



Xuesong Jin *, Huiyuan Sun and Yuhang Zhang D

School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China; sunhuiyuan2021@163.com (H.S.); zhangyh@s.hrbcu.edu.cn (Y.Z.)

* Correspondence: shy@s.hrbcu.edu.cn

Abstract: This research is based on an intra-frame rate control algorithm based on the Versatile Video Coding (VVC) standard, considering that there is the phenomenon of over-allocating the bitrate of the end coding tree units (CTUs) in the bit allocation process, while the front CTUs are not effectively compressed. Fusing a Canny-based edge detection algorithm, a color contrast-based saliency detection algorithm, a Sum of Absolute Transformed Differences (SATD) based CTU coding complexity measure, and a Partial Least Squares (PLS) regression model, this paper proposes a CTU-level bit allocation improvement scheme for intra-mode code rate control of the VVC standard. First, natural images are selected to produce a lightweight dataset. Second, different metrics are utilized to obtain the significance and complexity values of each coding unit, the relatively important coding units in the whole frame are selected, which are adjusted with different weights, and the optimal adjustment multiplicity is supplemented into the dataset. Finally, the PLS regression model was used to obtain regression equations to refine the weights for adjusting the bit allocation. The proposed bit allocation scheme improves the average rate control accuracy by 0.453%, Y-PSNR by 0.05 dB, BD-rate savings by 0.33%, and BD-PSNR by 0.03 dB compared to the VVC standard rate control algorithm.

Keywords: video coding; Versatile Video Coding (VVC); intra rate control; bit allocation; significance detection

check for updates

Citation: Jin, X.; Sun, H.; Zhang, Y. Research on VVC Intra-Frame Bit Allocation Scheme Based on Significance Detection. *Appl. Sci.* **2024**, *14*, 471. https://doi.org/10.3390/ app14010471

Received: 17 December 2023 Revised: 29 December 2023 Accepted: 3 January 2024 Published: 4 January 2024



Copyright: © 2024 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

With the development and progress of science and technology, digital video has gradually become an important medium for disseminating information. At the same time, access to video information is no longer only dependent on fixed devices, the development of mobile devices such as cell phones and tablet computers has also led to progress in the video industry. Furthermore, the China Internet Network Information Center (CNNIC) in Beijing has determined in its 52nd China Internet Development Report that the expansion of the Internet is a significant factor in the advancement of the video industry [1]. As of June 2023, the number of Internet users in China reached 1.079 billion, and the Internet penetration rate reached 76.4%. Digital video transmission is in high demand for a variety of purposes, including webcasting, video conferencing, security surveillance, and personal entertainment. With technological innovations, higher resolution [2] videos are gradually used in people's productive lives, so it is especially important to compress the volume of video data more effectively without affecting the content of the video [3]. Video coding and decoding technologies can reduce the cost of video storage and transmission needs while increasing transmission efficiency. To further enhance coding efficiency, expand versatility, and alleviate the strain on the Internet, a succession of global video coding standards have been established, with Versatile Video Coding (VVC) being the most recent iteration of these standards.

The video transmission bandwidth is usually subject to certain limitations, and bandwidth differences in network communication affect the transmission bitrate of video and video reconstruction quality. the relationship between the coded output bitrate and the compressed distortion is mutually constraining and contradictory. To assure the playback quality of video while effectively transmitting data, despite constraints on channel bandwidth and transmission delay, rate control of the video coding process is necessary.

In past research, Li et al. [4] proposed the R- λ model based on the Lagrange multiplier λ according to the relationship between rate and distortion, which achieved excellent results and was adopted by the High Efficiency Video Coding (HEVC) standard and carried over to the VVC standard. In the process of rate control in intra-frame mode, front CTUs tend to have lower coding costs due to the absence of inter-frame information as a reference. On the other hand, the last CTU often needs to increase bit cost significantly to reach the target bitrate, resulting in uneven cost allocation for each CTU in intra-frame mode. To address this issue, this research employs two significance detection methods and a complexity calculation method to guide the bit allocation process at the CTU level, achieving a more reasonable allocation of bits.

The rest of the paper is organized as follows, Section 2 introduces the video coding standards and related work on rate control, respectively; Section 3 analyzes the characteristics of existing rate control algorithms in terms of bit allocation and proposes an optimization scheme for bit allocation based on significance detection; Section 4 shows the experimental results and analyzes them; and finally, it concludes the work as well as proposes further optimization schemes.

2. Related Work

2.1. Video Coding Standard

A wide variety of video applications have given rise to multiple video coding methods from the very beginning. To ensure interoperability of encoded streams and standardization of decoding across a broad spectrum, international standards for video coding typically embody the latest advancements in video coding technology from the same generation [5]. Figure 1 shows the development of each family of video coding standards.



Figure 1. Development process of video coding standards in various series.

The H.261 [6] coding standard is predominantly designed for low-bit rate and high real-time video transmission. Being the inaugural iteration of video coding standards, it holds significant guiding importance within the domain of video coding. In 1995, the International Telecommunications Union-Telecommunications Standardization Sector (ITU-T) and the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) organizations collaborated for the first time to develop the

H.262/MPEG-2 [7] standard, which improves the computational power and complexity of video coding algorithms compared to the H.261 standard. The H.263 [8] standard was developed for the transmission of low-bit-rate video communication, which provides more efficient video coding than the H.261 coding standard and better single-frame images in the video. Advanced Video Coding (AVC) [9] greatly improves video compression performance while having the Bit Error Rate (BER) resistance to transmit the information over the network. HEVC [10] follows the basic coding framework of AVC, with updated substitutions of techniques in the basic modules. Compared to the AVC coding standard, the HEVC standard reduces the BER and increases the correctness of real-time transmission. The VVC standard [11] is similar to previous international video coding standards in that it still employs a hybrid coding framework, i.e., multiple methods are used for coding at the same time, but more advanced coding techniques are used in key modules [12]. Under the same Peak Signal to Noise Ratio (PSNR) conditions, the performance of the VVC standard saves an average of 50% bitrate compared to HEVC, while the decoding complexity is no more than twice that of the HEVC standard [13]. So far, the VVC test model (VTM) has been released up to version 22.2.

