



Heriot-Watt University
Research Gateway

Towards a Mixed Reality System for Construction Trade Training

Citation for published version:

Bosché, F, Abdel-Wahab, MS & Carozza, L 2016, 'Towards a Mixed Reality System for Construction Trade Training', *Journal of Computing in Civil Engineering*, vol. 30, no. 2, 04015016.
[https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000479](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000479)

Digital Object Identifier (DOI):

[10.1061/\(ASCE\)CP.1943-5487.0000479](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000479)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Journal of Computing in Civil Engineering

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Towards a Mixed Reality System for Construction Trade Training

Dr. Frédéric Bosché^{1,*}, Dr. Mohamed Abdel-Wahab²,

Dr. Ludovico Carozza³

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. We report 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 our 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 into the envisioned

¹Assistant Professor, School of the Built Environment, Heriot-Watt University.

²Assistant Professor, School of the Built Environment, Heriot-Watt University.

³Research Associate, School of the Built Environment, Heriot-Watt University.

* Corresponding author: f.n.bosche@hw.ac.uk

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

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

27 **Introduction**

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

35

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

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

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

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

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

74

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

78

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

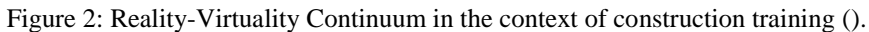
92 The paper commences with a literature review of the current applications of MR in
93 construction training, which leads to identification of the need for different MR systems
94 better suited to the needs of construction trade training. We then present the on-going
95 development of such an MR system. The current system is only a VR system, but includes
96 several of the functional components that will be required in the final MR system. We
97 particularly focus on our main computational contribution that is a robust, cost-effective 6-

98 DOF Head Tracking system. The performance of the current system is experimentally
99 assessed in challenging scenarios. Finally, strategies are discussed for the completion of the
100 envisioned MR system.

101 **'Reality-Virtuality' continuum of construction training**

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

107

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

109

110 **Real Environment**

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

116 However, training on real construction projects is constrained by the type of activity taking
117 place on site, project duration, in addition to (occupational) health and safety (H&S) risks.

118 Trainees may not be allowed to perform certain tasks on real projects as this can cause delays

119 and errors can be costly, especially when it comes to high profile projects such as the
120 Olympics. To address this issue, attempts have been made in recent years to ‘simulate’ real
121 project environments where trainees can conduct real tasks without compromising project
122 performance and H&S.

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

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

136 **Virtual Reality (VR)**

137 At the other end of the ‘Reality-Virtuality’ continuum (Figure 2), *Virtual Reality* (VR) is
138 increasingly used for construction training. VR development boomed in the 1990’s and VR is
139 in fact still under intense development, with education and training an important area of
140 application. Mikropoulos and Natsis (2011) define a Virtual Reality Learning Environment
141 (VRLE) as “*a virtual environment that is based on a certain pedagogical model, incorporates*
142 *or implies one or more didactic objectives, provides users with experiences they would*

143 *otherwise not be able to experience in the physical environment and can support the*
144 *attainment of specific learning outcomes.”*

145

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

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

159

160 Equipment simulators enable training in realistic construction project scenarios with high-
161 fidelity, which is made possible by force feedback mechanisms, and without exposing
162 trainees or instructors to occupational H&S risks. They support fast and efficient learning
163 thereby increasing trainees’ motivation (Volvo CE, 2011; TSPIT, 2011). For example, the
164 ITAE simulator, employed in mining equipment operation training, is used to ensure that
165 apprentices can demonstrate a certain skill level prior to working in mines. The manufacturer
166 claims that the simulator has proved to be effective in modifying and improving operators’

167 behaviour as well as enhancing the existing skills levels and performance of employees
168 (Immersive Technologies, 2008).

