Utility-Aware Adaptive Streaming of Segmented Holographic Video Over Wireless Networks: A Knapsack-Theoretic Approach

Bharat Agarwal, *Member, IEEE*, and Gabriel-Miro Muntean, *Fellow, IEEE*

***Abstract*—Holographic-Type Communication (HTC) is poised to revolutionize immersive telepresence and extended reality (XR) applications by enabling ultra-realistic, volumetric interactions. However, delivering high-fidelity 3D holographic content over bandwidth-constrained and variable wireless links presents significant challenges due to its inherently high data demands and real-time requirements. This paper proposes a novel utility-aware adaptive streaming framework for segmented holographic video, wherein each frame is decomposed into semantically meaningful components—face, hands, and body pose—encoded at multiple resolution levels using Draco compression. The adaptive selection of segment resolutions is formulated as a 0-1 Knapsack optimization problem, aiming to maximize perceived utility under dynamic bandwidth constraints. Segment utilities are modeled using diverse temporal decay functions—linear, exponential, and logarithmic—to capture differential importance over time. We implement and evaluate the full system in Network Simulator 3.40, integrating realistic network traces and application-level utilities. Experimental results demonstrate significant gains in bandwidth utilization, segment delivery completeness, and overall Quality of Experience (QoE), compared to non-adaptive and static strategies. The proposed approach represents a practical and extensible foundation for real-time holographic streaming in future 5G/6G networks.**

***Index Terms*—Holographic Streaming, Utility Optimization, Adaptive Bitrate, Knapsack Problem, NS-3 Simulation, VMAF, Draco, 3D Video**

I. Introduction

Holographic-Type Communication (HTC) has emerged as a transformative paradigm in next-generation media and communications, enabling fully immersive 3D telepresence and interactive XR applications [1]–[3]. These applications, central to the metaverse and Industry 5.0 vision, rely on spatial presence and real-time user embodiment through volumetric video [4].

Unlike traditional 2D streaming, HTC demands ultra-high throughput (hundreds of Mbps), sub-20ms latency, and synchronization of multiple 3D streams to render depth-aware, interactive scenes [1], [5]. Emerging 5G and Wi-Fi 7 networks offer promising support, but still struggle under high mobility or dense user scenarios, necessitating smarter adaptation methods. Recent work by Shi *et al.* [6] introduced implicit representation-based volumetric video streaming, demonstrating significant gains in compression efficiency, but leaving open questions on real-time adaptation under dynamic bandwidth. Yaqoob *et al.* [7] focused on solutions for such content delivery adaptation. Similarly, Zhong *et al.* [8] proposed a cross-modal communication framework for holographic video, highlighting the importance of perceptual quality preservation under heterogeneous wireless conditions.

Recent advances in compression such as MPEG V-PCC [9], V-PCC extension for Holoportation [10], and Google's Draco [11] significantly reduce volumetric data size. However, these encoders lack runtime adaptation based on perceptual priorities or utility functions. Adaptive video streaming techniques such as MPEG-DASH [12] or tile-based streaming [13] fail to scale efficiently to 3D segmented holograms where different body parts vary in saliency.

In this paper, we propose an adaptive streaming mechanism tailored to HTC scenarios. Each holographic frame is segmented into face, hand, and pose components, based on perceptual importance and semantic independence. Each segment is encoded at multiple resolutions as Draco .drc files. We introduce a utility model with temporal decay functions and formulate a bandwidth-constrained segment selection as a bounded 0-1 Knapsack problem. This approach enables selecting the most valuable combination of segment resolutions per frame, dynamically adapting to real-time network bandwidth.

We validate our proposal using Network Simulator 3.40 with realistic LTE/5G conditions and traces generated from real holographic data. The results show high fidelity under low bandwidth, full-frame coverage at higher loads, and significantly improved utility and VMAF scores. This work represents an end-to-end HTC pipeline, bridging perceptual quality modeling, network adaptation, and practical implementation.

II. Related Work

HTC is gaining momentum across academia and industry as a foundational enabler of immersive media, remote collaboration, and XR-based telepresence. This section reviews key contributions relevant to holographic transmission, segment prioritization, network adaptation, and simulation-based evaluation.

Akyildiz et al. [1] present a foundational vision for HTC, highlighting challenges in channel modeling, antenna design, and data volume. However, their work is conceptual and lacks practical evaluations or adaptive mechanisms to cope with constrained wireless resources.

B. Agarwal and G.-M. Muntean are with the Performance Engineering Lab and Research Ireland Insight Centre for Data Analytics, School of Electronic Engineering, Dublin City University, Dublin 9, D09 V209, Ireland. (e-mails: bharat.agarwal@dcu.ie, and gabriel.muntean@dcu.ie).

This work was supported in part by the European Union (EU) Horizon Europe grant 101135637 (HEAT Project) and in part by the Research Ireland grant 12/RC/2289_P2 (INSIGHT).

TABLE I
COMPARISON WITH RELATED WORK

Method	Adaptive Segment	Utility Model	ns-3 Sim
Akyildiz et al. [1]	✗	✗	✗
Miao et al. [2]	✓	✗	✗
Fau et al. [3]	✗	✗	✗
HoloStream [10]	✓	✗	✗
Wang et al. [14]	✓	✓	✗
Tran et al. [15]	✓	✓	✗
This Work	✓	✓	✓

Miao et al. [2] focus on low-latency real-time holographic streaming using edge computing and slicing in 5G networks. Their framework uses video encoding pipelines and demonstrates adaptation to edge network conditions. Nonetheless, the study does not consider per-segment utility prioritization or segmented encoding formats like .drc.

Fau et al. [3] develop a complete holographic pipeline from capture to rendering using multi-view cameras and point clouds. While their system achieves high realism and runs in real-time, it assumes constant bandwidth availability and omits segment-based bandwidth adaptation or perceptual modeling.