2.2. Conventional Rate Control Algorithms

Due to the critical nature of the rate control module in video coding standards, each standard recommends a distinct rate control model, such as TMN8 for H.263 [14], JVT-G012 for AVC [15], JCTVC-K0103 for HEVC, and JVET-K0390 for VVC [16]. From a mathematical modeling point of view, the bit rate control schemes can be classified into three types of models, namely, Q-domain models, λ -domain models, and ρ -domain models. The three models initially establish a correlation between bit rate and distortion; however, scholars have progressively transformed this correlation into one with the parameters of the proposed model, which subsequently dictate the quantization parameters.

The Q-domain model explores the relationship between the code rate R and the distortion D in the QP domain and is the most widely used model for rate control. Lee et al. [17] proposed a more accurate second-order Rate-Distortion (R-D) model based on Q-domain. In it, a sliding window is used to select the data, which can effectively reduce the impact of scene changes. Meanwhile, shape coding artifacts are eliminated using adaptive thresholding, which achieves accurate bit allocation with lower latency and limited buffer. This scheme has been adopted by the MPEG-4 standard. Wu et al. [18] proposed a low-delay HEVC rate control scheme based on a temporal prediction structure. Firstly, a quadratic R-Q model is established and a method for determining the QP of the first frame is proposed. Then a frame-level bit allocation scheme is proposed, and finally, the CTU-level QP is predicted. In the 8th meeting of JCT-VC, Choi et al. [19] proposed a rate control algorithm based on the R-Q model, which captures the relationship between the target rate R and the QP value of a pixel, and then unifies the hierarchical relationship of the coding units in the rate control, which means that the model can be used for Group of Pictures (GOP), frame or CTU, and the algorithm is also adopted by the AVC coding standard.

However, according to the research of JCTVC-I0426 [20], the slope λ of the R-D curve has a more important position in the determination of the bit rate. Li et al. [4] proposed the K0103 rate control proposal, which has been incorporated into the HEVC standard, and established the R- λ and λ -QP models to ascertain the quantization parameters of the coding unit in accordance with Shannon's rate-distortion theory via the Lagrange multiplier λ , which signifies the rate-distortion model's slope, as a link. Karc et al. [21] proposed an intraframe rate control scheme adapted to the R- λ model based on SATD, which provides better rate control at the intra-level and the slice level. In addition to this, Guo et al. [22] proposed a frame-level bit allocation scheme in HEVC, which presents an improved R-D model to obtain higher accuracy with lower bit rate and mismatch rate by utilizing the information of the coded frames more fully. To achieve the most optimal R-D performance on a global scale, an optimal frame-level bit allocation formula and a GOP-level Lagrange multiplier are implemented. Chen et al. [23] proposed a new R-D quadratic model using VTM-2.0 as a platform to model the distortion as a quadratic function of the logarithm of the bit rate and derive the R-D relationship from it. In addition, this new GOP-level and frame-level bit allocation schemes are proposed to implement the frame-level rate control algorithm for VVC. The study conducted by Li et al. [24] examined the impact of skip blocks on the estimation of rate-distortion parameters and the correlation between inter-frame quality. The findings of this research served as inspiration for a redesign of frame-level bit allocation, which further reduces the prediction error of the R- λ model at the frame-level and CTU-level by utilizing the statistical information of skip block ratios and disregarding the parameter updating information of the skip blocks.

2.3. Visual Saliency-Based Video Coding

Visual saliency detection is a fundamental method of image analysis within the domain of computer vision. It operates by employing intelligent algorithms to simulate the attributes of the human visual system to enable the extraction of salient regions from an image. These regions are of particular interest to humans, and their extraction involves the prediction of eye movements and visual gaze points. With the development of related technologies, researchers have gradually combined the field of video coding with saliency detection techniques, actively sacrificing a portion of non-significant video scene regions to achieve better-quality coding results. Currently, the forms of saliency modules combined with video coding techniques are roughly divided into three categories: adjusting the quantization parameter QP and the allocation of target bits in the rate control according to the saliency; replacing the current rate-distortion metrics with distortion metrics based on the characteristics of human vision; and replacing the current rate-distortion metrics with Just Noticeable Distortion (JND) based Perceptual Video Coding (PVC) technique.

Bai et al. [25] determined that the current $R-\lambda$ rate control model fails to adequately capture the characteristics of the video content. To address this, they proposed a prominence mapping-based rate control scheme that is applicable at both the framelevel and CTU-level. Additionally, they implemented a new feedback mechanism to improve the correlation between the weights of coding units and the weight gains. This mechanism further influences the frame-level λ and the quantization parameter (QP), resulting in improved performance. Luz et al. [26], in the context of the lack of video coding schemes specifically designed for 360-degree panoramic videos, proposed an adaptive coding solution that improves performance by coding the most visually salient video regions with higher quality in the process of quantization parameter control, and a machine-learning based saliency detection model for detecting higher quality coding applicable to the situation. Milani et al. [27] proposed a saliency map that identifies the key elements of the detector and maps the saliency to the value of the quantization parameter to be used by the video encoder, which improves the accuracy with respect to the default algorithm. A saliency model based on deep learning was suggested by Pelurson et al. [28] as a means to enhance the effectiveness of video coding. Following the introduction of a saliency-guided preprocessing filtering phase, the two methods are combined. Both objective and subjective assessments indicate that a substantial reduction in bit cost can be achieved without compromising the visual quality. Li et al. [29] proposed a PVC framework consisting of a fast coding unit (CU) division algorithm and a quantization control algorithm compliant with the VVC optimization scheme in the context of perceptual video coding. First, based on the visual saliency model, a fast CU segmentation scheme is proposed, which includes re-determining the CU segmentation depth by calculating the Scharr operator and variance; and executing the decision for the Intra-Subdivision (ISP) to reduce the coding complexity. Second, a multilevel classification based on significant values at the CU level reduces the bitrate by adjusting the quantization parameters. This algorithm significantly reduces computational complexity and saves bitrate with reasonable peak SNR loss and almost identical subjective perceptual quality.

