A Multidimensional Media Adaptation Framework for Live Holographic Communication

Carl Udora, Peng Qian, Jingxuan Men and Ning Wang, *Senior Member, IEEE*

***Abstract*—The rapidly increasing popularity of immersive multimedia services such as live holographic communication represents the future trend of extended reality (XR) applications. However, the realization of such immersive and interactive experiences is limited by the lack of fundamental understanding of how different user behaviours and environmental factors jointly affect the overall quality of experience (QoE). In particular, compared with the media adaptation mechanisms applied in conventional video applications, considerably more independent factors may influence user QoE in these applications, including both human- and network-related factors. In this paper, we investigate the fundamental design principles of dynamic media adaptation methods for live holographic communication by holistically considering these factors. Specifically, a machine learning-based scheme is introduced to facilitate intelligent adaptation of both the frame quality (resolution level) and frame rate according to specific contexts, such as the user intent/behaviour (including object motion patterns and user movements) and real-time network conditions. Extensive real-world experiments are conducted to assess the feasibility and performance of the proposed method, and comparisons with state-of-the-art methods are performed. The results indicate that the proposed approach can effectively satisfy user intent, with increased user QoE.**

***Index Terms*—User Intent, Holographic Communication, Immersive Media, Intelligent Adaptation, Quality of Experience.**

I. Introduction

THE increased demand for immersive and interactive user-driven multimedia experiences has been addressed through the use of consumer-friendly multimedia devices such as head-mounted displays (HMDs) and immersive media applications. One such application, projected to be critical for multimedia services [1], [2], is holographic communication [3]. This application provides immersive user-driven experiences through photorealistic object representations. These representations are accessed through unrestricted user movements and object views across six degrees of freedom (6DoF) in extended reality (XR) space.

The functionalities of holographic communication systems can be realized through existing 5G networks and edge computing facilities; however, existing systems are considered low level, with limited functions [4]. In particular, widespread utilization is limited owing to strict computational, interactive and network requirements [5].

Carl Udora, Peng Qian and Jingxuan Men are with the Institute for Communication Systems, University of Surrey, GU2 7XH Guildford, United Kingdom, (E-mails: [cu00029, peng.qian, j.men]@surrey.ac.uk)
Ning Wang is with the University of Bristol, BS8 1QU, Bristol, United Kingdom, (Email: n.wang@bristol.ac.uk)
This manuscript was edited for English language by American Journal Experts (AJE).

Immersive multimedia applications provide a sense of presence, copresence, and interaction, thus facilitating more immersive user experiences [6]–[8]. On the other hand, it can be inferred that a variety of independent factors may influence user quality of experience (QoE) in a complex fashion. These factors can be broadly categorized into three influence factor (IF) classes: human IFs, system IFs and context IFs [8], [9]. Here, environmental IFs collectively refer to system and context IFs due to contextual overlaps [8]. Human-oriented factors include influential human properties [8], which, in holographic communication, include source object behaviour and user behaviour and intent. In this context, human locomotion and navigation are highly important factors, as they enable more immersive experiences [10] and significantly impact the QoE [11], [12]. Furthermore, source object behaviours are important factors, including both static and dynamic behaviours, as they are contextually influential environment-oriented factors [8]. Moreover, in holographic communication, the network environment is critical, as strict latency requirements are implemented in such applications [1], [5]. Thus, network parameters are highly important, as such parameters could be associated with adverse network conditions and substantially impact the user QoE [13]. It is therefore essential to incorporate the aforementioned human-oriented factors and network-oriented factors in adaptation approaches for live holographic communication to further improve the immersive experience, satisfy user intent and enhance user QoE.

For the very first time, this paper presents a multidimensional media adaptation framework for live holographic communication to address the fundamental limitations of existing schemes, which are tailored for traditional media adaptation. Specifically, in the proposed framework, the uncertainties associated with unpredictable user and object behaviours in heterogeneous network environments are considered, as these uncertainties present significant problems for immersive multimedia applications such as live holographic communication. In this context, artificial intelligence (AI) technologies are leveraged to consider multiple human- and environment-oriented uncertainties. Furthermore, we introduce an AI-driven user and source object behavioural and network awareness strategy for immersive multimedia applications. Extensive experiments involving representations of various quality, network delays and packet losses are performed to generate machine learning models, which are applied in the proposed framework. Moreover, to our knowledge, we are the first to explore multidimensional adaptation strategies, particularly in relation to QoE performance metrics such as the quality of the

representations, or the resolution, and the number of frames per second (FPS) for immersive multimedia applications. The results of thorough experiments demonstrate the feasibility and good performance of the proposed framework compared with those of state-of-the-art frameworks, indicating its suitability to satisfy user intent while maintaining user QoE.

Although more immersive multimedia systems are being used, extremely immersive and interactive holographic communication is still considered to be in the early stages of development, with more advancements expected in the future. The technical contributions of this paper can be summarized as follows:

- To our best knowledge, this is the first piece of work in the literature to systematically elucidate how different human-oriented and environment-oriented factors affect specific QoE performance metrics such as resolution and frame rates. Furthermore, we explore physical user locomotion and static and dynamic source object behaviours and evaluate their impact on various performance metrics. Moreover, we explore the impact of adverse network conditions on adaptive holographic communication strategies and show that network awareness should be considered in adaptation frameworks to maintain satisfactory QoE for the user.
- We propose an AI-driven adaptation strategy to intelligently optimize the user QoE in immersive holographic communication applications. We introduce novel behavioural and network awareness facets to account for uncertainties in user and source object behaviour and adverse network conditions. Furthermore, owing to the agnostic design approach, the proposed method has widespread applicability and can be employed in holographic communication applications that incorporate similar QoE performance metrics. Moreover, to our knowledge, we are the first group to propose a multidimensional adaptation strategy based on QoE performance metrics, particularly for emerging holographic communication.
- We conducted real-world experiments to demonstrate the effectiveness of the proposed approach, accounting for user and source object behaviours and, in particular, adverse network conditions caused by network delays and packet losses. The results of thorough evaluations revealed the feasibility and good performance of the proposed method. The detailed analysis results indicate that the proposed method can effectively satisfy user intent while accounting for various uncertainties and maintaining the user QoE. Additionally, comparisons with relevant state-of-the-art methods indicate that the proposed approach achieves the best performance.

II. Literature Review

In immersive and volumetric multimedia applications such as holographic communication, crucial factors related to user behaviour, environmental factors and the multimedia presentation strategy are critical for ensuring an immersive user experience. The influence of such factors has drawn considerable interest, with research dedicated to elucidating the impact of factors such as user navigation strategies, object/environment complexity, network limitations, compression performance, and interactivity on the immersive experience of the user and, more broadly, the QoE performance metrics of immersive and holographic multimedia systems [1], [2], [14]–[17].

For example, Park et al. [18] introduced an optimized rate-utility scheme for volumetric media streaming. Moreover, Van Der Hooft et al. [19] proposed an adaptive rate scheme for streaming compressed point cloud objects. The proposed scheme is a dynamic adaptive streaming over HTTP (DASH)-compliant framework [20]. The study of Gudumasu et al. [21] is similar to that of [19], who proposed a DASH-compliant scheme. The proposed scheme has some unique differences, with the primary content type being visual volumetric video-based coding (V3C). Other DASH-compliant schemes have been proposed in the literature, with Subramanyam et al. [22] proposing a DASH-compliant user-centred adaptive streaming framework for dynamic point clouds. Similarly, Farrugia et al. [23] proposed a DASH-compliant adaptive streaming framework for web-based 3D content. This framework was tailored for textured mesh-based 3D objects, incorporating compression and camera path navigation strategies and an index for perceivable quality [24].

Multiple studies have aimed to evaluate user behaviours and QoE in adaptive streaming applications. For example, Rossi et al. [12] investigated user behaviour in virtual reality (VR) environments and its subsequent influence on VR-based adaptive streaming systems. Wang et al. [25] proposed a QoE-driven framework for point cloud-based adaptive streaming. A tile-based adaptive approach, accounting for the user view and distance via a proposed QoE model, is employed in this framework. Furthermore, Minh et al. [11] conducted subjective experiments to evaluate the impact of several factors on mixed reality-based adaptive streaming. In addition, Subramanyam et al. [26] performed subjective assessments of a user-centric tiling-based adaptive framework for dynamic point cloud streaming, and Van Der Hooft et al. [14] focused on objective and subjective quality evaluations for adaptive point cloud streaming applications.

Although various studies have presented adaptive streaming approaches for immersive and volumetric media systems, four important matters must still be addressed. First, few works have considered adverse network conditions. Second, although several works have considered user behaviour/motion via predetermined camera paths, few studies have considered the uncertainties associated with both user and object behaviours. Third, few works have focused primarily on live streaming scenarios. Finally, few works have explored the impact of multidimensional adaptation strategies, with most works focusing only on the quality of the representations.

We aim to address these knowledge gaps through a holistic intelligent adaptation approach for immersive multimedia applications such as live holographic communication, accounting for the influence of user and object behavioural uncertainties and adverse network conditions. Furthermore, the proposed framework facilitates multidimensional adaptation. Importantly, the focus of this work is not compression; rather, we primarily consider user and object behaviours and adverse

network conditions.

III. Proposed Framework

A. Motivation and Problem Formulation

To address the knowledge gaps in the literature, we explore the impact of various behaviours and their associated uncertainties on user QoE. Here, we considered the user, source object, and dynamic network behaviours.

Previous studies have shown that user behaviour and the resulting changes, which could be considered the user intent [27], influence user experience. This conclusion was reached considering the characterization of the influence of user behaviour in previous works. For example, in [11], subjective studies were performed to evaluate the impact of user motion and the associated changes on the object viewing distance. The authors revealed that the perceived user experience is better at larger distances with reduced object quality. However, when closer objects are viewed, the object quality must be increased to achieve a similar perceived quality [11]. Thus, it is necessary to account for such behaviours to effectively satisfy user intent while maintaining optimal QoE.

To account for changes in user behaviour, which in this scenario is based on user motion, the relative user motion or distance must be quantified. To achieve this, a viewer threshold relative to the source object position can be utilized, as represented by Equation (1):

$$V_t = V_p - O_p \tag{1}$$

where V_t is the viewer threshold, V_p is the viewer position and O_p is the object position. Here, the viewer threshold is utilized to confirm whether the user intent has been satisfied according to I_{ct} and I_{ft}, which are the close and far proximity intent thresholds, respectively, as shown in Algorithm III-F1. On the basis of the viewer and intent thresholds, adaptations could be made so that the source object has higher quality when the viewer is closer and lower quality when the viewer is farther away. Thus, the user intent can be satisfied, and the optimal QoE can be maintained.

Furthermore, previous studies have revealed the influence of the characterization of the visual content on the QoE. Specifically, in the case of visual media, the behaviour of the media object, particularly its motion, influences the QoE [28]. Furthermore, when considering factors beyond object motion, the associated frame rate also impacts perceptive quality. For example, [29], [30] revealed that an increased frame rate increases the perceptual quality for visual media objects/content with fast camera and background motions. Similarly, the techniques with reduced frame rates are suitable for use in the case of visual media objects or content with slow motion [31]. Thus, the motion of the visual media content must be considered to ensure an optimal user experience. Building on this, we explore such behaviours in the context of holographic communication. The association between the frame rate and media object motion can be reflected through a simplified FPS threshold, as shown in Equation (2)

$$F_o = a.exp^{k.O_b(t)} \tag{2}$$

where F_o is the FPS threshold for an object, a and k are constants and $O_b(t)$ is the object behaviour at time *t*. Here, the object behaviour or object motion information is utilized to determine the FPS threshold. In particular, faster motion and a larger FPS threshold would result in a higher FPS for the object, whereas a slow or static object motion would result in a lower FPS for the object.

In addition, the network parameters and dynamic network conditions notably impact user QoE [1], [2], [13]. Specifically, the dynamic nature of network environments could lead to adverse and erroneous network conditions, and these adverse network conditions can significantly degrade user QoE [32]. Moreover, the impact of such deteriorative behaviour on key performance metrics such as quality and frame rate must be considered. In this context, it is necessary to consider dynamic network behaviours to facilitate an optimal user experience. For example, a previous study [32] indicated that in the case of adverse network conditions, a stable FPS is more important than the object quality level.

