Towards Immersive Communications in 6G

Xuemin (Sherman) Shen [1], Jie Gao [2], Mushu Li [3,*], Conghao Zhou [1], Shisheng Hu [1], Mingcheng He [1], Weihua Zhuang [1]

[1] *Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada*
[2] *School of Information Technology, Carleton University, Ottawa, Ontario, Canada*
[3] *Department of Electrical, Computer, and Biomedical Engineering, Toronto Metropolitan University, Toronto, Ontario, Canada*

Correspondence*:
Mushu Li
mushu1.li@ryerson.ca

ABSTRACT

The sixth generation (6G) networks are expected to enable immersive communications and bridge the physical and the virtual worlds. Integrating extended reality, holography, and haptics, immersive communications will revolutionize how people work, entertain, and communicate by enabling lifelike interactions. However, the unprecedented demand for data transmission rate and the stringent requirements on latency and reliability create challenges for 6G networks to support immersive communications. In this survey article, we present the prospect of immersive communications and investigate emerging solutions to the corresponding challenges for 6G. First, we introduce use cases of immersive communications, in the fields of entertainment, education, and healthcare. Second, we present the concepts of immersive communications, including extended reality, haptic communication, and holographic communication, their basic implementation procedures, and their requirements on networks in terms of transmission rate, latency, and reliability. Third, we summarize the potential solutions to addressing the challenges from the aspects of communication, computing, and networking. Finally, we discuss future research directions and conclude this study.

Keywords: 6G Networks, Immersive Communications, Extended Reality, Haptic Communication, Holographic Communication.

1 INTRODUCTION

Ever since its birth, the communication technology has been a symbol of the modernization of human society, and the evolution of communication technology has accompanied the advance of civilization. The commercialization of electrical telegraph and telephone during the second industrial revolution boosted globalization by facilitating finance and trade overseas (Wenzlhuemer, 2013). The debut of vehicle-mounted mobile radio systems (“car phones”) and the analog first generation (1G) mobile telecommunication systems from the 1950s to 1980s enabled voice calls on the go (del Peral-Rosado et al., 2018). The second generation (2G) mobile communication systems, which introduced roaming and preliminary data services in the form of text messages, emerged amidst and as a part of the third industrial revolution (i.e., the digital revolution) (Billström et al., 2006). Then, the next two decades witnessed the proliferation of mobile Internet and mobile multimedia services brought by the third and fourth generation (3G and 4G) mobile communication technology, which revolutionized how people communicate and changed the world.

Nowadays, the fifth generation (5G) mobile communication systems are reshaping industries by facilitating the fourth industrial revolution (i.e., Industry 4.0) towards smart inter-connectivity and automation (Chen et al., 2021a).

Accustomed to the convenience brought by the latest communication technologies, many people may not realize that ordinary daily activities such as video calls or zoom meetings were nothing more than science fiction merely three decades ago. Indeed, from the so-called "telephot" in the pioneering novel "Ralph 124C 41+" to the video call scene in the classic movie "Back to the Future", the simultaneous transmission of live image and sound was considered as a "technology of the future" in the most part of the twentieth century (Gooday, 2005; Fowler et al., 1986). When the fantasy of the past has become a reality, a question that naturally arises is: what will be the next revolutionary form of communications, potentially in the era of the sixth generation (6G)? Fortunately, we may again find clues in science fiction, with examples ranging from the famous scene of Princess Leia's three-dimensional (3D) holographic message in "Star Wars" (Conti, 2008) to the virtual world "OASIS" in the metaverse presented in the recent film "Ready Player One" (Sparkes, 2021). The fact that such scenes created a long-lasting impact on a vast audience reflected people's desire for more lifelike, immersive, and interactive communications (Xu et al., 2022).

Unfolding exactly as depicted in science fiction or not, immersive communications will come to reality and shift the current communication paradigm in three aspects. First, rather than two-dimensional (2D) images displayed on a flat screen, immersive communications will deliver 3D images with parallax information. Second, in addition to audiovisual information, immersive communications will involve haptic information. Third, the pursuit of immersive experiences will further blur the boundary between the physical and the virtual worlds, allowing new forms of interactions across the two worlds. These paradigm shifts can significantly enrich communication experiences of users and enable a plethora of new use cases such as holographic 3D telepresencing (Zhang et al., 2022a), ultra-realistic online interactive sports (Next G Alliance, 2022), and immersive learning in education (Pellas et al., 2020), to name a few. In particular, immersive communications can also enable human-machine collaboration in industrial environments and propel the next industrial revolution, i.e., industrial 5.0 (Maddikunta et al., 2022). As a result, immersive communications are expected to have a profound impact on the landscape of communication industries and impact how people study, work, and entertain in the years to come.

Motivated by the potentials of immersive communications, scientists and engineers over the world have been working on the development of related technologies, products, and platforms. Significant progress has been made in recent years, including but not limited to advancements in sensor systems and data capture techniques (Meyer et al., 2022; Dahiya et al., 2019), data processing and computing frameworks (Song et al., 2022; Petkov et al., 2022; Qian et al., 2022), and rendering and display devices (Hirayama et al., 2019; Xiong et al., 2021; Schmitz et al., 2020). Some component development of immersive communications is progressing faster than others, leading to the establishment of testbeds, prototypes, or even commercial products. Virtual reality (VR), as an example, has gained popularity, especially in the gaming industry (Jung et al., 2020). Devices such as VR headsets and haptic glove development kits are available in the market (Chen et al., 2022; Kugler, 2021), while researchers are building testbeds for extended reality (XR) (Huzaifa et al., 2022) and human-machine interaction with haptic feedback (Gokhale et al., 2020). As the aforementioned progress and efforts are paving the way for realizing immersive communications, advancements in communication and networking technologies will be indispensable.

Despite the advent of 5G systems and the accompanying advancements in network capabilities, there are still many challenges to achieving immersive communications in various aspects of communications, networking, and computing. The data rate required to transmit live 3D images can be so large, e.g., on the

Table 1. List of main acronyms

1G - 6G	First Generation - Sixth Generation	2D	Two-dimensional
3D	Three-dimensional	3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence	AR	Augmented Reality
D2D	Device-to-device	DetNet	Deterministic Networking
DoF	Degree of Freedom	FoV	Field-of-view
HI	Haptic Interface	IMU	Inertial Measurement Unit
IRS	Intelligent Reflecting Surface	JND	Just-noticeable Difference
LDPC	Low-density Parity-check	LFV	Light-field Video
LIDAR	Light Detection and Ranging	LoS	Line-of-sight
LSTM	Long Short-term Memory	MEC	Mobile Edge Computing
MIMO	Multiple-input and Multiple-out	MR	Mixed Reality
MSE	Mean Squared Error	MTP	Motion-to-photon
MVC	Multi-view Coding	NOMA	Non-orthogonal Multiple Access
O-RAN	Open Radio Access Network	PoV	Persistence of Vision
QoE	Quality of Experience	QoS	Quality of Service
THz	Terahertz	TSN	Time-sensitive Networking
VR	Virtual Reality	XR	Extended Reality

level of terabits per second (Tbps), that even 5G cannot support it, especially for high-resolution and 360° videos. The required end-to-end delay for delivering haptic information can be as low as a few milliseconds for a satisfactory user experience (Maier and Ebrahimzadeh, 2019; Sim et al., 2021). The synchronization of data streams from multiple cameras or sensors and that of audiovisual and haptic information in data transmission also create new challenges. The storing and processing of massive data for immersive communications demand new architectures and techniques for caching and computing (Glushakov et al., 2020; Taleb et al., 2021; Liu et al., 2021). Moreover, artificial intelligence (AI) is necessary both for supporting applications such as human-machine collaboration and user viewpoint/gesture prediction, and for orchestrating network resources to satisfy the demanding requirements of immersive communications (Tataria et al., 2021; Zawish et al., 2022; Maier et al., 2018).

Recognizing the importance of immersive communications, its recent developments, and the technical challenges, we present a comprehensive review of immersive communications in this article. With a focus on the communication, networking, and computing perspectives, we illustrate the challenges and potential solutions to immersive communications in the era of 6G communications. In specific, we focus on immersive communications by looking into its three main forms, i.e., XR, haptic communication, and holographic communication in the remainder of this article. Section 2 introduces representative use cases of immersive communications to illustrate its promising prospect. Section 3 presents the concepts, basic implementation procedures, and requirements of XR, haptic communication, and holographic communication to paint an overall picture of immersive communications. Section 4 focuses on the challenges and the state-of-the-art solutions towards realizing each of the three forms of immersive communications. Section 5 discusses some open issues regarding immersive communications in 6G, and Section 6 concludes this article.

2 USE CASES

There are many potential use cases for immersive communications, relating to both commercial and enterprise scenarios and ranging from gaming to industrial control. In this section, we detail four representative use cases to illustrate the promising prospect of immersive communications. A list of representative use cases is given in Table 2.

2.1 Immersive Gaming and Entertainment

XR provides the ultimate gaming and entertainment experience by presenting convincing gaming environments through XR devices such as VR headsets or smartphones. Players can interact with each other without feeling a barrier between the virtual and the physical worlds (Bastug et al., 2017). XR devices display the virtual world of the game to players and capture their actions such as eye movements to allow them to interact with the virtual world (Elbamby et al., 2018b). With the success of advanced XR gaming consoles and headsets, e.g., Oculus and PlayStation VR, as well as games and platforms, e.g., Pokemon Go and Roblox, game developers are striving to offer more flexible XR experiences with wireless XR devices (Maimone and Wang, 2020). Through wireless XR devices, players can interact freely with other players or virtual objects, e.g., in XR sporting (Kim et al., 2018). Furthermore, haptic communication devices can be combined with XR to significantly enhance the immersive gaming experience. Transducer arrays, which can be attached to XR devices, can capture haptic data from players. As a result, XR devices can fuse haptic information into the virtual world and provide haptic feedback to players by mapping motions in the game to players' sensations. Players can use haptic devices, such as gloves, to control objects in the game (Hashimoto and Ishibashi, 2006) or synchronize their sensations with other players (Mauve, 2000).

2.2 Telesurgery

In telesurgery, surgeons remotely manipulate robotic arms to operate on patients by utilizing control panels and real-time display of the surgical scenes. Telesurgery is beneficial in removing the barrier of distance among surgeons and patients, tackling the scarcity of surgeons in remote or difficult-to-reach areas such as countryside, battlefields and spacecraft, and facilitating the collaboration of surgeons at different locations (Choi et al., 2018; Mohan et al., 2021). The assistance of robotic arms can enhance the performance of surgeries by detecting and canceling out the physiological tremors of surgeons' hand motions (Kumar et al., 2020), performing delicate surgical operations and minimizing the surgical incision areas for reducing blood loss and incision-related complications (Diana and Marescaux, 2015). To guarantee the performance of surgeons, the display of surgical scenes to them should be highly precise and informative. To this end, 3D video of the surgical scenes with depth information, can be displayed to the surgeons, e.g., by using passive polarized glasses, and an eye-tracking mechanism can be used to quickly center the area where the surgeon is viewing in the visual display (Stark et al., 2015). In addition, augmented reality (AR) can be leveraged to overlay medical images such as ultrasound images and computed tomography (CT) images onto the video of surgical scenes (Liu et al., 2016b). Besides visual information, haptic information in the surgeries, such as the texture of tissues and the tension in tying surgical sutures, can be captured by the haptic devices on the robotic arms and then transmitted to and reproduced by the haptic devices at the surgeons' side (Patel et al., 2022; El Rassi and El Rassi, 2020).

2.3 Immersive Learning

Immersive learning integrates emerging technologies, including XR and haptic technologies, into teaching to provide students or trainees an interactive and engaging learning experience (Affan et al., 2021; Laamarti et al., 2014). During the recent COVID-19 pandemic, traditional methods of teaching, e.g., online courses, encountered the problem of engaging students in the learning process (Jumreornvong et al., 2020; Fitzek et al., 2021). To this end, immersive learning, as a potential solution to boost student engagement, is receiving increasing attention, especially from primary and secondary schools. With immersive learning, avatars of students and teachers can be created in the virtual world (Gupta et al., 2019), and each student is allowed to interact with the avatars of teachers and other students via the senses of sight, hearing, and touch. Such interactions can keep students' attention in learning process. Immersive learning is categorized as either asynchronous or synchronous. Training some skills, such as sports skills and cooperative tele-operation skills for industrial robots, requires real-time interactions, which can encourage active participation in the learning process (Kaluschke et al., 2021; Lee et al., 2021). Utilizing XR, haptic communication, and holography communication technologies, teachers can check whether the moves and actions of their students are correct, and provide immediate corrections if not, regardless of their physical distance from each other. For the skills that do not need real-time interactions, information regarding teachers' positions, velocities, and applied forces can be recorded and displayed to students via XR and haptic devices asynchronously (Tan et al., 2020). Such "record-and-replay" strategy can allow a much larger number of students to learn at their own pace, despite the absence of real-time interactions (Steinbach et al., 2018; Yokokohji et al., 1996a,b).

2.4 Holographic Teleconference

Teleconference is a convenient choice for users to remotely collaborate with each other. In the current video teleconferencing, remote participants can only be displayed on flat screens, which results in a very different perception in a virtual conference from that in an on-site conference. In order to provide an immersive experience in teleconferences, holographic teleconferences depict realistic 3D presence for people by projecting 3D images of remote participants as holograms (Jiang et al., 2021; Zhang et al., 2022a; Zhou et al., 2022b). Specifically, when a remote participant joins the holographic teleconference, 3D visual information and the corresponding audio information of the participant can be captured by multiple sensors, transmitted, and then reconstructed as a hologram on the side of other participants to provide 3D audiovisual information for real-time interactions among participants (Strinati et al., 2019). In this case, holographic teleconference can reduce the impact on participants of the separation between the virtual and the physical worlds. In addition to the audio and video information, participants in a holographic teleconference are able to obtain haptic information from others to achieve an immersive experience with the sense of physical contacts (Tataria et al., 2021). For example, a participant with haptic sensors can sense a handshake with others, thereby enabling an immersive experience similar to in-person interactions.

3 IMMERSIVE COMMUNICATIONS: CONCEPTS AND REQUIREMENTS

The use cases for immersive communications and their potential importance in 6G are intuitive. Understanding immersive communications beyond the use cases, however, requires answers to the question "what are immersive communications?". Since the research of immersive communications is in an early stage, there is no commonly-agreed definition yet.

We consider immersive communications as a communication paradigm along with the supporting technologies that allow users to have lifelike experiences in the physical world, the virtual world, or both, with interactions via 3D audiovisual and/or haptic information exchange. In this section, we focus on the three main forms of immersive communications as illustrated in Fig. 1, i.e., XR, haptic communication, and holographic communication.[1] Via introducing the concept, basic implementation procedure, and the network requirements for each of the three forms, we aim to sketch an overall picture of immersive communications.

---

[1] Note that the three forms may co-exist since a use case may involve more than one form, and additional forms of immersive communications may exist or emerge.

Table 2. Representative use cases of immersive communications

Use Cases	References
Gaming	(Carroll and Yildirim, 2021; Hu et al., 2019)
E-learning	(Kavanagh et al., 2017; Harvey et al., 2021; Ahmad et al., 2021) (Freina and Ott, 2015)
Teleconference	(Kantonen et al., 2010; Zhang et al., 2022a)
Tele-operation	(Lee et al., 2021; Choi et al., 2018; McCloy and Stone, 2001)
E-heath	(Jumreornvong et al., 2020; Riva, 2000)
E-commerce	(Ornati, 2022; Speicher et al., 2017)
Smart home	(Zhu et al., 2020; Lertlakkhanakul et al., 2008)
Manufacturing	(Aijaz and Sooriyabandara, 2018)
Tourism and travel	(Guttentag, 2010)
Metaverse	(Wang et al., 2022b)

3.1 Extended Reality

In this subsection, we introduce the concept of XR and investigate two respective XR technologies: VR and AR. Then, we examine their implementation procedure and service requirements for 6G.

3.1.1 Concept

XR covers a range of technologies, including VR, AR, mixed reality (MR), and everything in between (Hu et al., 2020). In general, XR combines the physical and virtual worlds through extensive video processing and data fusion. Using XR devices, users can interact with virtual avatars and access XR content. Under the umbrella of XR, a variety of technologies are defined depending on the level of virtuality. Two representative technologies in XR are AR and VR. With the lowest level of virtuality, AR focuses on constructing artificial objects according to the objects (e.g., buildings, faces, or vehicles) residing in the physical world and enabling users to interact with them. Conversely, with the highest level of virtuality, VR creates an entirely artificial scenery and allows users to interact with the objects in a completely artificial environment generated by the headsets. In MR, the concepts of VR and AR can be combined to create different levels of virtuality. In spite of the variety of XR technologies, the methods to provide immersive experiences to users are similar, which combine sensory data with virtual environments to produce artificial sceneries, from either the physical or virtual worlds, using headsets or portable display devices.

The first VR flight simulator was developed in 1970s to train pilots for flights without exposing them to risks of flying (Earnshaw, 1993). In the early stage, VR headsets were cumbersome, and processing VR content required large supercomputers. Nowadays, VR technologies have gained momentum due to recent advances in computing and display technologies. The headsets, such as Oculus head-mounted displays and HTC Vive, are affordable and can support ultra-high resolutions (3840×2160 in Pimax 8K) and refresh rates (up to 120 Hz) (Hu et al., 2020). At the present time, most VR content is processed and rendered by user devices. Rendering content with a high level of virtuality requires extensive computing power. For a VR headset, a console is required to supply additional computing power to the headset, while a wired connection restricts the user to a workstation. Therefore, wireless VR is the primary focus of VR research

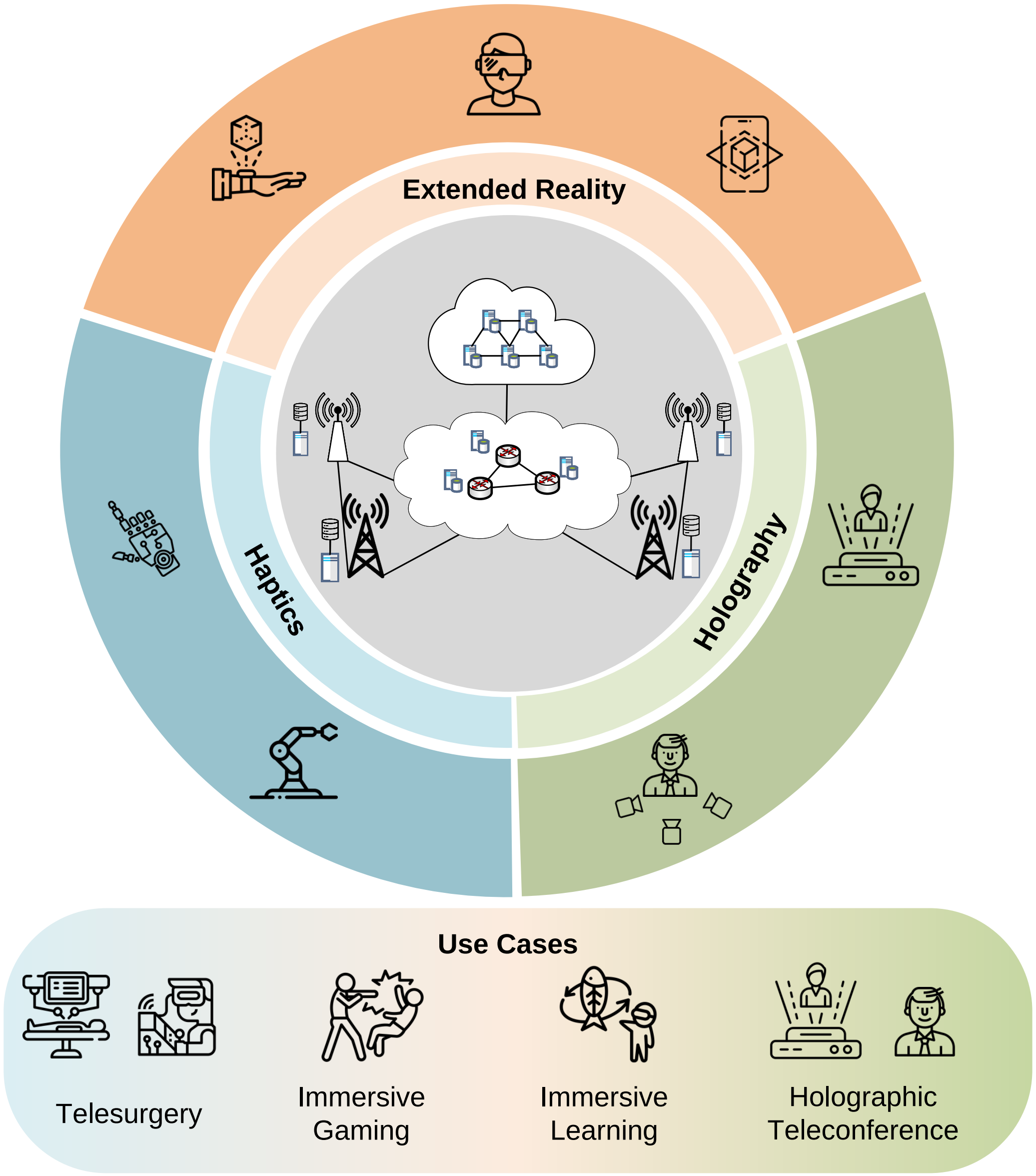

Figure 1. Main forms and exemplary use cases of immersive communications.

now (Elbamby et al., 2018a). In addition, multi-sensory XR, as another future vision of XR, integrates human senses and perception, including visual, auditory, olfactory, and tactile into XR content, enabling a truly immersive experience. This requires the confluence of multiple disciplines, including artificial intelligence, computer vision, biology, ultra-low-latency networking, etc., while linking the real and virtual worlds (Hu et al., 2021; Wang and Li, 2022).

