



Context-Aware Holographic Communication Based on Semantic Knowledge Extraction

Agata Manolova¹ · Krasimir Tonchev¹ · Vladimir Poulkov¹ · Sudhir Dixit² · Peter Lindgren³

Accepted: 4 May 2021 / Published online: 3 June 2021
© The Author(s) 2021

Abstract

Augmented, mixed and virtual reality are changing the way people interact and communicate. Five dimensional communications and services, integrating information from all human senses are expected to emerge, together with holographic communications (HC), providing a truly immersive experience. HC presents a lot of challenges in terms of data gathering and transmission, demanding Artificial Intelligence empowered communication technologies such as 5G. The goal of the paper is to present a model of a context-aware holographic architecture for real time communication based on semantic knowledge extraction. This architecture will require analyzing, combining and developing methods and algorithms for: 3D human body model acquisition; semantic knowledge extraction with deep neural networks to predict human behaviour; analysis of biometric modalities; context-aware optimization of network resource allocation for the purpose of creating a multi-party, from-capturing-to-rendering HC framework. We illustrate its practical deployment in a scenario that can open new opportunities in user experience and business model innovation.

Keywords Holographic communication · Semantic knowledge · Deep neural networks · Business model innovation

✉ Agata Manolova
amanolova@tu-sofia.bg

Krasimir Tonchev
k_tonchev@tu-sofia.bg

Vladimir Poulkov
vkp@tu-sofia.bg

Sudhir Dixit
sudhir.dixit@ieee.org

Peter Lindgren
peterli@btech.au.dk

¹ Technical University of Sofia, Sofia, Bulgaria

² University of Oulu, Oulu, Finland

³ Aarhus University, Århus, Denmark

1 Introduction

The current methods of remote interaction between human beings are becoming obsolete, as new forms of interaction are being developed leading to a true immersion into a distant or virtual environment. Five dimensional (5D) communications and services, integrating all human senses information (sight, hearing, touch, smell and taste), are expected to emerge, together with holographic communication, thus providing a truly immersive experience. This technology opens the door to new heights of interpersonal communication, but it also presents a lot of challenges in terms of digital data gathering and transmission. Holographic communication using multiple view cameras will require transmission rates in Tbps. Even with broadband channels, streaming big amounts of data will lead to a latency which will influence the natural perception of the communication.

The 5G technology inherently brings aspects such as high-speed, ultra-low latency and high bandwidth - all in wireless communication networks. These networks are architected to support service categories such as: Enhanced mobile broadband (eMBB) and Ultra-reliable low latency communication (URLLC). Enabling true ubiquitous connectivity to remote computing power is one of the essential selling points of 5G. So for the holographic communication to realize its full potential it must adopt the significant opportunities offered by 5G [1].

For the effective deployment of a holographic communication architecture, Artificial Intelligence (AI) algorithms running under severe delay constraints will be needed. They are expected to play a key role in various aspects of the communication process such as: self-optimization of network resource allocation, possibly adopting proactive strategies based on network learning and prediction; development of applications, that learn from user's behaviour and act as a context-aware virtual intelligent assistant; development of semantic inference algorithms and semantic communication strategies to incorporate knowledge representation in communication strategies.

Augmented reality (AR), mixed reality (MR) and virtual reality (VR) technologies will facilitate and make communication more effective for business, teachers, technologists, people of different professions anywhere in the world, with all the benefits of physical presence, but without any geographical restrictions. Holographic communication is a human-computer-machine interface combining AR and VR presence, creating realistic 3D models of the human body, rendered in MR environments, used for real time communication and interaction between remote users [2]. It can include all five senses, as well as objects and elements of the environment that are identical to those in the remote location. One of the major problems with holographic communication is how to overcome the limitations of the communication channel when huge amounts of data generated by the capture process need to be transmitted in real time and with negligible latency. Even with ultra-high-speed communication access and backbones, transmitting large amounts of data at a long distance will be subject to latency, which will affect the user's natural perception. 5G and the future 6G systems will be able to deliver a full immersive MR experience that captures all senses based on their ability to provide very low MR latency for high-speed applications. Truly immersive technology requires a collaborative design that incorporates not only these communication technologies, big data and advanced signal processing requirements, but also human perceptions related to high-level semantic knowledge [3]. It depends upon advances in intelligent ways of presenting the data in order to design a holographic communication architecture to maintain a sense of reality and naturalness of the face-to-face interaction. There is an urgent need to extend the classical communication model to characterize not

