International Journal of Computer Applications (0975 - 8887)
Volume 187 - No.151, November 2026

ArcEncode: A Content-Aware Adaptive Video Compression Framework for Perceptual Quality and Storage Efficiency

Pronojit Mallick Jit
Broadcast Engineer, Independent Television Limited
Dhaka, Bangladesh

ABSTRACT

Modern satellite television stations, visual media production facilities, OTT platforms and digital video archives handle massive amount of high-volume video files. These files often tend to be in different formats, codecs, resolutions and bit-rates. A large number of these files are also received in raw and uncompressed status. These high-volume of video files require extensive amount of storage which is really expensive and hard to manage. Reducing the file size of these video files without experiencing perceptual quality degradation and encoding latency remains an overwhelming challenge. Standard transcoding workflows commonly apply fixed, content-agnostic encoding profiles to all material, which wastes storage on simple content or degrades complex content. Content-adaptive approaches such as per-title and complexity-based encoding have addressed this for streaming delivery [1, 2], but they are not designed for archival ingest of heterogeneous broadcast masters. This paper presents the ArcEncode software platform, an integrated perceptual video optimization desktop-based application designed for heterogeneous hardware architectures and broadcast engineering requirements. The designed program introduces a multi-stage optimization pipeline containing: automated spatio-temporal visual complexity characterization via discrete Laplacian variance, inter-frame temporal differences, and residual noise estimation. A deterministic recommendation engine is implemented that maps visual complexity metrics, source video stream-characteristics. ArcEncode also implements hardware acceleration APIs to optimal rate-control parameters (CRF/CQP) while preventing re-quantization inflation; dynamic hardware support across NVIDIA NVENC, Intel Quick Sync (QSV), AMD AMF, and multi-core CPU backends; automated closed-loop objective quality assessment combining Netflix VMAF, SSIM, and PSNR; and post-process cryptographic SHA-256 checksums [3] with EBML Matroska packaging [4]. Experimental evaluations across 130 broadcast sequences confirm an aggregate storage reduction of 84.73% (with category-mean reductions ranging between 68.50% and 92.22%) with category-mean VMAF 93.16–96.10, SSIM 0.9844–0.9973, PSNR 35.78–42.33 dB.

General Terms

Algorithms, Measurement, Performance, Reliability.

Keywords

Video Compression Methodology, Perceptual Optimization, Recommendation Engine, VMAF, SSIM, PSNR, HEVC, Hardware Acceleration, Broadcast Engineering.

1. INTRODUCTION

Digital video production facilities, news channels, sporting events, broadcast archives etc. generate massive quantities of high-bitrate and intra-frame mezzanine media files daily. Raw & master formats such as Apple ProRes 422, Avid DNxHD/HR, Sony XAVC-I, Panasonic AVC and other uncompressed broadcast grade video containers preserve pristine quality across multi-generation editing cycles. However, they require substantial network throughput and multi-terabyte storage footprints. HEVC offers substantially better coding efficiency than earlier standards [5, 6], which makes it a practical candidate for re-encoding intra-frame mezzanine material. Long-term video repository preservation requires a more aggressive compression approach. But, media and broadcast workflows enforce rigorous constraints. The output video streams must remain perceptually close to the source, maintain 10-bit color fidelity, safeguard multichannel audio integrity, and provide verifiable cryptographic provenance to be used as on-air ready content and long term archival process. Also, the files need to be less storage consuming and have to be preserved and retained quickly.

In production grade broadcast environments, most video compression pipelines encounter two fundamental operational limitations:

—**Content-Agnostic Parameterization:** Most of the encoding programs or workflows apply unified Constant Rate Factor (CRF) or fixed bitrate targets across all kinds of footages. However, a static studio broadcast stream requires significantly fewer bits to attain perceptual transparency than highly dynamic scenes such as a sporting event or a cinematic clip. When content characteristics are not taken into account, either substantial storage capacity is occupied or noticeable visual degradation (blocking, ringing, blurring and banding), which are well-documented distortions of hybrid block-based codecs [7], is introduced and observed.