To account for dynamic network conditions, we represent the dynamic behaviour using a qualifier, as displayed in Equation (3)

$$Ci(t) = f(N_i(t)) \tag{3}$$

where $Ci(t)$ is the network impairment categorization at time *t* and $N_i(t)$ is the current network state for a given set of parameters i at time *t*. Using the impairment categorization, which is based on the network parameter information at a given time, the changes in both $QR_d(t)$ and $F_d(t)$, which are the adaptive changes in the quality of the representations and FPS at time *t*, can be represented using Equations (4) and (5), respectively.

$$QR_d(t) = f(V_t, Ci(t)) \tag{4}$$

$$F_d(t) = f(F_o, Ci(t)) \tag{5}$$

The aim is to optimally satisfy user intent considering the various behaviours and changes expressed in Equations (2) to (5). We mathematically represent this problem as shown in Equation (6) [33]:

$$P = optimal(VI = f(QR_d(t), F_d(t))) \tag{6}$$

where VI is the user intent, which depends on the changes in the resolution and FPS $QR_d(t)$ and $F_d(t)$. Owing to limitations in effectively obtaining a suitable solution using standard numerical optimization methods [33], the formulation P presents an issue. Similar issues are encountered for Equations (2) to (5). To adequately address these issues, machine learning is employed within the proposed framework, enabling various behaviours and their associated uncertainties to be effectively considered. In addition, analyses and quantitative assessments of the proposed framework are conducted using key performance indicators, the FPS and resolution, as both are quantitative performance metrics relevant to the user experience that also reflect the capability of the proposed system to satisfy the

user intent.

We aimed to explore the influence of different human-oriented and environment-oriented factors on the QoE performance metrics of the proposed framework. Fig. 1 shows the dependence of various performance metrics on these factors. The factors considered are the object motion, user behaviour and the dynamic nature of the network. Each factor clearly influences the adaptable parameters. The object behaviour directly influences the change in the FPS, with a lower FPS utilized for static objects and a higher FPS utilized for dynamic objects. The user behaviour is correlated with the user intent and directly impacts the quality of the representations, represented by the resolution. In this context, when the user has a specific intent, for example, when the user is close to or far from the hologram, the quality is increased or decreased appropriately. The change made corresponds to the literature, which provides information on the influence of the user distance and physical locomotion on the QoE. Finally, the dynamic network properties significantly influence the QoE because these properties impact both the quality of the representations and the number of FPS. Thus, the dynamic network properties are prioritized, which ensures that the network conditions that might impact other adaptation decisions are sufficiently considered. Specifically, this approach ensures that the user intent can be accurately satisfied. Furthermore, the proposed approach is flexible, as it considers performance trade-offs between the number of FPS and the quality of the representations, namely, the resolution. This enables more diverse and nuanced changes to be made, making the proposed approach more robust.

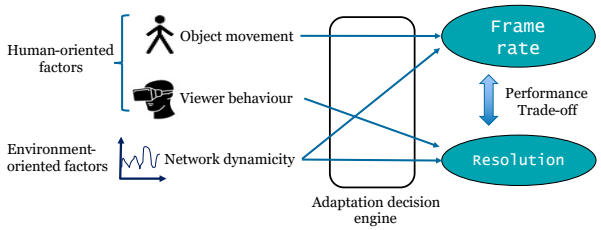

Fig. 1. Dependency of the performance metrics on external factors

B. System Overview

We aimed to design an effective framework, and the system configuration depicted in Fig. 2 was utilized. This setup displays the configuration for live holographic communication over a network, which occurs via the internet. The source object is placed within the range of a single (RGB-D) camera sensor, which is utilized to capture both the colour and depth information of the object. The colour and depth information of the source object is processed in conjunction with additional tracking data collated at the client to prepare the frames of the captured object in a 3D format. Additionally, the translation and orientation information of the camera are available from its calibration process, which is used to determine the relative position of the object in the environment. The client generates frames with a point cloud format. The point cloud frames consisting of both the spatial and colour information of each point are then prepared for transmission and sent to the server, at which point the necessary rendering and adaptation processes are performed. The server also maintains frame requests throughout the session to enable live holographic communication. The complete 3D photorealistic hologram of the object is then transmitted to the viewer equipment, such as an HMD, allowing the viewer to have unrestricted 6DoF views in the XR space.

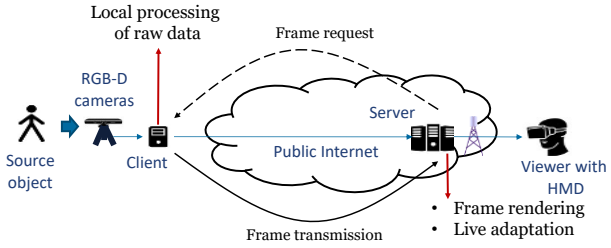

Fig. 2. Holographic communication application platform

C. Handling Viewer Behaviours

In live holographic communication, the user can view the environment with 6DoF. This creates a problem with respect to uncertainties in user behaviour, as this behaviour cannot be adequately predicted because of the spontaneous nature of the user. With the proposed framework, we attempt to address this issue by incorporating the user intention in the design of our adaptive framework. We apply an agnostic coding approach to account for user behaviour. Various behaviours or intentions might have different realization thresholds, which could then be used to determine whether a specific behaviour or intention has been carried out. This agnostic approach is implemented in the proposed framework via the introduction of intention thresholds that can be met on the basis of user behaviour.

For simplicity, we explain the proposed approach in the case of one user and one source object. The user intention and the physical motion of the user are considered for two main reasons. First, freedom of movement is a key requirement for immersive 6DoF environments, thus increasing the likelihood of such behaviour being performed by users. Second, user movement is important because of its impact on immersive media adaptive streaming and, in particular, the QoE [11], [22], [34]. Thus, user movement must be one of the fundamental behaviours considered when accounting for user intention. Fig. 3 depicts the user intention based on the user movement.

To realize this type of user intention, we exploit the availability of real-time user location information facilitated by 6DoF tracking and spatial mapping, as well as the intention tracking systems deployed in HMDs such as HoloLens 2 [27]. This information is utilized to calculate the user threshold relative to the intention threshold. When the intention threshold is met, the server and the intelligent adaptation framework are informed that the user intention has been triggered, and user behaviour-based adaptation is performed to satisfy the user intent.

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

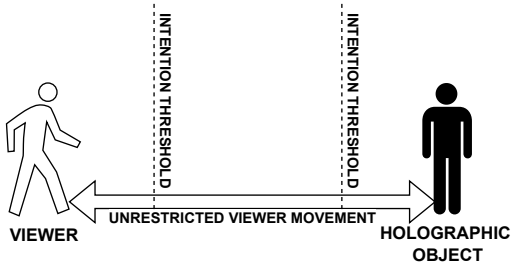

Fig. 3. User intention based on the user movement

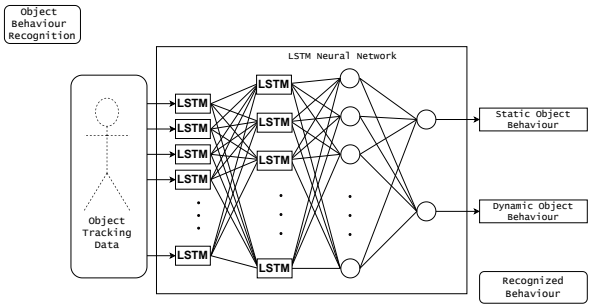

Fig. 4. Topology of the Object activity recognition neural network

D. Handling Source Object Movement Patterns

To account for the object behaviour in immersive multimedia applications such as live holographic communication, which is the focus of this paper, a machine learning approach is adopted. Furthermore, we strictly consider animated objects or participants, as such objects are primarily used in live holographic communication applications. We attempt to consider object awareness and focus on two primary object motion modes[1]. These two object motion modes are static and dynamic object behaviours, both of which are considered in the design of the neural network.

1) Design of the Object Behaviour Recognition Model: To adequately consider animated object behaviours, object behaviour recognition is assumed to be a human activity recognition (HAR) problem in our live holographic communication application. We utilize recurrent neural networks (RNNs) in our object motion mode recognition approach. Specifically, we use long short-term memory (LSTM) [35] neural networks to address this issue. We selected this type of network due to the suitability of LSTM networks for HAR [36]–[40].

We design a stacked LSTM neural network for object activity recognition. The topology of the network is illustrated in Fig. 4. Here, the network is made up of two connected LSTM layers and an additional hidden layer. To streamline the proposed framework, we use object tracking data rather than video data. This is facilitated via the utilization of Microsoft's Azure Kinect DK camera sensor and Azure Kinect Body Tracking SDK, allowing for the extraction of skeletal tracking data. The extracted skeletal data are used to form "skeletal frames" that consist of key points and their corresponding X, Y, and Z coordinate data, similar to data obtainable with other body tracking sensors. Offline model training utilizing frame data is conducted. The class label vector is fed into a *label encoder* (LE), with reshaping and normalization processes then performed. The skeletal frames that serve as inputs are fed into the neural network, which contains 3 layers and 318 neurons. The *ReLU* activation function [41] is utilized, and the *tanh* activation function [42] is also employed within certain layers. The *softmax* activation function [43] is utilized in the output layer, with the *Adam* [44] optimizer and $sparse\_categorical\_cross-entropy$ loss function applied. The outputs are limited to two classes at this stage; $static$ and $dynamic$ object behaviours. The object behaviours identified with the network depicted in Fig. 4 provide additional information for the framework, thereby enhancing the performance of the multidimensional adaptation strategy.

[1]In this paper, we consider only two simple modes, namely, static and dynamic modes, for simplicity and clarity. However, the proposed framework can be easily extended to support more fine-grained motion modes.

E. Handling Dynamic Network Conditions

To account for the dynamic network conditions in the proposed framework, a machine learning-based approach is employed. Error-induced transmission facilitated via the Netem network emulator [45] is conducted. Furthermore, various quality representation metrics, FPS values, which are the target values for the camera sensor, and delay and packet loss values are considered in the erroneous streaming conditions to adequately account for adverse network conditions and their impact on user-related metrics. Moreover, this approach enables effective handling of the dynamic network properties, with the information obtained via dedicated network monitoring [46] utilized for adaptation-related decision-making.

1) Design of the Network-Aware Adaptation Model: To effectively perform network aware adaptation, we aimed to develop an intelligent network aware adaptation model that uses the aforementioned key parameters to provide multiple quality representation recommendations. As such, we consider this problem to be a multiclass recommendation problem.

To solve this problem, offline model training is performed. The output vector, which contains each class value, is reshaped by feeding the vector into an LE, after which the subsequent output is then processed via a *one-hot encoding* approach. This step is performed to convert the text-based label data, which cannot be directly used in the neural network, to numeric data. Furthermore, one-hot encoding is applied for effective training [47]. The input parameters are fed into a *multilayer perceptron* (MLP) neural network with multiple hidden layers for training. The fully connected network contains six hidden layers, with 280 neurons, with the four input parameters projected to 128 neurons. In each hidden layer, the *ReLU* activation function [41] is used. The activation function was chosen because of its computational simplicity and suitability for training *feedforward* networks [41], [48]. The output layer has three output values, each corresponding to an output parameter. The *softmax* activation function [43] is utilized in the output layer, facilitating prediction probabilities, and the output parameter with the largest probability is recommended.

