THE SITUATION ENGINE: A HYPER-IMMERSIVE PLATFORM FOR CONSTRUCTION WORKPLACE SIMULATION AND LEARNING

Sidney Newton, Russell Lowe, Rosamond Kember, Rui Wang and Stephen Davey
The University of New South Wales, Australia

ABSTRACT: *The prospect of being able to place an individual within an entirely interactive, simulated environment has long been held, but only recently is it being realized. Flight simulators were the first to provide a hyper-immersive experience using a combination of very detailed and accurate models of aircraft systems, high-resolution visualization and motion platforms. More recently, advanced video game technologies have been coupled with augmented reality systems and sophisticated tracking technologies to provide hyper-immersive experiences of battlefield conditions, crime scenes, operating theatres, industrial processes, etc. A key problem for developers of any hyper-immersive environment is the significant overhead costs of modeling, programming, display technologies and motion simulation.*

The Situation Engine is an application platform that provides for specific and managed building and construction experience to be made available using low-cost, advanced digital technologies. The same engine can drive a multitude of learning situations. Multiple users collectively occupy the same simulated workplace but experience that situation individually by individual movement through the space. Head tracking, gesture recognition, voice communication, 3D head-mounted displays, location-based sound and embedded learning resources have all been incorporated into the Situation Engine at minimal cost. The total enabling technology cost per participant is currently around $600 Australian.

This paper will focus on the hyper-immersive nature of the Situation Engine. In particular, the distinction between immersion (as a quantitative measure of sensory fidelity) and presence (as a qualitative perception of 'being there') will be articulated and clarified. The paper also highlights one of the various ways in which hyper-immersion is manifested in the Situation Engine: gestural control. Gestural control has been implemented using a Microsoft Kinect™ and proprietary gesture detection algorithms to monitor a range of gestures in parallel, including gestures that are context dependent.

KEYWORDS: *Simulation, Hyper-Immersion, Cost, Situation Engine, Gestural Control.*

1. CONSTRUCTION WORKPLACE SIMULATION AND LEARNING

Programs of study in the higher education sector are being shaped increasingly by the imperative of employability. A recent review of graduate employability in Australia highlighted the need to provide students with effective "exposure to professional settings" (Cleary *et al*, 2007:10). A national framework for addressing employability skills in Australia subsequently characterised the individual performance requirements in terms of their ability to: "work with increasing levels of autonomy; cope with increasing complexity and uncertainty; and adapt with increasing ease and effectiveness to unfamiliar contexts". (Goodwin *et al*, 2012:15). The focus on employability is casting new light on the broad notion of practice-based experiences, the higher cognitive levels of learning and how this all might best be integrated into higher education programs of study (Billett and Henderson, 2011). Practice-based experiences include such activities as practicums, industry placements, case studies, role play and site visits.

An effective practice-based experience is one in which the learning outcomes are realised in terms that are both deliberate and intentional (Washbourn, 1996): deliberate in the sense that the process of learning is managed effectively and controlled from a pedagogical perspective; intentional in the sense that the particular skills to be developed through that experience are made explicit, can be demonstrated by the student and are assessable. However, the many practice-based experiences that do provide for explicit student development and assessment also tend to be problematic when it comes to exercising control. For example, an industry placement can develop

Citation: Newton, S., Lowe, R., Kember, R., Wang, R. & Davey, S. (2013). The situation engine: a hyper-immersive platform for construction workplace simulation and learning. In: N. Dawood and M. Kassem (Eds.), Proceedings of the 13th International Conference on Construction Applications of Virtual Reality, 30-31 October 2013, London, UK.

important practical skills but the resources required and opportunities available to target particular skills are often prohibitive. Those practice-based experiences that might lend themselves to more direct control also tend to be more abstract in their learning outcomes. For example, a case study can be selected carefully to address a particular issue or skill but does so vicariously or once removed from the actual activity itself.

The current context of higher education in Australia, as it is elsewhere, is one of broadening access and participation in a climate of structural and organisational change (Bradley *et al*, 2008). Higher education must cater for increasing numbers of students and do so with a tightening of the available resources. In that context the integration of effective practice-based learning is increasingly problematic. The potential for practice-based learning to be exercised using relatively inexpensive and scalable new simulation and learning technologies is a compelling one. This paper will describe the technical configuration and functionality for a new concept in simulation and learning technology, The Situation Engine. The Situation Engine is an application platform that provides for specific and managed building and construction experience to be made available using low-cost, advanced digital technologies.