3.1.2 Basic Implementation Procedure

While XR comprises several technologies with different levels of virtuality, its implementation procedure can be summarized into three steps: content transmission, rendering, and feedback collection. For each of the above three steps, communication networks can play an important role.

In the step of content transmission, VR content generated by VR content providers is transmitted from content servers and VR devices. VR devices play 360° spherical videos, which can be mapped to

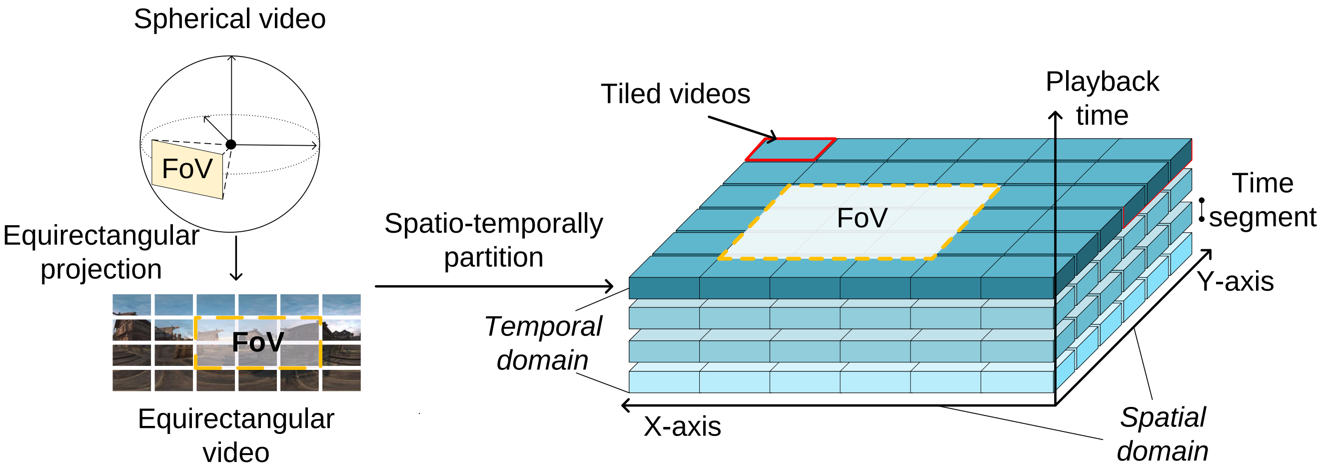

Figure 2. VR video projection and partition.

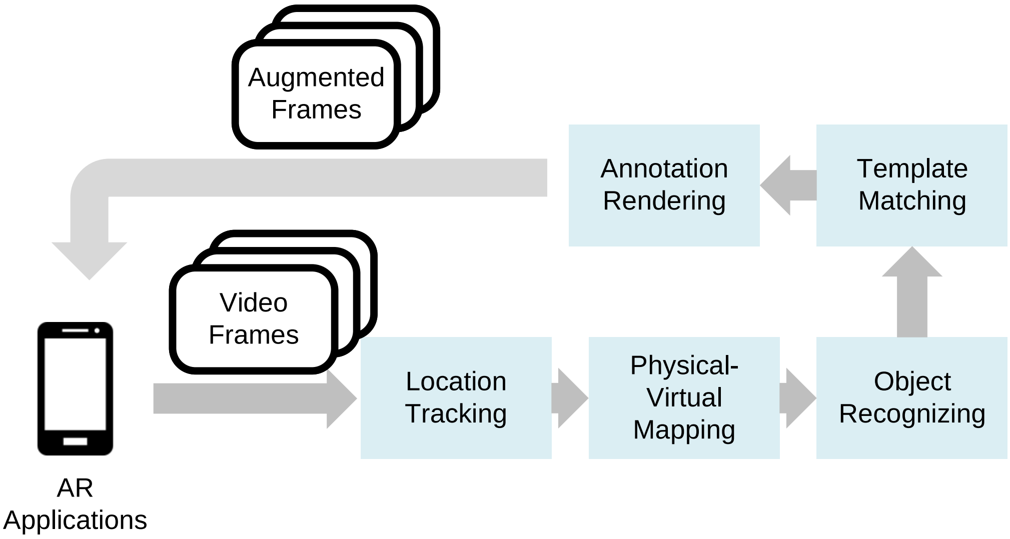

Figure 3. Content processing and rendering for AR applications.

equirectangular videos. During VR content playback, these equirectangular videos are mapped onto a sphere, in which the user is situated at the center, to provide a 3D stereoscopic experience. The key feature of VR video is the ultra-high spatial resolution. A VR video has a resolution of up to 12K (11,520 × 6,480), while the conventional video normally has a resolution of 4K or less. Transmitting full equirectangular videos from content servers requires an ultra-high data rate. Thus, tile-based transmission is usually adopted in VR video delivery. As shown in Fig. 2, a content server can divide equirectangular videos spatio-temporally into video chunks, i.e., tiled videos, and only the tiled videos within a user's field-of-view (FoV) is delivered (Yadav and Ooi, 2020; Son et al., 2018). In this way, VR content can be delivered in a significantly reduced data size. However, the tile-based solution requires VR headsets to detect and estimate user viewpoints to determine the region of FoV. Content servers should select which tiled videos to be delivered to users based on both the user's real-time viewpoint and network conditions (Zare et al., 2016). In terms of AR, AR devices generate raw content by the sensors at the local devices, such as cameras in smartphones (Ren et al., 2020). In contrast to VR devices, which download content from a content server, AR devices can upload raw content to the server for further processing. Specifically, raw videos captured by AR devices are clipped into frames with a specific image format, and those frames can be offloaded to the server. The processed content is then delivered to and played on the AR devices.

In the step of content rendering, tiled VR videos transmitted to VR devices are stitched together, and computing resources are required to project 2D stereoscopic videos to 3D stereoscopic videos, i.e., generating two different videos for the left and right eyes respectively. This content rendering step can be performed on VR devices once all the required content has been received. In addition, due to the limited computing capability of VR devices, the workload of content rendering can be offloaded to adjacent edge servers enabled by mobile edge computing (MEC) (Dai et al., 2020; Dang and Peng, 2019; Sukhmani et al., 2018). Content processing and rendering are more complex in AR than in VR, where AR processing procedure is shown in Fig. 3. Once the raw AR content, i.e., video frames, is captured by an AR device, a location tracking step determines the device's location and position according to the captured frames. Then, a mapping step establishes a virtual coordinate of the environment based on the result of the tracker, and an object recognizing step detects the objects to process in the video frames (Qiao et al., 2018; Ren et al., 2019). Based on the identified objects, the augmented data is retrieved from the local cache or network servers and attached to the frames accordingly. Specifically, a template matching step attaches the augmented data to the frames, and an annotation rendering step renders the processed frames at AR devices. The computing workload for conducting the above functions can be fully or partially offloaded from AR devices to network servers to minimize computing latency or improve energy efficiency at AR devices.

After receiving and playing XR content, XR devices collect user feedback to select the content to deliver next. VR and AR devices have similar methods for feedback collection, with sensors or cameras attached to the devices to capture users' actions and motions. Moreover, VR requires additional feedback regarding the user's viewpoint. A user's viewpoint determines which tiled videos to deliver to render the FoV of the user. The real-time viewpoint can be captured by motion tracking modules on a VR device. Additionally, motion emulation can be used to simulate a user's viewpoint movement on VR devices. VR devices can request the content proactively based on the emulation results to avoid playback performance degradation, such as rebuffering (R. Yao and Hoberman, 2017). In addition, for interactive applications such as XR gaming, the sensors connected to XR devices, such as inertial measurement units (IMUs), haptic gloves, etc., gather inputs from the users. Depending on the inputs, the XR devices can either process the inputs locally or upload the inputs to content servers for computing and updating.

3.1.3 Requirements

In general, XR has stringent latency requirements for accurate and smooth content playback based on user motions. In terms of VR, motion-to-photon (MTP) delay is the most important delay metric, which measures the time difference between the user's viewpoint movement and corresponding reflections at the output of the VR headset. If MTP delay is larger than 20 ms, VR users may feel spatially disoriented and dizzy, referred to as VR sickness (R. Yao and Hoberman, 2017). Current VR industries target lower MTP delay (below 15 ms) for ideal user experience (Mangiante et al., 2017). In addition, for VR applications requiring extensive interactions, the requirement of response time for rendering the interactions into VR content can be longer than the MTP delay requirement. For example, in VR gaming, a latency of up to 50 ms for responding to player actions can be noticeable yet currently acceptable (Zhang et al., 2017). In terms of AR, the content is mainly captured by local devices. The MTP delay in AR can be minimized by playing the raw content captured by AR devices before the content is processed. However, users' immersive experiences can be adversely affected by delayed processing for rendering the user's motions into AR content. The delay requirements for reproducing user interactions in AR content are 75 ms for online gaming and 250 ms for telemetry based on the sensitivity of the human vestibular system (Mohan et al., 2020).

Furthermore, in order to achieve low content delivery latency, an ultra-high data transmission rate is required for delivering XR content. Specifically, users view VR videos on headsets placed a few centimeters from their faces. Therefore, high-resolution videos are required for VR applications to improve user experience. Although tile-based content transmission can reduce the data size in VR content delivery, data rate requirements can still be 2.35 gigabits per second (Gbps) or above for VR video delivery, which is more than 100 times higher than the data rate for current high-definition video streaming (Mangiante et al., 2017). For interactive XR applications, such as VR gaming and AR, extensive video processing is required. The computing capability of both network servers and user devices dominates the performance of interactive XR applications, and limited computing capability in the network can be another bottleneck for XR content delivery (Elbamby et al., 2018b).

3.2 Haptic Communication

In this subsection, we first provide the concepts of haptics and haptic communication. Then, we detail the implementation procedure and service requirements of haptic communication in the 6G era.

3.2.1 Concept

The term haptics initially referred to interactions between humans and objects in the physical world that involve the sense of touch, e.g., swiping a phone screen (Steinbach et al., 2012). The development of tele-operation technologies over the past few decades have expanded the definition of haptics to all forms of interactions involving the sense of touch, including interactions between humans and virtual objects in the virtual world or the tele-operated machines in the physical world (Tan et al., 2020; O'malley and Gupta, 2008). The information conveying the sense of touch in such interactions is referred to as haptic information. The sense of touch relates to different types of mechanoreceptors in human skin and muscles, and the haptic information can be broadly classified into tactile and kinesthetic information (Abiri et al., 2019). Specifically, tactile information is related to the sense of surface texture, friction, and temperature felt by the human skin when in contact with objects, and kinesthetic information is related to the sense of position and motion of limbs along with the associated forces (Srinivasan and Basdogan, 1997; Steinbach et al., 2012). A device that supports such haptic interactions and the transmission of haptic information is referred to as haptic interface (HI) or haptic device (Culbertson et al., 2018). An HI is comprised of haptic sensors and haptic actuators responsible for capturing and displaying haptic information, respectively (Antonakoglou et al., 2018). An HI can capture and display a variety of haptic information, and the number of independent coordinates used by the HI to specify the haptic information is referred to as the degrees of freedom (DoF) of the HI (Promwongsa et al., 2020).

Haptic communication refers to the process in which humans communicate and interact through the sense of touch over a communication network (Steinbach et al., 2012). The communication network supporting haptic communication is named as *Tactile Internet* in some existing works (Ali-Yahiya and Monnet, 2022).[2] With the use of HIs and the transmission of haptic information over communication networks, users can interact with virtual objects in the virtual world or remotely operate machines in the physical world (Steinbach et al., 2012). The transmission of haptic information can be unilateral, bilateral, or multilateral, depending on the number of users participating in the transmission. In the cases of one

---

[2] Haptic communication and the Tactile Internet are related as a service and a medium as in the case of voice over IP (VoIP) services and the Internet (Aijaz et al., 2016).

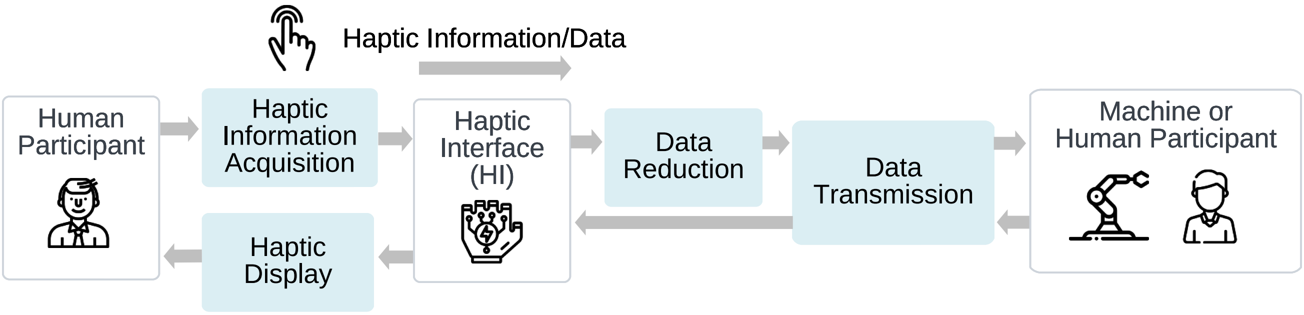

Figure 4. Implementation procedure of bilateral haptic communication.

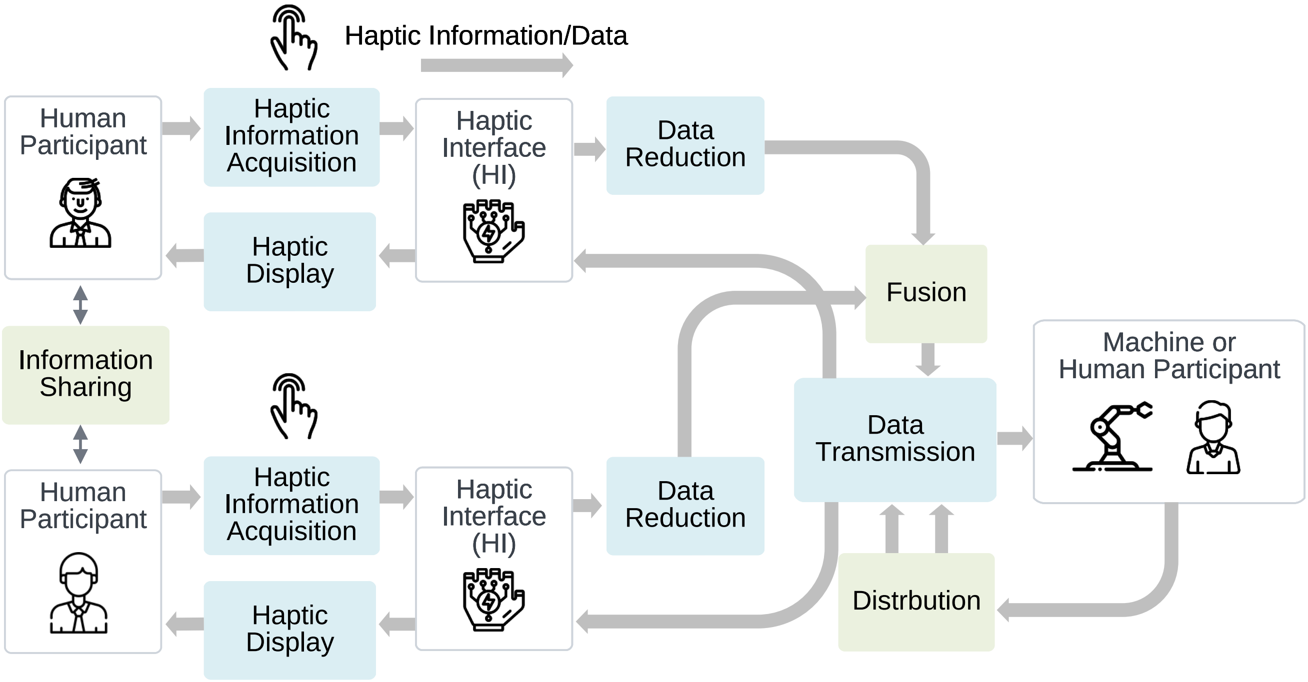

Figure 5. Implementation procedure of multilateral haptic communication.

user manipulating a remote machine or two users interacting with each other, the haptic communication is unilateral (i.e., an HI either sends or receives haptic information) or bilateral (i.e., an HI both sends and receives haptic information). In other cases, haptic information can be transmitted multilaterally, e.g., in cooperative tele-operations involving multiple users. This is because the behavior of each user may have an effect on other users, resulting in interconnections and couplings in the exchanges of haptic information (Shahbazi et al., 2018; Feth et al., 2009). Since haptic communication centers on humans, some studies examine the *human-in-the-loop* nature of haptic communication and predict a paradigm shift from content delivery to skillset delivery, as a result of the emergence of haptic communication (Simsek et al., 2016; Ali-Yahiya and Monnet, 2022).

