Holoportation of human avatars: Combining 3D imaging and haptics to enable multi-modal communications in 5G

Agata Manolova, Nikolay Neshov, Krasimir Tonchev, Vladimir Poulkov, Sudhir Dixit and Peter Lindgren

Abstract— Nowadays, the way people see the world and interact between each other is changing; new technologies are paving the way of innovative solutions for communication between humans. The recent research work on relevant topics and the rapid development of powerful hardware and software implementations allow for building of new communications pathways for lifelike interaction between people in different locations. One very new technology may offer the practical application of human bond communication is holoportation, but it also presents a lot of challenges in terms of digital data gathering and transmission even in a dedicated environment in 5G. By combining appropriate technics, we propose an approach for holoportation combined with haptic technology between same controlled environments. We present a conceptual model of a holoportation architecture for real time communication based on highly accurate 3D modelling of the human face and body, recognition and prediction of human actions and facial expressions to achieve realistic communications. Designed this way the proposed conceptual model of the holoportation system addresses the challenges from the information transmission aspect where real time constraints and narrowband channels are imposed, while big amounts of data, such as 3D body models of humans need to be transmitted. We illustrate its practical deployment in a particular use case scenario for Business Model Innovation.

INTRODUCTION

A. Manolova, N. Neshov, K. Tonchev, O. Boumbarov and V. Poulkov are with the Faculty of Telecommunications at the Technical University – Sofia, Bulgaria, email : {amanolova, nneshov, k_tonchev, olb, vkp}@tu-sofia.bg
Sudhir Dixit is with the University of Oulu, Finland and the Basic Internet Foundation, USA
P. Lindgren is with the Department of Business Development and Technology at the Aarhus University, email : peterli@btech.au.dk

The way people see the world and interact between themselves is changing; new technologies are paving the way of innovative solutions for communication between humans. The current communication technology involves the transmission of aural and optical data between subjects. Thus the users can use only two of their five senses (optic, auditory, olfactory, gustatory, and tactile) to interact with others from a distance. Without doubt video communication tools such as Skype and FaceTime are useful for many applications and tasks. However, text, 2D image and voice are not enough of a satisfying alternative to the personal interaction because the real eye contact is missing, touch is not possible between interlocutors and there is a limited feeling of presence (sharing the same space). These facts bring a sense of incompleteness and dissatisfaction from the communication process to the user.

The 5G technology will inherently bring aspects such as high-speed, ultra-low latency and high bandwidth — all in a 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 multi-modal communication industry to realize its full potential it must adopt the significant opportunities offered by 5G [1].

Currently in their infancy, augmented reality (AR), mixed reality (MR) and virtual reality (VR) technologies offer great potential to include all five human senses in the communication process to make it more meaningful for all participants. These technologies are already changing the way people interact between themselves and the machines and eventually will radically restructure the work process and the way of living. They will make it easier and more effective for executives, teachers, technologists, people with different professions to communicate in a more natural way anywhere around the globe, with all the advantages of physical presence but without the limitations of the current 2D video communication systems.

Holoportation is a complete human-computer-machine interface combining AR and VR telepresence thus creating real-life digital scans and realistic 3D avatars of subjects, displayed in MR environment, used for real time communication and interaction between remote users [2]. Holoportation can incorporate all five senses technologies with objects and elements from the home environment perceptually identical to the ones in remote site. 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. One major problem in holoportation is how to overcome the limitations of the communication channel when streaming the huge amounts of data generated by the capture process. Even with broadband channels, streaming big amounts of data will lead to a latency which will influence the natural perception of the communication. In the initial holoportation system presented in [2], a capture of a 3D frame including mesh geometry is about 5MB with compression so for real time performance, an average per frame transmission of 1-2Gbit/sec with overhead for compression (< 10ms) is a must. At the begging of 2018 initial experiments for holoportation between participants in 50 km range were presented by the University of Surrey (5GIC) requiring the high bandwidth and guaranteed low latency aspects of 5G. So to ensure such end-to-end throughput especially in the case of long distance international communications and in addition to the requirement for low latency is not an easy task. The holoportation being a MR application siting at the crossroads between eMBB and URLLC, needs multiple gigabits per second of data subject to serious latency constraints. Developing an intelligent way to “compress” the data so that current rate commodity channels can be utilized and exploit the potential of 5G is the main challenge of long distance real time holoportation.