3. Model and Proposed Method

3.1. Analysis of Rate Control Model

The rate control problem for a video sequence can be described as determining the optimal quantization parameters for each coding unit such that the total distortion is minimized under the condition that the total number of coded bits is less than or equal to R_{cr} i.e.,

$$Q^* = (Q_1^*, \dots, Q_N^*) = \arg\min_{(Q_1, \dots, Q_N)} \sum_{i=1}^N D_i \ s.t. \sum_{i=1}^N R_i \le R_c$$
(1)

Take the image as a coding unit as an example, where *N* is the number of images contained in the sequence, D_i is the distortion of the ith image, $Q^* = (Q_1^*, \ldots, Q_N^*)$ is the optimal quantization parameter for each image. However, the complexity of determining the quantization parameters of the coding unit directly based on (1) is extremely high, so Li et al. [4] developed the R- λ and λ -QP models and proposed the K0103 code rate control proposal. The classical R- λ model in this proposal is followed in the code rate control algorithm of VVC. Therefore, the code rate control scheme is usually divided into two parts:

- Target Bit Allocation: Considering the correlation of video in the air and time domains, the optimal number of target bits for each coding unit is determined according to the total number of target bits in accordance with the levels of GOP, Frame, Slice, CTU, and so on.
- Quantization parameter decision: based on the model of the relationship between coding rate and quantization parameter, its quantization parameter is determined independently for each coding unit according to its target bit number.

where the number of target bits for each CTU is determined by (2). For an image containing N_L CTUs, the target bit number $T_L(m)$ of the m-th CTU to be coded is:

$$T_L(m) = \frac{T_f - H_f - R_{L,c}}{\sum_{k=m}^{N_L} \omega_L(k)} \omega_L(m)$$
⁽²⁾

 T_f is the bit cost allocated by the encoder for the current frame. $R_{L,c}$ is the sum of the actual bit costs used by all encoded CTUs in the current frame, H_f is the predicted value of the number of bits of header information for this image, and $\omega_L(m)$ is the weight assigned to the bits of each CTU. Experiments in the literature [4] have shown that the hyperbolic function can express the relationship between code rate and distortion well:

$$D(R) = C_R^{-K} \tag{3}$$

where *C* and *K* are model parameters related to video content characteristics, and the Lagrange factor λ is:

$$\lambda = -\frac{\partial D}{\partial R} = CKR^{-K-1} = \alpha \cdot R^{\beta} \tag{4}$$

The rate R is defined in practical coding as the number of bits per pixel (bpp) assigned to a particular coding unit. However, for improved intra-coding bitrate control, the VVC standard incorporates the complexity measure *C* of the coding unit into the R- λ model, which is founded on the SATD, i.e.,

$$\lambda = \alpha \cdot \left(\frac{C}{bpp}\right)^{\beta} \tag{5}$$

As an example, a 128×128 CTU contains 16,384 pixels and corresponding transform coefficients inside. In order to calculate the coding complexity of this CTU, we need to perform the following steps:

- Apply the Hadamard transform to the original CTU, traversing each pixel within this CTU and taking the absolute value of the resulting transform coefficients for each pixel.
- The sum of the absolute values of the transform coefficients of all pixels within the CTU yields the codestream complexity of the CTU. Assuming that the result of the calculation is SATD.
- Calculate the code rate complexity for each pixel. Divide the code stream complexity N by the square of the number of pixels within this CTU, i.e., by 16,384 × 16,384, to obtain the per-pixel code rate complexity C.

$$SATD = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |h_{i,j}|, N = 128$$
(6)

$$C = \frac{SATD}{N_{pixels}} \tag{7}$$

Finally, based on the code rate complexity *C* and the number of remaining bits for each pixel, the appropriate QP value can be determined to achieve the goal of rate control, QP is determined as shown in (8).

$$QP = 4.2005 ln\lambda + 13.7122 + 0.5 \tag{8}$$

There are typically two modes of operation for intra-frame rate control:

- Using the constant λ for the entire frame. Parameters α and β can be updated after each frame is encoded.
- Parameters α and β are kept constant throughout the frame, but bits are allocated to each CTU based on the number of remaining bits allocated to the frame. The value of λ is calculated for each CTU using Equation (5).

While the HEVC coding standard adopts the first mode of rate control as its intramode, the VVC standard uses the second mode for its intra-mode, which is also known as Bitrate-based Rate Control. In this mode, the value of λ for each CTU will be calculated based on the number of remaining bits of that CTU and the parameters α , β , and γ (limiting the number of bit counts at the frame and image level). In this way, the remaining bit count will be equally distributed to each CTU to achieve the target bit rate for the whole frame. This ensures that the image quality is balanced across CTUs and reduces the space occupation. However, the research in [30] mentions the phenomenon that in the actual coding process, the bit cost of the last one of a set of CTUs to be assigned is often much larger than the bit cost of all the preceding CTUs. Some of the results of the verification using the official test sequences provided by the VVC standard with the default rate control algorithm turned on in the coding mode of all intra are shown in Figure 2. In the figure, we use some of the official test sequences to observe the bit cost allocation per CTU under different bit cost conditions in intra mode.

From Figure 2, we can see that the bpp of all the previous CTUs is maintained at a low level, while the bpp of the last CTU is significantly higher than the others. This phenomenon only occurs in all intra modes, which is because the predefined intra-frame rate control parameters of the default intra-frame rate control algorithm are always useless for coding, and since each coding unit is coded independently, the encoder is unable to utilize the information of the inter-frame prediction for code rate allocation. As a result, all the preceding CTUs may be allocated fewer bits because the encoder tries to save as many bits as possible while maintaining the image quality. While the last CTU increases the number of coding bits due to the need to reach the set target bit rate, this may result in the last CTU being allocated significantly more code rate than all the previous CTUs, leading to an overall uneven code rate distribution. Since frames undergoing intra-frame mode coding are used as reference frames by subsequent P- and B-frames to continue coding, especially the first frame of the whole sequence, improving the efficiency of intra-frame

code rate control at the target bit rate remains a challenge. This experiment addresses this characteristic by introducing two significance detection algorithms and the coding complexity originally used in the process of determining the quantization parameters for the R- λ model into the bit allocation process, resulting in an even distribution of the bit cost of the coding units at the same level. The overall flow of the program is shown in Figure 3.



Figure 2. The bpp situation of all CTUs in the first frame of intra-frame mode under the conditions of the default rate control algorithm. (a) BQMall; (b) Cactus; (c) BQSquare; (d) Fourpeople.