3.2.2 Basic Implementation Procedure

The implementation procedure of haptic communication depends on how the haptic information is transmitted. For bilateral haptic communication, the implementation procedure mainly consists of four steps: haptic information acquisition, data reduction, data transmission, and haptic display, as shown in

Fig. 4.[3] In the first step, haptic information, including tactile and kinesthetic information, can be acquired by haptic sensors in HIs. In terms of tactile information, force sensors, thermistors, and laser scanners are mainly used in the measurement or evaluation of friction and hardness, warmth, and macroscopic roughness, respectively (Liu et al., 2017; Lederman and Klatzky, 2009; Okamoto et al., 2012; Fishel and Loeb, 2012). Haptic sensors such as IMUs are responsible for the acquisition of kinesthetic information, e.g., tracking the position, velocity, and angular velocity of sensors positioned at different parts of a human (Steinbach et al., 2018). The haptic sensors of interest can be dynamically selected, and only the haptic information captured by the selected haptic sensors needs to be collected for efficient haptic information acquisition (Van Den Berg et al., 2017). Due to the potentially high DoF of an HI, data reduction is adopted in the second step to reduce the amount of haptic data without degrading the users' immersive experience too much. Specifically, waveform-based representation and feature extraction algorithms can be used in the compression of tactile information, and perceptual coding techniques based on perceptual masking phenomenon can be applied for compressing kinesthetic information (Jayasankar et al., 2021; Steinbach et al., 2010). In addition, predictive methods (also called predictive coding techniques) can be leveraged to reduce the amount of transmitted haptic data by inferring upcoming haptic information (Steinbach et al., 2018). Haptic data reduction can be carried out at either HIs or network servers (Fitzek et al., 2021; Steinbach et al., 2012). Existing methods of haptic data reduction are detailed in Section 4.2. In the third step, the haptic data is transmitted over a communication network, resulting in a haptic data stream between the two HIs. The haptic data stream can consist of multiple haptic data substreams, each of which corresponds to a type of haptic information. Data traffic patterns and QoS requirements can vary across different haptic data substreams due to the differences in the sensitivity of human perception, such as reaction time and the range of perception (Fitzek et al., 2021). The respective QoS requirements of haptic data substreams should be satisfied, and the haptic data substreams should be synchronized in transmission. Moreover, a haptic data stream should be synchronized with audiovisual data streams in the case of immersive communications involving multiple modalities (Cizmeci et al., 2017). In the last step, i.e., haptic display, haptic actuators in the HI stimulate human mechanoreceptors to create realistic haptic sensations when an HI receives haptic data (Wang et al., 2019). In general, haptic display includes tactile display, e.g., adjusting the temperature, and kinesthetic display, e.g., creating motion and changing muscle tension (Ozioko et al., 2020; Steinbach et al., 2018; Pacchierotti et al., 2017). In the case when haptic data transmission is unreliable or delayed, predictive methods can be leveraged at the receiver side to estimate the haptic data not received timely for smooth haptic display.

In the case of multilateral haptic communication, three additional steps take place besides the aforementioned four steps, especially for cooperative tele-operation applications (Feth et al., 2009). The implementation procedure of multilateral haptic communication is shown in Fig. 5, and the three additional steps are highlighted with green rectangles. First, even if there is no direct haptic interaction between two users, they can still share haptic information (Takagi et al., 2017). The data format and content of the transmitted haptic information in such haptic information sharing may differ from those of the transmitted haptic information in direct haptic interactions (Shahbazi et al., 2018). Second, it is necessary to properly fuse the haptic information from multiple users, e.g., the weighted sum, when their behaviors affect other users (Thanh et al., 2012; Fujimoto et al., 2008). Third, when one user's behavior affects multiple users at the same time, distributing haptic information to multiple users according to their different behaviors is required to achieve precise haptic display for individual users, e.g., different reaction forces are applied to tele-operators (Chen et al., 2016).

---

[3] In unilateral haptic communication, either the step of haptic information acquisition or the step of haptic display is skipped depending on whether an HI is sending or receiving haptic information.

3.2.3 Requirements

The data transmission rate requirement of haptic communication is determined by the packet rate and size of haptic data. The packet rate is the number of packets transmitted by an HI per second, which depends on the information update rate. For the smoothness and fidelity of haptic perception, haptic information typically needs to be updated at a rate above 1,000 times per second (Choi and Tan, 2004). If each update of haptic information is packetized and transmitted, the corresponding packet rate of haptic data is above 1,000 packets per second (Xu et al., 2015). The packet size of haptic data largely depends on the DoF of the haptic data (Holland et al., 2019). For kinesthetic data, controlling one movable component (e.g., a joint) on a tele-operator (e.g., a robotic arm) needs six coordinates to be specified to achieve 6 DoF, with three coordinates specifying the transitional motion in the 3D space and the other three specifying the rotational motion including roll, pitch and yaw, respectively (Promwongsa et al., 2020). Since a human hand consists of multiple movable components (e.g., finger joints and wrist joints), its kinesthetic data can be described by a 24-DoF model (Cobos et al., 2008). In addition, for reproducing tactile information with high fidelity, a dense array of haptic sensors/actuators needs to be deployed on a user (Hoggan et al., 2007). For example, for reproducing vibrotactile data, four actuators are deployed around one fingertip (Baik et al., 2020). As a result, tactile data can involve even higher DoF than kinesthetic data (Holland et al., 2019).

The delay tolerance of haptic communication can be as low as 1 ms since the packet rate of haptic data can be above 1,000 packets per second (Fettweis et al., 2014). In practice, the delay requirement of haptic communication is determined by factors including the perceptual sensitivity of receivers, the dynamics of haptic interaction, and specific operation or interaction. First, higher perceptual sensitivity for haptic information generally indicates the need for a higher packet rate and thus a stricter delay requirement (Chaudhuri and Bhardwaj, 2018). For example, while touring a virtual museum of natural history, archaeologists can have a stricter delay requirement than the majority of visitors due to their higher perception sensitivities of artifacts and specimens. Second, similarly, higher dynamics of haptic interaction generally call for a higher packet rate and a lower delay. Specifically, the delay requirement when such dynamics is high (e.g., in tele-soccer), medium (e.g., in telerehabilitation) and low (e.g., in tele-maintenance) is 1-10 ms, 10-100 ms and 100-1,000 ms, respectively (Holland et al., 2019). Third, a delay below 2 ms is required for remote machine manipulation, while a delay below 50 ms is required for remote machine monitoring and maintenance (Aijaz and Sooriyabandara, 2018).

The reliability of haptic communication can be evaluated in terms of bit error rate, packet loss rate, delay-bound violation probability, or prediction error when haptic data prediction is adopted (Promwongsa et al., 2020). The requirement for the reliability depends on factors such as the specific communication scenario and whether or not haptic reduction is used. First, in terms of delay-bound violation probability, the reliability of haptic communication in immersive gaming is required to be above 99.9% (Holland et al., 2019). In contrast, when critical operation tasks are performed based on haptic information, higher reliability of haptic communication is required. For example, the reliability of above 99.999% is required for haptic communication in telesurgery and remote machine manipulation, (Gupta et al., 2019; Aijaz and Sooriyabandara, 2018). Second, when haptic data reduction is adopted, the same packet loss or bit error rate can cause more degradation in the haptic information (Steinbach et al., 2010). As a result, the use of haptic data reduction can result in a stricter requirement for the reliability of haptic communication. For example, the reliability above 99.999% is required in immersive gaming when haptic data reduction is adopted (Holland et al., 2019).

3.3 Holography and Holographic Communication

In this subsection, we introduce holography and holographic communication, beginning from presenting the concept and different types of holography, followed by the basic implementation procedure of holographic communication, and ending with the data transmission rate and delay requirements.

3.3.1 Concepts

As the name suggests, holographic communication depends on holography technology, which has made significant progress in the past decade. There are different stages in the development of holography technology. *Optical holography* generates holograms via recording and recreating optical wavefront, and the corresponding holograms are recorded interference patterns (e.g., on photographic emulsions) of an "object wave" and a "reference wave". When the recorded interference pattern is illuminated by the reference wave, a 3D light field can be recreated using diffraction. The original idea of hologram was developed in 1940s, and real breakthrough was made in 1960s thanks to the development of laser (Gabor, 1972). Later, with advances in electronic devices, *digital holography* emerged, which uses image sensors to capture interference patterns. In digital holography, recording is done optically, while a 3D image is reproduced via numerical calculation of light wave diffraction using methods such as Fourier transform (Tahara et al., 2018). The latest development of holography is *computer-generated holography*, in which both the interference pattern and the 3D image in display are generated digitally using a computer (Sahin et al., 2021). With computer-generated holography, the object to be displayed does not have to be physically present, which yields great flexibility at the cost of high computational complexity (Shimobaba et al., 2022). Despite of the advance in recent years, generating dynamic 3D holograms in real time is challenging. As a result, alternative approaches to displaying 3D images emerge, which are sometimes referred to as "false holography". Such approaches use glass panes or other "tricks" to create illusions of 3D images (Jones et al., 2007; Kerrigan, 2018). Among the false holography techniques, *volumetric display* has attracted significant interest in the field of computer-aided design and medical imaging (Favalora, 2005). Volumetric display, an umbrella term for many different techniques, renders volume-filling 3D images via the generation, absorption, and scattering of illumination in a confined space, e.g., a cube or cone (Yang et al., 2016). The study of volumetric display is active with exciting experiments (Smalley et al., 2018), and commercial products are also available (Gibney, 2019). Other approaches to imitate 3D display include the use of multiple projectors and a human-size retroreflective cylinder (Gotsch et al., 2018).

Based on either true holography or "false holography", holographic communication is about transferring data representing dynamic 3D images of a physical object over a network and displaying the object in 3D at the receiver.[4] Integrating 3D data capturing, processing, transmission, and rendering, holographic communication is expected to enable exciting new services in 6G (Strinati et al., 2019; Clemm et al., 2020). At the moment, there is no consensus on the scope of holographic communication in the literature, and some researchers consider the transferring and rendering of 3D data in AR/VR as a type of holographic communication (Essaili et al., 2022). In this review, holographic communication refers to data transfer for autostereoscopic 3D display, i.e., 3D images that can be viewed by naked eye without the aid of eyewear or headsets and, ideally, are different when viewed from different positions, angles, or tilts. The 3D display at the receiver can be rendered via real holography, false holography such as volumetric display, or other techniques as long as the objective of autostereoscopic 3D display is achieved. Similar to existing

---

[4] Note that the term "holographic communication" is also used in the literature of massive MIMO and IRS but with a different and unrelated meaning (Dardari and Decarli, 2021).

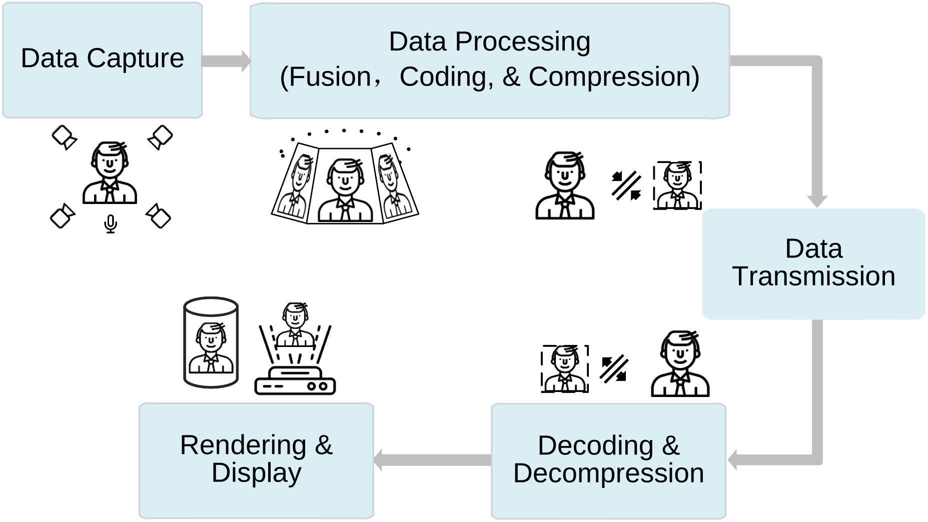

Figure 6. Implementation procedure of holographic communication.

multimedia communications, the content of holographic communication can be either generated in real time or recorded, and the communication mode can be unicast, multicast, or broadcast.

3.3.2 Basic Implementation Procedure

Although various approaches for holographic communication differ in the implementation procedure, the general process includes the steps of data capture, processing, transmission, and rendering.

Except for computer-generated holography, a capture system is required to record 3D images of a physical object. An ideal capture system for holographic communication would capture the light field, i.e., all the information of each light ray, in the target scene (Jiang et al., 2021). In practice, capture is conducted with visual sensors such as a camera array (Nakamura et al., 2019) or light detection and ranging (LIDAR) sensors (Fratz et al., 2021). The depth information of the object of interest is either directly captured (e.g., in the case of a capture system with LIDAR sensors) or computed in the subsequent data processing step (e.g., in the case of a capture system with a camera array). The performance of the visual capture system depends on factors such as the number of sensors and the camera sampling rate (Jiang et al., 2021).

In the data processing step, the depth information of target objects in the scene is computed (if not directly captured), and the output from capture sensors is fused to form a composite 3D representation of the captured scene (Javidi et al., 2005). For example, in digital holography, a computer can process 2D images taken from different angles and tilts by a camera array to form a single 3D representation of the captured scene (Essaili et al., 2022). The fusion of images may help achieve visualization enhancement in the rendered 3D images such as improvement in the resolution and contrast (Javidi et al., 2005), and it can be conducted either solely at the transmitter side or with the help of an edge server. In addition, the data processing step is responsible for the compression of the fused data to speed up the transmission and reconstruction, and reduce the required data transmission rate and storage in holographic communication (Kurbatova et al., 2015; Cheremkhin and Kurbatova, 2019). The compressed data for the 3D representation is then encoded and transmitted over a network.

At the receiver side, the received data is decoded using one or multiple chosen codecs and decompressed. The captured scene is then reconstructed, possibly with the help of an edge server, and rendered on a display device. An ideal display device for holographic communication would regenerate the light field in the captured scene to create an illusion that the user is placed in the scene. In practice, creating such an illustration is difficult as it requires each point (e.g., each pixel) of the display device to emanate different light rays in different directions. However, given the limitations of human perception, the feeling of visual immersion can be created by using equipment such as a cylindrical light field display (Gotsch et al., 2018), a persistence of vision (PoV) display (Gately et al., 2011), or a static volumetric display device (Kumagai et al., 2021). Such devices render 3D images by using a large curved display to fill the user's FoV, exploiting the phenomena of a lingering afterimage on the retina, and dynamic turning on/off of voxels in a confined 3D space, among other methods for creating illusions of 3D images.

It is worth noting that holographic communication may also involve audio data capture, processing, and rendering. In such a case, capturing the sound field in the target scene and ensuring audio and video synchronization are important for users to enjoy an immersive holographic communication experience (Jiang et al., 2021).

3.3.3 Requirements

Holograms mainly come in two types, namely volumetric-based holograms and image-based holograms. The transmission of the two types of holograms requires different data rates, ranging from hundreds of Mbps up to Tbps (Clemm et al., 2020). For volumetric-based holograms, a physical object is represented as a set of 3D pixels or voxels, such as a point cloud. Transmitting a point cloud targeting an object requires a data rate on the level of hundreds of Mbps to several Gbps, depending on the resolution of the 3D content (FG-NET2030, 2020). For example, to fully represent a human, the point cloud in each frame typically consists of 10^5 to 10^6 points, while each point needs 15 bytes of data to represent the color and 3D coordinate of the point. In the case of 30 frames per second, the data rate requirement is between 300 Mbps and 3 Gbps (Essaili et al., 2022; Selinis et al., 2020). For image-based holograms, such as light-field video (LFV), an object is presented by an array of images captured at different angles, tilts, and/or positions. An LFV-based hologram can be more precise as compared with a volumetric-based hologram, especially in high resolution when a large number of images from different tilts, angles, and positions are used per frame (Jiang et al., 2021). For example, if the 3D representation of an object requires a separate image every $0.3°$, a hologram with an FoV angle range of $30°$ and a tilt range of $10°$ needs 3,300 separate 2D images. In order to transmit an LFV-based hologram for a human-sized object, the required data rate should be between 100 Gbps and 2 Tbps (Clemm et al., 2020).

To support real-time holographic communication, the overall delay, including data capturing, processing, transmission, and rendering delay, should be less than 100 ms (He et al., 2023). In addition to low delay, synchronization is important to holographic communication. Generally, the hologram of objects or humans may be sampled by multiple sensors from different angles and different distances. In this case, data from different sensors should be synchronized in transmission (Strinati et al., 2019). Taking holographic teleconference as an example, as multiple participants can join the teleconference from different locations, multi-source synchronization is necessary for them to have good quality of experience (QoE) in holographic communication. Otherwise, a part of the rendered hologram can be slightly ahead or behind relative to the rest of hologram for some users, resulting in poor QoE (Lesniak and Tucker, 2018). Moreover, holographic communication can involve multi-sensory information, e.g., the haptic, audio, and video information (Taleb et al., 2021). In this case, the synchronization of different sensory information in transmission is also important for a participant to see the hologram, hear the voice, as well as receive touch-sensory feedback from others without a degradation of the immersive experience due to out-of-sync issues. For holographic communication involving the transmission of audiovisual and haptic data, the tolerable difference in the delay of different types of data should be lower than 80 ms for satisfactory QoE (Montagud et al., 2018).

Table 3. Requirements of use cases in immersive communications

	Use Cases	**Requirements**
XR	360° video playback	$<$ 20 ms MTP delay (R. Yao and Hoberman, 2017), 2.35 Gbps data rate (Mangiante et al., 2017)
	Interactive applications (e.g., VR gaming)	$<$ 50 ms response time (Zhang et al., 2017)
	Collaborative virtual applications (e.g., teleconference)	$<$ 150 ms virtual feedback (Jay et al., 2007) 12.5 Tbps/km^2 upload capacity
Haptics	Telesurgery	$>$ 99.999% reliability, $<$ 1 ms delay (Gupta et al., 2019)
	Remote machine manipulation	$>$ 99.999% reliability, $<$ 2 ms delay (Aijaz and Sooriyabandara, 2018)
	Haptic interaction-based rehabilitation	$>$ 99.999% reliability, $<$ 50 ms delay (Holland et al., 2019)
Holography	Volumetric-based hologram (e.g., point cloud)	$>$ 300 Mbps data rate (Essaili et al., 2022; Selinis et al., 2020)
	Image-based hologram (e.g., LFV)	$>$ 100 Gbps data rate (Clemm et al., 2020)
	Real-time holographic video transmission	$<$ 100 ms delay (He et al., 2023)

4 IMMERSIVE COMMUNICATIONS: CHALLENGES AND SOLUTIONS

After introducing the concepts, implementation procedures, and requirements of immersive communications, we now discuss challenges in XR, haptic communication, and holographic communication, as well as the state-of-the-art solutions, with the most important ones summarized in Fig. 7. Note that our review here focuses on the challenges and solutions related to the communication, computing, and networking aspects of immersive communications.

4.1 Extended Reality

The main challenge of XR is delivering the required content to users on time, given the limited transmission resources and computing capability in a network. A variety of network functions and resources contribute to the performance of content delivery. Systematic solutions involving data processing, rendering, transmission, etc., have been developed to address these challenges. We summarize the solutions for implementing XR in three aspects: content selection, transmission improvement, and computing optimization.

4.1.1 Content Selection

The fundamental step in supporting XR applications is to identify which content needs to be processed and transmitted. This step focuses on minimizing the overall data size of the content to deliver at the cost of tolerable performance degradation, thus reducing the delivery time.

