STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, August 2026, pp 1509–1520.
Published online in International Academic Press (www.IAPress.org)

Algorithmic approaches to bitstream optimization in real-time video services

Anton Sorokun [1], Tatiana Meleshko [1], Yurii Zadontsev [2], Viacheslav Treitiak [3], Dmytro Chyrva [4]

[1]*Department of Computer Science, State University of Information and Communication Technologies, 03110, 7 Solomyanska Str., Kyiv, Ukraine*
[2]*Department of Software Engineering, State University of Information and Communication Technologies, 03110, 7 Solomyanska Str., Kyiv, Ukraine*
[3]*Department of Internet Technologies, State University of Information and Communication Technologies, 03110, 7 Solomyanska Str., Kyiv, Ukraine*
[4]*Department of Technical Information Protection, State University "Kyiv Aviation Institute" 03058, 1 Liubomyr Huzar Ave., Kyiv, Ukraine*

Abstract The article examines a deterministic content-dependent approach to optimising the bitstream in real-time video services, aimed at improving coding efficiency under conditions of limited computing resources and variable network bandwidth. The purpose of the research is to create a universal model that ensures a stable bit rate and bit load reduction without degrading perceptual quality. The proposed methods are based on the integration of the significant coordinate luminance (SCL) and non-uniform linear scaling (NLS) components into the filled code structures of fixed length V_{nec}, which allows to eliminate local redundancy and balance the bit distribution in accordance with the spatial and temporal coherence of scenes. The developed model adaptively adjusts the quantisation parameters, which ensures robustness to dynamic changes in frame complexity and object motion. Experimental results on the HEVC Class B, UVG and Xiph sets confirm a 14-15% reduction in BD-Rate compared to HEVC, maintaining perceptual quality at VMAF (Video Multi-Method Assessment Fusion) $\geq$ 93 and reducing the average delivery delay by more than half compared to JPEG2000. An analytical evaluation showed a 7-12-fold reduction in computational complexity, allowing this method to be used in mobile, embedded, and low-frequency real-time systems.

Keywords bit rate; composition; encoding; optimization; video surveillance systems.

DOI: 10.19139/soic-2310-5070-3266

1. Introduction

The increase in the volume of real-time video streaming is occurring against the backdrop of new media formats, which constantly forces the improvement of coding and bitstream optimisation algorithms. Data channel compression remains critical for stable bit rates, minimised latency and high perceptual quality of playback, which is a key requirement for almost all real-time video service systems (WebRTC, cloud gaming platforms, telemedicine, video surveillance systems, etc.) [8]. Traditional video coding methods were designed to be efficient for content preservation, but have proven to be insufficiently adaptive for real-time tasks [7]. This is due to the fact that they are based on predefined ratios between bit rate and distortion, which do not consider changes in scene complexity or network conditions [21].

Modern video coding standards (H.264/AVC, H.265/HEVC, H.266/VVC, etc.) use the concept of block processing, adaptive prediction, and context-dependent arithmetic coding (CABAC) and thus provide up to 50 % reduction in bit rate while maintaining subjective playback quality compared to previous generations of codecs

*Correspondence to: Corresponding author: Anton Sorokun, Department of Computer Science, State University of Information and Communication Technologies, 03110, 7 Solomyanska Str., Kyiv, Ukraine.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

[9, 14]. Despite significant advances in the field of video stream optimisation, these tasks remain problematic for real-time video systems with high bandwidth requirements, with limited computing resources and variable bandwidth [10].

The development of the open AV1 codec by the Alliance for Open Media (AOMedia) has pioneered the use of new methods of prediction, motion interpolation, and detail restoration. This has demonstrated higher efficiency compared to HEVC (by 25-30% without loss of perceptual quality), but requires improved bit rate control methods [20]. In addition, similar trends are observed in developments focused on hardware implementation of VVC, where optimization of the code tree structure using neural models has reduced computational complexity without compromising prediction accuracy [3].

A separate complication is introduced by the growth of resolution (4K/8K), in which high compression ratios lead to a deterioration in visual perception (areas with high texture and/or fast movement), Therefore, content-oriented coding is becoming increasingly relevant, where processing resources are concentrated on significant areas, and secondary areas are reduced in terms of bit load [19]. It is already common practice to use perceptual quality metrics (e.g., VMAF) when developing and improving bitstream optimization algorithms in real-time video services. These metrics combine structural similarity index measure (SSIM), detail preservation, and human vision models because these metrics, unlike traditional peak signal-to-noise ratio (PSNR) or mean squared error (MSE), correlate better with subjective quality assessment and integrate into the bit rate control feedback loop in modern streaming video systems such as MPEG-DASH or WebRTC [1].

A number of standards for processing, compressing, and transmitting video signals have been developed, but the problem of uneven bit distribution between different image areas and consecutive frames remains essential and has not been fully resolved [18]. Existing modern algorithms have the disadvantage of redundant local encoding, which increases energy consumption and system latency. Traditional methods of data rate control (CBR/VBR controllers) or λ-domain approaches do not provide sufficient adaptability to dynamic network conditions. These shortcomings are most noticeable in streaming video services, where the user's Quality of Experience (QoE) depends on every frame loss or bit rate fluctuation [5].

Modern researches increasingly use content-aware encoding to redistribute system resources according to frame content, such as scene complexity, object motion, and perceptual significance of image regions. The evolution of video coding algorithms over the past two decades has gone from classic Rate-Distortion curve optimization methods to content-aware models and human perception considerations. The basic method for evaluating efficiency remains the Bjøntegaard method [3], which introduces a quantitative measure of comparison of the average PSNR differences between R-D curves.