2. THE SITUATION ENGINE

We define the Situation Engine as:

"An application that provides for specific and managed practical building and construction experience to be made available to students through advanced digital technologies."

Figure 1 begins to un-wrap this definition in more functional terms. Each specific situation is comprised of: certain environmental conditions (weather, time, location, etc.); objects and their properties (buildings, equipment, materials, etc. with dimensions, mass, movement, density, etc.); actors and their behaviours (characters, interfaces, avatars, etc. with behaviours, scripts, intelligence, etc.); and data feeds (web, video, motion, devices, etc.). Various combinations of environments, objects, actors and feeds constitute a particular situation. Each situation is then articulated as a series of interactions. The interactions are not prescribed, but rather emerge from the basic physics and decision-making that governs the behaviour of environments in certain conditions, objects with certain properties, actors with certain behaviours and feeds with certain data manipulation. Howsoever the complex interactions resolve themselves at any given moment in time, is then rendered to the user as a display of some form (screen, goggles, digital cave, 3D, soundscape, etc.). The user interface needs to deliver an immersive, first-person experience of the situation to the user as it unfolds. The more realistic the immersive experience the better Bystrom *et al* (1999). First-person engagement is critical to an immersive experience in this context, as the specific situation is then presented as a person would typically engage with the world. Clearly, the same Situation Engine is intended to drive a multitude of tailored learning situations.

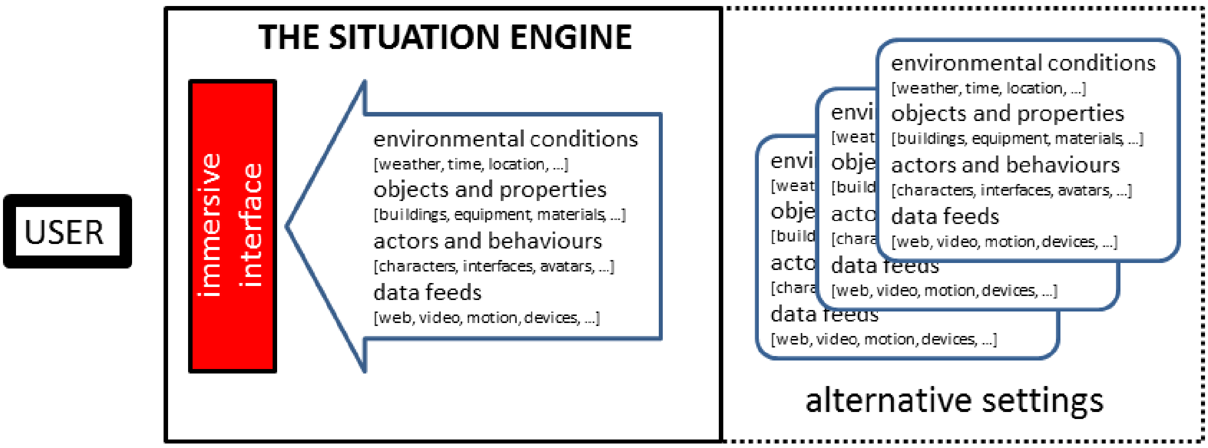

Fig. 1: Concept Structure for a Situation Engine

The target curriculum for the current Situation Engine development is the 1st year course of a 4 year program of undergraduate study in construction management and property in Australia. The course is the first in the program of study to introduce students to construction technology. It deals with the functional requirements and construction methods specific to residential/domestic construction typical in Australia. As such the course examines a range of key technical aspects, including: brick and timber frame construction methods and materials; domestic joinery; staircase construction; finishes; plumbing, drainage and electrical services; methods of setting out and supervision. The course also involves developing skills in on-site observation and the production of housing site reports.

A formal process of human factor analysis using focus groups and task analysis has been undertaken, along with an analysis of the learning needs of current students (Newton, 2012). For instance, the learning needs were assessed by reviewing the performance of several hundred students in their end-of-year examinations, to identify those topics where students were having problems and the typical mistakes they were making specific to construction technology. A small reference group of users has been established to trial prototype systems and evaluate various implementations. Formal evaluation of the system is ongoing.