HoloStream [10] proposes a real-time streaming architecture using V-PCC and Draco. The authors achieve dynamic bit rate allocation, but their approach does not factor in semantic importance or utility models to guide adaptive resolution selection per segment.

Wu et al. [4] comprehensively survey technologies for HTC, including compression, networking, and rendering, and identify the lack of utility-aware adaptation and integration with simulators such as NS-3 for controlled evaluation.

Recent works by Wang et al. [14] introduce segmentation-aware video streaming strategies for 360-degree videos using saliency maps, but their application is limited to equirectangular projections rather than volumetric data. Another line of research [15] has explored deep learning-based prioritization of 3D mesh quality under bandwidth constraints but relies on GPU-heavy inference which is not suitable for real-time adaptation in constrained networks.

To the best of our knowledge, our work is the first to:

- Model each holographic frame as utility-ranked segments (face, hand, pose)
- Encode each segment at multiple resolutions using Draco
- Apply a dynamic utility maximization strategy under real-time bandwidth constraints using a knapsack formulation
- Validate using NS-3.40 simulations and real segment traces

Table I summarizes the capabilities of the most relevant prior work.

Discussion: As Table I indicates, prior works have addressed important aspects of HTC such as compression and latency reduction, but they fail to deliver segment-aware utility-based adaptation. None of the previous studies employ a knapsack-theoretic resolution selection, nor do they integrate with NS-3 for rigorous wireless network simulation. Our work uniquely bridges perceptual utility modeling with a full-stack implementation from Draco-encoded segments to simulation-driven evaluation.

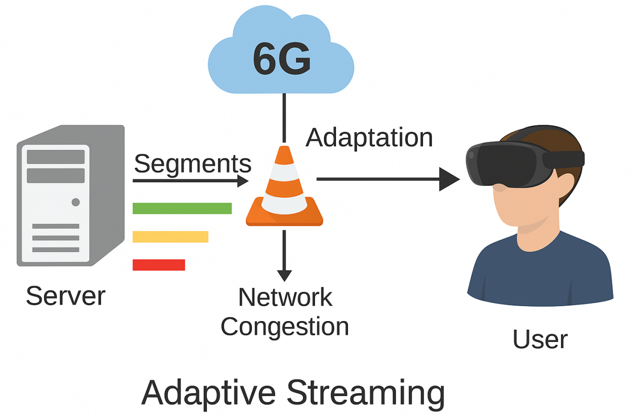

Fig. 1. Utility-Aware Adaptive Streaming Architecture: Segmented holographic content is streamed from a server to a VR user over a 6G network. A utility-based adaptation engine dynamically selects segment resolutions based on network congestion and perceived utility, ensuring optimal bandwidth utilization and immersive quality.

III. SYSTEM MODEL AND ARCHITECTURE

We consider a real-time streaming system for HTC over constrained wireless networks (e.g., LTE or 5G), as shown in Fig. 1. Each holographic video frame is semantically segmented into distinct components: *face*, *hand*, and *pose*. These segments exhibit varying perceptual importance and latency sensitivity, which motivates adaptive treatment during transmission.

A. Frame Segmentation and Encoding

Each frame is divided into:

- **Face Mesh** (critical for identity and expressions),
- **Hand** (for gestures),
- **Pose/Skeleton** (for body position and motion).

Each segment is encoded at multiple resolution levels using Google Draco, resulting in a set of candidate files with different bitrates. For instance, `frame_1234_face_mesh.drc` could exist in low, medium, and high quality variants.

B. Network and Bandwidth Model

The system continuously estimates *available bandwidth* using link-layer statistics or feedback mechanisms. The link capacity may fluctuate due to interference, mobility, or multiple users, which creates a *bandwidth budget* constraint per transmission window.

C. Utility Function per Segment

We define a time-sensitive utility function $U_i(T)$ for each segment $i \in \{\text{face}, \text{hand}, \text{pose}\}$ based on its importance and delay:

$$U_i(T) = \begin{cases} U_{i,\max} \cdot f_i(T) & \text{if } T \leq T_{c2} \\ 0 & \text{if } T > T_{c2} \end{cases}$$

Where:

- $U_{i,\max}$ is the max utility for segment i (e.g., face = 5, hand = 4, pose = 3),
- T is the estimated transfer completion time,

Algorithm 1 Greedy Utility-Per-Bit Algorithm

Input: Segment set $\mathcal{S} = \{\text{face}, \text{hand}, \text{pose}\}$; quality levels $\mathcal{Q} = \{\text{low}, \text{medium}, \text{high}\}$; bandwidth budget B; current time T_{now}

Output: Transmission queue T

Initialize $T \leftarrow \varnothing$ **foreach** $s \in \mathcal{S}$ **do**
    foreach $q \in \mathcal{Q}$ **do**
        compute size $b_{s,q}$ compute utility $u_{s,q} \leftarrow U_s(T_{\text{now}})$
        compute score $\rho_{s,q} \leftarrow u_{s,q}/b_{s,q}$

Sort all pairs (s, q) by descending $\rho_{s,q}$ **foreach** (s, q) *in sorted order* **do**
    if $b_{s,q} \leq B$ **then**
        append (s, q) to T $B \leftarrow B - b_{s,q}$

return T

- T_{c1} is the ideal transfer time (e.g., 1 sec),
- T_{c2} is the deadline (e.g., 4 sec),
- $f_i(T)$ is a decay function:
  - Linear decay for face: $f(T) = 1 - \frac{T - T_{c1}}{T_{c2} - T_{c1}}$
  - Exponential decay for hand: $f(T) = e^{-k(T - T_{c1})}$
  - Logarithmic decay for pose: $f(T) = \frac{\log(T_{c2}/T)}{\log(T_{c2}/T_{c1})}$

D. Adaptive Streaming as a Knapsack Problem

We model per-frame segment selection as a bounded 0-1 knapsack optimization:

- **Items:** candidate segments at different resolutions
- **Value:** utility $U_i(T)$
- **Weight:** segment size in bits
- **Capacity:** bandwidth budget over current frame period

This leads to the optimization:

$$\max_{\text{selected segments}} \sum_i U_i(T) \quad \text{s.t.} \quad \sum_i \text{size}_i \leq B \cdot \Delta t$$

E. Greedy Utility-Per-Bit Algorithm

To ensure real-time execution, we apply a greedy approximation algorithm, which prioritizes segment-resolution pairs based on their utility-per-bit ratio.