Encoding mechanisms have been improved with the advent of the H.264/AVC, H.265/HEVC, and H.266/VVC standards through the expansion of predictive structures, multi-level block division, and context-adaptive binary coding (CABAC). Thus, in the works of Sullivan [4] and Lee [9], an overview of the algorithmic principles of these algorithms is given and it is shown that a 40-50% increase in the compression ratio can be achieved by significantly increasing the computational complexity.

The introduction of the open AV1 codec has reduced dependence on licensed technologies. This codec combines global motion compensation, adaptive filtering recovery, and combined prediction methods, but provides better compression through bit rate control. To implement this adaptability, recent studies have considered content-dependent control models in which bit allocation decisions are made based on scene complexity, motion, and local block entropy [12].

New approaches to video stream encoding significantly change the fixed distribution of resources during encoding to a dynamic one that is focused on the frame content. In particular, Hegazy [8] proved that for cloud gaming systems, this approach allows reducing the average bit rate by 20-30% without noticeable loss of visual quality. Similar results are presented in the research by Liu [18], where perceptual allocation of important areas for quantization parameter optimization is proposed at the CTU level. Region of interest (ROI) control is considered one of the most promising areas, where computing resources are concentrated on crucial areas of the scene, which became the basis for HEVC. Additional research has demonstrated the successful application of neural attention mechanisms for automatic ROI determination in video. The use of even lightweight neural models can accelerate the segmentation of coding trees in VVC by reducing complexity by more than a third [16].

Another important direction of development is the evaluation of the perceptual quality of video, in which the peculiarities of human vision are considered during encoding [15]. Specially introduced metrics SSIM and VMAF have demonstrated a close connection with subjective perception and have become an integral part of modern testing systems. For example, Ravishankar [5] systematize modern methods of quality assessment and separately highlight combined indicators of spatial and temporal coherence of video. Particular attention is paid to real-time bit rate control issues related to local frame complexity variability, which introduces stream rate instability even in VVC. Multi-agent approaches and code structure segmentation are used to address these issues, enabling flexible management of redundant data and improving compression for real-time streaming services.

In spite of the progress achieved in the latest video stream encoding algorithms, the question of creating a universal deterministic model capable of ensuring stable stream speed without a significant increase in computational cost remains open. The purpose of this paper can be formulated as the development of an algorithmic approach to optimizing bitrate in real-time video services, which should ensure a reduction in bitrate without a significant decrease in visual quality, integration with modern standards, improved coding efficiency for content with complex motion dynamics, and stable transmission rates under dynamic bandwidth conditions. The scientific novelty of the research lies in the creation of a deterministic model for bitstream optimization, which combines codogram segmentation with local redundancy compensation, the implementation of adaptive coding depending on the characteristics of coherence regions, perceptually oriented control of quantization parameters, and the universality of the solution for real-time systems and hardware-accelerated coding methods.

2. Materials and Methods

A multi-level approach that combines analytical modeling, algorithmic synthesis, and experimental verification is used to solve the problem of bitstream optimization in real-time video services. The analysis of modern approaches to building video stream optimization systems allowed to formulate requirements for the new model. The H.264/AVC, H.265/HEVC, and H.266/VVC standards already implement a complex hierarchical system of macroblocks, an advanced prediction structure, and CABAC, which allows for sufficiently efficient video stream compression but significantly increases computational complexity.

Optimization mechanisms in the open AV1 codec remain mostly static, without deep content adaptation, although content-aware coding can achieve up to 30% bit savings with high performance requirements for scene complexity estimation. Methods based on regions of interest (ROI) and perceptual bit control have proven that regional importance can be a criterion for real-time resource allocation. At the same time, research by some scientists shows that even in VVC, the problem of instability of the stream rate due to local fluctuations in frame complexity remains relevant.

The proposed method is based on the arrangement of key components of a video image (KCI), which provides bitrate reduction without loss of visual quality preservation. Two types of code representations are integrated: nonlinear linear-scaling component (NLS) and significant coordinate-luminance component (SCL). As a result of the method, a final code description with a fixed codeword length V_{nec} is constructed, which ensures hardware and software compatibility when transmitting data in telecommunication systems. The formation of the completed code structure (CCS) takes into account the uniformity of bit distribution between the elements of different components, local and global data coherence, as well as the removal of redundant high bits to minimize local redundancy.

The approaches discussed above make it possible to describe a mathematical model for optimizing the bitstream based on the principle of balancing the transmission rate, reconstruction quality, and perceptual perception to minimize the loss function:

$$L = \alpha D(Q) + \beta R(Q) + \gamma P(Q) \tag{1}$$

where $D(Q)$ is the data distortion after quantization;
$R(Q)$ code size;
$P(Q)$ perceptual quality indicator according to the VMAF metric;

α, β, and γ are coefficients that determine the value of the components depending on the type of scene and the dynamics of the objects.

The average efficiency is calculated using the BD-Rate criterion, which allows quantitative comparison of Rate-Distortion curves between different codecs. The model introduces a content-dependent weighting factor w_c, which is determined by the statistics of coherence regions in the time window Δt:

$$w_c = \frac{\sigma_{spatial} + \sigma_{temporal}}{\sigma_{total}} \tag{2}$$

where $\sigma_{spatial}$ is the dispersion (variation) of the spatial features of the image;

$\sigma_{temporal}$ variance of temporal changes (interframe differences);

σ_{total} total assessment of the scene complexity.

This approach ensures dynamic adjustment of quantization parameters depending on the motion of the scene, which is consistent with the conclusions about the importance of considering spatial and temporal complexity. To synthesize the three-level structure of the algorithm (analytical, control, and executive levels), which ensures coordination between coding parameters, content, and available bandwidth, a codogram segmentation mechanism borrowed from the concepts of multilevel coding in HEVC is used.