only the sequences of bits but also the meaning behind these bits. Through the incorporation of semantic knowledge in communications we can extend the capacity of a communication channel beyond Shannon's limits. What we want to achieve is the correct transmission of human activity of the body and face, including voice data, based not only on their syntactic notions, but also on their meaning in a given context and thus to realize a context-aware holographic communication based on semantic knowledge extraction, i.e. to add an additional "semantic" capacity to the communication channel.

In this paper we present a context-aware holographic communication architecture based on semantic knowledge extraction, relying on highly accurate 3D modelling of the human face, body and clothes, recognition and prediction of human actions and facial expressions. Such architecture can empower 5G communications and address some of the challenges imposed by real time constraints and channel limits when transmitting huge and heterogeneous amounts of data. The described scenario is related to "holoportation" of humans and their interactions among a network of globally connected hexagonal closed rooms, called "Bee cubes", for the needs of Business Model Innovation (BMI) process.

The rest of the paper is organized as follows: the second section presents brief state of the art of holographic communication systems; the third section describes the key features/building blocks of the proposed holographic architecture. An insight into a use case scenario with application to BMI is given with its opportunities and challenges. The final section identifies the challenges for such an architecture, and concludes the paper with suggestions for future work.

2 Current State of the Art

The architecture of the holographic platform reflects its multidisciplinary character and consists of three steps: (1) capture and reconstruction, (2) data compression and transmission, and (3) reconstruction and visualization. The hardware components always include an integrated multi-sensor imaging system and displays for visualization. One way to implement a holographic system is through synthetic avatars, with the user's movement captured and transmitted in real time [4]. However, with such representations, one fundamental aspect is missing, the natural appearance of the user. Using the latest advances in 3D real time reconstruction, realistic deformable and parametric 3D models can be created to facilitate interaction and contain important information such as facial features, clothing movements and body postures. This leads to an increased level of realism, which makes the whole experience natural, thus increasing the consumer's experience and participation.

In order to generate a realistic representation of the user, several current methods for reconstruction from RGB cameras could be found in the literature, but they are not able to reconstruct detailed and realistic face and hands and often expensive photo studios are used for capturing. Unfortunately most of the methods are not applicable to our idea of processing and communication in real time because of their low productivity. A solution is possible by creating a platform of several RGB-d sensors to create three-dimensional images by triangulating depth maps captured by each sensor and generating colour information coded as a colour-per-vertex attribute and skin texture. In order to interact remotely and in real time, the holographic communication system should focus on 3D reconstruction and real time transmission, where full 3D geometry and appearance are generated for each participant, enabling a realistic experience. However, such systems require complex installation configurations unsuitable for rapid deployment, so

the focus of this paper is to use methods to implicitly compress data realized by detecting and sending only semantic information of the human body and the parameters of face, hands and voice instead of the entire 3D model of the body.

Especially important for the development of holographic communication system is the analysis of human behaviour. Therefore, it is essential to develop and integrate reliable and accurate modelling techniques for human behaviour that seek to learn and predict human behaviour based on semantic knowledge combined with deep architectures.

Capturing human behaviour through modelling techniques is extremely difficult because of the complex physiological, psychological and behavioural aspects of human beings. So our paper addresses some common challenges, such as user-specific metrics: facial characteristics, body skeleton and skeletal joints, assessment of changes in human behaviour over time, and the sensory system created to sense the relevant aspects of human behaviour. Sustainable holographic communication systems are likely to require predictive models to avoid network latency issues. Proper modelling and analysis of the results of these systems will require multimodal and multidimensional analyses. The result of the analysis will be useful for multisensory communications and in particular for semantic compression of data. Techniques that model a particular aspect of human behaviour are currently very application-specific, such as predicting Facebook usage [5], modelling the behaviour of occupants of buildings [6] or computer-based assessment of personality [7].

