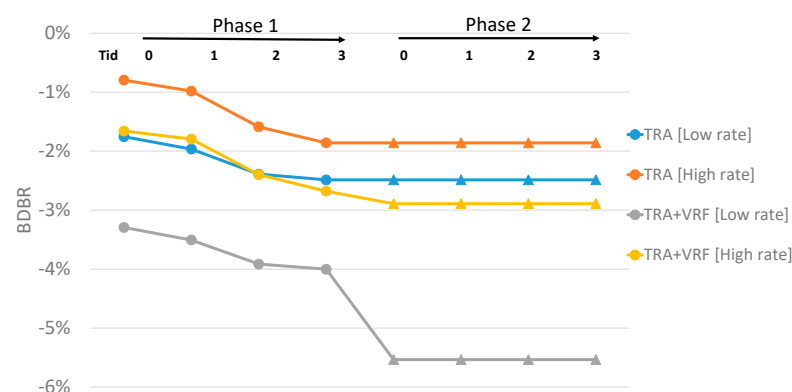


Figure 8. BDBR performance of TRA on training dataset for class B(1080p) at Phase 1 with $\Delta QP \in \{\pm 1, 0\}$ and Phase 2 with $\Delta QP \in \{\pm 2, \pm 1, 0\}$, only one iteration is used.

4.2. Compression Performance Assessment

The coding performance on test sequences is evaluated in Table 3 for BDBR and Table 4 for BDPSNR comparisons. We compared our TRA method with and without VRF to HEVC with (TRA) and without VRF (TRA + VRF), and the most relevant adaptive quantization method [3] as well as the standard HEVC. Similar to the training set, low-rate and high-

rate BDBR are used for comparison. Overall, both TRA and VRF provide a consistent improvement over the baseline HEVC.

Table 2. Training performance (Y-BDBR % at low rate/high rate).

| Resolutions | Best ΔQP for Each Class Sequences | | | | Best ΔQP for All Training Sequences | | | |
|-------------|---|-------------|----------------------------|-------------|---|-------------|----------------------------|-------------|
| | Phase 1: $\pm 1, 0$ | | Phase 2: $\pm 2, \pm 1, 0$ | | Phase 1: $\pm 1, 0$ | | Phase 2: $\pm 2, \pm 1, 0$ | |
| | TRA | TRA + VRF | TRA | TRA + VRF | TRA | TRA + VRF | TRA | TRA + VRF |
| 1080p | −2.49/−1.86 | −4.00/−2.68 | −2.53/−2.03 | −4.05/−2.85 | −2.49/−1.86 | −4.00/−2.68 | −3.77/−2.25 | −5.27/−3.12 |
| 720p | −1.48/−0.84 | −2.18/−0.84 | −1.72/−1.53 | −2.43/−1.50 | −0.36/−0.72 | −1.12/−0.84 | −2.53/−1.50 | −2.43/−1.50 |
| 480p | −1.44/−1.13 | −3.38/−2.02 | −1.49/−1.50 | −3.38/−2.43 | −1.44/−1.13 | −3.38/−2.02 | −2.03/−1.41 | −4.08/−1.93 |
| 4SIF | −1.25/−3.40 | −3.40/−1.00 | −1.56/−0.60 | −3.78/−1.56 | −1.07/−0.55 | −3.14/−1.48 | −2.03/−0.41 | −4.08/−1.93 |
| 240p | −0.76/−0.33 | −3.49/−1.49 | −1.05/−0.84 | −3.54/−1.59 | −0.83/−0.76 | −3.40/−1.69 | −1.63/−0.96 | −4.10/−1.72 |
| CIF | −1.21/−0.84 | −2.67/−1.23 | −1.26/−1.16 | −2.71/−1.43 | −1.07/−0.95 | −2.67/−1.23 | −1.68/−1.35 | −1.53/−1.45 |
| Average | −1.58/−1.01 | −3.17/−1.61 | −1.73/−1.40 | −3.30/−1.96 | −1.26/−1.07 | −2.85/−1.68 | −2.50/−1.45 | −3.80/−1.96 |

Table 3. Y-BDBR (%) coding performance of various methods for test sequences, full frames.

