AN IMMERSIVE HYBRID REALITY SYSTEM FOR CONSTRUCTION TRAINING

Ludovico Carozza, Frédéric Bosché, Mohamed Abdel-Wahab
School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, UK

ABSTRACT: *Mixed Reality (MR) technologies are inspiring a wide range of applications in the Construction industry, such as visual simulation for urban and construction planning and services' maintenance. Current MR systems offer only limited interaction modalities between the user (the trainee), the real surrounding world (including working tools and materials) and the virtual environment. Such interactions are essential for construction training applications.*
Current approaches aim at providing such capabilities through the virtualization of real world to be fused with the virtual one, and usually require the complex set-up and integration of external marker-based tracking systems, immersive displays (e.g. CAVEs) and haptic devices. Moreover such approaches, being based on pre-fetched avatar models of the real objects, often result in a lack of realism and flexibility.
In this paper we present the concept of an immersive Hybrid Reality (iHR) system which allows the trainee to experience in real time a realistic perception of self-embodiment within the virtual worksite as well as realistic interaction with both real and virtual objects. Our system fuses in real time the visual and structural information of the real world, acquired by an egocentric RGB-D camera integral with a wearable immersive head mounted display, within the virtual representation of a work environment. The system is based on a recently developed commodity hardware, which is affordable and easily set-up, and therefore it has the potential for supporting a wider application across vocational training courses.
The paper outlines the implementation of the main components of the system and the benefits achieved in terms of enhanced simulation capabilities. Promising preliminary results within the context of a training scenario are discussed, together with limitations of the current implementation and directions for further improvement.

KEYWORDS: *Immersive experience, Mixed Reality, Construction training, self-embodiment*

1. INTRODUCTION

Recent advances in Mixed Reality (MR) technologies are fostering an unprecedented development of new applications exploiting their simulation capabilities in a wide range of different sectors, spanning from entertainment to enhanced visualization and training. In the reality-virtuality continuum of Milgram et al. (1999), MR is defined as all the various types of `mixed' reality visualization solutions ranging from virtual reality to the real-world reality. In the AEC/FM (Architecture, Engineering and Construction & Facilities Management) sector, MR technologies are employed especially for visual simulation in urban and construction planning (Carozza et al. 2014a), facilities management and on-site information retrieval (Bae et al. 2013), enhanced visualization of built environment with augmentation of existing underlying structures (Côté et al. 2013). Recently, Augmented Reality has been used for guided maintenance and assisted training of service technicians for machinery plants by adding consistent multi-modal (audio/visual) aids and training material to the real environment (Bleser et al. 2011, Webel et al. 2013, Zhiwei Zhu 2014). But, extension to *training* in the AEC/FM sector has seen little application so far.

Simulation capabilities offered by MR technologies can have a huge impact on the *variety* of training scenarios and on the *quality* of the offered training, while at the same time reducing or eliminating health and safety risks normally present on the construction workplace (e.g. working at heights). Ideally, such training systems should be able to simulate challenging, hazardous or even extreme working conditions while the trainee is conducting a real procedural task, which could involve interaction, assembly or maintenance with real equipment or tools (Bosché et al. 2015, Webel et al. 2013). However, such functionalities imply a number of significant technical challenges. In particular, such training systems would require a high sense of presence for the user, or *immersiveness*, during movements, as well as realistic and consistent *interaction* with both the *real* and *virtual* training environments, which have to be seamlessly blended at both *perception* and *interaction* levels. In that regard, the realistic perception by the user of his own body within the virtual environment, i.e. *self-embodiment,* has been shown to be crucial for perception, e.g. of heights, within the virtual worksite (according to the *visual cliff paradigm* (Gutekunst et al. 2014, Lin et al. 2013)). The recent work of McGill et al. (2015) on interaction with real world objects (e.g. keyboard) while operating in immersive VR without removing the HMD shows that the user experience in an immersive VR is strongly affected by the realistic perception – *visualization* and *interaction* - of real objects with respect to the virtual scene. Accordingly, realistic and consistent interactions between the user, the surrounding real environment and the virtual worksite must be properly perceived, as they constitute an essential part of the learning process for an effective training experience. As a consequence, distances must be perceived correctly and occlusions between real and virtual entities (both ways) must be handled properly, overcoming one of the main limitations of conventional Augmented Reality (AR) systems, which generally simply overlay the virtual content (a virtual tag or model) on the real world's view, without taking account of the 3D structure (depth) of the real world.