3. A HYPER-IMMERSIVE PLATFORM

The most sophisticated interactive virtual reality simulation environments with practical application to teaching and learning are to be found in video games. Of particular relevance is the recent emergence of providers making the game engines themselves (the kernel of coding used to drive a collection of actual game implementations) available on an open-source basis. The most powerful game engines are now typically free to acquire for teaching and learning purposes, they allow third party modifications and are supported online by a significant and committed community of users and developers. The Situation Engine has been implemented using the proprietary video game engine CryENGINE®3. This engine features easily the most advanced graphical, physical and animation technologies available (see: http://www.crytek.com/cryengine).

The quality of the visual rendering in CryENGINE®3 is illustrated in Figure 2, which is a screen grab from the current implementation of The Situation Engine. It shows a situation where the construction project has progressed through excavation, foundations and is nearing completion of timber framing. The site shows various workers, plant and equipment, facilities and materials – as seen through the eyes of the user avatar. The same site can be used to present alternative situations, at different stages of construction, with different activities and other configurations of material storage, signage, waste management, security, etc. Multiple users can be represented and experience the situation collectively. At various points a user can interact with the building as it is constructed – checking the placement of reinforcement and formwork just prior to pouring the slab, for example. Users get to see how the work at different stages of construction has been prepared, measure and check sizes and distances, assess details against building codes and best-practice guides, etc. Assessment tasks can test a students' understanding of related issues, such as safe work practices, material storage and handling considerations, site security, environmental protection, wet-weather hazards, noise pollution, etc.