| | | Low Rate (QP = 32, 37, 42, 45) | | | | High Rate (QP = 22, 27, 32, 37) | | | |
|---------------------|---------------------|--------------------------------|-------|-------|-----------|---------------------------------|-------|-------|-----------|
| | | QPA [3] | TRA | VRF | TRA + VRF | QPA [3] | TRA | VRF | TRA + VRF |
| Class A and Class B | BasketballDrive | −0.44 | −1.38 | −1.84 | −3.64 | −0.43 | −1.51 | −0.91 | −2.21 |
| | Kimono | −0.49 | −0.99 | −3.09 | −4.82 | −1.69 | −1.14 | −1.75 | −2.65 |
| | Cactus | −0.17 | −1.69 | −3.33 | −6.06 | −0.09 | −1.20 | −1.87 | −2.96 |
| | PeopleOnStreet | N/A | −1.11 | −8.35 | −9.67 | N/A | −1.28 | −4.56 | −5.69 |
| | Traffic | N/A | −1.89 | −4.14 | −7.98 | N/A | −1.95 | −4.05 | −4.68 |
| HD 720p | FourPeople | −0.71 | −1.67 | −2.71 | −4.53 | −0.68 | −0.40 | −2.38 | −3.02 |
| | KristenAndSara | −0.66 | −1.75 | −2.56 | −4.37 | −0.56 | −0.68 | −1.92 | −2.77 |
| | Vidyo1 | −0.70 | −2.03 | −4.45 | −6.43 | −0.88 | −0.55 | −3.26 | −4.03 |
| | Vidyo3 | −0.73 | −2.24 | −3.35 | −5.83 | −0.61 | −0.63 | −1.85 | −2.70 |
| | Vidyo4 | −0.90 | −1.48 | −3.23 | −4.92 | −0.64 | −0.29 | −1.84 | −2.35 |
| | ChinaSpeed | N/A | −1.98 | −1.60 | −3.65 | N/A | −0.81 | −0.64 | −1.44 |
| | SlideEditing | N/A | −0.01 | 0.02 | 0.07 | N/A | 0.27 | −0.06 | 0.26 |
| Class C 480p | BQMall | −0.77 | −1.39 | −3.96 | −5.33 | −0.88 | −0.94 | −1.86 | −2.78 |
| | PartyScene | −0.57 | −2.80 | −1.70 | −4.44 | −0.61 | −1.18 | −0.65 | −1.83 |
| | BasketballDrillText | N/A | −2.24 | −2.22 | −4.37 | N/A | −1.61 | −1.11 | −2.69 |
| 4SIF | Ice | −0.68 | −0.80 | −9.33 | −9.90 | −0.97 | −0.77 | −4.23 | −4.60 |
| Class D | BasketballPass | −0.43 | −0.81 | −4.52 | −5.99 | 0.84 | 0.27 | −1.80 | −2.33 |
| | BlowingBubbles | −0.57 | −2.42 | −1.49 | −4.63 | 0.26 | 0.71 | −0.65 | −1.00 |
| CIF | Hall | −0.68 | −2.07 | −2.04 | −4.30 | −0.88 | −1.92 | −0.92 | −3.21 |
| | Foreman | −0.41 | −1.48 | −4.03 | −5.50 | −0.49 | −1.25 | −2.12 | −3.49 |
| | Football | −0.82 | −0.41 | −1.89 | −2.99 | −0.72 | −0.45 | −0.56 | −1.07 |
| Average | | −0.61 | −1.55 | −3.32 | −5.20 | −0.56 | −0.82 | −1.86 | −2.73 |

In addition, to explore the effectiveness of the proposed TRA method, we conducted further experiments for relevant frame-based QP adaptation methods, notably the QP- λ model proposed in [34] where the QP and λ are modeled as a linear function and our recent work in [35] where the QP and RD cost is modeled as a polynomial function. The model parameters were selected as in [34,35]. We also slightly modified the TRA to change the QP for only the Tid 0, named TRA_0. The BDBR comparison is illustrated in Table 5.

Table 4. Y-BDPSNR (dB) coding performance of various methods for test sequences, full frames.