Despite recent progresses, achieving an acceptable level of user experience still represents an open issue and often requires the integration of different technologies into complex and expensive dedicated facilities and systems. Current approaches generally employ Infrared (IR) marker-based optical tracking system to track the user's body, his parts and other real objects (usually within a CAVE environment), so that *virtualized* versions of them can be consistently inserted in the virtual environment. However, virtualization is often based on avatar models acquired off-line (for example, through a body scan), which can affect realism (the virtual representation of the body is not convincing) and flexibility to dynamic changes during interactions. These approaches subsequently need dedicated procedures, such as *retargeting*, to animate the character-based avatar according to the tracked movements of the real user (Lin et. al 2013). Manipulation and interactions are usually based on gesture recognition or dedicated hardware, such as haptic gloves or recently Leap Motion device[1].

In this paper, we present a proof of concept of a novel immersive Mixed Reality system, which enables the trainee to experience in real time a realistic perception of self-embodiment within the virtual worksite as well as realistic interaction with both real *and* virtual objects. We refer to this further level of augmentation of interactive immersive MR as *immersive hybrid reality*, naming our system *iHR* accordingly. Our system fuses in real time the visual and structural information of the real world, acquired by an egocentric RGB-D camera integral with an Oculus Rift[2] immersive head mounted display (HMD), within the virtual representation of a worksite. The noisy data acquired from the RGB-D camera are efficiently filtered and the remaining data represented as 3D meshes, taking into account the relative offset with respect to the HMD, estimated according to an off-line calibration procedure. Then, the real-time integration of this 3D structural information inside a game engine allows correct perception and interaction of both real and virtual entities. In contrast with comparable approaches, detailed in Sect. 2, our system has the following advantages:

- It is based on commodity hardware and a compact light-weight wearable set-up (Fig. 1). Such immersive hardware is becoming widely available at a low price, and this would potentially ease the diffusion of such systems. Moreover, being egocentric, our system is less prone to self-occlusions.
- In our system, the self-through virtual embodiment perception, or virtual ownership illusion (that is, the realistic visual representation of oneself in the Virtual Environment), is not achieved through the creation of a user's virtual avatar. Instead, our approach ego-captures the real 3D data (of oneself) and ego-renders it within the virtual environment.
- Since our system ego-captures in real time the 3D structure of the surrounding real world, including other users and multiple objects (without the need of tracking them explicitly), occlusions of real and virtual objects can be handled correctly (similarly to (Carozza et al. 2014a), but in real time), overcoming one of the main limitations of conventional AR systems. Potentially, imminent collisions can be detected to issue warnings to trainees in real time, which would be of value in health and safety contexts.
- Interactions between real and virtual objects can be simulated more consistently, achieving a near-physical interaction modality. Free hand object manipulation would not require additional equipment (such as motion capture gloves).

[1] https://www.leapmotion.com/product/vr

[2] https://www.oculus.com/

The main stages of our approach and details of its implementation are presented in Sect 3. Preliminary qualitative results, with a focus on potential application to training in the AEC/FM sector, are illustrated in Sect. 4. Finally, Sect. 5 discusses the current limitations and highlights future steps to overcome them.

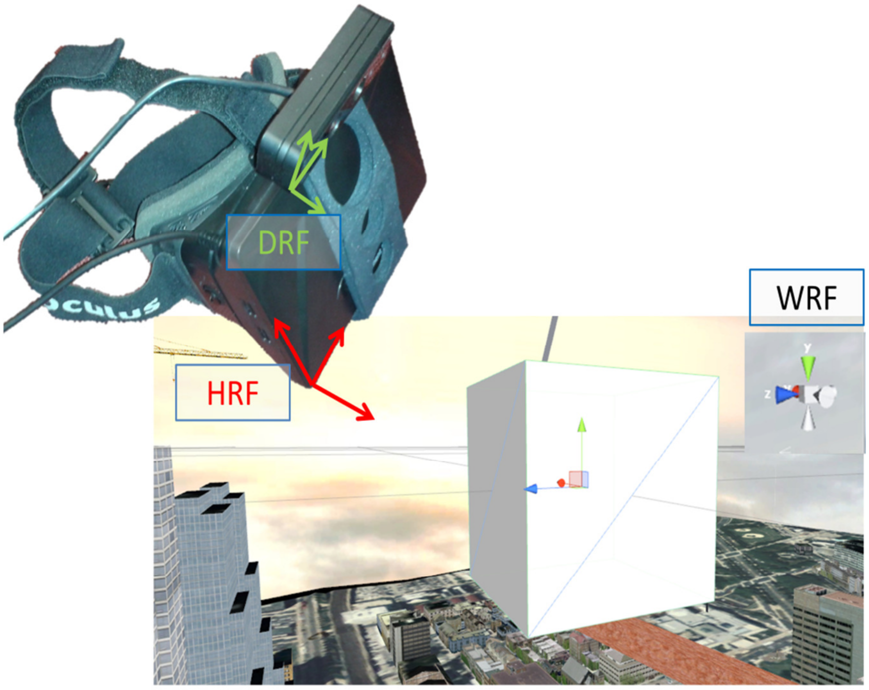

Fig. 1: Set-up of our iHR system and relevant reference systems used for calibration.