For holographic communication, latency is one of the biggest challenges that must be overcome. The system should achieve communication between users in no more than 16 ms, which would be possible by adding a step to predict human behaviour. People can experience time delays of approximately 16 ms or greater. 3D holographic communication will allow users to communicate remotely with realistic interactivity. The data requirement of 3D holograms is assumed to be at terabytes. Real time holographic transmission will require 10 Gbps or higher using current compression techniques [8].

In summary, semantic information can significantly improve the communication effectiveness of the holographic system by setting different priorities for different data on their semantics and using each form of shared knowledge to enable semantic based decision-making.

3 Architecture of a Context-Aware Holographic Communication System Based on Semantic Knowledge Extraction

The proposed conceptual system architecture for real time holographic communication between two identical closed environments is illustrated in Fig. 1. For simplicity, the architecture's pipeline is presented in two blocks: one for the offline stage and one for real time communication process where the two sides of the communication channel are presented.

When working in a controlled environment identical in both sending and receiving side, there is no need to send the complete information about the surrounding scene. Immutable objects such as walls, tables and even chairs can be considered static and not relevant to the communication process. Only objects possessing dynamical properties and information of the human interlocutors need to be transmitted between two or more closed environments. The computational load for extracting, processing, predicting and decision making are distributed between fixed backbone systems installed at the home and the remote site.

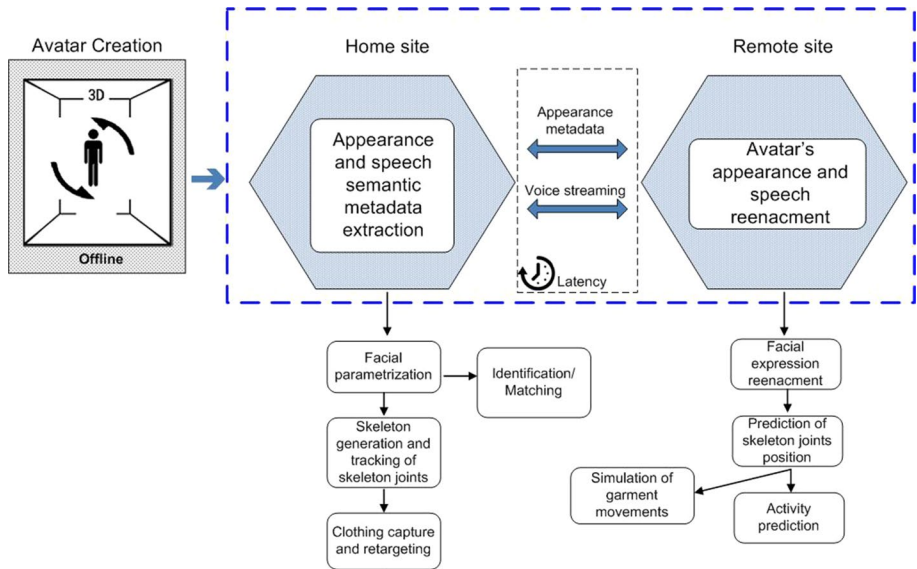


Fig. 1 Conceptual architecture for a context-aware holographic communication system based on semantic knowledge extraction

3.1 Modelling of Human Body and Face—Avatar Creation

One of the main tasks of the proposed architecture concerns the parametric modelling of the human body. In holographic communication, the human figure is a central element in video sequencing. Understanding its posture, hand movements and facial expressions used for non-verbal communication and interaction with the world is critical to the overall understanding of the communication process. However, to extract semantic knowledge of human behaviour, more than the basic body trays need to be captured - a full 3D surface of the body, hands and face is required, as well as the possibility of differentiation between the female and the male body. It is necessary to construct a sufficiently accurate model of an already existing complex object in order to be able to recreate its view from different perspectives in the most realistic way, which will also help for the purposes of recognition. Automatically constructing geometric models of 3D human body involves three basic steps: (1) collecting data, (2) capturing images from different views, and (3) integrating. Data collection involves obtaining brightness or depth information about the object from multiple perspectives. In many cases, complex transformations are required to obtain accurate geometric relationships in 3D space from 2D images. Thus, integrating data from multiple sensors is not only based on the description of the model from the individual views, but also requires knowledge of the transformations between the data from these sensors. The purpose of the registration is to find the transformations that link the data from the individual images and thus bring the shared regions into one aggregate model. The integration step integrates data from multiple views using the calculated transforms from each view to create a unique surface representation in a common coordinate system.