In this paper we present a model of a holoportation architecture including haptics addressing real time communication needs, based on highly accurate 3D modelling of the human face and body, recognition and prediction of human actions and facial expressions. We illustrate its practical deployment in a particular use case scenario. The scenario is related to holoportation of humans and their interactions among a network of globally connected hexagonal controlled closed environments, called “Bee cubes”, for the needs of business model innovation (BMI). It addresses the challenges from the information transmission aspect where real time constraints and narrowband channels are imposed, while heterogeneous amounts of data need to be transmitted.

The rest of the paper is organized as follows: the next section presents the key features/building blocks of the proposed holoportation system. An insight into a use case scenario with application to BMI will be given with its opportunities and challenges. The final section pinpoints the challenges for such an architecture, draw the conclusion and suggest the scope of future work.

Architecture of a Holoportation System for Long Distance Real Time Communication

One of the main challenges for holoportation is the real time transfer of scene data. Short distance communications such as LAN can easily handle data rates of 1-2Gbps, but long distance real time holoportation seems rather impossible for current packet switched networks. A uniform experience is paramount for the holoportation. Lag, stutter, and stalls are unacceptable for user’ s quality of experience, service and comfort. Consistent quality, no disruptions from buffering, no reduction in quality from fluctuating bitrates, anywhere usage, reliable service even in challenging environments, high mobility are a must for the long distance real time holoportation.

A closed controlled environment alleviates the need for communicating the complete information about the surrounding scene. In particular, only objects possessing dynamical properties need to be transmitted over the channel. Such objects in the “Bee cube” are the people participating in the process of BMI. Parts of the scene such as walls, tables and even chairs can be considered static and not relevant. Moreover, they are exactly the same across all environments, so there is no need to holoport them. However, sending 3D human models can still require data rates in the order of hundreds of megabits, but the idea of human avatar provides clever hints towards metadata based compression. In our architecture for real time holoportation, we propose to send metadata about the scene and in particular, metadata related to the participants' visual appearance and movement. To be more specific, we propose to send the human skeleton data and facial parameters over the channel as well as the type of human activity, when recognized. Skeleton data consists of the 25 joints positions in 3D space and can capture the human pose as well as the person's position. The facial parameters on the other hand capture the facial data including expression, head pose and illumination: in total 265 parameters. Combining the skeletal with the facial data results in approx. 15 Kbit/sec data stream, considering 30fps and 12 bytes per point in 3D space. Audio and haptics data are transmitted using standard data rates for speech transmission. This information is negligible for modern communication channels and allows real-time holography to be realized through the proposed framework. Of course, high compression requires high computing power, but this can be achieved with modern GPU-based computing systems. We also achieve good Quality of Service and near zero latency thanks to human action recognition and prediction step [3]. Processing time including measuring depth map and 3D positions of 25 body joints is less than 33 ms so we can estimate motion with 0.5 seconds in advance. Thanks to this adaptation the network jitter tolerance increases and this information is enough for delivering a holoportation with good Quality of Experience (QoE).

The proposed conceptual system architecture for real time holoportation is illustrated in Fig.1. For simplicity, the holoportation 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.

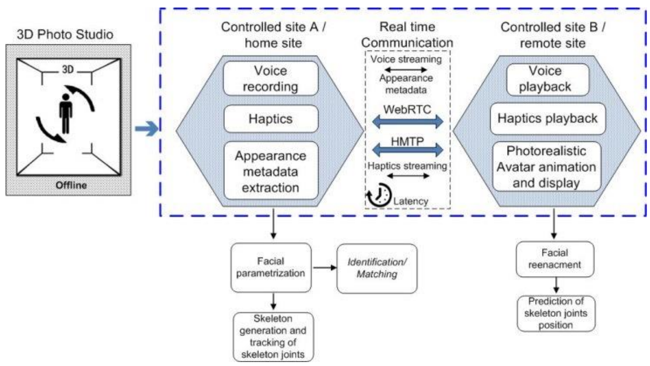

Figure 1. Conceptual architecture for a two-way real time holoportation system.