2. PREVIOUS WORK

Based on the well-known Milgram reality-virtuality paradigm (Milgram et al. 1999), both Augmented Reality and Augmented Virtuality (AV) approaches have been employed to simulate fusion of virtual and real content. In this paragraph we discuss recent immersive approaches using consumer hardware based set-ups.

The approach described in (Khattak et. al 2014) employs, in addition to the Oculus Rift and the RGB-D camera, a Leap Motion device to track hands and fingers, and an AR marker placed nearby to track the head in a small volume, as an alternative to an external optical tracking system. With respect to conventional AR, their AV approach can handle properly a virtual object occluding a real one (hands) as well as hands occluding a virtual object, so to perform better object manipulation tasks. However, only augmentation with hands is possible, and requires the Leap Motion controller to track and recognize hand gestures, so that interaction and manipulation can be performed. This results in a more complex set-up and limited accuracy in interaction (inaccurate gesture recognition).

The set-up presented in (Steptoe et al. 2014), called AR-RIFT, provides a video see-through AR by mounting two monocular cameras on an Oculus Rift in correspondence to the two eye positions, so that the real world can also be "seen" through the Oculus Rift. This is made possible through the large field of view (FOV) achieved through the AR-RIFT system set-up and the cameras' high resolution, resulting in noticeable visual quality and immersiveness. However, this set-up acquires the real scene and performs the rendering of the virtual content independently on the two camera images without any stereo matching, so that just side-by-side (and not full) stereoscopy is achieved. Accordingly, real depth information is not available and necessarily virtual/real object interaction cannot be fully consistent. Indeed, in such an AR approach, the real entities (the user's hands) always result behind the virtual objects.

The same set-up is used in (Gutekunst et al. 2014), where an AV system (the virtual content being predominant) is employed to investigate height perception (simulating the user "looking" down into a pit). In this work, hands/arms only are extracted from the real scene separately on the two camera images based on a color model. As mentioned by the authors, the main limitations of this approach are the low frame rate (14 fps) and most importantly the lack of real depth information, which jeopardize correct occlusion handling, so that in this AV approach real entities (the user's hands) always occlude the virtual objects.

Very recently, McGill et al. (2015) have proposed a novel AV approach that partially *blends* reality and virtuality, and study the effect of varying the amount of blending of real objects (incorporating reality into a virtual scene). Preserving the sense of touch on real objects while being immersed in a virtual environment is considered important for user presence awareness when performing a task. Immersion itself can be increased through natural interactions with virtual objects as they occur in real world. Due to an improved sense of presence and proximity to real entities, including other users, they see interaction through AV as a potential alternative to gestural interfaces (Leap Motion) and/or interfaces suitable for use without sight (tangible controllers, keyboards). The authors' implementation makes use of chroma-key as the standard for AV, by means of background subtraction within a physical environment judiciously coloured in green. The presented set-up uses a single camera mounted in front of the Oculus Rift, and only monocular depth cues. Accordingly, as pointed out by the authors, since the system is not able to perform true stereoscopy, real depth cannot be measured and the user can just roughly orientate himself within the real environment. This could also jeopardize consistent occlusion handling, as already discussed for the works of (Steptoe et al. 2014) and (Gutekunst et al. 2014).

The system described in (Tecchia et al. 2014) employs a set-up very similar to the one used in our work, with 6-DOF tracking in addition. Hands of the user are reconstructed from the RGB-D camera data, and discrete hand markers attached to fingers are tracked to simulate simplified grabbing of virtual objects. However, their system is conceived for VR experience only, and does not encompass interaction with real objects.

3. THE PROPOSED SYSTEM

The proposed iHR system integrates the 3D structural information of the real scene acquired by the RGB-D camera within the virtual world rendered on the HMD in a consistent reference frame. The system setup is detailed in Sect. 3.1. In order to deliver a consistent user experience, off-line and on-line stages are involved:

- *Off-line stage*: Initially, the system needs to be set-up and calibrated off-line, so that the information acquired in the RGB-D reference frame can be expressed in the HMD reference frame, on which everything is rendered. The calibration procedure is detailed in Sect 3.2.
- *On-line stage*: During on-line operations, the RGB-D camera data are filtered and integrated inside the game engine to simulate correct interaction between real and virtual entities. Filtering is necessary because RGB-D data provided by Time-of-Flight (ToF) technology is noisy. This is followed by a mesh reconstruction step. The reconstructed mesh, expressed in the HMD's reference frame according to the calibration transformation, can interact with the virtual environment through the physics engine of the game engine. It is worth noting that this stage must present real-time performance in order to reduce latency effects. Its implementation is detailed in Sect. 3.3.