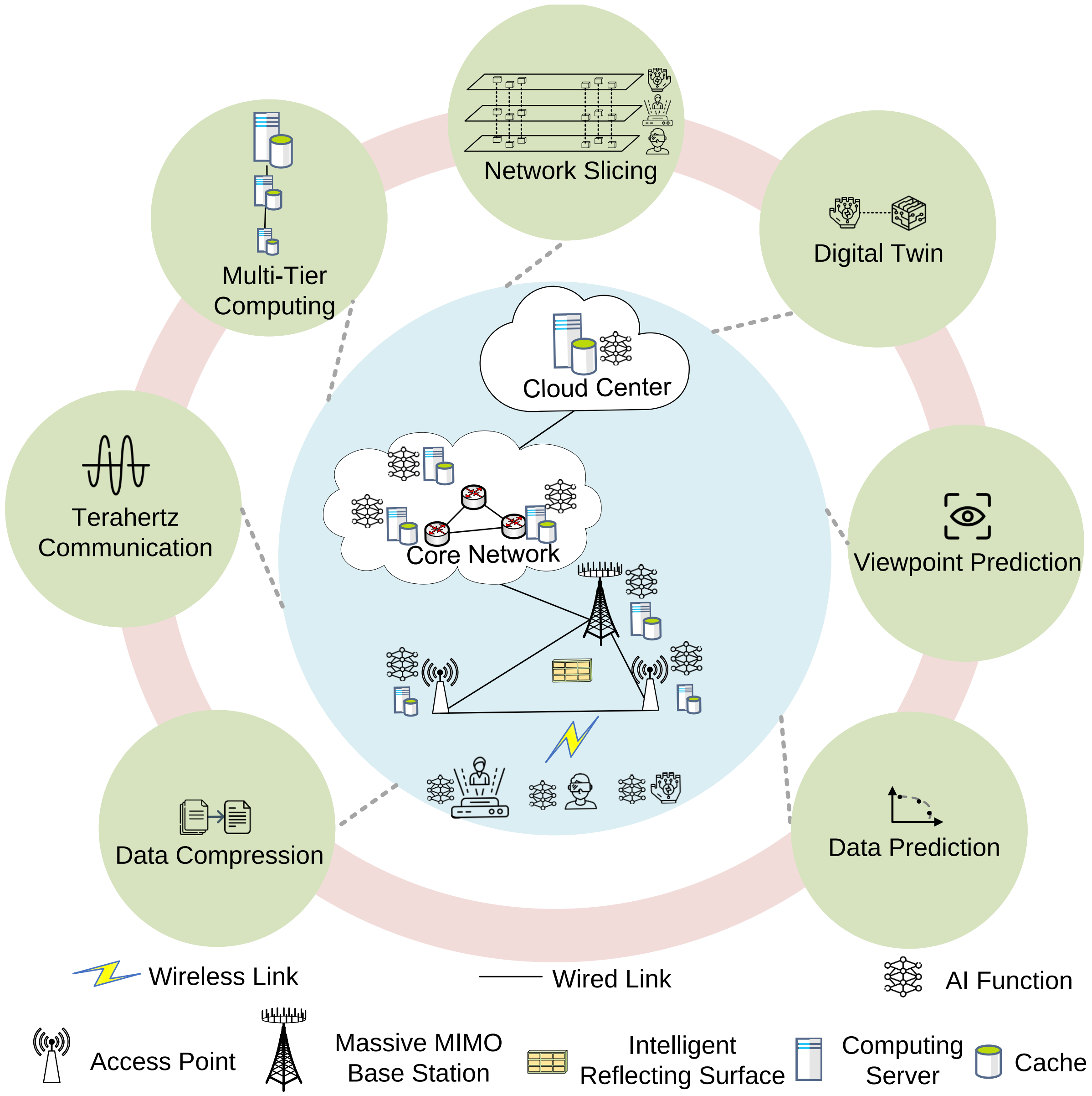

Figure 7. Potential solutions to immersive communications.

In VR services, proactive content delivery is commonly used to meet MTP delay requirements. Thus, in tile-based content transmission, the primary research challenge is how to predict user viewpoints accurately so as to determine which tiled videos to deliver to users. The prediction of user viewpoints can be achieved by sequential learning and data analysis methods based on the user's viewpoint trajectory, such as linear regression (He et al., 2018; Nasrabadi et al., 2020), and long short-term memory (LSTM) (Hou et al., 2018). A lightweight viewpoint prediction function can be deployed at the VR headset for local viewpoint prediction. Alternatively, the viewpoint trajectory can be updated to a network server (e.g., edge server), in which a more advanced machine learning model can be applied for accurate prediction (Hou et al., 2021). If the viewpoints are predicted by the network server, the prediction can be conducted based on not only current viewpoint trajectories for a group of users (Sun et al., 2020) but also the historical viewpoint trajectory data to further improve the prediction accuracy (Xu et al., 2018b; Feng et al., 2019). Although viewpoint prediction enables proactive tile-based content delivery, perfect prediction cannot be achieved due to the dynamics of user viewpoint movement. Even if viewpoints are known in advance, dynamic network environments such as data traffic load and processing time require adaptive resource management to ensure playback performance. With stochastic decision-making methods, such as reinforcement learning, it is possible to identify the dynamics of user viewpoint movement and determine which tiled videos to deliver to the corresponding VR device (Hu et al., 2022).

AR devices capture raw content, i.e., video frames, which can be offloaded to network servers for prompt content processing. Once the raw content is offloaded, the server detects and processes the objects within video frames captured by users' cameras, then returns the processed content to the AR devices. Though it is easier to satisfy the MTP delay requirement in AR than VR, accurate and rapid content processing (e.g., object detection) requires sufficient bandwidth to provide low-latency two-way transmission for satisfactory QoE. To balance transmission bandwidth usage for computing offloading and content processing performance, current solutions mainly focus on using machine learning techniques to adjust the number of frames offloaded by an AR device per unit time, based on the network environment and AR device movement. Specifically, offloading more video frames to a network server can improve object detection accuracy, especially when the AR device moves quickly and generates new content frequently. However, the bandwidth usage increases accordingly due to a large number of frames to offload (Liu et al., 2018a). Taking AR device mobility and network dynamics into account, adaptive frame rate adjustment is investigated in (Chen et al., 2021b). A deep reinforcement learning approach is used to study how mobility dynamics affect AR service performance and to determine the optimal uploading frame rate for maximal object detection accuracy and playback fluency.

XR content is expected to be further enriched in the era of 6G. Digital twins can incorporate AI to collect environmental information, characterize physical objects, and construct digital models of the physical objects accordingly. Digital models from digital twins can be used for XR applications as a new type of XR content that can be accessed by XR devices (Zhang et al., 2022b). For example, in an industrial Internet-of-Things scenario, designers and workers can use XR devices to interact with the digital models of machines and products in a simulated virtual environment. In addition, XR devices can collect the interactions from designers and workers. Based on the interactions, digital twins can adaptively configure their settings, such as data collection frequency (Aheleroff et al., 2021). The combination of XR and digital twins can support emerging applications such as metaverse. However, synchronizing among the physical world, digital twins, and XR content requires considerable network resources. Game theoretic methods are adopted in (Han et al., 2022) to adjust the synchronization rate between the physical world and digital twins based on the demand of virtual service providers that provide content to XR devices. A network slicing-based solution is proposed for providing metaverse services (Liu et al., 2022), which allocates multi-dimensional resources for content synchronization to improve the fidelity of digital twins and the QoE of XR users.

4.1.2 Transmission Improvement

As discussed in Subsection 3.1.3, the main bottleneck for VR video delivery is a limited data rate. Therefore, a straightforward solution to overcome the bottleneck is to increase the data rate with advanced communication techniques. As a key technology in 5G, millimeter wave (mmWave) communications can facilitate VR content delivery due to their high data rate and ultra-low propagation latency (Abari et al., 2016). In 6G, the transmission rate can be further improved by the physical layer technologies of terahertz (THz) transmission and intelligent reflecting surface (IRS), which can be applied in VR video delivery, especially between an edge server and VR devices (Chaccour et al., 2020; Du et al., 2020). However, communication links using ultra-high frequency bands, such as mmWave and THz, are prone to outage as they require line-of-sight (LoS) channels. Physical obstacles in the environment, including the user's body, may break the communication links and severely degrade the communication quality. To address this issue, a sub-6 GHz frequency band can be used as a backup if the mmWave or THz bands does not provide satisfactory channel quality. However, dynamic frequency band switching can result in a time-varying data transmission rate, thereby degrading the content delivery performance. The work (Liu et al., 2019b) models communication link state transitions corresponding to switching different frequency bands (e.g., mmWave and sub-6 GHz bands) in VR content delivery as a Markov chain. Content processing policies are adjusted to compensate for transmission delays when channel state transitions occur. In addition to adapting to channel dynamics, the reliability of mmWave or THz communication links can be improved by establishing multiple communication links between a device and several edge servers for VR content delivery (Yang et al., 2022a; Gu et al., 2022).

In addition, at the network layer, a network virtualization-based solution is proposed for VR content delivery, in which network controllers can create private logic networks for VR applications to satisfy their service requirements and dynamically adapt the routing schemes according to the mode of content delivery (i.e., uni-cast or multi-cast) (Huawei Technologies Co., 2016). The transmission protocols are designed according to the features of VR content delivery. The transmission protocol based on quick UDP Internet connections (QUIC) is proposed in (Yen et al., 2019) to prioritize important tiled videos, such as the videos in the center of the user's FoV or the videos to be played soon, in transmission over a QUIC connection, in order to minimize the ratio of missing tiles in VR video playback.

4.1.3 Computing Optimization

Supporting wireless XR requires networks to have sufficient computing capability for processing and rendering the content, especially for interactive applications such as VR gaming. Processing the content locally at the XR devices can be time-consuming and energy-inefficient due to their limited computing capability. Instead, the computing workload can be fully or partially offloaded to network servers, and multi-tier computing can be a potential solution to reduce computing time and bandwidth consumption when providing computing services to XR devices. Accordingly, computing strategies should base on the features of diverse network servers to improve resource utilization and service performance.

In MEC, edge servers can provide additional computing capability for resource-limited devices to reduce content processing latency for mobile XR content delivery. Specifically, in VR, edge servers can project monoscopic videos to stereoscopic videos when content is transmitted from the content provider's cloud server to VR devices. Such MEC-assisted content delivery can reduce bandwidth consumption compared to delivering stereoscopic videos from the cloud server directly, and computing time can be reduced compared to projecting the videos at the local devices (Mangiante et al., 2017). In AR, devices can offload captured content to an edge server to minimize processing latency (Siriwardhana et al., 2021). In addition, edge servers can cache the processed XR content to further reduce the content delivery and processing time (Sukhmani et al., 2018). Joint computing, caching, and communication resource management for VR video delivery is investigated in (Sun et al., 2019; Dang and Peng, 2019), which studies the tradeoff between computing and caching resource allocation for minimizing content delivery delay, given stochastic content processing time and popularity. Deep reinforcement learning methods are adopted to allocate computing resources at an edge server for individual content delivery requests in (Liu et al., 2019b; Liu and Deng, 2021), aiming to minimize content delivery delay while adapting to dynamic network environments and user viewpoint movement.

Nonetheless, the computing capability at edge servers may not always be sufficient for processing XR content. Compared to cloud servers, edge servers usually have limited storage resources for caching XR content. Targeting 6G, a multi-tier computing architecture provides a potential solution for further accelerating XR content delivery by coordinating computing and storage resources among cloud servers, fog servers (e.g., servers at the gateway), and edge servers across the network. By integrating computing resources across the entire network, content processing workloads can be optimally distributed among multiple servers, and storage capacity among servers can be utilized to satisfy offloaded computing demands. However, optimizing XR performance by multi-tier computing can be complicated when there are a multitude of computing offloading and caching options to choose from. The computing and caching resource coordination between the cloud server and edge servers is studied in (Mehrabi et al., 2021) and (Al-Abbasi et al., 2019). Based on the information of a static network environment, e.g., transmission rate and XR computing demand, mixed integer nonlinear programming is investigated. Considering dynamic network environments and user mobility, the work (Zhou et al., 2022a) utilizes digital twins of end users to characterize network dynamics and statuses. The meta-learning method is adopted to jointly allocate computing and caching resources at servers on different tiers of a network for context-based applications, including XR, based on the captured network statuses from digital twins.

In addition to jointly allocating computing and caching resources at network servers, computing performance can be further enhanced by scheduling computing tasks at edge servers. Edge servers can provide location-based content to users, which can contribute to computing optimization for XR applications. Specifically, in AR, users at close locations may offload and require similar content, and therefore, raw content offloaded from the nearby users can be processed together for improving computing efficiency (Jia and Liang, 2018). Furthermore, rendering pipelines can be optimized based on real-time communication and computing performance of network servers and local devices when part of the workloads for content rendering are offloaded. A collaborative rendering pipeline is investigated in (Xie et al., 2021), which dynamically arranges the execution order of sub-tasks in content rendering on both the edge server and XR devices, based on network characteristics, to facilitate parallel computing and improve content rendering efficiency.

4.2 Haptic Communication

The main challenge in haptic communication is to satisfy the stringent delay and reliability requirements in the delivery of haptic data, especially when the data packet rate is high. To tackle this challenge, solutions have been developed in three aspects, including haptic data reduction to reduce the packet size or the packet rate, advanced communication and networking techniques to reduce delay and improve reliability, and haptic data prediction to compensate for excessive delay and packet loss over communication networks.

4.2.1 Haptic Data Reduction

To improve the fidelity of haptic perception, the number of haptic sensors/actuators deployed on an HI has been increasing (Steinbach et al., 2018). For example, electronic skin (e-skin) can be attached to prosthetic limbs for sensing haptic information, or to human skin for virtual social interaction (Yu et al., 2019; Dahiya, 2019). To reproduce the function of human skin, sensors/actuators need to be densely deployed on e-skin, for example, 25 sensors/actuators per 1 cm^2 (Liu et al., 2020). In addition, the required packet rate for haptic data can be above 1,000 packets per second. As a result, with a large number of devices and a high packet rate, the required data rate of haptic communication can be high. To tackle this challenge, one solution is haptic data reduction, which is to reduce the packet size or rate of haptic data.

For reducing the packet size of haptic data, floating-point compression in the time domain or quantization of haptic data in the frequency domain can be exploited. In floating-point compression, one degree of freedom in the haptic information (e.g., the direction of the transitional movement in an axis) can be represented by a 32-bit floating-point number, and only the bits different from those in the previous haptic data are transmitted (You and Sung, 2008). Using time-frequency transformation algorithms such as discrete cosine transform, a sequence of haptic data packets in the time domain can be transformed into the data in the frequency domain, which are then quantized and transmitted (Tanaka and Ohnishi, 2009; Zeng et al., 2020). For reducing the packet rate of haptic data, the perceptual masking phenomenon is widely exploited, which suggests that a human cannot perceive the difference of haptic information below the just-noticeable difference (JND). According to the Weber's law, the JND of haptic information is proportional to the currently perceived value of the information, and the proportion is referred to as the Weber fraction (Steinbach et al., 2018). In this regard, the perceptual haptic reduction method is to transmit an updated haptic data packet only when the difference is larger than a threshold (e.g., JND) (Steinbach et al., 2010). Moreover, the perceptual masking phenomenon in both time and frequency domains can be jointly exploited to achieve a higher data reduction ratio and lower data deviation (Wei et al., 2022).

The use of haptic data reduction should adapt to the type of haptic data, the delay requirement and the reliability requirement for haptic communication. First, haptic data can exhibit different Weber fractions in the JND, e. g., $7\% \sim 15\%$ for force data and $13\% \sim 28\%$ for stiffness data, which results in different thresholds in perceptual haptic reduction (Chaudhuri and Bhardwaj, 2018). Second, data reduction in the frequency domain results in high processing delay since it is based on a sequence of data packets in the time domain. It is suitable for use cases with high delay tolerance, such as the passive perception and exploration of remote/virtual objects (Sachs et al., 2018). In contrast, data reduction in the time domain, implemented in real time, is suitable for use cases with low delay tolerance such as immersive gaming, which involves extensive interactions between the players (Holland et al., 2019). Third, haptic data reduction may not be suitable for use cases requiring high reliability. As discussed in Subsection 3.2.3, with the use of haptic data reduction, the required reliability of haptic communication increases. In this regard, for use cases with a high-reliability requirement (e.g., 99.999% for telesurgery), the reliability requirement can be difficult to satisfy if haptic data reduction is used.

4.2.2 Communication and Networking Solutions

To satisfy the ideal communication delay of below 1 ms for haptic communication, physical-layer delay of less than 0.1 ms is desired (Aijaz et al., 2016). For reducing the queuing delay, haptic data may be allowed to preempt the data of other types in the downlink transmission (Ji et al., 2018). For uplink transmission, a grant-based user scheduling mechanism can take 0.3-0.4 ms for exchanging the scheduling request and transmission grant (Ji et al., 2018). Besides such delay, the signaling overhead, resulting from network control or grant-based scheduling, reduces the efficiency of data transmission (Ding et al., 2021). Therefore, grant-free user scheduling has been exploited, which periodically pre-reserves transmission resources to avoid real-time scheduling, and the same resources can be pre-reserved to multiple haptic devices for improving resource utilization (Ali et al., 2021; Gao et al., 2021). For improving the communication reliability of haptic communication, several approaches have been adopted in the literature. First, considering the small size of a haptic data packet, short block-length channel codes with strong error correction capabilities, such as low-density parity-check (LDPC) codes and short polar codes, have been investigated for haptic communication (Yuan et al., 2022; Miloslavskaya and Vucetic, 2020). Second, spatial diversity can be exploited by massive multiple-input and multiple-output (MIMO), IRS, and multi-connectivity techniques (Tarneberg et al., 2017; Tang et al., 2020; Anwar et al., 2021). Third, time diversity can be exploited by a K-repetition mechanism, in which a haptic device can automatically transmit K repetitions of a packet over consecutive slots, thereby avoiding the delay caused by waiting for a retransmission request from the receiver (Yang et al., 2021).

To guarantee low delay and high reliability for haptic communication, network slicing, which allows multiple isolated virtual networks to be constructed over a shared physical network infrastructure, has been exploited (Polachan et al., 2020). The perceptual masking phenomenon of haptic information, as introduced in Subsection 4.2.1, can be exploited to accurately capture the maximum tolerable delay of haptic communication requests, which facilitates resource reservation in the network slice for haptic communication (Ge et al., 2019). For multiple tele-operation slices, diverse stability control capabilities of tele-operators in the presence of delay should be considered for customized transmission resource reservation (Liu et al., 2018b). Moreover, by exploiting AI-based learning methods, traffic patterns of haptic devices can be accurately captured, and efficient resource reservation can then be facilitated (Shen et al., 2020).

4.2.3 Haptic Data Prediction

The delay requirement of haptic communication can impose a constraint on the distance between two users. For example, to satisfy a delay requirement of 10 ms, the distance between a transmitter HI and a receiver HI must be smaller than 3,000 km since the propagation speed is upper-bounded by the speed of light. This can create an issue for applications such as VR gaming with haptic interactions of players across continents. In addition, it is impossible to eliminate the loss of data packets or the violation of delay requirement in haptic communication (Aijaz and Sooriyabandara, 2018). To improve user experience considering the above facts, haptic data prediction can be exploited.

For haptic data prediction, model-based or model-free prediction algorithms can take historical haptic data and other correlated data as the input. In tele-operation, the force feedback from the tele-operator is predicted by evaluating the previous force feedback through an auto-regressive model (Sakr et al., 2007). In the tele-operated needle insertion, the force/torque feedback from the patient is predicted by inputting the force/torque commands of the surgeon to the hidden Markov model (HMM) (Boabang et al., 2020). Audiovisual data collected in the interaction with a surface material are input to a neural network-based semantic learning algorithm to predict the texture of the surface material (Wei et al., 2021).

Haptic data can be predicted either at the receiver side or at the transmitter side to compensate for an excessive delay or packet loss. The receiver can predict the haptic data from the transmitter when an excessive delay occurs (Maier and Ebrahimzadeh, 2019). For example, digital twin-based prediction can be used by the receiver for real-time interactions (El Saddik, 2018). Alternatively, the transmitter can predict its future haptic data and transmit the predicted data to compensate for the transmission delay (Hou et al., 2019). In this case, the prediction of whether haptic interaction is about to occur can assist to determine whether the haptic data prediction and the subsequent transmission are necessary (Mondal et al., 2020).