The scheme first constructs a lightweight dataset of video sequences then decides the set of improved CTUs in the frame based on the Canny edge detection-based saliency detection algorithm, color contrast-based saliency detection algorithm, and SATD-based complexity measure together, adjusts the improved CTUs with different multiplicity to expand the dataset, and finally fits the dataset data with a regression equation using the PLS model. The significance module and the regression equation are jointly embedded in the encoder to complete the rate control. Algorithm details will be presented in subsequent sections.

3.2. Dataset Production

This experiment aims to build a generalizability regression equation by selecting videos with different conditions and selecting certain of these CTUs for cost enhancement or reduction at different multiplicities. To achieve this goal, we use a PLS regression model and generate a lightweight small sample dataset to fulfill the experimental requirements. In the experiments of this research, we decided to use a lightweight small sample dataset instead of a large-scale dataset. Despite the small number of datasets, they were carefully designed and selected to be able to cover the key conditions and parameters in video coding

while providing some generalization. This allows us to conduct experiments on a relatively small scale and obtain meaningful results. One of the benefits of using lightweight small sample datasets is the time and cost savings in data collection and processing. Compared to large-scale datasets, we need to invest fewer resources to generate and manage these samples. In addition, small sample datasets can be more flexibly tuned and controlled to better understand the multiplicity weight values in video coding. This helps us to conduct experiments quickly and efficiently and gain valuable insights from them. Although the lightweight small-sample dataset is limited in size, we can still obtain reliable and representative results through proper design and sample selection. This provides an effective method for in-depth research on the characteristics and influencing factors of intra-patterns in the VVC standard and provides valuable reference and guidance for practical applications.



Figure 3. The overall flow of the proposed algorithm.

In this research, 18 natural images with different resolutions, different scenes, and other conditions were selected as base images on the public dataset platform, each base image was flipped horizontally to expand the dataset, respectively, and finally, the images were converted into single-frame YUV videos using the ffmpeg multimedia runtime library for each image. The parameters of the base pictures are shown in Table 1.

Table 1. Parameters and coding costs for 18 base images.

Sequences (.yuv)	Resolution	Content Scene	Target Bitrate (kbps)
mv1, mv2, mv3	2560×1600	subdistrict	12,000/30,000/40,000/50,000
mv4, mv5, mv6	1920×1080	high-speed movement	1500/2000/2500/3000
mv7, mv8, mv9	832 imes 480	Indoor shopping malls	20,000/30,000/40,000/50,000
mv10, mv11, mv12	416×240	field movement	5000/8000/10,000/12,000
mv13, mv14, mv15	1280 imes 720	living environment	5000/6000/8500/10,000
mv16, mv17, mv18	832 imes 480	Basketball (text masking)	2000/4000/6000/8000

In Table 1, mv1–mv3 use overhead views of street pedestrians, and the video content belongs to the category of scenes with high complexity and variability; mv4–mv6 belong to scenes of sports and physical activities, and video sequences in this category may

contain fast movements, which challenge the motion estimation and motion compensation capabilities of the coding algorithms; mv7–mv9 belongs to videos that refer to indoor scenes such as commercial plazas or shopping centers, these video sequences usually have more details and higher complexity, which puts certain requirements on the detail retention and compression effect of the coding algorithm; mv10–mv12 belongs to the scenes referring to animal races or animals in the wild, which are different from mv4–mv6 in terms of resolution, coding cost, and video content; mv13–mv15 belong to the scenes of character activities in daily life, which are usually restricted to a smaller scope; mv16–mv18 are the expansion of mv4–mv6, which add text masking of video scenes based on sports, such as scoreboards, advertisement spots and so on. These six categories basically generalize the video sequence scenes in the general case and have good generalization ability.

3.3. Options for Improving CTUs

The purpose of this experiment is to select certain CTUs (hereinafter referred to as improved CTUs) and perform an active process to enhance the bit cost of their default algorithms and to reduce the bit cost of other CTUs during the bit allocation process of the default rate control algorithm at the CTU level. This section will focus on the determination process of the improved CTUs, while the determination of the adjustment multiplier will be presented in Section 3.4.

Once all the videos in the dataset have been acquired, the saliency and complexity information for each CTU in each video (i.e., each frame) must be obtained so that the set of CTUs that require improvement in this frame can be determined. The saliency detection algorithm based on Canny edge detection is able to accurately detect the edges of the image and also has good resistance to noise. The process of saliency extraction using the Canny algorithm needs to complete the following steps:

- Grayscaling: While the encoder enters the frame-level encoding, the complete pixel values of the luma component of the frame, i.e., the grayscale image, are extracted.
- Gaussian Filtering: Smooth the grayscale map with Gaussian filtering to reduce the effect of noise.
- Calculate the gradient of the image: Use the Sobel operator to calculate the gradient of the grayscale map in the horizontal and vertical directions, so as to obtain the gradient intensity and direction of each pixel point.
- Non-extremely large value suppression: For each pixel, compare the size of the gradient values of two neighboring pixels in the gradient direction, and keep only the pixel with the largest gradient value in the gradient direction, which is conducive to eliminating the blurring effect on the edges.
- Dual Thresholding: Classify pixel points into three categories: strong edges, weak edges, and non-edges. When the gradient value of a pixel point is higher than the higher threshold, it is categorized as a strong edge. When the gradient value of a pixel point is below a lower threshold, it is categorized as non-edge. When the gradient value of a pixel point is between the higher and lower thresholds, it is classified as a weak edge.
- Edge Joining: Join weak edges and their surrounding strong edges to form complete edges.

Since the saliency map obtained by the Canny-based algorithm is a binary image, we use the sum of the saliency values of all white pixel points within the pixel range of each CTU to characterize the saliency value of that CTU. At the same time, we use a color contrast-based saliency detection algorithm to simultaneously guide the CTU saliency. First, the image is converted to grayscale after applying the Laplace convolution kernel to convolve the grayscale map and produce a gradient map in which the salient regions have higher gradient values and the points larger than a set threshold are labeled as edge points. In the actual test, we use the default threshold of the algorithm as the edge threshold. Since the resulting saliency map is not a purely binary image, we extracted the CTU-level saliency values by averaging the saliency values of all pixel points for each CTU to characterize the saliency value of the CTU. To eliminate the impact of specific areas of

high luminance on CTUs, all pixel points within a CTU should be counted in the average value category instead of non-black pixel points. The Canny-based saliency map with the saliency information obtained based on color contrast for the test sequence mv5 is shown in Figure 4.