3.1 System Setup

The system we propose (Fig. 1) is based on an RGB-D camera (Softkinetic DepthSense 325) mounted integrally with an Oculus-Rift DK1 (HMD) by means of a 3D printed camera bracket (http://www.shapeways.com/). The system is compact and light-weight (approximately 600 g), and is based on off-the-shelf consumer hardware components, which makes it affordable (approximately $500).

The wide field of view (FOV) of Oculus Rift (110º, diagonal), together with high frequency head orientation tracking, allows experiencing a higher degree of immersiveness. Note that, in the work reported here, just head rotations are tracked, not positional translations, so that only static-position HR experiences are simulated (6 DOF tracking is part of on-going development).

The depth camera (30 fps, 320 x 240 pixels) works in long range mode, nominally within the depth range of [0.15- 2.9] meters, and a field of view (FOV) of 74º x 58º x 87º (H x V x D). As described in Sect.3.2, additional volume culling has been implemented inside the system in order to specify the action range of the depth camera within the selected simulated training scenario. Typically, we choose to operate in the range [0.15 - 1.5] meters, that is the volume within which objects are reachable by the user with their hands.

RGB values are acquired from the color camera (30fps, 1280 x 720 pixels, FOV: 63.2º (H) x 49.3º (V) x 75.2º (D)) and *mapped* to the corresponding depth values using the registration transformation (UV map) between the two cameras provided by the manufacturer. For both cameras, a standard (pin-hole) projective camera model (Heikkila and Silven 1997) is applied, employing intrinsic and lens distortion parameters provided by the manufacturers.

As a result, for each acquired frame a RGB-D map, $\mathrm{M(u,v)} \rightarrow [D(u,v), RGB(u,v)]$, is created with consistent depth and color values for each pixel (u,v) in the depth image domain. Necessarily, due to the smaller FOV of the RGB camera, the depth values not belonging to the color camera frustum do not have corresponding color information associated.

3.2 Calibration

The depth values are acquired in a reference frame integral with the RGB-D camera (Depth Camera Reference Frame, DRF), which differs from the global (*world*) reference frame (WRF) of the scene adopted by the game engine and according to which everything must be expressed. Moreover, during on-line operations, the HMD must be tracked in order to know the *viewpoint* for the rendering of the virtual environment, i.e. the pose of the HMD reference frame (HRF) with respect to the WRF must be estimated. Fig. 1 illustrates all the reference frames involved.

Let $p^{DRF} = (X^{DRF}, Y^{DRF}, Z^{DRF}, 1)^T$ be the generic 3D point acquired by the RGB-D camera and expressed in homogeneous coordinates in the DRF, it is transformed into the WRF according to the depth camera pose $T_{DRF}^{WRF} = (R_{DRF}^{WRF}, t_{DRF}^{WRF})$, where R_{DRF}^{WRF} represents the rotation matrix and t_{DRF}^{WRF} the translation vector, through the equation:

$$p^{WRF} = \begin{bmatrix} R_{DRF}^{WRF} & t_{DRF}^{WRF} \\ \mathbf{0}^T & 1 \end{bmatrix} p^{DRF} = T_{DRF}^{WRF} p^{DRF} \qquad \text{(Eq. 1)}$$

For the HRF, the transformation can be similarly written:

$$p^{WRF} = T_{HRF}^{WRF} p^{HRF} \qquad \text{(Eq. 2)}$$

From (Eq. 1) and (Eq. 2), we achieve:

$$p^{HRF} = (T_{HRF}^{WRF})^{-1} T_{DRF}^{WRF} p^{DRF} = T_c \, p^{DRF} \qquad \text{(Eq. 3)}$$

where T_c is the calibration transformation to estimate in order to refer the measured 3D points to the HRF and finally to the WRF (by substituting the HMD pose T_{HRF}^{WRF} from head tracking into (Eq. 2)). Accordingly, T_c retains the relative transform between DRF and HRF, which depends only on the mechanical set-up.

Our calibration procedure aims at estimating R_c. The translation vector t_c can be adjusted subsequently to the null vector by artificially moving (using the game interface) the viewpoint of the virtual camera, i.e. the origin of the HRF, until the user does not perceive any difference between the 3D reconstruction (in the WRF, according to (Eq. 1)) of a real reference pattern, positioned at a known fixed distance, and a virtual representation of the reference pattern, rendered in the virtual world at the same fixed position (in equivalent units).