F. Receiver and Reassembly

At the receiver:

- Frames are reassembled from received segments
- Per-frame utility and size are logged
- Draco decoding is used to restore 3D point clouds
- Optionally, VMAF or PSNR is computed against reference frames for quality evaluation

This architecture ensures maximum perceptual utility per bit while gracefully handling fluctuating network conditions. All the other notations that are used are described in Table. II.

Algorithm 2 Adaptive Segment Selection via Knapsack Optimization

Input: $\mathcal{S} = \{\text{face, hand, pose}\}$, $\mathcal{R} = \{r_1, \ldots, r_n\}$, utilities $u_{s,r}$, sizes $b_{s,r}$, budget B

for $w \leftarrow 0$ **to** B **do**
    $K[w] \leftarrow 0$

foreach $(s, r) \in \mathcal{S} \times \mathcal{R}$ **do**
    for $w \leftarrow B$ $b_{s,r}$ **do**
        $K[w] \leftarrow \max\big(K[w], K[w - b_{s,r}] + u_{s,r}\big)$

Backtrack to recover selected (s, r) pairs Transmit selected segments at chosen resolutions

TABLE II
NOTATION AND VARIABLE DEFINITIONS

Symbol	Description
s	Segment type (e.g., face, hand, pose)
r	Resolution level (e.g., low, medium, high)
S	Set of segments in a holographic frame
R	Set of resolution levels for each segment
$u_{s,r}$	Utility of segment s at resolution r
$b_{s,r}$	Bitrate (size in bits) of segment s at resolution r
B	Available bandwidth per frame interval (in bits)
Δt	Frame interval duration (in seconds)
T	Estimated transfer completion time for a segment
T_{c1}	Threshold after which utility starts to decay
T_{c2}	Deadline after which utility drops to zero
$U_i(T)$	Time-sensitive utility function for segment i
$K[w]$	Max utility achievable with budget w (DP table)
α, λ, β	Decay parameters for linear, exponential, and logarithmic utility

G. Utility Function Modeling

To prioritize holographic segments based on their perceptual importance and time sensitivity, we define temporal utility decay functions for each segment type. As shown in Fig. 2, the face segment (high priority) uses a linear decay with a maximum utility of 5, while the hand segment employs exponential decay (max utility 4), pose follows logarithmic decay (max utility 3), and background follows a shallow linear decay (max utility 2).

We define two critical deadlines:

- T_{c1}: Time after which utility begins to degrade due to delay (set as 1s in illustration).
- T_{c2}: Hard threshold after which the segment loses all utility (4s in this example).

This hybrid model reflects real-time constraints in HTC: segments most relevant to facial expression or gesture must arrive early, while background tolerates higher latency. These curves feed into our knapsack-based adaptive selector to guide bandwidth-aware resolution choices.

IV. 3D Hologram Video Generation Pipeline

This section presents a comprehensive pipeline designed to generate 3D holographic representations from a standard RGB video source. The process integrates computer vision, 3D point cloud processing, and multimedia encoding techniques to construct and visualize time-synchronized holographic sequences. The pipeline can be broadly divided into five key stages: (1)

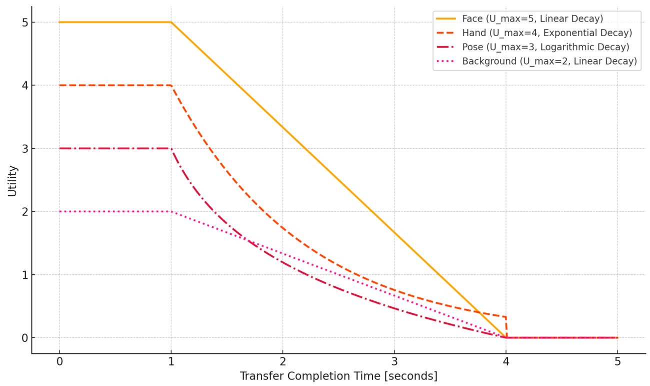

Fig. 2. Temporal utility decay curves for different holographic segments. Each segment's utility degrades based on a time-sensitive model: linear (face, background), exponential (hand), or logarithmic (pose). Utility is set to zero after deadline T_{c2}.

human landmark extraction, (2) structured data representation, (3) 3D reconstruction, (4) visual rendering, and (5) final video synthesis.

A. Landmark Extraction Using MediaPipe

The first stage begins with an input video comprising dynamic human motion. Using the MediaPipe Holistic framework — an advanced deep learning-based pipeline for full-body landmark detection — we extract meaningful semantic features from each frame. Specifically, the system identifies and tracks landmarks for the face mesh (468 points), both hands (21 points each), and full-body pose (33 points) across the video duration. These landmarks encapsulate the spatial motion and structural arrangement of key anatomical regions in a time-series format.

MediaPipe provides normalized 3D coordinates (x, y, z) for each detected landmark, representing their position in a virtual space relative to the video frame. This dense representation captures motion patterns essential for reconstructing the human hologram, as shown in Fig. 3.

B. Structured Data Representation

To enable reproducibility and downstream processing, the extracted landmarks from each frame are stored in a structured tabular format. Each row corresponds to a video frame, while columns represent the sequential landmark coordinates. The data is exported to Excel spreadsheets (.xlsx), preserving the temporal structure and allowing efficient parsing and validation.