Figure 4. Significance information extraction using mv5 as an example. (a) original image; (b) grayscale image; (c) significant map based on Canny algorithm; and (d) significant map based on color contrast algorithm.

It should be noted that although the Laplace operator emphasizes high-frequency information and regions with large brightness variations, this saliency detection algorithm still has some problems. First, it is sensitive to noise and requires some noise suppression processing; second, the discretization matrix of the Laplace operator introduces a sharp-ening effect when processing the image, which may lead to some unnecessary details in the saliency map. Therefore, this experiment uses the Canny algorithm with SATD-based coding complexity together with the Laplacian operator to guide the selection of improved CTUs. The Canny significance value, color contrast significance value, and coding complexity information obtained for mv5 as an example are shown in Figure 5. From the metrics obtained in the three ways, it can be seen that the metrics converging to half of all CTUs in the frame are the more obvious ones, and they should be assigned more than the average number of bits, whereas all three metrics of the last CTU are significantly low, not even up to the average, and it is clearly not reasonable to assign the extreme bit cost to the last CTU.

After obtaining the relevant information about each CTU using each of the three methods, they are sorted in descending order and the CTU serial number that co-occurs in the first *K* values is selected as the serial number of the improved CTU, with the value of *K* being determined by the following:

$$K = round\left(\frac{1}{5}\left(\frac{1}{2}N_{CTU} + \frac{1}{4}\sum_{i=1}^{4}T_{F,i}\frac{1}{24000}\right)\right) \cdot 5$$
(9)

where N_{CTU} is the total number of CTUs in the frame, and $T_{F,i}$ is the actual number of target bits allocated by the encoder for the frame under the condition of 4 different target bit numbers. The value of *K* will be taken as a multiple of 5 of the closest *K*. Subsequently, the serial number of the improved CTUs will be determined by *K*. In order to prevent the



occurrence of overflow or imbalance between the number of allocated bits and the number of CTUs, they are trimmed after the decision of *K*. The trimming is shown in Table 2.

Figure 5. Feature information of mv5 as an example. (a) coding complexity based on SATD; (b) significance value based on Canny algorithm; (c) significance value based on color contrast algorithm.

Table 2. Cutting of K.

Conditions	Adjustment of K Value
<i>N_{CTU}</i> > 100	$K = round\left(\frac{1}{5}min\left(\frac{N_{CTU}}{2} + \frac{1}{4}\sum_{i=1}^{4}T_{F,i}\frac{1}{24000}, \frac{N_{CTU}}{2} + \frac{N_{CTU}}{10}\right)\right) \cdot 5$
$10 \le N_{CTU} \le 100$	$K = round\left(\frac{1}{5}\left(\frac{1}{2}N_{CTU} + \frac{1}{4}\sum_{i=1}^{4}T_{F,i}\frac{1}{24000}\right)\right) \cdot 5$
$N_{CTU} < 10$	$K = round\left(\frac{1}{5}\left(\frac{1}{2}N_{CTU}\right)\right) \cdot 5$

3.4. Determining Optimal Weight Values: Decision Process

After determining the decision-making approach for the improved CTU in this experiment, while improving CTU for bit allocation, the bit cost assigned by the default rate control algorithm is adjusted by multiplying the weight parameter as shown in (10).

$$f(x) = \begin{cases} \frac{T_f - H_f - R_{L,c}}{\sum_{k=m}^{N_L} \omega_L(k)} \omega_L(m) \omega_U(m), & m \in K \\ \frac{T_f - H_f - R_{L,c}}{\sum_{k=m}^{N_L} \omega_L(k)} \omega_L(m) \omega_D(m), & m \in (U - K) \end{cases}$$
(10)

Among them, $\omega_U(m)$ is the lifting parameter, $\omega_D(m)$ is a parameter for reduction, K is the set of all improved CTUs, *U* is the set of all CTUs for the frame, and the other parameters

have been explained in (2). For the determination of the two, this experiment was fitted using a PLS regression model, which in turn determined the adjusted magnification of the improved CTUs. PLS regression is a multivariate statistical method for solving covariance problems. Its applicable scenario is when the number of two sets of variables is large, both of them are multi-correlated, and the number of observed data are small The model built by partial least squares regression has the advantages that traditional methods such as classical regression analysis do not have. The dataset first needs to be further expanded and information extracted by enhancing the bit allocation of the improved CTUs in the video in each dataset with different multiplicities and reducing the other CTUs in the same frame, respectively. After the coding of all the videos is completed, the multiplier that results in the most efficient improvement in the rate control of each video is used as the expansion of the dataset, which is the output of the PLS regression model. The improvement in rate control efficiency for the sequence mv5 in the dataset is shown in Figure 6, where the x-axis is the boosting multiplier, the y-axis corresponds to the decreasing multiplier, and the z-axis corresponds to the BD-PSNR, which is used to characterize the reconstruction pixel quality improvement. The highest value in the figure is 0.043 dB, which corresponds to the red data point, for the case of boosting multiplicity 1.2 and decreasing multiplicity 1.



Figure 6. Relationship between multiplicity and rate control efficiency for the example of mv5.

From the feature inputs of the model, the proportion of improved CTUs, average framelevel bit cost, and saliency and complexity information are selected for this experiment, as shown in Table 3.

Features	Calculation Method
Proportion of improved CTUs	$ratio = rac{N_c}{N_{CTU}}$
Average frame-level target bitrate	$T_{F,avg} = rac{1}{4}\sum_{i=1}^{4}T_{F,i}$
Improving CTU's average complexity	$T_{C,avg} = rac{1}{N_c}\sum\limits_{i=1}^{N_c} C_i$
Mean Canny significance values for improved CTUs	$T_{st1,avg} = rac{1}{N_c}\sum\limits_{i=1}^{N_c}S_{1,i}$
Mean color contrast significance values for improved CTUs	$T_{st2,avg} = \frac{1}{N_c} \sum_{i=1}^{N_c} S_{2,i}$

Where, N_c is the number of improved CTUs, C_i is the coding complexity of each CTU, $S_{1,i}$ is the Canny significance value of each CTU, and $S_{2,i}$ is the color contrast significance value of each CTU. To assess the predictive capability of the models, 30% of the 36 features extracted during data processing comprised the optimal multiplicity mapping data. The remaining 30% was allocated to the training set. The number of PLS principal components defined in this experiment is 5. It is worth noting that the boosted magnification is regressed using the same features as the reduced magnification. The regression analysis of the model with improved CTU enhancement magnification is shown in Figure 7a–c, and the analysis with reduced magnification is shown in Figure 7d–f.