| | | Low Rate (QP = 32, 37, 42, 45) | | | High Rate (QP = 22, 27, 32, 37) | | |
|---------------------|---------------------|--------------------------------|------|-----------|---------------------------------|------|-----------|
| | | TRA | VRF | TRA + VRF | TRA | VRF | TRA + VRF |
| Class A and Class B | BasketballDrive | 0.04 | 0.06 | 0.12 | 0.03 | 0.02 | 0.05 |
| | Kimono | 0.03 | 0.11 | 0.17 | 0.03 | 0.05 | 0.08 |
| | Cactus | 0.06 | 0.11 | 0.21 | 0.02 | 0.04 | 0.07 |
| | PeopleOnStreet | 0.05 | 0.40 | 0.46 | 0.06 | 0.20 | 0.25 |
| | Traffic | 0.08 | 0.16 | 0.34 | 0.07 | 0.31 | 0.16 |
| HD 720p | FourPeople | 0.09 | 0.14 | 0.24 | 0.02 | 0.09 | 0.11 |
| | KristenAndSara | 0.09 | 0.13 | 0.22 | 0.02 | 0.06 | 0.09 |
| | Vidyo1 | 0.10 | 0.22 | 0.33 | 0.02 | 0.11 | 0.13 |
| | Vidyo3 | 0.11 | 0.17 | 0.30 | 0.02 | 0.06 | 0.09 |
| | Vidyo4 | 0.06 | 0.14 | 0.22 | 0.01 | 0.06 | 0.07 |
| | ChinaSpeed | 0.09 | 0.07 | 0.16 | 0.04 | 0.03 | 0.08 |
| | SlideEditing | 0.00 | 0.00 | -0.01 | -0.04 | 0.01 | -0.04 |
| Class C 480p | BQMall | 0.06 | 0.17 | 0.23 | 0.04 | 0.07 | 0.11 |
| | PartyScene | 0.10 | 0.06 | 0.16 | 0.05 | 0.03 | 0.08 |
| | BasketballDrillText | 0.10 | 0.10 | 0.19 | 0.07 | 0.05 | 0.12 |
| 4SIF | Ice | 0.04 | 0.44 | 0.46 | 0.03 | 0.14 | 0.16 |
| Class D | BasketballPass | 0.03 | 0.18 | 0.24 | -0.01 | 0.09 | 0.11 |
| | BlowingBubbles | 0.08 | 0.06 | 0.16 | -0.03 | 0.03 | 0.04 |
| CIF | Hall | 0.11 | 0.11 | 0.23 | 0.06 | 0.03 | 0.10 |
| | Foreman | 0.06 | 0.17 | 0.23 | 0.05 | 0.09 | 0.15 |
| | Football | 0.01 | 0.07 | 0.11 | 0.02 | 0.03 | 0.06 |
| Average | | 0.07 | 0.15 | 0.23 | 0.03 | 0.08 | 0.10 |

Table 5. Y-BDBR (%) coding performance for TRA and other frame-based methods.

| Sequence | QPA [35] | QP— λ [34] (a = 4.2005, b = 13.7112) | QP— λ [34] (a = 4.2005, b = 9.7112) | QPA [3] | TRA_0 | Proposed TRA |
|----------------|----------|---|--|---------|-------|--------------|
| BasketballPass | 3.01 | 11.89 | 3.58 | -0.43 | -0.60 | -0.81 |
| BlowingBubbles | 5.16 | 8.39 | 3.37 | -0.57 | -1.48 | -2.42 |
| Football | 2.37 | 15.65 | 3.72 | -0.82 | -0.22 | -0.41 |
| Foreman | 3.90 | 10.59 | 2.63 | -0.41 | -1.35 | -1.48 |
| Hall | 5.39 | 7.27 | 2.91 | -0.68 | -1.62 | -2.07 |
| Average | 3.97 | 10.76 | 3.24 | -0.58 | -1.05 | -1.44 |

From the obtained testing performance shown in Tables 3–5, some conclusions can be obtained as:

- Overall, the compression performance of the HEVC with proposed TRA and VRF methods outperforms both HEVC with and without QPA benchmarks;
- The HEVC with TRA and VRF methods achieved a significant coding improvement for all test sequences, notably by 5.2% and 2.7% of BDBR on average for the low and high rates regions, respectively;

- The TRA method provides around 1.55% and 0.82% of BDBR saving for test sequences at the low and high rate regions, respectively, while the VRF method provides around 3.32% and 1.86% of BDBR saving;
- The proposed methods, both TRA and VRF, achieve better compression performance for the low rate region than for the high rate region;
- In most cases, the combination of TRA and VRF methods achieves even better compression performance than a simple addition method where the compression gain of TRA is added with that of VRF. This composed effect motivates the use of both TRA and VRF in improving HEVC performance;
- Experimental results also show that no compression gain is achieved for screen content videos such as *SlideEditing*. This may come from the fact that the training set does not include any video with this content, and the VRF creation may also not work well for screen-captured videos;
- Compared to other frame-based QPA algorithms, the proposed method achieved better BDBR (see Table 5). It should be noted that the QPA proposed in [35] was mainly designed for surveillance video content and its high-order polynomial model is highly sensitive to the selected parameters. Similarly, the QP- λ linear model proposed in [34] is also unable to achieve good BDBR performance even with the original model parameters used in [34] or our new parameters;
- A similar compression achievement is also observed for the BDPSNR comparison.