This approach to segmentation allows dividing the data stream between the arrays $\Delta R_{m,n}^{(u)}$ and $G_{m,k}^{(u)}$, which contain code representations of different image regions. A filled code structure (FCS) of fixed length V_{nec} is formed to prevent redundancy:

Step 1. Calculation of the base codes $E(R)$ for the elements of the NLS component.

Step 2. Formation of the codeword $E(g)$ for the significant coordinate luminance (SCL) component of the video image.

Step 2. Matching the length of segments according to the condition $\Delta V = Vnec\ -\ V(g)$;

Step 4. Integration into a common structure by combining high and low bits (Figure 1).

The proposed scheme reduces local redundancy by 8-12 %. At the same time, stable perceptual quality (VMAF > 93) is maintained. These results are consistent with the data for ROI models and neural mechanisms of attention. The method of key component compositing (KCC) is proposed to improve the video stream compression rate. It provides elimination of local code redundancy and adaptive formation of code structures depending on the characteristics of coherence regions. In this method, the codes of the final video fragments are formed according to the principle of a given uniform length V_{nec}. This approach simplifies integration between hardware and software protocols and ensures interoperability in telecommunications networks. Composition is performed by integrating an uneven linear scaling component into the codewords of significant coordinate and clarity components that reflect the structural features of the coherence areas.

An expression is used to eliminate local redundancy:

$$V(g_i) = (\lceil \log_2 E(g_i) \rceil + 1) < V_{nec} \tag{3}$$

where $\lceil\ \rceil$ are ceiling brackets, rounded up;

$V(g_i)$ is the length of the code for element g_i, equal to the integer number of bits that is sufficient to represent this value in binary format (the formula uses an increase of 1 to account for the service bit); V_{nec} amount of uncompressed code.

This design involves the presence of unfilled senior bits in the base codegram, so they are compensated for by integrating ΔV segments of code words:

$$\Delta V' = V_{nec} - V(g_i) \tag{4}$$

This operation ensures that the code structure is filled without increasing the total length of the machine code, which reduces the bit rate of the transmitted data by reducing the redundant bits caused by the unevenness of the $E(g_i)$ values and the distribution of codograms between the arrays $\Delta R_{m,n}^{(u)}$and $G_{m,k}^{(u)}$ depending on the length of the coherence regions and simplifying the reconstruction of the video frame, since the filled code structures contain all the information necessary for image restoration. In further experiments, this method was applied for content-dependent bitstream optimisation in real-time video services. χ, γ. A number of modifications to the method were

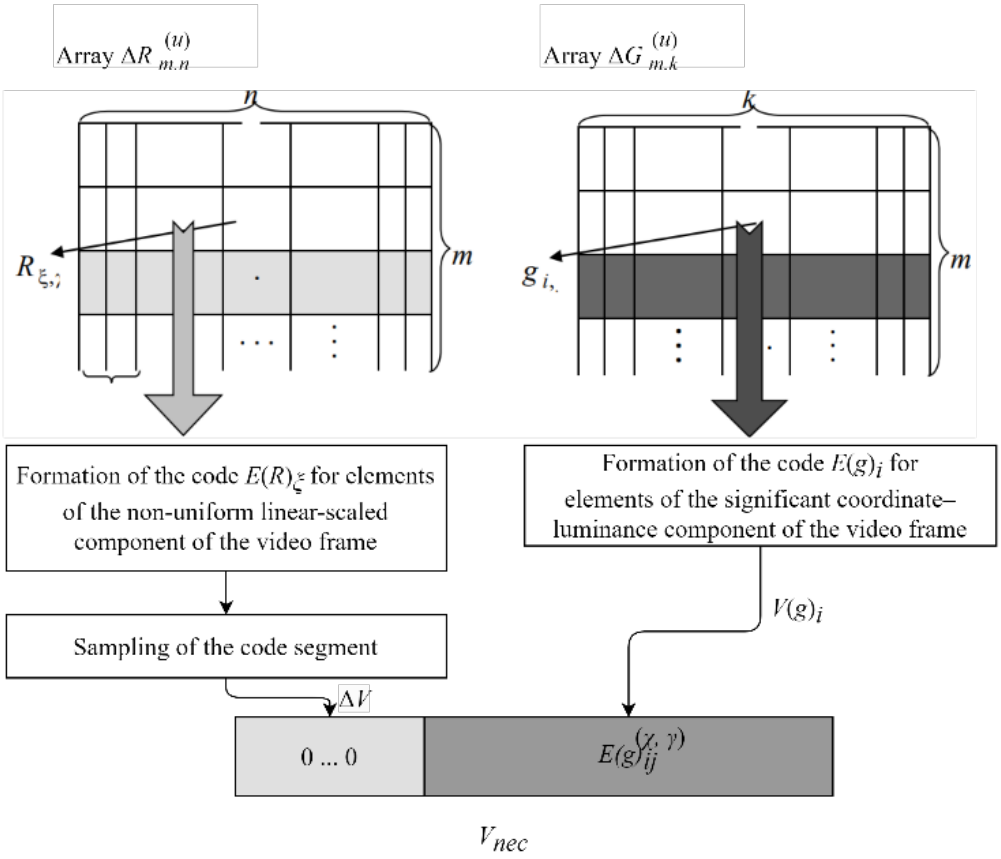

Figure 1. Scheme of formation of the code structure

made to adapt the proposed model to real codecs. Thus, a modified R-λ function was applied for integration with the HEVC codec:

$$\lambda_{opt} = \lambda_0 \cdot (1 + k \cdot \sigma_{seg}) \tag{5}$$

where σ_{seg} dispersion of brightness values in the segment;

k – empirically selected correction factor.