In this scenario we consider two communicating sites, home and remote with exactly the same arrangement and equipped with the same type of sensors for capturing visual, audio and haptic data and VR/AR glasses. Standard RGB-d sensors are mounted on the walls, calibrated and used to capture video and scene depth data. There can be two scenarios:

- One-way communication: where on the remote site an observer will wear VR/AR glasses and monitor the interactions between the participants at the home site in real time. To achieve this, the participants will be holoported at the remote location; The requirements for network latency are more relaxed as haptic data will not be transmitted.
- Two-way communication: where participants from both sites will wear VR/AR glasses and interact between each other. To achieve this participants' data including voice, face, body movement and haptics will be sent both ways. The requirements for end-to-end latency are very strict. Any delay between an action and the perception of its effects might lead to inconsistencies between action and stimulus presented by the head-mounted display and the haptic device, leading to unpleasant jitter in the holoportation experience.

To start the actual communication, a preliminary offline step is needed, in which a 3D model of the participants is built including all the data necessary for personal identification. This 3D model is available at the home and the remote sites and serves for the avatar which is animated. The animation at the remote site is done using the metadata sent by the home site, where the position of the person at the home site will determine the position of the avatar at the remote site, the skeleton data is used to animate the pose of the avatar and the facial key points are used to re-enact the facial expression. This metadata is generated at the home site using RGB-d cameras and dedicated computer vision algorithms. Additionally, voice and haptic data is captured by specialized sensors at home and transmitted simultaneously and synchronously with the metadata. Once the avatar at the remote is animated and synchronized in time with the audio and haptics information, it is displayed on the VR/AR glasses enhanced with audio and haptics playback.

Regarding the communication aspects, considerations have to be taken into account regarding the type of information sent. In order to accomplish real time, life-like holoportation, information is transmitted by both sides. The most important property of the streaming information is that the data follows strict order, arranged in time, so mechanisms for packet order reconstruction are utilized at the recipient. Delivering real time performance depends on two parameters of the network: capacity and latency. Since only metadata, audio and haptics data are transmitted this should not be a problem for the current state of Internet in terms of capacity. Latency could be a problem and we add another mechanism for latency compensation based on short term prediction of body movement.

3D human body model acquisition/human avatar creation: At the offline step, the human body model is captured using high quality 3D body scanner. Such scanner can be implemented in different ways, one of which is multiple camera systems placed in a controlled environment called studio. These systems developed for human shape reconstruction can run online or offline. Online systems must perform fast and robust real-time shape reconstruction but often they use off-the-shelf camera equipment and the cost of the scan is low. Offline scanning systems create much more accurate and detailed human avatars thanks to high quality cameras with a higher cost per scan but this fact can be counterbalanced by the scan being performed only once. For the holoportation to have lifelike experience is mandatory so more realistic avatar created by the offline system is a very good solution. In our case we consider a total of 72 off- the-shelf 8MP cameras arranged in cylinder formation with diameter 3m and height 2m similar to that presented in [4]. Each participant in the communication process must have a corresponding 3D model, which we build before the actual interaction (i.e. in offline mode, presented in Fig.1). The 3D models are available to the remote site beforehand and are stored in a library or the equivalent of an address book. In Fig. 2 are shown the essential reconstruction steps, which are used to acquire the 3D avatar.

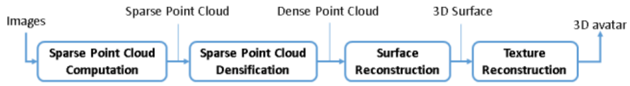

Figure 2. Pipeline of offline 3D avatar creation in multi-camera studio with predefined parameters such as illumination, uniform background and camera calibration.

The captured images form the cameras are used to create a detailed 3D human model with real textures such as skin, facial expressions and clothing. Creating such model based on images is usually done in multiple stages. As soon as the input images are captured, they are processed by the Sparse Point Cloud Computational algorithm. Initially in this algorithm, for each image a set of feature key points including color and illumination is extracted which are further utilized to solve the correspondence problem. This data is employed to solve the Structure from Motion (SfM) problem which output is a 3D sparse point cloud of the human figure. The densification step relies on the sparse point cloud to increase resolution/diminish the distance between the points and can be done, for example, using the patch-match stereo method [5]. Its goal it so reconstruct the 3D model of the human in more details. The texture reconstruction step is important for the holoportation application since this texture will be later used for the avatar's full body, face and clothing animation. There are typically two major steps in the Texture Reconstruction block: at first, a selection of view (or views) is performed in order to texture each face. Next, the texture is optimized for consistency to avoid seams between adjacent texture patches. Together with each avatar, identification data is stored, to help with the matching process performed at the begging of each conversation. After the lifelike 3D model with texture of each interlocutor is created offline, it is stored in the libraries on both sites using the equivalent of an address book. The network load for communicating the model is not considered as relevant, because it is not part of the actual communication process, but a preliminary step instead.