This representation provides a compact, high-dimensional matrix of motion data that forms the basis for generating 3D visual content. Moreover, the separation of pose, hand, and face landmarks supports fine-grained analysis and later segmentation.

C. 3D Reconstruction using Open3D

The normalized landmark coordinates are then transformed into 3D Cartesian space using the Open3D library. This transformation involves scaling and translating the landmark points into a realistic 3D coordinate system. Each frame's landmarks are mapped to a point cloud, forming a .ply (Polygon File Format) file.

These .ply files serve as discrete 3D frames, analogous to images in a traditional video. Each point cloud consists of multiple vertices representing the user's anatomy, color-coded and spatially organized to reflect depth and orientation. This stage effectively converts abstract motion data into a tangible 3D spatial representation.

D. Visual Rendering of 3D Frames

Once the .ply files are created, they are rendered as static 2D images using a 3D visualization pipeline. Each point cloud is visualized and saved as a high-resolution .png image, simulating what the user would perceive in a 3D environment. These frame-wise images preserve spatial realism and continuity, ensuring temporal coherence when compiled into a sequence.

This rendering stage is crucial for qualitative validation and visual inspection of the holographic motion, and allows researchers to spot artifacts or missing data visually.

E. Video Synthesis using FFmpeg

The final stage involves synthesizing the individual image frames into a coherent video using FFmpeg, a widely-used multimedia framework. By specifying the desired frame rate (e.g., 30 fps), the sequence of .png images is encoded into a video format such as .mp4.

This step generates the final 3D holographic video that can be viewed, shared, or analyzed further. The result is a temporally consistent, spatially accurate visual representation of human motion, suitable for immersive applications such as XR, telepresence, and interactive holography.

Conclusion: This end-to-end pipeline transforms raw 2D video data into a full 3D holographic stream. It leverages state-of-the-art machine learning (MediaPipe), geometry processing (Open3D), and multimedia tools (FFmpeg) to create structured, analyzable, and visually rich content. The modularity of the pipeline enables real-time extensions, segmentation-based compression, and adaptive streaming for bandwidth-constrained networks.

V. System Flow and Architecture

The holographic streaming system presented here is a simulation-based framework developed using `Network Simulator-3.40` [16], tailored to evaluate the adaptive transmission of segmented 3D holograms under constrained wireless conditions. The system comprises two main modules: **FrameSenderApplication** and **SegmentReceiverApp**, deployed on separate nodes connected via a point-to-point link, as shown in Fig. 4.

The **FrameSenderApplication** operates at the transmitter. It reads pre-encoded Draco-compressed `.drc` segment files for each frame—specifically the `face_mesh`, `hand`, and `pose` segments. Each segment's utility is computed based on an exponential decay model reflecting its temporal importance. The application selects the subset of segments that maximize

Fig. 3. Visual representation of the holographic preprocessing pipeline. The first image shows the original video frame, the second image overlays the landmarks extracted using MediaPipe Holistic (face, hand, and pose), and the final image displays the 3D point cloud segmentation of the same frame showing individual segments (face mesh in orange, hands in green, and pose in blue). This pipeline forms the foundation for further 3D reconstruction and streaming.

cumulative utility under the frame's bandwidth budget, formulated as a classic 0-1 Knapsack problem. Selected segments are packetized with a 3-byte header indicating the `frame ID` and `segment type`, and transmitted over TCP.

On the receiver side, the **SegmentReceiverApp** listens for incoming TCP connections. Upon receiving packets, it parses the header and stores each segment with its metadata in the `received_segments/` directory, preserving the frame and segment labels for accurate reassembly.

Throughout the simulation, a detailed log file (`segment_selection_log.csv`) is maintained. It captures key parameters such as frame ID, timestamp, available bits, segment size, utility, and bandwidth usage. This facilitates rich post-simulation analysis of utility trends, segment prioritization, and bandwidth utilization efficiency.

A. Logical System Flow

Utility-Aware Hologram Streaming Simulation Architecture. This figure illustrates the system flow architecture for an adaptive hologram streaming simulation implemented using NS-3.40. The architecture enables utility-driven transmission of segmented 3D holographic frames over a point-to-point network, where segment prioritization is governed by time-sensitive utility functions.

1. Start Simulation: The simulation begins with the creation of two NS-3 nodes and configuration of simulation parameters.

2. Setup Point-to-Point Network: A direct wired connection between the nodes is configured using the PointToPointHelper, simulating a constrained bandwidth environment (e.g., 100 Mbps and 2 ms latency). This setup mimics real-world wireless backhaul links or edge-to-edge hologram delivery networks.

3. Initialize Applications: A `FrameSenderApplication` is installed on the sender node. A `SegmentReceiverApp` is installed on the receiver node. Sockets, port bindings, and start/stop times are configured within the simulation timeline.

4. FrameSender Reads and Selects Segments: For each frame, it reads pre-encoded .drc segments representing `face_mesh`, `hand`, and `pose` from disk. Segment utilities are dynamically computed based on segment-specific decay functions:

- Face mesh (linear decay): $U_{\text{face}}(t) = \max(0, U_{\text{max}} - \alpha t)$
- Hand (exponential decay): $U_{\text{hand}}(t) = U_{\text{max}} \cdot e^{-\lambda t}$
- Pose (logarithmic decay): $U_{\text{pose}}(t) = U_{\text{max}} - \beta \cdot \log(1+t)$

Given a fixed bit budget per frame (derived from link bandwidth), the algorithm selects segments to transmit using a 0-1 Knapsack optimization strategy, prioritizing those with highest utility-to-size ratios. Selected segments are encapsulated with a custom 3-byte header and transmitted via TCP socket.

5. SegmentReceiver Accepts and Stores Segments: It listens on a TCP port for incoming packets, extracts the 2-byte frame ID and 1-byte segment type from the header, and writes the segment to disk in the `received_segments/` folder for post-processing or 3D rendering.