The problem of reconstructing the 3D geometry of a human face from a set of facial images in multiple views is a very up-to-date task related to the creation of realistic human models [9]. Given the drawbacks of single-view algorithms, we reconstruct the faces and

their facial features based on 3D deformable models, with a set of multi-view facial images given as input. We propose an approach for regressing parameters from 3D deformable models with multiple views with a convolutional neural network (CNN). Multi-view geometric constraints are considered when training the network by matching different views and balancing the view alignment error. By minimizing the loss of view alignment, 3D shapes can be better reconstructed so that the synthetic projection from one view to another can be better aligned with the observations. We used an approach for 3D hand finger positions in real time from RGB-d images using 3D convolutional neural networks. The approach uses a 3D volumetric representation of the hands, which can capture its 3D spatial structure. To further improve the accuracy of the assessment, we apply to the 3D deep network architecture, the overall surface of the hands as an intermediate learning element for learning 3D hand postures from deep images.

One of the challenges to create a realistic model of the human body includes simulating high quality movements of garments, a very important element for visually plausible presentation of the model. The highly realistic physical simulation of clothing on the human body in motion is complex: clothing is difficult to design; patterns must be scaled so that they can be sized for different attributes and the physical parameters of the fabric must be known. Current 3D clothing capture methods are accurate and detailed enough to compete with physical simulation [10]. The main issues that need to be addressed include high-quality imaging, segmentation, tracking of surface shape, as well as body shape and posture evaluation during real time movement.

3.2 Semantic Knowledge of Human Activity

Unlike the low level features, semantic describes inherent characteristics of human activity. Therefore, semantic annotation is necessary for reliable recognition of activities. A semantic space has to be defined that includes the most popular semantic characteristics of activity, namely the human body (posture and poselet), attributes, related objects, and context of the scene. We use the human knowledge to create descriptors that capture intrinsic properties of context-aware activities. The attributes describe the spatial and temporal movements of the actor. A deep model of CNN is developed that not only learns the attributes but also high-level semantic functions to better represent the activities and interactions in the group of actors. The results are used to better predict the activity including facial expressions in the context of holographic communication.

To be able to recreate realistically the movements of the user from one controlled environment to the other, we employ 3 to 6 calibrated RGB-d sensors, attached to the walls of the controlled environment on both locations. They are used to generate skeleton data of the moving users in real time. To avoid self-occlusion, multiple precisely calibrated RGB-d sensors are necessary. The skeleton of the human body is described by a number of joints such as hands, feet and facial features. These 2D features overlay a polygonal detailed 3D pattern that has $N = 10,475$ peaks and $K = 54$ joints that include neck, jaw, eyeballs, and finger joints. We use VNect [11] to capture the full global 3D skeletal pose of a human body. Its main idea is to combine CNN based pose regressor with kinematic skeleton fitting to estimate the 2D and 3D joint locations. We improve the method by employing more than one RGB-d sensor and capturing the movement of the user from different views at the same time. The gathered skeletal joint data is used in the next steps of the process: activity recognition and prediction.

Traditionally, in activity recognition and prediction tasks or behavior modelling, inputs are represented as base coordinate vectors of the skeletal joints. The problem with this type of presentation is that it does not contain information about the significance of the activity. Using a unitary vector alone, it is not possible to calculate how similar the two activities are and this information is not available for the model that will use the activity. The solution to this is to use embedding to represent the activity. While the base coordinate vectors are thinned and the model's characteristics increase with the size of the dictionary, the embedding is denser and more computationally efficient, with the number of functions being constant, regardless of the number of activities. Most importantly for the proposed model is that embedding gives a semantic meaning to the presentation of activities. Each action is presented as a point in a multidimensional plane that sets them apart from the other activities, thus providing similarity and meaning between them. To create a probabilistic model for predicting behaviour, we use a deep neural network architecture based on recurrent neural networks, in particular, long short-term memory (LSTM) [12]. LSTMs are versatile in the sense that they can theoretically calculate everything a computer can, given enough network units. These types of networks are particularly well suited for modelling problems where temporary relationships are appropriate and event intervals are unknown. LSTMs have also been shown to be suitable for sequential data structures. In activity modelling, the prediction of an activity label depends on the activities previously recorded. Recurrent LSTM memory management allows us to model the problem, given these consistent dependencies.