In addition, similar to other temporal rate allocation methods, TRA and TRA + VRF show better performance for slow-motion sequences while still greatly improve performance for fast-moving sequences such as *BasketballDrive*. It should be noted that the testing sequences are different from the training sequences, thus validating the generalization of our TRA scheme.

To further visualize the performance of our proposed method, Figure 9 shows the RD comparison between the original HEVC and the proposed TRA + VRF.

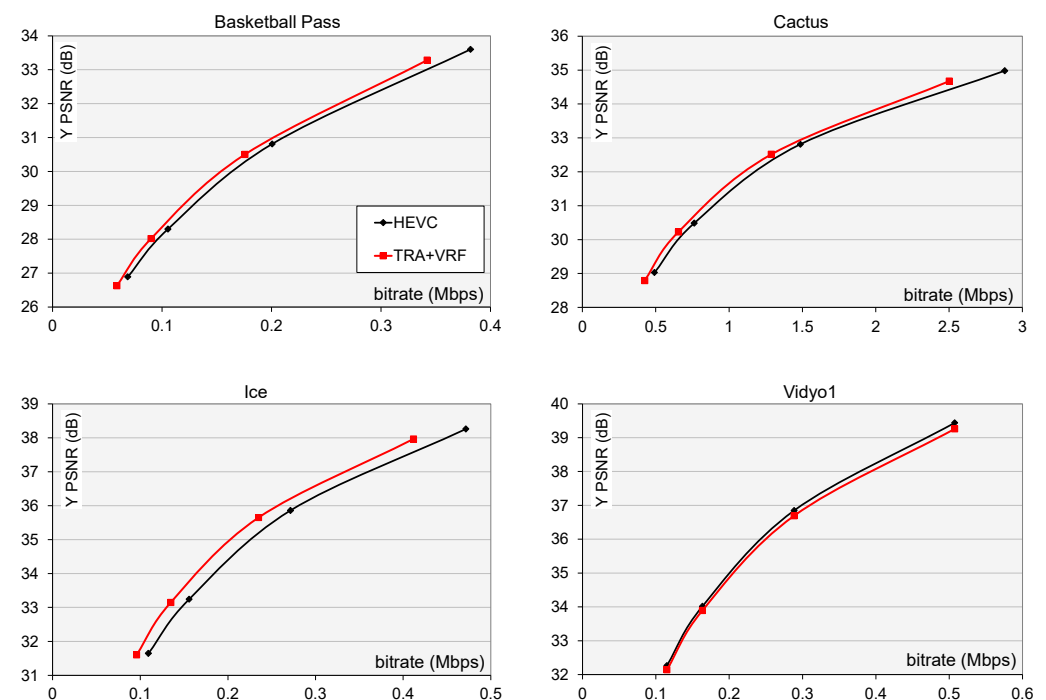


Figure 9. RD performance comparison between the original HEVC and the proposed methods for various sequences, full frames, low-rate.

4.3. Rate Allocation Assessment

To explore the rate allocation performance with the proposed method, the rate distribution with different temporal layers is measured and shown in Table 6 and Figure 10 for both low rate and high rate.

Table 6. Rate distribution (%) between temporal index.

| | Tid | HEVC | HEVC + Proposed Methods | | |
|----|-----|-------|-------------------------|-------|-----------|
| | | | VRF | TRA | TRA + VRF |
| HR | 0-I | 23.95 | 24.07 | 30.06 | 30.23 |
| | 0-P | 32.03 | 32.19 | 33.43 | 33.61 |
| | 1-B | 17.27 | 17.27 | 16.43 | 16.41 |
| | 2-B | 20.27 | 20.16 | 11.69 | 11.57 |
| | 3-B | 6.50 | 6.30 | 8.39 | 8.19 |
| LR | 0-I | 45.78 | 46.27 | 51.44 | 52.04 |
| | 0-P | 34.57 | 34.94 | 31.24 | 31.58 |
| | 1-B | 8.20 | 8.17 | 7.42 | 7.36 |
| | 2-B | 7.82 | 7.51 | 5.81 | 5.48 |
| | 3-B | 3.63 | 3.11 | 4.09 | 3.54 |