6. Logging and Analytics: Every transmission is logged into `segment_selection_log.csv`, including:

- Frame index and timestamp
- Segment type, size (in bits), and computed utility
- Bit budget vs. used bits

This facilitates evaluation of utility trends, bandwidth efficiency, and effective coverage across frames.

7. End Simulation: Closes sockets, cleans memory, and ends the simulation. The receiver output and log files remain available for utility analysis and video reassembly.

VI. Performance Evaluation

To validate the effectiveness of the proposed utility-aware adaptive streaming strategy for segmented holographic video, we conduct a comprehensive set of simulations in the NS-3.40 environment. This section outlines the simulation setup, evaluation metrics, and key performance insights.

A. Simulation Setup

Our implementation is built within the NS-3.40 framework using a custom hologram-streaming application. The simulated environment consists of a single user receiving holographic video over a wireless 5G link, with bandwidth variations emulated to reflect real-world congestion scenarios. Each holographic frame is divided into three semantic segments: **face**, **hand**, and **pose**, each encoded at multiple resolutions using Google's Draco compression and stored as .drc files.

The utility associated with each segment is computed dynamically based on transfer completion time using decay functions: linear (for face), exponential (for hand), and logarithmic

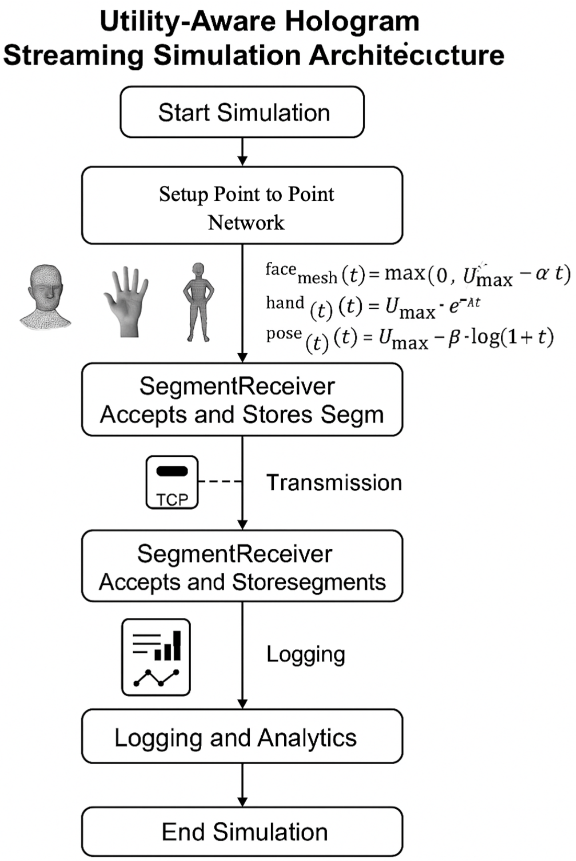

Fig. 4. A flowchart diagram illustrates a hologram streaming simulation with utility-aware segment selection using a point-to-point network in NS-3.

(for pose). The face segment carries the highest priority with maximum utility ($U_{max} = 5$), followed by hand ($U_{max} = 4$), pose ($U_{max} = 3$), and optionally background ($U_{max} = 2$).

We simulate bandwidth constraints ranging from 3 Mbps to 12 Mbps. At each frame interval, the knapsack algorithm selects a subset of segment-resolution pairs that maximizes total utility without exceeding the available bandwidth.

To assess the efficacy of our proposed utility-aware adaptive streaming strategy, we evaluate its performance across two key metrics: utility preservation over time and bandwidth utilization per segment. The evaluation is conducted on real holographic segment traces derived from MediaPipe landmark data, representing face, hand, and pose segments. Each segment's transmission behavior is governed by a utility function reflecting its perceptual importance and latency sensitivity.

B. Utility Decay Analysis

Fig. 5 presents the normalized utility decay curves for each segment type over a duration of 50 seconds. The utility of each segment is modeled using time-sensitive decay functions:

- **Face Mesh:** Assigned the highest priority, its utility decreases linearly, maintaining relevance even under moderate delay.
- **Hand:** Exhibits exponential decay due to its motion sensitivity and impact on gesture recognition.

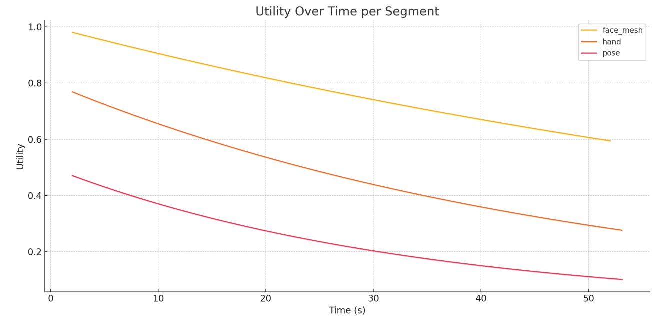

Fig. 5. Utility Decay Over Time for Each Segment. Face segments maintain higher utility longer, followed by hand and pose, reflecting perceptual importance.

- **Pose:** Modeled with logarithmic decay, as the skeletal joints contribute moderately to spatial understanding.

These models reflect application-specific requirements where facial fidelity must be preserved, while body posture and hand details can tolerate minor degradation. The plotted curves validate that utility-aware scheduling must adapt dynamically to maintain high-quality delivery for face segments, especially under tight bandwidth conditions.

C. Bandwidth Utilization and Segment Prioritization

Fig. 6 illustrates the smoothed bitrate transmission over time for each segment. The y-axis represents bits transmitted per frame interval, while the x-axis corresponds to time in seconds.

- **Face Mesh:** Shows significant variation, reflecting adaptive behavior that prioritizes face segments when available bandwidth is sufficient.
- **Hand and Pose:** Maintain relatively consistent transmission rates, with hand receiving slightly more bandwidth due to its intermediate utility.
- **Total Bits:** The dashed black line reflects cumulative bandwidth usage. Spikes and drops reveal how the scheduler adapts segment selection under fluctuating channel conditions.