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

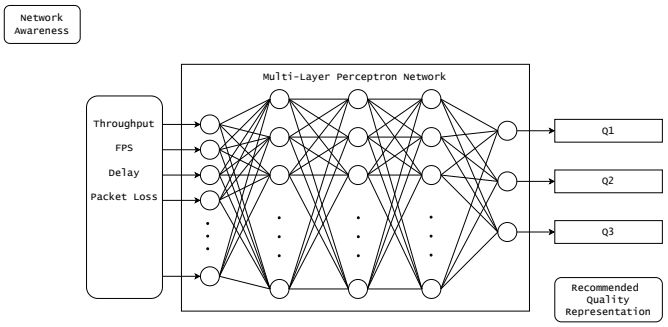

Fig. 5. Basic topology of the network-aware model

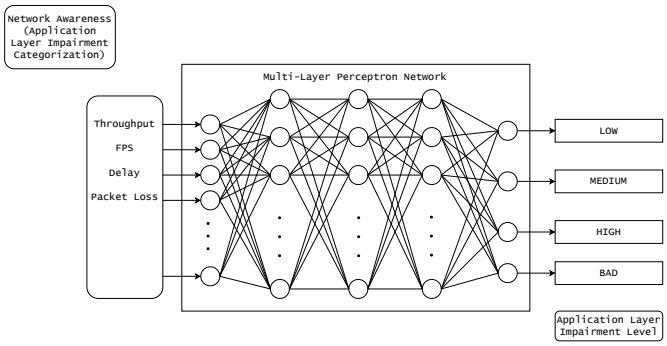

Fig. 6. Basic topology of the impairment categorization neural network

The loss function used in the network is the logarithmic loss function $categorical\_cross - entropy$, with the *Adam* [44] optimizer utilized for gradient-based optimization owing to its efficiency and simplicity. The basic topology of the neural network is depicted in Fig. 5.

Fig. 5 and Table I clearly show that the trained model focuses primarily on providing network-aware recommendations for adaptations based on the quality of the representations. The output parameters Q1, Q2 and Q3 represent varying quality levels or point densities. In relation to the framework and the overall system, these quality levels directly correspond to varying resolution levels, with Q1, Q2 and Q3 corresponding to the high definition (HD), full high definition (FHD) and wide quad high definition (WQHD) levels, respectively. Additional details related to the resolution/quality levels are presented in Table III [4]. To facilitate this multidimensional adaptation process, the network aware adaptation model is implemented in conjunction with the impairment categorization network described in subsection III-E2.

TABLE I
NETWORK AWARE MODEL PARAMETERS

Input Features	Throughput	FPS	Delay	Packet Loss
Output Parameters	Q1	Q2	Q3	

2) Design of the Impairment Categorization Model : To facilitate the effective performance of the network aware adaptation model, an impairment categorization model is designed. This model is used to enable adaptation of different output parameters. The model proposed in Subsection III-E1 strictly focuses on quality representation-based adaptation, limiting the ability of the framework to perform multidimensional adaptation, which we refer to as simultaneous quality representation and FPS adaptation. To address this issue, the impairment categorization model informs the framework of the top-level impairment categorization, allowing for better adaptation with the introduction of a second parameter, namely, the number of FPS. This model is very similar to the base network, with slight differences in the topology and output parameters. The model described in Subsection III-E1 outputs a recommended quality representation on the basis of the network awareness, but the model described in this subsection outputs an impairment category, as listed in Table II. The impairment level indicates the degree to which the normal or optimal performance of the application layer is impacted when considering the severity of the erroneous network conditions. For example, a low categorization implies that the system performance is stable and normal, whereas a bad categorization indicates that the erroneous network conditions have severe impacts on performance. Fig. 6 depicts the basic topology of the impairment categorization neural network. With the impairment categorization model and quality recommendation model, the proposed framework can handle user behaviours and dynamic network properties effectively in the network awareness-based adaptation process. The network-aware recommended quality ensures that the proposed framework can effectively make optimal adaptation decisions, particularly when the influence of user behaviour is considered. Specifically, this approach ensures that the proposed framework provides the most effective adaptation decisions while attempting to satisfy user intent. Furthermore, the impairment level provides additional context for the proposed framework, increasing its robustness and flexibility in making more fine-grained adaptation decisions to account for and address unexpected and atypical scenarios.

TABLE II
IMPAIRMENT CATEGORIZATION MODEL PARAMETERS

Input Features	Throughput	FPS	Delay	Packet Loss
Output Categorization	Low	Medium	High	Bad

F. Proposed Intelligent Adaptation Framework

Building on the approaches described in Subsections III-C, III-E and III-D, a novel user intention-driven network-aware intelligent framework is proposed. In the proposed framework, machine learning strategies are leveraged, a novel multidimensional adaptation approach for holographic communication is introduced, and trade-offs are optimized to effectively satisfy user intentions. The shrewd handling of the various behaviours ensures that the proposed framework appropriately accounts for the uncertainties associated with the user and object behaviours and the network environment. Furthermore, the optimization of the trade-offs considering the adaptable parameters ensures that the optimal user experience is prioritized. The design of the proposed framework is depicted in Fig. 7.

The figure displays the multifaceted awareness approach used within the proposed framework. The framework accounts

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

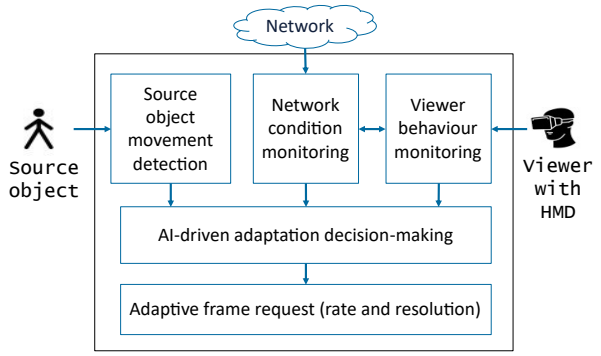

Fig. 7. Proposed framework

for human-oriented and environment-oriented uncertainties, leveraging AI for intelligent adaptation-related decision-making. Furthermore, the proposed framework is quite robust, as the various awareness-related strategies are independent, as observed with the source object movement detection and user behaviour monitoring processes. Additionally, because of its flexibility, the proposed framework enables simultaneous changes in the frame rate and resolution, which facilitates complex decision-making. To prevent logical conflicts in decision making that may arise due to the independence of the awareness-related strategies, an additional paradigm, the awareness dependency strategy, is incorporated in the proposed framework.

1) Awareness Dependency: One potential issue with the multidimensional awareness approach for the proposed adaptation framework is the possibility of logical conflicts in decision making due to inconsistent adaptation decisions. To prevent this issue, an awareness dependency strategy is utilized in the proposed framework. Here, the output of one awareness-related network influences the output of the other networks, such as the network aware and user intention aware strategies shown in Fig. 7. When the physical movement of the user is considered, adaptation decisions might be made on the basis of the proximity of the user to the object. However, the network might not be able to support significant increases in the quality of the object representation. To overcome this issue, the network-aware strategy provides the highest quality representations that the network can support, ensuring that the user intent can always be satisfied. Moreover, the impairment categorization model allows for greater flexibility in responding to adverse network conditions, thereby ensuring a stable user experience.

Building on the proposed framework design shown in Fig. 7, we leveraged the awareness dependencies to implement the algorithm outlined in Algorithm III-F1. Here, two conditions must be met. The first condition, $V_p - O_p > 0$, is necessary, as it is used to confirm that user movement-based behaviours are present. This step also enables the user threshold V_t to be calculated. The second condition, $SessionActive == true$, is required to ensure that the necessary inputs can be acquired. An animated object must be captured by the camera sensor to generate object tracking data O_{td}, and the server must have an active connection with both the client and the HMD. Provided that adverse network conditions are absent when $Ci(t) == Low$, the algorithm satisfies the user intention by providing the appropriate quality representation that can be supported by the network. In the case of adverse network conditions, with $Ci(t) \neq Low$, the framework remains flexible owing to the impairment awareness model, allowing for changes in both the quality of the representation $QR_d(t) = QR_C$ and the number of FPS $F_d(t) = F_C$ on the basis of the impairment level $Ci(t)$. The priority of the network-aware adaptation decisions under such conditions is determined on the basis of the awareness dependency strategy and network-aware qualifier N_AQ, which is a generic parameter utilized primarily within network-driven adaptation applications. With this approach, the framework attempts to provide a satisfactory experience for the user. In this context, a QoE-based model for holographic communication [32] is employed. Furthermore, the object behaviour awareness component based on the recognized activity $O_b(t)$ facilitates effective interactions and resource utilization through changes in the number of FPS.

IV. Experimental Environment

To assess the feasibility and performance of the proposed approach accurately, the configuration in Fig. 8 is employed. The proposed framework is implemented at the server, and extensive experiments are conducted. The figure shows the complete experimental configuration, and an extensively modified version of the LiveScan3D platform [49] was utilized for the experiments. A camera sensor placed in front of the animated source object, which in this scenario is a human, is used to capture relevant data required for the generation of a 3D hologram. The data are captured in point cloud format, with each point having unique spatial information (X, Y, and Z coordinates) and colour information (RGB value ranging from 0-255). The camera sensor used in the experiments was Microsoft's Azure Kinect DK. This camera was chosen as it can capture the relevant depth and colour information of the source object. Additionally, the camera sensor supports multiple quality representations and frame rates, as detailed in Table III [4]. The camera sensor, which is connected to a computer, referred to as the client, facilitates tracking of the source object. Additional processing is carried out at the client level to generate the relevant skeletal data and point cloud frames, which are continuously transmitted to the server to resolve the frame request by the server. The server leverages its powerful computing resources to perform the relevant management processes, and the point cloud data are processed via transformation and rendering operations, with the output sent to the HMD. The server plays an additional important role, as it houses the proposed framework, incorporating the user, source object and various network-aware intelligent adaptation models. In this setup, network links with a 1 Gbps bandwidth capacity are utilized to facilitate connectivity. Frame requests are continuously sent to maintain the live streaming session. Furthermore, a middle box is introduced in the experimental configuration to enable controlled experiments to be performed under adverse network conditions. The 6DoF views of the 3D holographic object allow unrestricted user movement;

Algorithm 1 Algorithm of the Proposed Framework

V_t = Viewer Threshold
I_{ct} = Close Proximity Intent Threshold
I_{ft} = Far Proximity Intent Threshold
V_p = Viewer Position
O_p = Object Position
$N_i(t)$ = Network aware input parameters at time *t*
O_{td} = Object Tracking Data
$O_b(t)$ = Object behaviour at time *t*
$Ci(t)$ = Application layer impairment categorization at time *t*
QR = Quality Representation
F = FPS
QR_N = Network aware recommended quality representation
QR_C = Impairment categorization recommended quality representation
F_C = Impairment categorization recommended FPS
$QR_d(t)$ = Quality representation adaptation decision at time *t*
$F_d(t)$ = FPS adaptation decision at time *t*
N_AQ = Network aware qualifier
N_AQ_{th} = Network aware qualifier threshold
Input: V_t, $N_i(t)$, O_{td}
Output: $QR_d(t)$, $F_d(t)$
Require: $V_p - O_p > 0$
Ensure: Session Active

while $SessionActive == true$ **do**
  if $Ci(t)$ == Low **then**
    if $V_t \leq I_{ct}$ **then**
      for $N_i(t)$, $i = 1, 2, ...N$ **do**
        $QR_d(t) = max(QR_N)$
      end for
    else if $I_{ft} \leq V_t$ **then**
      $QR_d(t) = min(QR)$
    end if
  else if $Ci(t) \neq Low$ **then**
    if $V_t \leq I_{ct}$ **then**
      for $N_i(t)$, $i = 1, 2, ...N$ **do**
        $QR_d(t) = QR_C$
        $F_d(t) = F_C$
      end for
    else if $I_{ft} \leq V_t$ **then**
      $QR_d(t) = min(QR)$
      $F_d(t) = F_C$
    end if
  end if
  while $N_AQ \geq N_AQ_{th}$ **do**
    for O_{td} *at time t* **do**
      if $O_b(t) == static$ **then**
        $F_d(t) = min(F)$
      else if $O_b(t) == dynamic$ **then**
        $F_d(t) = max(F)$
      end if
    end for
  end while
end while

thus, the uncertainties associated with the user and object behaviours and the impact of adverse network conditions are both considered in the experiments.

TABLE III
RESOLUTION LEVEL DETAILS

Resolution Level	Display Pixels	FPS	Point Density
HD	1280 x 720	5,15,30	418000
FHD	1920 x 1080	5,15,30	940000
WQHD	2560 x 1440	5,15,30	1671000

V. RESULTS

A. Framework Performance Analysis

To assess the effectiveness of the framework for supporting increases in the user QoE, experiments are conducted in which the user and object behaviours are considered. The network condition is controlled according to specific parameter settings so that the performance evaluations can be conducted under different network conditions.