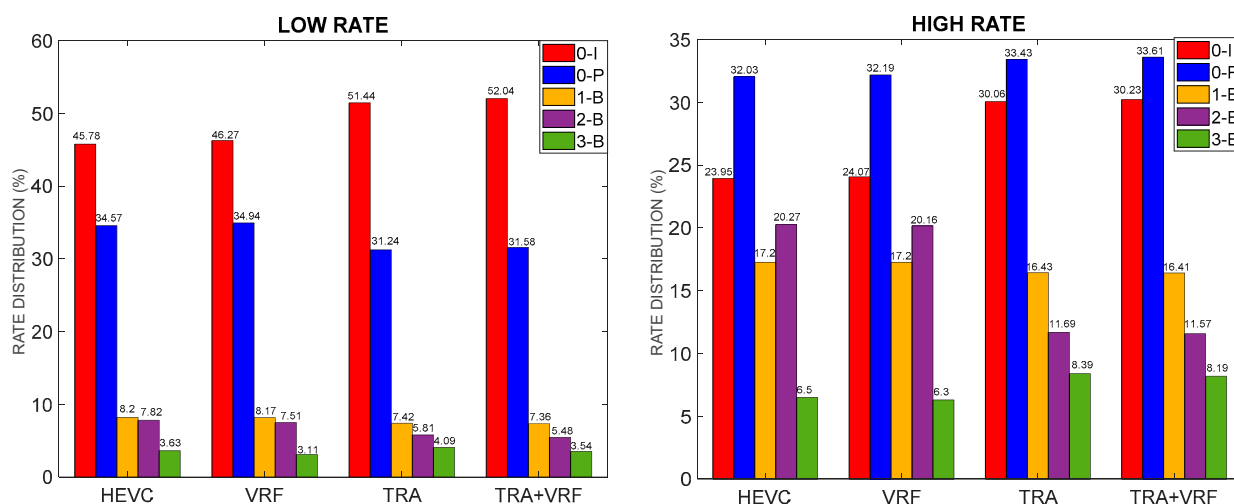


Figure 10. Rate allocation among coded pictures in GOP and for several codecs.

The proposed TRA redistributed the rate among temporal layers. The proposed method tends to give more bits for pictures at the lower temporal layer indexes, i.e., Tid 0. In fact, the pictures at these positions have a higher impact than the remaining ones as they can be referred to when coding the pictures at the higher Tid. In general, our proposed TRA tends to reduce the rate while maintaining a similar level of quality.

On the other hand, VRF shows a virtually similar rate distribution compared to the baseline HEVC. Therefore, the rate redistribution is only impacted by our TRA. Details on the rate distribution can be seen in Table 6.

4.4. Complexity Assessment

To complete the performance evaluation, we evaluated the encoding time of our proposed TRA with and without VRF. The time increase with the proposed tools was assessed and is shown in Table 7. As obtained, The TRA method does not affect the computational complexity of the encoder.

Table 7. Encoding time (%) compared to HEVC.

| Class | Low Rate | | High Rate | |
|---------|----------|-----------|-----------|-----------|
| | TRA | TRA + VRF | TRA | TRA + VRF |
| 1080p | 99.7 | 675.90 | 99.1 | 561.5 |
| 720p | 100.8 | 447.10 | 101.9 | 463.2 |
| 480p | 87.0 | 231.30 | 85.6 | 281.7 |
| 4SIF | 105.9 | 231.19 | 96.9 | 243.4 |
| 240p | 109.3 | 148.55 | 106.9 | 166.7 |
| CIF | 96.1 | 165.50 | 96.9 | 180.8 |
| Average | 99.8 | 316.60 | 97.9 | 316.2 |

On the other hand, TRA can even reduce the encoding time, especially at a high rate. The main reason is due to the reduction in total rate, which greatly impacts other coding tools. However, the computation time caused by the VRF method introduces nearly three times the encoding increase to the proposed HEVC. This mainly comes from the creation of additional references, and more reference frames also lead to more encoding time for motion estimation. This complexity–compression performance trade-off may prevent the wide deployment of VRF in some video coding applications where the computational complexity is constrained. However, as shown in Table 4, BDBR with the combination of TRA and VRF methods is mostly higher than the total BDBR of TRA and VRF when they are individually applied. This effect motivates to include in HEVC not only the rate control solution but also other VRF creation approach in future video coding development. Additionally, the complexity issue in VRF can be addressed by looking for low complexity VRF creation methods.

4.5. VRF Assessment

Finally, although VRF has shown an impressive compression performance when adopted in the HEVC architecture, it also introduced a remarkable complexity overhead. To further explore the creation of VRF in the proposed HEVC, this sub-section will assess the visual quality and the computational complexity of VRF.

