

REVERIE: NATURAL HUMAN INTERACTION IN VIRTUAL IMMERSIVE ENVIRONMENTS

Julie Wall, Ebroul Izquierdo^{*} LEMONIA ARGYRIOU[◇]
David S. Monaghan, Noel E. O'Connor[†] Steven Poulakos, Aljoscha Smolic[•] Rufael Mekuria[▽]

^{*} Queen Mary, University of London, UK [◇] CTVC Ltd, UK

[†] Dublin City University, Ireland [•] Disney Research Zurich, Switzerland

[▽] Centrum voor Wiskunde en Informatica, Netherlands

ABSTRACT

REVERIE (REal and Virtual Engagement in Realistic Immersive Environments [1]) targets novel research to address the demanding challenges involved with developing state-of-the-art technologies for online human interaction. The REVERIE framework enables users to meet, socialise and share experiences online by integrating cutting-edge technologies for 3D data acquisition and processing, networking, autonomy and real-time rendering. In this paper, we describe the innovative research that is showcased through the REVERIE integrated framework through richly defined use-cases which demonstrate the validity and potential for natural interaction in a virtual immersive and safe environment. Previews of the REVERIE demo and its key research components can be viewed at www.youtube.com/user/REVERIEFP7.

Index Terms— tele-immersion, avatars, virtual human interaction, 3D audio and visual content

1. INTRODUCTION AND MOTIVATION

For many people, social networking is a way of life. More and more of our interactions with others are online. But to date, on-line interactions have been a poor substitute for real human interactions. For younger users especially, it is crucial that an on-line environment allows safe and enjoyable collaborative interaction bringing together realistic inter-personal communication with 3D media creation. In such an environment, a user can meet, socialise and share experiences. Furthermore, they can do so using well-known commercial depth-sensing equipment such as Microsoft Kinects and stereoscopic displays. In this paper we describe a technology platform to enable this, including not only sensing and analysis tools but also a range of content creation tools targeting such applications. Given such a platform, the possibilities are endless: take your class on a field trip to a realistic virtual model of the EU parliament; participate in interactive and entertaining role playing games; then invite your friends to hangout with you using their existing social networking channels.

REVERIE integrates cutting-edge technologies related to areas as diverse as 3D data acquisition and processing,

spatial sound processing, autonomous avatars, networking, real-time stereoscopic rendering, and physical interaction and emotional engagement in virtual worlds. This introduces technological challenges that are combined into two use cases or demonstrators that provide the basis for technical integration. These use cases also demonstrate the validity and potential socio-economic benefits of REVERIE's vision for the future of social networking: immersive social networks and natural interaction in an immersive environment.

2. SCIENTIFIC AND TECHNICAL DESCRIPTION

The main architectural components of the REVERIE prototype framework can be grouped by social network; 3D reconstruction and user activity analysis; networking for immersive communication; avatar reasoning, interaction and autonomy; and composition and visualization.

The REVERIE social network (reveriefp7social.eu/) provides a web portal with functionality for user account creation, authentication, and allows users to import their contacts and likes from existing social networks. Integration with social networking allows existing contacts to communicate in a new way by exploring exciting and educational locations and hanging out together in a realistic setting.

For the REVERIE multi-party immersive framework, methods for capturing, reconstruction and transmitting multiple photo-realistic users in real-time in the shared virtual spaces were developed with capturing devices such as Microsoft Kinect sensors.

These real-time 3D reconstructions of the user's body, termed 'replicants', include both 3D geometry and texture [2]. Novel methods for capturing and reconstruction include support for scalable coloured point clouds/meshes and low-poly meshes with multiple textures; two point-cloud smoothing modules, measurement confidence-based smoothing and fast moving least squares smoothing; and the simulation of real or virtual acoustic environments with dynamic reverberation.

Users can interact with each other and the elements of the shared environment through the development of novel solu-

tions for activity analysis and gesture recognition by means of differing modalities, e.g. WIMUs or Kinects [3]. Many machine learning techniques were investigated, and it was found that regularised logistic regression classifiers were the best fit for the user activities appropriate for the REVERIE framework. In addition, algorithms were developed for reconstruction of 3D models from images/video exploiting wide baseline stereo matching; and disparity estimation by applying spatio-temporal filtering to video data.

