Towards a Mixed Reality System for Construction Trade Training

Dr. Frédéric Bosché[1],*, Dr. Mohamed Abdel-Wahab[2],

Dr. Ludovico Carozza[3]

Abstract

Apprenticeship training is at the heart of government skills policy worldwide. Application of cutting edge Information and Communication Technologies (ICTs) can enhance the quality of construction training, and help in attracting youth to an industry that traditionally has a poor image and slow in up-taking innovation. This article reports on the development of a novel Mixed Reality (MR) system uniquely targeted for the training of construction trade workers, i.e. skilled *manual* workers. From a general training viewpoint, the system aims to address the shortcomings of existing construction trades training, in particular the lack of solutions for enabling trainees to train in realistic and challenging site conditions whilst eliminating Occupational Health and Safety risks. From a technical viewpoint, the system currently integrates state of the art Virtual Reality (VR) goggles with a novel cost-effective 6 degree-of-freedom (DOF) head pose tracking system supporting the movement of trainees in room-size spaces, as well as a game engine to effectively manage the generation of the views of the virtual 3D environment projected on the VR goggles. Experimental results demonstrate the performance of the proposed 6-DOF head pose tracking system, which is the main computational contribution of the work presented here. Then, preliminary results reveal its value to enable trainees to experience construction site conditions, particularly being at height, in different settings. Details are provided regarding future work to extend the system

[1] Assistant Professor, School of the Built Environment, Heriot-Watt University.
[2] Assistant Professor, School of the Built Environment, Heriot-Watt University.
[3] Research Associate, School of the Built Environment, Heriot-Watt University.
* Corresponding author: f.n.bosche@hw.ac.uk

into the envisioned full MR system whereby a trainee would be performing an actual task, e.g. bricklaying, whilst being immersed in a virtual project environment.

Keywords: Apprenticeship; construction; trade; training; mixed reality; occupational health and safety; work at height; productivity monitoring.

Introduction

Given the on-going development in new technologies (such as, Building Information Modelling (BIM) and green technologies), investment in training becomes essential for addressing the industry's evolving skills needs. It is also imperative to ensure that there are sufficient numbers of new entrants joining the construction industry to support its projected growth. Latest figures from the UK Office of National Statistics (ONS) reveal a 2.8% growth in the third quarter (Q3) of 2013 (ONS, 2013). A sustained investment in construction apprenticeship training thus becomes essential.

In the UK, the Construction Industry Training Board (CITB) retains a unique position by administering a Levy/Grant scheme (LGS) on behalf of the construction industry – as mandated by the Industrial Training Act 1964. It raises approximately £170m annually from training levies which is re-distributed to the industry in the form of training grants. Approximately 50% of the levy is spent on training grants for apprenticeships in order to attract, retain and support new entrants into the industry. However, the UK Government's 'Skills for Growth' white paper similarly called for: 1) Improving the quality of provision at Further Education (FE) colleges and other training institutions, and 2) Developing a training system that provides a higher level of vocational experience; one that promotes a greater mix of work and study (Department for Business, Innovation and Regulatory Reform, 2009). And recently, the UK Minister for Universities and Science, David Willetts, announced the introduction of tougher standards to drive up apprenticeship quality – a view which was echoed by the Union of Construction, Allied Trades and Technicians (UCATT) (BIS, 2012; and Davies 2008).

Globally, the International Labour Organization (ILO) urges governments worldwide to upgrade the skills of master crafts-persons and trainers overseeing the apprenticeships and ensure that apprenticeships provide a real learning experience (ILO, 2012). Clearly, enhancing the quality of apprenticeship training in-line with the industry's evolving skills needs is paramount for supporting its future development and prosperity.

Along with other researchers and experts, the authors argue that novel technology can enhance trainee experience, improve training standards, eliminate or reduce health and safety risks, and in turn induce performance improvements on construction projects. For example, simulators for equipment operator training allow testing trainees to ensure that they can demonstrate a certain skill level prior to start working. A company developing novel technologies for the mining industry has claimed that, as a result of using simulators, there was a 20% improvement in truck operating efficiency and reduction in metal-to-metal accidents (Immersive Technologies, 2008).

Yet, the construction industry has been traditionally slow in the uptake of innovation, particularly in areas such as ICT (Egan Report, 1998). For this reason, innovation in construction continues to be at the top of the UK government (UK Government, 2011; UK Government, 2013).

This article reports on the development of a novel Mixed Reality (MR) system using state-of-the-art Head-Mounted Display (HMD) and 6 Degree-Of-Freedom (DOF) head motion tracking technologies. The overarching aim of the MR system is to enable construction trade trainees to safely experience virtual construction environment while conducting real tasks, i.e. while conducting real manual activities using their actual hands and tools, just as they currently do in college workshops. Figure 1 illustrates the concept of the MR system where the trainee experiences height in a virtual environment whilst performing the task of bricklaying.

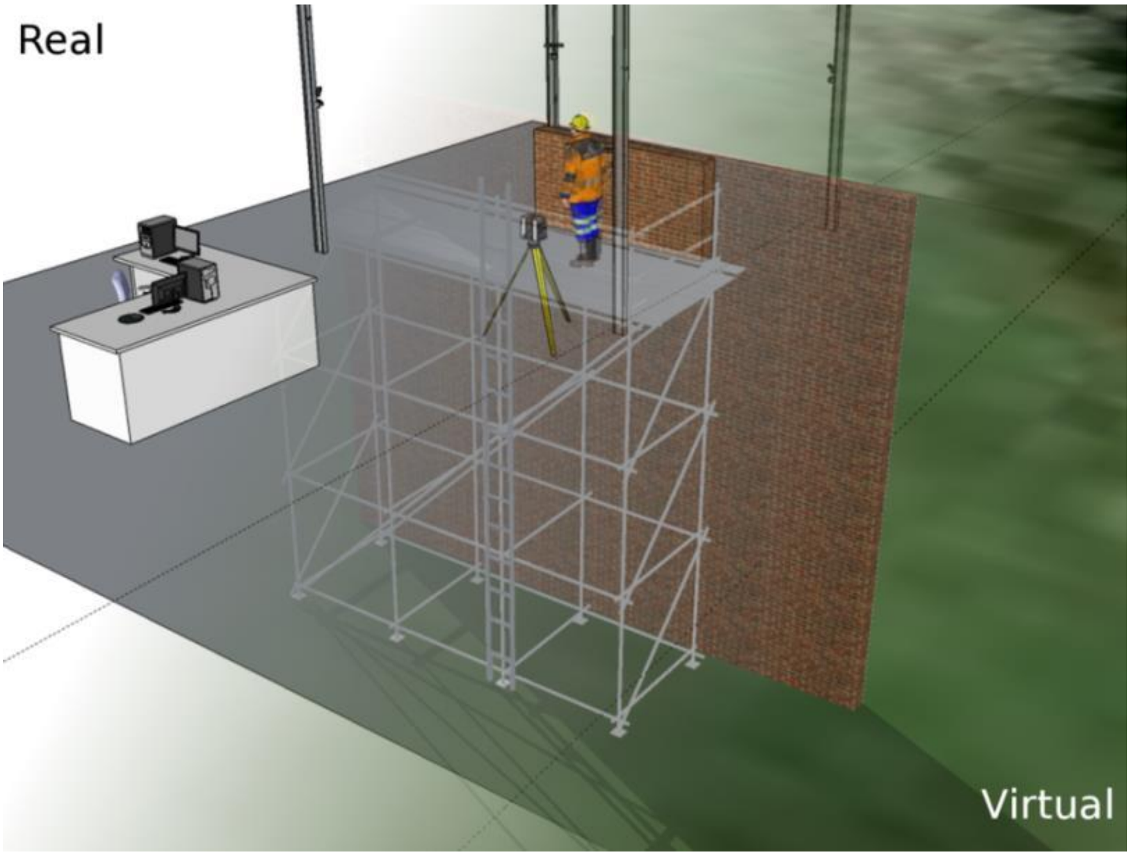


Figure 1: Illustration of the use of the proposed MR environment to immerse trainees and their work within a "work at height" situation. Here the trainee conducts bricklaying works on the floor of the college lab (safe), but experiences conducting the activity on a high scaffold (situation with safety risks).

The piloting of the proposed MR system mimics working at height in a construction site environment. The focus is on height simulation as falling from height accounts for nearly 50% of the fatalities in the UK with falls from edges and opening account for 28% of falls, followed by falling from ladders (26%), and finally scaffolding and platforms (24%) (HSE, 2010). Similarly, in the USA, the most common types of falls from heights in the construction industry are falling from a scaffold and ladder (Rivara and Thompson, 2000). The construction sector is particularly impacted because many construction-related trades involve working at height, such as scaffolders, roofers, steel erectors, steeple-jacking, painting and decorating. Furthermore, ironically for H&S reasons, colleges can often not train trainees at heights above 8m. It is hoped that the proposed system enhances the quality of training provision by providing trainees an exposure to construction site conditions through simulation, so that they are better prepared to working on site and the likelihood of accidents is reduced (through better perception of hazards on site).