—**Fragmentation Across Hardware and Quality Assurance:** Both commercial and open-source transcoding software typically operate in an open-loop manner. They transcode media

files without systematically validating perceptual quality metrics (e.g., VMAF [8] or SSIM [9]). As a result, post-encoded clips lack automated fallback across heterogeneous hardware acceleration APIs (NVIDIA NVENC, Intel QSV, AMD AMF, and multi-threaded CPU). Traditional encoding software also do not generate verifiable cryptographic archival certificates.

To address these challenges, this paper presents an integrated perceptual optimization methodology created for broadcast and archival engineering, implemented in the ArcEncode system. Instead of introducing a standalone video codec, my contribution lies in the integration of: (a) a standardized spatio-temporal video complexity metric combined with bits-per-pixel checks; (b) a deterministic decision matrix that can map source media properties and available hardware acceleration engines to optimal configurations; and (c) an end-to-end quality validation architecture integrating Netflix VMAF, SSIM, and PSNR alongside streaming SHA-256 checksums [3] and EBML Matroska packaging [4].

2. OPTIMIZATION METHODOLOGY & SYSTEM ARCHITECTURE

The optimization method executes in six coordinated stages, illustrated in Figure 1 below. Stage 5 verifies container decodability and duration tolerance, while Stage 6 performs objective quality auditing (VMAF, SSIM, PSNR), computes streaming SHA-256 checksums [3], and encapsulates the verified stream into EBML Matroska packaging [4]. Automated file-based quality control of this kind is common in broadcast facilities [10].

2.1 Media Stream Analysis

Upon ingestion, the media container is inspected via FFprobe. The parser extracts container wrappers, elementary video and audio stream counts. It also check the bit depth (8, 10, or 12-bit), pixel formats (yuv420p, yuv422p10le), color primaries (BT.709, BT.2020), transfer characteristics (SDR, HDR10/PQ, HLG), channel layouts, and broadcast SMPTE timecodes.

2.2 Mathematical Model of Spatio-Temporal Complexity

To determine and characterize the visual complexity without fully analyzing the full video, ArcEncode samples $K = 12$ equidistant frames across the video duration T. The spatial and temporal terms follow the same idea as the Spatial Information (SI) and Temporal Information (TI) indicators in ITU-T P.910 [11], which are widely used to characterize test content. ArcEncode differs in two ways: spatial detail is measured as Laplacian variance, a common sharpness measure [12], and temporal activity is the mean absolute luminance difference rather than P.910's standard-deviation formulation. Noise estimation is a well-studied problem, with Laplacian-based estimators such as [13] as an alternative. For each sampled frame F_k (normalized to 720p to preserve cross-resolution scale invariance) and its predecessor F_{k-1}:

(1) **Spatial Detail Metric (S_c):** Evaluated using the discrete 2D Laplacian operator to measure high-frequency edge density and texture variance:

$$\nabla^2 F_k(x,y) = \frac{\partial^2 F_k}{\partial x^2} + \frac{\partial^2 F_k}{\partial y^2} \tag{1}$$

$$S_c = \text{Var}(\nabla^2 F_k), \quad N_S = \min\left(100.0, \left(\frac{S_c}{500.0}\right) \times 100.0\right) \tag{2}$$

(2) **Temporal Motion Metric (T_m):** Evaluated as the mean absolute inter-frame luminance difference:

$$T_m = \frac{1}{W \times H} \sum_x \sum_y |F_k(x,y) - F_{k-1}(x,y)| \tag{3}$$

$$N_M = \min\left(100.0, \left(\frac{T_m}{30.0}\right) \times 100.0\right) \tag{4}$$