169 VRLEs have also been developed for *supervision/management training*. The first UK
170 construction management simulation centre has opened at Coventry University in 2009 and is
171 known as ACT-UK (Advanced Construction Technology Simulation Centre). The centre is
172 aimed at already practicing foremen and construction managers, and potentially students
173 (Austin and Soetanto, 2010; ACT-UK, 2012). Similar centres exist with the Building
174 Management Simulation Center (BMSC) in The Netherlands (De Vries *et al.*, 2004; BMSC,
175 2012) or the OSP VR Training environment collaboratively developed as part of the
176 Manubuild EU project (Goulding *et al.*, 2012). In these VRLEs, trainees can be partially
177 immersed in simulated construction site environments to safely expose them to situations that
178 they must know how to deal with appropriately. These may include H&S, work planning and
179 coordination, or conflict resolution scenarios (Harpur, 2009; Ku, 2011; Li, 2012). Other
180 VRLEs have also been investigated for other applications for enhancing communication and
181 collaboration during briefing, design, and construction planning (Duston, 2000; Arayici,
182 2004; Bassanino, 2010).

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

188 Nonetheless, despite the general agreement on the potential of VRLEs to enhance education,
189 Mikropoulous (2011) and Wang and Dunston (2005) noted that there is a general lack of
190 thorough demonstration of the value-for-money achieved by those systems, which may be

191 due to implementation cost, but possibly also to the quantity and quality of training scenarios
192 that could be developed and their impact on learning and practice.

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

198 **Mixed Reality (MR)**

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

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

214 display devices, motion tracking, and conformal mapping of the virtual and real worlds
215 (Martin *et al.*, 2011).

216

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

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

233

234 To the knowledge of the authors, no work has been reported to date on developing MR
235 systems for the training of construction trades, (e.g. roofing, painting and decorating,
236 bricklaying, scaffolding, etc.). The particularity of those trades is that the trainee must be in
237 direct manual contact with tools and materials. Immersing their work thus requires specific

238 interfaces for tracking the limbs of trainees (particularly the arms and hands), and integrating
239 the manipulations with virtual environments.

240 Research has been widely conducted to develop such interfaces. Haptic gloves or other worn
241 devices are investigated (Tzafestas, 2003; Buchmann *et al.*, 2004), but are invasive. Non-
242 invasive vision-based body tracking solutions have also been considered (Hamer *et al.*,
243 2010), but are usable only within very small spaces. Thus, despite continuous improvements,
244 current solutions for manual interactions with virtual environments do not provide the
245 richness and interactivity required for effective trade training.

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

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

257 **Need Identification, Functional Analysis, and Current System**

258 It was concluded in the previous section that construction trade training can benefit from MR
259 by employing it solely to visually immerse trainees, while they conduct training activities
260 with real tools and materials. Referring to the taxonomy of Milgram *et al.* (1994; 1999), the

261 type of system required appears to correspond to MR systems they classify as *Class 3* or
262 *Class 4* (see Table 1). However, we also observe that, from a visualization viewpoint, this
263 more specifically requires that the trainee be able to see their real body and real work (tools,
264 material), and see these immersed within a virtual world. This means that the system would
265 have to calculate in real-time in which parts of the user's field of view the virtual world must
266 be overlaid on the real world, and in which parts it should not. In other words, the system
267 needs to deliver AR functionality with (local) occlusion handling, which requires that the 3D
268 state of the real world be known accurately and in real-time (the 3D state of the virtual world
269 is naturally already known). Referring again to the taxonomy of Milgram *et al.* (1994; 1999),
270 the type of system required thus needs to have an *Extent of (Real) World Knowledge (EWK)*
271 where the depth map of the real world from the user's viewpoint is completely modelled (see
272 Figure 3).

273

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

276

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

279

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

- 282 • *6-DOF head tracker*: provides the 3D pose (i.e. location and orientation) of the user's
283 head in real-time;

- 284 • *Depth sensor*: provides a depth map of the environment in the field of view of the
285 user;
- 286 • *Virtual World Simulator / Game Engine*: simulates the virtual 3D environment and is
287 used to generated views of it from given locations;
- 288 • *Processing Unit*: uses the information provided by the three components above to
289 calculate the user’s views of the mixed real and virtual worlds to be displayed in the
290 HMD in real-time;
- 291 • *HMD (preferably, but not necessarily, see-through)*: is used to display the views
292 generated by the Processing Unit.

293

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

295

296 In the following, we present our progress to date that involves the implementation of four of
297 the five components above:

- 298 • *6-DOF Head Tracker*: The 6-DOF head tracking (i.e. localization) is probably the
299 most critical functionality to be delivered by real-time MR systems. Localization is
300 even more critical for MR systems than for VR systems, because poor pose tracking is
301 far more disturbing in MR scenarios since these require the virtual display content to
302 be very accurately aligned with the reality. Robust localization is critical to user
303 experience.