The paper commences with a literature review of the current applications of MR in construction training, which leads to identification of the need for different MR systems better suited to the needs of construction trade training. This is then followed by the presentation of the on-going development of such an MR system. The current system is only a VR system, but includes several of the functional components that will be required in the final MR system. The presentation particularly focuses on the main computational contribution that is a robust, cost-effective 6-DOF Head Tracking system. The performance of the current system is experimentally assessed in challenging scenarios. Finally, strategies are discussed for the completion of the envisioned MR system.

'Reality-Virtuality' continuum of construction training

Figure 2 depicts a 'Reality-Virtuality' continuum in the context of construction training, highlighting the training environments where construction training takes place. This section summarizes developments that have been made at different stages within this continuum, starting with training in real environments, followed by training using Virtual Reality systems, and finally training using Mixed Reality systems.

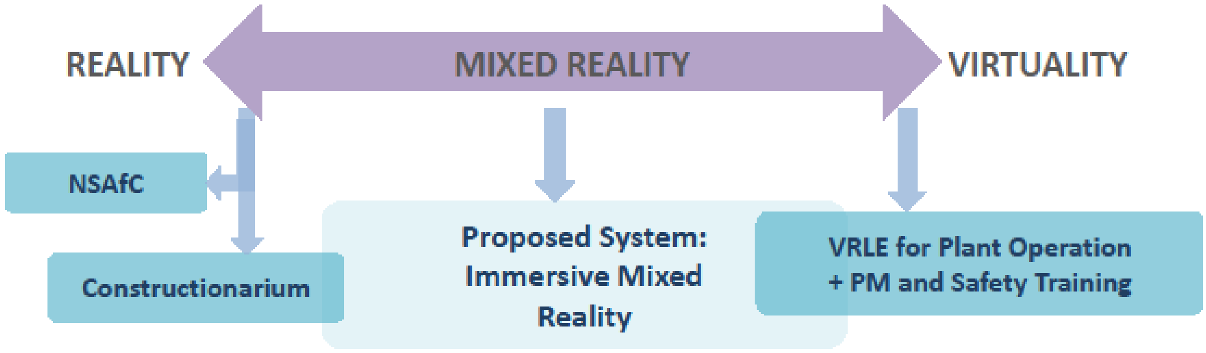


Figure 2: Reality-Virtuality Continuum in the context of construction training ().

Real Environment

At one end, there is training within *'Real'* construction project environment. For example, the UK CITB has set-up the National Skills Academies for Construction (NSAfC) with the aim of providing project-based training that is driven by the client through the procurement process. NSAfC included projects such as the 2012 Olympic which provided 460 apprenticeship opportunities.

However, training on real construction projects is constrained by the type of activity taking place on site, project duration, in addition to (occupational) health and safety (H&S) risks. Trainees may not be allowed to perform certain tasks on real projects as this can cause delays and errors can be costly, especially when it comes to high profile projects such as the Olympics. To address this issue, attempts have been made in recent years to 'simulate' real project environments where trainees can conduct real tasks without compromising project performance and H&S.

An example is 'Constructionarium' in the UK which is a collaborative framework where university, contractor and consultant work together to enable students to physically construct scaled-down versions of buildings and bridges (Ahearn, 2005). This enables students to experience the various construction processes and associated challenges that cannot be learned in a traditional classroom setting. Auburn University in the US, and the University of Technology Sidney in Australia have run similar schemes (Burt, 2012; Forsythe, 2009).

As for construction trade training, apprentices typically train in a FE college's workshop. The FE college training counts towards their attainment of a vocational qualification, which also includes work placement. However, it must be noted that training at FE's workshop is constrained by the space provided at the college and the requirements set-out in the National Occupational Standards – whereby trainees can only experience heights up to 8m, which is not representative to working at higher heights on many construction projects, such as high-rise buildings or skyscrapers.

Virtual Reality (VR)

At the other end of the 'Reality-Virtuality' continuum (Figure 2), *Virtual Reality* (VR) is increasingly used for construction training. VR development boomed in the 1990's and VR is in fact still under intense development, with education and training an important area of application. Mikropoulos and Natsis (2011) define a Virtual Reality Learning Environment (VRLE) as "*a virtual environment that is based on a certain pedagogical model, incorporates or implies one or more didactic objectives, provides users with experiences they would otherwise not be able to experience in the physical environment and can support the attainment of specific learning outcomes.*"

VRLEs must demonstrate certain characteristics that were summarized by Hedberg and Alexander (1994) as: *immersion*, *fidelity* and *active learner participation*. Other terms employed to refer to these characteristics are *sense of presence* (Winn and Windschitl, 2000) and *sense of reality*.

VRLEs can be classified as: *Desktop*, where the user interacts with the computer generated imagery displayed on a typical computer screen; or *Immersive*, where the computer screen is replaced with a HMD or other technological solutions attempting to better 'immerse' the participant in the (3D) virtual world (Bouchlaghem *et al.*, 1996). Most current *simulators* are VRLEs that are commonly developed for *plant operation training* (e.g. tower cranes, articulated trucks, dozers and excavators). For example, Volvo Construction Equipment (Volvo CE, 2011) and Caterpillar have developed simulators for training on their range of heavy equipment, such as excavators, articulated trucks and wheel loaders (Immersive Technologies, 2010).

Equipment simulators enable training in realistic construction project scenarios with high-fidelity, which is made possible by force feedback mechanisms, and without exposing trainees or instructors to occupational H&S risks. They support fast and efficient learning thereby increasing trainees' motivation (Volvo CE, 2011; TSPIT, 2011). For example, the ITAE simulator, employed in mining equipment operation training, is used to ensure that apprentices can demonstrate a certain skill level prior to working in mines. The manufacturer claims that the simulator has proved to be effective in modifying and improving operators' behaviour as well as enhancing the existing skills levels and performance of employees (Immersive Technologies, 2008).

VRLEs have also been developed for *supervision/management training*. The first UK construction management simulation centre has opened at Coventry University in 2009 and is known as ACT-UK (Advanced Construction Technology Simulation Centre). The centre is aimed at already practicing foremen and construction managers, and potentially students (Austin and Soetanto, 2010; ACT-UK, 2012). Similar centres exist with the Building Management Simulation Center (BMSC) in The Netherlands (De Vries *et al.*, 2004; BMSC,

2012) or the OSP VR Training environment collaboratively developed as part of the Manubuild EU project (Goulding *et al.*, 2012). In these VRLEs, trainees can be partially immersed in simulated construction site environments to safely expose them to situations that they must know how to deal with appropriately. These may include H&S, work planning and coordination, or conflict resolution scenarios (Harpur, 2009; Ku, 2011; Li, 2012). Other VRLEs have also been investigated for other applications for enhancing communication and collaboration during briefing, design, and construction planning (Duston, 2000; Arayici, 2004; Bassanino, 2010).

VRLEs can generally provide significant benefits over traditional ways of training and learning. The main benefit is to enable trainees to "*cross the boundary between learning about a subject and learning by doing it, and integrating these together*" (Stothers, 2007). A simulated working environment enables skills to be developed in a wide range of realistic scenarios, but in a safe way (Stothers, 2007; Austin and Soetanto, 2010).

Nonetheless, despite the general agreement on the potential of VRLEs to enhance education, Mikropoulous (2011) and Wang and Dunston (2005) noted that there is a general lack of thorough demonstration of the value-for-money achieved by those systems, which may be due to implementation cost, but possibly also to the quantity and quality of training scenarios that could be developed and their impact on learning and practice.

It is interesting to note that VRLEs and Constructionarium are two learning approaches at the opposite ends of the continuum and may be regarded as complementary. Arguably, a blended learning approach can be employed whereby VRLEs are used for initial learning exercises, and approaches like Constructionarium are used for subsequent more real learning-by-doing activities and thereby supporting the transition before going on-site.

Mixed Reality (MR)

Within the Reality-Virtuality continuum, *Mixed Reality* (MR), sometimes called Hybrid Reality, refers to the different levels of combinations of virtual and real objects that enable the production of new environments and visualisations where physical and digital objects co-exist and interact in real time (De Souza e Silva and Sutko, 2009). Two main approaches are commonly distinguished within MR. *Augmented Reality* (AR) specifically refers to situations when computer-generated graphics are overlaid on the visual reality, while *Augmented Virtuality* (AV) specifically refers to when real objects are overlaid on computer graphics (Milgram and Colquhoun, 1999).

MR has a distinct advantage over VR for delivering both immersive and interactive training scenarios. The nature and degree of interactivity offered by MR systems can provide a richer and superior user experience than purely VR systems. In particular, in contrast to VR, MR systems can support more direct (manual) interaction of the user with real and/or virtual objects, which is key to achieve active learner participation and skill acquisition (Wang and Dunston, 2005; Pan *et al.*, 2006). However, developments in MR are more recent and still in their infancy, essentially because of the higher technical challenges surrounding specific display devices, motion tracking, and conformal mapping of the virtual and real worlds (Martin *et al.*, 2011).

