Abstract

The Internet is used for distributed shared experiences such as video conferencing, voice calls (possibly in a group), chatting, photo sharing, online gaming and virtual reality. These technologies are changing our daily lives and the way we interact with each other. The current rapid advances in 3D depth sensing and 3D cameras are enabling acquisition of highly realistic reconstructed 3D representations. These natural scene representations are often based on 3D point clouds or 3D Meshes. Integration of these data in distributed shared experiences can have a large impact on the way we work and interact online. Such shared experiences may enable 3D Tele-immersion and Mixed reality that combine real and synthetic contents in a virtual world. However, it poses many challenges to the existing Internet infrastructure. A large part of the challenge is due to the shear volume of reconstructed 3D data. End-to-End Internet connections are bandlimited and currently cannot support real-time end-to-end transmission of uncompressed 3D point cloud or mesh scans with hundreds of thousands of points (over 15 Megabytes per frame) captured at a fast rate (over 10 frames per second). Therefore the volume of the 3D data requires development of methods for efficient compression and transmission, possibly taking application and user specific requirements into account. In addition, sessions often need to be setup between different software and devices such as browsers, desktop applications, mobile applications or server side applications. For this reason interoperability is required. This introduces the need for standardisation of data formats, compression techniques and signalling (session management). In the case of mixed reality in a social networking context, users may use different types of reconstructed and synthetic 3D content (from simple avatar commands, to highly realistic 3D reconstructions based on 3D Mesh or Point Clouds). Therefore such signalling should take into account that different types of user-setups exist, from simple to very advanced, that can each join shared sessions and interact.

This thesis develops strategies for compression and transmission of reconstructed 3D data in Internet infrastructures. It develop three different approaches for the compression of 3D meshes and a codec for time varying 3D point clouds. Further, it develops an integrated 3D streaming framework that includes session management and signalling, media synchronization and a generic API for sending streams based
on UDP/TCP based protocols. Experiments with these components in a realistic inte-
TEGRATED mixed reality system with state of art rendering and 3D data capture inves-
TIGATES the specific system and user experience issues arising in the integration of
these sub components.

The first mesh codec is based on taking blocks of the mesh geometry list based on
local per block differential encoding and coding the connectivity based on a repeti-
tive pattern resulting from the reconstruction system. The main advantage of this
approach is simplicity and parallelizability. The codec is integrated in an initial pro-
TOTYPE for 3D immersive communication that includes a communication protocol
based on rateless coding based on LT codes and a light 3D rendering engine that
includes an implementation for global illumination.

The second mesh codec is a connectivity driven approach. It codes the connectivity
in a similar manner as the first codec but with entropy encoding added based on
deflate/inflate (based on the popular zlib library). This addition makes the connec-
tivity codec much more generically applicable. Subsequently it traverses the connec-
tivity to apply differential coding of the geometry. The differences between con-
Nected vertices are then quantized using a non linear quantizer. We call this approach
delayed quantization step late quantization (LQ). This approach resulted in reduced
encoding complexity at only a modest degradation in R-D performance compared to
the state of the art in standardized mesh compression in MPEG-4. The resulting
codec performs over 10 times faster encoding in practice compared to the latter. The
codec is used to achieve real-time communication in a WAN/MAN scenario in a
controlled IP network configuration. This includes real-time rendering and rateless
packet coding of UDP Packet data. The streaming pipeline has been optimized to run
in real time with varying frame rates that often occur in 3D Tele-immersion and
mixed reality. In addition it was tested in different network conditions using a LIFO
(Last in First Out) approach that optimizes the pipeline. In addition, it has been inte-
grated with highly realistic rendering and 3D capture.

The third codec is based on a geometry driven approach. In this codec the geometry
is coded first in an octree fashion and then the connectivity representation is con-
verted to a representation that indexes voxels in the octree grid. This representation
introduces correlation between the indices that is exploited using a vector quantiza-
tion scheme. This codec enables real-time coding at different levels of detail (LoD)
and highly adaptive bit-rates. This codec is useful when the 3D immersive virtual
room is deployed in the Internet when bandwidths may fluctuate heavily and are more restricted compared to the controlled WAN/MAN scenario. In addition, it is suitable for 3D representations that can be rendered at a lower level of detail, such as participants/objects rendered at a distance in the 3D Room.

Next, the focus shifts towards 3D Point Clouds instead of 3D meshes. 3D Point Clouds are a simpler representation of the 3D reconstructions. The thesis develops a codec for time-varying point clouds. It introduces a hybrid architecture that combines an octree based intra codec with lossy inter-prediction and lossy attributes coding based on mapping attributes to a JPEG image grid. It also introduces temporal inter-prediction. The predictive frames are reduced up to 30% and the colours up to 90% in size compared to the current state of the art in real-time point cloud compression. Subjective experiments in a realistic mixed reality virtual world framework developed in the Reverie project showed no significant degradation in the resulting perceptual quality.

In the last phase of this thesis, the complete 3D tele-immersive streaming platform is further developed. Additions include signalling support for 3D streaming (session management) that supports terminal scalability for light clients (render only) up to very heavy clients. This is done by signalling the local modular configuration in advance via the XMPP Protocol. Further, the streaming platform architecture presents an API where different stream types suitable to different 3D capture/reconstruction platforms (i.e. 3D audio, 3D visual, 3D animation) can be created. As the platform includes a distributed virtual clock, mechanisms to perform inter-stream and inter-sender media synchronization can be deployed at the application layer. Therefore, synchronization of compressed 3D audio streams in an audio playout buffer was implemented in a 3D audio rendering module. We also implemented a mesh and point cloud playout buffer in the module for synchronized rendering. This mesh playout buffer enables inter-sender synchronization between different incoming visual streams. In addition, the system includes simple publish and subscribe transmission protocol for small messages based on web socket (through a real-time cloud broker). In addition publish and subscribe based on the XMPP and UDP protocols was implemented. These publish and subscribe messages are particularly suitable for 3D animation commands and AI data exchange.

All components developed throughout this thesis have been integrated with 3D capture/rendering modules and in a social networking context in the larger Reverie 3D
Tele-immersive framework. Field trials of this system in different scenarios have shown the benefits of highly realistic live captured 3D data representations. This further highlights the importance of this work. The components developed in this thesis and their integration outlines many of the significant challenges encountered in the next generation of 3D tele-presence and mixed reality systems. These insights have contributed to the development of requirements for new international standards in the consortia MPEG (Moving Picture Experts Group) and JPEG (Joint Picture Experts Group). In addition, the developed codec and quality metrics for point cloud compression have been accepted as a base reference software model for a novel standard on point cloud compression in MPEG and are available in the MPEG code repository and online on github.

Keywords

3D Tele-immersion, 3D Mixed Reality, 3D Networking, 3D Mesh Compression, 3D Point Cloud Compression, 3D Rendering