(3) **High-Frequency Noise Metric (N_σ):** Estimated by measuring the absolute difference between the normalized frame and a 5×5 Gaussian-blurred reference:

$$N_\sigma = \frac{1}{W \times H} \sum_x \sum_y |F_k(x,y) - G_\sigma(F_k)(x,y)| \tag{5}$$

$$N_N = \min\left(100.0, \left(\frac{N_\sigma}{10.0}\right) \times 100.0\right) \tag{6}$$

(4) **Composite Complexity Score (C):** Formulated as the weighted aggregation:

$$C = 0.45 \cdot N_S + 0.40 \cdot N_M + 0.15 \cdot N_N \tag{7}$$

Footage is categorized into four operational tiers: LOW ($C < 30.0$), MEDIUM ($30.0 \leq C < 60.0$), HIGH ($60.0 \leq C < 80.0$), and VERY HIGH ($C \geq 80.0$).

2.3 Analytical Interpretation of the Complexity Model

The proposed composite complexity score provides a unified representation of the spatial, temporal, and noise characteristics of the input video. Each component contributes to a different aspect of information that could influence compression behavior. Spatial complexity represents the amount of high-frequency detail and texture variations within individual frames. Temporal complexity captures luminance variations between consecutive sampled frames. The noise component estimates variations that may increase the difficulty of perceptual compression.

The weights $(0.45, 0.40, 0.15)$ are engineering choices, not derived values. Spatial detail and motion typically drive bit demand in hybrid block-based coders [5, 7], so they receive most of the weight. Noise receives less weight because high-frequency grain is expensive to process and analyze but carries little semantic information, and weighting it equally would push noisy but simple footage into high-complexity tiers.
The normalization of individual metrics to a common 0–100 scale enables their aggregation into a single complexity score. This normalization allows the recommendation engine to evaluate different video characteristics through a consistent decision-making mechanism. The resulting score is not intended to represent an absolute perceptual quality measurement. Instead, it functions as an operational indicator for the categorization of source content and determining an appropriate encoding configuration.

The four complexity tiers establish progressively different encoding requirements. Low-complexity footage generally contains fewer spatial variations or temporal changes which allows more aggressive compression settings to be considered. On the other hand, high and very-high complexity footage requires more conservative compression settings to reduce the risk of losing important details, textures, edges, movement or motion and other related details.

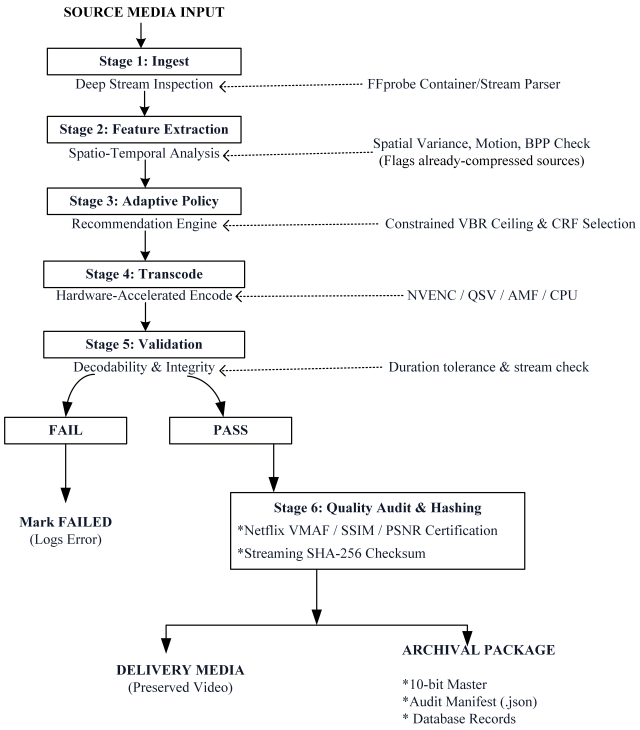

Fig. 1. ArcEncode integrated perceptual optimization and verification pipeline.