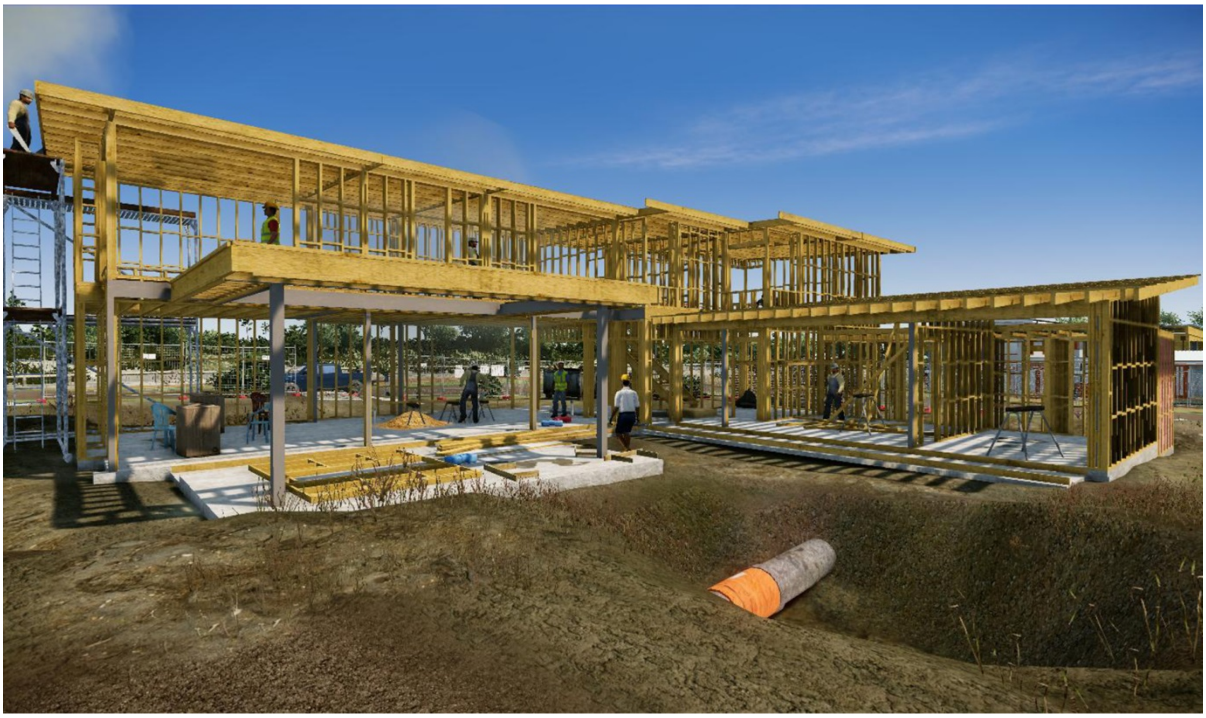

Fig. 2: Screen Capture of Construction Site Rendering in The Situation Engine

Immersion is a critical feature of virtual reality simulation. Slater and Wilbur (1997) define immersion as “an objective description of aspects of the system such as field of view and display resolution”. The term is still used to refer to the quantifiable aspects of display technologies (such as vividness, resolution and display dimensions). With advances in virtual reality technologies and the blurring of display with other aspects of the virtual reality simulation, immersion is now also used to refer to a broader compass of representation technologies, including the self-representation of the user/avatar, the physics of the models, the sound quality, etc. This broader range of immersive qualities is critical because they each contribute to a users’ overall sense of presence.

In contrast to immersion, presence is used to refer to the subjective phenomenon of ‘being there’ (Heeter, 1992). Presence is a product of the mind and independent of any specific type of technology or technological context. A greater sense of presence is often associated with improved user performance (Bystrom *et al*, 1999).

The relationship and distinction between immersion, presence and performance can be confusing, as the terms are often conflated in the literature. Figure 3 seeks to clarify the meanings used in this paper. In this model, the level of immersion is solely determined by the technical qualities of the virtual reality simulation technology. The better the quality of the rendering, the way objects behave, the soundscape, the currency of the information, the social interaction, bodily engagement, haptic feedback, etc., the greater the level of immersion. But the measure of immersion is expressed entirely in technical terms of frame rate, resolution, lag, frequency, etc. What Bystrom *et al* (1999) refer to as the ‘sensory fidelity’ of the technology.

Beyond a measure of the sensory fidelity is the sense of presence experienced by a user. It is generally presumed that a greater sensory fidelity will lead to a greater sense of presence, but the issue is more complex (van den Hoogen *et al*, 2009). Steuer (1992) notes that in addition to the sensory fidelity (what he refers to as the vividness and interactivity measures of a virtual reality) the sense of presence will be directly influenced by the individual characteristics of the user and the particular task activity in which they are engaged at the time. In other words, the cultural and experiential background of the user and the purpose and process with which they employ a given virtual reality simulation will significantly influence the sense of presence they perceive.