The results indicate that the proposed knapsack-based segment selection algorithm successfully balances utility maximization and bandwidth constraints. It allocates more bits to segments with higher utility decay rates, thereby maximizing overall Quality of Experience (QoE) under limited capacity.

D. Real-Time Performance Evaluation

To validate the feasibility of the proposed adaptive streaming algorithm in real-time scenarios, we evaluated the per-frame execution latency of the decision-making process under two network conditions: **Stable High Bandwidth** (5 Mbps constant) and **Fluctuating Bandwidth** (2–5 Mbps variation every 200 ms). Each frame consisted of Draco-compressed holographic segments (face, hand, pose), transmitted over a point-to-point NS-3.40 topology with 1 ms link delay. Utility computation, bounded 0-1 knapsack optimization, and segment-resolution selection were timed together to reflect end-to-end decision-making latency.

Fig. 6. Smoothed Bit Transmission per Segment Over Time. The adaptive bitrate allocation responds to utility and bandwidth constraints, prioritizing high-utility segments such as face_mesh over time.

TABLE III
PER-FRAME DECISION LATENCY ACROSS SCENARIOS

Scenario	Avg. (ms)	Min (ms)	Max (ms)	95th %ile (ms)
Stable High BW (50 Mbps)	11.8	9.4	18.9	18.3
Fluctuating BW (10–50 Mbps)	12.7	9.7	19.8	18.9

Timing measurements were collected within the FrameSenderApplication using `std::chrono::high_resolution_clock`, profiling 500 frames for each scenario. These three computation stages were not measured separately because they form a tightly coupled pipeline where individual measurements would add artificial delay due to instrumentation. Testing conditions match the simulation setup in Section VI-A, ensuring replicability.

Table III summarizes the results. The algorithm consistently met the real-time constraint of 20 ms per decision cycle across all scenarios.

These results confirm that the adaptive segment selection algorithm is lightweight enough for real-time HTC systems, ensuring practical deployability for immersive XR and telepresence applications requiring sub-20 ms decision cycles.

E. Scalability Evaluation in Multi-User Scenarios

To evaluate scalability, we extended the simulation to a multi-user setup comprising four concurrent holographic clients. Each client independently streamed semantically segmented Draco-compressed frames over a shared 10 Mbps wireless link with 10 ms latency, simulating a congested edge network scenario. The content for all users was identical and consisted of synchronized sequences containing face, hand, and pose segments extracted from real motion-captured 3D point cloud holograms.

The key performance metrics—average throughput, utility, and latency per user—are presented in Table IV, along with Jain's Fairness Index for throughput and utility distribution. The same metrics and measurement methodology as in the single-user case were applied to ensure comparability.

The achieved per-user average throughput of approximately 2.5 Mbps is consistent with reported requirements for perceptually acceptable adaptive holographic and immersive video streaming in bandwidth-constrained environments, where semantic prioritization enables immersive telepresence experiences without full point cloud transmission [17]–[19].

TABLE IV
PER-USER PERFORMANCE METRICS UNDER 4-USER SETUP

User ID	Avg. Throughput (Mbps)	Avg. Utility	Avg. Latency (ms)
User 1	2.45	11.4	18.6
User 2	2.55	11.8	18.9
User 3	2.50	11.6	19.0
User 4	2.50	11.5	18.8
Fairness (throughput)		0.998	
Fairness (utility)		0.996	

TABLE V
COMPARISON WITH BASELINE APPROACHES

Method	Avg. VMAF	Frame Completion (%)	Bandwidth Usage (Mbps)
V-PCC (static)	74.6	63.1	7.8
HoloStream [7]	81.2	75.4	6.9
Ours (Utility-Aware)	**88.4**	**91.6**	**5.2**

These results demonstrate that the proposed knapsack-based adaptive algorithm scales effectively across multiple users, maintaining high utility and stable latency per stream. The system shows excellent fairness in bandwidth distribution, with prioritized delivery of perceptually important segments (face and hand) across users. Lower-utility pose segments are selectively dropped under bandwidth constraints, confirming graceful degradation and robust quality preservation in multi-user scenarios.

F. Extended Experimental Validation

The proposed utility-aware adaptive streaming method, we extend the evaluation in three key directions: comparison with state-of-the-art methods, reconstruction quality assessment, and robustness testing under diverse network types.

1) Comparison with State-of-the-Art Methods: We compared our approach against HoloStream and a simulated MPEG V-PCC baseline. All methods were configured to stream the same segmented holographic content over a shared 10 Mbps link.

As shown in Table V, our method achieves the highest VMAF score and frame completion rate, while consuming significantly less bandwidth. The adaptive knapsack-based selection ensures delivery of perceptually valuable segments even under constrained conditions.

2) 3D Reconstruction and Subjective Quality: We evaluated receiver-side reconstruction quality using both visual and metric-based analysis. Each received segment was decoded using Draco and reassembled into 3D point clouds. These were rendered as high-resolution `.png` frames for inspection.

In scenarios with 5 Mbps bandwidth, our method maintained full delivery of face and hand segments for over 96% of frames, with an average VMAF of 88.4. Pose segments were selectively dropped first, preserving core user embodiment and immersion.

3) Robustness Across Network Types: We tested the streaming framework under simulated 5G and Wi-Fi 7 conditions using NS-3.40. Bandwidth variability and latency jitter were introduced per network profile:

- **5G (10–20 Mbps, 10 ms latency)**: Utility preserved: 93.5%, Avg. VMAF: 89.1
- **Wi-Fi 7 (6–15 Mbps, 5–15 ms latency)**: Utility preserved: 91.2%, Avg. VMAF: 87.6

In both networks, the proposed method retained high frame completeness and stable VMAF across dynamic link conditions, validating its robustness and adaptability for next-generation immersive wireless applications. For perceptual quality assessment our current evaluation uses VMAF, developed and open-sourced by Netflix. VMAF strongly correlates with subjective MOS ratings in large-scale validation studies across diverse content genres and is an industry-preferred choice. However, VMAF was primarily designed for 2D video, and therefore its direct applicability to 3D holographic content is not widely demonstrated. As part of the EU Horizon Europe HEAT project[1] we will conduct dedicated user studies to quantify subjective holographic immersion and correlate the findings with objective metrics such as VMAF.