Figure 7. Regression model analysis for improved CTUs.

(a) depicts the variation of the mean square error (MSE) of Y computed by crossvalidation on the training set with the number of PLS components, with the lowest MSE value and the best fit when the number of principal components is 5. (b) and (c) describe the relationship between the actual response values and the predicted values of the model in the training and test sets, respectively, where the horizontal coordinates represent the resultant data in the training set and the vertical coordinates represent the predicted values of the model. Ideally, the data points should be arranged around a straight line with a slope of 1, indicating that the predicted values match the actual values, which reflects the model's ability to generalize over unknown data, at the same time, the slope of the best-fit

Table 3. PLS model feature selection.

line (red solid line) differs from 1 by a distance that is within an acceptable margin of error. Comparison of images from (b) with (c) and (d) with (e), respectively, shows that the fitting effect of using the model for the training set and the test set is close to each other, which proves that the model has a better generalization ability. Finally, in the actual bit allocation process of rate control, the feature information corresponding to the sequence is brought into the regression equation fitted by RLS to obtain the tuned CTU allocation.

4. Results

4.1. Experimental Condition

In order to compare the performance of the proposed bit allocation scheme with the default rate control scheme, this experiment is compared with the official encoder VTM14.0 provided by JVET, and all the test sequences need to be tested using the official CTC [31] test standard, and we select the 1st frame from 6 sequences with different scenarios, resolutions, and content richness to be tested under the all intra mode, and the detailed information of each test sequence is shown in Table 4.

Table 4. Detailed information about the test sequence.

Sequences	Resolution	Frame Number	Frame Rate	Bit Depth	Number of CTUs per Frame
PeopleOnStreet	2560×1600	300	30	8	260
BasketballDrive	1920×1080	500	50	8	135
BQMall	832 imes 480	600	60	8	28
RaceHorses	416 imes 240	300	30	8	8
FourPeople	1280×720	300	60	8	60
BasketballDrillText	832 imes 480	500	50	8	28

Widely used quality evaluation metrics in video coding are Bjontegaard–Delta rate (BD-rate), Peak Signal-to-Noise Ratio (PSNR), and BD-PSNR. BD-rate denotes the rate increase in the optimized algorithm compared to the original algorithm for the same objective quality of the video, and a negative value indicates that the coding performance of the optimized algorithm has been improved. PSNR is an evaluation standard for video quality by calculating the average of the sum of squared errors of each pixel of the original and compressed videos and converting it to logarithmic form in dB, which is often used for comparing the performance of different compression algorithms and evaluating the quality of the compressed video. In general, the higher the PSNR value, the higher the similarity between the compressed video and the original video, and the better the video quality. The PSNR is calculated as shown in (11) and (12).

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} || I(i,j) - K(i,j) ||^2$$
(11)

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right)$$
(12)

where I(i, j) and K(i, j) are the pixel values of each pixel point of the original and reconstructed images, respectively. *MSE* defines the mean square error of the two images and MAX_I denotes the maximum value of the pixel value of each pixel point of the image. BD-PSNR evaluates coding effectiveness by comparing the PSNR values of different coding schemes at the same bit rate. A lower BD-PSNR value indicates that one coding scheme is more efficient compared to the other. The BD-PSNR is calculated as shown in (13).

$$BD - PSNR = 10 \cdot \log_{10} \left(\frac{R_{old}}{R_{new}} \right)$$
(13)

where R_{old} is the PSNR of the original video and R_{new} is the PSNR of the reconstructed video. In addition to this, another important metric in rate control is the rate control

accuracy, which describes the extent to which the actual number of bits output in the rate control deviates from the target number of bits set before encoding. The purpose of video coding is to maintain the cost of coding at a target level while ensuring the quality of the reconstructed video, so the appropriateness of the bit rate control is a common measure of the effectiveness of the bit rate control module. The bit rate control accuracy is calculated as follows:

$$M = \frac{|B-T|}{T} \times 100\% \tag{14}$$

B is the actual number of output bits, *T* is the set target number of bits, and *M* is the rate control precision error.

4.2. Analysis of Results

Figure 8 shows the difference in the bit allocation of each test sequence compared to the pre-adjustment period. Table 5 shows the coded output bit count and rate control accuracy error of the proposed method with the VTM-14.0 rate control algorithm. Table 6 shows the results of the comparison of the Y-PSNR. Table 7 shows the results of the BD-rate and BD-PSNR results.



Figure 8. Bit tuning for a test sequence with a set maximum bit cost. (a) PeopleOnStreet; (b) BasketballDrive; (c) BQMall; (d) RaceHorses; (e) FourPeople; (f) BasketballDrillText.

Sequences	Target Bitrate	VTM14.0 RC	Bit Error (%)	Proposed Method	Bit Error (%)
PeopleOnStreet	12,000	12,070.56	0.588	12,074.6	0.622
-	30,000	30,083.04	0.279	30,071.3	0.238
	40,000	40,084.8	0.212	40,103.3	0.258
	50,000	50,078.88	0.158	50,089.9	0.180
BasketballDrive	1500	1516.4	1.093	1518	1.200
	2000	2004	0.2	2005.2	0.260
	2500	2477.6	0.896	2473.2	1.072
	3000	2977.6	0.747	2969.2	1.027

Table 5. Comparison results of the proposed method with VTM14.0.

Sequences	Target Bitrate	VTM14.0 RC	Bit Error (%)	Proposed Method	Bit Error (%)
BQMall	20,000	19,884.96	0.575	20,017.9	0.090
	30,000	28,863.36	3.789	29,267.5	2.442
	40,000	39,685.92	0.785	40,003.7	0.009
	50,000	47,296.32	5.407	48,072.5	3.855
RaceHorses	5000	4780.8	4.384	4742.64	5.147
	8000	7029.12	12.136	7042.08	11.974
	10,000	7925.52	20.745	7985.04	20.150
	12,000	8960.16	25.332	9168.96	23.592
FourPeople	5000	5086.08	1.722	5042.4	0.848
-	6000	6011.52	0.192	6025.44	0.424
	8000	7919.52	1.006	7932.48	0.844
	10,000	9825.12	1.749	9845.76	1.542
BasketballDrillText	2000	1942.8	2.86	1978.4	1.080
	4000	3705.6	7.36	3728.8	6.780
	6000	5555.6	7.407	5584	6.933
	8000	7348.4	8.145	7494.8	6.315
Average			4.49		4.037

Table 5. Cont.