One challenge for the holographic communications is how to capture and reproduce accurately semantic traits such as expressions, age, gender, ethnicity, etc. in the case where the parties in the communication process wear VR/AR glasses which occlude the majority of the face. The most successful technique for "real-time facial reenactment that can transfer facial expressions and realistic eye appearance" in our opinion is HeadOn [13].

HeadOn is based on the idea of having prior scans or video data of the faces of the interlocutors so their facial characteristics can be parameterized. The parameterization of the whole head is done under general uncontrolled illumination based on a multi-linear face and an analytic illumination model. Features such as rigid head pose, geometric identity, surface reflectance, facial expressions and illumination form a feature vector describing the head with very high dimensionality. These unique facial characteristics can be used for facial matching to identify the correct avatar from the library in both locations. For the purpose of holographic communication this feature vector including data for the gaze tracking and semantic information of audio data are sent to the remote location to complete the process of facial reenactment of the avatar with photo-realistic rendering of the face region including opening of the mouth when speaking perfectly synchronized with specific speech information, blinking of the eyes and gaze tracking.

The last task of the proposed architecture is the real time avatar reenactment visualized at the remote site, based on the metadata captured at the home site and the semantic information gathered. The created avatar needs to be rigged with the captured skeleton hierarchy and appropriate texture maps for skin and clothes. To bind the actual 3D mesh of the avatar to the skeleton joint setup, we employ skinning process. The process entails that the joints have influence on the vertices of 3D model and move them according to the articulated motion, and most joints have influence on only certain parts of the 3D mesh of the model. A skeleton based animation strategy is employed for robustly and accurately fit the avatar to the skeleton and then larger scale deformations and movements are applied in real time. Thanks to the multiple RGB-d sensors, all the joints of the skeleton are visible.

Additionally, we use the semantic information of the recognized activity, to perform short term prediction of the skeleton movements which helps to compensate for network latency.

3.3 Context-Dependent Holographic Communication

Semantic models are independent of each other, i.e. the semantics of human activities, the semantics of 3D models, and the semantics of facial expressions do not belong to one space. We build dependencies between different semantics, which unites them into one semantic. This association is constructed using the context of the holographic communication.

The technical solution is based on multi-task inductive training and the construction of undirected graphs. In the first step, common layers are introduced in the deep architectures that encompass knowledge of all modalities used for training and thus separate the context-dependency of each of the modalities. In the second step, graphs are constructed describing the dependencies between the different semantics, and the weights of the graph edges indicate the severity of the dependencies. Using manifold learning, low-dimension cliques that are loosely coupled are removed from the semantic model, thus eliminating context-independent semantics.

In the proposed highly sophisticated holographic communication system, semantic and context-dependent information is an important part of the communication process to ensure near-zero latency. The communication sides of the framework will share in real time audio, semantic knowledge of face, body, hands and speech, even in the future, haptic signals. The use of semantic information in the context of a communication task requires quantitative assessment of the information as such. This allows us to evaluate the data compression that is achieved when using semantic information relative to its raw type or using standard compression methods. As the amount of semantic information depends on the interpretation of the meaning, not on the characters themselves transmitting the message, the end result may be that the number of symbols of the semantically shorter information is larger than the shorter message, but with significantly more information. This is precisely what necessitates the use of a modification of the standard measure of entropy, namely semantic entropy. At its core, semantic entropy is conditional entropy that exploits the dependencies between different messages. We use the semantic models created in the previous steps to define the main blocks of semantic entropy. These blocks are basic knowledge and models based on conditional probabilities describing dependencies. Once defined, an analysis and exploration of the amount of semantic information in the context of holographic communication is performed. Additional attention is paid to the compression of semantic information by detecting and removing semantically equivalent messages, i.e. reducing semantic redundancy at the source.