To accomplish our main goal to develop an appropriate data compression schemes for creating a multi-party from-capturing-to-rendering holoportation architecture suitable for the current telecommunication networks we employ several state of the art computer vision and machine learning algorithms.

Skeleton detection and tracking for human action recognition and prediction: For the purpose of the holoportation, 3 to 6 RGB-d sensors, attached to the walls of the controlled environment on both locations, are used to generate skeleton data of all participants in those locations. Multiple sensors are necessary to avoid self-occlusion, which is a common problem among most vision-based sensing systems. The skeleton of the human body is described by a number of joints representing key body parts such as head, neck, shoulders, elbows, wrist, torso, hip, knee and ankles. Each joint is represented by its 3D coordinates in a common coordinate system attached to the environment. The skeleton detection is done by combining the measurements of all RGB-d sensors. However combining the measurements from the different sensors creates a new issue known as the **data fusion problem** solved by employing Kalman filtering framework so the sensors can work together to correct any inaccurately captured joint data. Due to the recent advance of Deep Learning more precisely in Convolutional Neural Networks (CNN), skeleton detection is quite mature now, and can even be considered as technology, running on of the shelf computing resources. VNect [6] is one example of a real-time method able to capture the full global 3D skeletal pose of a human body. The main idea is to combine CNN based pose regressor with kinematic skeleton fitting to estimate 2D and 3D joint locations. Thanks to the existence of more than one RGB-d sensor capturing at the same time the movement of the user from different sides, the performance of the method on poses with significant amount of self-occlusion will not be a serious problem as it is with one camera.

Human action recognition can add value when animating the avatar at the remote site. For this task we utilize a spatio-temporal long short-term memory (ST-LSTM) network in the temporal and spatial domains [7]. ST-LSTM were chosen because of their ability to predict the input and compare it with the actual incoming input thanks to their long-term memory mechanism to learn when and how to remember and forget the contents of the memory cell as to minimize the estimation error.

Real-time facial identification and Gaze-aware facial reenactment: The biggest challenge for the holoportation is how to capture and represent accurately the facial characteristics and expressions especially if it is needed the parties in the communication process to wear large physical devices such as VR/AR glasses which occlude the majority of the face. So in order to identify the user and capture in detail the facial expressions, the facial data will be gathered and transmitted separately. We parametrize the facial characteristics of the participants during the teleconference. These facial key points transmitted in real time are used to reanimate the avatar's face. FaceVR [8] is currently one of the most successful "real-time facial reenactment approaches that can transfer facial expressions and realistic eye appearance" between two remote locations for users wearing head-mounted devices. This technology is based on the idea of having prior scans or video data of the faces of the interlocutors so the face 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. Parameters such as rigid head pose, geometric identity, surface reflectance, facial expressions and illumination form a feature vector describing the head with dimensionality R^{265}. This unique facial data can be used for facial matching with the scanned face of the avatar so the correct avatar will be chosen from the library in both locations. In such a case only this feature vector including data for the gaze tracking are sent to the remote location to complete the process of facial reenactment of the avatar with photo-realistic re-rendering of the face region including opening of the mouth when speaking, blinking of the eyes and gaze tracking.

Photorealistic real time animation of the avatar: The most important task for the proposed holoportation will be the real time avatar animation visualized at the remote site, based on the metadata captured at the home site. The created 3D model needs to be rigged with the captured skeleton hierarchy and appropriate texture maps [9]. To bind the actual 3D mesh of the avatar to the skeleton joint setup that we have created, we employ skinning process. This process entrails that the joints will have influence on the vertices of 3D model and move them according to the articulated motion, and most joints will need to have influence on only certain parts of the 3D mesh of the model. For enhanced realism, the skinning is realized as a combination of Linear Blend Skinning and Dual quaternion Linear Blending [10]. A skeleton based animation strategy is employed for robustly and accurately fitting the avatar to the skeleton and then large scale deformations and movements are applied in real-time. To achieve translation between two coordinate systems: one of the rigged avatar and the other of the captured skeletal data from the RGB-D sensors in real time, we subtract the offset vector from the mean vertex location of the stomach from the kinematically animated template to the mean of the waist and torso skeletal joints of the RGB-D parameters. Thanks to the multiple RGB-D sensors employed, there are no occlusions of joints. Additionally, we use the recognized activity if any, to perform short term prediction of the skeleton movements to compensate for network latency.