3D content coding, transmission and synchronization technologies are being developed and integrated to support the specific 3D data types effectively through the network [4, 5, 6, 7]. The framework distinguishes between semi-static (avatars, off-line user generated objects, videos) and real-time content (3D audio, reconstructed 3D visual/motion data and scene/clock sync messages). Semi-static content is stored in the 3D content servers and can be effectively retrieved with a DASH 3D content client. Real-time content is transmitted through a specific content delivery protocol based on the UDP protocol and optional application layer fec (AL-FEC) based on linear rate-less codes.

To enable prioritization and efficient synchronization of the streams, the traffic is divided into four classes with different delay, jitter, and bandwidth requirements: 1. High in volume visual data, such as the user's replicant representation, 2. 3D audio compressed with a wideband audio codec, 3. Motion and gesture data streams, and 4. Scene composition / renderer updates and clock synchronization messages. A component has been developed to manage all these streams jointly over four different channels with heuristics to achieve good receiver side synchronization and frame rates given current network conditions. The result is a content aware adaptive real-time streaming engine supporting the various classes of 3D data.

REVERIE's form of avatar reasoning serves to control the autonomous agents and semi-autonomous avatars that act in REVERIE's virtual environment with a framework which will drive their coherent and natural behaviour [8]. It supports interaction of both user-driven as well as autonomous virtual characters that engage in understandable social interactions in an emotional and expressive manner. Autonomous agents can perform specific actions at certain locations or times and behave autonomously triggered by user response [9, 10, 11]. For semi-autonomous avatars, which are midway between a passive avatar and an autonomous agent, the framework provides consistent control of aspects of the avatar not directly controlled by the user. The appeal of the reasoning framework in REVERIE is the complete chain that forms the basis for humanoid behaviour: observation of the users; reasoning about the world; and finally action.

Finally, all of the REVERIE 3D content (simple avatars, replicants, audio, environments) has to be visualised, [12]. The audio and visual streams from all the technologies described heretofore are rendered into one consistent and syn-

chronised output according to the mono- or stereoscopic display devices available. Some innovative REVERIE research has impacted the state-of-the-art in visualisation, e.g. fast global illumination computation algorithms and efficient deferred shading pipelines have been developed [13].

3. IMPLEMENTATION AND USE

Use Cases and Interactional Aspects: A user-centric approach is adopted in REVERIE by addressing several challenges that take into account what the end user actually wants from collaborative and on-line human interaction. Lifelike, naturalistic and real-time representations of the user enable natural communication and emotional interaction in a virtual environment. Collaborative gaming is facilitated in a realistic form, fostering both educational and entertaining outcomes. Finally, interactive services are responsive and adaptive to individual users as well as groups of users.

Two distinct use cases were identified to support the development, evaluation and demonstration of the REVERIE framework. The first provides strong integration of many people with an emphasis on social networking. It refers to the use of REVERIE mainly by students and teachers to provide educational services. In this scenario, student users registered to the REVERIE social network are invited by their teacher to an educational virtual field trip to the Brussels EU Parliament. An autonomous agent guides the students in an explanatory tour of the parliament scene. A virtual debate session then takes place where each student can present their personal topic to their fellow students. REVERIE allows its users to enhance their presentations by selecting to stream a video over the internet which will be rendered on the 3D virtual projectors in the parliament scene. The students have access to an avatar authoring tool [14] which automatically generates a new face for their avatar portraying their appearance and their facial expressions are captured in real-time through their web cam and adapted to their character's representation. The students move through the environment using an on-board navigation system, or under control of the autonomous agent via a flocking functionality. The users' attention and emotional status is analysed throughout the whole virtual experience and the system detects and recognises their head nods and shakes, as captured by a web cam, in order to understand their affective status. In turn, the autonomous agent demonstrates pre-scripted and autonomous body and facial animations (clapping, waving, happy and angry expressions, etc.) and speech triggered in response to the user's status. The students can take and share photos to their social media channels as they would do in a common social networking experience.

The second use case shows highly realistic visualizations and aims for the look and feel of real physical presence and interaction for entertainment. Users are reconstructed as they appear at their local REVERIE terminal. This results in a high quality representation of natural expression. The first

scenario, REVERIE Hangout, involves a replicant user immersed in a virtual 3D environment. The aim is to provide a high quality visual experience similar to Skype, however with each user reconstructed in a common virtual volume. Gesture-based interaction enables each user to navigate the environment and interact with graphical interfaces. The ability to capture and represent natural gestures provides the expressivity required for immersive social interaction.