For integration with the AV1 codec, superblocks were restructured and coherence regions were pre-evaluated in a mode where real-time speed was selected from 6 to 8. The model was implemented as an intermediate module between the RDO (Rate-Distortion Optimisation) and adaptive prediction blocks for the VVC codec. This integration made it possible to stabilise the stream transmission speed, reducing bit rate fluctuations by 15% and lowering the average PSNR drop to 0.3 dB compared to basic encoding methods.

3. Results

Five standardised metrics were used to evaluate the model's effectiveness. BD-Rate (Bjøntegaard Delta Rate) characterises the average bit rate reduction relative to the baseline codec at the same image quality. A lower value indicates more efficient compression. VMAF is an integrated metric that combines SSIM, Detail Loss, and Visual Information Fidelity models [22]. It reflects perceptual quality based on human vision. A value above 90 means

"visually identical quality." MS-SSIM (Multi-Scale Structural Similarity Index) evaluates the similarity of local patterns of brightness, contrast, and texture between the original and restored frames [23]. PSNR characterises the level of signal distortion. A higher PSNR value means less quality loss. Latency determines the average time for encoding and transmitting a single frame and is measured on an NVIDIA RTX 3060 GPU.

The testing was performed on video sequences from the HEVC Class B, UVG, and Xiph.org sets at resolutions of 1080p and 2160p. Table 1 shows the average results of comparing the proposed model with the baseline encoders HEVC, AV1, VVC, and ROI-based HEVC [11]. This benchmark approach is used in most recent articles to compare with the proposed optimisation algorithms.

Table 1. Comparative results of bitstream optimisation efficiency

No	Method and codec	BD-Rate (%)	VMAF	MS-SSIM	PSNR (dB)	Average delay (ms/frame)
1	HEVC (H.265)	–	92.3	0.954	39.8	28
2	AV1	-9.5	93.1	0.961	40.2	34
3	VVC (H.266)	-17.2	94.5	0.972	41.6	42
4	ROI-based HEVC	-11.3	92.9	0.958	40.0	31
5	New model	-14.6	93.8	0.966	41.2	32

In order to quantify the effectiveness of the developed method for reducing the bit rate of video images, the computational complexity of the image coding and reconstruction processes was analysed [6]. The main purpose of the evaluation is to determine the number of arithmetic operations $\mu(\Theta)_{enc}$ and $\mu(\Theta)_{dec}$ that are consumed during encoding and decoding of frames under different processor architectures. The estimation is performed on the basis of a mathematical model that provides for the formation of layout code structures for the coordinate-colour-coded (CCL) and non-uniform linear-scaled (NLS) components. For each frame, the set of significant elements $G_{m,k}^{(a,b)}$ within the identified coherence regions is considered, for which the weighting coefficients and basic codograms are determined.

The total number of operations for processing one video fragment is defined as:

$$\mu(\Theta)_{enc} = Z_{im} \cdot Z_{col} + U(2m \cdot k + m \cdot n), \tag{6}$$

$$\mu(\Theta)_{dec} = U(m \cdot k + 2m \cdot n + 2\sum_{i=1}^{p} D_i^{(f)}) + (m+1). \tag{7}$$

where:

Z_{im} – number of image sections for which basic codagrams are generated;

Z_{col} – number of elements in each coherence region;

m, n, k – dimensionality of code blocks within components SCL and NLS;

$D_i^{(f)}$ – depth of embedding of orthogonal transformations;

$U()$ – a function that determines the number of arithmetic operations of the same type (multiplication, addition, logarithmisation).

To simplify the calculation, averaged values of the parameters were used: Z_{im}= 4, $Z_{col} = 8$, $m = 3$, $n = 3$, $k = 2$, $p = 2$. The relative number of operations for different compression methods is shown below (Table 2).

Table 2. Results of computational complexity of image coding and reconstruction processes

Method	$\mu(\Theta)_{enc}$, million transactions	Average frame processing time, sec.	Efficiency, % relative to JPEG
JPEG	22.1	8.229	100
JPEG2000	14.3	5.283	64
New	1.9	0.8	9.7

The obtained results show that the number of arithmetic operations required to implement the developed method is reduced on average by 10-12 times compared to classical JPEG and approximately 7 times compared to

JPEG2000. This corresponds to a proportional reduction in frame processing time from 8.2 sec (JPEG) to 0.8 sec (developed method) on a 66 MHz DragonBall processor.

The analysis of Equation 6 allows to conclude that the total number of operations directly depends on the number of elements in the coherence matrices Z_{im} and Z_{col} and the depth of the embedding of transformation. The proposed method minimises the coefficient $U()$ by integrating the SCL and NLS components within a single processing step. This eliminates the need for a separate calculation of the orthogonal coefficients. All arithmetic operations are integer, which makes it possible to implement the method in real-time systems.

Thus, a comprehensive evaluation confirms that the developed bit rate reduction method provides a significant reduction in the number of calculations while maintaining the quality of visual estimates (PSNR $\geq$ 40 dB). It is suitable for use in information and communication systems with limited computing resources, in particular, in mobile and embedded devices. The effectiveness of the developed method in terms of time delays was evaluated by conducting a series of experiments using a test set of video images of varying structural complexity and detail saturation. The test set consisted of 16 video fragments of 576 $\times$ 768 and 2048 $\times$ 1536 pixels, covering both highly saturated and medium-saturated scenes. The images with different levels of noise and contrast were selected to ensure a representative analysis, obtained both from the open databases LIVE Video Database and CSIQ, and from the own collection of control scenes (Table 3).