With regard to construction training, MR systems reported to date mainly focus on equipment operator training, with human-in-the-loop simulators. According to the definitions above, these simulators can be considered as AV systems. For example, Keskinen *et al.* (2000) developed a training simulator for hydraulic elevating platforms that integrates a real elevator platform mounted on 6-DOF Stewart platform with a background display screen for visualization of the virtual environment. Standing on the platform, the operator moves it within the virtual environment using its actual command system and receives feedback stimuli through the display and the Stewart platform.

Noticeably, this and other similar AV-type systems are not fully immersive and thus, from a visual perspective, do not provide a full sense of presence. In an attempt to address this limitation, Wang *et al.* (2004) have proposed an AR-based Operator Training System (AR OTS) for heavy construction equipment operator training. In this system, the user operates a real piece of equipment within a large empty space, and feels that s/he and the piece of equipment are immersed in a virtual world (populated with virtual materials) displayed in AR goggles. However, this system appears to have remained a concept, with no technical progress reported to date.

To the knowledge of the authors, no work has been reported to date on developing MR systems for the training of construction trades, (e.g. roofing, painting and decorating, bricklaying, scaffolding, etc.). The particularity of those trades is that the trainee must be in direct manual contact with tools and materials. Immersing their work thus requires specific interfaces for tracking the limbs of trainees (particularly the arms and hands), and integrating the manipulations with virtual environments.

Research has been widely conducted to develop such interfaces. Haptic gloves or other worn devices are investigated (Tzafestas, 2003; Buchmann *et al.*, 2004), but are invasive. Non-invasive vision-based body tracking solutions have also been considered (Hamer *et al.*, 2010), but are usable only within very small spaces. Thus, despite continuous improvements, current solutions for manual interactions with virtual environments do not provide the richness and interactivity required for effective trade training.

In addition, there is a strong argument that MR should not (yet) be used for virtualizing ‘manual’ tasks; traditional training approaches using real manipulation of real materials and tools must remain the standard. Instead, MR could be solely focused on enabling existing students training in college workshops to develop their skills within challenging realistic site conditions, such as working at height. In other words, MR should be used to immerse both ‘trainees and their manual tasks’ in varying and challenging virtual environments.

As mentioned earlier, construction site experience is a vital and integral part of apprenticeship training and therefore MR technology could help in preparing trainees for actual site conditions. However, it should be viewed as complementary to real site experience and not a replacement. It could be used as a transition to establish the trainees’ readiness before they can actually go on-site.

Need Identification, Functional Analysis, and Current System

It was concluded in the previous section that construction trade training can benefit from MR by employing it solely to visually immerse trainees, while they conduct training activities with real tools and materials. Referring to the taxonomy of Milgram *et al.* (1994; 1999), the type of system required appears to correspond to MR systems they classify as *Class 3* or *Class 4* (see Table 1). However, the authors also observe that, from a visualization viewpoint, this more specifically requires that the trainee be able to see their real body and real work (tools, material), and see these immersed within a virtual world. This means that the system would have to calculate in real-time in which parts of the user’s field of view the virtual world must be overlaid on the real world, and in which parts it should not. In other words, the system needs to deliver AR functionality with (local) occlusion handling, which requires that the 3D state of the real world be known accurately and in real-time (the 3D state of the virtual world is naturally already known). Referring again to the taxonomy of Milgram *et al.* (1994; 1999), the type of system required thus needs to have an *Extent of (Real) World Knowledge (EWK)* where the depth map of the real world from the user's viewpoint is completely modelled (see Figure 3).

Table 1: Some major differences between classes of Mixed Reality (MR) displays: reproduced from Milgram *et al*. (1994).

Class of MR System	Real (R) or Computer Generated (CG) world	Direct (D) or Scanned (S) view of substrate	Exocentric (EX) or Egocentric (EG) reference	Conformal mapping (1:1) or not (1:k)
1. Monitor-based video, with CG overlays	R	S	EX	1:k
2. HMD-based video, with CG overlays	R	S	EG	1:k
3. HMD-based optical see-through, with CG overlays	**R**	**D**	**EG**	**1:1**
4. HMD-based video see-through, with CG overlays	**R**	**S**	**EG**	**1:1**
5. Monitor/CG-world, with video overlays	CG	S	EX	1:k
6. HMD/CG-world with video overlays	CG	S	EG	1:k
7. CG-based world with real object intervention	CG	D, S	EG	1:1

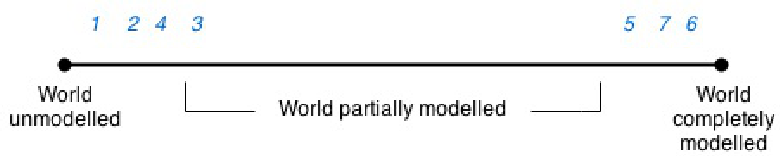


Figure 3: Extent of World Knowledge (EWK) dimension; reproduced from Milgram *et al*., (1994).

From this analysis, the authors have derived a system's process that includes five specific functionalities and corresponding components (Figure 4):

- *6-DOF head tracker*: provides the 3D pose (i.e. location and orientation) of the user's head in real-time;
- *Depth sensor*: provides a depth map of the environment in the field of view of the user;
- *Virtual World Simulator / Game Engine*: simulates the virtual 3D environment and is used to generated views of it from given locations;
- *Processing Unit*: uses the information provided by the three components above to calculate the user's views of the mixed real and virtual worlds to be displayed in the HMD in real-time;
- *HMD (preferably, but not necessarily, see-through)*: is used to display the views generated by the Processing Unit.

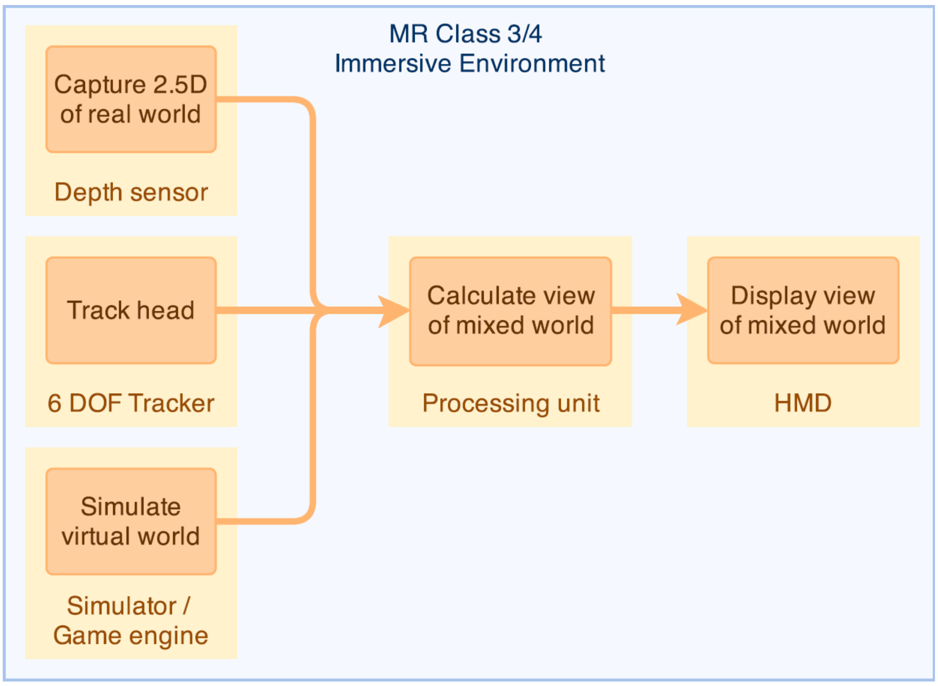


Figure 4: Process and associated components for delivering the envisioned immersive MR environment.

In the following, progress to date is reported, that involves the implementation of four of the five components above:

- *6-DOF Head Tracker*: The 6-DOF head tracking (i.e. localization) is probably the most critical functionality to be delivered by real-time MR systems. Localization is even more critical for MR systems than for VR systems, because poor pose tracking is far more disturbing in MR scenarios since these require the virtual display content to be very accurately aligned with the reality. Robust localization is critical to user experience.
  Guaranteeing continuous operation while the user is moving is already a challenge; doing it without requiring complex and expensive set-up, is an even greater one. The main contribution reported in this paper is an original cost-effective visual-inertial 6-DOF head tracker. The system is detailed in the section below, and its performance is particularly assessed in the experiments reported later on.
- *Game Engine*: The proposed 6-DOF Head Tracking system is integrated as a third-party component to the Unity 3D game engine (Unity 3D, 2014). This gives the proposed approach a wider applicability and scalability to a range of different training scenarios, thus providing flexibility to different operative trades. Game engines also have the important advantage of already providing optimized capabilities for high-quality rendering and user interaction within complex virtual environments.
- *HMD*: Our system currently employs the Oculus Rift (Oculus, 2013) that is a non-see-through HMD, i.e. VR, device that offers great immersive experience with a 110° field of view.
- *Processing Unit*: as discussed below, the Depth Sensing component has not been implemented yet. As a result, our current system can only deliver VR functionality, not AR. Therefore, the Processing Unit is currently only partially implemented, as it

only calculates views of the virtual 3D environment (managed by the Game Engine) to be displayed on the HMD.