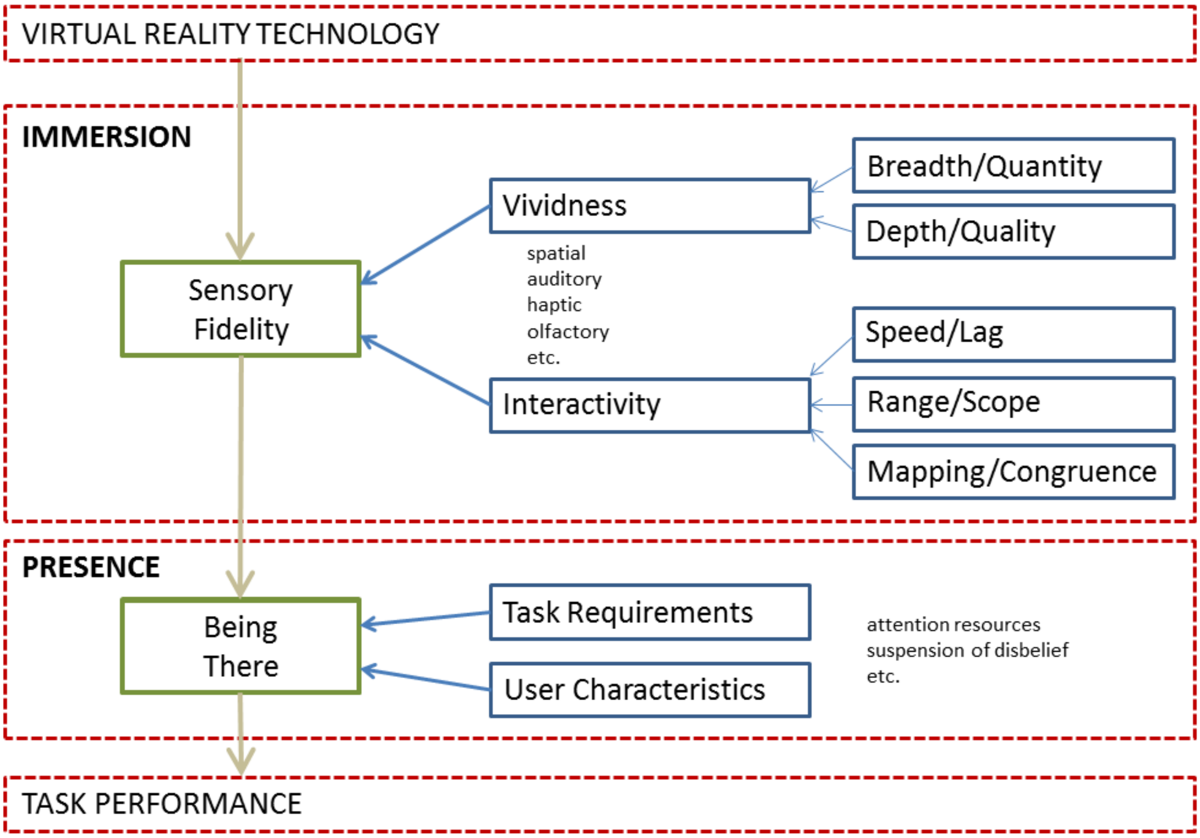

Fig. 3: Elements of Immersion and Presence (adapted from Steuer, 1992 and Bystrom *et al*, 1999)

Notwithstanding the complicating nature of presence on immersion, or other things being equal, increased levels of sensory fidelity do equate with increased immersion. A hyper-immersive platform is defined here as “a virtual reality simulation system where the fidelity of the various sensory channels is maximized within current technical capacities”. That is to say, spatial modeling provides near-photorealism for indoor and wide-open outdoor environments, auditory modeling provides 3D locational sound reproduction and syncronised voice animation, haptic modeling includes full body tracking and gestural control, etc.

For example, in terms of bodily immersion the Situation Engine has been developed to enable the user to control the behavior of their avatar in the system using bodily movements. Control of this kind is termed ‘gestural control’. Gestural control is an important aspect of immersion because gestures are often more intuitive for users, particularly those inexperienced with established gaming interface technologies. Gestural controls have the significant advantage over traditional controls (such as keyboard and mouse) because they can map user actions directly to their congruent actions in the virtual world – a swing of the user arm equates with a swing of the virtual arm. Key issues for the development of improved gestural control relevant to the Situation Engine include:

(i) The need for low latency, real-time interactivity. This is required to ensure that gestures are interpreted into actions in the virtual environment with the same immediacy as for real-world actions. This required working directly with the base-level coding and software development tools.

(ii) Design of the gesture detection algorithm to recognize and interpret natural gestures appropriately. This required careful filtering of, and focus on, relevant body parts to reduce the total computation load. For example, many gestures could be limited to tracking and analysis of the upper arm and forearm movements only.

(iii) Capacity to perform multiple gestures in parallel. Walking, turning, leaning and swinging a hammer, are all possible at the same time in the Situation Engine's gesture control system. The system monitors body movements for all and any registered gestures in parallel.

The Microsoft Kinect™ provides skeletal tracking data for 21 virtual bones, including head, arms, legs, hips, neck, and spine. The skeletal framework is approximated to fit the body of the user as it is scanned for shape and depth.

Data on the movement and position of each virtual bone is parsed in real-time by specially developed detection algorithms, which interpret sequences of movement into registered gestures. Gestures currently detected include turning, jumping, leaning in any direction, moving at gesture-controlled speeds on a smooth and continuous spectrum from walking to running. More sophisticated, context-dependent gestures are also monitored - reaching out with the hand to open a door, entering the cab of an excavator or grabbing a small hand tool, for example. Further gestures are being added, with the intention to exhaustively yet efficiently map a full compass of potential user actions.