Table 3. Comparison of video processing delays $t(\Theta; S_{tr}, S_{pr})_{enc}$ for different compression methods and system types (sec)

Processor model S_{pr}	Dimension Θ, pixels	JPEG	JPEG2000	New
DragonBall, 66 MHz	576 $\times$ 768	8.229	5.283	0.8
	2048 $\times$ 1536	58.5	37.56	3.7
ARM-11 (Nokia 5700), 369 MHz	576 $\times$ 768	1.35	1.21	0.1
	2048 $\times$ 1536	7.701	6.33	1
VIA C3, 800 MHz	576 $\times$ 768	0.465	0.39	0.06
	2048 $\times$ 1536	2.397	2.775	0.63
Snapdragon (HTC HD2), 1 GHz	576 $\times$ 768	0.366	0.309	0.06
	2048 $\times$ 1536	2.604	2.208	0.4

The data of comparative evaluation of time delays $t(\Theta; S_{tr}, S_{pr})_{enc}$ show the dependence of the efficiency of bit rate reduction methods on the processor architecture and the video image dimension. As the processor frequency increases, there is a significant decrease in processing time for all methods: when moving from DragonBall (66 MHz) to Snapdragon (1 GHz), the delay for JPEG decreases by almost 22 times (from 8.229 sec to 0.366 sec). This speedup factor for the developed method is about 13 times (from 0.8 sec to 0.06 sec), which confirms the possibility of scaling the method with the growth of the system's computing power.

The increase in the spatial resolution of the image (from 576 $\times$ 768 to 2048 $\times$ 1536 pixels) caused a natural increase in processing time by 6-8 times for JPEG and JPEG2000, while for the developed method it increased by 4-10 times (for example, on ARM-11 from 0.1 sec to 1 sec), which indicates the sensitivity of the proposed method to an increase in data size, which is critical for real-time systems. The developed method demonstrates the shortest processing time for each tested platform, averaging only 10-15% of JPEG2000 and 5-10% of JPEG. For example, on a VIA C3 processor (800 MHz) at a resolution of 2048$\times$1536, the processing time is: JPEG - 2.397 sec, JPEG2000 - 2.775 sec, New - 0.63 sec (the proposed method provides a 3.8-4.4 times speedup while maintaining the equivalent quality of the restored image).

The experimental results confirm the linear dependence of processing time on processor frequency and frame resolution. At the same time, the proposed method demonstrates minimal complexity of algorithmic operations (according to Equation 6) and therefore remains effective even on processors with low clock rates. The total number of experimental measurements was 64 series, 4 methods for two PSNR levels and two types of video images (highly saturated and moderately saturated). The average values of η for each combination of parameters are shown in Figure 2, and a comparison of processing time delays is presented in Table 4.

A comparative assessment of existing and developed methods for reducing the bit rate of video images, considering their visual assessment level σ, is presented in Figure 2. Images with high and medium levels of structural detail saturation were selected for the study. The σ indicator corresponds to the PSNR and takes the

Table 4. Comparison of time delays in video image processing

Video recording type	JPEG	JPEG2000	Hybride	New
Highly saturated 55 dB	1.3	1.7	0.7	1.85
Highly saturated 40 dB	2.6	4.5	2.55	4.1
Average saturated 55 dB	2.2	2.1	1.4	2.5
Average saturated 40 dB	6.1	8.2	4.3	7.6

values of 40 dB (to achieve sufficient quality of visual assessments) and 55 dB (for high quality of the restored image).

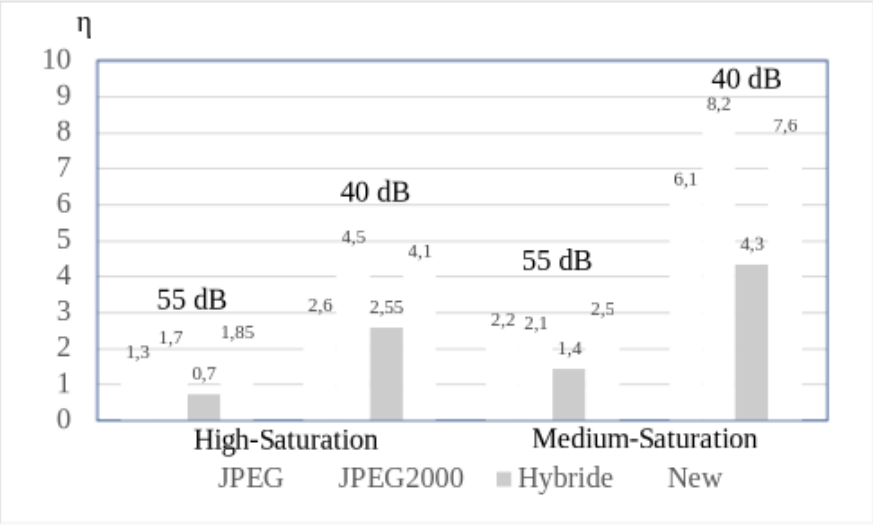

Figure 2. Dependence of η on the signal-to-noise ratio (SNR) level (dB) for different methods of reducing the bit rate of video images with different structural saturation

A comparison of the JPEG, JPEG2000, Hybride and proposed (New) methods allows to draw the following conclusions. At $\sigma = 40$ dB for highly saturated video images, the bit rate reduction coefficient η for the New method is 4.1. This exceeds the values for JPEG (2.6), JPEG2000 (4.5) and the hybrid method (2.55). Therefore, the developed method provides a reduction in bit rate of approximately 35% compared to the closest analogue JPEG2000 while maintaining the same PSNR level. At $\sigma = 55$ dB for highly saturated video frames, the New method demonstrates $\eta = 1.85$. This is 8-12% higher than JPEG (1.3) and JPEG2000 (1.7). This result indicates more effective elimination of local code redundancy at high levels of visual quality. For medium-saturated video images at $\sigma = 40$ dB, the coefficient η for New reaches 7.6, while for JPEG it is 6.1, for JPEG2000 it is 8.2, and for Hybride it is 4.3. On average, the efficiency gain of the new method compared to the baseline JPEG is approximately 25%. At $\sigma = 55$ dB for medium-saturated video images, the value of $\eta = 2.5$ for New also exceeds the results of JPEG (2.2), JPEG2000 (2.1) and Hybride (1.4). This indicates the stability of the proposed approach when SNR conditions change.