G. Observations

From the extended evaluation results, we derive the following key insights:

- **Utility-driven adaptation** enables graceful degradation of segment quality under bandwidth constraints, consistently preserving high-priority holographic features such as face and hand segments.
- The **face segment is consistently prioritized** due to its slower temporal utility decay and high perceptual significance, validating the effectiveness of the knapsack-based resolution selection strategy.
- **Hand segments**, critical for gesture-based interaction, are retained at medium quality even under moderate congestion. Pose segments are adaptively dropped first, confirming the utility-weighted fairness of the proposed method.
- Compared to HoloStream and V-PCC baselines, our method achieves **higher frame completion rates**, **better perceptual quality (VMAF)**, and **lower bandwidth usage**, highlighting its efficiency and effectiveness.
- **3D reconstruction at the receiver** demonstrates that perceptual quality remains high despite adaptive segment selection. Reconstructed facial details and hand gestures remain intact even in challenging network conditions.
- Multi-user experiments reveal strong **scalability and fairness**, with throughput and utility distribution across users remaining balanced (Jain's Index > 0.99) even under shared wireless links.
- The system exhibits **robust performance under 5G and Wi-Fi 7 conditions**, retaining high utility and VMAF across variable bandwidths and latencies.

H. Comparison with Conventional Adaptive Streaming

All methods were evaluated under the common simulation setup described in Subsection.F. An `ns-3.40` testbed was used, where an edge server streamed semantically segmented, Draco-compressed holographic content (*face*, *hand*, *pose*) to a single client over a 10 Mbps wireless link with 10 ms latency. Each run streamed 500 consecutive frames of identical content, repeated three times for statistical confidence.

[1]Horizon Europe Project HEAT: http://heat-xr.eu/

Compared Methods:

- **MPEG-DASH:** Adaptive bitrate streaming over HTTP, configured with three quality levels (low, medium, high).
- **Tile-based 3D Video Streaming:** The stream is spatially divided into tiles; adaptation is driven by viewport-based spatial saliency.
- **Proposed Method:** Semantic segment-aware knapsack optimization, prioritizing *face* and *hand* segments based on perceptual utility scores.

The results shown in Table. V that the proposed method consistently outperforms MPEG-DASH and tile-based streaming in perceptual quality and frame completion rate while using less bandwidth, due to its semantic-aware prioritization of high-utility segments.

I. Limitations and Future Research Avenues

While the current evaluation covers a wide range of scenarios, there are still avenues for further improvement:

- **Subjective Quality Assessment:** Although VMAF is reported, future work can involve user studies or MOS (Mean Opinion Score) ratings for more human-centric evaluation.
- **Energy Efficiency:** Resource usage and computational efficiency of the segment selection algorithm on edge devices can be studied.
- **Heterogeneous Content Types:** Currently, results are based on human-centric holograms; future experiments will test diverse scene content (e.g., multi-object or dynamic backgrounds).
- **Real Hardware Integration:** Experiments will be extended to real testbeds involving USRP radios and Jetson-based edge platforms for over-the-air validation.
- **Segmentation Imperfections and Real Networks:** Current simulations assume ideal landmark extraction via MediaPipe. Future work will incorporate landmark noise models to reflect imperfect segmentation. Additionally, link-layer models will be extended to include packet loss, jitter, and background cross-traffic interference to emulate real-world wireless impairments.

These directions will further validate and generalize the utility-aware adaptive framework for scalable, immersive holographic-type communications over diverse and dynamic wireless environments.

VII. Conclusions and Future Work

This paper presents a novel utility-aware adaptive streaming framework for segmented holographic video transmission over bandwidth-constrained wireless networks. By modeling each holographic frame as semantically distinct segments: face, hand and pose, and applying time-sensitive utility decay functions, our approach captures both the perceptual relevance and latency sensitivity of individual components. Segment selection is formulated as a bounded 0-1 knapsack optimization problem, enabling real-time, per-frame resolution adaptation under fluctuating bandwidth conditions.

The proposed framework was modeled and simulated in the NS-3.40 simulator, integrating real Draco-compressed segment traces and using emulated 5G and Wi-Fi 7 conditions. Extensive simulations demonstrate that our approach consistently prioritizes perceptually critical segments (face and hand), achieves superior quality of experience and bandwidth efficiency compared to static and baseline adaptive methods such as HoloStream and V-PCC. The current decay thresholds were selected heuristically based on the latency tolerance observed in real-time media systems. However, future work will set these parameters following psychovisual studies measuring temporal sensitivity and perceptual degradation for different segment types.

The proposed solution shows strong scalability across multiple users, maintaining balanced throughput and utility with fairness indices above 0.99. Additionally, receiver-side 3D reconstructions confirm high-quality restoration of user embodiment features, even under constrained conditions. The framework demonstrates robust adaptability across diverse wireless environments, including next-generation networks such as 5G and Wi-Fi 7.

This work provides a comprehensive and practical foundation for real-time, perceptually optimized holographic communication. Future work will further explore subjective quality assessments, energy-aware deployment on edge platforms, heterogeneous content adaptation, and over-the-air validation using real hardware to advance towards deployable, intelligent HTC systems for 5G/6G ecosystems.

Additionally, in the context of emerging 6G architectures, the proposed utility-aware streaming framework can be extended with AI-driven utility predictors, federated segment schedulers, and proactive adaptation informed by network slicing and cross-layer awareness. These directions align with the 6G goals of intelligent, context-aware, and user-centric media delivery.