4 Context-Aware Holographic Communication in the Bee-Cubes Network: A Use Case Scenario

To illustrate a practical deployment of the proposed architecture, we develop a holographic framework of humans and their interactions among a network of globally connected hexagonal closed space, called “Bee cubes”. The Bee cube is conceived as a dedicated environment for enhancing and supporting BMIs.

The Bee cubes are hexagonal soundproofed rooms with diameter 4.23 m equipped with different business modelling tools, smart TV screen and their own controlled illumination.

They are equipped with advanced mobile and wireless sensors, both environmental and wearable by the participants. These modern technological advancements assist the processes ongoing inside. Their goal is to speed up the information flow between the participants, facilitate the observers in their objectives and help faster build of new business ideas and solutions. They can be put into any physical, digital or virtual business challenge and can enable any business, network of businesses, schools and universities to do any BMI—anywhere, anytime, with anybody—either in a physical, digital, virtual or integrated way.

The objective is to create a holographic communication between two or more of these Bee-cubes and to enable the participants to communicate between them no matter their physical location, to share, present and discuss business ideas but also the inter-actions between them to be observed in a passive way. In Fig. 2 is illustrated the conceptual model of the Bee-cube environment for holographic communication including all sensors deployment. In each cube there are 3 calibrated KinectV2 sensors with active microphone arrays for facial characteristics and skeleton joints tracking, loud speakers, one or more pairs of AR glasses - Microsoft HoloLens, one work station for data gathering, semantic knowledge extraction, processing and decision making.

The first steps towards context-awareness of the Bee-cube environment were done in direction to observe, analyze and predict human behaviour with goal to model the particular human behaviour and cognitive processes into semantic and logical expressions that are related to the specificity of BMI process [14].

In a 5G scenario major delays will be due to the computational complexity of the processing algorithms and the AR/VR head mounted display reacting to head movement

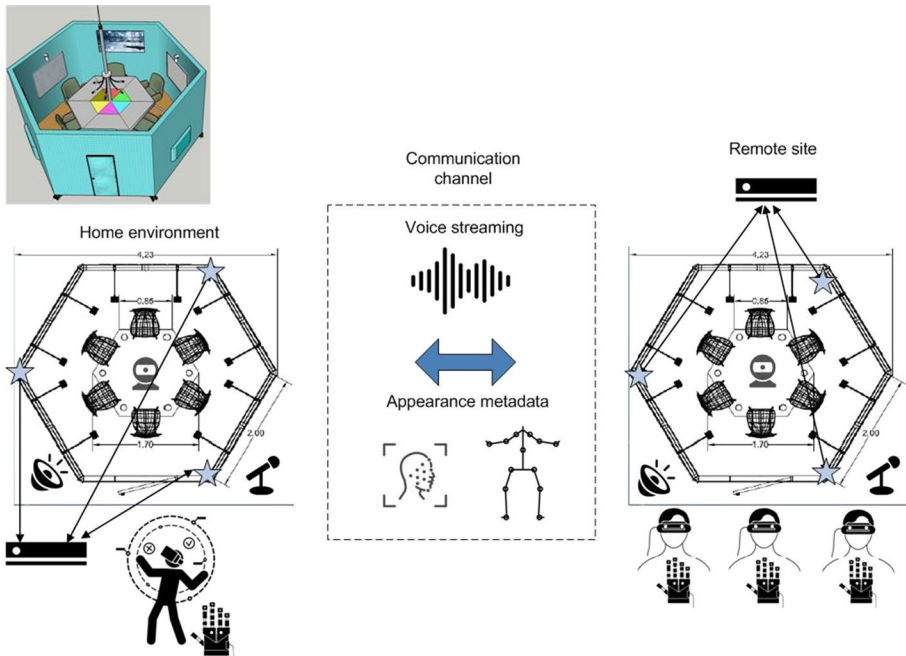


Fig. 2 Conceptual model of the Bee-cube environment for holographic communication including all sensors deployment for two-way communication where participants from both locations wear VR/AR glasses and data is shared

(user's changing views). To overcome these challenges we propose a distributed architecture where the processing is shared between the cubes in the network thus shuffling the computational resources to the edge of the communication network. Such an approach will be inherent in 5G and 6G networks to achieve communication-efficient distributed inference [15].