To estimate R_c, we use a simple inertial (accelerometer) based approach. Indeed, the RGB-D camera is equipped with an accelerometer which provides estimates of the gravity vector for each acquired frame. Due to our system set-up, the yaw offset between the HRF and the DRF can be neglected, and only the relative roll and pitch angles must be estimated. This is equivalent to estimate the misalignment between the vertical expressed in the two reference frames (HRF and HRF) as measured from the HMD IMU and the RGB-D camera accelerometer.

During calibration, the system is placed still on a horizontal plane, e.g. the desk, used as Gravity Reference Frame (GRF). For each frame the pitch and roll angles measured by the RGB-D accelerometers and the HMD IMU are employed to compute the corresponding rotation matrices R_{DRF}^{GRF} and R_{HRF}^{GRF}, and infer R_c from (Eq. 4):

$$R_c = \left(R_{HRF}^{GRF}\right)^{-1} R_{DRF}^{GRF} \qquad \text{(Eq. 4)}$$

Many such measurements can be arranged over a stack of N_c frames into a linear system and solved with the Least Squares Method. Fig. 2 shows an example of the rectification achieved through this calibration procedure.

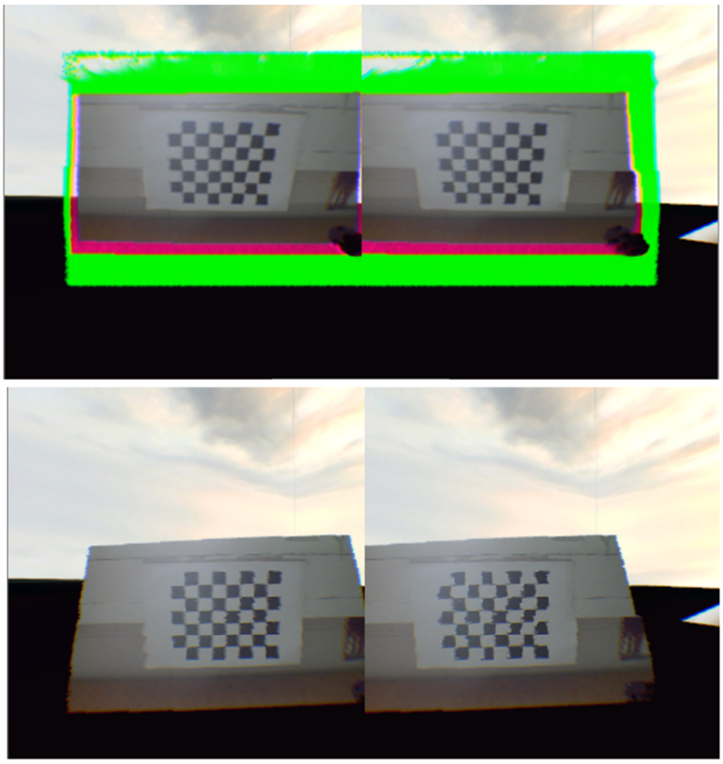

Fig. 2: Result of the calibration procedure. First row: stereoscopic view of unrectified point cloud. Second row: stereoscopic view of the mesh reconstructed from the rectified point cloud after the calibration procedure. Vertical lines on the checkerboard become vertical in the rectified views.

3.3 Implementation

The 3D data acquired from the depth camera need to be fused consistently with the virtual entities in order to deliver a consistent user/training experience. To this purpose, we implemented a software data processing framework that can be integrated as a C++ plugin for Unity4.6 game engine. The pipeline of this framework involves two main steps: *filtering* and *reconstruction*.

Current ToF technology is affected by several artifacts resulting in depth measurement errors (increasing with distance), depth outliers and *holes* (due to reflectivity of materials and surface curvatures), high noise around depth discontinuities (*flying pixels*), in combination with motion blur (Lefloch et al. 2013). In order to mitigate this effect, the depth map $D(u, v)$ and color map $RGB(u, v)$ are both initially filtered by using a median kernel with size W=3, which can be computed fast and reduces the impulse noise. Moreover, a validity binary mask $\mathrm{V(u, v)}$ (1 for valid pixels, 0 for invalid pixels) is created by applying a confidence threshold (function of the IR amplitude) and depth threshold (volume culling). These parameters are configurable, with real-time feedback, through the game graphical user interface, and saved and loaded as presets for different training scenarios.

Next, a two-stage combination of morphological opening and closing operations with increasing size of the structuring element (W=1, W=3) is applied on the validity binary mask $\mathrm{V(u, v)}$ to remove small unconnected regions and fill small holes.