As presented in Section 3.1.1, the VRF creation adopted in this paper includes three main steps, (i) the hierarchical motion estimation (HME) to initialize the motion vector field, (ii) the motion vector refinement (MVR) to mitigate the quantization errors which frequently happen in decoded references and (iii) motion compensation (MC) to interpolate the VRF. To reveal the contribution of these modules, we compute and illustrate in Table 8 the percentage of computation time (%) measured for VRF over overall HEVC ($\%VRF_{HEVC}$), and for the HME, MVR, and MC over the VRF ($\%HME_{VRF}$, $\%MVR_{VRF}$, $\%MC_{VRF}$), while Figure 11 illustrates the visual quality comparison for the proposed VRF with and without MVR and the related work in [25].

$$\%VRF_{HEVC} = Time_{VRF} \times 100 / Time_{HEVC} \quad (5)$$

$$\%HME_{VRF} = Time_{HME} \times 100 / Time_{VRF} \quad (6)$$

$$\%MVR_{VRF} = Time_{MVR} \times 100 / Time_{VRF} \quad (7)$$

$$\%MC_{VRF} = Time_{MC} \times 100 / Time_{VRF} \quad (8)$$

The complexity results obtained in Table 8 shows that the VRF consumes around 11.70% of computational complexity in overall HEVC encoding. In VRF, the HME required the largest computational complexity, i.e., 95.24%. In this case, reducing the time of the HME is critical to make the VRF more potential for widespread adoption in HEVC.

As observed in Figure 11, both HME and MVR contribute to the final quality of VRF. Naturally, HME exploits the video content to initialize the motion vector information; the good MV can mitigate the hole and occlusion problems in the interpolated frame. Meanwhile, the MVR can partly reduce the quantization noise and blocking artifacts in interpolated pictures by adaptively refining the MVF obtained in the HME. It should be noted that the blocking artifact typically happens in HEVC compressed videos due to the block-based coding approach, while the translational motion model adopted in BiME may also introduce further blocking artifact (see Figure 11 visual quality for VRF w/o MVR). In this case, WVMF can eliminate the “outlier” motion information, resulting in smoother MVF for final VRF creation (see Figure 11 visual quality for proposed VRF).

Table 8. VRF Complexity assessment.

| Video | %VRF _{HEVC} | %HME _{VRF} | %MVR _{VRF} | %MC _{VRF} |
|---------------------|----------------------|---------------------|---------------------|--------------------|
| BasketballDrive | 12.71 | 96.06 | 2.37 | 1.56 |
| Kimono | 12.41 | 96.09 | 2.38 | 1.53 |
| Cactus | 14.12 | 96.03 | 2.37 | 1.59 |
| PeopleOnStreet | 14.17 | 96.85 | 1.87 | 1.28 |
| Traffic | 18.47 | 96.81 | 1.86 | 1.33 |
| FourPeople | 13.70 | 95.21 | 2.85 | 1.94 |
| KristenAndSara | 12.71 | 96.06 | 2.37 | 1.56 |
| Vidyo1 | 13.72 | 95.22 | 2.82 | 1.96 |
| Vidyo3 | 13.15 | 95.26 | 2.82 | 1.93 |
| Vidyo4 | 13.24 | 95.24 | 2.85 | 1.91 |
| ChinaSpeed | 10.24 | 95.24 | 2.87 | 1.89 |
| SlideEditing | 13.73 | 95.22 | 2.83 | 1.95 |
| BQMall | 10.12 | 94.90 | 3.10 | 2.00 |
| PartyScene | 8.68 | 94.83 | 3.15 | 2.02 |
| BasketballDrillText | 9.75 | 94.84 | 3.13 | 2.02 |
| Ice | 10.71 | 94.76 | 3.11 | 2.12 |
| BasketballPass | 10.26 | 94.24 | 3.50 | 2.26 |
| BlowingBubbles | 8.13 | 94.33 | 3.46 | 2.21 |
| Hall | 10.85 | 94.25 | 3.28 | 2.46 |
| Foreman | 9.00 | 94.36 | 3.53 | 2.12 |
| Football | 5.85 | 94.30 | 3.42 | 2.28 |
| Average | 11.70 | 95.24 | 2.85 | 1.90 |



Figure 11. VRF visual quality illustration (*highlighting the improvement in the red-circle*): (a) Ice; (b) Vidyo1; (c) Football; (d) BQMall.