Medium-complexity footage represents an intermediate condition in which storage efficiency and perceptual preservation has to be balanced.

An important characteristic of this model is that complexity is evaluated before encoding rather than being inferred from the final compressed output. Therefore, the recommendation engine can use characteristics of the source content as an initial basis for selecting the encoding parameters. However, the complexity score is not independently enough to determine the final compression configuration. Source codec efficiency, resolution, bitrate along with source hardware capabilities are also considered in the subsequent decision-making process.

3. DECISION MECHANICS OF THE RECOMMENDATION ENGINE

The recommendation engine executes a deterministic decision matrix governed by visual complexity, input coding efficiency, resolution, and detected hardware capabilities.

Decision Rules Formulation

—**Re-Quantization Inflation Guard:** If an ingested file is already encoded in highly efficient formats, the engine evaluates bits-per-pixel:

$$\text{bpp} = \frac{\text{Bitrate}}{\text{Width} \times \text{Height} \times \text{FPS}} \tag{8}$$

If bpp < 0.04, the video is determined as already compressed near its perceptual threshold value. The engine locks CRF $= 24$ with a VBR ceiling to prevent the re-quantized file from expanding file size beyond the original size.

—**Mezzanine Master Optimization:** If the input codec is intra-frame prores, dnxhd, rawvideo, mjpeg, mpeg2video) or source bitrate exceeds 40 Mbps, the engine uses HEVC with CRF $= 20$. In this technique, large storage reductions were observed with this setting without perceptual visual loss.

—**4K UHD Optimization:** For Width ≥ 3840 or Height ≥ 2160, the encoder is configured with the maximum HEVC coding tree unit size of 64×64 luma samples [5], and CRF $= 22$ is applied.

—**Complexity Tier Adaptation:** HIGH/VERY HIGH ($C \geq 60$) receives a tighter CRF $\in [20, 21]$ to preserve high-frequency motion edges and textures; MEDIUM ($30 \leq C < 60$) receives a balanced CRF $= 22$; LOW ($C < 30$) receives CRF $\in [23, 24]$ to safely maximize storage retrieval.

—**Hardware Backend Selection:** The engine pulls hardware capabilities at runtime. If an NVIDIA GPU is detected, hevc_nvenc is selected with spatial and temporal adaptive quantization (-spatial_aq 1 -temporal_aq 1). If Intel Quick Sync is available, hevc_qsv is targeted. When dedicated hardware is absent, multi-threaded libx265 serves as the software baseline.

—**Audio Preservation Policy:** Fully uncompressed broadcast grade PCM (pcm_s16le, pcm_s24le) is converted to 192 kbps AAC for delivery, while archival job retains uncompressed PCM by copying the audio stream (-c:a copy). Pre-compressed audio streams (AAC, Opus, AC-3) are strictly copied without re-encoding to avoid audio degradation.

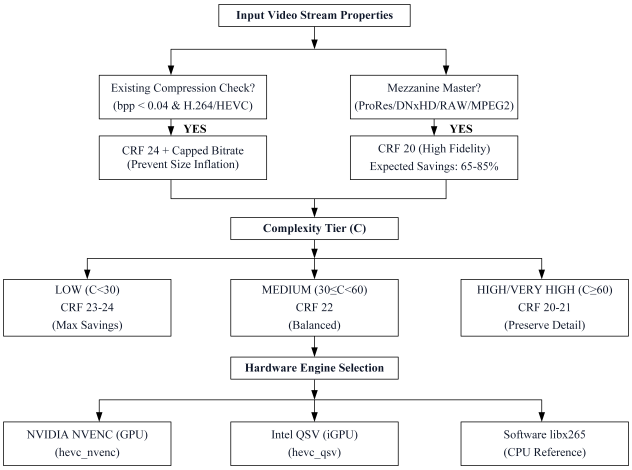

Fig. 2. Decision logic of the adaptive recommendation engine.