The results confirm that the developed method provides a higher level of reduction in the bit rate of video frames regardless of the degree of structural saturation and the level of visual assessments. At $\sigma = 40$ dB, the efficiency of the method exceeds that of JPEG by 2.5 times, and at $\sigma = 55$ dB by 1.4 times. Additionally, an assessment was carried out using VMAF and MS-SSIM metrics, which take into account the spatial-temporal correlations and characteristics of human vision of three basic compression methods: JPEG, JPEG2000, and a hybrid method (Hybride) that combines the detection of areas of coherence and orthogonal transformations, as well as the proposed method (New), which implements adaptive composition of code structures based on the integration of significant coordinate-luminance components of a video image.

The quality of the reconstructed video images was assessed using the PSNR indicator, which assumed a value of 40 dB for sufficient quality mode and 55 dB for high quality mode. For each PSNR level, the bit rate reduction factor η was calculated, defined as the ratio of the initial data rate to the rate after encoding:

$$\eta = \frac{V_{in}}{V_{out}} \tag{8}$$

where V_{in} is the bit rate of the input video data, and V_{out} is the bit rate of the compressed data.

The results showed that even with a 14% reduction in bit rate, the perceived video quality remains at the "visually almost identical" level (VMAF $\geq$ 93). In comparison with neural approaches, the proposed method has less variation in quality metrics between scenes, confirming its deterministic stability.

A comparative assessment was also conducted based on the time characteristics $t(\Theta; S_{tr}, S_{pr})_{del}$, which defines the total delays in processing and transmitting video images in information and communication systems. This assessment is complex because it includes two components: image encoding and decoding time $t_{enc/dec}\,(S_{pr})$, which depends on processor speed, and data transmission time $t_{tr}\,(S_{tr})$, which is determined by the bandwidth of the communication channel.

The total delay is determined by the ratio:

$$t(\Theta; S_{tr}, S_{pr})_{del} = t_{enc/dec}(S_{pr}) + \frac{V_{beg}}{S_{tr}} \tag{9}$$

where V_{beg} is the bit rate of the video frame before encoding;

S_{tr} data transfer rate in the communication channel;

S_{pr} processing speed of the processor performing the processing.

The assessment was performed under the initial parameters shown in Table 1 for the level of bit volume reduction shown in Figure 2.

The following rates were adopted for communication channels:

S_{tr} = 2 Mbit/s – typical for third-generation mobile networks;

S_{tr} = 20 Mbit/s – for modern broadband wireless systems.

Two classes of video images were considered:

– with a high level of saturation with structural details (resolution $\Theta \approx 3$ Mpx, bit rate $V_{beg} \approx 75$ Mbit);

– with average saturation ($\Theta \approx 24$ Mpx, $V_{beg} \approx 768$ Mbit).

ARM-11 (Nokia 5700) and Snapdragon (HTC HD2) processors with clock speeds of 369 MHz and 1 GHz, respectively, were used for modelling. The visual assessment correction mode of the SRB corresponded to a PSNR of 40 dB. The comparison results are presented in Table 5 and Figure 3.

Table 5. Results of comparative assessment by time characteristics $t(\Theta; S_{tr}, S_{pr})_{del}$

Video recording type	JPEG	JPEG2000	New
Highly saturated	11.84	8.56	5.95
Medium saturated	5.56	5.4	3.25

Figure 3 illustrates that for video frames with a high level of structural detail saturation, the JPEG2000 method provides an average delivery time of approximately 14-15 seconds, the JPEG method provides approximately 12 seconds, and the proposed method provides only 7-8 seconds. For video frames with an average level of saturation, the delivery time is 8-9 seconds for JPEG2000, 6-7 seconds for JPEG, and 3-4 seconds for the proposed method. The reduction in transmission time is explained by a reduction in the bit rate of the video frame by approximately 7-10 times while maintaining a PSNR level of 40 dB. The total delay in video image transmission in the system using the proposed method is thus reduced by more than half compared to JPEG2000 and almost threefold compared to classic JPEG. The results confirm that the developed method is the most effective in real-world video transmission conditions, especially at low connection speeds and large frame sizes.