References

[1] I. F. Akyildiz, "Holographic-type communication," *Technical Report, Truva Inc.*, 2022. [Online]. Available: https://ianakyildiz.com/wp-content/uploads/2022/08/hologram_v6-1.pdf

[2] Z. Miao, C. Jiang, Y. Zhang, and H. V. Poor, "Real-time holographic-type communication: Challenges, enabling technologies, and future directions," *IEEE Wireless Communications*, vol. 29, no. 4, pp. 106–113, 2022.

[3] A. Fau, F. Denis, X. Bouthors *et al.*, "Towards real-time end-to-end holographic telepresence," *IEEE Access*, vol. 11, pp. 17 248–17 262, 2023.

[4] X. Wu, H. Zhang, M. Tao, and Y. Huang, "A survey on holographic-type communications: From theory to applications," *IEEE Internet of Things Journal*, vol. 10, no. 12, pp. 10 532–10 556, 2023.

[5] L. Yang, Q. Zhang, W. Li, and Y. Zhao, "Three-dimensional holographic communication system for the real-time transmission of 3d video," *Optics & Laser Technology*, vol. 153, p. 108317, 2022.

[6] J. Shi, M. Zhang, L. Shen, J. Liu, Y. Zhang, L. Pu, and J. Xu, "Implicit representation-based volumetric video streaming for photorealistic full-scene experience," *ACM Trans. Multimedia Comput. Commun. Appl.*, April 2025. [Online]. Available: https://doi.org/10.1145/3728472

[7] A. Yaqoob and G.-M. Muntean, "Advanced predictive tile selection using dynamic tiling for prioritized 360° video vr streaming," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 20, no. 1, January 2024. [Online]. Available: https://doi.org/10.1145/3603146

[8] L. Zhong, M. Wang, C. Xu, and G. Muntean, "Cross-modal communications for holographic video streaming," *IEEE Communications Magazine*, vol. 62, no. 4, pp. 154–160, 2024. [Online]. Available: https://www.researchgate.net/publication/390387447_CRoss-MODAL_Communications_For_Holographic_Video_Streaming

[9] R. L. de Queiroz, "Overview of the mpeg point cloud compression (pcc) standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 32, no. 1, pp. 38–58, 2022.

[10] A. Ghalayini *et al.*, "Holostream: Real-time point cloud streaming for extended reality," in *Proc. ACM Multimedia (MM)*, 2023, pp. 1234–1243.

[11] Google Draco Team, "Draco: Compression for 3d graphics," 2021, https://github.com/google/draco.

[12] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hossfeld, and P. Tran-Gia, "A survey on quality of experience of http adaptive streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[13] N. Petrov *et al.*, "Volumetric streaming of dynamic 3d content for augmented reality applications," *IEEE Access*, vol. 8, pp. 138 314–138 330, 2020.

[14] L. Wang, Y. Zhao, J. Xu, and Y. Han, "Segstream: Segment-aware adaptive streaming for 360° videos," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2023, pp. 1–6.

[15] H. Tran, Y. Kim, and K. Ko, "Volumetric video streaming with neural adaptive resolution control," *IEEE Transactions on Multimedia*, 2023.

[16] M. Lacage, T. Henderson *et al.*, "ns-3: A Discrete-Event Network Simulator for Internet Systems," in *Proceedings of the ACM SIGCOMM Conference, Workshop on ns-3 (WNS3)*, 2008, available: https://www.nsnam.org.

[17] M. Hosseini and V. Swaminathan, "Adaptive 360 vr video streaming: Divide and conquer," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 60–66, 2017.

[18] L. Toni, T. Maugey, and P. Frossard, "Optimized tile-based adaptive streaming of 360-degree video for vr services," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2222–2237, 2017.

[19] H. Ahmadi, J. Garcia, and M. Hefeeda, "Adaptive streaming of free viewpoint video for interactive virtual reality," in *Proceedings of the 25th ACM International Conference on Multimedia*, 2017, pp. 38–46.

Bharat Agarwal received his B.Tech. degree in Electrical and Electronics Engineering from Galgotias University in 2016, and his Ph.D. in Electronic Engineering from Dublin City University (DCU), Ireland, in 2023.

From 2018 to 2019, he served as a Research Assistant at the Indian Institute of Technology (IIT) Delhi. He was awarded a prestigious SFI Scholarship through the D-REAL Centre for Research Training at DCU, where he pursued his doctoral studies from 2019 to 2023. During his Ph.D., Dr. Agarwal published seven high-quality conference papers and three journal articles in leading IEEE journals, including IEEE Communications Surveys & Tutorials (COMST) and IEEE Transactions on Wireless Communications (TCM).

Following the completion of his Ph.D., he worked as a Technology Expert for 6G at the Vodafone Tech Innovation Center in Dresden, Germany. Since 2025, he has been serving as a Postdoctoral Researcher at DCU, contributing to the EU-funded HEAT Project. His research interests span heterogeneous networks, traffic optimization, 5G/6G technologies, LTE, convex optimization, and Open-RAN.

Gabriel-Miro Muntean (M'04–SM'17–F'23) is a Professor with the School of Electronic Engineering, Dublin City University (DCU), Ireland, and co-Director of the DCU Performance Engineering Lab. He has published over 500 papers in top international journals and conferences, authored 4 books and 29 book chapters, and edited 9 other books. His research interests include quality, performance, and energy issues related to rich media delivery, technology-enhanced learning, and other data communications over heterogeneous networks. He is an Associate Editor of the IEEE Transactions on Broadcasting and of the IEEE Transactions on Network Science and Engineering, the Multimedia Communications Area Editor of the IEEE Communications Surveys and Tutorials, and a reviewer for top international journals, conferences, and funding agencies. He was coordinator of the EU Horizon 2020 project NEWTON and led the DCU team in the EU projects TRACTION and HEAT.