Haptic data prediction algorithms, such as AI-based ones, can be computing-intensive. To this end, they can be implemented using computing resources in the network to satisfy the stringent delay requirements (Simsek et al., 2016; Sukhmani et al., 2018). In a tele-operation scenario, each of the two interacting haptic devices is associated with one edge server which caches the haptic interaction data, trains and implements the LSTM network-based prediction algorithm, and delivers the predicted haptic data to its associated haptic device (Li et al., 2021b). Furthermore, with close proximity, auxiliary robots can be deployed around haptic devices to implement haptic data prediction and deliver the results to the devices using device-to-device (D2D) communications (Yu et al., 2022).

In addition to compensating for the delay or packet loss, haptic data prediction can be used to reduce the packet rate of haptic data (Antonakoglou et al., 2018). Specifically, the haptic transmitter can implement the haptic data prediction and evaluate the prediction deviation, and only transmit the data when the prediction deviation is higher than the JND of the receiver. If the haptic data has not been transmitted, the receiver can predict it based on the prediction algorithm shared with the transmitter.

4.3 Holographic Communication

In holographic communication, users are able to view 3D holograms from different angles, tilts, and positions. As a result, a hologram synthesized with information from more viewpoints can produce more detailed and continuous visual information for users, thereby creating a more realistic immersive experience (Liu et al., 2019a). This however requires the transmission of a large amount of data. The main challenge in holographic communication is its stringent data rate and delay requirements. In this subsection, we focus on potential solutions for tackling this challenge in the aspects of data processing, communication and networking.

4.3.1 Content Selection, Compression and Prediction

A high data rate is essential for holographic communication, and the demand for data rate can vary from hundreds of Mbps to several Tbps depending on the type of transmitted data, e.g., volumetric-based or image-based holograms. One way to relax the data rate requirement is to reduce the data size, for example, by transmitting only the most essential parts of a hologram through viewpoint-based content selection in holographic communication (Clemm et al., 2020). Since some parts of the hologram may not be observed depending on the user's viewpoint and position, as well as the presence of obstacles, those parts may not need to be transmitted. However, two issues remain even with the selective transmission. First, for an immersive experience in holographic communication, the 6 DoF (yaw, pitch, roll, up/down, left/right, forward/backward) need to be considered when a user views a hologram, which makes content selection based on the user's viewpoint a complex problem. In addition, without head-mount devices such as VR headsets, tracking the position and viewpoint of the user is challenging and requires mechanisms such as full-body tracking (Xu et al., 2018a) or eye tracking (Zhang et al., 2019a).

Another solution for reducing the required data rate is to apply data compression. For a 2D real-time video, current media codecs can achieve a compression ratios from 250:1 to 1,000:1 (Selinis et al., 2020; Essaili et al., 2022). Similarly, format conversion and data compression can be applied to reduce the data size in holographic communication. The authors in (Mekuria et al., 2017) propose a lossy real-time color-encoding method by exploiting the inter-frame redundancy of point clouds. Moreover, considering the strong correlation among different views in a hologram, multi-view coding (MVC) for LFV-based streaming is proposed in (Xiang et al., 2016), which improves the compression rate by analyzing both the horizontal and vertical correlations of images in adjacent angles and tilts. Meanwhile, many efforts have been made by standardization groups for the compression of holograms. For example, the Moving Picture Experts Group (MPEG) defined the video point cloud compression (V-PCC) by converting point clouds into two separate video sequences that capture the geometry and texture information, respectively (Schwarz et al., 2019). The Joint Photographic Experts Group (JPEG) intended to provide a standard representation framework to facilitate the compression of LFV- or point cloud-based content for holographic communication (Schelkens et al., 2019).

Retransmissions due to data packet loss result in additional delay. To avoid the retransmission delay, the lost data packets can be recovered based on predicted data according to historical information of an object such as its trajectory. For example, packets can be recovered from an LSTM-based prediction of human actions and movements in 3D (Liu et al., 2016a) or a short-term prediction by analyzing the actions, movements, or gestures of users (Manolova et al., 2021). By predicting content, data packets can be generated at the receiver side in the event of packet loss to reduce the delay in holographic communication (Strinati et al., 2019).

4.3.2 Communication and Networking Solutions

In addition to data processing, some communication and networking solutions have been investigated for satisfying delay and data rate requirements of holographic communication, including computing architecture, transport protocol, and physical layer technology.

In holographic communication, data captured from different sensors needs to be processed to form a 3D representation of the object, which is then rendered and reconstructed at the receiver side (Javidi et al., 2005). However, the limited computing capability of local devices may lead to a long processing delay due to the high workload of data fusion and rendering (Hu et al., 2017). Cloud computing is introduced to support high computing workloads for data processing in holographic communication. However, transmitting massive data to the cloud may result in a high communication delay (Wang et al., 2022a), which is not suitable for real-time holographic communication. One promising solution is to offload computing tasks to MEC servers for data processing, since MEC servers possess considerable computing capability and are placed close to users (Gupta et al., 2021). Thanks to network function virtualization (NFV), functions such as data fusion, data compression, and data rendering can be virtualized and flexibly deployed for MEC servers. In this case, captured data from different sources can be aggregated, fused, and synchronized at an MEC server before rendering (Qian et al., 2022). Moreover, a multi-tier computing scheme is proposed for 6G networks, which can be utilized for holographic communication by integrating computing resources at cloud servers, MEC servers, and local devices, to achieve a low delay for data transmission and high computing capacity for data processing with collaboration among different servers (Yang et al., 2018; Wang et al., 2022a). By integrating computing resources on different tiers, content can be processed at different servers to effectively utilize computing resources, and flexible computing resource management should be developed to facilitate multi-tier computing for holographic communication. For example, split rendering is introduced for an MEC server and a local device to cooperatively decode and render holograms according to the content (Essaili et al., 2022).

To satisfy the stringent delay and high reliability requirements of holographic communication, transport layer optimizations are also crucial. Current transport protocols, such as transmission control protocol (TCP) and user datagram protocol (UDP), can hardly satisfy the requirements of holographic communication. To improve the reliability and delay performance in real-time communication, new protocols based on UDP are introduced, such as QUIC over HTTP/3 (Seufert et al., 2019). Currently, the research on QUIC mainly focuses on traditional 2D video streaming services, while QUIC can serve as a potential solution for holographic communication, providing a quality-managed low-delay streaming option (Clemm et al., 2020). Moreover, the transmission of a hologram may consist of multiple substreams corresponding to different viewpoints, while the QoS requirement and the priority of each substream may be different. In this case, the transmission of the most essential substreams needs to be prioritized. To achieve this target, a new transport protocol is designed in (Rozen-Schiff et al., 2021) for holographic communication to satisfy different QoS requirements of different flows by providing flow-level granular control. In addition, an adaptive retransmission mechanism based on TCP is designed to reduce retransmissions by analyzing and differentiating packets (Clemm et al., 2020). For example, only important data, such as the data used for rendering the part of the hologram in the center of the user's FoV, will be retransmitted if the related packets are lost, to reduce retransmissions.

Finally, physical layer solutions are important to supporting a high data rate for holographic communication. In order to transmit high-resolution LFV-based holograms, holographic communication requires a data rate of several Tbps, while current 5G networks cannot support it (Shahraki et al., 2021; David and Berndt, 2018). Featuring higher frequency and larger bandwidth compared with mmWave in 5G, THz communications have the potential to support holographic communication with Tbps-level data rate (Elayan et al., 2019; Chen et al., 2019). To overcome the severe propagation loss of THz communication, dense deployment of access points and extremely narrow beams can be adopted to improve connection density and communication reliability (Zhang et al., 2019b). Considering the absorption and reflection properties in the THz regime (Aazhang et al., 2019), the deployment of the THz base stations and the prediction of user motion require further investigation to provide sustainable LoS links for holographic communication (Chaccour et al., 2022).

5 IMMERSIVE COMMUNICATION: OPEN ISSUES AND FUTURE DIRECTIONS

Despite an increasing amount of studies and solutions for supporting XR, haptic communication, and holographic communication, there exist many open issues to address before immersive communications can popularize. To name a few, synchronization of multi-modal communications, user QoE modeling and enhancement, and intelligent network management for immersive communications remain to be challenging problems. In this section, we present some major open issues in immersive communications and potential future directions to address these issues.

5.1 Multi-Modal Communications

While immersive communications have the potential to enhance user engagement and facilitate immersive interactions, effective network resource management for ensuring synchronized multi-modal perception in highly dynamic network environments is an open issue. The synchronization of multi-modal perception consists of two aspects: inter-stream (cross-modal) and intra-stream. First, the transmission of auditory, visual, and haptic data results in multiple data streams that should be synchronized in order to prevent motion sickness. For example, the time interval between perceived visual and tactile movement should not exceed 1 ms (Van Den Berg et al., 2017). Second, to enhance the immersive experience, a data stream can include multiple data substreams corresponding to different sensations, e.g., temperature and pressure, which also need synchronization. Data substreams corresponding to different DoF of an HI should be synchronized to maintain the stable perception of simultaneity, and data substreams transmitted from LIDAR sensors placed at different locations should be synchronized to render a 3D hologram precisely. There are many works that enable either intra-stream or inter-stream synchronization from the perspective of a single network layer (Zhang et al., 2018; Cizmeci et al., 2017). However, in order to synchronize multi-modal perception, both network-related and application-related information is necessary. This is because network resource management for multi-modal communications is affected by not only different data packet formats, data traffic patterns, and QoS requirements, but also different sensitivities of human perception. The cross-layer design of network protocols for multi-modal communications, which can support information sharing among different layers for efficient use of network resources, is a potential solution (She et al., 2020; Kumar and Muhammad, 2018). A higher-layer approach to synchronizing multi-modal information can benefit from information on network conditions at lower layers, e.g., adaptively changing the priority of modalities in transport-layer multiplexing according to real-time physical-layer data rates. In addition, lower-layer approaches can take into account application-related information for efficient network resource management, e.g., timely adjusting the amount of radio resources allocated to a user in response to the dynamic sensitivity of the user's perception.

5.2 AI-Native Immersive Communications

AI techniques have demonstrated outstanding performance in identifying data correlations and analyzing device dynamics. As a result, some application functions using AI techniques, i.e., AI-enabled functions, have been developed for exploring unknown device states in immersive communications, such as viewpoint predictions in VR devices and haptic data prediction (Wu et al., 2022). To support increased service demands on immersive communications in 6G, AI-enabled functions will be deployed at network servers, i.e., cloud and edge servers (Li et al., 2021a). Accordingly, the network should support the entire lifecycle of AI for the functions, including data collection, data pre-processing, AI model training, inference, and AI model evaluation. By taking AI-enabled functions as the built-in component for supporting immersive communications, several potential future research directions should be investigated. First, AI-enabled functions can be configured according to network management policies for supporting immersive communications. For example, in haptic communication, the prediction horizon, i.e., the time window for the predicted information, of tactile and kinesthetic information can be adjusted adaptively according to real-time network communication delay, AI-based prediction accuracy, and service reliability requirements. Second, efficient data management schemes can be developed, in which low-signaling-overhead and grant-free network management can be achieved by sharing the data obtained from AI-enabled functions. For example, in VR video delivery, network controllers can use a viewpoint prediction model or results from the viewpoint prediction function and allocate sufficient downlink communication resources to users with highly dynamic viewpoint movements. Additionally, effective resource management solutions should be developed to support AI model training in real-time, so that AI-enabled functions can be updated according to user behavior dynamics, where sufficient network resources should be allocated for supporting data collection and processing at edge and cloud servers.

5.3 Time-Sensitive and Deterministic Networking

The existing solutions mentioned in Section 4 can help reduce transmission delay in immersive communications. However, satisfying the stringent delay and reliability requirements of XR, haptic communication, and holographic communication, especially ms-level end-to-end delay, remains a challenge. Fortunately, the ongoing efforts of 3GPP, IEEE, and IETF in supporting time-sensitive networking (TSN) and deterministic networking (DetNet) (Messenger, 2018; Nasrallah et al., 2019) provide solutions to meet the requirements of immersive communications (Rost and Kolding, 2022). The current efforts largely focus on the link and network layers (i.e., layers 2 and 3) and mostly target industrial networks (Rost and Kolding, 2022). Therefore, the corresponding solutions may not be readily applicable to all use cases of immersive communications. Potential future directions of TSN and DetNet for immersive communications include the followings. First, a comprehensive solution integrating existing TSN and DetNet designs for delay minimization can be important to immersive communications. For example, the joint design of coordinated sensing/capturing and communication (on the physical layer), traffic shaping and scheduling (on the link layer), flow identification and packet treatment (on the network layer), and viewpoint/haptic data prediction (on the application layer) can help reduce the end-to-end delay in immersive communications. Second, instead of treating different data streams in a mutli-modal communication separately, joint prioritization and resource orchestration for different types of data given their respective delay and jitter requirements is another promising direction. Third, integrating environment-aware and service-oriented network management paradigms can potentially enable TSN and DetNet for immersive communications. An example is to incorporate adaptive radio access network (RAN) function splitting, network slicing, and AI-driven network management to minimize delay and jitter by customizing for a specific service and adapting to the network environment.

5.4 QoE-Oriented Networking

While QoS provisioning from a network perspective benefits the transmission of XR content, haptic information, and holograms, as detailed in Section 4, evaluating and guaranteeing individual users' QoE is crucial in providing them an immersive experience. This is because many factors, besides communication network conditions, can affect user experience in immersive communications, including coding, compression, and human perception. Therefore, QoE-oriented networking from users' perspective is a promising network management paradigm to support immersive communications in the 6G era, including two potential aspects: personalized QoE modeling and QoE-oriented network resource management. First, existing works on immersive communications have limitations on personalizing QoE models for individual users. Conventional QoE modeling are based on either subjective tests or objective quality assessments (Tasaka, 2022). The former, conducted in relatively static laboratory environments, is costly and inapplicable in dynamic network environments, whereas the latter, evaluated by empirical human perception models, does not differentiate individual users (Barakabitze et al., 2019; Ruan and Xie, 2021). Finding a way to model personalized QoE while adapting to dynamic network environments remains an open issue. Second, managing network resources to guarantee the QoE of individual users in immersive communications necessitates user-level information. Even if several users request the same service, they may have different resource demands for improving their QoE (Kougioumtzidis et al., 2022). For example, due to the difference in the sensitivity of haptic perception, e.g., reaction time, the haptic sensors of interest and the scan time for each haptic sensor may differ in supporting different users, yielding different communication and computing resource demands (Coutinho and Boukerche, 2022). In the 6G era, the paradigm of digital twins can be a potential solution for QoE-oriented networking. Specifically, individual users can be characterized by creating user digital twins, including user data profiles that contain extensive well-organized user data, and a variety of digital twin functions that support flexible and customized data collection and analysis (Shen et al., 2021). Both personalized QoE modeling and QoE-oriented network resource management for immersive communications can benefit from extensive timely updated and fine-grained user-level information (Wang et al., 2021).

5.5 New Network Architecture

Network architecture innovation is indispensable for a widespread realization of immersive communications, and innovations building on recent developments for 6G architecture are promising future directions. The need for new architectures manifests in several aspects. First, the computing-intensive nature of immersive communications, rooted from processing and compressing 3D data, predicting of viewpoints and haptics data, and reconstructing 3D objects, demands a network architecture with extensive computing resources and reliable computing service provisioning. As a result, a heterogeneous network with multi-tier computing architecture (Yang, 2019; Zhou et al., 2022a), featuring on-demand and collaborative computing task offloading and scheduling across the network, is important to immersive communications yet open to investigation at the moment. Second, as networks become increasingly complex and the requirements of immersive communications become exceedingly stringent, supporting immersive communications in 6G requires a network architecture with unprecedented scalability, flexibility, and adaptivity. A 6G architecture integrating digital twins, network slicing, and pervasive AI (Shen et al., 2021) can be a foundation to immersive communications. Third, considering the diverse delay requirements of different XR, haptic communication, and holographic communication use cases, the Open-RAN (O-RAN) architecture featuring realtime, near-realtime, and non-realtime layers can benefit service differentiation in RAN management for immersive communications (Abdalla et al., 2022). Last, considering different user preferences and diverse user devices, a new architecture enabling user-centric networking, such as the everyone-centric architecture in (Yang et al., 2022b), has a potential to empower immersive communications. However, as none of the above architectures is developed specifically for immersive communications, new designs and customizations based on them for supporting immersive communications are open for investigation.

6 CONCLUSION

In this article, we have delved into immersive communications towards 6G and presented a comprehensive review of the related concepts, representative use cases, technical challenges and potential solutions, and future directions. Focusing on XR, haptic communication, and holographic communication, we have illustrated their general procedures, network requirements, and recent developments in the context of a vision for 6G. Despite abundant emerging use cases and exciting recent advancements, we have shown that many challenges are yet to be conquered before the envisioned prosperity of immersive communications can occur. In particular, the exceeding transmission rate, delay, and reliability requirements, further complicated by the multi-modal and computing-intensive features of immersive communications, indicate the necessity of an unprecedented amount of communication and computing resources as well as novel paradigms such as AI-native communication, multi-tier computing, and user-centric networking.

The paradigm shift to immersive communications is truly exciting and inspiring, especially when viewed in the context of the evolution toward 6G. Many opportunities exist, and more will emerge for researchers and engineers in the fields of communications, networking, and computer science to realize immersive communications. We hope this review inspires further interest among fellow researchers and provides fundamental knowledge on related research, thereby contributing to this much-anticipated paradigm shift and making immersive communications the next reality.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Each author of this paper made a direct and important contribution to this work and approved its publication.