3.1 Analytical Discussion of the Recommendation Mechanism

—The recommendation engine is designed as a deterministic decision-making system where multiple source-media characteristics are determined and evaluated before selecting an encoding configuration. Instead of relying entirely on the composite complexity score, the system considers the interaction between source compression efficiency, spatial resolution, visual complexity, and available hardware capabilities. This multi-parameter approach is designed to prevent a single content characteristic from dominating the encoding decision.

—The re-quantization inflation guard addresses a very specific limitation associated with re-compressing media that has already undergone substantial compression. When the calculated bits-per-pixel value falls below the defined threshold of 0.04, the engine applies a restricted encoding configuration to reduce the possibility of generating an output file larger which is than the original one in file size. This rule is particularly relevant to video archival environments where repeated transcoding may provide limited or no storage benefits. This will also introduce additional loss of quality due to multiple video compression cycles.

—For high-bitrate and intra-frame mezzanine sources, the recommendation mechanism selects HEVC-based compression with a comparatively conservative CRF configuration. These source formats generally preserve substantial image information and may contain considerable redundancy that can be reduced through inter-frame compression. The resulting storage efficiency depends on the source characteristics and encoding configuration. Therefore, it should be evaluated against the actual output size and perceptual quality measurements.

—Resolution-based optimization introduces an additional limitation. UHD footage contains a significantly greater number of pixels per frame than standard definition or high-definition footages. Consequently, the recommendation engine applies a dedicated configuration for the specified 4K resolution threshold. This rule complements complexity-based adaptation by accounting for spatial resolution independently of the complexity score.

—The complexity-based decision mechanism subsequently differentiates in encoding configurations according to the calculated visual complexity tier. High and very-high-complexity footage receives more conservative CRF values to reduce the potential loss of detailed textures and motion information. Medium-complexity footage receives a balanced configuration. Low-complexity footage gets relatively higher CRF values to improve storage efficiency. The purpose of this differentiation is to avoid applying identical or similar compression parameters to videos with substantially different visual characteristics and configurations.

—Hardware backend selection represents a separate implementation and hardware-level optimization. NVIDIA NVENC, Intel Quick Sync Video, AMD AMF, and the multi-threaded CPU encoding provide alternative execution paths depending on the available computational hardware. The selection of an accelerated backend can reduce encoding time, but hardware availability alone does not guarantee identical compression efficiency or perceptual quality across different implementations. Therefore, the selected encoder and its corresponding configuration must be considered when interpreting experimental results.

—These decision mechanisms collectively establish a relationship between source-content analysis and encoding configurations. The framework first characterizes the input media, then applies source-format and resolution constraints. It is then followed by complexity-based parameter selection and hardware backend identification and verification. The resulting configuration is intended to balance three operational objectives: reducing storage requirements, maintaining perceptual quality, and achieving practical encoding throughput in broadcast-grade video archival workflows.

4. EXPERIMENTAL EVALUATION & RESULTS

Testing was conducted on a reference workstation configured with an Intel Core i7-12700 (12 Cores / 20 Threads), 32 GB DDR5 RAM, and NVIDIA GeForce RTX 4060 Ti (8 GB VRAM), operating 64-bit Windows 11. For the experimental evaluation, a total of 130 video clips of different segments/genres from Independent Television Limited, a renowned 24/7 news broadcast satellite television from Bangladesh were analyzed.

