



Spatio-Temporal Indexing, Content-Aware Compression, and Adaptive Delivery for Bandwidth-Efficient Live Video Streaming: A State-of-the-Art Review

¹Mohammed Monday Isah, ²Paschal Uchenna Chinedu

¹Department of Computer Engineering, Edo State University, Iyamho, Edo State.

²Department of Information Systems and Technology, Southern Delta University, Ozoro, Delta State

ABSTRACT

Live video streaming accounts for a substantial share of contemporary internet traffic; however, efficiently utilizing constrained bandwidth while preserving temporal continuity and perceptual quality remains a significant challenge. Addressing this challenge requires efficient video compression, timely delivery, and adaptive quality management under varying content complexity and network conditions. Conventional video codecs exploit spatial and temporal redundancies to reduce bitrate, while adaptive bitrate (ABR) streaming systems dynamically select encoded representations based on estimated network capacity and playback conditions. Similarly, keyframe extraction and video indexing identify representative frames, scene transitions, and temporal structures to facilitate content retrieval and analysis. Despite their complementary capabilities, these approaches address distinct objectives, and their integration into a unified live-streaming architecture introduces challenges related to computational overhead, end-to-end latency, coding dependencies, and temporal continuity. This paper presents a state-of-the-art review of established video coding standards, content-aware frame selection, spatio-temporal indexing, compression-domain analysis, and adaptive live-streaming architectures. Based on a critical synthesis of the literature, it identifies unresolved research gaps and highlights promising directions, including lightweight content-aware analysis, dependency-aware temporal selection, joint rate-distortion-latency optimization, and cross-layer feedback between encoding and delivery. The review provides a conceptual foundation for future research into integrated, bandwidth-efficient live video streaming without assuming that the proposed directions have been experimentally validated.

ARTICLE INFO

Article History

Received: February, 2026

Received in revised form: April, 2026

Accepted: June, 2026

Published online: September, 2026

KEYWORDS

Adaptive Bitrate Streaming; Spatio-Temporal Indexing; Video Compression; Keyframe Selection; Low-Latency Streaming

INTRODUCTION

Live video streaming has become an integral component of contemporary digital communication, supporting applications such as remote education, telemedicine, video surveillance, online broadcasting, industrial monitoring, and interactive communication [1-3]. The growing demand for high-quality, real-time video delivery places considerable pressure on network bandwidth, computational resources, and latency budgets [4]. These challenges are

particularly pronounced in bandwidth-constrained environments, where maintaining acceptable visual quality and temporal continuity requires careful coordination between video compression and delivery mechanisms. Although aggressive compression can reduce transmission requirements, it may introduce perceptual distortion, while excessive buffering and processing delays can compromise the responsiveness of live-streaming applications [5].

Corresponding author: Mohammed Monday Isah

isah.m.m@gmail.com

Department of Computer Engineering, Edo State University, Iyamho, Edo State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



Modern video coding standards address bandwidth limitations by exploiting spatial and temporal redundancies through prediction, transform coding, quantization, and entropy coding. Standards such as Versatile Video Coding (VVC/H.266) incorporate advanced coding tools intended to improve compression efficiency across diverse applications, including real-time communication and streaming [6-7]. However, the practical benefits of these techniques must be considered alongside encoding complexity, processing overhead, and latency constraints, particularly in resource-limited streaming environments. Consequently, codec efficiency alone does not fully determine the effectiveness of a live-streaming system.

Adaptive bitrate (ABR) streaming complements video compression by adjusting the representation delivered to a client according to prevailing network conditions and playback requirements [8]. Dynamic Adaptive Streaming over HTTP (DASH) enables the selection of suitable encoded representations, while low-latency delivery architectures, including those based on the Common Media Application Format (CMAF), seek to reduce delays associated with encoding, packaging, segment availability, and playback. Nevertheless, effective adaptation depends not only on network conditions but also on the characteristics of the video content and the ability of the streaming pipeline to respond to changing conditions without introducing excessive overhead or instability [9].