To add to the photorealistic part of the avatar, we employ a data-driven clothing capture approach to more easily dress the virtual models. The first step is to capture the garment geometry in motion on a body during the initial scan, then to estimate the body shape and pose under clothing, segment and extract the clothing pieces. Then the captured clothing can be transferred to new body shapes and poses [11].

Haptics: The sense of touch is conveyed to the human brain through the haptic sensory system and from the point of view of sensor replication is much easier to reproduce then the olfactory and gustatory senses. Touch is also another way to communicate emotions and help intensify interpersonal communication [12]. Haptics seek to add tactile sensations to the interpersonal interaction using device such as a haptic glove equipped with sensors for application of force, the feeling of texture; sense the bending of the fingers and movements of the hand or for full body haptic jacket to experience for example hugging sensations. For the holoportation architecture both options are viable. The gloves are fitted with sensors attached to microcontrollers, and a communication module for sending and receiving data. When the fingers on one hand move, the signal is sent via the network to the sensors in the other glove, which will vibrate to recreate the same motion. The jacket is equipped with an array of vibrating motors to simulate the sense of touch. The goal is to allow the feeling of handshake or hug between remote interlocutors, to overcome the deficit of the standard video communication and to make it more realistic, engaging and personal. Currently there are several transport and application layer protocols that deal with streaming of haptic data [13]. One of these transport layer protocols is developed especially to be employed in virtual environments for haptic collaborative use – the Hybrid Multicast Transport Protocol (HMTP), but we must note that this is a connection oriented protocol. Several attempts to create application layer protocols have been proposed for 3D tele-immersion environments, but one seems more promising in the context of holoportation: the MPEG Media Transport application layer transport protocol. For the proposed architecture, haptic, audio and appearance metadata are sent in separate data streams (Fig. 1). So a multiplexing and synchronizing scheme using different protocols, codecs and procedures for establishing the connection between two or more communication points is developed and customized in the holoportation context.

WebRTC multimedia communication protocol: The communicating parties will share in real-time between them audio, facial parameters, skeletal data, haptic signals, matching and localization data via customized Web Real-Time Communication stream. So using as a basis the WebRTC protocol [14], the holoportation system can guarantee the stability of the communication by adapting the data streaming to the state of the network. WebRTC enables peer-to-peer audio, video, and data sharing between peers in real-time with near-zero communication latency. Latency being one of the biggest challenges to overcome when speaking of the feeling of presence, the framework must achieve communication between the users in less than 16 milliseconds. Humans can detect time delays of approximately 16 milliseconds and greater. The proposed protocol runs over a normal internet connection and requires 3-5 Mbit/sec bitrate that even works below 1 Mbit/s thanks to the adaptive, depth encoding and streaming. The connection is established via a routing mechanism to connect peer-to-peer for the best transfer rates. Thanks to the implementation of secure communication protocols, platform independency and near-zero latency, WebRTC can be an ideal network platform for multi-user holoportation.

Real-time human bond communication: Finally to achieve real-time human bond communication and a rich immersive feeling in 3D virtual and augmented reality that users perceive the virtual body of their interlocutors as something natural all the above mentioned technologies will be combined. The computational load for extracting, processing, predicting and decision making are distributed between fixed backbone systems installed at the home and the remote site. A second alternative such as fog or edge computing can be a logical and adaptable approach to provide low latency real-time computing, where the computation resources are positioned around the network's edge close to the end users. The haptic, audio and appearance metadata including user identification are sent in separate data streams through the communication channel using standard codecs and protocols for these types of data but the multiplexing and synchronizing scheme is customized in the holoportation context. Thanks to the identical form and construction of the two communication sites the user's position in relation to the surrounding objects can be easily calculated at the remote site.

Holoportation in the Bee-Cubes Network: A Use Case Scenario

The Bee cube is conceived as a dedicated environment for enhancing and supporting BMIs. The Bee cubes are hexagonal soundproofed closed rooms with diameter 4,23 m equipped with different business modelling tools, TV screen and their own controlled illumination. They can be 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 goals 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 goal is to create a digital link 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 interactions between them to be observed in a passive way. In Fig. 3 is illustrated the conceptual model of the Bee-cube environment for holoporation including all sensors deployment. In each cube there are 3 calibrated KinectV2 sensors with active microphone arrays for facial characteristics and skeleton joints tracking, speakers, one or more pairs of AR glasses – Microsoft HoloLens, one work station for data gathering, processing and decision making.