304 Guaranteeing continuous operation while the user is moving is already a challenge;
305 doing it without requiring complex and expensive set-up, is an even greater one. Our
306 main contribution in this paper is an original cost-effective visual-inertial 6-DOF head

307 tracker. The system is detailed in the section below, and its performance is
308 particularly assessed in the experiments reported later on.

- 309 • *Game Engine*: we integrated our 6-DOF Head Tracking system as a third-party
310 component to the Unity 3D game engine (Unity 3D, 2014). This gives our approach a
311 wider applicability and scalability to a range of different training scenarios, thus
312 providing flexibility to different operative trades. Game engines also have the
313 important advantage of already providing optimized capabilities for high-quality
314 rendering and user interaction within complex virtual environments.
- 315 • *HMD*: Our system currently employs the Oculus Rift (Oculus, 2013) that is a non-see-
316 through HMD, i.e. VR, device that offers great immersive experience with a 110°
317 field of view.
- 318 • *Processing Unit*: as discussed below, the Depth Sensing component has not been
319 implemented yet. As a result, our current system can only deliver VR functionality,
320 not AR. Therefore, the Processing Unit is currently only partially implemented, as it
321 only calculates views of the virtual 3D environment (managed by the Game Engine)
322 to be displayed on the HMD.

323 At this stage, we have not implemented any solution for the Depth Sensing component.
324 However, a solution is proposed in the *Future Work* section at the end of this paper.
325 Similarly, our envisioned system needs to deliver AR, not just VR functionality. Our
326 proposed approach to achieve this is also discussed in the *Future Work* section.

327

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

332 **6-DOF Head Tracker**

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

336 **Introduction**

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

350 In an effort to address these limitations, we have been investigating an alternative visual-
351 inertial approach for 6-DOF position tracking that integrates an IMU and a markerless vision-
352 based system. Visual-inertial *ego-motion* approaches have been conceived in general to
353 represent an affordable technology, also usually requiring limited set-up. Complementary
354 action of visual and inertial data can increase robustness and accuracy in determining both

355 position and orientation even in response to faster motion (Welch and Foxlin 2002, Bleser
356 and Stricker 2008). Our specific approach, detailed in the following section, has been
357 designed to handle system outages and deliver continued tracking at the required quality.

358 **Our Approach**

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

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

375 **Off-line Reconstruction Stage**

376 The *off-line reconstruction stage* (Figure 5 left) is performed in advance, once and for all, by
377 automatically processing pictures of the training room, acquired by the camera from different

378 viewpoints, according to the Structure from Motion *Bundler* framework (Snavely 2008). The
379 training room has been textured in advance by using posters (Figure 5 (a)) – with a random
380 layout – so that a *3D map of visual references* can be reliably reconstructed (Figure 5 (b)).
381 The reconstructed point cloud is then used as reference for the alignment of the virtual
382 training scenario with the (real) world reference frame (Figure 5 (c)).

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

388 The result of the process above is a database of repeatable visual descriptors, referred in the
389 3D space, or world reference frame (WRF), and that is used for the subsequent on-line
390 localization stage.

391 **On-line Localization Stage**

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

402 However, the global matching approach can be (a) not sufficiently precise and robust, due to
 403 image degradation during fast movements, or (b) not sufficiently efficient for real-time
 404 performance (due to query search overhead over the whole database). Accordingly, a *feature*
 405 *tracking* strategy is used together with the IMU data for the subsequent frames. A frame-to-
 406 frame tracking approach based on the Kanade-Lucas-Tomasi (KLT) tracker (Shi and Tomasi
 407 1994) is employed between consecutive frames, with the advantage of being very efficient
 408 and exploiting spatio-temporal contiguity to track faster motions. More details about the
 409 feature tracking approach, and in particular *tracker reinitialization* to allow tracking over
 410 long periods, can be found in (Carozza *et al.*, 2013). Note that a pin-hole camera model is
 411 considered throughout all the stages of the vision-based approach, taking into account also
 412 lens radial distortion.