More so, spatio-temporal indexing and keyframe extraction techniques facilitate the identification of representative frames, scene transitions, motion patterns, and semantically significant events. These techniques are widely relevant to video retrieval, summarization, and content analysis and may also provide useful information for content-aware compression and resource allocation [10]. However, their direct application to continuous live streaming requires careful consideration. A frame that contributes little to a summary may still be necessary to preserve temporal smoothness or maintain inter-frame prediction dependencies. Similarly, high-motion or visually complex frames may require

different coding resources from relatively static frames, although motion and visual complexity alone do not necessarily indicate semantic importance. These distinctions highlight the need to examine content analysis, compression, and adaptive delivery as complementary but functionally distinct components.

Despite advances in video coding, content-aware analysis, and adaptive streaming, coordinating these components within a unified live-streaming framework remains challenging. In particular, the interaction between frame importance, coding dependencies, rate-distortion optimization, computational complexity, and end-to-end latency requires further investigation. An effective integration must exploit content information without disrupting temporal continuity, exceeding processing deadlines, or undermining the responsiveness of the delivery system.

Against this background, this paper presents a state-of-the-art review of video coding standards, spatio-temporal indexing, adaptive frame selection, content-aware compression, and adaptive live-streaming architectures. Through a critical synthesis of the existing literature, it examines the principles, capabilities, and limitations of these approaches, with particular emphasis on their suitability for bandwidth-constrained and latency-sensitive environments. The review further identifies unresolved technical challenges and discusses promising research directions, including lightweight content-aware analysis, dependency-aware temporal selection, joint rate-distortion-latency optimization, and cross-layer coordination between encoding and delivery. By bringing these complementary research areas into a coherent analytical framework, the paper establishes a foundation for future investigations into more efficient and quality-aware live video streaming.

STATE-OF-THE-ART APPROACHES

Video Coding Standards and Compression

Conventional video coding standards employ hybrid coding architectures to exploit spatial and temporal redundancies, thereby reducing the amount of data required to represent



video sequences [11-12]. These architectures combine predictive coding, transform coding, quantization, and entropy coding to achieve a balance between bitrate and reconstructed video quality. In particular, inter-frame prediction exploits temporal correlations by referencing previously reconstructed pictures, allowing the encoder to represent changes between frames rather than independently encoding every frame in its entirety. Although this approach can substantially improve compression efficiency for temporally correlated content, it introduces dependencies between pictures. Consequently, the loss or removal of reference pictures may affect the reconstruction of subsequent frames and compromise temporal continuity.

Versatile Video Coding (VVC/H.266) extends the established hybrid-coding framework through advanced partitioning, prediction, and transform-coding tools designed to improve compression efficiency across a range of video applications [12]. Its versatility makes it relevant to bandwidth-constrained streaming; however, the practical benefits depend on encoding configurations, content characteristics, computational resources, and latency requirements. In particular, improved compression efficiency does not necessarily translate into reduced encoding complexity or lower end-to-end latency, as the computational demands of advanced coding tools may introduce additional processing overhead. More so, modern codecs provide increasingly sophisticated mechanisms for reducing video bitrate, codec-level compression represents only one component of bandwidth-efficient live streaming.

Three closely related but functionally distinct operations must be considered. First, video compression reduces the number of bits required to represent a sequence while maintaining an acceptable level of reconstructed quality. Second, frame selection and spatio-temporal indexing identify frames, temporal intervals, or content structures that are relevant to a particular analytical, retrieval, or representation task. Third, adaptive streaming determines which encoded representation should be delivered according to prevailing network conditions and

playback requirements. Although these operations can complement one another, they are not interchangeable. In particular, removing frames does not constitute an improvement in codec compression efficiency and may introduce temporal discontinuities or disrupt inter-frame prediction dependencies if performed without appropriate safeguards.