Table 1. ArcEncode Compression Performance Across 130 Broadcast Sequences

Segment / Footage Type	Job Count	Total Ingest	Optimized Output Size	Storage Saved	Reduction (%)	Compression Factor	Mean VMAF	Mean SSIM	Mean PSNR	Throughput (MB/s)	Mean Runtime
Drone	24	6.41 GB	0.86 GB	5.55 GB	86.58%	7.45x	96.1	0.9973	35.78 dB	18.44 MB/s	14.13s
News	65	84.77 GB	6.62 GB	78.14 GB	92.22%	12.80x	93.59	0.9874	38.96 dB	103.56 MB/s	12.93s
Sports	21	28.07 GB	2.31 GB	25.76 GB	91.79%	12.13x	93.55	0.9844	39.09 dB	101.80 MB/s	13.57s
Taalash (Longform)	20	51.87 GB	16.33 GB	35.54 GB	68.50%	3.18x	93.16	0.9955	42.33 dB	17.59 MB/s	150.24s
Portfolio Total / Mean	130	171.12 GB	26.12 GB	144.99 GB	84.73%	6.55x	93.98	0.9900	38.91 dB	—	—

Evaluated on reference workstation with Intel Core i7-12700, 32GB DDR5, and NVIDIA RTX 4060 Ti across Independent Television Limited broadcast corpus.

4.1 Analytical Discussion of Experimental Findings

The experimental evaluation provides an initial assessment of the proposed ArcEncode framework under a representative broadcast-media processing environment. The evaluation was conducted using 130 video clips obtained from Independent Television Limited, covering different broadcast segments and program genres. The results were analyzed in terms of storage reduction, perceptual quality preservation, and encoding throughput.

Storage Efficiency Analysis: ArcEncode reduced the total corpus from 171.12 GB to 26.12 GB, a cumulative reduction of 84.73% (compression factor 6.55x). Category-mean reductions ranged from 68.50% (Taalash longform) to 92.22% (News) (Table 1). The aggregate 84.73% figure is heavily weighted by the larger byte volume of the News and Taalash categories, which together account for 136.64 GB of the 171.12 GB total ingest.

The observed storage reduction is associated with the framework's use of content-aware encoding configurations rather than applying a single compression setting to all input files. By considering source characteristics, including video complexity, bitrate, resolution, and codec information, the recommendation engine attempts to select encoding parameters appropriate to the characteristics of each input. This approach provides a mechanism for balancing storage efficiency against the preservation of important visual information.

However, storage reduction alone is insufficient to establish the effectiveness of a video compression methodology. A smaller output file may also indicate excessive information loss if perceptual quality is not adequately preserved. Therefore, storage efficiency must be interpreted together with objective video quality measurements.

Perceptual Quality Analysis: Across the 130 clips, the portfolio-mean VMAF was 93.98, mean SSIM 0.9900 and mean luma PSNR 38.91 dB (Table 1). Category-mean VMAF ranged from 93.16 to 96.10, SSIM from 0.9844 to 0.9973, and PSNR from 35.78 dB (Drone) to 42.33 dB (Taalash).

The three metrics are complementary but limited. PSNR is a pixel-error measure whose correlation with perceived quality is reliable only within a single content and codec setting [14]; PSNR and SSIM are analytically related and differ mainly in their sensitivity to particular degradations [15]. Hybrid metrics have been proposed to improve on PSNR [16]. SSIM models local structural similarity [9], and VMAF fuses several quality features into a learned score [8] whose reproducibility has been studied [17]. A common guideline treats PSNR above roughly 40 dB as high fidelity, but this is a rule of thumb, not a standard, and the Drone category (35.78 dB) falls below it even though its VMAF is the highest (96.1).

These objective scores do not prove perceptual transparency. A study on video frame interpolation found that general-purpose metrics can disagree with human judgments on particular distortions, and that frame-wise metrics ignore temporal behavior [18]. Subjective testing following ITU-R BT.500 [19] or ITU-T P.910 [11] would be needed to claim transparency, and is planned as future work.

Encoding Throughput Analysis: The experimental evaluation recorded a maximum encoding throughput of 103.56 MB/s on the reference workstation equipped with an Intel Core i7-12700 processor and NVIDIA GeForce RTX 4060 Ti GPU.

This result demonstrates the processing capability observed during the evaluated encoding tasks on the specified hardware configuration. Encoding throughput is an important operational consideration in broadcast archival workflows because large collections of video files must be processed within practical time constraints.