The second scenario, Role Playing Games, explores the interactivity between users as well as autonomous agents. An example game Simon Says, provides a simple mechanism to evaluate a variety of human gestures. An autonomous agent, Simon, issues instructions for physical movement and observes the activity of each user. This simple game mechanism enables evaluation of a variety of human gestures. Finally, a graphical user interface was designed in order for the user to be able to access and interact with many of the above-mentioned functionalities in the rendered world.

4. PROPOSED DEMONSTRATION

The proposed demo will involve a live interactive software demonstration of the *educational virtual field trip* and *hang-out* scenarios from the REVERIE use cases. These will be show-cased in a real-time, engaging and diverting demonstration where user's can interact while being immersed in an exciting virtual world.

5. CONCLUSIONS AND FUTURE DEVELOPMENTS

To summarise, the main objective of REVERIE is the development of an advanced framework which enables immersive media capturing, representation, encoding and semi-automated collaborative content production, along with transmission and adaptation to heterogeneous displays. The framework acts as a key instrument to push social networking towards immersive collaborative environments that support realistic inter-personal communication. There are many potential avenues for utilizing this framework, and potential scenarios for both entertainment and education have been outlined in this paper.

6. ACKNOWLEDGEMENTS

The research that led to this paper was supported in part by the European Commission under the Contract FP7-ICT-287723 REVERIE.

7. REFERENCES

[1] REVERIE (REal and Virtual Engagement in Realistic Immersive Environments), <http://www.reveriefp7.eu/>

- [2] D.S. Alexiadis, D. Zarpalas, P. Daras, 'Real-time, realistic, full 3-D reconstruction of moving humans from multiple Kinect streams', *IEEE T Multimedia*, 15(2):339-358, 2013.
- [3] P. Kelly, C. Conaire et al., 'Motion synthesis for sports using unobtrusive lightweight body-worn and environment sensing', *Comput Graph Forum*, 32(8):48-60, 2013.
- [4] M. Sanna, E. Izquierdo, 'Live scalable video streaming on peer-to-peer overlays with network coding', *IEEE Latin-American T*, 11(3):962-968, 2013.
- [5] R. Mekuria, M. Sanna et al., 'A 3D tele-immersion system based on live captured mesh geometry', *4th ACM Multimedia Systems Conference*, pp. 24-35, 2013.
- [6] T. Zahariadis A. Papadakis, N. Nikolaou, 'Simulation framework for adaptive overlay content caching', *35th IEEE International Conference on Telecommunications Signal Processing*, pp. 62-66, 2012.
- [7] U. Ijaz, E. Quacchio, D. Alfonso, 'A web based framework for compressed 3D objects download and rendering', *10th International Conference on Frontiers of Information Technology*, pp. 129-133, 2012.
- [8] F. Kuijk, S. Van Broeck et al., 'A framework for human-like behaviour in an immersive virtual world', *18th International Conference on Digital Signal Processing*, pp. 1-7, 2013.
- [9] H. Wei, P. Scanlon et al., 'Real-time head nod and shake detection for continuous human affect recognition', *14th International Workshop on Image Analysis for Multimedia Interactive Services*, pp. 1-4, 2013.
- [10] B. Ravenet, M. Ochs, C. Pelachaud, 'From a user-created corpus of virtual agents non-verbal behavior to a computational model of interpersonal attitudes', *Intelligent Virtual Agents*, pp. 263-274, 2013.
- [11] J. Wall, E. Izquierdo, Q. Zhang, 'Fuzzy ensembles for embedding adaptive behaviours in semi-autonomous avatars in 3D virtual worlds', *18th International Conference on Digital Signal Processing*, pp. 1-6, 2013.
- [12] A. Ozerov, A. Liutkus et al., 'Coding-based informed source separation: Non-negative tensor factorization approach', *IEEE T Audio Speech*, 21(8):1699-1712, 2013.
- [13] M. Eitz, R. Richter et al., 'Photosketcher: Interactive sketch-based image synthesis', *IEEE Comput Graph*, 31(6):56-66, 2012.
- [14] K.C. Apostolakis, P. Daras, 'RAAT: The REVERIE avatar authoring tool', *18th International Conference on Digital Signal Processing*, pp. 1-6, 2013.