Spatio-Temporal Redundancy and Adaptive Frame Selection

Spatial and temporal redundancies constitute fundamental sources of compression in video sequences. Spatial redundancy arises from correlations among neighbouring pixels within a frame, whereas temporal redundancy results from similarities and dependencies between successive frames [13]. Modern video codecs exploit these characteristics through techniques such as motion estimation, motion compensation, and inter-frame prediction, reducing the amount of information required to represent video without necessarily discarding frames. The effectiveness of these techniques, however, depends on the degree of spatial and temporal correlation within the content, which may vary considerably across scenes and motion patterns.

In contrast to codec-based redundancy reduction, frame-selection techniques identify representative, informative, or task-relevant frames for content analysis, retrieval, and summarization. Conventional approaches employ uniform temporal sampling, inter-frame differences, colour histograms, motion descriptors, and scene-change detection thresholds. Although computationally efficient, these methods may fail to capture brief but significant events, gradual scene transitions, or complex temporal relationships. Learning-based approaches seek to address some of these limitations by incorporating higher-level visual features, semantic information, and temporal context.

Existing research work on video summarization, spatio-temporal frame indexing proposed by [14-15] provide a useful foundation for understanding the principal methodological categories and their underlying approaches.

Corresponding author: Mohammed Monday Isah

✉ isah.m.m@gmail.com

Department of Computer Engineering, Edo State University, Iyamho, Edo State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved

Spatio-temporal indexing facilitates the identification of representative frames and the characterization of video content. However, content complexity, semantic importance, and coding dependencies represent distinct but

interrelated characteristics that must be considered when selecting frames for content-aware video processing. Figure 1 illustrates the conceptual relationship.

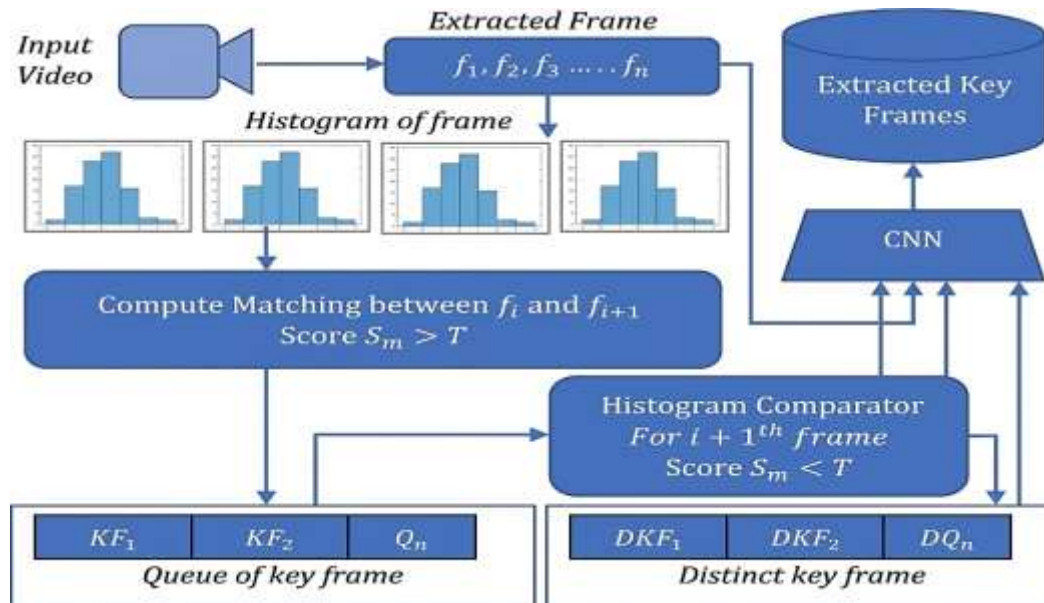


Figure 1: Conceptual relationship between spatio-temporal indexing, content complexity, semantic importance, and coding dependencies in video sequences

As illustrated in Figure 1, a visually complex frame is not necessarily semantically important, while a representative frame may not be sufficient to preserve inter-frame prediction dependencies. Therefore, effective content-aware frame selection should consider both the analytical importance of video content and the coding requirements necessary to maintain temporal continuity. The suitability of frame-selection methods for live-streaming applications depends on their computational requirements, responsiveness, and intended function. Uniform temporal sampling offers low computational overhead and predictable sampling intervals but may overlook transient events occurring between selected frames.