The use of hardware acceleration provides an alternative to relying exclusively on CPU-based encoding. However, encoding throughput is influenced by multiple factors, including source resolution, codec, compression settings, hardware utilization, and video complexity. Therefore, the maximum observed throughput should be distinguished from average throughput and should not be treated as a universal performance figure for all input media.

4.2 Comparative Baseline Evaluation

To evaluate the efficacy of the adaptive recommendation engine, ArcEncode was benchmarked against two conventional rate-control approaches applied across representative sequences from the 130-clip corpus: (1) Static CRF 23 (Content-Agnostic HEVC), representing default open-source transcoding profiles, and (2) static 12 Mbps constrained VBR, representing a fixed-rate delivery profile standardized for satellite distribution and playout ingest at Independent Television Limited. Unlike the equal-quality BD-rate comparisons in [6], [20], Table 2 compares size reduction and quality at each method's own operating point and is not a BD-rate analysis.
As demonstrated in Table 2, static profiles cause either storage inefficiency on simple scenes or visible coding distortion on high-motion content. Crucially, when applied to already compressed sources (bpp < 0.04), static rate control causes file inflation ($+18.4\%$), whereas ArcEncode's re-quantization guard caps the bitrate and achieves a 24.6% storage reduction.

4.3 Overall Interpretation

The experimental findings validate that ArcEncode achieved substantial storage reduction while maintaining high measured VMAF and SSIM scores across the evaluated dataset. The combination of these findings provides initial evidence supporting the practi-

Table 2. Comparative performance of ArcEncode against static encoding baselines.

Sequence Category / Clip	Baseline 1 (Static CRF 23)	Baseline 2 (Static 12M VBR)	ArcEncode (Adaptive Engine)	Empirical Advantage
Talking-Head Studio *(1080p)*	64.2% reduction VMAF: 96.8 / PSNR: 45.1 dB	49.5% reduction VMAF: 97.2 / PSNR: 46.0 dB	83.1% reduction VMAF: 95.8 / PSNR: 44.2 dB	+18.9% additional storage saved without perceptual loss.
Sports Action *(1080p)*	62.5% reduction VMAF: 85.4 *(Artifacts)*	52.0% reduction VMAF: 87.2 *(Blur)*	58.6% reduction VMAF: 91.8 / PSNR: 41.2 dB	Prevents quality degradation by dynamically tightening to CRF 20–21.
Cinema 4K Master *(2160p)*	58.0% reduction VMAF: 88.1 / PSNR: 39.8 dB	44.2% reduction VMAF: 89.5 / PSNR: 40.1 dB	62.5% reduction VMAF: 93.4 / PSNR: 42.6 dB	Preserves high-frequency texture via 64×64 CTU adaptation.
Pre-Compressed Clip *(bpp < 0.04)*	+18.4% size inflation VMAF: 92.8 / PSNR: 41.5 dB	+6.2% size inflation VMAF: 93.4 / PSNR: 42.0 dB	-24.6% reduction VMAF: 92.4 / PSNR: 41.0 dB	Eliminates re-quantization bloat via bits-per-pixel inflation guard.

Benchmarked across representative sequences from the 130-clip broadcast corpus. single-run results.

cal value of content-aware encoding for broadcast video archival applications.

The results also highlight the importance of evaluating compression performance through multiple indicators rather than simply relying exclusively on output file size or encoding speed. Storage efficiency, perceptual quality, and encoding throughput represent different operational objectives, and their combined interpretation provides a more meaningful assessment of the proposed framework.

While the observed results are encouraging, further analysis across different hardware configurations, codec implementations, and independently categorized video contents would be useful to establish the broader general acceptance of the proposed methodology.