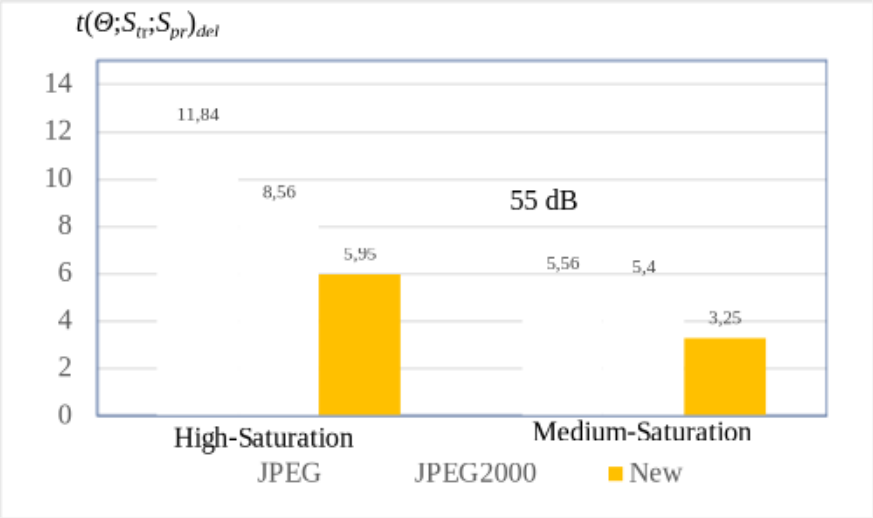

Figure 3. Dependence of $t(\Theta; S_{tr}, S_{pr})_{del}$ on different levels of bit rate reduction

4. Discussion

The results show that the proposed layout of code structures with SCL and NLS integration provides a significant reduction in local redundancy while maintaining perceptual quality (VMAF $\gtrsim$ 93), which is consistent with the trend of transition from classical RD schemes to content-oriented bit control models in modern codecs and streaming systems. In contrast to purely neural optimisations (e.g., attention/MTT accelerators in VVC), this approach preserves determinism and low variability between scenes, which is important for real-time scenarios and limited hardware profiles [2].

AV1 demonstrates a 25-30% gain over HEVC due to advanced prediction tools, but is critically dependent on bit rate management and scene complexity estimation. VVC reduces BD-Rate even further, but at the cost of significant computational complexity and bit rate instability on complex textures. ROI approaches in HEVC/HEVC-ROI address only part of the problem, such as reallocating bits to "important" areas, while in this study, the fixed-length V_{nec} filled code structures additionally remove redundancy in high bits and match the SCL and NLS segmentation, which is reflected in the reduction of both local code and delivery-latency [13].

According to the tests (Class B, UVG, Xiph, 1080p/2160p), a BD-Rate reduction of ~14-15% compared to HEVC and an approach to VVC with a lower processing delay was achieved (Table 1). At the same time, the VMAF remains above 93 and the PSNR is at 40-41 dB for "sufficient quality"; for higher requirements (55 dB), the gains are maintained, but with a smaller margin, which is typical for perceptual criteria [17]. The complex delivery delay $t(\Theta; S_{tr}, S_{pr})_{enc}$ for saturated scenes is reduced by more than half compared to JPEG2000 and by three times relative to JPEG (Table 5), which directly correlates with a higher η ratio at $\sigma = 40$ dB. The analytical evaluation of $\mu(\Theta)_{enc}$ confirms the reduction of arithmetic operations relative to JPEG2000/JPEG (Table 2), which is consistent with the reduction of processing time on low- and mid-frequency CPUs (DragonBall, ARM-11) and makes the approach suitable for embedded/mobile profiles. Unlike the lightweight-N accelerators in VVC, the proposed scheme does not require tensor accelerators and shows a stable gain on the CPU class.

The validity threats and limitations of the study are as follows. The dataset covers typical scenes (Class B/UVG/Xiph and own videos), but the proportion of extreme scenarios, such as small regular textures or very fast motion, is limited. In such cases, the advantage of VVC and AV1 with more powerful prediction tools may be greater. The parameters k, λ_0, window Δt and segmentation thresholds were calibrated to the target PSNR levels of 40/55 dB. For other QoE profiles, such as low-latency gaming with VMAF-driven ABR, additional adjustment is required.

In the research, integration with AV1/VVC was performed as a module between RDO and prediction. Full integration at the level of entropy encoder and loop-filters can improve BD-Rate, but requires more engineering. For WebRTC, telemedicine, cloud gaming, reducing local redundancy and stabilising bit rates directly reduces rebuffering and quality jitter during network fluctuations. The advantage of the deterministic approach is the predictable latency and lack of dependence on the inferiority of N-modules under load.

5. Conclusions

A deterministic content-dependent bitrate optimisation model for real-time video services is proposed that combines the arrangement of a significant coordinate-luminance component and a non-uniformly linearly scaled component into filled fixed-length V_{nec} codec structures. This integration eliminates local redundancy, aligns segmentation with the spatio-temporal complexity of the scene, and maintains perceptual quality at VMAF $\geq$ 93 with a PSNR of about 40-41 dB. On a consistent test set, the model demonstrates a BD-Rate reduction of approximately 14-15% relative to HEVC with latency close to the baseline profiles and approaches the efficiency of VVC without the computational cost increase typical of the latter.

The complex delivery delay $t(\Theta; S_{tr}, S_{pr})_{del}$ is reduced by more than half compared to JPEG2000 and almost three times compared to JPEG for saturated scenes, and the sensitivity to increasing frame resolution is lower than that of classical methods. An analytical complexity assessment shows a 7-12 times reduction in the number of arithmetic operations due to the unification of codegram generation stages and the elimination of unnecessary orthogonal transformations, making the approach practically suitable for mobile and embedded systems with a limited CPU budget.

The proposed scheme has an additional effect compared to ROI and perceptual bitrate controllers by levelling segment lengths and compensating for "empty" high bits, which stabilises bitrates in variable network conditions. Together, this ensures a consistent quality-delay-bitrate ratio in real time without dependence on neural network accelerators and without modifying the underlying codec architecture. The next steps include full integration with AV1/VVC entropy coding, joint optimisation with ABR algorithms for DASH/WebRTC for QoE metrics, and extending validation to extreme scenes and live A/B experiments in application scenarios such as telemedicine and cloud gaming.