At this stage, no solution for the Depth Sensing component has been implemented. However, a solution is proposed in the *Future Work* section at the end of this paper. Similarly, the authors' envisioned system needs to deliver AR, not just VR functionality. The authors' proposed approach to achieve this is also discussed in the *Future Work* section.

As mentioned above, out of the four components implemented to date, the 6-DOF Head Tracking component is the most challenging. The approach developed is a significant computational contribution, and this paper thus particularly focuses on presenting it and assessing its performance. The following section presents the approach.

6-DOF Head Tracker

This section is divided in two sub-sections. The first sub-section provides a short review of prior works on localization methods, identifying their strengths and weaknesses. The second sub-section presents the proposed visual-inertial approach.

Introduction

Numerous absolute position tracking technologies exist, but some either do not work indoor (e.g. GNSS; e.g. see the work of Kamat et al. (Talmaki and Kamat, 2014)) or do not provide the level of accuracy necessary for MR applications (e.g. UWB, RFID, Video, depth sensors) (Teizer and Vela, 2009; Gong and Caldas, 2009; Cheng et al, 2011; Yang et al., 2011; Escorcia et al., 2012; Ray and Teizer, 2012; Teizer et al., 2013). In construction, Vision-based approaches with multiple tracked markers, such as commonly considered Infrared-Red vision-based systems, can provide accurate 6-DOF data, but require significant infrastructure (cost), line-of-sight, and are somewhat invasive. Inertial Measurement Units (IMU), that integrate numerous sensors like gyroscopes, accelerometers, compass, gravity sensor, and magnetometer, are mainly used to track orientation. Although IMUs can theoretically also be used to track translation, the authors' experience (see Section *Experimental Results*), as well as that of others (e.g. see (Borenstein et al., 2009)), is that this is prone to rapid divergence, hence unreliable information.

In an effort to address these limitations, an alternative visual-inertial approach for 6-DOF position tracking has been investigated that integrates an IMU and a markerless vision-based system. Visual-inertial *ego-motion* approaches have been conceived in general to represent an affordable technology, also usually requiring limited set-up. Complementary action of visual and inertial data can increase robustness and accuracy in determining both position and orientation even in response to faster motion (Welch and Foxlin 2002, Bleser and Stricker 2008). The proposed specific approach, detailed in the following section, has been designed to handle system outages and deliver continued tracking at the required quality.

Proposed Approach

The proposed head tracking system relies on the complementary action of visual and inertial tracking. The authors have conceived an *ego-motion (or inside-out)* localization approach, which integrates visual data of the surrounding environment (training room), acquired by a monocular camera mounted integral with the HMD *Oculus Rift* (the first version is used), together with inertial data provided by the IMU embedded into the HMD *Oculus Rift*. A dedicated computing framework robustly integrates this information, providing in real-time a stable estimation of the position and the orientation of the trainee's head.

As far as the visual approach is concerned, it provides *global references* that can be used for localizing from scratch the trainee's head within the training room, also recovering its pose in case of system outage. Following the general markerless vision-based approach proposed in (Carozza *et al.*, 2014a), the method proposed here puts in place new computational strategies in order to increase the robustness (e.g., for fast motion) and the responsiveness of the system. Indeed, in order to deliver a consistent user experience, system outages, as well as drift and jitter effects, must be minimized for general motion patterns. The proposed method follows two main stages, i.e. an *off-line reconstruction stage* and *on-line localization stage*, as outlined in Figure 5.

Off-line Reconstruction Stage

The *off-line reconstruction stage* (Figure 5 left) is performed in advance, once and for all, by automatically processing pictures of the training room, acquired by the camera from different viewpoints, according to the Structure from Motion *Bundler* framework (Snavely 2008). The training room has been textured in advance by using posters (Figure 5 (a)) – with a random layout – so that a *3D map of visual references* can be reliably reconstructed (Figure 5 (b)). The reconstructed point cloud is then used as reference for the alignment of the virtual training scenario with the (real) world reference frame (Figure 5 (c)).

A multi-feature framework has been developed so that it is possible to associate different *visual descriptors*, with flexible performance in terms of robustness and time processing, to the reconstructed 3D point cloud. Based on the recent comparative evaluation of visual features' performance (Gauglitz 2011), SURF (Bay *et al.* 2008) and BRISK (Leutenegger *et al.* 2011) descriptors have been evaluated.

The result of the process above is a database of repeatable visual descriptors, referred in the 3D space, or world reference frame (WRF), and that is used for the subsequent on-line localization stage.

On-line Localization Stage

At the beginning of *on-line* operations, visual features extracted from the images acquired by the camera mounted on the HMD (Figure 5 (d)) are robustly and efficiently matched with the visual features stored in the map, so that the *global pose* of the camera can be estimated from the resulting 2D/3D correspondences (Figure 5 (e), left) by means of *camera resectioning (*Hartley and Zissermann, 2003*)*. In particular, for each frame the set of query descriptors is matched through *fast approximate nearest neighbour* search over the whole room map, and the 3-point algorithm (Haralick, 1994) is applied on the set of inliers resulting from a robust RANSAC (Fischler and Bolles, 1981) filtering stage. In this way, the system is *initialized* to its starting *absolute pose* $P_{WRF}^{-} = (p_{WRF}, R_{WRF})$, where p_{WRF} and R_{WRF} are respectively the position vector and the orientation matrix with respect to the WRF.

However, the global matching approach can be (a) not sufficiently precise and robust, due to image degradation during fast movements, or (b) not sufficiently efficient for real-time performance (due to query search overhead over the whole database). Accordingly, a *feature tracking* strategy is used together with the IMU data for the subsequent frames. A frame-to-frame tracking approach based on the Kanade-Lucas-Tomasi (KLT) tracker (Shi and Tomasi 1994) is employed between consecutive frames, with the advantage of being very efficient and exploiting spatio-temporal contiguity to track faster motions. More details about the feature tracking approach, and in particular *tracker reinitialization* to allow tracking over long periods, can be found in (Carozza *et al.*, 2013). Note that a pin-hole camera model is considered throughout all the stages of the vision-based approach, taking into account also lens radial distortion.

Inertial data are used jointly with the visual data in an Extended Kalman Filter (EKF) framework (Figure 5 (e)). This framework is necessary to filter the noise affecting both information sources and provide a more stable and smoother head trajectory. A *loosely-coupled sensor fusion* approach has been implemented, which initially processes *separately* inertial and visual data to achieve a robust estimate of the *orientation* and a set of *visual inliers*. Then, this information is fused together into the EKF to estimate the *position*. The *measurement equations* used in the EKF involve the visual 2D/3D correspondences according to the camera (non-linear) projective transformation, $\Pi(P_{WRF}^{-})$, related to the predicted pose $P_{WRF}^{-} = (p_{WRF}, R_{WRF})$, by computing the *predicted projections* m^{-} of the 3D points X onto the image plane:

$$m^{-} = \Pi(P_{WRF}^{-})X$$

The *loosely-coupled approach* has the advantage of decoupling position and orientation noises, so that the system is inherently more immune to pose divergence possibly rising from non-linearities inherent in the projective model.

However, in order to be fused consistently with the visual data, the inertial data must be referred to the same absolute reference frame of the visual data (i.e. the training room). The authors developed an on-the-fly *camera-IMU calibration* routine, which automatically processes the first N_{calib} pairs of visual and inertial data following the very first successful initialization to estimate the *calibration matrix* relating the inertial reference frame to the global reference frame. The calibration method we employ is similar to the classic *hand-eye* calibration (see Lobo *et al.* 2007), but it can be employed on-line since the relative translation between the camera and the IMU centres does not need to be estimated (it is not taken into account into the subsequent calculations).

It is worth noting that the IMU measures represent the only data available in case of outage of the visual approach, due to image degradation, poor texturing, or occlusion, for example. In these cases, the method relies on the sole orientation information measured by the IMU (*Tracking_IMU*), while data measured from the accelerometers are not directly employed to estimate position, which would rapidly result in positional drift. Among the different approaches applicable in this situation, the authors decided to assume the position fixed and invoke frequently a *relocalization* routine.

During the *relocalization* stage, the matching approach employed for *initialization* is applied on the map points only within an expanded camera frustum, computed from the last successfully computed pose. This guided search has the advantage of being significantly faster. If the *relocalization* fails, the system enters the *Tracking_IMU* state for N_{lost} consecutive relocalization attempts at maximum, then invoking the *inizialization*.