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

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

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

426 However, in order to be fused consistently with the visual data, the inertial data must be
427 referred to the same absolute reference frame of the visual data (i.e. the training room). We
428 developed an on-the-fly *camera-IMU calibration* routine, which automatically processes the
429 first N_{calib} pairs of visual and inertial data following the very first successful initialization to
430 estimate the *calibration matrix* relating the inertial reference frame to the global reference
431 frame. Our calibration method is similar to the classic *hand-eye* calibration (see Lobo *et al.*
432 2007), but it can be employed on-line since the relative translation between the camera and
433 the IMU centres does not need to be estimated (it is not taken into account into the
434 subsequent calculations).

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

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

447 In Figure 6, the state diagram of the adopted 6-DOF tracking framework summarizes the
448 main transitions occurring during on-line operations among the different stages encountered
449 above. These transitions illustrate at a high level the continued operation of the system over

450 long periods from the initialization to the response and recovery from different system
451 outages.

452

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

455

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

458

459 Finally, for each frame, once the head pose is estimated, any 3D graphic model/virtual
460 environment can be rendered consistently with the estimated viewpoint. For example, Figure
461 5 (f) shows the rendered views of a virtual model of the training room corresponding to the
462 head locations estimated using the two head-mounted camera views shown in Figure 5 (d).

463

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

470 **Experimental Results**

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

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

481 **Head Tracking**

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

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

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

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

504

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

509

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

511

512 The different performance for the BRISK and SURF methods is also the result of the
513 different frequency of *relocalization* following tracking failure. Indeed, because SURF
514 matching is slower (Table 2, third column), *relocalization* using SURF cannot be invoked too
515 often, when compared to BRISK, in order not to impact time performance (and so minimize

516 latency). As a result, with SURF, the system is exposed to longer periods of lack of positional
517 information (remaining in the *Tracking_IMU* mode), leading potentially to positional drift.

518

519 In Figure 8 the views of the virtual model of the room, rendered according to the estimated
520 viewpoints, are shown for both methods (second and third columns) together with the real
521 images (i.e. ground truth) acquired by the head-mounted camera (first column) for two
522 significant sample time instants. It can be seen that, even in the presence of image
523 degradation due to fast movements, the real and the virtual views generally appear in good
524 visual agreement. However, as expected from the considerations above, the BRISK approach
525 shows a better robustness and limited long-term drift. Furthermore, being a looping path
526 sequence, the corresponding 3D loop closure error (the measured distance between the initial
527 and final position) can be used as a measure of the drift effect. It has been estimated to be
528 0.09 m for the BRISK method, and 0.13 m for the SURF method. A longer four-minute
529 sequence, with the user free to walk but returning three times to the same predefined location,
530 has shown an average error of 0.18 m for BRISK and 0.88 m for SURF. That second
531 sequence presents challenging motion patterns similar to the ones encountered in the first
532 sequence, showing a similar behaviour for recovering after system outages and reinitializing
533 the system. Further results confirming the robustness of the system during continued
534 operation, particularly when using BRISK features, can also be found in (Carozza *et al.*,
535 2014b) and (Carozza *et al.*, 2014c).

536

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

540

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

545 **Application: Experiencing Height**

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

553

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

556

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

561

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

572 **Conclusion and Future Work**

573 The construction industry has traditionally shown poor levels of investment in R&D and
574 innovation and as such is slow in the uptake of new technologies, in particular when it comes
575 to the application of new technologies for education and training (CIOB, 2007). It is claimed
576 that “*courses do not prepare students for the realities of construction sites or even the basics*
577 *of health and safety and there is a bias towards the traditional trades and sketchy provision*
578 *for new technologies*” (Knutt, 2012). This underlines the need for investment in new
579 technologies to support construction training. If colleges want to become part of future
580 education they should create change rather than waiting for it to happen to them (Hilpern,
581 2007).

582

583 The system presented in this paper is a novel approach that has the potential to transform
584 construction trade training. The current VR Immersive Environment enables trainees to

585 experience height, without involving any actual work. This simple exposure already enables
586 trainees to experience such heights and assess their comfort in standing and eventually
587 working in such conditions. Ultimately, it could even enable them to start accustom
588 themselves to such conditions.