Table 6. PSNR Comparison Results.

Sequences	Target Bitrate	VTM 14.0 RC Y-PSNR	Proposed Method Y-PSNR	ΔPSNR
PeopleOnStreet	12,000	32.8213	32.8479	0.0266
	30,000	37.4595	37.4576	-0.0019
	40,000	38.9225	38.9233	0.0008
	50,000	40.1155	40.1172	0.0017
BasketballDrive	1500	31.9553	31.899	-0.0563
	2000	33.0822	33.1002	0.018
	2500	33.9155	33.9622	0.0467
	3000	34.6552	34.6369	-0.0183
BQMall	20,000	40.4325	40.4504	0.0179
	30,000	42.6537	42.7425	0.0888
	40,000	44.5173	44.5839	0.0666
	50,000	45.8516	46.002	0.1504
RaceHorses	5000	42.7031	42.7139	0.0108
	8000	46.2087	46.3002	0.0915
	10,000	47.5341	47.6092	0.0751
	12,000	48.8587	49.1383	0.2796
FourPeople	5000	35.0326	34.9933	-0.0393
	6000	36.0668	36.1002	0.0334
	8000	37.7986	37.8055	0.0069
	10,000	39.1359	39.139	0.0031
BasketballDrillText	2000	31.6835	31.8176	0.1341
	4000	34.928	34.9855	0.0575
	6000	36.8878	36.9174	0.0296
	8000	38.2579	38.3774	0.1195
Average				0.05

Figure 8 shows the difference in bit allocation for each CTU in the six test sequences compared to the pre-adjustment one, using the black dashed line. The horizontal coordinate in the figure is the CTU serial number and the red solid line is Y = 0. This is used to see whether the bit cost is enhanced or weakened for the adjusted CTUs compared to the default algorithm. In the figure, it can be clearly observed that, through the joint guidance of the two significance detection algorithms and a complexity calculation method, and excluding the influence of image noise, only when the significance and complexity are within the limit boosted, it makes the allocation of the bit cost of the "really important" CTUs more even

compared with the default algorithm. At the same time, due to the nature of the default bit allocation algorithm, which allocates the remaining CTUs at the remaining bit cost, the CTUs at the end of the frame will have fewer CTUs to allocate, and a more balanced situation is reached, which is characterized by much less than 0 in the image. In Table 5, although the error is higher than the default algorithm in some conditions, most of the results are better than the default algorithm, which reduces the error from 4.49% to 4.037% on average; similarly, the average Y-PSNR is improved by 0.05 dB in Table 6. Most of the data in Table 6 performs well, but there are still some negatively optimized data, such as when the cost is 30,000 kbps in PeopleOnStreet, when the cost is 1500 and 3000 kbps in the BasketballDrive sequence, and so on. After our analysis, the reason for this phenomenon is partly due to the characteristics of the sequences; sequences similar to the PeopleOnStreet and BasketballDrive categories have complex content and fast movement of the character scenes, and therefore are more difficult to encode, which may lead to similar situations; and secondly, it may be due to the fact that fewer combinations of parameters were used to determine the optimal tuning multiplicity for each test sequence. Secondly, this may be due to the fact that the combination of parameters used to determine the optimal adjustment multiplier for each test sequence was small, and not entirely precise to the combination that would give the best results, so this phenomenon may be elaborated. In Table 7, we are comparing the average reconstructed pixel quality under the same bit cost improved by 0.03 dB, and the consumed bit cost can be reduced by 0.33% with the same reconstruction quality, which is a better result than the default rate control, and it can be shown that this model can fulfill the rate control task with improved efficiency.

Table 7. BD-rate and BD-PSNR results.

Sequences	BD-PSNR (dB)	BD-Rate (%)	
PeopleOnStreet	0.00658	-0.11867	
BasketballDrive	0.01045	-0.25067	
BQMall	0.00655	-0.08705	
RaceHorses	0.10506	-0.87276	
FourPeople	0.00560	-0.09022	
BasketballDrillText	0.02654	-0.54218	
Average	0.03	-0.33	

5. Conclusions

In this paper, a bit allocation scheme based on saliency and coding complexity is proposed in accordance with the VVC standard, taking into account the unevenness of bit allocation at the CTU level in the intra-rate control scheme. The construction is completed through a fusion operation that incorporates a color contrast-based significance detection algorithm, a Canny-based significance detection algorithm, a self-constructed video dataset, and a PLS principal component regression model. The findings of this study indicate that intra-rate control is enhanced by the CTC test conditions for rate control on official test sequences; the average rate control error is decreased to 4.037%, the average BD-rate is decreased by 0.33%, and the average Y-PSNR is enhanced by 0.03 dB. This outcome subtly illustrates the potential benefits and feasibility of utilizing this model to resolve analogous challenges in subsequent endeavors. The experimental procedure and results indicate that the model still has space for improvement, including the method for dividing the number and type of videos in the self-built dataset, as well as further refinement of the video enhancement of the dataset and a reduction in the multiplier. In addition to this, the lack of publicly available datasets of image-mapped video sequences poses some challenges for the experiments. In order to continue to validate the model's adaptability and application value in different video categories, it will continue to expand the video dataset as well as the corresponding multiplicity, and design more excellent bit allocation schemes. It is hoped that this research can provide good suggestions and new ideas for researchers. In future work, this paper will further focus on the rate control scheme of the VVC standard

and seek to find innovative work on the basic algorithms in different coding modes as well as in more domains.