The proposed framework enables effective and rapid adaptations, as it sufficiently accounts for the uncertainties associated with the user and object behaviours. Fig. 9 depicts the number of FPS, throughput and quality representation metrics of the proposed framework for diverse viewer and object behaviours. For clarity, we use resolution to represent the quality of the representation, with QR1, QR2 and QR3 corresponding to HD, FHD and WQHD, respectively. The resolution levels correlate with the number of display pixels, with HD corresponding to the worst quality, FHD corresponding to a higher quality level, and WQHD corresponding to the best quality. Furthermore, these resolutions indicate the object point density. Table III [4] lists the various resolutions or quality representations and their associated number of display pixels and relevant point densities. The FPS plot in the figure displays the flexibility of the framework in responding to different object behaviours. The results demonstrate that the framework can adequately identify multiple object behaviours and appropriately adjust in response to such behaviours. This is inferred from the algorithm-controlled reduction in the number of FPS, which occurs when a static object behaviour is detected, and the subsequent increase in the number of FPS, with the maximum FPS chosen when the object displays dynamic movements. The resolution results demonstrate the ability of the framework to satisfy user intentions. Here, when the user approaches the holographic object, the conditions associated with the user intention are satisfied, and the quality or resolution of the object is increased to the highest possible level in accordance with the user behaviour. This adaptation is facilitated by the stability of the network, with the resolution decreasing to the lowest level when the user is far from the holographic object. Furthermore, the FPS and resolution results highlight the feasibility of the multioutput parameter adaptation approach, as both the number of FPS and the resolution are adapted in response to the user and object behaviours. This is evident from the FPS results, which indicate that when the object is static and the user is close, the lowest number of FPS and highest resolution levels are chosen. In contrast, when

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

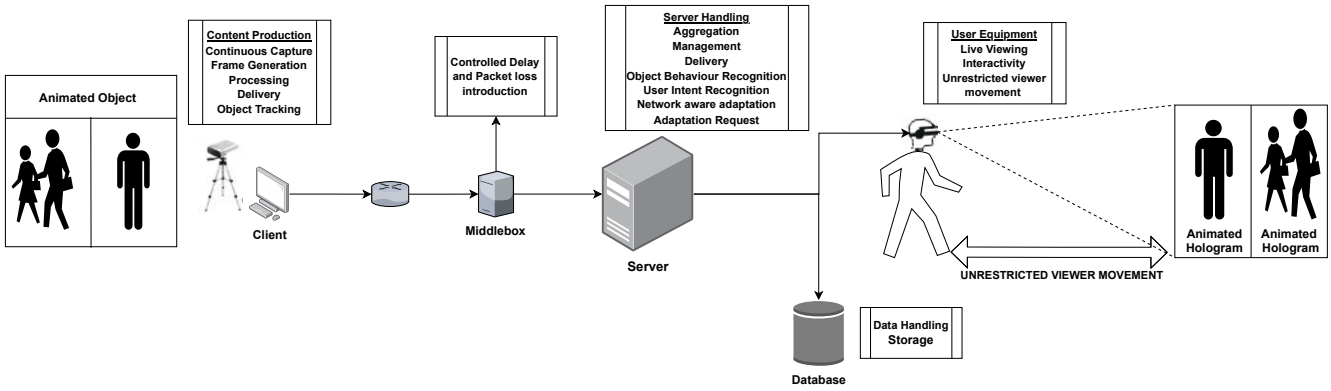

Fig. 8. Experimental configuration for the assessment of the proposed framework

the user is close and the object has dynamic movements, the highest number of FPS and resolution are chosen. The same response is observed when the user is far from the object, with the lowest resolution chosen and the number of FPS dependent on the object behaviour. Furthermore, the throughput plot confirms the effectiveness of the proposed framework. In particular, the throughput metric increases when the user is close to the object, with the throughput levels also dependent on the object behaviour. Moreover, the results demonstrate the effectiveness of the resources utilized by the framework, with less bandwidth utilization for static object behaviours and more bandwidth utilization for dynamic object behaviours. Additionally, the results support the feasibility of the framework for multioutput adaptation with no logical conflicts in satisfying the user intention while considering the object behaviour.

1) Analysis of the Effectiveness of the Facets Integrated in the Proposed Framework: To further evaluate the effectiveness of the proposed framework, additional experiments are conducted to investigate the importance of the integrated behaviour handling strategies. Fig. 10 shows the performance of the proposed framework without the source object behaviour handling and user behaviour handling strategies. The figure highlights the necessity of considering the source object behaviour. This is reflected in the inability of the proposed framework without the source object behaviour handling strategy to make suitable adaptations in response to the object behaviour, thus reducing its performance. In particular, the absence of adaptive FPS changes required for the proposed framework without the source object behaviour handling strategy indicates that this framework cannot adequately address changes in object behaviour, which in this case is static. This conclusion is further supported by the throughput results, which show significant increases in the throughput in the static object case when the source object behaviour handling algorithm is not applied in the adaptation process. Thus, the exclusion of the source object behaviour handling algorithm limits the flexibility and adaptability of the proposed framework, resulting in higher throughput for the same static object compared with that achieved with the complete framework. Although user intention can be somewhat accounted for by including the user behaviour and network handling algorithms, as shown by the resolution results, the omission of the source object behaviour handling component considerably degrades framework performance.

The second set of plots in Fig. 10 illustrates the performance of the framework without the user behaviour handling algorithm. The results demonstrate that the inclusion of this component is critical for ensuring the performance of the proposed framework. This conclusion is reached on the basis of the resolution results, which show that without this component, the proposed framework cannot adapt to effectively satisfy the user intent. This conclusion is further supported by the throughput results, which show no increases in throughput in response to the changes in the resolution. Moreover, although the inclusion of the source object behaviour handling algorithm in this scenario allows the proposed framework to account for the source object behaviour, the performance of the proposed framework remains significantly degraded.

Fig. 11 shows the effective performance of the proposed network when the dynamic network conditions are not considered. The figure shows that the framework can still account for the source object behaviour, which is dynamic in this case, and provides relevant adaptations to satisfy the user intent. This is inferred from the change in the number of FPS to 30 FPS, as well as the increases in resolution. However, under such conditions, the impact of the erroneous network conditions and the requirement for the inclusion of the dynamic network condition adaptation strategy is more apparent, with significant decreases in the number of FPS observed. In this particular scenario, the absence of the dynamic network condition adaptation strategy hinders the ability of the framework to make suitable and effective trade-offs in the adaptation process, limiting its ability to provide stable performance while satisfying user intent. This is further confirmed by the resolution and throughput results, which indicate that although the resolution increases to the WQHD level, the throughput remains fairly unstable. Thus, the source object behaviour, user behaviour and dynamic network condition adaptation strategies must all be included to achieve optimal framework performance and effectively satisfy user intent.

2) Visualization of the Framework Functionality: Fig. 12 depicts the rendered source object before and after the resolution adaptation strategy based on user intent is applied.

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

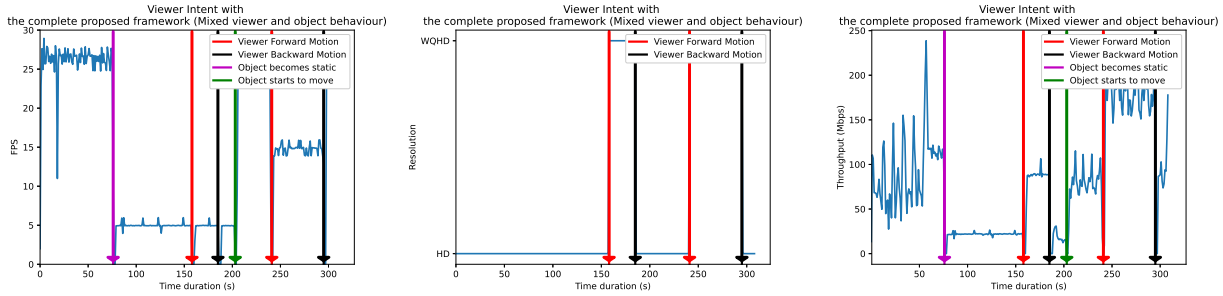

Fig. 9. Experimental results: Framework performance in the case of diverse user and object behaviours

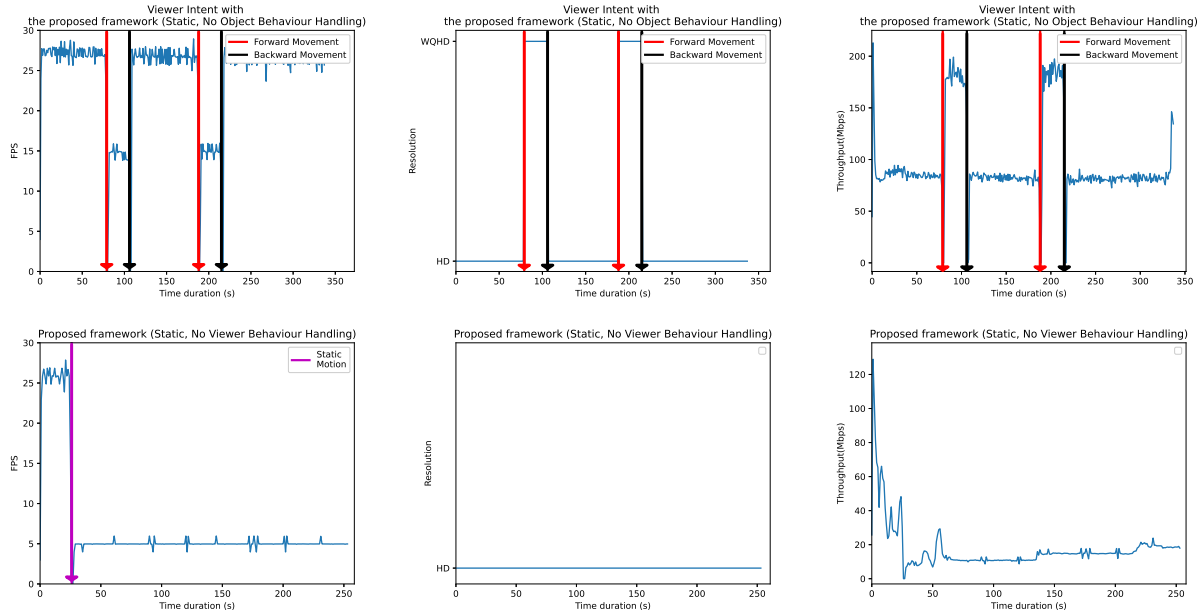

Fig. 10. Experimental results: Effective performance of the individual components of the proposed framework (source object movement detection, viewer behaviour monitoring)

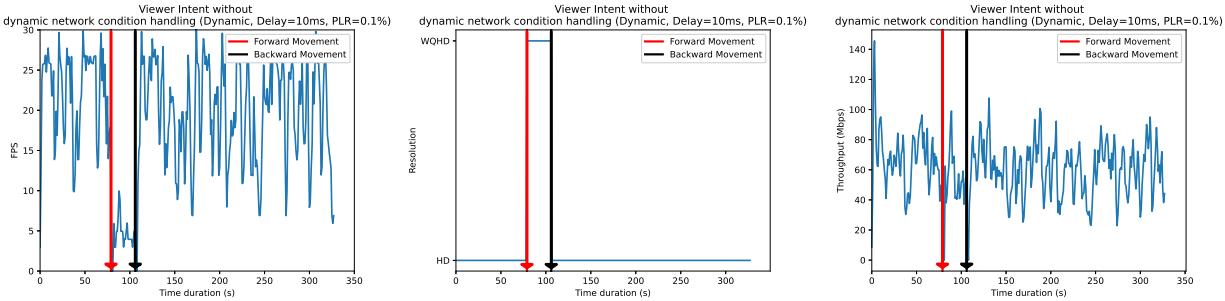

Fig. 11. Experimental results: Effective performance of the proposed framework without considering dynamic network conditions

The figure clearly shows that the proposed framework can effectively satisfy user intent while accounting for different human and environmental factors. The point density significantly increases after the resolution increases, as shown in Figs. 12(b) and 12(c). This is reflected in the visibility of finer details in the image, such as the clarity of the face. Additionally, Fig. 12(a) shows that if the user is far from the object, a lower object resolution is acceptable, as details are harder to perceive in such circumstances. Thus, the framework can effectively satisfy user intent while accounting for the uncertainties associated with different human and environmental behaviours.