Scene-change detection can identify abrupt transitions and support temporal segmentation; however, simple threshold-based methods may be sensitive to gradual transitions,

camera movement, and rapid changes in visual content. Motion-aware selection can direct analytical attention towards dynamically active intervals, although motion magnitude alone is insufficient to determine semantic importance or the number of bits required for adequate representation. Learning-based selection can capture more complex relationships between visual content and temporal importance, but its practical deployment may be constrained by inference latency, computational cost, training-data requirements, domain shift, and the stability of its decisions under changing streaming conditions.

These limitations become particularly important when frame selection is considered alongside video compression. A frame selected as representative of a scene is not necessarily suitable as a coding reference picture, just as a frame considered less important for summarization may remain essential for

Corresponding author: Mohammed Monday Isah

isah.m.m@gmail.com

Department of Computer Engineering, Edo State University, Iyamho, Edo State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



maintaining temporal continuity or supporting inter-frame prediction.

Spatio-Temporal Indexing and Content-Aware Analysis

Spatio-temporal indexing organizes video content into searchable temporal units, including individual frames, shots, scenes, and event intervals. Each unit may be associated with metadata such as timestamps, visual features, motion descriptors, semantic labels, or learned feature embeddings. The primary purpose of video indexing is to facilitate content retrieval, temporal organization, and content analysis rather than to directly reduce the bitrate of a live video stream. Nevertheless, the information generated through indexing may provide useful contextual input for compression and streaming decisions [16].

In continuous live-streaming environments, spatio-temporal indexing and content-aware analysis may support three complementary functions. First, complexity estimation characterizes intervals that may require additional coding resources because of high motion, intricate textures, or frequent scene changes. Second, importance estimation identifies frames or temporal intervals associated with significant events for which perceptual degradation may be particularly undesirable. Third, temporal organization identifies scene boundaries and event transitions that may provide suitable points for reassessing encoder configurations, analysis windows, or delivery policies. When appropriately integrated, these functions can help distinguish variations in coding demand from changes in the perceptual or semantic significance of video content [16].

However, complexity, importance, and temporal structure represent different dimensions of video analysis and should not be treated as interchangeable indicators. A visually complex frame is not necessarily semantically important, while a significant event may occur within a visually simple scene. Similarly, a scene transition may indicate a change in content characteristics without necessarily requiring a change in encoding parameters or delivery policy.

Consequently, systems that use a single descriptor as a proxy for all three dimensions risk making inappropriate resource-allocation decisions, potentially reducing compression efficiency or perceptual quality.

A central challenge in applying content-aware analysis to live streaming is the need to generate sufficiently informative descriptors within strict computational and temporal constraints. Delays between content analysis and the execution of encoding or delivery decisions may reduce the relevance of descriptors, particularly during rapid scene transitions or abrupt changes in motion. Conversely, computationally intensive semantic-analysis models may introduce additional processing overhead and compete with the resources required for real-time encoding. These constraints highlight the importance of lightweight, low-latency indexing methods that can provide timely content information without undermining the responsiveness of the streaming pipeline.

Thus, potential contribution of spatio-temporal indexing to bandwidth-efficient live streaming lies not in treating indexing as a direct compression mechanism, but in investigating how timely information about content complexity, semantic importance, and temporal structure can inform encoding and delivery decisions. Effective integration requires these descriptors to be evaluated according to their distinct purposes while accounting for processing overhead, coding dependencies, perceptual quality, and end-to-end latency.