Explicit surface-based modelling has been chosen to represent the 3D structure of the acquired scene in real time inside the game engine. A variation of the method described in (Turk and Levoy 1994) has been implemented to this purpose. For each valid pixel of the filtered depth map $\overline{D}(u, v)$, the depth values of the window corners $\overline{D}(u + \Delta, v)$, $\overline{D}(u, v + \Delta)$, $\overline{D}(u + \Delta, v + \Delta)$, $\Delta = 2$, are examined and their mutual distances computed. They are linked together if they are 'closer' than a distance threshold, and triangular meshes are built accordingly from these vertices on the smaller diagonal edge. As a consequence, zero, one or two triangles can be generated for each valid pixel. Efficient strategies have been put in place to maintain real-time performance, quickly skipping invalid vertices. Per-vertex coloring is then applied to each vertex by reading the corresponding value in the filtered color map.

The resulting mesh is passed on to the Unity game engine standard graphical mesh representation. Positioning of the reconstructed mesh inside the Unity scene is then employed not only for visualization, i.e. properly taking into account occlusions, but also for interaction in the Unity physics engine.

4. RESULTS

The proposed system has been preliminarily tested with regard to a potential training scenario, i.e. training for crane operators. The user is positioned inside the cabin of a crane model[3] overlooking a virtual representation of the city on Philadelphia[4], letting him interact with virtual and real objects. The system runs at 30fps on a Dell Aurora Alienware PC (Intel i7-3280 @ 3.6GHz, 16GB RAM, NVIDIA GTX 680).

In Fig. 3 the views (left eye screenshots) rendered through the Oculus Rift at four sample times are shown as an example. One can notice the good visual quality for representing hands and other real objects, e.g. the hard hat. Note that the user can also see his legs immersed within the environment, which further enhances the sense of presence.

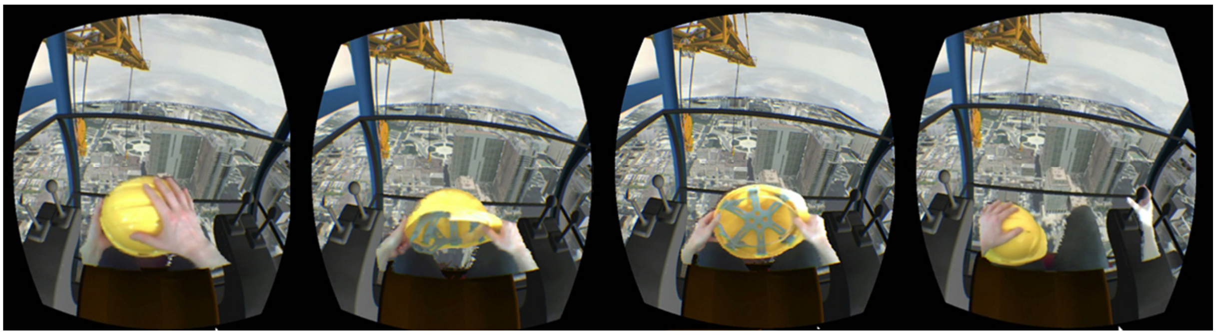

Fig. 3: Screenshots of the left-eye rendered view at different times. Distortion is due to Oculus pincushion compensation.

In Fig. 4, zoomed views of the user hand trying to grasp the virtual crane joystick show correct occlusion handling in both the cases when the hand and the joystick occlude each other.

The user can rotate the crane in two ways: (1) by virtually "colliding" with the virtual joystick, i.e. the user keeps his hand close (within the virtual cube) near the right virtual joystick for a sufficient amount of time; or (2) b operating through a real joystick, which is "positioned" consistently within the virtual scene. Fig. 5 shows

[3] http://cad-unigraphics-projects.blogspot.co.uk/2011/07/tower-crane-3d-model-free-cad-download.html

[4] Model by courtesy of ESRI.

screenshots related to the first mode, with the virtual cube passing from red when the hand enters the virtual cube, to green if the hand "holds" the cube for a sufficient time, and finally becoming blue when the cube is "released". Fig. 6 shows screenshots corresponding to the rotation of the crane controlled by using the real joystick. The limited latency of our system as well as an accurate calibration procedure facilitate the reach-and-grab task of the real objects, i.e. the real joystick in this case.

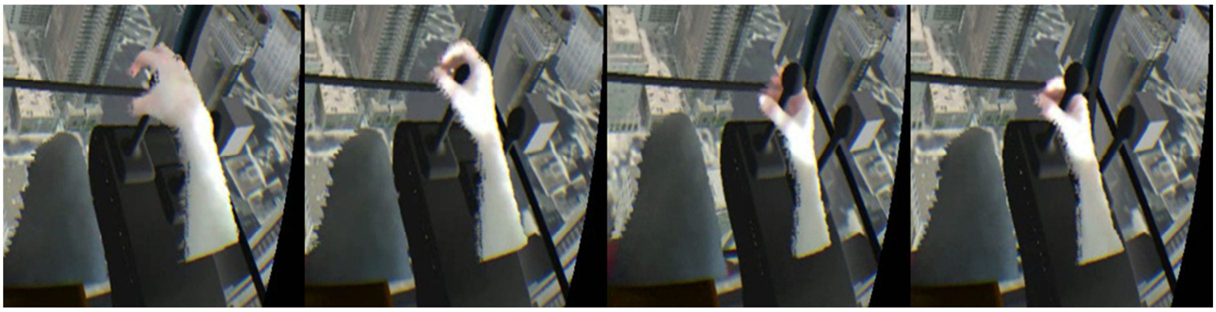

Fig. 4: Details of (left-eye) views of the user hand trying to grasp the virtual joystick, with correct occlusion handling.