3) Framework Performance with Static Object Behaviour Under Adverse Network Conditions: Fig. 13 displays the performance metrics of the proposed framework with static object behaviour under adverse network conditions with a 50 ms delay and 0.1% packet loss. The figure clearly shows that under adverse network conditions, the proposed framework is effective. The resolution results demonstrate that the proposed framework can sufficiently satisfy user intentions under such adverse conditions. In this context, the proposed framework, via the incorporation of network awareness algorithms, can provide recommendations of appropriate FPS and resolution levels to satisfy user intentions. This is inferred from the trade-off provided by the proposed approach, with the FPS reduced to ≈ 5, allowing for corresponding increases in quality levels to increase stability and satisfy user intentions. The FPS results for the proposed framework indicate that fairly stable performance is achieved even under such circumstances. This finding also implies that the proposed framework is suitable for maintaining the user QoE owing to its ability to remain stable in live streaming scenarios while satisfying user intentions. The resolution results provide further insight, indicating that the network cannot effectively satisfy user intentions at the highest resolution levels. However, the need to satisfy user intentions remains. As such, the proposed framework makes intelligent adaptations that account for the network state. The results demonstrate this, as shown in the resolution changes between the HD and FHD levels. The throughput results confirm these conclusions, with the changes in throughput corresponding to the adaptation decisions.

When the performance of the proposed framework is compared with that of the anchor implementation method (which considers only user behaviour) under adverse network conditions, it is evident that although the anchor implementation strategy aims to satisfy user intentions under such conditions, it cannot do so effectively. This is inferred from the FPS results of the anchor implementation method, as the FPS results are very unstable, with irregular non-adaptive changes in the number of FPS ranging from 0-30 FPS. Furthermore, the areas of the plot highlighted in light grey represent significant regions of concern. These sections outline the areas where adaptations are made to satisfy user intentions. The highlighted areas demonstrate the incompatibility of the user intention with the implementation results. This incompatibility is due to the inability of the network to adequately support adaptations made to satisfy user intent and is characterized by frequent zero or near zero FPS values. The zero or near zero FPS values of the anchor implementation strategy, particularly compared with those of the proposed framework under the same conditions, indicate that the proposed framework significantly outperforms the anchor implementation method. Moreover, the decrease in the number of FPS with the anchor implementation strategy indicates that this method could degrade or negatively impact the user QoE under such adverse network conditions. The throughput results confirm this assessment, as although the resolution results indicate that the highest resolution levels have been chosen to satisfy user intent, the reduced throughput values imply that the outputs are very unstable when the maximum resolution level is chosen.

4) Framework Performance with Dynamic Object Behaviour Under Adverse Network Conditions: Fig. 14 illustrates the performance of the proposed framework in the case of dynamic object behaviours under adverse network conditions with a 50 ms delay and 0.1% packet loss. The figure shows that the proposed framework can satisfy user intentions under adverse network conditions. The FPS results of the proposed framework indicate that although there are occasional decreases in the FPS value due to the impact of adverse network conditions, the proposed framework can achieve a somewhat stable FPS. This outcome is facilitated by the network awareness components and the inclusion of the awareness dependency strategy in the framework. Although the framework can accurately recognize the dynamic object behaviour, the utilization of the awareness dependency strategy allows the proposed framework to prioritize the network awareness component, thereby providing the most stable performance while satisfying user intent. This allows the framework to adapt as needed, resulting in reductions and increases in the FPS value to provide a more stable performance. Furthermore, the resolution results indicate the sufficient suitability of the proposed framework, demonstrating the flexibility of the framework in both decision-making and adaptations. This is characterized by increases in the resolution level with simultaneous reductions in the number of FPS, with the number of FPS reduced to 5 and the resolution increased to the FHD level. Furthermore, the framework can decrease the resolution to the lowest level, HD, as well as reduce the number of FPS to improve stability. Moreover, the throughput results confirm that the framework can provide somewhat stable performance while satisfying user intentions.

In comparison, the performance of the anchor implementation strategy, which considers only user behaviours, shows that although this method can adapt on the basis of user behaviours, it cannot sufficiently satisfy the user intent. This determination is made on the basis of the FPS results of the anchor implementation strategy. The FPS results of the anchor implementation strategy are unstable and erratic, with the FPS value ranging from 0-30 FPS, which differs from the FPS results of the proposed method under the same network conditions. Moreover, similar to Fig. 13, the anchor implementation strategy displays an intent incompatibility problem, as observed by the highlighted section with zero or near zero FPS levels. Furthermore, although the resolution results of the anchor implementation method indicate that this strategy can perform the relevant adaptations, the throughput

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

(a) Before adaptation(Far proximity, low resolution) (b) Before adaptation(Close proximity, low resolution) (c) After adaptation(Close proximity, High resolution)

Fig. 12. Experimental results: Framework functionality visualization

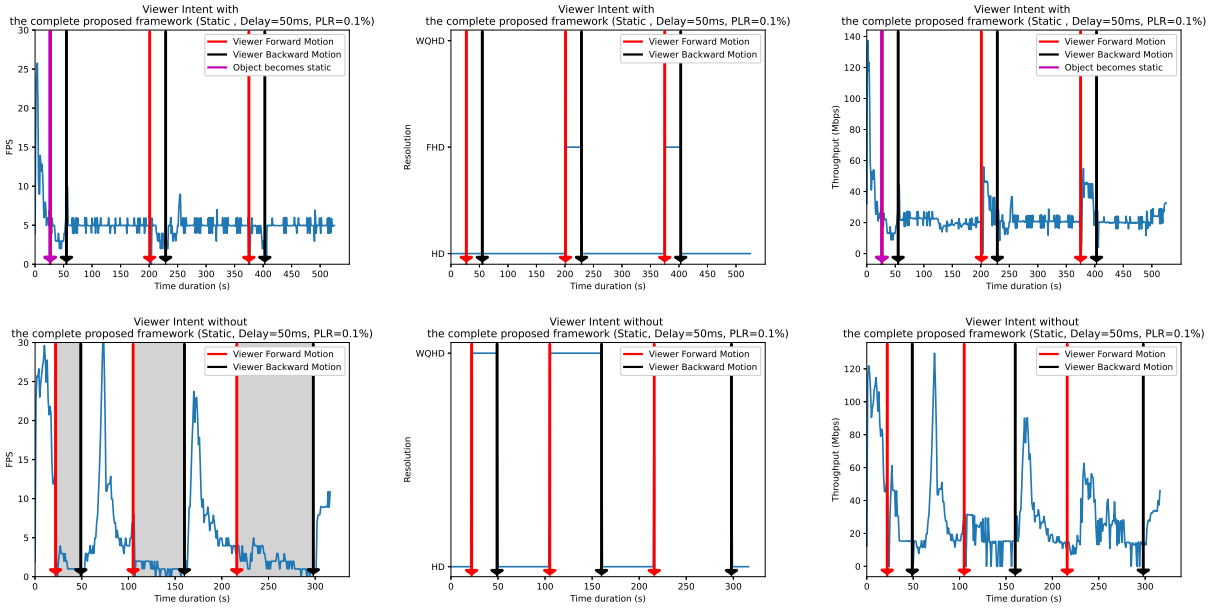

Fig. 13. Experimental results: Framework performance under different network conditions with static object behaviour (50 ms, 0.1% PLR)

results indicate that the anchor implementation method cannot suitably satisfy the user intention. A quantitative summary of the performance of the methods with both static and dynamic object behaviours can be found in Table IV.

TABLE IV
QUANTITATIVE SUMMARY OF THE PERFORMANCE OF THE PROPOSED FRAMEWORK IN SATISFYING USER INTENT (50 MS DELAY, 0.1% PLR)

Adaptation scheme	Adaptation	Resolution	FPS	Expected FPS
Proposed framework	Resolution, FPS	HD, FHD	≈ 5	5 (Adaptation)
Without proposed framework	Resolution	WQHD	0 - < 5	15 (WQHD Max)

B. Computational Complexity

To determine the impact of the proposed framework on the computational load and its subsequent complexity, a computational complexity analysis is performed. In this context, the increase in the execution time of the proposed framework is compared with the execution time of a simple anchor implementation strategy and is represented as a percentage. This percentage is calculated using Equation (7).

$$Complexity = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{PT_i - AT_i}{AT_i}\right) \times 100\% \qquad (7)$$

where PT_i and AT_i are the execution times of the proposed framework and anchor implementation strategies for a given iteration i, respectively.

TABLE V
COMPUTATIONAL COMPLEXITY OF THE PROPOSED FRAMEWORK

Adaptation scheme	Adaptation	Complexity (%)
Proposed framework	Resolution, FPS	20.02%

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

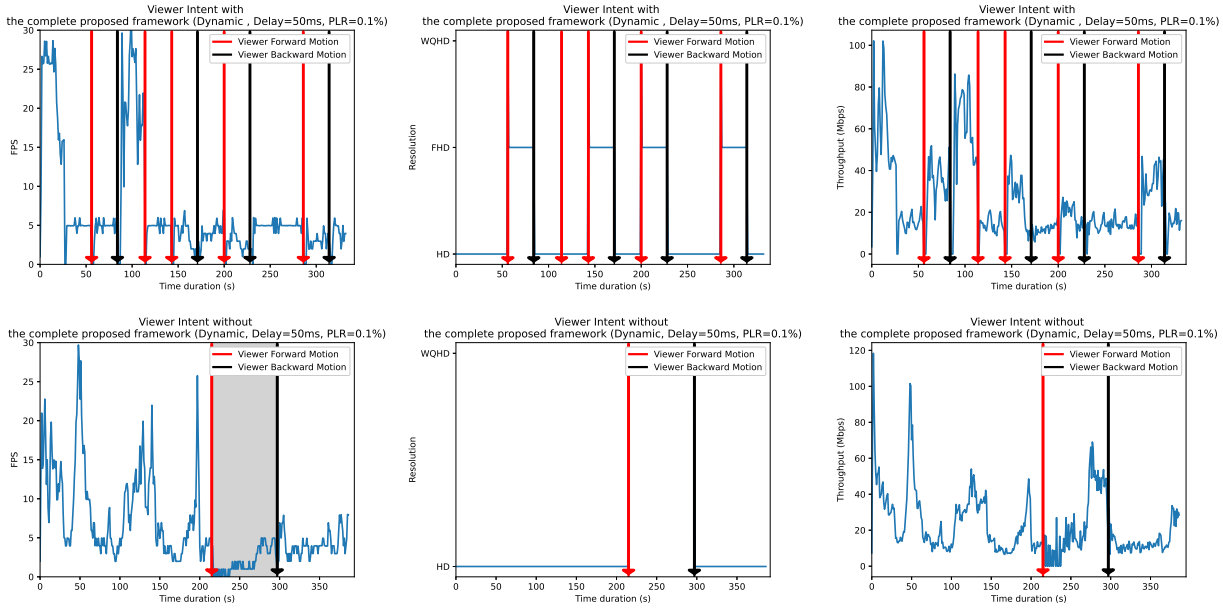

Fig. 14. Experimental results: Framework performance under different network conditions with dynamic object behaviour (50 ms, 0.1%PLR)

Table V displays the computational complexity of the proposed framework. The table clearly shows that the proposed framework is somewhat complex, with a 20.02% increase in computational complexity compared with that of the anchor implementation strategy. However, the increase is not extremely significant. This is due to the utilization of some multithreading strategies in the implementation of the proposed framework. Furthermore, the increase in complexity can be countered with further optimizations and improved computing capabilities.

C. Comparative Analysis

To analyse the performance of the proposed framework effectively and thoroughly, a comparative analysis is conducted. In this context, performance comparisons with state-of-the-art methods are presented. The state-of-the-art methods were selected for comparison owing to their recency and relevance, focusing on user intent/behaviour-centric volumetric media and adaptation strategies. The comparison approaches are listed below.