Adaptive Live-Streaming Architectures

Dynamic Adaptive Streaming over HTTP (DASH) provides a framework for delivering video through addressable media segments and multiple encoded representations described in a Media Presentation Description (MPD) [17]. During playback, the client adaptation logic selects an appropriate representation based on available information, including estimated network throughput, playback-buffer occupancy, and segment characteristics. The DASH framework supports both live and on-demand streaming, enabling clients to adjust video quality in response

Corresponding author: Mohammed Monday Isah

✉ isah.m.m@gmail.com

Department of Computer Engineering, Edo State University, Iyamho, Edo State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



to changing network conditions and playback requirements. However, the effectiveness of adaptation depends on the accuracy of available estimates, the responsiveness of the selection strategy, and the operational characteristics of the encoding and delivery pipeline.

Adaptive bitrate (ABR) algorithms employ different control strategies to balance video quality, bandwidth utilization, and playback stability. Throughput-based approaches estimate available network capacity to guide representation selection, whereas buffer-based approaches use playback-buffer occupancy to manage the risk of rebuffering. Model-based approaches combine network and playback information to anticipate the consequences of adaptation decisions. Learning-based approaches, in turn, use observed system states and outcomes to derive representation-selection policies. Authors in [17-18], demonstrates the application of reinforcement learning to adaptive bitrate selection. Although these approaches provide alternative mechanisms for managing streaming quality, live delivery introduces additional constraints, including strict latency budgets, continuously evolving content, and limited opportunities to recover from delayed or inappropriate adaptation decisions.

The Common Media Application Format (CMAF) supports fragmented media delivery and can be used with chunk-oriented streaming workflows in which portions of a media segment become available before the complete segment has been produced. This capability can reduce delays associated with segment completion and availability, thereby supporting low-latency delivery. Nevertheless, CMAF does not independently guarantee low end-to-end latency. The resulting delay also depends on factors such as encoder lookahead, group-of-pictures (GOP) structure, packaging behaviour, transport conditions, player-buffering policies, and audio-

video synchronization. Achieving low latency therefore requires coordination across the encoding, packaging, transmission, and playback stages rather than reliance on a single format or delivery mechanism.

These considerations have important implications for integrating content-aware analysis into adaptive live-streaming architectures. Frame-level or scene-level information can contribute to more informed encoding and representation-selection decisions only when it is generated and acted upon within the relevant processing and delivery deadlines. If content analysis introduces excessive computational overhead or produces decisions too late, its potential benefits may be offset by increased latency or reduced system responsiveness. Furthermore, frequent changes to encoding parameters, representation characteristics, or temporal processing policies may introduce additional overhead and complicate coordination across representations, segment boundaries, and playback timelines.

Consequently, effective content-aware live streaming requires a balance between the potential benefits of fine-grained content analysis and the operational constraints of adaptive delivery. A promising research direction is to investigate how content descriptors can inform encoding and adaptation policies without destabilizing playback, disrupting synchronization, or exceeding end-to-end latency budgets. This perspective reinforces the need to evaluate content-aware mechanisms as components of an integrated streaming pipeline rather than as isolated optimization techniques. Figure 2 presents a conceptual architecture illustrating how spatio-temporal analysis, video encoding, adaptive bitrate selection, and network feedback could be coordinated to improve bandwidth utilization while maintaining perceptual quality and low latency.