In Figure 6, the state diagram of the adopted 6-DOF tracking framework summarizes the main transitions occurring during on-line operations among the different stages encountered above. These transitions illustrate at a high level the continued operation of the system over long periods from the initialization to the response and recovery from different system outages.

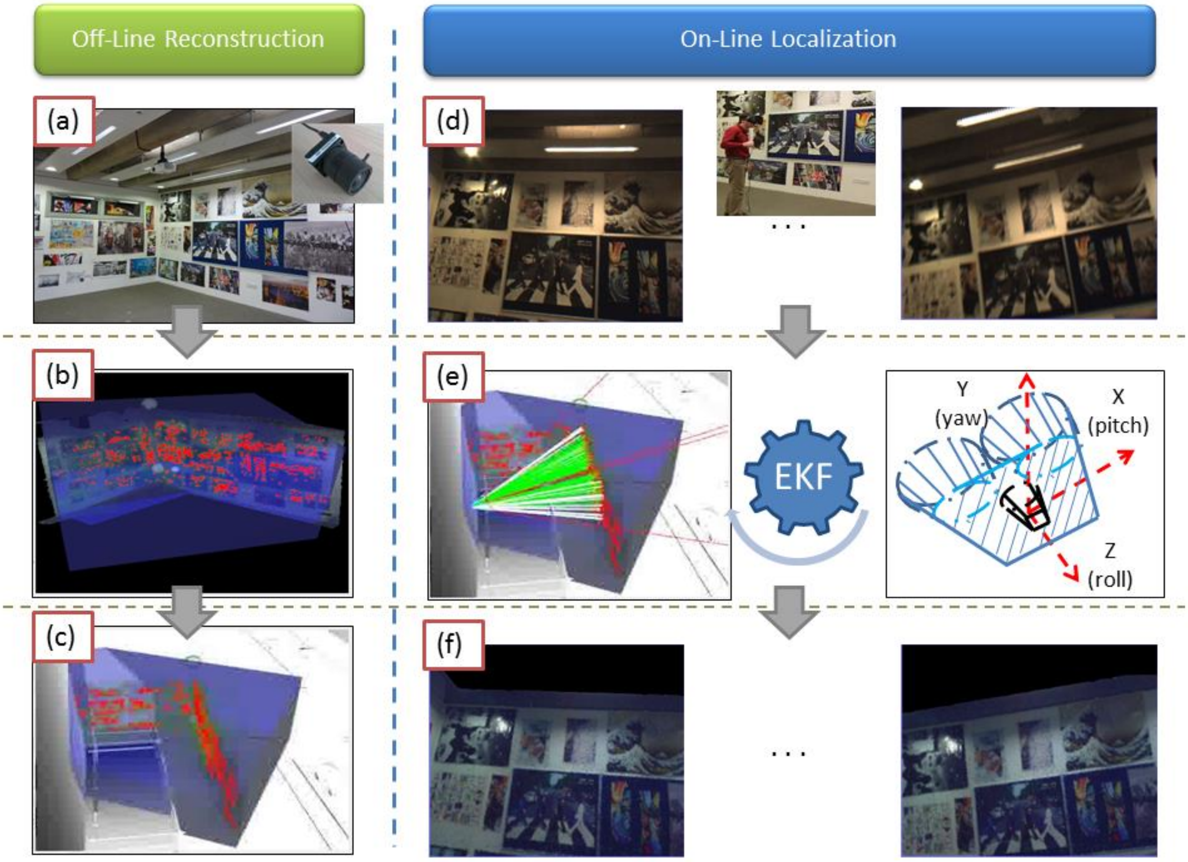


Figure 5: An overview of the main components of the proposed approach to 6-DOF head tracking and HMD-based immersion.

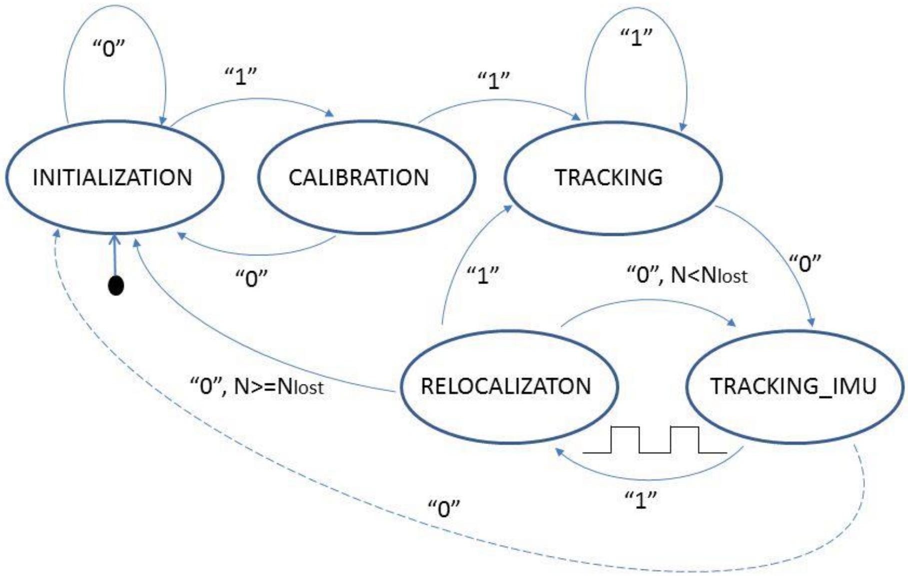


Figure 6: State diagram of the visual inertial 6-DOF tracking framework. “1” and “0” represent successful or unsuccessful state execution, respectively.

Finally, for each frame, once the head pose is estimated, any 3D graphic model/virtual environment can be rendered consistently with the estimated viewpoint. For example, Figure 5 (f) shows the rendered views of a virtual model of the training room corresponding to the head locations estimated using the two head-mounted camera views shown in Figure 5 (d).

The authors acknowledge that vision-based location systems have the limitation of requiring line-of-sight to sufficiently textured surfaces. However, the proposed system is targeted towards controlled environments for which the surrounding boundary walls can be appropriately textured as needed. Furthermore, the inertial system increases the robustness of the system by taking over orientation tracking upon failure of the vision-based system (that is reinitialized as frequently as possible).

Experimental Results

In this section, results are first reported on the performance of the proposed 6-DOF head tracking system. This is then followed by results from the current full system in action, that integrates the head tracking system with a VR Immersive Environment that uses the Unity game engine to manage the virtual 3D model (game environment / simulation) and generate the views of it in real-time, and the *Oculus Rift* to display these views.

All the experiments were performed in a rectangular room of size 3.75 m x 5.70 m with walls covered with posters arranged according to a random layout. Note, however, that these experiments are only part of a series of experiments that have been conducted in different rooms with varying poster arrangements and geometrical structures, that have shown no substantial difference in performance (e.g. see (Carozza *et al.*, 2013)).

Head Tracking

The proposed 6-DOF head tracking approach has been tested on several different live sequences, showing real-time performance (30 fps on the average on a Dell Alienware Aurora PC) and an overall good robustness to user movements, as detailed below.

The off-line reconstruction process has led to a *map* of 3,277 SURF and 2,675 BRISK descriptors, respectively, which present different spatial accuracy and distribution.

To assess localization performance, a virtual model of the room has been reconstructed by re-meshing a laser-scan acquisition of the room and aligning this mesh with the 3D feature database. This virtual model enables the rendering of the view of the room for each computed location, which can then be visually compared with the real view of the room from the camera image to assess localization performance (Figure 5, left, third row).

Table 2 presents the statistics related to the on-line performance for a looping path sequence of 2 minutes (3,600 frames) for BRISK and SURF features, respectively (shown in Figure 7). The sequence contains significant motion patterns (e.g. rapid head shaking and bending) to assess the robustness of the method while the user is free to move. The table lists, for the two different types of visual features, the number of frames ($\#F_{Loc}$) successfully localized by the visual-inertial sensor fusion approach as well as the number of frames ($\#F_{IMU}$) for which the visual information is deemed unreliable (e.g. due to fast motion blur, occlusion, poor texturing) and the IMU information only is used (*Tracking_IMU*). The table also provides the computational times achieved for visual matching (i.e. *initialization* and *relocalization*) (T_M), and visual-inertial tracking (T_T). As it can be seen, the BRISK approach provides in general better resilience to visual outages, also because of its better computational performance (T_M) during visual matching (third column of Table 2).

Table 2: Statistics related to the on-line performance for a looping path sequence of 2 minutes (3,600 frames), using either BRISK or SURF features. The table lists the number of frames localized by the sensor fusion approach ($\#F_{Loc}$), and in the TRACKING_IMU mode ($\#F_{IMU}$), together with related timings (in ms, mean±std.dev.) for visual matching (T_M) and visual-inertial tracking (T_T).