To achieve the connection of the cubes in a network and facilitating the control, data access and transfer but also to make possible the remote interpersonal communication, the following scenario is considered: two way communication process where all the participants from the home and remote site wear VR/AR glasses to be able to "holoport". They can see face to face and experience the feeling of "presence" with eye contact and facial expressions visible. In this case semantic information data will be transferred both ways. Thanks to the proposed context-aware holographic communication architecture, image artefacts and latency problems are minimized, thus empowering overall communication.

5 Conclusion and Future Work

Holographic communication applications have been considered as one of the most resource-demanding in the context of the 5G and the future 6G networks. Currently all major internet giants and corporations are developing holographic applications with fully immersive AR/VR experience and near-real personal communications with life-like holograms. Full immersion holographic telepresence systems will be achieved when all human senses can be included, but will require extremely high data rates (in the order of Gbps or even Tbps) to convey the rich and immersive content and even lower latency (less than 13 ms) for real-time user interaction [16]. The current holographic telepresence systems are still in their beginner stage. None of them can support large-scale communications over the global networks, due to the system's requirement on severely high data rate and furthermore the lack of agility in managing complex and ever-changing network conditions. This paper presents a model of a context-aware holographic communication architecture based on semantic knowledge extraction to overcome latency, limit the dependency on network resources and enhance current and future wireless technologies. The proposed approach can be considered as the way to a practical realization of an AI empowered wireless network that will give the opportunity to overcome current limitations in holographic communication. The benefit of the proposed context-aware holographic system is that it allows practically real time user interactions - anywhere, anytime, with anybody - either in a virtual or integrated way offering the feeling of personal interactivity and the feeling of shared space. Including all five senses in such architecture will bring us closer to achieving a full human bond communication [17]. Exploiting temporal consistency, different compressing technics, assuring Quality of Experience and incorporating cloud-based infrastructures will be the next steps in the proposed holographic communication process.

Acknowledgements This work was supported by the contract KP-06-H37/8 from 06.12.2019 for research project: "Inference algorithms for semantic knowledge extraction based on deep architectures for context-aware holographic communication" of the Bulgarian Research Fund of the Ministry of Education and Science, Bulgaria.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