Author Contributions: Conceptualization, H.S.; methodology, H.S.; software, H.S.; validation, H.S.; formal analysis, H.S.; investigation, H.S.; resources, H.S.; data curation, H.S.; writing—original draft preparation, H.S.; writing—review and editing, H.S., X.J. and Y.Z.; visualization, Y.Z.; supervision, H.S.; project administration, X.J.; funding acquisition, X.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Collaborative Innovation Achievement Program of Double First-class Disciplines in Heilongjiang Province grant number LJGXCG2023-074.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the fact that the dataset production is part of the innovation point of this paper, the dataset will not be open-sourced at the first time, we will continue to explore new possibilities in the subsequent experimental research, and then the dataset will be made public along with the related code.

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

References

- 1. Statistical Report on the Development of Internet in China; China Internet Network Information Center (CNNIC): Beijing, China, 2023.
- Zhou, Y.; Hu, X.; Guo, X. Research on Key Technologies of VVC Intraframe Prediction. *Radio TV Broadcast Eng.* 2019, 46, 62–69. [CrossRef]
- Zhu, X.; Tang, G. Progress in Learning Based Video Coding Technology. J. Nanjing Univ. Posts Telecommun. (Nat. Sci. Ed.) 2022, 42, 1–12. [CrossRef]
- 4. Li, B.; Li, H.; Li, L.; Zhang, J. Rate Control by R-Lambda Model for HEVC. ITU-T/ISO/IEC JCT-VC Document JCTVC-K0103. 2012; Volume 10.
- 5. Liu, Y. Research on H.266/VVC complexity analysis and inter frame algorithm optimization. *Taiyuan Univ. Technol.* 2020. [CrossRef]
- 6. ITU-T Recommendation H.261; Video Codec for Audiovisual Services at px64 Kbit/s. Geneva, Switzerland, 1990.
- ITU-T and ISO/IEC JTC 1; Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video, ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2 Video), version 1. 1994.
- 8. Rijkse, K. 263: Video coding for low-bit-rate communication. IEEE Commun. Mag. 1996, 34, 42–45. [CrossRef]
- 9. Richardson, I.H. 264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia; John Wiley & Sons: Hoboken, NJ, USA, 2004.
- 10. Sullivan, G.; Ohm, J.; Han, W. Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.* 2012, 22, 1649–1668. [CrossRef]
- 11. Versatile Video Coding Editorial Refinements on Draft 10 [EB/OL]. Available online: https://jvet-experts.org/doc_end_user/ current_document.php?id=10540 (accessed on 1 January 2023).
- Saldanha, M.; Sanchez, G.; Marcon, C. Complexity Analysis of VVC Intra Coding. In Proceedings of the 2020 IEEE International Conference on Image Processing, Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 3119–3123.
- 13. Bross, B.; Wang, Y.; Ye, Y. Overview of the Versatile Video Coding (VVC) Standard and its Applications. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 3736–3764. [CrossRef]
- Tsai, J.C.; Shieh, C.H. Modified tmn8 rate control for low-delay video communications. *IEEE Trans. Circuits Syst. Video Technol.* 2004, 14, 864–868. [CrossRef]
- Li, Z.; Pan, F.; Lim, K.; Feng, G.; Lin, X.; Rahardja, S. Adaptive Basic Unit Layer Rate Control for jvt (jvt-g012). In Proceedings of the Joint Video Team (JVT) 7th Meeting, Pattaya, Thailand, 7–14 March 2003.
- 16. Li, Y.; Chen, Z. Rate control for vvc. In Proceedings of the Joint Video Experts Team (JVET) 11th Meetting, Ljubljana, Slovenia, 10–18 July 2018.
- 17. Lee, H.; Chiang, T.; Zhang, Y. Scalable rate control for MPEG-4 video. IEEE Trans. Circuits Syst. Video Technol. 2000, 10, 878–894.
- 18. Wu, W.; Liu, J.; Feng, L. Novel rate control scheme for low delay video coding of hevc. ETRI J. 2016, 38, 185–194. [CrossRef]
- 19. Choi, H.; Nam, J.; Yoo, J. Rate Control Based on Unified RQ Model for HEVC. ITU-T/ISO/IEC JCT-VC Document JCTVC-H0213. 2012; Volume 2.
- 20. Li, B.; Zhang, D.; Li, H. QP Determination by Lambda Value. ITU-T/ISO/IEC JCT-VC Document JCTVC-I0426. 2012; Volume 5.
- 21. Karczewicz, M. Intra Frame Rate Control Based on Satd. In Proceedings of the Joint Collab—Orative Team on Video Coding (JCT-VC) 13th Meetting, Incheon, Republic of Korea, 18–26 April 2013.

- 22. Guo, H.; Zhu, C.; Li, S. Optimal bit allocation at frame level for rate control in HEVC. *IEEE Trans. Broadcast.* **2018**, 65, 270–281. [CrossRef]
- Chen, Y.; Kwong, S.; Zhou, M. Intra frame rate control for versatile video coding with quadratic rate-distortion modelling. In Proceedings of the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 4422–4426.
- 24. Li, Y.; Liu, Z.; Chen, Z. Rate Control for Versatile Video Coding. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 1176–1180.
- Bai, L.; Song, L.; Xie, R. saliency based rate control scheme for high efficiency video coding. In Proceedings of the 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, Republic of Korea, 13–16 December 2016; pp. 1–6.
- Luz, G.; Ascenso, J.; Brites, C. Saliency-driven omnidirectional imaging adaptive coding: Modeling and assessment. In Proceedings of the 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Luton, UK, 16–18 October 2017; pp. 1–6.
- Milani, S.; Bernardini, R.; Rinaldo, R. A saliency-based rate control for people detection in video. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 2016–2020.
- Pelurson, S.; Cozanet, J.; Guionnet, T. AI-Based Saliency-Aware Video Coding. SMPTE Motion Imaging J. 2022, 131, 21–29. [CrossRef]
- 29. Li, W.; Jiang, X.; Jin, J. Saliency-Enabled Coding Unit Partitioning and Quantization Control for Versatile Video Coding. *Information* **2022**, *13*, 394. [CrossRef]
- Yang, Z.; Luo, Y.; Lin, Y. Convolutional neural network-based optimal R-λ intra rate control in Versatile Video Coding. J. Electron. Imaging 2022, 31, 063011. [CrossRef]
- Bossen, F. common conditions and software reference configurations (jvt-j1100). In Proceedings of the Joint Collaborative Team on Video Coding (JCT-VC) 8th Meeting, San Jose, CA, USA, 1–10 February 2012.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.