- Ref [22]: To provide a more accurate comparison analysis of our work with [22], we emulate this approach. To do this, we replicate the client-based selection of different content quality levels on the basis of the available network throughput (using bitrate targets) in our system. Additionally, as adaptation performed based on unrestricted user interaction/movement was considered, we factored this in our emulation. Furthermore, in this work, the FPS was kept at 30 FPS, with no discussion of FPS-based adaptation. Moreover, the tiling approach was not considered, as it extends beyond the scope of this work. Thus, to better compare the performance of the different approaches, our emulation of [22] provides adaptation results based on user movement, and the quality level is chosen on the basis of the available network throughput.
- Ref [27]: In the comparison analysis of our work with [27], we consider a user intention scenario, namely, a user proximity-driven resolution change scenario for online user intention processing. Furthermore, the variability of different paths falls outside the scope of this approach; thus, we assume that each path has an identical profile, with the changes occurring in one path reflected in the other paths. Thus, the emulation of the work presented in [27] considers online user intention processing and adaptation by adapting the resolution on the basis of the user proximity to the object.

Fig. 15 depicts a comparison of the performance of the proposed framework with that of the state-of-the-art method described in [27] in the case of static object behaviour under adverse network conditions with a 50 ms delay and 0.1% packet loss. The results demonstrate that the proposed framework outperforms the previous approach [27] in terms of its ability to satisfy user intent while maintaining stable performance. The FPS results indicate that although the impact of the erroneous network conditions can be observed, the proposed framework can make optimal adaptations to satisfy user intent. This differs from the results of the previous method described in [27], for which erratic FPS results are obtained, with 0 FPS and near 0 FPS values observed and significant decreases when considering user intention based on the proximity of the user to the object. Furthermore, although the resolution results indicate that both approaches made adaptations to satisfy user intent, the 0 and near 0 FPS values of the method described in [27] in the grey areas compared with the expected FPS values of 15 for the WQHD level highlight the resulting incompatibility of the intention with the output and the inability of this method [27] to effectively satisfy user intent under such adverse conditions. The suitability of the proposed framework is demonstrated by its multidimensional adaptation capability, leveraging the

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

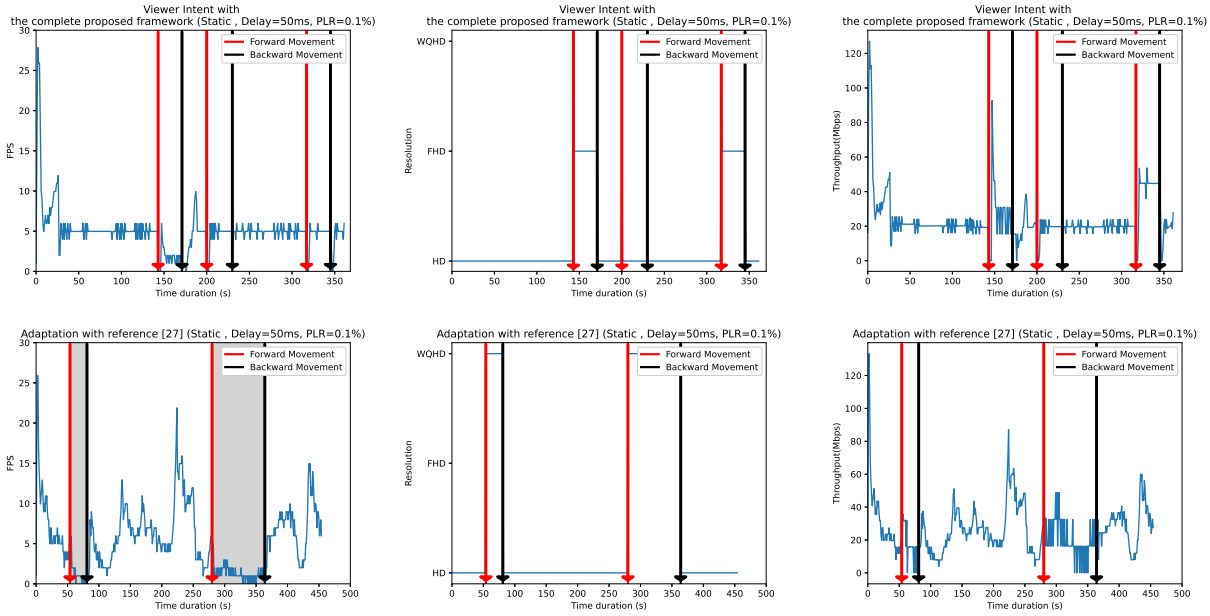

Fig. 15. Experimental results: Comparison of the performance of the proposed framework with that of state-of-the-art methods (Ref [27]) in the case of static object behaviour (50 ms, 0.1% PLR)

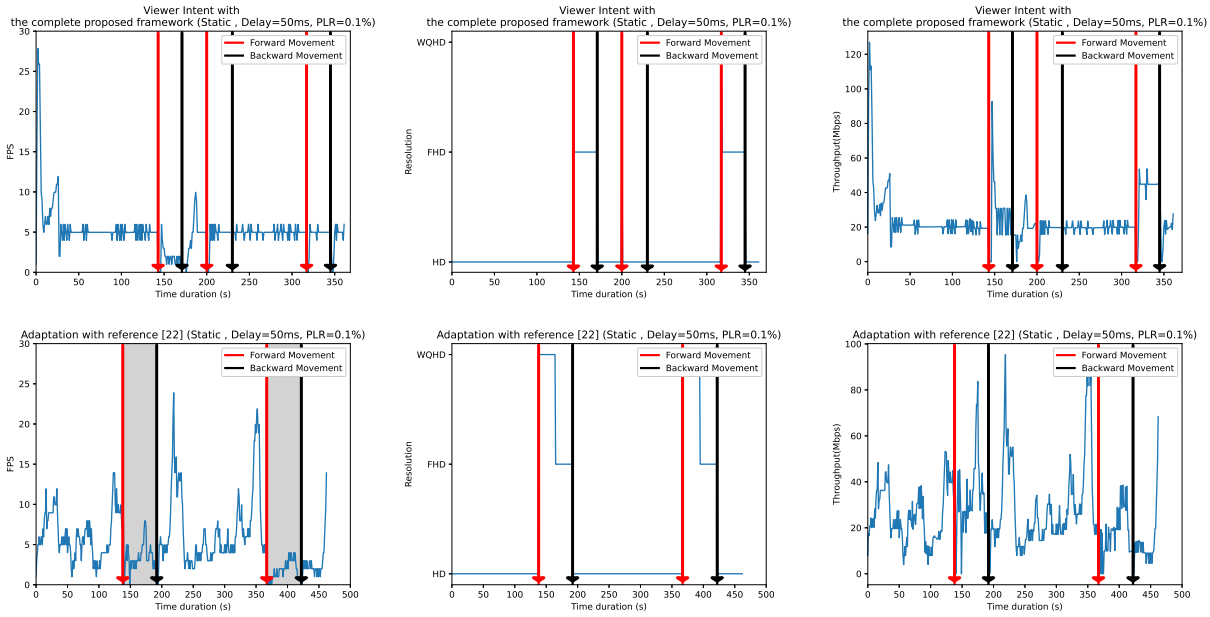

Fig. 16. Experimental results: Comparison of the performance of the proposed framework with that of state-of-the-art methods (Ref [22]) in the case of static object behaviour (50 ms, 0.1% PLR)

network awareness components in adapting to the FHD level rather than the WQHD level, as shown by the resolution results for the previous method [27], as well as the decrease in the FPS value to 5. Moreover, after the initial adaptation with a suboptimal performance, the proposed framework displays additional flexibility, emphasizing the stability of the user experience with further adaptation to the HD level, ensuring a satisfactory FPS under adverse network conditions. Then, the resolution is again changed to the FHD level while retaining a stable FPS value of 5, illustrating the intelligent capabilities of the proposed framework in making the most suitable and stable adaptations. The throughput results further confirm the good performance of the proposed framework, showing its stable performance and increases in throughput. Similarly, the unsuitable performance of the previous method [27] is confirmed, with decreases in throughput when making adaptation decisions.

Fig. 16 shows a comparison of the performance of the proposed framework with that of the state-of-the-art method described in [22] in the case of static object behaviour under adverse network conditions with a 50 ms delay and 0.1% packet loss. In contrast to the method in [27], the approach

in [22] can better satisfy user intent under such adverse conditions. This is observed from the ability of this method [22] to make relevant adaptations through its use of network-based adaptation strategies, as shown by its ability to adapt from the WQHD resolution level to the FHD resolution level on the basis of the network conditions. Although the method in [22] displays better performance than the method in [27], it does not outperform the proposed framework. This is inferred from the FPS results, which indicate that although the resolution can be effectively adapted, the FPS values of the method in [22] remain erratic, with 0 FPS and near 0 FPS values observed. Moreover, the decreases in the FPS value to $\leq$ 6 during periods in which the user intent is satisfied compared with the expected FPS values of 15 for the WQHD level and 25–30 for the FHD level indicate that this method [22] cannot effectively satisfy user intent under such adverse conditions. As such, the decrease in the FPS results confirm the severe degradation of the overall performance this method [22]. However, the proposed framework displays better performance than this previous approach [22]. This outcome is attributed to its multidimensional adaptation capability, as well as its flexibility in making such adaptations. In the proposed framework, an intelligent approach and other components are used to make the most suitable trade-offs, allowing it to provide stable performance and experience while satisfying user intent. This is characterized by the fairly stable FPS results due to its simultaneous adaptation of the resolution and FPS. This result is supported by the throughput results, confirming that the framework can provide somewhat stable performance while satisfying user intent. A quantitative summary of the comparison performance results is presented in Table VI.

TABLE VI
QUANTITATIVE SUMMARY OF THE PERFORMANCE OF VARIOUS METHODS IN SATISFYING USER INTENT (50 MS DELAY, 0.1% PLR)

Adaptation scheme	Adaptation	Resolution	FPS	Expected FPS
Proposed framework	Resolution, FPS	HD, FHD	≈ 5	5 (Adaptation)
Ref. [27]	Resolution	WQHD	0 - ≤ 3	15 (WQHD Max)
Ref. [22]	Resolution	WQHD, FHD	0 - ≤ 6	15(WQHD Max) 25-30(FHD)

VI. CONCLUSION

With the introduction of new features in the emerging holographic communication, handling user QoE has become significantly more challenging compared to traditional media. This new direction requires simultaneous consideration of a wide variety of factors including user intentions, source object behaviours and network conditions that may jointly affect user experiences. Considering this progress, in an effort to meet such demands, we propose a multidimensional media adaptation framework for live holographic communication.

The proposed framework was designed to accommodate the uncertainties associated with both user and object behaviours while also considering the influence of heterogeneous and adverse network conditions. In the design of the proposed framework, multiple machine learning approaches were employed to provide intelligent decision-making capabilities and enable the use of multidimensional adaptation approaches. Extensive experiments to assess the performance of the proposed framework under adverse network conditions with varying object and user behaviours were conducted. The results indicate that the proposed method can effectively satisfy user intentions under both stable and adverse network conditions, outperforming the current state-of-the-art methods, particularly under adverse conditions. The capabilities of the proposed framework will be extended and enhanced in future work. In particular, a larger range of user and object behaviours and dynamic network conditions could be considered, enhancing the flexibility and adaptability of the proposed framework and increasing its applicability in various scenarios. These improvements would further improve user experiences. Furthermore, the proposed method could be extended to support multisource streaming scenarios to satisfy user intentions in such cases. Finally, subjective experiments to assess the perceived user experience could be performed to elucidate the suitability of the proposed framework.

REFERENCES

[1] A. El Essaili, S. Thorson, A. Jude, J. C. Ewert, N. Tyudina, H. Caltenco, L. Litwic, and B. Burman, "Holographic communication in 5g networks," *Ericsson Technology Review*, vol. 2022, no. 5, pp. 2–11, 2022.

[2] I. F. Akyildiz and H. Guo, "Holographic-type communication: A new challenge for the next decade," *ITU Journal on Future and Evolving Technologies.*, 2022.

[3] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters, and F. D. Turck, "Toward truly immersive holographic-type communication: Challenges and solutions," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 93–99, 2020.

[4] P. Qian, V. S. H. Huynh, N. Wang, S. Anmulwar, D. Mi, and R. R. Tafazolli, "Remote production for live holographic teleportation applications in 5g networks," *IEEE Transactions on Broadcasting*, vol. 68, no. 2, pp. 451–463, 2022.