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

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

597

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

606 From a technical viewpoint, our next step is to develop the depth sensing component and
607 review the world mixing component, so that trainees can see their own body and selected
608 parts of the surrounding real world, which is necessary to enable them to conduct actual

609 construction tasks within varying virtual environments. For depth sensing, we propose to
610 integrate a 3D camera (e.g. SoftKinetic *DepthSense 325* that provides range sensing up to
611 1.5m (SoftKinetic, 2013)), on top of the HMD and use the depth information to calculate in
612 real-time the parts of the views of the virtual 3D environments that should be displayed on
613 top of the real view, and those that should not be shown (i.e. the parts of the user's view
614 where s/he should still be able to view the real world). For the AR viewing functionality (i.e.
615 AR HMD), two approaches are possible. The first is to attach two cameras to the HMD and
616 use the real-time imagery provided by these to create the mixed reality views, as recently
617 demonstrated by Steptoe et al. (Steptoe, 2014). Alternatively, see-through HMDs, i.e. AR
618 HMDs, can be employed that prevent altogether the need to acquire, process and consistently
619 display views of the real world. In our system, we propose to use of the META *Spaceglasses*
620 (META, 2013), a device that will be available in 2014. It is interesting to note that the META
621 *Spaceglasses*, just like the *Oculus Rift*, integrate a high-frequency IMU (see discussion in the
622 following bullet). But, even more interesting is the fact that the *Spaceglasses* also integrate a
623 *DepthSense 325* camera. The *Spaceglasses* thus seem to already deliver many of the
624 functionalities required by our envisioned system.

625

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

630 Acknowledgements

631 The authors are grateful to the Construction Industry Training Board (CITB) for funding the
632 development of the proposed MR system and for our stakeholders, Edinburgh College and
633 the National Construction College (NCC), for supporting our project.

634 References

- 635 Ahearn, A., Wise, C., McCann, E., and Goring P. (2005). Constructionarium: Building to
636 Learn. *CEBE Transactions*, Vol. 2(1), pp. 6-16.
- 637 ACT-UK, (2012). *ACT-UK Simulation Centre*. <http://www.act-uk.co.uk>, last accessed on 20
638 December 2013.
- 639 Arayici, Y., and Aouad, G. (2004). DIVERCITY: Distributed virtual workspace for
640 enhancing communication and collaboration within the construction industry.
641 *Proceedings of the European Conference on Product and Process Modelling in the*
642 *Building and Construction Industry (ECPPM)*, Istanbul, Turkey, pp. 695-706.
- 643 Austin, S., and Soetanto, R. (2010). The use of ACT-UK Virtual Reality Simulation Centre to
644 enhance the learning experience of undergraduate building students. *Engineering*
645 *Education*, Vol. 5(1), pp. 2-10.
- 646 Bassanino, M., Wu, K.-C., Yao, J., Khosrwohahi, F., Fernando, T., and Skjaerbaek, J.
647 (2010). The impact of immersive virtual reality on visualization for a design review in
648 construction. *IEEE 14th International Conference on Information Visualization (IV)*,
649 London, UK, pp. 585-589.

650 Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. (2008) SURF: Speeded Up Robust
651 Features, *Computer Vision and Image Understanding (CVIU)*, Vol. 110(3), pp. 346-
652 359, 2008

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

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

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

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

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

665 Buchmann, V., Violich, S., Billinghurst, M., and Cockburn, A. (2004). FingARtips: gesture
666 based direct manipulation in augmented reality. In *GRAPHITE'04: Proceedings of the*
667 *2nd international conference on Computer graphics and interactive techniques in*
668 *Australasia and South East Asia*, pp. 212–221.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

713 Gauglitz S., Hollerer, T., and Turk, M. (2011). Evaluation of interest point detectors and
714 feature descriptors for visual tracking. *International Journal of Computer Vision*, Vol.
715 94(3), pp. 335–360.

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

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

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

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

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

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

732 HSE (2010) *Construction Intelligence Report: Analysis of Construction Injury and ill health*
733 *intelligence*, Health and Safety Executive (HSE), available at:
734 <http://www.hse.gov.uk/construction/pdf/conintreport.pdf>, last accessed on 23 October,
735 2013.