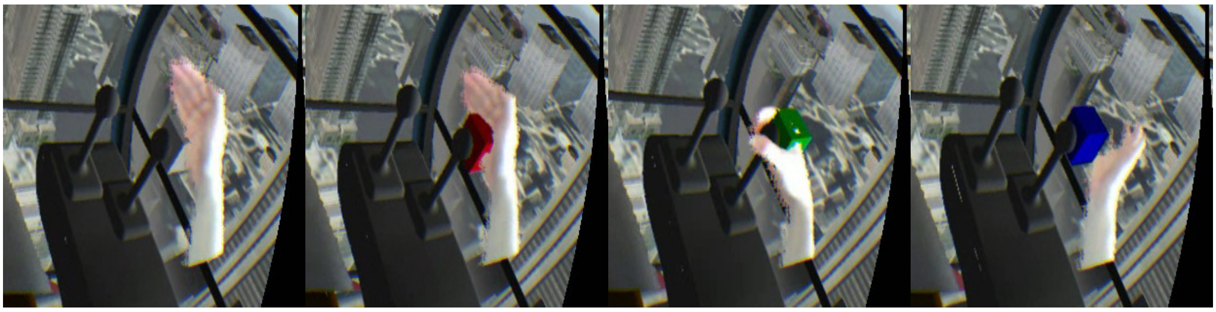

Fig. 5: Screenshots of the user interacting with virtual joystick by proximity-based (through the virtual cube) "collision detection".

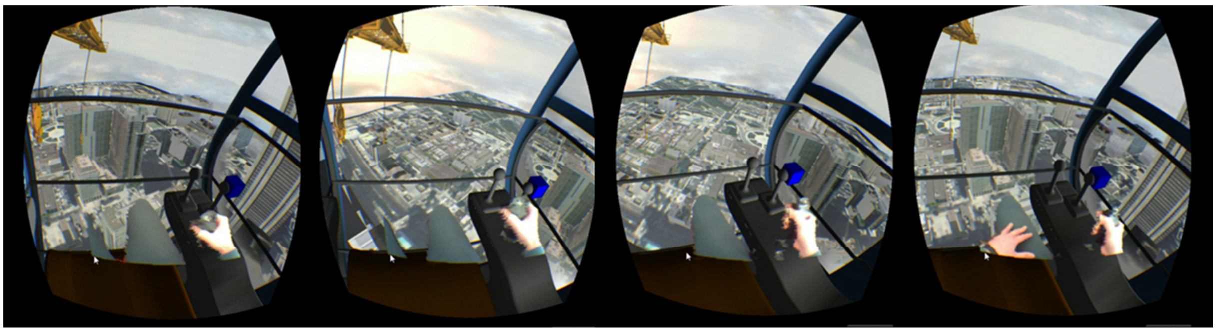

Fig. 6: Views related to the simulated rotation of the crane controlled by means of the real joystick. First and second column: rotation to left. Third and fourth column: rotation to right.

5. CONCLUSIONS

This paper presented the early development of an immersive Hybrid Reality (iHR) system based on commodity hardware that is able to consistently simulate interaction between virtual and real objects. The system is based on the real-time integration of 3D structural data of the real entities, reconstructed through a RGB-D camera, with the virtual content rendered on an immersive stereo HMD. The system takes advantage of the reconstructed 3D data and game engine simulation capabilities to properly compute occlusions and simulate (proximity-based) physical interactions between virtual and real entities. These features can bring a number of benefits in terms of improved sense of presence and a wider range of interaction modalities. In addition, 3D information can be employed for health and safety training, for example detecting imminent collisions or dangerous practices. All these capabilities make our approach particularly suitable for training in the Construction sector.

We demonstrated the potential of our approach through a significant case, i.e. a training simulator for crane operators. Despite being still at an early stage, our system proves to handle effectively a number of crucial issues, e.g. occlusion and real-virtual interaction. We are working to extend it to a complete crane simulator, with real (e.g. joystick) and virtual (e.g. cockpit) parts, interacting among themselves, with a real instructor possibly intervening or providing hands-on aid.