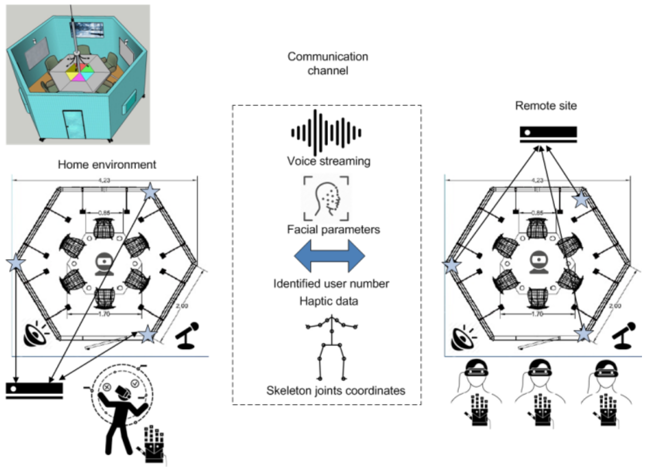

Figure 3. Conceptual model of the Bee-cube environment for holoporation including all sensors deployment for two-way communication where participants from both locations where VR/AR glasses and data is sent both ways.

The first steps towards digitalization of the Bee-cube environment were done in direction to observe, analyze and predict human behavior with goal to model human behavior and cognitive processes into logical expressions that can be digitized and automated to facilitate the BMI development [15]. In this network the delays are due to nature of the network channel and the latency of the AR/VR head mounted display of reacting to head movement (user's changing views). As a consequence we use a distributed architecture where the processing is done by a fixed backbone system installed at the home and the remote site. To achieve the connection of all available cubes in a network and facilitating the control, data access and transfer but also to make possible the remote interpersonal communication, two scenarios were considered:

• **Remote monitoring (One-way communication):** the ongoing meeting in cube can be remotely monitored by specialists, observing passively the interaction between the participants. This scenario represents a one way communication where an observer will observe unobtrusively the interactions between the participants in the cube. In this case the specialist or psychologist wearing VR/AR glasses will be able to monitor in real-time the interpersonal communication in the remote site. The avatars of the participants will be holoported and animated according to the body movements and facial expressions of the participants in the remote site at the same location in the controlled environment. The task of facial reconstruction and animation will be much easier because none of the participants will wear VR/AR glasses. In this case the biggest chunk of data will be transferred from the remote site including skeleton joints coordinates, facial parameters, voice streaming and at the start of the session a unique identification number of all participants will be sent to help identify the avatar recorded in the library. For this scenario the haptics are not necessary for the experience of the observer. He will be able to observe unobtrusively the behavior of the participants, their facial and body expressions, their attention levels and voice pitch.

• **3D Video teleconferencing/holoportation (Two-way communication):** this is a two or multiple way communication process. In this case all the participants from the home and remote site will wear VR/AR glasses to be able to holoport/experience the lifelike meeting where thanks to information from the haptic glove a handshake between interlocutors can be exchanged. They can see face to face and experience the feeling of "presence" with eye contact, facial expressions visible and touch possible between interlocutors. In this case scenario data will be transferred both ways. The conceptual model illustrated on Fig. 3 represents this scenario in detail. The participants wear VR/AR glasses so the facial reenactment of the avatar will follow the procedure described in the previous section. Thanks to the proposed real time communication architecture, image artefacts and latency problems will be minimized.

Conclusion and Future Work

The benefit of the proposed holoportation system is that it allows users to interact in real time - anywhere, anytime, with anybody – either in a virtual or integrated way offering the feeling of personal interactivity and the feeling of shared space. Holoportation offers the opportunity to address current limitations in communication combining 3D capture technology and mixed reality. Including all five senses will bring us closer to achieving a full human bond communication. But a lot of work is still needed to make this system an everyday reality for the society. Exploiting temporal consistency and different compressing technics of haptic, audio appearance metadata will be the next step in the holoportation process.