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

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

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

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

749

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

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

757 Ku, K., and Mahabaleshwarkar, P. (2011). Building Interactive Modeling for Construction
758 Education in Virtual Worlds. *Journal of Information Technology in Construction*, Vol.
759 16, pp. 189-208.

760 Leutenegger, S., Chli, M., and Siegwart, R.Y. (2011). BRISK: Binary Robust invariant
761 scalable keypoints, *IEEE International Conference on Computer Vision (ICCV)*, pp.
762 2548-2555.

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

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

- 767 Lobo, J., and Dias, J. (2007). Relative pose calibration between visual and inertial sensors,
768 *International Journal of Robotic Research (IJRR)*, Vol. 26(6), pp. 561–575.
- 769 Martin, S., Diaz, G., Sancristobal, E., Gil, R., Castro, M., and Peire, J. (2011). New
770 technology trends in education: Seven years of forecasts and convergence. *Computers
771 & Education*, 57, pp. 1893-1906.
- 772 META (2013) *Spaceglasses*, <https://www.spaceglasses.com/> , last accessed 23rd November
773 2013.
- 774 Mikropoulous, T., and Natsis, A. (2011). Educational Virtual Environments: A ten-year
775 review of empirical research (1999-2009). *Computers & Education*, Vol. 56, pp. 769-
776 780.
- 777 Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1994). Augmented Reality: A class
778 of displays on the reality-virtuality continuum. *Proceedings of Telem manipulator and
779 Telepresence Technologies*. pp. 2351–2385.
- 780 Milgram, P., and Colquhoun, H. (1999). A Taxonomy of Real and Virtual World Display
781 Integration, *IEICE Transactions on Information and Systems*, Vol. 77(12), pp. 1321-
782 1329.
- 783 Oculus (2013), *Oculus Rift*, <http://www.oculusvr.com/>, last visited: 10th December 2013.
- 784 Pan, Z., Cheok, A. D., Yang, H., Zhu, J., and Shi, J. (2006). Virtual reality and mixed reality
785 for virtual learning environments. *Computer & Graphics*, 30, pp. 20-28.
- 786 Ray, S.J., and Teizer, J. (2012). Real-time construction worker posture analysis for
787 ergonomics training. *Advanced Engineering Informatics*, Vol. 26(2), pp. 439-455.

788 Rivara, P.F., and Thompson, D.C. (2000) Prevention of falls in the construction industry:
789 evidence for program effectiveness. *American Journal of Preventive Medicine*, VOL.
790 18(48), pp. 23-26.

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

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

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

798 SoftKinetic (2013), *DepthSense 325*, information available at
799 <http://www.softkinetic.com/Store/tabid/579/ProductID/6>, last visited: 10th December
800 2013.

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

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

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

811 Talmaki, S. and Kamat, V. (2014). Real-Time Hybrid Virtuality for Prevention of Excavation
812 Related Utility Strikes. *ASCE Journal of Computing in Civil Engineering*, Vol. 28(3),
813 04014001.

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

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

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

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

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

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

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

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

- 835 Wang, X., Dunston, P.S., and Skibniewski, M., (2004). Mixed Reality Technology
836 Applications in Construction Equipment Operator Training, *Proceedings of the 21st*
837 *International Symposium on Automation and Robotics in Construction (ISARC)*, Jeju,
838 Korea, Sep 23rd.
- 839 Wang, X., and Dunston, P.S., (2005) Heavy Equipment Operator Training via Virtual
840 Modeling Technologies. *Proceedings of the ASCE Construction Research Congress*
841 *(CRC)*, 1-10, San Diego, USA.
- 842 Welch, G., and Foxlin, E. (2002). Motion Tracking: No Silver Bullet, but a Respectable
843 Arsenal, *IEEE Comput. Graph. Appl.*,22(6), pp 24-38.
- 844 Winn, W., and Windschitl, M. (2000). Learning science in virtual environments: the interplay
845 of theory and experience. *Themes in Education*, 1(4), pp. 373-389.
- 846 Yang, J., Cheng, T., Teizer, J., Vela, P.A., and Shi, Z.K. (2011). A performance evaluation of
847 vision and radio frequency tracking methods for interacting workforce, *Advanced*
848 *Engineering Informatics*, 25(4), pp. 736–747.