Map	*$\#F_{Loc}$*	*$\#F_{IMU}$*	*T_M (ms)*	*T_T (ms)*
SURF	2660 (74%)	940 (26%)	299±22	19±3
BRISK	2858 (80%)	742 (20%)	130±27	20±3

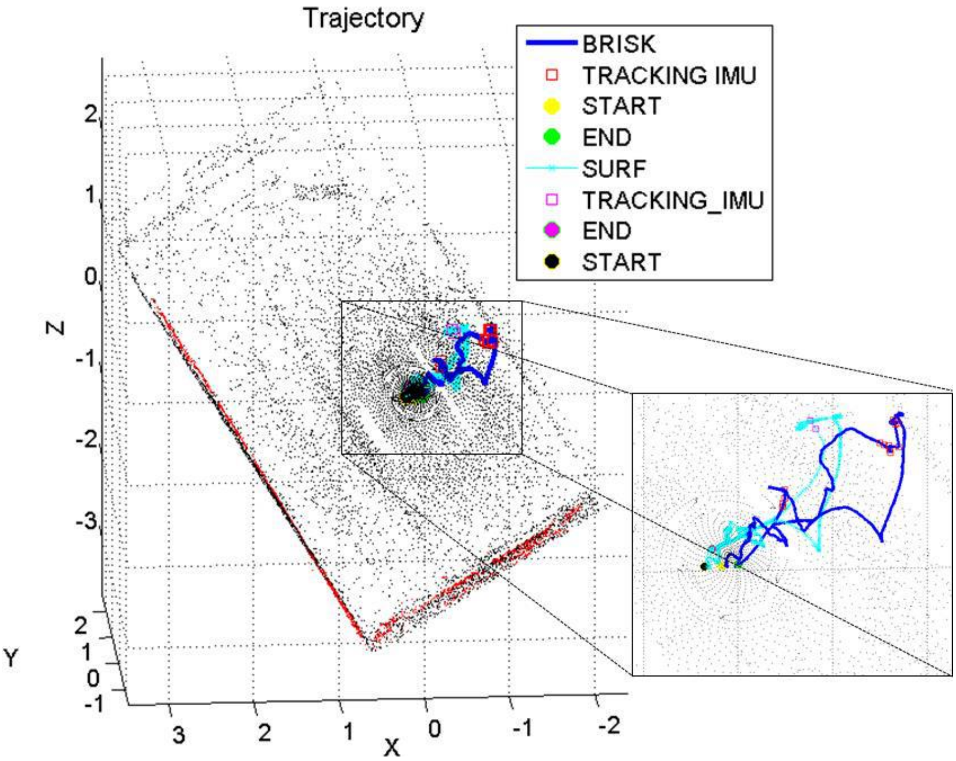


Figure 7: Trajectories (top view) estimated by the head tracking method for BRISK and SURF.

The different performance for the BRISK and SURF methods is also the result of the different frequency of *relocalization* following tracking failure. Indeed, because SURF matching is slower (Table 2, third column), *relocalization* using SURF cannot be invoked too often, when compared to BRISK, in order not to impact time performance (and so minimize latency). As a result, with SURF, the system is exposed to longer periods of lack of positional information (remaining in the *Tracking_IMU* mode), leading potentially to positional drift.

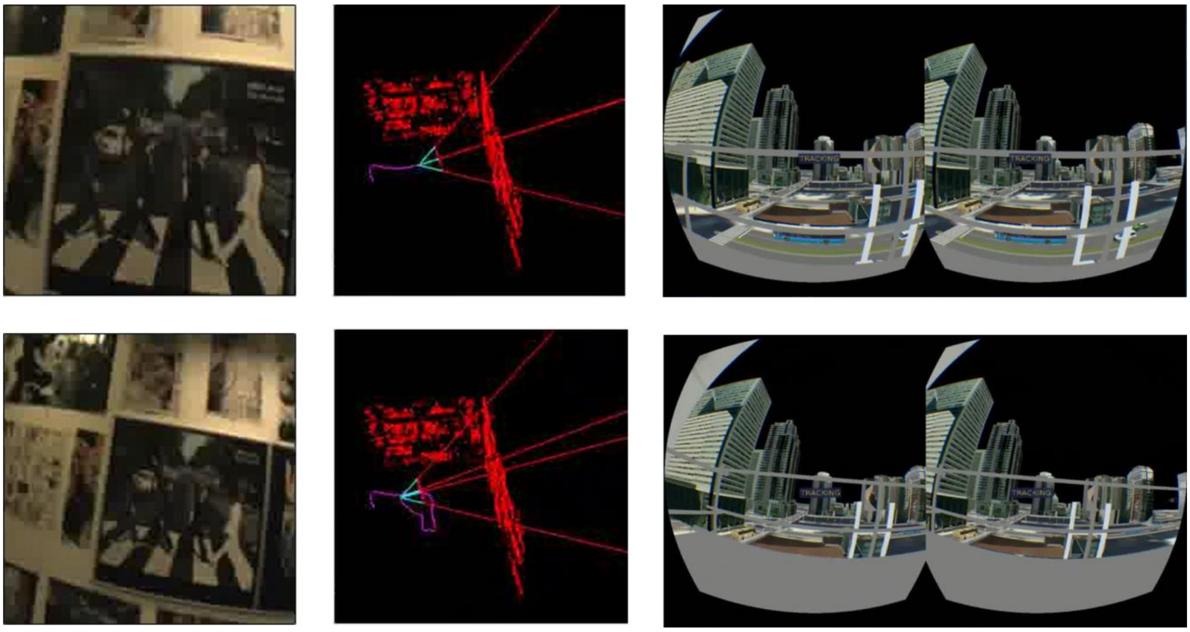
In Figure 8 the views of the virtual model of the room, rendered according to the estimated viewpoints, are shown for both methods (second and third columns) together with the real images (i.e. ground truth) acquired by the head-mounted camera (first column) for two significant sample time instants. It can be seen that, even in the presence of image degradation due to fast movements, the real and the virtual views generally appear in good visual agreement. However, as expected from the considerations above, the BRISK approach shows a better robustness and limited long-term drift. Furthermore, being a looping path sequence, the corresponding 3D loop closure error (the measured distance between the initial and final position) can be used as a measure of the drift effect. It has been estimated to be 0.09 m for the BRISK method, and 0.13 m for the SURF method. A longer four-minute sequence, with the user free to walk but returning three times to the same predefined location, has shown an average error of 0.18 m for BRISK and 0.88 m for SURF. That second sequence presents challenging motion patterns similar to the ones encountered in the first sequence, showing a similar behaviour for recovering after system outages and reinitializing the system. Further results confirming the robustness of the system during continued operation, particularly when using BRISK features, can also be found in (Carozza *et al.*, 2014b) and (Carozza *et al.*, 2014c).

Figure 8: Comparison between real images acquired live by the camera (after lens distortion compensation) - at first row: frame #525, second row: frame #1368 - and views of the virtual training room model rendered according to the viewpoint estimated using BRISK and SURF features, for fast motion.

These experimental results show good promise. However, the complete validation of the head tracking system will only be achieved once it will be integrated within an AR display system, which will enable the much more clear identification of drift and other pose estimation errors, and their actual impact on the overall system's usability.

Application: Experiencing Height

The research team was able to already employ the overall VR system to enable construction trainees to experience height. As mentioned earlier, for H&S reasons trainees in colleges cannot be physically put at heights above approx. 8m, so that many trainees may not have experienced common work-at-height situations prior to their first day on the job, and hence do not really know how well they can cope. Two scenarios have been considered: standing and moving on a scaffold at 10m height, and sitting on a structural steel beam at 100m height. Figure 9 illustrates users immersed in the two scenarios.



(a)

(b)

Figure 9: Application of the localization approach to two virtual scenarios: (a) standing and moving on a 10m scaffold; (b) sitting on a beam at 100m height (virtual model of the city courtesy of ESRI).

Early presentation of the system to FE college students and trainers received positive feedback, confirming that such a system could play a role in enabling trainees to safely experience different working conditions at height, to develop their readiness to such situations that they may later encounter in the real construction project environment.

Yet, it is interesting to discuss issues surrounding motion sickness. Indeed, users of VR goggles like Oculus have expressed concerns regarding motion sickness even after short utilisation (although it has also been reported that this sickness can disappear after some adaptation time). However, the authors note that those sicknesses appear to be reported in the case of current gaming scenarios where the user remains seated the whole time, in which case the visualized body motion does not match the actual motion felt through other body senses. As shown in previous studies (Laviola, 2000; Stanney, 2002; Chen et al, 2013), the authors believe that an additional advantage of 6-DOF motion head tracking systems like the one proposed here is that the visualized body motion directly and consistently relates to actual body motion, which should reduce the risk of motion sickness.

Conclusion and Future Work

The construction industry has traditionally shown poor levels of investment in R&D and innovation and as such is slow in the uptake of new technologies, in particular when it comes to the application of new technologies for education and training (CIOB, 2007). It is claimed that “*courses do not prepare students for the realities of construction sites or even the basics of health and safety and there is a bias towards the traditional trades and sketchy provision for new technologies*” (Knutt, 2012). This underlines the need for investment in new technologies to support construction training. If colleges want to become part of future education they should create change rather than waiting for it to happen to them (Hilpern, 2007).