[5] Y. Huang, Y. Zhu, X. Qiao, X. Su, S. Dustdar, and P. Zhang, "Toward holographic video communications: A promising ai-driven solution," *IEEE Communications Magazine*, vol. 60, no. 11, pp. 82–88, 2022.

[6] B. Dalgarno and M. J. W. Lee, "What are the learning affordances of 3-d virtual environments?," *British Journal of Educational Technology*, vol. 41, no. 1, pp. 10–32, 2010.

[7] G. B. Petersen, G. Petkakis, and G. Makransky, "A study of how immersion and interactivity drive vr learning," *Computers Education*, vol. 179, p. 104429, 2022.

[8] A. Perkis, C. Timmerer, S. Baraković, J. B. Husić, S. Bech, S. Bosse, J. Botev, K. Brunnström, L. Cruz, K. De Moor, *et al.*, "Qualinet white paper on definitions of immersive media experience (imex)," *arXiv preprint arXiv:2007.07032*, 2020.

[9] S. Jain and D. Werth, "Current state of mixed reality technology for digital retail: A literature review," in *HCI in Business, Government and Organizations. eCommerce and Consumer Behavior* (F. F.-H. Nah and K. Siau, eds.), (Cham), pp. 22–37, Springer International Publishing, 2019.

[10] S. Rossi, I. Viola, L. Toni, and P. Cesar, "A new challenge: Behavioural analysis of 6-dof user when consuming immersive media," in *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 3423–3427, 2021.

[11] M. Nguyen, S. Vats, S. Van Damme, J. Van Der Hooft, M. T. Vega, T. Wauters, C. Timmerer, and H. Hellwagner, "Impact of quality and distance on the perception of point clouds in mixed reality," in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 87–90, 2023.

[12] S. Rossi, C. Ozcinar, A. Smolic, and L. Toni, "Do users behave similarly in vr? investigation of the user influence on the system design," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 2, pp. 1–26, 2020.

[13] S. Anmulwar, N. Wang, V. S. H. Huynh, S. Bryant, J. Yang, and R. Tafazolli, "Holosync: Frame synchronisation for multi-source holographic teleportation applications," *IEEE Transactions on Multimedia*, pp. 1–14, 2022.

[14] J. van der Hooft, M. T. Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, "Objective and subjective qoe evaluation for adaptive point cloud streaming," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, 2020.

[15] Y. Gao, X. Wei, J. Chen, and L. Zhou, "Toward immersive experience: Evaluation for interactive network services," *IEEE Network*, vol. 36, no. 1, pp. 144–150, 2022.

[16] G. Kougioumtzidis, V. Poulkov, Z. Zaharis, and P. Lazaridis, "Qoe assessment aspects for virtual reality and holographic telepresence applications," in *Future Access Enablers for Ubiquitous and Intelligent Infrastructures* (D. Perakovic and L. Knapcikova, eds.), (Cham), pp. 171–180, Springer International Publishing, 2022.

[17] I. Selinis, N. Wang, B. Da, D. Yu, and R. Tafazolli, "On the internet-scale streaming of holographic-type content with assured user quality of experiences," in *2020 IFIP Networking Conference (Networking)*, pp. 136–144, 2020.

[18] J. Park, P. A. Chou, and J.-N. Hwang, "Rate-utility optimized streaming of volumetric media for augmented reality," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 149–162, 2019.

[19] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner, "Towards 6dof http adaptive streaming through point cloud compression," in *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, (New York, NY, USA), p. 2405–2413, Association for Computing Machinery, 2019.

[20] I. Sodagar, "The mpeg-dash standard for multimedia streaming over the internet," *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, 2011.

[21] S. Gudumasu, G. Maury, A. Glasroth, and A. Hamza, "Adaptive streaming of visual volumetric video-based coding media," in *Proceedings of the 15th International Workshop on Immersive Mixed and Virtual Environment Systems*, MMVE '23, (New York, NY, USA), p. 30–33, Association for Computing Machinery, 2023.

[22] S. Subramanyam, I. Viola, A. Hanjalic, and P. Cesar, "User centered adaptive streaming of dynamic point clouds with low complexity tiling," in *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, (New York, NY, USA), p. 3669–3677, Association for Computing Machinery, 2020.

[23] J.-P. Farrugia, L. Billaud, and G. Lavoue, "Adaptive streaming of 3d content for web-based virtual reality: an open-source prototype including several metrics and strategies," in *Proceedings of the 14th Conference on ACM Multimedia Systems*, MMSys '23, (New York, NY, USA), p. 430–436, Association for Computing Machinery, 2023.

[24] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on graphics (TOG)*, vol. 30, no. 4, pp. 1–14, 2011.

[25] L. Wang, C. Li, W. Dai, S. Li, J. Zou, and H. Xiong, "Qoe-driven adaptive streaming for point clouds," *IEEE Transactions on Multimedia*, pp. 1–1, 2022.

[26] S. Subramanyam, I. Viola, J. Jansen, E. Alexiou, A. Hanjalic, and P. Cesar, "Subjective qoe evaluation of user-centered adaptive streaming of dynamic point clouds," in *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, 2022.

[27] P. Qian, N. Wang, C. H. Foh, C. Udora, and R. Tafazolli, "Enabling user intent-based network path adaptation for live volumetric streaming," in *2024 International Federation for Information Processing (IFIP) Networking Conference*, 2024.

[28] T. Lyko, Y. Elkhatib, M. Sparks, R. Ramdhany, and N. Race, "Qoe assessment for multi-video object based media," in *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–4, 2022.

[29] N. Van den Ende, H. de Hesselle, and L. Meesters, "Towards content-aware coding: User study," in *Interactive TV: a Shared Experience* (P. Cesar, K. Chorianopoulos, and J. F. Jensen, eds.), (Berlin, Heidelberg), pp. 185–194, Springer Berlin Heidelberg, 2007.

[30] Y. Shen, Y. Liu, N. Qiao, L. Sang, and D. Yang, "Qoe-based evaluation model on video streaming service quality," in *2012 IEEE Globecom Workshops*, pp. 1314–1318, 2012.

[31] G.-M. Su, X. Su, Y. Bai, M. Wang, A. V. Vasilakos, and H. Wang, "Qoe in video streaming over wireless networks: perspectives and research challenges," *Wireless networks*, vol. 22, pp. 1571–1593, 2016.

[32] C. Udora, P. Qian, S. Anmulwar, A. Fernando, and N. Wang, "Quality of experience modelling and analysis for live holographic teleportation," in *2024 International Conference on Computing, Networking and Communications (ICNC): Multimedia Computing and Communications*, 2024.

[33] W. Wen, J. Yan, Y. Zhang, Z. Huang, L. Liang, and Y. Jia, "Adaptive cooperative streaming of holographic video over wireless networks: A proximal policy optimization solution," *arXiv preprint arXiv:2406.08806*, 2024.

[34] S. Van Damme, I. Mahdi, H. K. Ravuri, J. van der Hooft, F. De Turck, and M. T. Vega, "Immersive and interactive subjective quality assessment of dynamic volumetric meshes," 2023.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] J. D. Domingo, J. Gómez-García-Bermejo, and E. Zalama, "Improving human activity recognition integrating lstm with different data sources: Features, object detection and skeleton tracking," *IEEE Access*, vol. 10, pp. 68213–68230, 2022.

[37] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, "Stacked lstm network for human activity recognition using smartphone data," in *2019 8th European Workshop on Visual Information Processing (EUVIP)*, pp. 175–180, 2019.

[38] S. W. Pienaar and R. Malekian, "Human activity recognition using lstm-rnn deep neural network architecture," in *2019 IEEE 2nd Wireless Africa Conference (WAC)*, pp. 1–5, 2019.

[39] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "Lstm networks for mobile human activity recognition," in *Proceedings of the 2016 International Conference on Artificial Intelligence: Technologies and Applications*, pp. 50–53, Atlantis Press, 2016/01.

[40] A. A. Aljarrah and A. H. Ali, "Human activity recognition using pca and bilstm recurrent neural networks," in *2019 2nd International Conference on Engineering Technology and its Applications (IICETA)*, pp. 156–160, 2019.

[41] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[42] M. H. Essai Ali, A. B. Abdel-Raman, and E. A. Badry, "Developing novel activation functions based deep learning lstm for classification," *IEEE Access*, vol. 10, pp. 97259–97275, 2022.

[43] D. Zhu, S. Lu, M. Wang, J. Lin, and Z. Wang, "Efficient precision-adjustable architecture for softmax function in deep learning," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 12, pp. 3382–3386, 2020.

[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[45] S. Hemminger *et al.*, "Network emulation with netem," in *Linux conf au*, vol. 5, p. 2005, Citeseer, 2005.

[46] G. Miranda, D. F. Macedo, and J. M. Marquez-Barja, "A qoe inference method for dash video using icmp probing," in *2020 16th International Conference on Network and Service Management (CNSM)*, pp. 1–5, 2020.

[47] A. Y. Hussein, P. Falcarin, and A. T. Sadiq, "Enhancement performance of random forest algorithm via one hot encoding for iot ids," *Periodicals of Engineering and Natural Sciences*, vol. 9, no. 3, pp. 579–591, 2021.

[48] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (G. Gordon, D. Dunson, and M. Dudík, eds.), vol. 15 of *Proceedings of Machine Learning Research*, (Fort Lauderdale, FL, USA), pp. 315–323, PMLR, 11–13 Apr 2011.

[49] M. Kowalski, J. Naruniec, and M. Daniluk, "Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors," in *2015 International Conference on 3D Vision*, pp. 318–325, 2015.

APPENDIX

A. Results

1) Framework Performance with Static Object Behaviour Under Stable Network Conditions: Fig. 17 shows the performance metrics of the proposed framework with static object behaviour under stable network conditions with a 0 ms delay and 0% packet loss. The figure clearly shows that the proposed framework, which considers user and object behaviours, can effectively perform intelligent adaptations to satisfy user intent. Moreover, the FPS results demonstrate the effectiveness of the proposed framework. Upon the detection of static object behaviour, the necessary adaptation, which in this scenario is reducing the number of FPS to 5, is made. Furthermore, the resolution results highlight the suitability of the proposed framework in satisfying user intent. In this context, the framework satisfies user intent while also accounting for object behaviour. The throughput results provide additional insight, with increases in throughput corresponding to forwards user movement, indicating that more data were utilized to provide higher-quality representations of the object.

We next compared the performance of the proposed framework with that of the anchor implementation strategy, in which only user movement is considered, under the same stable network conditions. We can infer that the anchor implementation strategy can also satisfy user intent under such conditions, as shown by the resolution results. The FPS results for the anchor implementation strategy indicate that this method does not account for the object behaviour, since a higher FPS of ≈ 30 is utilized for the same static object. This conclusion was confirmed by the throughput results, showing the increased bandwidth under these conditions. Thus, the proposed method provides more effective resource utilization than the anchor implementation method.

2) Framework Performance with Dynamic Object Behaviour Under Stable Network Conditions: Fig. 18 shows the performance of the proposed framework in the case of dynamic object behaviour under stable network conditions with a 0 ms delay and 0% packet loss. The figure results indicate that the proposed framework can effectively satisfy user intent via the intelligent adaptation strategies while accounting for user and object behaviours under such ideal conditions. The FPS results provide further insight. After the object behaviour is determined, the maximum FPS supported for the corresponding resolution is provided. Furthermore, the resolution results indicate the suitability and flexibility of the proposed framework in attempting to satisfy user intent. The flexibility of the proposed framework is displayed by its ability to freely choose between the WQHD resolution, which allows for the highest quality but not the highest FPS of $\geq$ 15 FPS, and the FHD resolution, which provides higher quality than the HD level but lower quality than the WQHD level with a higher FPS of $\approx 25+$ in an effort to satisfy user intent. The throughput results validate these findings, showing increases in throughput, indicating that the proposed framework can effectively satisfy user intent under such stable conditions.

A comparison of the performance of the proposed framework with that of the anchor implementation strategy, which accounts for only user behaviour, indicates that the anchor implementation strategy can also satisfy user intentions under such stable conditions. However, the anchor implementation strategy does not account for object behaviour. Furthermore, the FPS and resolution results indicate that the anchor implementation strategy is not flexible, thus limiting its ability to satisfy user intent.