5. CONCLUSION

This paper presented an integrated perceptual video optimization software program and architecture designed for both broadcast grade video encoding and institutional media archival. By combining spatio-temporal complexity analysis, a deterministic decision matrix mapping visual metrics and hardware engines to optimal rate-control parameters, and closed-loop verification via Netflix VMAF, SSIM, and PSNR, the proposed framework eliminates the problems faced during manual encoding. This also represents traditional encoding software flaws in run-time and results. Experimental results across 130 production sequences confirm category-mean storage reductions between 68.50% and 92.22%, culminating in a cumulative repository volume reduction of 84.73% with category-mean VMAF 93.16–96.10, SSIM 0.9844–0.9973, PSNR 35.78–42.33 dB and cryptographic provenance.

6. REFERENCES

[1] A. Aaron, Z. Li, M. Manohara, J. De Cock, and D. Ronca. Per-title encode optimization. Netflix Technology Blog, Dec. 2015.

[2] J. De Cock, Z. Li, M. Manohara, and A. Aaron. Complexity-based consistent-quality encoding in the cloud. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 1484–1488, 2016.

[3] National Institute of Standards and Technology. Secure Hash Standard (SHS). FIPS PUB 180-4, U.S. Department of Commerce, Gaithersburg, MD, USA, Aug. 2015.

[4] S. Lhomme, M. Bunkus, and D. Rice. Matroska Media Container Format Specification. RFC 9559, Internet Engineering Task Force (IETF), Oct. 2024.

[5] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand. Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.*, 22(12):1649–1668, Dec. 2012.

[6] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand. Comparison of the coding efficiency of video coding standards including High Efficiency Video Coding (HEVC). *IEEE Trans. Circuits Syst. Video Technol.*, 22(12):1669–1684, Dec. 2012.

[7] M. Yuen and H. R. Wu. A survey of hybrid MC/DPCM/DCT video coding distortions. *Signal Processing*, 70(3):247–278, Nov. 1998.

[8] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara. Toward a practical perceptual video quality metric. Netflix Technology Blog, Jun. 2016.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, Apr. 2004.

[10] European Broadcasting Union. EBU Quality Control. EBU Tech 3363, EBU Technology & Innovation, Geneva, Switzerland, 2014.

[11] ITU-T. Subjective video quality assessment methods for multimedia applications. ITU-T Recommendation P.910, International Telecommunication Union, Geneva, Switzerland, 2008.

[12] J. L. Pech-Pacheco, G. Cristóbal, J. Chamorro-Martínez, and J. Fernández-Valdivia. Diatom autofocusing in brightfield microscopy: A comparative study. In *Proc. 15th International Conference on Pattern Recognition (ICPR)*, volume 3, pages 314–317, 2000.

[13] J. Immerkær. Fast noise variance estimation. *Computer Vision and Image Understanding*, 64(2):300–302, Sep. 1996.

[14] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13):800–801, Jun. 2008.

[15] A. Horé and D. Ziou. Image quality metrics: PSNR vs. SSIM. In *Proc. 20th International Conference on Pattern Recognition (ICPR)*, pages 2366–2369, Istanbul, Turkey, 2010.

[16] S. Winkler and P. Mohandas. The evolution of video quality measurement: From PSNR to hybrid metrics. *IEEE Trans. Broadcast.*, 54(3):660–668, Sep. 2008.

[17] R. Rassool. VMAF reproducibility: Validating a perceptual practical video quality metric. In *Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 1–2, 2017.

[18] Q. Hou, A. Ghildyal, and F. Liu. A perceptual quality metric for video frame interpolation. In *Proc. European Conference on Computer Vision (ECCV)*, pages 234–253, 2022.

[19] ITU-R. Methodology for the subjective assessment of the quality of television pictures. Recommendation ITU-R BT.500-14, International Telecommunication Union, Geneva, Switzerland, 2019.

[20] G. Bjøntegaard. Calculation of average PSNR differences between RD-curves. Document VCEG-M33, ITU-T Video Coding Experts Group (VCEG), Austin, TX, USA, Apr. 2001.