IRIDiuM: Immersive Rendered Interactive Deep Media

Babis Koniaris, Ivan Huerta, Maggie Kosek, Karen Darragh, Charles Malleson
Joanna Jamrozy, Nick Swafford, Jose Guitian, Bochang Moon, Ali Israr, Kenny Mitchell *
Disney Research

Abstract

Compelling virtual reality experiences require high quality imagery as well as head motion with six degrees of freedom. Most existing systems limit the motion of the viewer (prerecorded fixed position 360 video panoramas), or are limited in realism, e.g. video game quality graphics rendered in real-time on low powered devices. We propose a solution for presenting movie quality graphics to the user while still allowing the sense of presence afforded by free viewpoint head motion. By transforming offline rendered movie content into a novel immersive deep media representation, we display the content in real-time according to the tracked head pose. For each frame, we generate a set of 360-degree images (colors and depths) using cameras placed in selected locations within a small view volume surrounding a central viewing position. We employ a parallax masking technique which minimizes the rendering work required for the additionally visible surfaces in viewing locations around the main viewpoint. At run-time, a decompression and rendering algorithm fetches the appropriate surface data in real-time and projects them to the eye positions as the user moves within the tracked view volume.

To further illustrate this ability for interactivity and embodiment within VR movies, we track the full upper body using our sparse sensor motion capture solver allowing users to see themselves in the virtual world. Here, both head and upper body are tracked in real-time using data from IMU (Inertial Measurement Unit) and EMG (Electromyogram) sensors. Our real-time solver, *Triduna Live* uses a physics-based approach to robustly estimate pose from a few sensors. Hand gesture and object grasping motions are detected from the EMG data and combined with the tracked body position to control gameplay seamlessly integrated within the deep media environment.

Keywords: real-time rendering, VR, viewer tracking, interaction

Concepts: •**Computing methodologies** → **Image manipulation;** *Computational photography;*

1 Virtual Reality Deep Media

Fixed viewpoint video-spheres or 360° renderings lack immersiveness since the viewer's head translation is not accounted for. For full immersion, the image shown to the viewer needs to be computed in real-time according to their exact head pose. However, real-time rendering - because of computation time constraints - is typically of lower quality than offline ray-traced rendering available for cinematic content production. Light field rendering is highly realistic, but is currently computationally expensive for real-time rendering. Our 360° VR 'deep media' video format allows free-viewpoint display of high-quality graphics with 6DOF motion parallax in real-time on commodity hardware.

Figure 1: *Interactive panoramic movie display with translational motion parallax.*

The 'deep media' video enhances the immersion factor through the motion parallax effect; for example, viewers can shift their heads left or right to see what's behind a column. With such viewing freedom, besides the increased immersion, we can also set viewer-dependent inter-pupillary distance (IPD) to prevent viewing discomfort, as well as allow viewers to tilt and rotate their heads with full freedom, as long as the eyes remain within a supported view volume (Figure 1).

The novelty in 360° deep video lies in both capturing the data and rendering them. For every animation frame, we generate the data by rendering a 360° video plus depth of the dynamic environment from several viewpoints, located within a view zone. One of these viewpoints lies at the center of this view zone and is used as the main dataset used for display. The rest of these viewpoints are used to infer dis-occluded regions, unseen from the main view, via our parallax masking approach. With this representation a multi-layer relief mapping approach may be employed [Policarpo and Oliveira 2006]. However, this requires closed topology of layers, therefore we developed a ray-casting approach that samples this set of 360° views to render a novel view from any location within the a given volume. The data that we capture are final colors and depths from the camera. The fill-in views are generated using a novel offline rendering optimization that only generates the data that will be used, resulting in faster media production and better data compression through elimination of redundancy (Figure 2).

Our method currently works with synthetic data, but it is possible to use real datasets, as long as color and depth inputs are provided. To obtain such data, we can exploit the growing interest in light fields for VR in recent works, e.g. spherical capture [Debevec et al. 2015] or depth estimation [Kim et al. 2015].