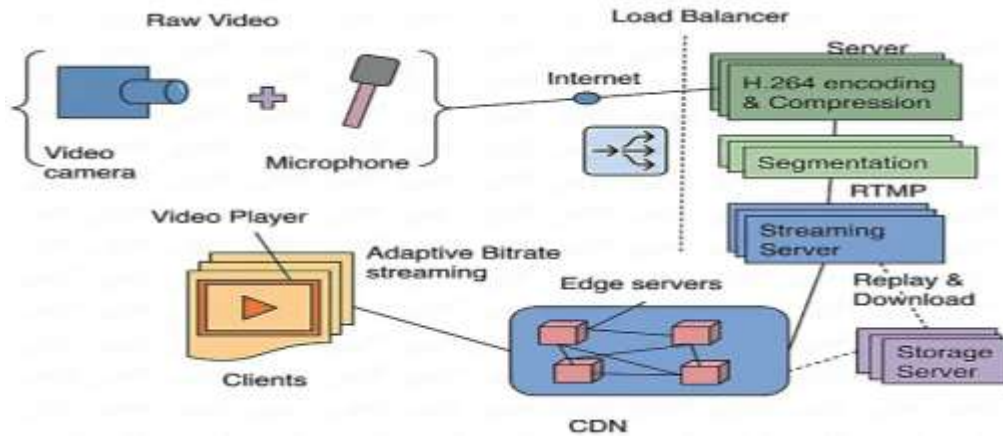


Figure 2: Conceptual Architecture for Content-Aware, Bandwidth-Efficient Live Video Streaming

As illustrated in Figure 2, content analysis can inform encoding decisions, while network and playback feedback guide adaptive bitrate selection. However, effective coordination must account for computational overhead, coding dependencies, and end-to-end latency. The architecture is conceptual and requires further evaluation to establish its practical benefits.

Content-Aware and Learning-Assisted Control

Content-aware compression and learning-assisted adaptation seek to improve streaming efficiency by using video characteristics, network conditions, and playback feedback to guide encoding and representation-selection decisions [19]. Content descriptors, such as motion activity, scene changes, and visual complexity, may help anticipate variations in coding demand, while learning-based methods can support adaptive bitrate selection under changing network conditions.

Integrating lightweight content analysis with established codec rate-control and adaptive bitrate (ABR) mechanisms presents a promising

research direction. However, such integration must account for computational overhead, coding dependencies, and end-to-end latency. The literature reviewed does not establish that a unified spatio-temporal indexing and compression controller consistently outperforms independently optimized encoding and ABR components. Its potential benefits and limitations therefore remain open for further investigation.

Critical Comparative Synthesis

The five state-of-the-art approaches reviewed address complementary aspects of bandwidth-efficient live video streaming. Video coding reduces the bitrate required to represent video, while frame selection and spatio-temporal indexing characterize temporal structure and content importance. Adaptive streaming manages representation selection under changing network conditions, whereas content-aware and learning-assisted control offer opportunities to coordinate these processes. Their strengths and limitations are compared in Table 1.

Table 1. Critical comparative synthesis of state-of-the-art approaches for bandwidth-efficient live video streaming

Approach	Core principles	Strengths	Limitations
Video coding standards and compression	Exploits spatial and temporal redundancy through prediction, transform coding, quantization, video	Reduces bitrate while preserving quality;	Advanced coding tools may increase encoding complexity and processing delay. Inter-

Corresponding author: Mohammed Monday Isah

isah.m.m@gmail.com

Department of Computer Engineering, Edo State University, Iyamho, Edo State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