A challenge to consider related to the QoE and Quality of Task is that apart from only streaming haptic data through the communication channel, we must synchronize this stream with transmission of both audio and appearance metadata without leaving the scope of real time interaction. So how to aggregate and manage of streams of appearance, audio and haptic data in order to be transported using a single data stream is the next step of the development of the holoportation architecture as to allow seamless multi-modal communication. Some of the challenges of holoportation are related to identifying a rate adaption mechanism and reliability requirements but also how the transport protocols should be adapted to different situations (ex. for updating user's position requiring very low latency; or for receiving a remote image, requiring less strict delay). In this case a future path to consider is that the WebRTC protocol can be complemented with UDP-based Data Transfer Protocol (UDT), developed to transfer large data and to provide delivery of all messages sent on high latency connection.

Acknowledgment

This work was supported by European Regional Development Fund and the Operational Program "Science and Education for Smart Growth" under contract UNITe № BG05M2OP001-1.001-0004 -C01 (2018-2023).

References

1. Al-Falahy, N. and Alani, O.Y., 2017. Technologies for 5G networks: Challenges and opportunities. *IT Professional*, *19*(1), pp.12-20.
2. Orts-Escolano, Sergio, et al. "Holoportation: Virtual 3d teleportation in real-time." Proceedings of the 29th Annual Symposium on User Interface Software and Technology. ACM, 2016.
3. Elbamby, M. S., Perfecto, C., Bennis, M., & Doppler, K. (2018). Toward Low-Latency and Ultra-Reliable Virtual Reality. IEEE Network, 32(2), 78-84.
4. Starck, Jonathan, et al. "The multiple-camera 3-d production studio." IEEE Transactions on circuits and systems for video technology 19.6 (2009): 856-869.
5. Shen, Shuhan. "Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes." IEEE transactions on image processing 22.5 (2013): 1901-1914.
6. Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H.P., Xu, W., Casas, D. and Theobalt, C., 2017. Vnect: Real-time 3d human pose estimation with a single rgb camera. ACM Transactions on Graphics (TOG), 36(4), p.44.
7. Liu, Jun, et al. "Spatio-temporal lstm with trust gates for 3d human action recognition." European Conference on Computer Vision. Springer, Cham, 2016.
8. Thies, Justus, et al. "Facevr: Real-time facial reenactment and eye gaze control in virtual reality." arXiv preprint arXiv:1610.03151 (2016).
9. Fechteler, Philipp, et al. "Real-time avatar animation with dynamic face texturing." Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016.
10. Fechteler Philipp et al., "Example-based body model optimization and skinning." In Proceedings of the 37th Annual Conference of the European Association for Computer Graphics: Short Papers (EG '16), Germany, 5-8, 2016.
11. Pons-Moll, Gerard, et al. "Clothcap: Seamless 4d clothing capture and retargeting." ACM Transactions on Graphics (TOG) 36.4 (2017): 73.
12. Eid, Mohamad A., and Hussein Al Osman. "Affective Haptics: Current Research and Future Directions." IEEE Access 4.2016 (2016): 26-40.
13. Antonakoglou, Konstantinos, et al. "Towards haptic communications over the 5G tactile Internet." IEEE Communications Surveys & Tutorials (2018).
14. Gunkel, Simon, et al. "WebVR meets WebRTC: Towards 360-degree social VR experiences." Virtual Reality (VR), 2017 IEEE. IEEE, 2017.
15. Tonchev, Krasimir, et al. "Digitizing human behavior in business model innovation." Wireless Summit (GWS), 2017 Global. IEEE, 2017.

Biographies

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 is laureate of Fulbright scholarship and an IEEE member.

Nikolay Neshov is assistant professor and young researcher with the Faculty of Telecommunications at TU-Sofia. His domains of interest are content based image retrieval systems, face and emotion recognition, 3D image processing and machine learning.

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, modeling of dynamical behavior, Bayesian modeling, 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 “Teleinfrastructure” and “Electromagnetic Compatibility of Communication Systems” R&D Laboratories, chairman of Bulgarian Cluster Telecommunications, Vice-Chairman of European Telecommunicaitons Standartization Institute (ETSI) General Assembly, Senior IEEE Member.

Ognian Boumbarov is longtime professor with the Faculty of Telecommunications, TU-Sofia. He is a renowned researcher in the field of computer vision in Bulgaria. His main scientific interests are 3D face image processing, face and emotion recognition, neural networks, EEG and ECG signal processing, biometrics and human behaviour analysis.

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.