REFERENCES

Aazhang, B., Ahokangas, P., Alves, H., Alouini, M.-S., Beek, J., Benn, H., et al. (2019). *Key drivers and research challenges for 6G ubiquitous wireless intelligence (white paper)* (Oulu: 6G Flagship, University of Oulu)

Abari, O., Bharadia, D., Duffield, A., and Katabi, D. (2016). Cutting the cord in virtual reality. In *Proceedings of the 15th ACM Workshop on Hot Topics in Networks* (ACM), 162–168

Abdalla, A. S., Upadhyaya, P. S., Shah, V. K., and Marojevic, V. (2022). Toward next generation open radio access networks–what O-RAN can and cannot do! *IEEE Network* to be published. doi:10.1109/MNET.108.2100659

Abiri, A., Pensa, J., Tao, A., Ma, J., Juo, Y.-Y., Askari, S. J., et al. (2019). Multi-modal haptic feedback for grip force reduction in robotic surgery. *Scientific Reports* 9, 5016. doi:10.1038/s41598-019-40821-1

Affan, M., Khan, H., Ahmed, S., Uddin, R., and Shirazi, M. A. (2021). Haptic-enabled virtual laboratory for hands-on E-learning: A technology for and beyond the pandemic era. In *IEEE International Conference on Robotics and Automation in Industry* (IEEE), 1–6

Aheleroff, S., Xu, X., Zhong, R. Y., and Lu, Y. (2021). Digital twin as a service (DTaaS) in industry 4.0: An architecture reference model. *Advanced Engineering Informatics* 47, 1–15. doi:10.1016/j.aei.2020.101225

Ahmad, A. S., Alomaier, A. T., Elmahal, D. M., Abdlfatah, R. F., and Ibrahim, D. M. (2021). EduGram: Education development based on hologram technology. *International Journal of Online & Biomedical Engineering* 17, 32–49. doi:10.3991/ijoe.v17i14.27371

Aijaz, A., Dohler, M., Aghvami, A. H., Friderikos, V., and Frodigh, M. (2016). Realizing the tactile Internet: Haptic communications over next generation 5G cellular networks. *IEEE Wireless Communications* 24, 82–89. doi:10.1109/MWC.2016.1500157RP

Aijaz, A. and Sooriyabandara, M. (2018). The tactile Internet for industries: A review. *Proceedings of the IEEE* 107, 414–435. doi:10.1109/JPROC.2018.2878265

Al-Abbasi, A. O., Aggarwal, V., and Ra, M.-R. (2019). Multi-tier caching analysis in CDN-based over-the-top video streaming systems. *IEEE/ACM Transactions on Networking* 27, 835–847. doi:10.1109/TNET.2019.2900434

Ali, R., Zikria, Y. B., Bashir, A. K., Garg, S., and Kim, H. S. (2021). URLLC for 5G and beyond: Requirements, enabling incumbent technologies and network intelligence. *IEEE Access* 9, 67064–67095. doi:10.1109/ACCESS.2021.3073806

Ali-Yahiya, T. and Monnet, W. (2022). *The tactile Internet* (Hoboken, NJ: John Wiley & Sons)

Antonakoglou, K., Xu, X., Steinbach, E., Mahmoodi, T., and Dohler, M. (2018). Toward haptic communications over the 5G tactile Internet. *IEEE Communications Surveys & Tutorials* 20, 3034–3059. doi:10.1109/COMST.2018.2851452

Anwar, W., Kumar, A., Franchi, N., and Fettweis, G. (2021). Physical layer abstraction for multi-connectivity communications: Modeling and analysis. *IEEE Transactions on Wireless Communications* 21, 1779–1793. doi:10.1109/TWC.2021.3106771

Baik, S., Han, I., Park, J.-M., and Park, J. (2020). Multi-fingertip vibrotactile array interface for 3D virtual interaction. In *Proceedings of the IEEE Haptics Symposium* (IEEE), 898–903

Barakabitze, A. A., Barman, N., Ahmad, A., Zadtootaghaj, S., Sun, L., Martini, M. G., et al. (2019). QoE management of multimedia streaming services in future networks: A tutorial and survey. *IEEE Communications Surveys & Tutorials* 22, 526–565. doi:10.1109/COMST.2019.2958784

Bastug, E., Bennis, M., Medard, M., and Debbah, M. (2017). Toward interconnected virtual reality: Opportunities, challenges, and enablers. *IEEE Communications Magazine* 55, 110–117. doi:10.1109/MCOM.2017.1601089

Billström, O., Cederquist, L., Ewerbring, M., Sandegren, G., and Uddenfeldt, J. (2006). Fifty years with mobile phones from novelty to no. 1 consumer product. *Ericsson Review* 3, 101–106

Boabang, F., Glitho, R., Elbiaze, H., Belqami, F., and Alfandi, O. (2020). A framework for predicting haptic feedback in needle insertion in 5G remote robotic surgery. In *Proceedings of the IEEE 17th Annual Consumer Communications & Networking Conference* (IEEE), 1–6

Carroll, M. and Yildirim, C. (2021). The effect of body-based haptic feedback on player experience during VR gaming. In *International Conference on Human-Computer Interaction* (Springer), 163–171

Chaccour, C., Soorki, M. N., Saad, W., Bennis, M., and Popovski, P. (2020). Risk-based optimization of virtual reality over Terahertz reconfigurable intelligent surfaces. In *IEEE International Conference on Communications* (IEEE), 1–6

Chaccour, C., Soorki, M. N., Saad, W., Bennis, M., Popovski, P., and Debbah, M. (2022). Seven defining features of terahertz (THz) wireless systems: A fellowship of communication and sensing. *IEEE Communications Surveys & Tutorials* 24, 967–993. doi:10.1109/COMST.2022.3143454

Chaudhuri, S. and Bhardwaj, A. (2018). Predictive sampler design for haptic signals. In *Kinesthetic Perception*, ed. J. Kacprzyk (Singapore: Springer). 29–53

Chen, J., Lu, J.-L., and Ochiai, Y. (2022). Simulation object edge haptic feedback in virtual reality based on dielectric elastomer. In *Proceedings of the International Conference on Human-Computer Interaction* (Springer), 3–9

Chen, K., Kamezaki, M., Katano, T., Ishida, T., Seki, M., Ichiryu, K., et al. (2016). Analysis of operation strategy in a multi-operator control system for four-arm disaster response robot OCTOPUS. In *IEEE/SICE International Symposium on System Integration* (IEEE), 514–519

Chen, K.-C., Lin, S.-C., Hsiao, J.-H., Liu, C.-H., Molisch, A. F., and Fettweis, G. P. (2021a). Wireless networked multirobot systems in smart factories. *Proceedings of the IEEE* 109, 468–494. doi:10.1109/JPROC.2020.3033753

Chen, N., Quan, S., Zhang, S., Qian, Z., Jin, Y., Wu, J., et al. (2021b). Cuttlefish: Neural configuration adaptation for video analysis in live augmented reality. *IEEE Transactions on Parallel and Distributed Systems* 32, 830–841. doi:10.1109/TPDS.2020.3035044

Chen, Z., Ma, X., Zhang, B., Zhang, Y., Niu, Z., Kuang, N., et al. (2019). A survey on Terahertz communications. *China Communications* 16, 1–35. doi:10.12676/j.cc.2019.02.001

Cheremkhin, P. and Kurbatova, E. (2019). Wavelet compression of off-axis digital holograms using real/imaginary and amplitude/phase parts. *Scientific Reports* 9, 1–13. doi:10.1038/s41598-019-44119-0

Choi, P. J., Oskouian, R. J., and Tubbs, R. S. (2018). Telesurgery: past, present, and future. *Cureus* 10. doi:10.7759/cureus.2716

Choi, S. and Tan, H. (2004). Effect of update rate on perceived instability of virtual haptic texture. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1–6

Cizmeci, B., Xu, X., Chaudhari, R., Bachhuber, C., Alt, N., and Steinbach, E. (2017). A multiplexing scheme for multimodal teleoperation. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 13, 1–28. doi:10.1145/3063594

Clemm, A., Vega, M. T., Ravuri, H. K., Wauters, T., and De Turck, F. (2020). Toward truly immersive holographic-type communication: Challenges and solutions. *IEEE Communications Magazine* 58, 93–99. doi:10.1109/MCOM.001.1900272

Cobos, S., Ferre, M., Uran, M. S., Ortego, J., and Pena, C. (2008). Efficient human hand kinematics for manipulation tasks. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE), 2246–2251

Conti, J.-P. (2008). In search of the princess Leia effect. *Engineering & Technology* 3, 66–69. doi:10.1049/et:20081613

Coutinho, R. W. and Boukerche, A. (2022). Design of edge computing for 5G-enabled tactile Internet-based industrial applications. *IEEE Communications Magazine* 60, 60–66. doi:10.1109/MCOM.001.21261

Culbertson, H., Schorr, S. B., and Okamura, A. M. (2018). Haptics: The present and future of artificial touch sensation. *Annual Review of Control, Robotics, and Autonomous Systems* 1, 385–409. doi:10.1146/annurev-control-060117-105043

Dahiya, R. (2019). E-skin: from humanoids to humans [point of view]. *Proceedings of the IEEE* 107, 247–252. doi:10.1109/JPROC.2018.2890729

Dahiya, R., Yogeswaran, N., Liu, F., Manjakkal, L., Burdet, E., Hayward, V., et al. (2019). Large-area soft e-skin: The challenges beyond sensor designs. *Proceedings of the IEEE* 107, 2016–2033. doi:10.1109/JPROC.2019.2941366

Dai, J., Zhang, Z., Mao, S., and Liu, D. (2020). A view synthesis-based 360° VR caching system over MEC-enabled C-RAN. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 3843–3855. doi:10.1109/TCSVT.2019.2946755

Dang, T. and Peng, M. (2019). Joint radio communication, caching, and computing design for mobile virtual reality delivery in fog radio access networks. *IEEE Journal on Selected Areas in Communications* 37, 1594–1607. doi:10.1109/JSAC.2019.2916486

Dardari, D. and Decarli, N. (2021). Holographic communication using intelligent surfaces. *IEEE Communications Magazine* 59, 35–41. doi:10.1109/MCOM.001.2001156

David, K. and Berndt, H. (2018). 6G vision and requirements: Is there any need for beyond 5G? *IEEE vehicular technology magazine* 13, 72–80. doi:10.1109/MVT.2018.2848498

del Peral-Rosado, J. A., Raulefs, R., López-Salcedo, J. A., and Seco-Granados, G. (2018). Survey of cellular mobile radio localization methods: From 1G to 5G. *IEEE Communications Surveys & Tutorials* 20, 1124–1148. doi:10.1109/COMST.2017.2785181

Diana, M. and Marescaux, J. (2015). Robotic surgery. *Journal of British Surgery* 102, e15–e28. doi:10.1002/bjs.9711

Ding, J., Nemati, M., Pokhrel, S. R., Park, O.-S., Choi, J., and Adachi, F. (2021). Enabling grant-free URLLC: An overview of principle and enhancements by massive MIMO. *IEEE Internet of Things Journal* 9, 384–400. doi:10.1109/JIOT.2021.3107242

Du, J., Yu, F. R., Lu, G., Wang, J., Jiang, J., and Chu, X. (2020). MEC-assisted immersive VR video streaming over Terahertz wireless networks: A deep reinforcement learning approach. *IEEE Internet of Things Journal* 7, 9517–9529. doi:10.1109/JIOT.2020.3003449

Earnshaw, R. A. (1993). *Virtual reality systems* (San Diego, CA: Academic press)

El Rassi, I. and El Rassi, J.-M. (2020). A review of haptic feedback in tele-operated robotic surgery. *Journal of Medical Engineering & Technology* 44, 247–254. doi:10.1080/03091902.2020.1772391

El Saddik, A. (2018). Digital twins: The convergence of multimedia technologies. *IEEE Multimedia* 25, 87–92. doi:10.1109/MMUL.2018.023121167

Elayan, H., Amin, O., Shihada, B., Shubair, R. M., and Alouini, M.-S. (2019). Terahertz band: The last piece of RF spectrum puzzle for communication systems. *IEEE Open Journal of the Communications Society* 1, 1–32. doi:10.1109/OJCOMS.2019.2953633

Elbamby, M. S., Perfecto, C., Bennis, M., and Doppler, K. (2018a). Edge computing meets millimeter-wave enabled VR: Paving the way to cutting the cord. In *IEEE Wireless Communications and Networking Conference* (IEEE), 1–6

Elbamby, M. S., Perfecto, C., Bennis, M., and Doppler, K. (2018b). Toward low-latency and ultra-reliable virtual reality. *IEEE Network* 32, 78–84. doi:10.1109/MNET.2018.1700268

Essaili, A. E., Thorson, S., Jude, A., Ewert, J. C., Tyudina, N., Caltenco, H., et al. (2022). Holographic communication in 5G networks. *Ericsson Technology Review.*

Available online at: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/holographic-communication-in-5g-networks [accessed August 25, 2022]

Favalora, G. (2005). Volumetric 3D displays and application infrastructure. *Computer* 38, 37–44. doi:10.1109/MC.2005.276

Feng, X., Swaminathan, V., and Wei, S. (2019). Viewport prediction for live 360-degree mobile video streaming using user-content hybrid motion tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1–22. doi:10.1145/3328914

Feth, D., Tran, B. A., Groten, R., Peer, A., and Buss, M. (2009). Shared-control paradigms in multi-operator-single-robot teleoperation. In *Human centered robot systems: Cognition, interaction, technology*, eds. H. Ritter, G. Sagerer, R. Dillmann, and M. Buss (Berlin, Heidelberg: Springer). 53–62

Fettweis, G., Boche, H., Wiegand, T., Zielinski, E., Schotten, H., Merz, P., et al. (2014). The tactile Internet-ITU-T technology watch report. Available online at: https://www.itu.int/dms_pub/itu-t/opb/gen/T-GEN-TWATCH-2014-1-PDF-E.pdf [accessed July 20, 2022]

FG-NET2030 (2020). Representative use cases and key network requirements for network 2030. Available online at: http://handle.itu.int/11.1002/pub/815125f5-en [accessed June 16, 2022]

Fishel, J. A. and Loeb, G. E. (2012). Bayesian exploration for intelligent identification of textures. *Frontiers in Neurorobotics* 6. doi:10.3389/fnbot.2012.00004

Fitzek, F. H., Li, S.-C., Speidel, S., Strufe, T., Simsek, M., and Reisslein, M. (2021). *Tactile Internet: With human-in-the-loop* (London: Academic Press)

Fowler, M. S., Halprin, A., and Schlichting, J. D. (1986). Back to the future: A model for telecommunications. *Federal Communications Law Journal* 38, 145–200

Fratz, M., Seyler, T., Bertz, A., and Carl, D. (2021). Digital holography in production: An overview. *Light: Advanced Manufacturing* 2, 283–295. doi:10.37188/lam.2021.015

Freina, L. and Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. In *The International Scientific Conference eLearning and Software for Education* (Carol I National Defence University Publishing House), 133–141

Fujimoto, T., Ishibashi, Y., and Sugawara, S. (2008). Influences of inter-stream synchronization error on collaborative work in haptic and visual environments. In *IEEE Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (IEEE), 113–119

Gabor, D. (1972). Holography, 1948-1971. *Science* 177, 299–313. doi:10.1126/science.177.4046.299

Gao, J., Zhuang, W., Li, M., Shen, X., and Li, X. (2021). MAC for machine-type communications in industrial IoT—Part I: Protocol design and analysis. *IEEE Internet of Things Journal* 8, 9945–9957. doi:10.1109/JIOT.2021.3051181

Gately, M., Zhai, Y., Yeary, M., Petrich, E., and Sawalha, L. (2011). A three-dimensional swept volume display based on LED arrays. *Journal of Display Technology* 7, 503–514. doi:10.1109/JDT.2011.2157455

Ge, X., Zhou, R., and Li, Q. (2019). 5G NFV-based tactile Internet for mission-critical IoT services. *IEEE Internet of Things Journal* 7, 6150–6163. doi:10.1109/JIOT.2019.2958063

Gibney, E. (2019). Star wars-style 3D images created from single speck of foam. *Nature* 575, 272–274. doi:10.1038/d41586-019-03454-y

Glushakov, M., Zhang, Y., Han, Y., Scargill, T. J., Lan, G., and Gorlatova, M. (2020). Invited paper: Edge-based provisioning of holographic content for contextual and personalized augmented reality. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops* (IEEE), 1–6

Gokhale, V., Kroep, K., Rao, V. S., Verburg, J., and Yechangunja, R. (2020). Tixt: An extensible testbed for tactile Internet communication. *IEEE Internet of Things Magazine* 3, 32–37. doi:10.1109/IOTM.0001.1900075

Gooday, G. J. (2005). Electrical futures past. *Endeavour* 29, 150–155. doi:10.1016/j.endeavour.2005.07.007

Gotsch, D., Zhang, X., Merritt, T., and Vertegaal, R. (2018). Telehuman2: A cylindrical light field teleconferencing system for life-size 3D human telepresence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (ACM), 1–10

Gu, Z., Lu, H., Hong, P., and Zhang, Y. (2022). Reliability enhancement for VR delivery in mobile-edge empowered dual-connectivity sub-6 GHz and mmWave HetNets. *IEEE Transactions on Wireless Communications* 21, 2210–2226. doi:10.1109/TWC.2021.3110099

Gupta, R., Reebadiya, D., and Tanwar, S. (2021). 6G-enabled edge intelligence for ultra-reliable low latency applications: Vision and mission. *Computer Standards & Interfaces* 77, 103521. doi:10.1016/j.csi.2021.103521

Gupta, R., Tanwar, S., Tyagi, S., and Kumar, N. (2019). Tactile-Internet-based telesurgery system for healthcare 4.0: An architecture, research challenges, and future directions. *IEEE Network* 33, 22–29. doi:10.1109/MNET.001.1900063

Guttentag, D. A. (2010). Virtual reality: Applications and implications for tourism. *Tourism management* 31, 637–651. doi:10.1016/j.tourman.2009.07.003

Han, Y., Niyato, D., Leung, C., Kim, D. I., Zhu, K., Feng, S., et al. (2022). A dynamic hierarchical framework for IoT-assisted digital twin synchronization in the Metaverse. *IEEE Internet of Things Journal* to be published, 1–18. doi:10.1109/JIOT.2022.3201082

Harvey, C., Selmanović, E., O'Connor, J., and Chahin, M. (2021). A comparison between expert and beginner learning for motor skill development in a virtual reality serious game. *The Visual Computer* 37, 3–17. doi:10.1007/s00371-019-01702-w

Hashimoto, T. and Ishibashi, Y. (2006). Group synchronization control over haptic media in a networked real-time game with collaborative work. In *Proceedings of 5th ACM SIGCOMM workshop on Network and System Support for Games* (ACM), 8–es

He, J., Qureshi, M. A., Qiu, L., Li, J., Li, F., and Han, L. (2018). Rubiks: Practical 360-degree streaming for smartphones. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (ACM), 482–494

He, L., Liu, K., He, Z., and Cao, L. (2023). Three-dimensional holographic communication system for the metaverse. *Optics Communications* 526, 128894. doi:10.1016/j.optcom.2022.128894

Hirayama, R., Martinez Plasencia, D., Masuda, N., and Subramanian, S. (2019). A volumetric display for visual, tactile and audio presentation using acoustic trapping. *Nature* 575, 320–323. doi:10.1038/s41586-019-1739-5

Hoggan, E., Anwar, S., and Brewster, S. A. (2007). Mobile multi-actuator tactile displays. In *Proceedings of the International Workshop on Haptic and Audio Interaction Design* (Springer), 22–33

Holland, O., Steinbach, E., Prasad, R. V., Liu, Q., Dawy, Z., Aijaz, A., et al. (2019). The IEEE 1918.1 "tactile Internet" standards working group and its standards. *Proceedings of the IEEE* 107, 256–279. doi:10.1109/JPROC.2018.2885541

Hou, X., Dey, S., Zhang, J., and Budagavi, M. (2018). Predictive view generation to enable mobile 360-degree and VR experiences. In *Proceedings of the 2018 Morning Workshop on Virtual Reality and Augmented Reality Network* (ACM), 20–26

Hou, X., Dey, S., Zhang, J., and Budagavi, M. (2021). Predictive adaptive streaming to enable mobile 360-degree and VR experiences. *IEEE Transactions on Multimedia* 23, 716–731. doi:10.1109/TMM.2020.2987693

Hou, Z., She, C., Li, Y., Zhuo, L., and Vucetic, B. (2019). Prediction and communication co-design for ultra-reliable and low-latency communications. *IEEE Transactions on Wireless Communications* 19, 1196–1209. doi:10.1109/ICC.2019.8762045

Hu, F., Deng, Y., and Aghvami, A. H. (2022). Cooperative multigroup broadcast 360° video delivery network: A hierarchical federated deep reinforcement learning approach. *IEEE Transactions on Wireless Communications* 21, 4009–4024. doi:10.1109/TWC.2021.3126147

Hu, F., Deng, Y., Saad, W., Bennis, M., and Aghvami, A. H. (2020). Cellular-connected wireless virtual reality: Requirements, challenges, and solutions. *IEEE Communications Magazine* 58, 105–111. doi:10.1109/MCOM.001.1900511

Hu, F., Deng, Y., Zhou, H., Jung, T. H., Chae, C.-B., and Aghvami, A. H. (2021). A vision of an XR-aided teleoperation system toward 5G/B5G. *IEEE Communications Magazine* 59, 34–40. doi:10.1109/MCOM.001.2000581

Hu, L., Tian, Y., Yang, J., Taleb, T., Xiang, L., and Hao, Y. (2019). Ready player one: UAV-clustering-based multi-task offloading for vehicular VR/AR gaming. *IEEE Network* 33, 42–48. doi:10.1109/MNET.2019.1800357