Approach	Core principles	Strengths	Limitations
	and entropy coding. Examples include VVC/H.266 and AV1 [11].	coding frameworks support a wide range of video applications.	frame dependencies also constrain unrestricted frame removal and may affect error propagation.
Spatio-temporal redundancy and motion descriptors, scene change detection, sampling, or event selection	Uses temporal differences, learned features to identify representative or relevant frames [14].	Supports temporal analysis, summarization, gradual transitions, and selection can disrupt temporal continuity if applied simply indiscriminately and must be methods can operate with distinguished from codec-level relatively low redundancy reduction. computational overhead.	May overlook brief events or transitions. Frame selection can disrupt temporal continuity if applied simply indiscriminately and must be methods can operate with distinguished from codec-level relatively low redundancy reduction. computational overhead.
Spatio-temporal indexing and content-aware analysis	Organizes video into frames, and associates them with timestamps, visual features, motion information, or semantic descriptors [14-15].	Enables efficient content retrieval and provides reduce bitrate. Descriptor accuracy, processing cost, and analysis delays can limit its usefulness in real-time streaming.	Indexing does not directly reduce bitrate. Descriptor accuracy, processing cost, and analysis delays can limit its usefulness in real-time streaming.
Adaptive live-streaming architectures	Uses DASH representations, throughput estimates, playback-buffer information, and conditions; adaptation policies to select quality suitable streaming quality. CMAF supports interoperable segmented-media delivery, while Pensieve illustrates learning-based ABR [17-19].	Adapts delivery to changing network conditions; supports quality management, or rebuffering. Low latency playback stability, and depends on coordinated streaming encoding, packaging, transport, and playback rather than the delivery format alone.	Throughput estimates may be inaccurate, and adaptation can cause quality fluctuations. Low latency playback stability, and depends on coordinated streaming encoding, packaging, transport, and playback rather than the delivery format alone.
Content-aware and learning-assisted control	Uses content descriptors, network feedback, and informed potentially learned policies to coordinate encoding decisions and representation selection [14,19].	Offers potential for more resource allocation by combining video characteristics with superiority of an integrated network and playback conditions.	May introduce computational overhead, prediction errors, and control instability. The compression controller over independently optimized components remains unestablished.

CONCLUSION

Bandwidth-efficient live video streaming requires a coordinated consideration of video compression, spatio-temporal indexing, content-aware analysis, and adaptive delivery. Although modern video codecs exploit spatial and temporal redundancies to reduce bitrate, spatio-temporal indexing techniques characterize video content,

while adaptive bitrate (ABR) mechanisms respond to variations in network conditions and playback requirements. These approaches address distinct yet complementary aspects of streaming efficiency. However, their effective coordination remains challenging, particularly in balancing compression efficiency, perceptual quality,

Corresponding author: Mohammed Monday Isah

isah.m.m@gmail.com

Department of Computer Engineering, Edo State University, Iyamho, Edo State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



temporal continuity, computational complexity, and end-to-end latency.

This review has examined the principles, capabilities, and limitations of these complementary approaches, highlighting the potential of spatio-temporal content descriptors to inform encoding and delivery decisions. It has also established the importance of distinguishing semantic importance, visual complexity, and coding necessity when considering content-aware frame selection and compression. In particular, the removal or prioritization of frames must account for inter-frame prediction dependencies and temporal continuity, while content-analysis mechanisms must operate within the computational and timing constraints of live-streaming environments. These considerations underscore the need for further investigation into lightweight content-aware analysis, dependency-aware temporal selection, joint rate-distortion-latency optimization, and coordinated feedback between encoding and adaptive delivery.

Ultimately, the review identifies the integration of spatio-temporal indexing, content-aware compression, and adaptive streaming as a promising research direction for improving bandwidth utilization in live video delivery. Nevertheless, the extent to which such integration can provide measurable advantages over independently optimized encoding and ABR mechanisms remains an open research question. By critically synthesizing existing approaches, identifying unresolved technical challenges, and outlining relevant research directions, this paper provides a conceptual foundation for the development and rigorous evaluation of more adaptive, perceptually efficient, and bandwidth-conscious live video-streaming architectures.