In the future we plan to tackle a number of important limitations still present. In this work just head rotations are tracked, leaving aside positional translations, and accordingly only static-position experiences are simulated. Oculus-Rift DK2 presents (limited) tracking capabilities for tracking translations within small volumes, and could thus be used to extend tracking to 6-DOF in small volumes. Alternatively, integration can be achieved with 6-DOF tracking solutions like traditional IR marker-based tracking systems, as in (Tecchia et al. 2014), or inside-out methods, as in (Carozza et al. 2014b). Quantitative evaluation of subjective user experience on a number of trainees would be used to assess the effectiveness of these future improvements.

6. ACKNOWLEDGMENTS

We are grateful to the Construction Industry Training Board (CITB) for funding this project, and Edinburgh College for their continuous support. However, the views and opinions expressed in this publication are those of the authors and do not necessarily reflect those of CITB and Edinburgh College.

7. REFERENCES

Bae H., Golparvar-Fard M., White J. (2013). High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (AEC/FM) applications. *Visualization in Engineering*, 1(1):1–13, 2013.

Bleser G., Hendeby G., Miezal M. (2011). Using egocentric vision to achieve robust inertial body tracking under magnetic disturbances. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, 103 –109.

Bosché F., Abdel-Wahab M., Carozza L. (2015). Towards a Mixed Reality System for Construction Trade Training, *ASCE Journal of Computing in Civil Engineering*, doi: 10.1061/(ASCE)CP.1943-5487.0000479.

Carozza, L., Tingdahl, D., Bosché, F., van Gool, L. (2014a), Markerless Vision-Based Augmented Reality for Urban Planning. Computer-Aided Civil and Infrastructure Engineering, 29: 2–17.

Carozza L., Bosché F., Abdel-Wahab M. (2014b). Robust 6-DOF immersive navigation using commodity hardware, *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology (ACM VRST 2014)*, Edinburgh, UK, 19-22.

Côté S., Barnard J., Snyder R., Gervais R. (2013). Offline spatial panoramic video augmentation for visual communication in the AEC industry. *In: N. Dawood and M. Kassem (Eds.), Proceedings of the 13th International Conference on Construction Applications of Virtual Reality*, 30-31 October 2013, London, UK.

Gutekunst M., Geuss M., Rauhöft G., Stefanucci J., Kloos U., Mohler B. (2014) A Video Self-avatar Influences the Perception of Heights in an Augmented Reality Oculus Rift, *International Conference on Artificial Reality and Telexistence*, 19th Eurographics Symposium on Virtual Environments (ICAT-EGVE 2014), 9-12.

Heikkila J. and Silven O. (1997). A four-step camera calibration procedure with implicit image correction, *Proceedings of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 1997),*1106-1112, 17-19 Jun 1997.

Khattak S., Cowan B., Chepurna I., Hogue A. (2014). A real-time reconstructed 3D environment augmented with virtual objects rendered with correct occlusion, *Games Media Entertainment (GEM), 2014 IEEE*, pp.1-8, 22-24 Oct. 2014.

Lefloch D., Nair R., Lenzen F., Schäfer H., Streeter L., Cree M.J., Koch R., Kolb A. (2013). Technical Foundation and Calibration Methods for Time-of-Flight Cameras, *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications, Lecture Notes in Computer Science*, vol. 8200, 3-24.

Lin Q., Rieser J.J., Bodenheimer B. (2013). Stepping off a ledge in an HMD-based immersive virtual environment. *Proceedings of the ACM Symposium on Applied Perception* (SAP '13). ACM, New York, NY, USA, 107-110.

McGill M., Boland M., Murray-Smith R., Brewster S. (2015). A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 2143-2152.

Milgram, P., and Colquhoun, H. (1999). A taxonomy of real and virtual world display integration. *Mixed reality: Merging real and virtual worlds,* 5–30.

Steptoe W., Julier S., Steed A. (2014). Presence and discernability in conventional and non-photorealistic immersive augmented reality, *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pp.213-218.

Tecchia F., Avveduto G., Brondi R., Carrozzino M., Bergamasco M., Alem L. (2014). I'm in VR!: using your own hands in a fully immersive MR system, *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (VRST '14), 73-76.

Turk G. and Levoy M. (1994) Zippered Polygon Meshes from Range Images, *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, 311-318.

Zhiwei Zhu, Branzoi, V., Wolverton M., Murray G., Vitovitch, N., Yarnall L., Acharya G., Samarasekera S., Kumar R. (2014). AR-mentor: Augmented reality based mentoring system. *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pp. 17-22, 10-12 Sept. 2014

Webel S., Bockholt U., Engelke T., Gavish N., Olbrich M., Preusche C. (2013). An augmented reality training platform for assembly and maintenance skills, *Robotics and Autonomous Systems*, Vol. 61, No. 4, April 2013, 398-403.