The system presented in this paper is a novel approach that has the potential to transform construction trade training. The current VR Immersive Environment enables trainees to experience height, without involving any actual work. This simple exposure already enables trainees to experience such heights and assess their comfort in standing and eventually working in such conditions. Ultimately, it could even enable them to start accustom themselves to such conditions.

From a technical viewpoint, the main contribution of this paper is the presentation of an original visual-inertial 6-DOF head tracking system whose performance is shown to be promising.

It is worth noting that the choice of the system components – making use of commodity hardware and requiring very limited set-up (e.g. no installation and calibration of markers and multiple camera systems) – as well as the computing strategies adopted for each system stage already make the current VR system a valid alternative to existing immersive systems, such as CAVE (Cruz-Neira *et al.*, 1992).

The next phase of technical work will aim to complete the development of the envisioned MR immersive environment where the trainee can experience site conditions whilst performing real tasks. The accrued benefits of the application of MR and motion tracking technologies can include: enhancing the experience of apprenticeship training, complementing industrial placement and establishing site readiness, skills transfer and enhancement, performance measurement, benchmarking and recording, low operational cost and transferability across the industry. However, all these claims will require further research for validation using actual data.

From a technical viewpoint, the team's next step is to develop the depth sensing component and review the world mixing component, so that trainees can see their own body and selected parts of the surrounding real world, which is necessary to enable them to conduct actual construction tasks within varying virtual environments. For depth sensing, it is proposed to integrate a 3D camera (e.g. SoftKinetic *DepthSense 325* that provides range sensing up to 1.5m (SoftKinetic, 2013)), on top of the HMD and use the depth information to calculate in real-time the parts of the views of the virtual 3D environments that should be displayed on top of the real view, and those that should not be shown (i.e. the parts of the user's view where s/he should still be able to view the real world). For the AR viewing functionality (i.e. AR HMD), two approaches are possible. The first is to attach two cameras to the HMD and use the real-time imagery provided by these to create the mixed reality views, as recently demonstrated by Steptoe et al. (Steptoe, 2014). Alternatively, see-through HMDs, i.e. AR HMDs, can be employed that prevent altogether the need to acquire, process and consistently display views of the real world. For the envisioned system, the authors propose to use of the META *Spaceglasses* (META, 2013), a device that will be available in 2014. It is interesting to note that the META *Spaceglasses*, just like the *Oculus Rift*, integrate a high-frequency IMU (see discussion in the following bullet). But, even more interesting is the fact that the *Spaceglasses* also integrate a *DepthSense 325* camera. The *Spaceglasses* thus seem to already deliver many of the functionalities required by the system envisioned by the authors.

Finally, from an application viewpoint, it would be interesting to conduct a comparative study between traditional forms of construction training delivery and assessment (in a conventional workshop or classroom setting) as opposed to when using MR in order to demonstrate the impact of employing such technologies on trainees' performance.

Acknowledgements

The authors are grateful to the Construction Industry Training Board (CITB) for funding the development of the proposed MR system and to the project's stakeholders, Edinburgh College and the National Construction College (NCC), for their support.

References

Ahearn, A., Wise, C., McCann, E., and Goring P. (2005). Constructionarium: Building to Learn. *CEBE Transactions*, Vol. 2(1), pp. 6-16.

ACT-UK, (2012). *ACT-UK Simulation Centre*. http://www.act-uk.co.uk, last accessed on 20 December 2013.

Arayici, Y., and Aouad, G. (2004). DIVERCITY: Distributed virtual workspace for enhancing communication and collaboration within the construction industry. *Proceedings of the European Conference on Product and Process Modelling in the Building and Construction Industry (ECPPM)*, Istanbul, Turkey, pp. 695-706.

Austin, S., and Soetanto, R. (2010). The use of ACT-UK Virtual Reality Simulation Centre to enhance the learning experience of undergraduate building students. *Engineering Education*, Vol. 5(1), pp. 2-10.

Bassanino, M., Wu, K.-C., Yao, J., Khosrwoshahi, F., Fernando, T., and Skjaerbaek, J. (2010). The impact of immersive virtual reality on visualization for a design review in construction. *IEEE 14th International Conference on Information Visualization (IV)*, London, UK, pp. 585-589.

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. (2008) SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, Vol. 110(3), pp. 346-359, 2008

BIS (Department of Business, Innovation and Skills) (2012) *Better standards for Apprenticeships*. available at https://www.gov.uk/government/news/better-standards-for-apprenticeships, last accessed on 20 December, 2013.

Bleser, G. and Stricker, D. (2008). *Advanced tracking through efficient image processing and visual-inertial sensor fusion*. . IEEE Virtual Reality Conference VR '08, pp. 137-144.

BMSC (2012). *Building Management Simulation Centre (BMSC)*. http://www.bmsc.nl/, last accessed on 30 December 2013.

Borenstein, J, Ojeda, L., & Kwanmuang, S. (2009). Heuristic Reduction of Gyro Drift in a Personal Dead-reckoning System. *Journal of Navigation*, Vol. 62(1), pp. 41-58.

Bouchlaghem, N., Thorpe, A., and Liyange, I.G. (1996). Virtual Reality Applications in the UK's Construction industry. *CIB W78 Construction on the Information Highway*, Bled, Slovenia.

Buchmann, V., Violich, S., Billinghurst, M., and Cockburn, A. (2004). FingARtips: gesture based direct manipulation in augmented reality. In *GRAPHITE'04: Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pp. 212–221.

Burt, R. (2012). Delivering Construction Education for the Net Generation. *COBRA Conference*, Las Vegas, US.

Carozza, L., Bosché, F., and Abdel-Wahab, M. (2013). Image-based Localization for an Indoor VR/AR construction training system. *13th International Conference on*

Construction Applications of Virtual Reality (CONVR 2013), pp. 363-372, London, UK.

Carozza, L., Tingdahl, D., Bosché, F. and van Gool, L. (2014a). Markerless vision-based augmented reality for urban planning. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 29(1), pp. 2–17.

Carozza L., Bosché F. and Abdel-Wahab M. (2014b). Visual-Inertial 6-DOF Localization for a Wearable Immersive VR/AR System, *International Symposium on Mixed and Augmented Reality (ISMAR) 2014*, Munich, Germany, pp. 257-258

Carozza L., Bosché F. and Abdel-Wahab M. (2014c). Robust 6-DOF Immersive Navigation Using Commodity Hardware, *Virtual Reality Software and Technology 2014*, Edinburgh, UK, *accepted*.

Chen, W., Plancoulaine, A., Ferey, N., Touraine, D., Nelson, J., and Bourdot, P. (2013). 6-DoF Navigation in Virtual Worlds: Comparison of joystick-based and head-controlled paradigms. *In Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology (VRST),* ACM, New York, NY,USA, pp. 111-114.

Cheng, T., Venugopal, M., Teizer, J., and Vela, P.A. (2011). Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments. *Automation in Construction*, Vol. 20(8), pp. 1173–1184.

CIOB (2007). *Innovation in Construction: Ideas are the currency of the future*. Chartered Institute of Building (CIOB), available at http://www.ciob.org/sites/default/files/Innovation%20in%20Construction.pdf, last accessed on 14 October, 2013.

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V. and Hart, J. C. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, Vol. 35(6), pp. 64–72.

Davies S. (2008). *Apprenticeships: A firm foundation*. Report for UCATT. London: UCATT.

De Souza e Silva, A., and Sutko, D.M. (2009). *Digital Cityscapes: merging digital and urban playspaces*. New York: Peter Lang Publishing, Inc.

De Vries, B., Verhagen, S., and Jessurun, A.J. (2004). Building management simulation centre. *Automation in Construction*, Vol. 13 (5), pp. 679-687.

Escorcia, V, Dávila, M., Golparvar-Fard, M., and Niebles, J. (2012). Automated Vision-Based Recognition of Construction Worker Actions for Building Interior Construction Operations Using RGBD Cameras, in *Proceedings of the ASCE Construction Research Congress*, pp. 879-888.

Fischler M. A. and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM,* Vol. 24(6), pp. 381-395.

Forsythe P. (2009). The Construction Game – Using Physical Model Making to Simulate Realism in Construction Education. J*ournal for Education in the Built Environment*, Vol. 4(1), pp. 57-74.

Gauglitz S., Hollerer, T., and Turk, M. (2011). Evaluation of interest point detectors and feature descriptors for visual tracking. *International Journal of Computer Vision*, Vol. 94(3), pp. 335–360.

Gong, J., and Caldas, C. (2011). Learning and Classifying Motions of Construction Workers and Equipment Using Bag of Video Feature Words and Bayesian Learning Methods. *ASCE Computing in Civil Engineering*, pp. 274-281.