5. Conclusions

This paper presented a novel trellis-based bitrate allocation algorithm for HEVC inter coding where a virtual reference frame is considered. The proposed VRF-HEVC relies on the decoded information available at both the encoder and decoder; thus, no syntax elements need to change in the standard specification, and no overhead bitrate needs to concern. To optimize the VRF quality, a statistical learning-based VRF frame creation is adopted. This paper is the first work that considers the impact of quantization error and the video content on the VR frame quality. In addition, to achieve higher HEVC compression performance, a novel set of quantization parameters is introduced for the VRF-HEVC framework based on a dynamic programming based bitrate allocation approach. Experimental results obtained for a rich set of test sequences revealed that the proposed TRA and VRF based HEVC solutions significantly outperform relevant HEVC improvement methods and the standard HEVC.

Author Contributions: Conceptualization, X.H. and T.N.C.; methodology, X.H.; software, L.D.T.H.; validation, X.H., T.N.C.; formal analysis, X.H.; investigation, L.D.T.H.; resources, T.N.C.; data curation, X.H.; writing—original draft preparation, X.H.; writing—review and editing, T.N.C. and X.H.; visualization, L.D.T.H.; supervision, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2020.15.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sullivan, G.J.; Ohm, J.-R.; Han, W.-J.; Wiegand, T. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 1649–1668. [[CrossRef](#)]
2. Bross, B.; Chen, J.; Ohm, J.-R.; Sullivan, G.J.; Wang, Y.-K. Developments in International Video Coding Standardization after AVC, With an Overview of Versatile Video Coding (VVC). *Proc. IEEE* **2020**. [[CrossRef](#)]
3. Papadopoulos, M.A.; Zhang, F.; Agrafiotis, D.; Bull, D. An adaptive QP offset determination method for HEVC. In Proceedings of the IEEE Conference on Image Processing, Phoenix, AZ, USA, 25–28 September 2016; pp. 4220–4224.
4. Chen, Y.; Murherjee, D.; Han, J.; Grange, A.; Xu, Y.; Liu, Z.; Parker, S.; Chen, C.; Su, H.; Joshi, U.; et al. An Overview of Core Coding Tools in the AV1 Video Codec. In Proceedings of the Picture Coding Symposium (PCS), San Francisco, CA, USA, 24–27 June 2018; pp. 41–45.
5. Xu, M.; Canh, T.N.; Jeon, B. Simplified Rate-Distortion Optimized Quantization for HEVC. In Proceedings of the IEEE 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Valencia, Spain, 6–8 June 2018; pp. 1–6.
6. Zhang, Y.; Tian, R.; Liu, J.; Wang, N. Fast rate distortion optimized quantization for HEVC. In Proceedings of the 2015 Visual Communications and Image Processing (VCIP), Singapore, 13–16 December 2015; pp. 1–4.
7. Sun, L.; Au, O.C.; Zhao, C.; Huang, F.H. Rate distortion modeling and adaptive rate control scheme for high efficiency video coding (HEVC). In Proceedings of the 2014 IEEE International Symposium on Circuits and Systems (ISCAS), Melbourne, VIC, Australia, 1–5 June 2014; pp. 1933–1936.
8. Zhao, L.; Wang, S.; Zhang, X.; Wang, S.; Ma, S.; Gao, W. Enhanced Motion-Compensated Video Coding With Deep Virtual Reference Frame Generation. *IEEE Trans. Image Process.* **2019**, *28*, 4832–4844. [[CrossRef](#)] [[PubMed](#)]
9. Zhang, Y.; Zhang, C.; Fan, R. Fast Motion Estimation in HEVC Inter Coding: An Overview of Recent Advances. In Proceedings of the 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA, 12–15 November 2018; pp. 49–56.
10. Djelouah, A.; Campos, J.; Schaub-Meyer, S.; Schroers, C. Neural Inter-Frame Compression for Video Coding. In Proceedings of the IEEE International Conference Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 6420–6428.
11. Zhao, L.; Wang, S.; Zhang, X.; Wang, S.; Ma, S.; Gao, W. Enhanced Ctu-Level Inter Prediction with Deep Frame Rate Up-Conversion for High Efficiency Video Coding. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 206–210.
12. Xu, M.; Canh, T.N.; Jeon, B. Simplified Level Estimation for Rate-Distortion Optimized Quantization of HEVC. *IEEE Trans. Broadcast.* **2019**, *66*, 88–99. [[CrossRef](#)]
13. Xu, M.; Canh, T.N.; Jeon, B. Rate-Distortion Optimized Quantization: A Deep Learning Approach. In Proceedings of the IEEE High Performance Extreme Computing Conference, Waltham, MA, USA, 25–27 September 2018.
14. Choi, H.; Yoo, J.; Nam, J.; Sim, D.-G. Pixel-wise unified rate-quantization model for multi-level rate control. *IEEE J. Sel. Top. Signal Process.* **2013**, *7*, 1112–1123. [[CrossRef](#)]
15. Wang, S.; Ma, S.; Wang, S.; Zao, D.; Gao, W. Quadratic Rho-domain based rate control algorithm for HEVC. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013.
16. Li, B.; Li, H.; Li, L.; Zhang, J. λ domain rate control algorithm for high efficiency video coding. *IEEE Trans. Image Process.* **2014**, *23*, 3841–3854. [[CrossRef](#)] [[PubMed](#)]
17. Wang, M.; Ngan, K.N.; Li, H. An efficient frame-content based intra frame rate control for high efficiency video coding. *IEEE Signal Process. Lett.* **2014**, *22*, 896–900. [[CrossRef](#)]
18. Li, S.; Xu, M.; Deng, X.; Wang, Z. Weight-based R- λ rate control for perceptual HEVC coding on conversational videos. *Signal Process. Image Commun.* **2015**, *38*, 127–140. [[CrossRef](#)]
19. Gao, W.; Kwong, S.; Zhou, Y.; Yuan, H. SSIM-based game theory approach for rate-distortion optimized intra frame CTU-level bit allocation. *IEEE Trans. Multimed.* **2016**, *18*, 988–999. [[CrossRef](#)]
20. Yan, T.; Ra, I.-H.; Zhang, Q.; Xu, H.; Huang, L. A Novel Rate Control Algorithm Based on ρ Model for Multiview High Efficiency Video Coding. *Electronics* **2020**, *9*, 166. [[CrossRef](#)]
21. Yuanzhi, Z.; Chao, L. A Highly Parallel Hardware Architecture of Table – Based CABAC Bitrate Estimator in an HEVC Intra Encoder. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 1544–1558.