2 User tracking and gameplay

Recent work has shown that user interaction is a valuable component of immersive experiences [Rheiner 2014; Han et al. 2015]. Therefore we improve not only the realism of the visualization, but

*e-mail:kenny.mitchell@disneyresearch.com

SIGGRAPH '16, July 24-28, 2016, Anaheim, CA,
ISBN: 978-1-4503-4377-0/16/07
DOI: http://dx.doi.org/10.1145/2929490.2929496

Figure 2: 360° *parallax video data for central view and mask for fill-in view.*

Figure 3: *Example of body motion capture using* Triduna Live.

also we add an embodiment of the user inside of the experience to interact with it. The movement of the game's main player is tracked in real-time by our motion capture and re-targeting technology *Triduna Live*. Inertial Measurement Units (IMUs) are used in combination with a physics-based body model to feed an inverse dynamics solver which solved for orientation constraints introduced by the sensors, as well as pose constraints from a trained motion prior. The player skeletal character rig is updated using a forward dynamics simulation to produce physically plausible motion trajectories for the body. Figure 3 shows an example of our body motion capture using Triduna Live.

In addition, the player's hand animation is driven in real-time using electromyogram (EMG) data. This EMG data is used to recognize hand poses to determine the player's actions. The open palm *magic force* gesture, causes a virtual weapon to move towards the palm. The *fist* gesture lets the player control the weapon. A *grasping* action is also recognized allowing the sensation of holding a real sword. Figure 4 shows the recognized gestures. Body presence, gestures and actions are visualized in the VR world thereby increasing the user presence/immersion in the gameplay. Figure 5 shows an example of the gameplay within Unreal Engine. Our *Triduna Live* solver is a cameraless, markerless, real-time approach for tracking and simulating the dynamic movements of a human subject using inertial sensor data and a state-of-the-art motion prior based on neural networks.

Acknowledgements

We would like to thank Desislava Markova and Fraser Rothnie for the Medieval Knight and Photoreal Sponza models, respectively.

References

DEBEVEC, P., DOWNING, G., BOLAS, M., PENG, H.-Y., AND URBACH, J. 2015. Spherical light field environment capture for virtual reality using a motorized pan/tilt head and offset camera. In *ACM SIGGRAPH 2015 Posters*, ACM, 30.

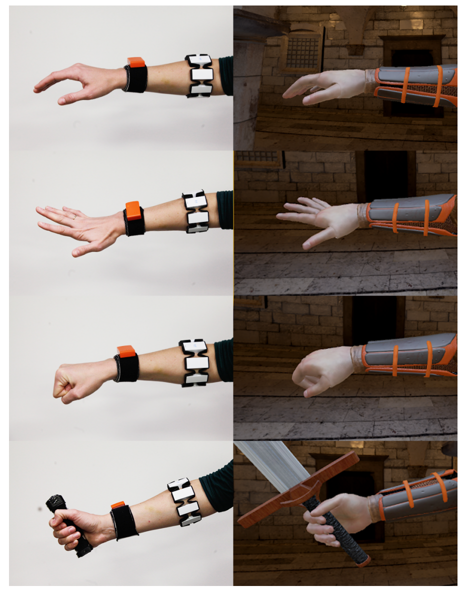

Figure 4: *Example of gesture and grasping recognition using EMG data.*

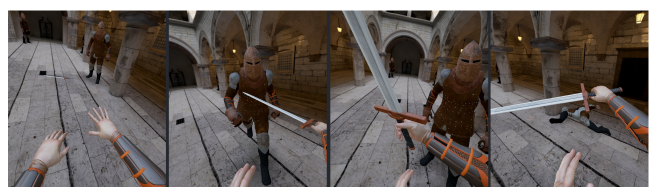

Figure 5: *Example of gameplay as seen through the VR headset.*

HAN, P.-H., HUANG, D.-Y., TSAI, H.-R., CHEN, P.-C., HSIEH, C.-H., LU, K.-Y., YANG, D.-N., AND HUNG, Y.-P. 2015. Moving around in virtual space with spider silk. In *ACM SIGGRAPH 2015 Emerging Technologies*, ACM, New York, NY, USA, SIGGRAPH '15, 19:1–19:1.

KIM, C., SUBR, K., MITCHELL, K., SORKINE-HORNUNG, A., AND GROSS, M. 2015. Online view sampling for estimating depth from light fields. In *Image Processing (ICIP), 2015 IEEE International Conference on*, IEEE, 1155–1159.

POLICARPO, F., AND OLIVEIRA, M. M. 2006. Relief mapping of non-height-field surface details. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, ACM, 55–62.

RHEINER, M. 2014. Birdly an attempt to fly. In *ACM SIGGRAPH 2014 Emerging Technologies*, ACM, New York, NY, USA, SIGGRAPH '14, 3:1–3:1.