Goulding, J., Nadim, W., Petridis, P., and Alshawi, M. (2012). Construction industry offsite production: a virtual reality interactive training environment prototype. *Advanced Engineering Informatics*, Vol. 26, pp. 103-116.

Hamer, H., Gall, J., Weise, T., and van Gool, L. (2010). An object-dependent hand pose prior from sparse training data. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, USA.

Haralick, R., Lee, C., Ottenberg K. and Nölle, M. (1994). Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem, *International Journal of Computer Vision*, Vol. 13 (3), pp. 331-356.

Hartley, R.I. and Zisserman, A. (2004). Multiple View Geometry in Computer Vision, second edition, Cambridge University Press.

Hedberg J., and Alexander S. (1994). Virtual reality in education: defining researchable issues. *Educational Media International*, Vol. 31(4), pp. 214-220.

HSE (2010) *Construction Intelligence Report: Analysis of Construction Injury and ill health intelligence*, Health and Safety Executive (HSE), available at: http://www.hse.gov.uk/construction/pdf/conintreport.pdf, last accessed on 23 October, 2013.

Hilpern, K. (2007) *What will colleges of the future look like?* The Independent, Available at: http://www.independent.co.uk/news/education/further/what-will-colleges-of-the-future-look-like-400366.html, last accessed on 11 November 2013.

Immersive Technologies (2008). PT Freeport Indonesia Incorporates Locals to the Work Force. *Mining Magazine*. Available at: http://magazine.mining.com/issues/0810/PTFreeportIndonesia.pdf, last accessed on 20 December 2013.

Immersive Technologies (2010) New PRO3 Simulator and Caterpillar 793F the perfect match in Tucson. Available at: http://www.immersivetechnologies.com/news/news2010/news_2010_08.htm, last accessed on 30 June 2014.

International Labour Organisation (ILO) (2012) The lure of apprenticeships in times of crisis, Available at: http://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_191357/lang--en/index.htm , last accessed on 6 June 2014.

Keskinen, E., Iltanen, M., Salonen, T., Launis, S., Cotsaftis, M., and Pispala, J. (2000). Main-in-loop Training Simulator for Hydraulic Elevating Platforms, *Proceedings of the 17th International Symposium on Automation and Robotics in Construction (ISARC)*, Taiwan, pp. 993-999.

Knutt, E. (2012). Generation Lost, *Construction Manager*, available at: http://www.construction-manager.co.uk/agenda/generation-lost/, last accessed on 11 November 2013.

Ku, K., and Mahabaleshwarkar, P. (2011). Building Interactive Modeling for Construction Education in Virtual Worlds. *Journal of Information Technology in Construction*, Vol. 16, pp. 189-208.

Leutenegger, S., Chli, M., and Siegwart, R.Y. (2011). BRISK: Binary Robust invariant scalable keypoints, *IEEE International Conference on Computer Vision (ICCV)*, pp. 2548-2555.

Laviola, JR., J. J. (2000). A discussion of cybersickness in virtual environments. *SIGCHI Bull.* Vol. 32, 1, pp. 47–56.

Li, H., Chan, G., and Skitmore, M. (2012). Visualizing safety assessment by integrating the use of game technology. *Automation in Construction*, Vol. 22, pp. 498-505.

Lobo, J., and Dias, J. (2007). Relative pose calibration between visual and inertial sensors, *International Journal of Robotic Research (IJRR)*, Vol. 26(6), pp. 561–575.

Martin, S., Diaz, G., Sancristobal, E., Gil, R., Castro, M., and Peire, J. (2011). New technology trends in education: Seven years of forecasts and convergence. *Computers & Education*, 57, pp. 1893-1906.

META (2013) *Spaceglasses*, https://www.spaceglasses.com/ , last accessed 23rd November 2013.

Mikropoulous, T., and Natsis, A. (2011). Educational Virtual Environments: A ten-year review of empirical research (1999-2009). *Computers & Education*, Vol. 56, pp. 769-780.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1994). Augmented Reality: A class of displays on the reality-virtuality continuum. *Proceedings of Telemanipulator and Telepresence Technologies*. pp. 2351–2385.

Milgram, P., and Colquhoun, H. (1999). A Taxonomy of Real and Virtual World Display Integration, *IEICE Transactions on Information and Systems*, Vol. 77(12), pp. 1321-1329.

Oculus (2013), *Oculus Rift*, http://www.oculusvr.com/, last visited: 10th December 2013.

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., and Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Computer & Graphics*, 30, pp. 20-28.

Ray, S.J., and Teizer, J. (2012). Real-time construction worker posture analysis for ergonomics training. *Advanced Engineering Informatics*, Vol. 26(2), pp. 439-455.

Rivara, P.F., and Thompson, D.C. (2000) Prevention of falls in the construction industry: evidence for program effectiveness. *American Journal of Preventive Medicine*, VOl. 18(48), pp. 23-26.

Shi J. and Tomasi, C. (1994). Good Features to Track, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593–600.

Skills Development Scotland (SDS) (2014) Our Skillsforce, Certificate of Work Readiness, available at: http://www.ourskillsforce.co.uk/funding-for-skills/certificate-of-work-readiness, last accessed on 9 June 2014.

Snavely, N., Seitz, S. M., and Szeliski, R. (2008) Modeling the world from internet photo collections. *Int. J. Comput. Vision*, 80(2), 189–210.

SoftKinetic (2013), *DepthSense 325*, information available at http://www.softkinetic.com/Store/tabid/579/ProductID/6, last visited: 10th December 2013.

Stanney, K. M., Kingdon, K. S., Graeber, D., and Kennedy, R. S. (2002). Human performance in immersive virtual environments: Effects of exposure duration, user control, and scene complexity. *Human Performance,* Vol. 15(4), pp. 339-366.

Steptoe, W., Julier, J. and Steed, A. (2014) Presence and Discernibility in Conventional and Non Photorealistic Immersive Augmented Reality, International Symposium on Mixed and Augmented Reality (ISMAR) 2014, Munich, Germany, September, Sept. 10-12, pp. 213-218.

Stothers, N. (2007). The use of Virtual Reality Simulation in the Education and Training of Construction Managers – Are the British 'going Dutch'? *2nd International Conference World of Construction Project Management*, TU Delft, Netherlands.

Talmaki, S. and Kamat, V. (2014). Real-Time Hybrid Virtuality for Prevention of Excavation Related Utility Strikes. *ASCE Journal of Computing in Civil Engineering*, Vol. 28(3), 04014001.

Teizer, J., and Vela, P.A. (2009). Personnel tracking on construction sites using video cameras. *Advanced Engineering Informatics*, Vol. 23, pp. 452-462.

Teizer, J., Cheng, T., and Fang, Y. (2013). Location tracking and data visualization technology to advance construction ironworkers' education and training in safety and productivity. *Automation in Construction*, Vol. 35, pp. 53-68.

TSPI (2011). *Construction equipment operator training simulators*. Training, Simulation, Performance and Improvement Technologies (TSPI). Available at: http://www.swri.org/4org/d07/tspi/construction.htm.

Tzafestas C.S. (2003). Whole-hand kinesthetic feedback and haptic perception in dextrous virtual manipulation. In *IEEE Transactions on Systems, Man and Cybernetics, Part A*, pp. 100-113.

UK Government (2011). *Government Construction Strategy*, Cabinet Office, available at: https://www.gov.uk/government/publications/government-construction-strategy, last accessed on 20 December 2013.

UK Government (2013). *Construction 2025*, UK HM Government, available at https://www.gov.uk/government/publications/construction-2025-strategy, last accessed on 20 December 2013.

Unity 3D (2014), *Unity 3D Game Engine*, information available at http://unity3d.com/, last visited October 2014

Volvo CE (Volvo Construction Equipment) (2011). *Real life comes to the classroom with Volvo's latest excavator training simulators*, press release, June 6.

Wang, X., Dunston, P.S., and Skibniewski, M., (2004). Mixed Reality Technology Applications in Construction Equipment Operator Training, *Proceedings of the 21st International Symposium on Automation and Robotics in Construction (ISARC)*, Jeju, Korea, Sep 23rd.

Wang, X., and Dunston, P.S., (2005) Heavy Equipment Operator Training via Virtual Modeling Technologies. *Proceedings of the ASCE Construction Research Congress (CRC)*, 1-10, San Diego, USA.

Welch, G., and Foxlin, E. (2002). Motion Tracking: No Silver Bullet, but a Respectable Arsenal, *IEEE Comput. Graph. Appl.*,22(6), pp 24-38.

Winn, W., and Windschitl, M. (2000). Learning science in virtual environments: the interplay of theory and experience. *Themes in Education*, 1(4), pp. 373-389.

Yang, J., Cheng, T., Teizer, J., Vela, P.A., and Shi, Z.K. (2011). A performance evaluation of vision and radio frequency tracking methods for interacting workforce, *Advanced Engineering Informatics*, 25(4), pp. 736–747.