22. Ascenso, J.; Brites, C.; Pereira, F. Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding. In Proceedings of the Conference on Speech and Image Process, Multimedia Communication and Services, Smolenice, Slovak, 29 June–2 July 2005.
23. Jeong, S.-G.; Lee, C.; Kim, C.-S. Motion-Compensated Frame Interpolation based Multihypothesis Motion Estimation and Texture Optimization. *IEEE Trans. Image Process.* **2013**, *22*, 4497–4505. [[CrossRef](#)] [[PubMed](#)]
24. Hoangvan, X. Statistical search range adaptation solution for effective frame rate up-conversion. *IET Image Process.* **2018**, *12*, 113–120. [[CrossRef](#)]
25. HoangVan, X.; Ascenso, J.; Pereira, F. Improving predictive video coding performance with decoder side information. In Proceedings of the IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012.
26. Fischer, T.R. Joint trellis coded quantization/modulation. *IEEE Trans. Commun.* **1991**, *39*, 17–176. [[CrossRef](#)]
27. Karczewicz, M.; Ye, Y.; Chong, I. Rate Distortion Optimized Quantization. In Proceedings of the ITU-T VCEG Meeting, Antalya, Turkey, 12–13 January 2008.
28. Bjontegaard, G. Calculation of average PSNR differences between RD curves. In Proceedings of the 13th ITU-T VCEG Meeting, Austin, TX, USA, 2–4 April 2001.
29. Bossen, F. Common HM Test Conditions and Software Reference Configuration. In Proceedings of the 14th meeting of the Joint Collaborative Team on Video Coding (JCT-VC), Vienna, Austria, 25 July–2 August 2013.
30. HEVC Reference Software. Available online: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/ (accessed on 8 June 2021).
31. Video Test Sequences. Available online: <ftp://hevc@ftp.tnt.uni-hannover.de/testsequences/> (accessed on 20 November 2020).
32. Available online: <ftp://vqeg.its.bldrdoc.gov> (accessed on 27 November 2020).
33. Xiph.org Video Test Media. Available online: <https://media.xiph.org/video/derf/> (accessed on 20 November 2020).
34. Li, B.; Xu, J.; Zhang, D.; Li, H. QP refinement according to lagrange multiplier for high efficiency video coding. In Proceedings of the IEEE International Symposium on Circuits and Systems, Beijing, China, 19–23 May 2013.
35. HoangVan, X. Adaptive Quantization Parameter Estimation for HEVC Based Surveillance Scalable Video Coding. *Electronics* **2020**, *9*, 915. [[CrossRef](#)]