Hu, P., Dhelim, S., Ning, H., and Qiu, T. (2017). Survey on fog computing: Architecture, key technologies, applications and open issues. *Journal of network and computer applications* 98, 27–42. doi:10.1016/j.jnca.2017.09.002

Huawei Technologies Co., L. (2016). Whitepaper on the bearer network VR-oriented requirement Available online at: https://www.huawei.com/en/technology-insights/industry-insights/technology/white-papers/whitepaper-on-the-vr-oriented-bearer-network-requirement [accessed August 10, 2022]

Huzaifa, M., Desai, R., Grayson, S., Jiang, X., Jing, Y., Lee, J., et al. (2022). ILLIXR: An open testbed to enable extended reality systems research. *IEEE Micro* 42, 97–106. doi:10.1109/MM.2022.3161018

Javidi, B., Ferraro, P., Hong, S.-H., Nicola, S. D., Finizio, A., Alfieri, D., et al. (2005). Three-dimensional image fusion by use of multiwavelength digital holography. *Optics Letters* 30, 144–146. doi:10.1364/OL.30.000144

Jay, C., Glencross, M., and Hubbold, R. (2007). Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment. *ACM Transactions on Computer-Human Interaction* 14, 8–es. doi:10.1145/1275511.1275514

Jayasankar, U., Thirumal, V., and Ponnurangam, D. (2021). A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications. *Journal of King Saud University-Computer and Information Sciences* 33, 119–140. doi:10.1016/j.jksuci.2018.05.006

Ji, H., Park, S., Yeo, J., Kim, Y., Lee, J., and Shim, B. (2018). Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects. *IEEE Wireless Communications* 25, 124–130. doi:10.1109/MWC.2018.1700294

Jia, M. and Liang, W. (2018). Delay-sensitive multiplayer augmented reality game planning in mobile edge computing. In *Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (ACM), 147–154

Jiang, W., Han, B., Habibi, M. A., and Schotten, H. D. (2021). The road towards 6G: A comprehensive survey. *IEEE Open Journal of the Communications Society* 2, 334–366. doi:10.1109/OJCOMS.2021.3057679

Jones, A., McDowall, I., Yamada, H., Bolas, M., and Debevec, P. (2007). Rendering for an interactive 360° light field display. In *Proceedings of the International Conference on Computer Graphics and Interactive Techniques* (ACM), 40–es

Jumreornvong, O., Yang, E., Race, J., and Appel, J. (2020). Telemedicine and medical education in the age of COVID-19. *Academic Medicine* 95, 1838–1843. doi:10.1097/ACM.0000000000003711

Jung, T., tom Dieck, M. C., and Rauschnabel, P. A. (2020). *Augmented reality and virtual reality: Changing realities in a dynamic world* (Switzerland: Springer)

Kaluschke, M., Yin, M. S., Haddawy, P., Srimaneekarn, N., Saikaew, P., and Zachmann, G. (2021). A shared haptic virtual environment for dental surgical skill training. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops* (IEEE), 347–352

Kantonen, T., Woodward, C., and Katz, N. (2010). Mixed reality in virtual world teleconferencing. In *IEEE Virtual Reality Conference* (ACM), 179–182

Kavanagh, S., Luxton-Reilly, A., Wuensche, B., and Plimmer, B. (2017). A systematic review of virtual reality in education. *Themes in Science and Technology Education* 10, 85–119

Kerrigan, S. (2018). 13 hologram projections around the world that are more than just a pretty display. Available online at: https://interestingengineering.com/innovation/13-hologram-projections-around-the-world-that-are-more-than-just-a-pretty-display [accessed June 20, 2022]

Kim, H., Yang, J., Choi, M., Lee, J., Yoon, S., Kim, Y., et al. (2018). Immersive 360° VR tiled streaming system for esports service. In *Proceedings of the 9th ACM Multimedia Systems Conference* (ACM), 541–544

Kougioumtzidis, G., Poulkov, V., Zaharis, Z., and Lazaridis, P. (2022). QoE assessment aspects for virtual reality and holographic telepresence applications. In *International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures* (Springer), 171–180

Kugler, L. (2021). The state of virtual reality hardware. *Communications of the ACM* 64, 15–16. doi:kugler2021state

Kumagai, K., Miura, S., and Hayasaki, Y. (2021). Colour volumetric display based on holographic-laser-excited graphics using drawing space separation. *Scientific Reports* 11, 1–9. doi:10.1038/s41598-021-02107-3

Kumar, A., Kumar, S., Kaushik, A., Kumar, A., and Saini, J. (2020). Real time estimation and suppression of hand tremor for surgical robotic applications. *Microsystem Technologies* 28, 305–311. doi:10.1007/s00542-019-04736-1

Kumar, A. and Muhammad, B. (2018). Multi-sensory framework for holistic communications. In *IEEE International Symposium on Wireless Personal Multimedia Communications* (IEEE), 399–404

Kurbatova, E., Cheremkhin, P., Evtikhiev, N., Krasnov, V., and Starikov, S. (2015). Methods of compression of digital holograms. *Physics Procedia* 73, 328–332. doi:10.1016/j.phpro.2015.09.150

Laamarti, F., Eid, M., and El Saddik, A. (2014). An overview of serious games. *International Journal of Computer Games Technology* 2014, 1–15. doi:10.1155/2014/358152

Lederman, S. J. and Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics* 71, 1439–1459. doi:10.3758/APP.71.7.1439

Lee, J., Zhang, X., Park, C. H., and Kim, M. J. (2021). Real-time teleoperation of magnetic force-driven microrobots with 3D haptic force feedback for micro-navigation and micro-transportation. *IEEE Robotics and Automation Letters* 6, 1769–1776. doi:10.1109/LRA.2021.3060708

Lertlakkhanakul, J., Choi, J. W., and Kim, M. Y. (2008). Building data model and simulation platform for spatial interaction management in smart home. *Automation in Construction* 17, 948–957. doi:10.1016/j.autcon.2008.03.004

Lesniak, K. and Tucker, C. S. (2018). Dynamic rendering of remote indoor environments using real-time point cloud data. *Journal of Computing and Information Science in Engineering* 18, 1–11. doi:10.1115/1.4039472

Li, M., Gao, J., Zhou, C., Shen, X., and Zhuang, W. (2021a). Slicing-based artificial intelligence service provisioning on the network edge: Balancing AI service performance and resource consumption of data management. *IEEE Vehicular Technology Magazine* 16, 16–26. doi:10.1109/MVT.2021.3114655

Li, X., Yuan, Z., Zhao, J., Du, B., Liao, X., and Humar, I. (2021b). Edge-learning-enabled realistic touch and stable communication for remote haptic display. *IEEE Network* 35, 141–147. doi:10.1109/MNET.011.2000255

Liu, J., Shahroudy, A., Xu, D., and Wang, G. (2016a). Spatio-temporal LSTM with trust gates for 3D human action recognition. In *European conference on computer vision* (Springer), 816–833

Liu, Q., Huang, S., Opadere, J., and Han, T. (2018a). An edge network orchestrator for mobile augmented reality. In *IEEE Conference on Computer Communications* (IEEE), 756–764

Liu, S., Li, M., Xu, X., Steinbach, E., and Liu, Q. (2018b). QoE-driven uplink scheduling for haptic communications over 5G enabled tactile Internet. In *Proceedings of the IEEE International Symposium on Haptic, Audio and Visual Environments and Games* (IEEE), 1–5

Liu, X. and Deng, Y. (2021). Learning-based prediction, rendering and association optimization for MEC-enabled wireless virtual reality (VR) networks. *IEEE Transactions on Wireless Communications* 20, 6356–6370. doi:10.1109/TWC.2021.3073623

Liu, X., Dohler, M., Mahmoodi, T., and Liu, H. (2017). Challenges and opportunities for designing tactile codecs from audio codecs. In *Proceedings of the European Conference on Networks and Communications* (IEEE), 1–5

Liu, X., Kang, S., Plishker, W., Zaki, G., Kane, T. D., and Shekhar, R. (2016b). Laparoscopic stereoscopic augmented reality: toward a clinically viable electromagnetic tracking solution. *Journal of Medical Imaging* 3, 045001. doi:10.1117/1.JMI.3.4.045001

Liu, Y., He, W., Wang, Y., and Yang, H. (2019a). Network-assisted neural adaptive naked-eye 3D video streaming over wireless networks. *IEEE Access* 7, 141363–141373. doi:10.1109/ACCESS.2019.2944437

Liu, Y., Liu, J., Argyriou, A., and Ci, S. (2019b). MEC-assisted panoramic VR video streaming over millimeter wave mobile networks. *IEEE Transactions on Multimedia* 21, 1302–1316. doi:10.1109/TMM.2018.2876044

Liu, Y., Zheng, H., Zhao, L., Liu, S., Yao, K., Li, D., et al. (2020). Electronic skin from high-throughput fabrication of intrinsically stretchable lead zirconate titanate elastomer. *Research* 2020. doi:10.34133/2020/1085417

Liu, Y.-J., Du, H., Niyato, D., Feng, G., Kang, J., and Xiong, Z. (2022). Slicing4Meta: An intelligent integration framework with multi-dimensional network resources for Metaverse-as-a-Service in Web 3.0. *ArXiv Preprint.* doi:10.48550/arXiv.2208.06081

Liu, Z., Li, Q., Chen, X., Wu, C., Ishihara, S., Li, J., et al. (2021). Point cloud video streaming: Challenges and solutions. *IEEE Network* 35, 202–209. doi:10.1109/MNET.101.2000364

Maddikunta, P. K. R., Pham, Q.-V., B, P., Deepa, N., Dev, K., Gadekallu, T. R., et al. (2022). Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration* 26, 100257. doi:10.1016/j.jii.2021.100257

Maier, M. and Ebrahimzadeh, A. (2019). Towards immersive tactile Internet experiences: Low-latency FiWi enhanced mobile networks with edge intelligence. *Journal of Optical Communications and Networking* 11, B10–B25. doi:10.1364/JOCN.11.000B10

Maier, M., Ebrahimzadeh, A., and Chowdhury, M. (2018). The tactile Internet: Automation or augmentation of the human? *IEEE Access* 6, 41607–41618. doi:10.1109/ACCESS.2018.2861768

Maimone, A. and Wang, J. (2020). Holographic optics for thin and lightweight virtual reality. *ACM Transactions on Graphics* 39, 1–14. doi:10.1145/3386569.3392416

Mangiante, S., Klas, G., Navon, A., GuanHua, Z., Ran, J., and Silva, M. D. (2017). VR is on the edge: How to deliver 360° videos in mobile networks. In *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network* (ACM), 30–35

Manolova, A., Tonchev, K., Poulkov, V., Dixir, S., and Lindgren, P. (2021). Context-aware holographic communication based on semantic knowledge extraction. *Wireless Personal Communications* 120, 2307–2319. doi:10.1007/s11277-021-08560-7

Mauve, M. (2000). Consistency in replicated continuous interactive media. In *Proceedings of ACM conference on Computer Supported cCooperative Work* (ACM), 181–190

McCloy, R. and Stone, R. (2001). Virtual reality in surgery. *British Medical Journal* 323, 912–915. doi:10.1136/bmj.323.7318.912

Mehrabi, A., Siekkinen, M., Kämäräinen, T., and yl¨ J¨¨ski, A. (2021). Multi-tier CloudVR: Leveraging edge computing in remote rendered virtual reality. *ACM Transactions on Multimedia Computing, Communications, and Applications* 17, 1–24. doi:10.1145/3429441

Mekuria, R., Blom, K., and Cesar, P. (2017). Design, implementation, and evaluation of a point cloud codec for tele-immersive video. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 828–842. doi:10.1109/TCSVT.2016.2543039

Messenger, J. L. (2018). Time-sensitive networking: An introduction. *IEEE Communications Standards Magazine* 2, 29–33. doi:10.1109/MCOMSTD.2018.1700047

Meyer, J., Wilm, T., Fiess, R., Schlebusch, T., Stork, W., and Kasneci, E. (2022). A holographic single-pixel stereo camera sensor for calibration-free eye-tracking in retinal projection augmented reality glasses. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (ACM), 1–7

Miloslavskaya, V. and Vucetic, B. (2020). Design of short polar codes for SCL decoding. *IEEE Transactions on Communications* 68, 6657–6668. doi:10.1109/TCOMM.2020.3014946

Mohan, A., Wara, U. U., Shaikh, M. T. A., Rahman, R. M., and Zaidi, Z. A. (2021). Telesurgery and robotics: an improved and efficient era. *Cureus* 13. doi:10.7759/cureus.14124

Mohan, N., Corneo, L., Zavodovski, A., Bayhan, S., Wong, W., and Kangasharju, J. (2020). Pruning edge research with latency shears. In *Proceedings of the 19th ACM Workshop on Hot Topics in Networks* (ACM), 182–189

Mondal, S., Ruan, L., Maier, M., Larrabeiti, D., Das, G., and Wong, E. (2020). Enabling remote human-to-machine applications with AI-enhanced servers over access networks. *IEEE Open Journal of the Communications Society* 1, 889–899. doi:10.1109/OJCOMS.2020.3009023

Montagud, M., Cesar, P., Boronat, F., and Jansen, J. (2018). *MediaSync: Handbook on multimedia synchronization* (Springer)

Nakamura, T., Yano, T., Watanabe, K., Ishii, Y., Ono, H., Tambata, I., et al. (2019). 360-degree transparent holographic screen display. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference* (ACM), 1–2

Nasrabadi, A. T., Samiei, A., and Prakash, R. (2020). Viewport prediction for 360° videos: A clustering approach. In *Proceedings of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video* (ACM), 34–39

Nasrallah, A., Thyagaturu, A. S., Alharbi, Z., Wang, C., Shao, X., Reisslein, M., et al. (2019). Ultra-Low Latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research. *IEEE Communications Surveys & Tutorials* 21, 88–145. doi:10.1109/COMST.2018.2869350

Next G Alliance (2022). 6G applications and use cases. Available online at: https://www.nextgalliance.org/white_papers/6g-applications-and-use-cases/ [accessed August 16, 2022]

Okamoto, S., Nagano, H., and Yamada, Y. (2012). Psychophysical dimensions of tactile perception of textures. *IEEE Transactions on Haptics* 6, 81–93. doi:10.1109/TOH.2012.32

Ornati, M. (2022). Fashion touch. Surface haptics in fashion E-commerce. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications* (Springer Nature), 464–467

Ozioko, O., Karipoth, P., Hersh, M., and Dahiya, R. (2020). Wearable assistive tactile communication interface based on integrated touch sensors and actuators. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 1344–1352. doi:10.1109/TNSRE.2020.2986222

O'malley, M. K. and Gupta, A. (2008). Haptic interfaces. In *HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional Interfaces*, ed. P. Kortum (San Francisco, CA: Morgan Kaufmann Publishers). 25–64

Pacchierotti, C., Sinclair, S., Solazzi, M., Frisoli, A., Hayward, V., and Prattichizzo, D. (2017). Wearable haptic systems for the fingertip and the hand: taxonomy, review, and perspectives. *IEEE Transactions on Haptics* 10, 580–600. doi:10.1109/TOH.2017.2689006

Patel, R. V., Atashzar, S. F., and Tavakoli, M. (2022). Haptic feedback and force-based teleoperation in surgical robotics. *Proceedings of the IEEE* 110, 1012–1027. doi:10.1109/JPROC.2022.3180052

Pellas, N., Dengel, A., and Christopoulos, A. (2020). A scoping review of immersive virtual reality in stem education. *IEEE Transactions on Learning Technologies* 13, 748–761. doi:10.1109/TLT.2020.3019405

Petkov, N., Christoff, N., Manolova, A., Tonchev, K., and Poulkov, V. (2022). Comparative study of latent-sensitive processing of heterogeneous data in an experimental platform for 3D video holographic communication. In *Proceedings of the Global Conference on Wireless and Optical Technologies* (IEEE), 1–6

Polachan, K., Turkovic, B., Prabhakar, T., Singh, C., and Kuipers, F. A. (2020). Dynamic network slicing for the tactile Internet. In *Proceedings of the ACM/IEEE 11th International Conference on Cyber-Physical Systems* (IEEE), 129–140

Promwongsa, N., Ebrahimzadeh, A., Naboulsi, D., Kianpisheh, S., Belqasmi, F., Glitho, R., et al. (2020). A comprehensive survey of the tactile Internet: State-of-the-art and research directions. *IEEE Communications Surveys & Tutorials* 23, 472–523. doi:10.1109/COMST.2020.3025995

Qian, P., Huynh, V. S. H., Wang, N., Anmulwar, S., Mi, D., and Tafazolli, R. R. (2022). Remote production for live holographic teleportation applications in 5G networks. *IEEE Transactions on Broadcasting* 68, 451–463. doi:10.1109/TBC.2022.3161745

Qiao, X., Ren, P., Dustdar, S., and Chen, J. (2018). A new era for Web AR with mobile edge computing. *IEEE Internet Computing* 22, 46–55. doi:10.1109/MIC.2018.043051464

R. Yao, A. D. T. N., T. Heath and Hoberman, P. (2017). Oculus VR best practices guide. Available online at: http://developer.oculusvr.com/best-practices [accessed July 17, 2022]

Ren, J., He, Y., Huang, G., Yu, G., Cai, Y., and Zhang, Z. (2019). An edge-computing based architecture for mobile augmented reality. *IEEE Network* 33, 162–169. doi:10.1109/MNET.2018.1800132

Ren, P., Qiao, X., Huang, Y., Liu, L., Dustdar, S., and Chen, J. (2020). Edge-assisted distributed DNN collaborative computing approach for mobile Web augmented reality in 5G networks. *IEEE Network* 34, 254–261. doi:10.1109/MNET.011.1900305

Riva, G. (2000). From telehealth to E-health: Internet and distributed virtual reality in health care. *CyberPsychology & Behavior* 3, 989–998. doi:10.1089/109493100452255

Rost, P. M. and Kolding, T. (2022). Performance of integrated 3GPP 5G and IEEE TSN networks. *IEEE Communications Standards Magazine* 6, 51–56. doi:10.1109/MCOMSTD.0001.2000013

Rozen-Schiff, N., Navon, A., Bruckman, L., and Pechtalt, I. (2021). PRISM based transport: How networks can boost QoS for advanced video services? In *Proceedings of the Workshop on Design, Deployment, and Evaluation of Network-assisted Video Streaming* (ACM), 1–7

Ruan, J. and Xie, D. (2021). A survey on QoE-oriented VR video streaming: Some research issues and challenges. *Electronics* 10, 2155. doi:10.3390/electronics10172155

Sachs, J., Andersson, L. A., Araújo, J., Curescu, C., Lundsjö, J., Rune, G., et al. (2018). Adaptive 5G low-latency communication for tactile Internet services. *Proceedings of the IEEE* 107, 325–349. doi:0.1109/JPROC.2018.2864587

Sahin, E., Stoykova, E., Mäkinen, J., and Gotchev, A. (2021). Computer-generated holograms for 3D imaging: A survey. *ACM Computing Surveys* 53. doi:10.1145/3378444

Sakr, N., Georganas, N., Zhao, J., and Shen, X. (2007). Motion and force prediction in haptic media. In *Proceedings of IEEE International Conference on Multimedia and Expo* (IEEE), 2242–2245. doi:10.1109/ICME.2007.4285132

Schelkens, P., Ebrahimi, T., Gilles, A., Gioia, P., Oh, K.-J., Pereira, F., et al. (2019). JPEG Pleno: Providing representation interoperability for holographic applications and devices. *Electronics and Telecommunications Research Institute Journal* 41, 93–108. doi:10.4218/etrij.2018-0509