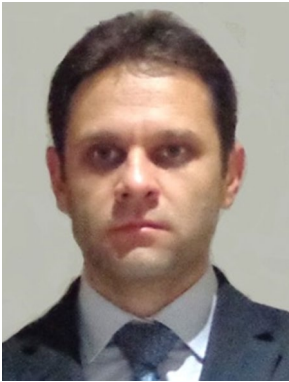
References

1. Al-Falahy, N., & Alani, O. Y. (2017). Technologies for 5G networks: Challenges and opportunities. *IT Professional*, 19(1), 12–20.
2. Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., & Kim, D., et al. (2016). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology* (pp. 741–754).
3. Saad, W., Bennis, M., & Chen, M. (2019). A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Network*, 34(3), 134–142.
4. Lin, C.-H., Sun, P.-Y., & Yu, F. (2015). Space connection: A new 3D tele-immersion platform for web-based gesture-collaborative games and services. In *2015 IEEE/ACM 4th international workshop on games and software engineering* (pp. 22–28). IEEE.
5. Sharma, S. K., Joshi, A., & Sharma, H. (2016). A multi-analytical approach to predict the Facebook usage in higher education. *Computers in Human Behavior*, 55, 340–353.
6. Yan, D., O'Brien, W., Hong, T., Feng, X., Gunay, H. B., Tahmasebi, F., & Mahdavi, A. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, 107, 264–278.
7. Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036–1040.
8. Thomas, M., Edwards, R. M., & Wang, Z. (2018). Consideration of IEEE 802.11 p and proposed 5G for Holograms in Vehicular Communication, 886–5.
9. Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 10975–10985).
10. Pons-Moll, G., Pujades, S., Hu, S., & Black, M. J. (2017). ClothCap: Seamless 4D clothing capture and retargeting. *ACM Transactions on Graphics (TOG)*, 36(4), 1–15.
11. Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H.-P., et al. (2017). Vnect: Real-time 3d human pose estimation with a single RGB camera. *ACM Transactions on Graphics (TOG)*, 36(4), 1–14.
12. Liu, J., Shahroudy, A., Xu, D., & Wang, G. (2016). Spatio-temporal LSTM with trust gates for 3d human action recognition. *European conference on computer vision* (pp. 816–833). Springer.
13. Thies, J., Zollhöfer, M., Theobalt, C., Stamminger, M., & Niessner, M. (2018). Headon: Real-time reenactment of human portrait videos. *ACM Transactions on Graphics (TOG)*, 37(4), 1–13.
14. Tonchev, K., Lindgren, P., Manolova, A., Neshov, N., & Poulkov, V. (2017). Digitizing human behavior in business model innovation. In *2017 global wireless summit (GWS)* (pp. 97–101). IEEE.
15. Letaief, K. B., Chen, W., Shi, Y., Zhang, J., & Angela Zhang, Y.-J. (2019). The roadmap to 6G: AI empowered wireless networks. *IEEE Communications Magazine*, 57(8), 84–90.
16. Selimis, I., Wang, N., Da, B., Yu, D., & Tafazolli, R. (2020). On the Internet-scale streaming of holographic-type content with assured user quality of experiences. In *2020 IFIP networking conference (networking)* (pp. 136–144). IEEE.
17. Prasad, R. (2016). Human bond communication. *Wireless Personal Communications*, 87(3), 619–627.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Agata Manolova is associate professor with the Faculty of Telecommunications at the Technical University of Sofia (TU-Sofia), Bulgaria and the head of the research laboratory “Electronic systems for visual information”. Her domains of interest are machine learning, pattern recognition, computer vision, image and video processing, biometrics, augmented and virtual reality. She has received her PhD from Université de Grenoble, France. She is laureate of Fulbright scholarship and an IEEE member.



Krasimir Tonchev a senior researcher leading research activities at the “TeleInfrastructure Lab”, Faculty of Telecommunications, TU-Sofia. His research interests include, on the theoretical side, large scale Kernel Machines, modelling of dynamical behaviour, Bayesian modelling, and on the application side, 2D and 3D facial analysis for soft biometrics, affective computing and general scene understanding from video. He is an IEEE member.



Vladimir Poulkov is a full time professor at the Faculty of Telecommunications at TU-Sofia. His expertise is in the field of information transmission theory, modulation and coding interference suppression, power control and resource management for next generation telecommunications networks, cyber physical systems. Currently he is Head of “Te-leInfrastructure” and “Electromagnetic Compatibility of Communication Systems” R&D Laboratories, chairman of Bulgarian Cluster Telecommunications, Vice-Chairman of European Telecommunications Standardization Institute (ETSI) General Assembly, Senior IEEE Member.



Sudhir Dixir is a Senior Fellow and Evangelist at the Basic Internet Foundation in Oslo, Norway. As a Docent at the University of Oulu, Finland, he also works on the Academy of Finland 6G Flagship Project. From 2015 to 2017 he was the CEO and Co-Founder of a start-up, Skydoot, Inc. From 2013 to 2015, he was a Distinguished Chief Technologist and CTO of the Communications and Media Services Business Unit for the Americas Region of Hewlett-Packard Enterprise Services in Palo Alto, CA, and from 2009 to 2013, he was the Director of Hewlett-Packard Labs India in Palo Alto and Bangalore. Before joining HP, he worked at Nokia, NSN, BlackBerry, and Verizon Communications. From 2010 to 2012, he was an Adjunct Professor of Computer Science at the University of California, Davis. A Life Fellow of the IEEE, IET and IETE, Dixit received Ph.D. from the University of Strathclyde, Glasgow, and M.B.A. from the Florida Institute of Technology, Melbourne.



Peter Lindgren holds a full Professorship in Multi business model and Technology innovation at Aarhus University, Denmark—Business development and technology innovation and is Vice President of CTIF Global Capsule (CGC). He is author to several articles and books about business model innovation in networks and Emerging Business Models. His research interests are multi business model and technology innovation in interdisciplinary networks, multi business model typologies, sensing and persuasive business models.