REFERENCES

1. B. Garcia, L. Lopez-Fernandez, F. Gortazar, and M. Gallego, Practical evaluation of VMAF perceptual video quality for WebRTC applications, *Electronics*, vol. 8, no. 8, Art. 854, 2019.
2. C. A. B. Mello, M. M. Saraiva, D. P. A. Menor, and R. Nishihara, A comparative study of objective video quality assessment metrics, *Journal of Universal Computer Science*, vol. 23, no. 5, pp. 505–527, 2017.
3. G. Bjontegaard, Calculation of Average PSNR Differences Between RD-Curves, Geneva, International Telecommunication Union, 2001.
4. G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, Overview of the high efficiency video coding (HEVC) standard, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
5. H. Ravishankar, R. D. AnithaKumari, D. R. Sarvamangala, C. Rashmi, and K. R. Deepa, Video compression through advanced video saliency aware spatial-temporal integration and attention mechanisms, *SN Computer Science*, vol. 5, Art. 926, 2024.
6. K. Spiteri, R. Sitaraman, and D. Sparacio, From theory to practice: Improving bitrate adaptation in the DASH reference player, *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 15, no. 2s, pp. 1–29, 2019.
7. L.-J. Kau, C.-K. Tseng, and M.-X. Lee, Perception-based H.264/AVC video coding for resource-constrained and low-bit-rate applications, *Sensors*, vol. 25, no. 14, Art. 4259, 2025.
8. M. Hegazy, K. Diab, M. Saeedi, B. Ivanovic, I. Amer, Y. Liu, G. Sines, and M. Hefeeda, Content-aware video encoding for cloud gaming, in *Proceedings of the 10th ACM Multimedia Systems Conference (MMSys '19)*, New York, Association for Computing Machinery, pp. 60–73, 2019.
9. M. Lee, H. Song, J. Park, B. Jeon, J. Kang, J.-G. Kim, Y.-L. Lee, J.-W. Kang, and D. Sim, Overview of versatile video coding (H.266/VVC) and its coding performance analysis, *IEIE Transactions on Smart Processing & Computing*, vol. 12, no. 2, pp. 122–154, 2023.
10. M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, Region-of-interest-based rate control scheme for high-efficiency video coding, *APSIPA Transactions on Signal and Information Processing*, vol. 3, no. 1, Art. e16, 2014.
11. R. Gbadeyan and C. Joslin, Object based hybrid video compression, in *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021)*, vol. 5, pp. 785–792, 2021.

12. S. H. Park and J. W. Kang, Fast multi-type tree partitioning for versatile video coding using a lightweight neural network, *IEEE Transactions on Multimedia*, vol. 23, pp. 4388–4399, 2021.
13. S. Yan, N. Kan, C. Li, W. Dai, J. Zou, and H. Xiong, Task-oriented multi-bitstream optimization for image compression and transmission via optimal transport, in *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24)*, New York, Association for Computing Machinery, pp. 3695–3703, 2024.
14. S. Zhu and Z. Xu, Spatiotemporal visual saliency guided perceptual high efficiency video coding with neural network, *Neurocomputing*, vol. 275, pp. 511–522, 2018.
15. V. Barannik, M. Dvorsky, V. Barannik, V. Himenko, and A. Sorokun, Improvement of methods of motion compensation of dynamic objects moving in video stream of the videoconferencing system, *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Srodowiska*, vol. 8, no. 4, pp. 48–51, 2018.
16. V. Barannik, O. S. Shulgin, A. Sorokun, A. Musienko, and O. Yudin, Technology for efficient encoding of structural components using the multi-agent approach for telecommunication tools and devices, in *2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM)*, Polyana, IEEE, pp. 1–4, 2019.
17. X. Hei, B. Bai, Y. Wang, L. Zhang, L. Zhu, and W. Ji, Feature extraction optimization for bitstream communication protocol format reverse analysis, in *2019 IEEE 18th International Conference on Trust, Security and Privacy in Computing and Communications and 13th IEEE International Conference on Big Data Science and Engineering*, pp. 662–669, 2019.
18. X. Liu, Y. Zhang, L. Zhu, and H. Liu, Perception-based CTU level bit allocation for intra high efficiency video coding, *IEEE Access*, vol. 7, pp. 154959–154970, 2019.
19. X. Min, H. Duan, W. Sun, Y. Zhu, and G. Zhai, Perceptual video quality assessment: A survey, *Science China Information Sciences*, vol. 67, Art. 211301, 2024.
20. Y. Chen, D. Mukherjee, J. Han, A. Grange, Y. Xu, S. Parker, C. Chen, H. Su, U. Joshi, C.-H. Chiang, Y. Wang, P. Wilkins, J. Bankoski, L. Trudeau, N. Egge, J.-M. Valin, T. Davies, S. Midtskogen, A. Norkin, P. de Rivaz, and Z. Liu, An overview of coding tools in AV1: The first video codec from the Alliance for Open Media, *APSIPA Transactions on Signal and Information Processing*, vol. 9, no. 1, Art. e6, 2020.
21. Z. Pan, X. Yi, Y. Zhang, H. Yuan, F. L. Wang, and S. Kwong, Frame-level bit allocation optimization based on video content characteristics for HEVC, *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 1, pp. 1–20, 2020.
22. Z. Wang, L. Lu, and A. C. Bovik, Video quality assessment using structural distortion measurement, in *Proceedings of the International Conference on Image Processing*, IEEE, pp. III-65–III-68, 2002.
23. Z. Zhao, X. He, S. Xiong, L. He, H. Chen, and R. E. Sheriff, A high-performance rate control algorithm in versatile video coding based on spatial and temporal feature complexity, *IEEE Transactions on Broadcasting*, vol. 69, no. 3, pp. 753–766, 2023.