REFERENCES

- [1] M. Alsader, A. A. Barakabitze, and I. H. Mkwawa, "QoE-driven adaptive video streaming: Architectures, techniques, and future research challenges toward 6G networks," *IEEE Access*, 2025.
- [2] A. Nauman, Y. A. Qadri, M. Amjad, Y. B. Zikria, M. K. Afzal, and S. W. Kim, "Multimedia Internet of Things: A comprehensive survey," *IEEE Access*, vol. 8, pp. 8202–8250, 2020.
- [3] K. Khan, "Advancements and challenges in 360 augmented reality video streaming: A comprehensive review," *International Journal of Computing*, vol. 13, no. 1, pp. 1–20, 2024.
- [4] M. O. Momoh, J. N. Ndunagu, V. K. Abanihi, S. M. Yusuf, and M. I. Abdullah, "A comparative analysis of video compression standard and algorithms: State of the art," *ATBU Journal of Science, Technology and Education*, vol. 8, no. 2, pp. 181–188, 2020.
- [5] R. Lee, S. I. Venieris, and N. D. Lane, "Deep neural network-based enhancement for image and video streaming systems: A survey and future directions," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–30, 2021.
- [6] K. Choi, "A study on fast and low-complexity algorithms for versatile video coding," *Sensors*, vol. 22, no. 22, Art. no. 8990, 2022.
- [7] B. Farhat, "Efficient hardware designs of the new Versatile Video Coding (VVC) tools for ASIC platforms," Ph.D. dissertation, INSA de Rennes, Rennes, France, 2022.
- [8] H. Amer, M. S. Hassan, and M. H. Ismail, "A review of learning-based methods for adaptive video streaming over HTTP," *IEEE Access*, vol. 13, pp. 111134–111162, 2025.
- [9] R. Viola, A. Gabilondo, A. Martin, J. F. Mogollón, M. Zorrilla, and J. Montalbán, "QoE-based enhancements of chunked CMAF over low latency video streams," in *Proc. 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2019, pp. 1–6.
- [10] M. Sima, "Key frame extraction for human action videos in dynamic spatio-temporal slice clustering," *Journal of Physics: Conference Series*, vol. 2010, no. 1, Art. no. 012076, 2021.
- [11] B. Bross, J. Chen, S. Liu, Y.-K. Wang, et al., "Overview of the Versatile Video Coding (VVC) standard and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, 2021, doi: 10.1109/TCSVT.2021.3101953.

Corresponding author: Mohammed Monday Isah

✉ isah.m.m@gmail.com

Department of Computer Engineering, Edo State University, Iyamho, Edo State.

© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



- [12] J. Žádník, M. Mäkitalo, J. Vanne, and P. Jääskeläinen, "Image and video coding techniques for ultra-low latency," *ACM Computing Surveys*, vol. 54, no. 11s, 2022.
- [13] M. O. Momoh, B. Yahaya, A. A. Abdulrazaq, H. M. Katagum, and U. Abubakar, "Towards achieving an optimal bandwidth utilization in live video streaming," in *Proc. Academic Conference of Cambridge Research and Publications International on 21st Century Challenges and Opportunities: Multidisciplinary Approaches*, vol. 19, no. 3, 2019, pp. 55–61.
- [14] P. Narwal, N. Duhan, and K. K. Bhatia, "A comprehensive survey and mathematical insights towards video summarization," *Journal of Visual Communication and Image Representation*, vol. 89, Art. no. 103670, 2022, doi: 10.1016/j.jvcir.2022.103670.
- [15] A. E. Adedokun, M. B. Abdulrazak, M. O. Momoh, H. Bello-Salau, and B. O. Sadiq, "A spatio-temporal based frame indexing algorithm for QoS improvement in live low-motion video streaming," *ATBU Journal of Science, Technology and Education*, vol. 7, no. 3, pp. 305–315, 2019.
- [16] S. Zhang, Y. Zhu, and A. K. Roy-Chowdhury, "Context-aware surveillance video summarization," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5469–5478, 2016.
- [17] H. Mao, R. Netravali, and M. Alizadeh, "Neural adaptive video streaming with Pensieve," in *Proc. ACM SIGCOMM 2017*, 2017, pp. 197–210, doi: 10.1145/3098822.3098843.
- [18] T. Lyko, M. Broadbent, N. Race, M. Nilsson, P. Farrow, and S. Appleby, "Evaluation of CMAF in live streaming scenarios," in *Proc. 2020 Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, 2020, pp. 21–26, doi: 10.1145/3386290.3396932.
- [19] T. Lyko, M. Broadbent, N. Race, *et al.*, "Improving quality of experience in adaptive low latency live streaming," *Multimedia Tools and Applications*, vol. 83, pp. 15957–15983, 2024, doi: 10.1007/s11042-023-15895-9.