To implement the gesture control system required basic modification to the CryENGINE®3 base-level coding (C^{++}). A multiplexer class was created such that data retrieved from the Microsoft Kinect™ could be served out to an arbitrary number of gesture detection algorithms with minimal impact on memory footprint and real-time performance. Gesture detection algorithms were implemented at the higher level as nodes in the CryENGINE®3 control flow graph (CFG). This provided ready access to gesture detection, as and when required. It also defined the related actions in the virtual environment that each gesture was intended to initiate.

Various other modifications and supplements to the immersive technologies in CryENGINE®3 have been developed to create a hyper-immersive environment. However, the abiding issue with any hyper-immersive virtual reality simulation is the associated costs. By way of context, the video game industry has now overtaken the film industry in terms of its overall investment and revenue values. Several "Call of Duty" titles, for example, have grossed over 1 billion dollars. The most recent release, "Black Ops", grossed that amount in just the 6 weeks following its release (see: http://www.vgchartz.com/article/250163/call-of-duty-a-sales-history/). In such a context it is hardly surprising perhaps that the budgets for developing hyper-immersive commercial games are also very large. Indeed, the development budget for "Black Ops" is believed to have been between $18-20 million, with an associated advertising and marketing campaign somewhere in the order of $250 million (see: http://agreatbecoming.com/2011/02/08/call-of-duty-black-ops-return-on-investment-is-4350/). How might the development costs for serious simulations be contained to anywhere near a more feasible level?

4. COSTS AND RESOURCES

The Situation Engine was developed with modest funding from the Australian Government Office for Learning and Teaching as a teaching project initiative. As a non-commercial development, the project was able to make use of the free TDK download of CryENGINE®3. This left the majority of direct development costs specific to the creation of environmental content. Content can broadly be split into two main categories: 3D models and 2D textures. 3D models include the visible geometry and collision volumes (required for physics calculations). 2D textures include the full complement of texture definition files needed to render specific materials (it is not uncommon to have diffuse, normal, detail, specular and bump maps all operating in concert to convincingly represent a material), particles and clouds (such as water, smoke and even insects), images and decals (overlay textures that are used to give basic character to surfaces or for labels and signage) and for various elements of the user interface. Costs for environmental content was kept to a minimum by, as far as was practical, modifying the existing game assets available from the gaming community and other online resources. Hundreds of unique characters, avatars, vehicles, construction equipment, building materials, signage and building elements were created in this way. However, given the need to represent authentic Australian construction practices and building codes, a significant effort was involved in the creation of tailored building elements and construction details. These were typically drafted using Sketch-Up® and Autodesk® 3ds Max® specific for the game engine, imported and textured. A range of leading community modeling and texturing software is supported directly into CryEngine. Alternative transfer protocols are supported using XML-based schema such as COLLADA™ (see: https://collada.org/). The full development workflow is illustrated in Figure 4.