3) Comparative Analysis: Figs. 19 and 20 display a comparison of the performance of the proposed framework with that of the state-of-the-art methods described in [27] and [22] in the case of static object behaviour under stable network conditions with a 0 ms delay and 0% packet loss. The figure shows that, similar to the proposed framework under such stable network conditions, the approaches described in [27] and [22] can provide the necessary adaptations to satisfy user intent. This is evident from the resolution results, which indicate that the quality level or resolution is increased on the basis of the user motion and proximity to the object, providing a satisfactory response. This conclusion is confirmed with increases in the throughput, corresponding to the adaptation decisions made. However, although the methods described in [27] and [22] display suitable performance in satisfying user intent, both adaptation schemes fail to effectively account for the source object behaviour. This is observed from the FPS results, which show that no adaptations based on the source object behaviour are made. The changes in the FPS for the methods described in [27] and [22] correspond with the maximum FPS of 15 for the WQHD level supported by the hardware. Thus, although the methods in [27] and [22] can satisfy user intent, the proposed framework provides more intelligent decisions and increased flexibility in addressing uncertainties associated with source object behaviour, as reflected in its multidimensional adaptability capacity.

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

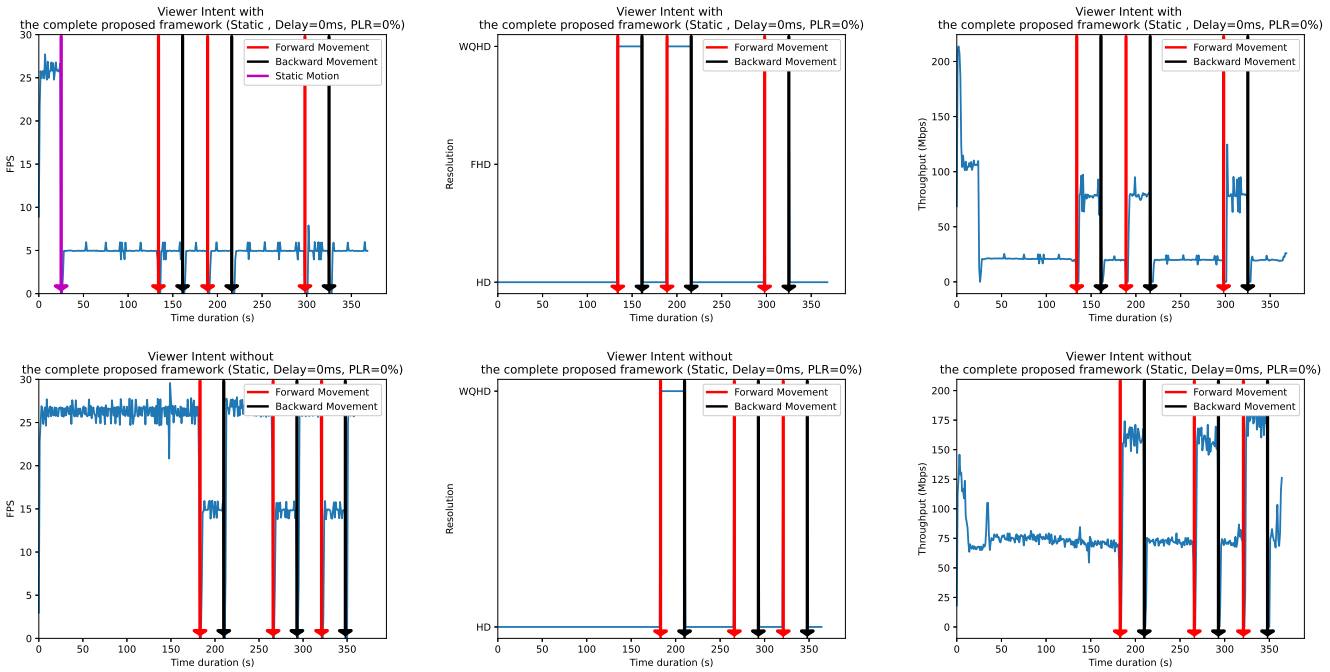

Fig. 17. Experimental results: Framework performance with static object behaviour under stable network conditions (0 ms, 0% PLR)

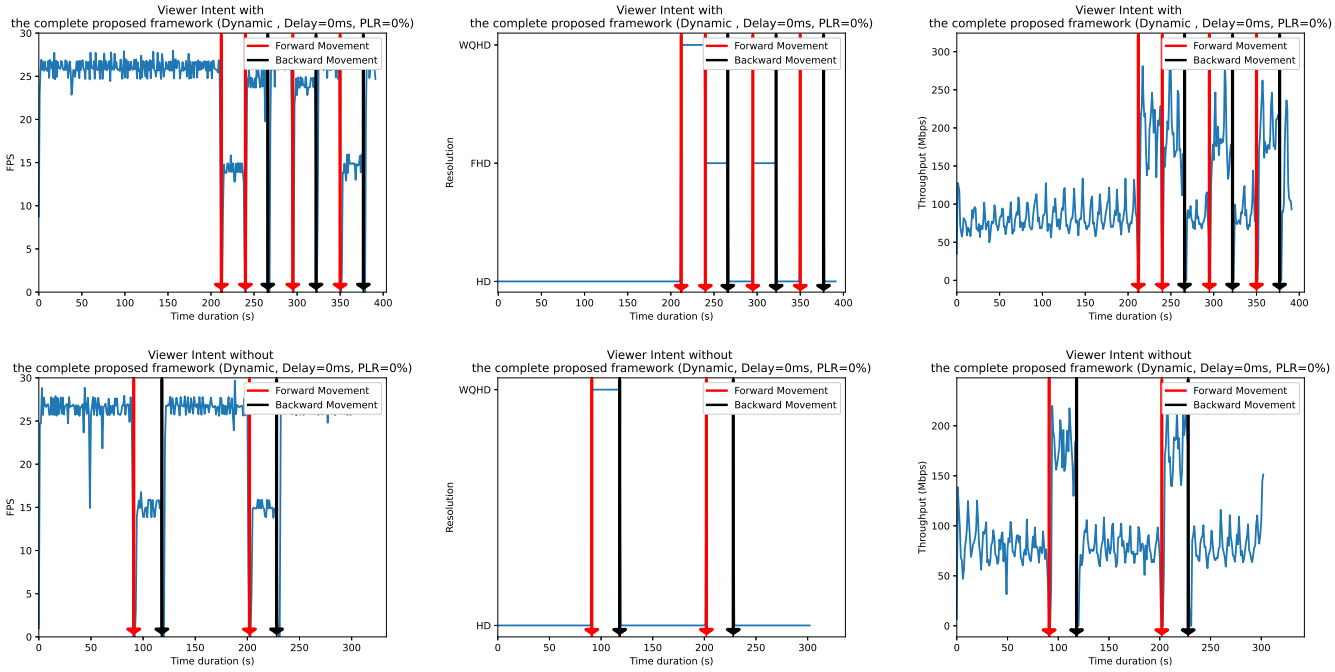

Fig. 18. Experimental results: Framework performance with dynamic object behaviour under stable network conditions (0 ms, 0% PLR)

This article has been accepted for publication in IEEE Transactions on Multimedia. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMM.2025.3623505

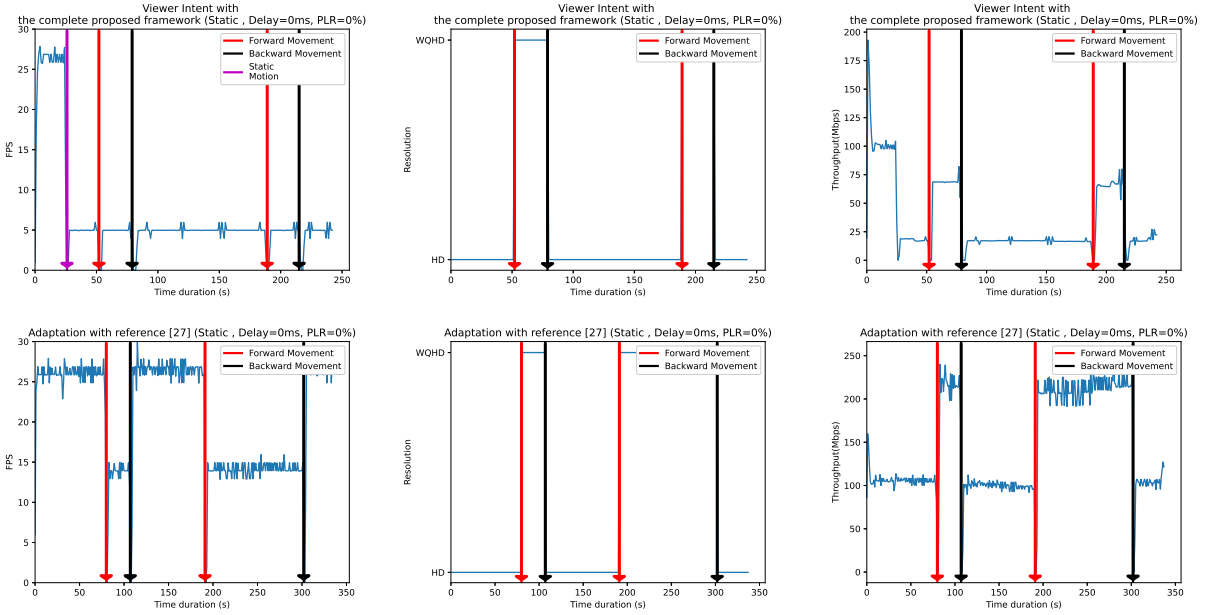

Fig. 19. Experimental results: Comparison of the performance of the proposed framework with that of state-of-the-art methods (Ref [27]) in the case of static object behaviour (0 ms, 0% PLR)

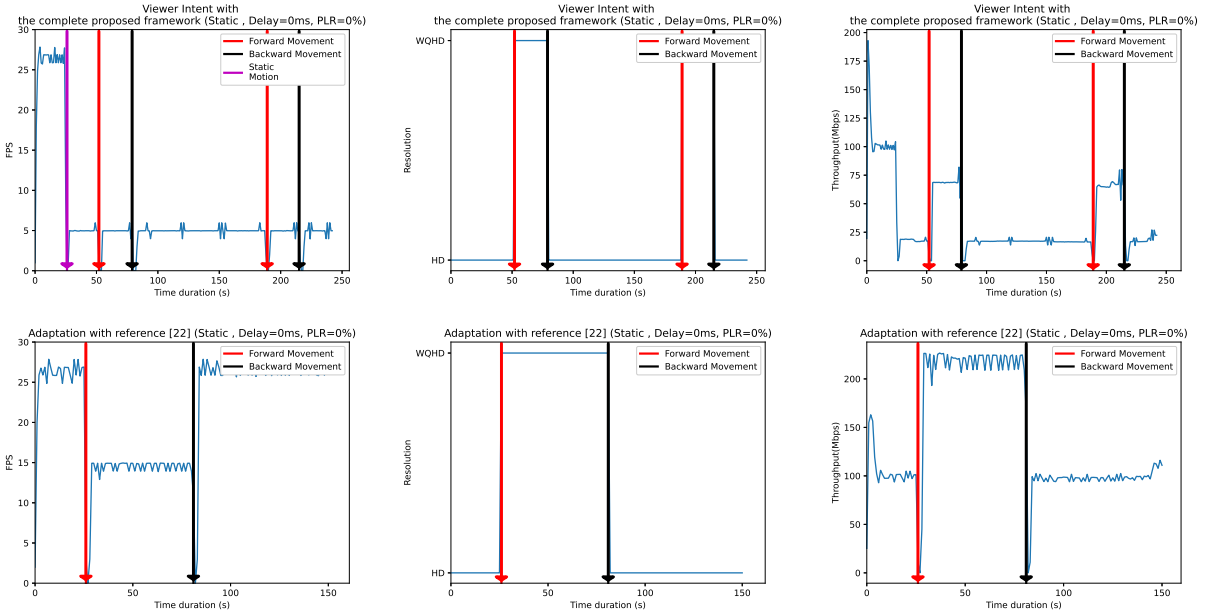

Fig. 20. Experimental results: Comparison of the performance of the proposed framework with that of state-of-the-art methods (Ref [22]) in the case of static object behaviour (0 ms, 0% PLR)