Schmitz, P., Blut, T., Mattes, C., and Kobbelt, L. (2020). High-fidelity point-based rendering of large-scale 3-D scan datasets. *IEEE Computer Graphics and Applications* 40, 19–31. doi:10.1109/MCG.2020.2974064

Schwarz, S., Preda, M., Baroncini, V., Budagavi, M., Cesar, P., Chou, P. A., et al. (2019). Emerging MPEG standards for point cloud compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 9, 133–148. doi:10.1109/JETCAS.2018.2885981

Selinis, I., Wang, N., Da, B., Yu, D., and Tafazolli, R. (2020). On the Internet-scale streaming of holographic-type content with assured user quality of experiences. In *IFIP Networking Conference* (IEEE), 136–144

Seufert, M., Schatz, R., Wehner, N., and Casas, P. (2019). QUICker or not? -an empirical analysis of QUIC vs TCP for video streaming QoE provisioning. In *Conference on Innovation in Clouds, Internet and Networks and Workshops* (IEEE), 7–12

Shahbazi, M., Atashzar, S. F., and Patel, R. V. (2018). A systematic review of multilateral teleoperation systems. *IEEE Transactions on Haptics* 11, 338–356. doi:10.1109/TOH.2018.2818134

Shahraki, A., Abbasi, M., Piran, M., Taherkordi, A., et al. (2021). A comprehensive survey on 6G networks: Applications, core services, enabling technologies, and future challenges doi:10.48550/ARXIV.2101.12475

She, C., Sun, C., Gu, Z., Li, Y., Yang, C., Poor, H. V., et al. (2020). A tutorial on ultra-reliable and low-latency communications in 6G: Integrating domain knowledge into deep learning. *ArXiv Preprint* doi:10.48550/arXiv.2009.06010

Shen, X., Gao, J., Wu, W., Li, M., Zhou, C., and Zhuang, W. (2021). Holistic network virtualization and pervasive network intelligence for 6G. *IEEE Communications Surveys & Tutorials* 24, 1–30. doi:10.1109/COMST.2021.3135829

Shen, X., Gao, J., Wu, W., Lyu, K., Li, M., Zhuang, W., et al. (2020). AI-assisted network-slicing based next-generation wireless networks. *IEEE Open Journal of Vehicular Technology* 1, 45–66. doi:10.1109/OJVT.2020.2965100

Shimobaba, T., Blinder, D., Birnbaum, T., Hoshi, I., Shiomi, H., Schelkens, P., et al. (2022). Deep-learning computational holography: A review (invited). *Frontiers in Photonics* 3. doi:10.3389/fphot.2022.854391

Sim, D., Baek, Y., Cho, M., Park, S., Sagar, A. S., and Kim, H. S. (2021). Low-latency haptic open glove for immersive virtual reality interaction. *Sensors* 21, 3682. doi:10.3390/s21113682

Simsek, M., Aijaz, A., Dohler, M., Sachs, J., and Fettweis, G. (2016). 5G-enabled tactile Internet. *IEEE Journal on Selected Areas in Communications* 34, 460–473. doi:10.1109/JSAC.2016.2525398

Siriwardhana, Y., Porambage, P., Liyanage, M., and Ylianttila, M. (2021). A survey on mobile augmented reality with 5G mobile edge computing: Architectures, applications, and technical aspects. *IEEE Communications Surveys & Tutorials* 23, 1160–1192. doi:10.1109/COMST.2021.3061981

Smalley, D., Nygaard, E., Squire, K., Van Wagoner, J., Rasmussen, J., Gneiting, S., et al. (2018). A photophoretic-trap volumetric display. *Nature* 553, 486–490. doi:10.1038/nature25176

Son, J., Jang, D., and Ryu, E.-S. (2018). Implementing motion-constrained tile and viewport extraction for VR streaming. In *Proceedings of the 28th ACM SIGMM Workshop on Network and Operating Systems Support for Digital Audio and Video* (ACM), 61–66

Song, P., Verinaz-Jadan, H., Howe, C. L., Foust, A. J., and Dragotti, P. L. (2022). Light-field microscopy for the optical imaging of neuronal activity: When model-based methods meet data-driven approaches. *IEEE Signal Processing Magazine* 39, 58–72. doi:10.1109/MSP.2021.3123557

Sparkes, M. (2021). What is a metaverse. *New Scientist* 251, 18. doi:https://doi.org/10.1016/S0262-4079(21)01450-0

Speicher, M., Cucerca, S., and Krüger, A. (2017). VRShop: A mobile interactive virtual reality shopping environment combining the benefits of on-and offline shopping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 1–31. doi:10.1145/3130967

Srinivasan, M. A. and Basdogan, C. (1997). Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers & Graphics* 21, 393–404

Stark, M., Pomati, S., D'Ambrosio, A., Giraudi, F., and Gidaro, S. (2015). A new telesurgical platform – preliminary clinical results. *Minimally Invasive Therapy & Allied Technologies* 24, 31–36. doi:10.3109/13645706.2014.1003945

Steinbach, E., Hirche, S., Ernst, M., Brandi, F., Chaudhari, R., Kammerl, J., et al. (2012). Haptic communications. *Proceedings of the IEEE* 100, 937–956. doi:10.1109/JPROC.2011.2182100

Steinbach, E., Hirche, S., Kammerl, J., Vittorias, I., and Chaudhari, R. (2010). Haptic data compression and communication. *IEEE Signal Processing Magazine* 28, 87–96. doi:10.1109/MSP.2010.938753

Steinbach, E., Strese, M., Eid, M., Liu, X., Bhardwaj, A., Liu, Q., et al. (2018). Haptic codecs for the tactile Internet. *Proceedings of the IEEE* 107, 447–470. doi:10.1109/JPROC.2018.2867835

Strinati, E. C., Barbarossa, S., Gonzalez-Jimenez, J. L., Ktenas, D., Cassiau, N., Maret, L., et al. (2019). 6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication. *IEEE Vehicular Technology Magazine* 14, 42–50. doi:10.1109/MVT.2019.2921162

Sukhmani, S., Sadeghi, M., Erol-Kantarci, M., and El Saddik, A. (2018). Edge caching and computing in 5G for mobile AR/VR and tactile Internet. *IEEE MultiMedia* 26, 21–30. doi:10.1109/MMUL.2018.2879591

Sun, L., Mao, Y., Zong, T., Liu, Y., and Wang, Y. (2020). Flocking-based live streaming of 360-degree video. In *Proceedings of the 11th ACM Multimedia Systems Conference* (ACM), 26–37

Sun, Y., Chen, Z., Tao, M., and Liu, H. (2019). Communications, caching, and computing for mobile virtual reality: Modeling and tradeoff. *IEEE Transactions on Communications* 67, 7573–7586. doi:10.1109/TCOMM.2019.2920594

Tahara, T., Quan, X., Otani, R., Takaki, Y., and Matoba, O. (2018). Digital holography and its multidimensional imaging applications: a review. *Microscopy* 67, 55–67. doi:10.1093/jmicro/dfy007

Takagi, A., Ganesh, G., Yoshioka, T., Kawato, M., and Burdet, E. (2017). Physically interacting individuals estimate the partner's goal to enhance their movements. *Nature Human Behaviour* 1, 1–6. doi:10.1038/s41562-017-0054

Taleb, T., Nadir, Z., Flinck, H., and Song, J. (2021). Extremely inter active and low-latency services in 5G and beyond mobile systems. *IEEE Communications Standards Magazine* 5, 114–119. doi:10.1109/MCOMSTD.001.2000053

Tan, H. Z., Choi, S., Lau, F. W. Y., and Abnousi, F. (2020). Methodology for maximizing information transmission of haptic devices: A survey. *Proceedings of the IEEE* 108, 945–965. doi:10.1109/JPROC.2020.2992561

Tanaka, H. and Ohnishi, K. (2009). Haptic data compression/decompression using dct for motion copy system. In *Proceedings of the IEEE International Conference on Mechatronics* (IEEE), 1–6

Tang, W., Dai, J. Y., Chen, M. Z., Wong, K.-K., Li, X., Zhao, X., et al. (2020). MIMO transmission through reconfigurable intelligent surface: System design, analysis, and implementation. *IEEE journal on selected areas in communications* 38, 2683–2699. doi:10.1109/JSAC.2020.3007055

Tarneberg, W., Karaca, M., Robertsson, A., Tufvesson, F., and Kihl, M. (2017). Utilizing massive MIMO for the tactile Internet: Advantages and trade-offs. In *Proceedings of the IEEE International Conference on Sensing, Communication and Networking* (IEEE), 1–6

Tasaka, S. (2022). An empirical method for causal inference of constructs for QoE in haptic–audiovisual communications. *ACM Transactions on Multimedia Computing, Communications, and Applications* 18, 1–24. doi:10.1145/3473986

Tataria, H., Shafi, M., Molisch, A. F., Dohler, M., Sjöland, H., and Tufvesson, F. (2021). 6G wireless systems: Vision, requirements, challenges, insights, and opportunities. *Proceedings of the IEEE* 109, 1166–1199. doi:10.1109/JPROC.2021.3061701

Thanh, N. T., Jiang, X., Abiko, S., Tsujita, T., Konno, A., and Uchiyama, M. (2012). Collaborative haptic interaction in virtual environment of multi-operator multi-robot teleoperation systems. In *Proceedings of SICE Annual Conference* (IEEE), 1585–1590

Van Den Berg, D., Glans, R., De Koning, D., Kuipers, F. A., Lugtenburg, J., Polachan, K., et al. (2017). Challenges in haptic communications over the tactile Internet. *IEEE Access* 5, 23502–23518. doi:10.1109/ACCESS.2017.2764181

Wang, D., Ohnishi, K., and Xu, W. (2019). Multimodal haptic display for virtual reality: A survey. *IEEE Transactions on Industrial Electronics* 67, 610–623. doi:10.1109/TIE.2019.2920602

Wang, K., Jin, J., Yang, Y., Zhang, T., Nallanathan, A., Tellambura, C., et al. (2022a). Task offloading with multi-tier computing resources in next generation wireless networks. *ArXiv Preprint.* doi:10.48550/arXiv.2205.13866

Wang, X., Liang, C.-J., Menassa, C. C., and Kamat, V. R. (2021). Interactive and immersive process-level digital twin for collaborative human–robot construction work. *Journal of Computing in Civil Engineering* 35, 04021023. doi:10.1061/(ASCE)CP.1943-5487.0000988

Wang, Y. and Li, Z. (2022). Living with smell dysfunction: A multi-sensory VR experience. In *ACM SIGGRAPH 2022 Immersive Pavilion* (ACM), 1–2

Wang, Y., Su, Z., Zhang, N., Xing, R., Liu, D., Luan, T. H., et al. (2022b). A survey on Metaverse: Fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials* to be published, 1–32

Wei, X., Shi, Y., and Zhou, L. (2021). Haptic signal reconstruction for cross-modal communications. *IEEE Transactions on Multimedia* to be publised, 1–12. doi:10.1109/TMM.2021.3119860

Wei, X., Yao, Y., Wang, H., and Zhou, L. (2022). Perception-aware cross-modal signal reconstruction: From audio-haptic to visual. *IEEE Transactions on Multimedia* to be published, 1–12. doi:10.1109/TMM.2022.3194309

Wenzlhuemer, R. (2013). *Connecting the nineteenth-century world: The telegraph and globalization* (Cambridge: Cambridge University Press)

Wu, W., Zhou, C., Li, M., Wu, H., Zhou, H., Zhang, N., et al. (2022). AI-native network slicing for 6G networks. *IEEE Wireless Communications* 29, 96–103. doi:10.1109/MWC.001.2100338

Xiang, W., Wang, G., Pickering, M., and Zhang, Y. (2016). Big video data for light-field-based 3D telemedicine. *IEEE Network* 30, 30–38. doi:10.1109/MNET.2016.7474341

Xie, C., Li, X., Hu, Y., Peng, H., Taylor, M., and Song, S. L. (2021). Q-VR: System-level design for future mobile collaborative virtual reality. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems* (ACM), 587–599

Xiong, J., Hsiang, E.-L., He, Z., Zhan, T., and Wu, S.-T. (2021). Augmented reality and virtual reality displays: emerging technologies and future perspectives. *Light: Science & Applications* 10, 1–30. doi:10.1038/s41377-021-00658-8

Xu, M., Ng, W. C., Lim, W. Y. B., Kang, J., Xiong, Z., Niyato, D., et al. (2022). A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges. *ArXiv Preprint* doi:10.48550/ARXIV.2203.05471

Xu, W., Chatterjee, A., Zollhöfer, M., Rhodin, H., Mehta, D., Seidel, H.-P., et al. (2018a). Monoperfcap: Human performance capture from monocular video. *ACM Transactions on Graphics* 37, 1–15. doi:10.1145/3181973

Xu, X., Cizmeci, B., Schuwerk, C., and Steinbach, E. (2015). Haptic data reduction for time-delayed teleoperation using the time domain passivity approach. In *Proceedings of IEEE World Haptics Conference* (IEEE), 512–518

Xu, Y., Dong, Y., Wu, J., Sun, Z., Shi, Z., Yu, J., et al. (2018b). Gaze prediction in dynamic 360° immersive videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE), 5333–5342

Yadav, P. K. and Ooi, W. T. (2020). Tile rate allocation for 360-degree tiled adaptive video streaming. In *Proceedings of the 28th ACM International Conference on Multimedia* (ACM), 3724–3733

Yang, L., Dong, H., Alelaiwi, A., and Saddik, A. E. (2016). See in 3D: state of the art of 3D display technologies. *Multimedia Tools and Applications* 75, 17121–17155. doi:10.1007/s11042-015-2981-y

Yang, L., Wu, D., and Zhou, L. (2021). Heterogeneous stream scheduling for cross-modal transmission. *IEEE Transactions on Communications* 69, 6037–6049. doi:10.1109/TCOMM.2021.3086522

Yang, P., Quek, T. Q. S., Chen, J., You, C., and Cao, X. (2022a). Feeling of presence maximization: mmWave-enabled virtual reality meets deep reinforcement learning. *IEEE Transactions on Wireless Communications* to be published, 1–15. doi:10.1109/TWC.2022.3181674

Yang, Y. (2019). Multi-tier computing networks for intelligent IoT. *Nature Electronics* 2, 4–5. doi:10.1038/s41928-018-0195-9

Yang, Y., Ma, M., Wu, H., Yu, Q., Zhang, P., You, X., et al. (2022b). 6G network AI architecture for everyone-centric customized services. *IEEE Network* to be published. doi:10.1109/MNET.124.2200241

Yang, Y., Wang, K., Zhang, G., Chen, X., Luo, X., and Zhou, M.-T. (2018). MEETS: Maximal energy efficient task scheduling in homogeneous fog networks. *IEEE Internet of Things Journal* 5, 4076–4087. doi:10.1109/JIOT.2018.2846644

Yen, S.-C., Fan, C.-L., and Hsu, C.-H. (2019). Streaming 360°videos to head-mounted virtual reality using DASH over QUIC transport protocol. In *Proceedings of the 24th ACM Workshop on Packet Video* (ACM), 7–12

Yokokohji, Y., Hollis, R. L., and Kanade, T. (1996a). What you can see is what you can feel-development of a visual/haptic interface to virtual environment. In *IEEE 1996 Virtual Reality Annual International Symposium* (IEEE), 46–53

Yokokohji, Y., Hollis, R. L., Kanade, T., Henmi, K., and Yoshikawa, T. (1996b). Toward machine mediated training of motor skills. Skill transfer from human to human via virtual environment. In *IEEE International Workshop on Robot and Human Communication* (IEEE), 32–37

You, Y. and Sung, M. Y. (2008). Haptic data transmission based on the prediction and compression. In *Proceedings of the IEEE International Conference on Communications* (IEEE), 1824–1828

Yu, B., Cai, Y., Zou, Y., Li, B., and Chen, Y. (2022). Can we improve the information freshness with prediction for cognitive IoT? *IEEE Internet of Things Journal* 9, 17577–17591. doi:10.1109/JIOT.2022.3155717

Yu, X., Xie, Z., Yu, Y., Lee, J., Vazquez-Guardado, A., Luan, H., et al. (2019). Skin-integrated wireless haptic interfaces for virtual and augmented reality. *Nature* 575, 473–479. doi:10.1038/s41586-019-1687-0

Yuan, Z., Kang, B., Wei, X., and Zhou, L. (2022). Exploring the benefits of cross-modal coding. *IEEE Transactions on Circuits and Systems for Video Technology* doi:10.1109/TCSVT.2022.3196586

Zare, A., Aminlou, A., Hannuksela, M. M., and Gabbouj, M. (2016). HEVC-compliant tile-based streaming of panoramic video for virtual reality applications. In *Proceedings of the 24th ACM International Conference on Multimedia* (ACM), 601–605

Zawish, M., Dharejo, F. A., Khowaja, S. A., Dev, K., Davy, S., Qureshi, N. M. F., et al. (2022). AI and 6G into the metaverse: Fundamentals, challenges and future research trends. *ArXiv Preprint*. doi:10.48550/ARXIV.2208.10921

Zeng, C., Zhao, T., Liu, Q., Xu, Y., and Wang, K. (2020). Perception-lossless codec of haptic data with low delay. In *Proceedings of the 28th ACM International Conference on Multimedia* (ACM), 3642–3650

Zhang, Q., Liu, J., and Zhao, G. (2018). Towards 5G enabled tactile robotic telesurgery. *ArXiv Preprint*. doi:10.48550/arXiv.1803.03586

Zhang, W., Chen, J., Zhang, Y., and Raychaudhuri, D. (2017). Towards efficient edge cloud augmentation for virtual reality MMOGs. In *Proceedings of the Second ACM/IEEE Symposium on Edge Computing* (ACM), 1–14

Zhang, X., Sugano, Y., Fritz, M., and Bulling, A. (2019a). MPIIGaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 162–175. doi:10.1109/TPAMI.2017.2778103

Zhang, Y., Yang, J., Liu, Z., Wang, R., Chen, G., Tong, X., et al. (2022a). Virtualcube: An immersive 3D video communication system. *IEEE Transactions on Visualization and Computer Graphics* 28, 2146–2156. doi:10.1109/TVCG.2022.3150512

Zhang, Z., Wen, F., Sun, Z., Guo, X., He, T., and Lee, C. (2022b). Artificial intelligence-enabled sensing technologies in the 5G/Internet of Things era: From virtual reality/augmented reality to the digital twin. *Advanced Intelligent Systems* 4, 1–23. doi:10.1002/aisy.202100228

Zhang, Z., Xiao, Y., Ma, Z., Xiao, M., Ding, Z., Lei, X., et al. (2019b). 6G wireless networks: Vision, requirements, architecture, and key technologies. *IEEE Vehicular Technology Magazine* 14, 28–41. doi:10.1109/MVT.2019.2921208

Zhou, C., Gao, J., Li, M., Shen, X., and Zhuang, W. (2022a). Digital twin-empowered network planning for multi-tier computing. *Journal of Communications and Information Networks* 7, 221–238. doi:10.23919/JCIN.2022.9906937

Zhou, F., Li, W., Yang, Y., Feng, L., Yu, P., Zhao, M., et al. (2022b). Intelligence-endogenous networks: Innovative network paradigm for 6G. *IEEE Wireless Communications* 29, 40–47. doi:10.1109/MWC.004.00320

Zhu, M., Sun, Z., Zhang, Z., Shi, Q., He, T., Liu, H., et al. (2020). Haptic-feedback smart glove as a creative human-machine interface (HMI) for virtual/augmented reality applications. *Science Advances* 6, eaaz8693. doi:10.1126/sciadv.aaz8693