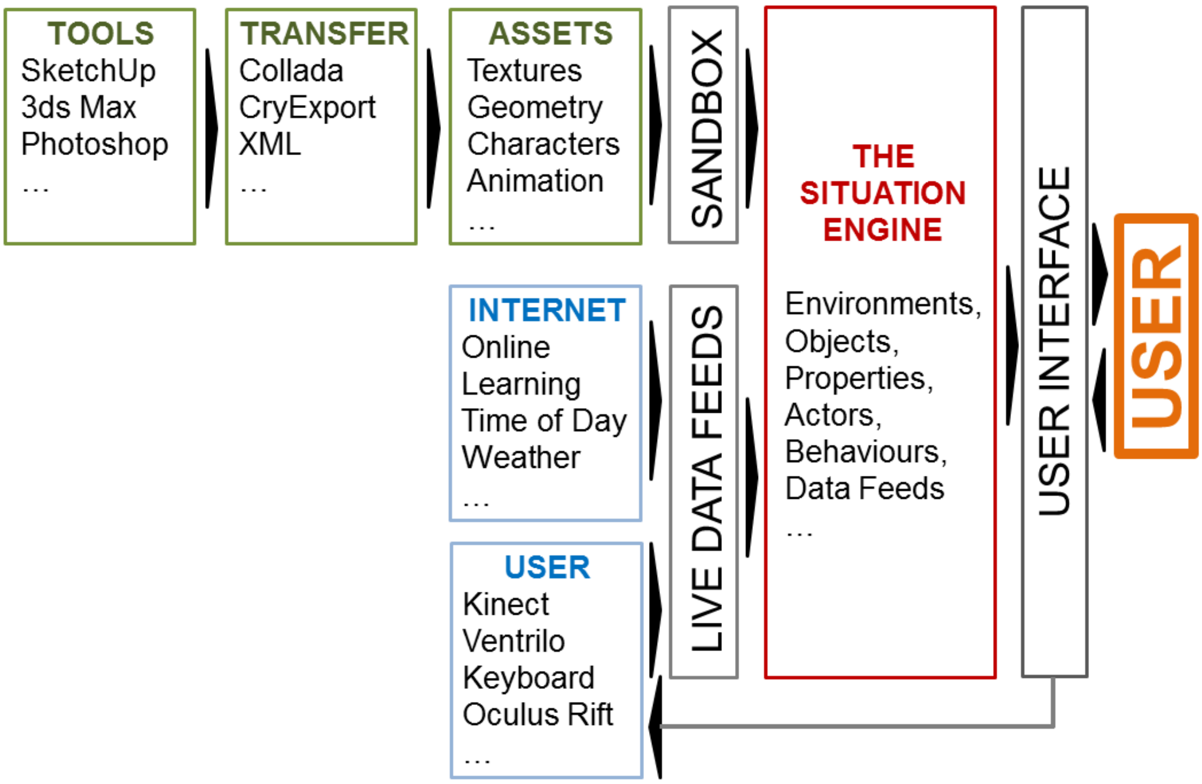

Fig. 4: Schematic of the Development and Deployment Workflow for The Situation Engine

Once the content has been modeled it must be animated and the interactivity defined. In the case of CryENGINE®3 the control of objects, agents and interactivity is managed using a system of control flow graphs (CFG's). The CFG is a visual representation, using graph notation, of the relationships and connections between different entities and events in the game. They define the routes available during execution by specifying how each individual entity or event will respond to given input or context variables. In a similar manner to visual programming, CFG's provide sophisticated programming capability without the need for specific programming language expertise. The WYSiWYP (What You See is What You Play) functionality in CryENGINE®3 is called 'The Sandbox', which provides full game authoring capabilities at a high level of definition. Functionality such as this has enabled the Situation Engine to be developed by non-computer scientists. Indeed, much of the development has been achieved by current undergraduate and recent graduate students in Architecture and Architectural Computing. Our estimate of the total time taken to develop the current Situation Engine implementation is 1-2,000 person hours, or somewhere around $50K. That does not include the supervisor time or time spent developing previous implementations, but is a reasonable measure of the effort required to progress from a blank, flat site to a fully operational, complex and accurate situation where all of the building content was modeled from scratch, based on actual construction drawings.

The development costs of each Situation Engine implementation may be relatively small compared to commercial block-buster games, but might still be prohibitive. Our experience in making the current system available to other academics has been that individuals tend to have a strong preference to tailor their student interaction to meet specific needs in particular ways. Two further resourcing efficiencies have been achieved:

(i) For staff or students with a basic technical competence in video game development (even basic CAD modeling skills), much of the existing content can be reused and modified through recombination and parametric adjustment. For example, the standard character available in CryENGINE®3 is a combat trooper. The standard character has over 130 default animations and movements available. It is possible to create or purchase at very low cost a huge range of custom meshes and clothing that can be simply added to the standard character rig (skeleton) to take advantage of the default movements available (see: http://www.the3dstudio.com/). Every characteristic imaginable, skin colour, gender, clothing, age, height, hair, etc., can be varied using the same standard character rig. Similar modifications can be made to existing vehicles to create entirely new forms of transport. For example, we have modified the standard military tank to look and perform as a hospital wheelchair.

(ii) For staff with no competence or interest in video game development, the same situation can be utilized in a variety of ways to teach a variety of topics – from construction technology to health and safety to site management. For each of these topics the situation can be the same, but the student interaction does need to be directed and supported differently. A recent innovation has been to incorporate live access to the internet using standard browser technologies placed in various locations within the situation – the browser can be mapped onto a computer or tv screen model located in the situation, or be a pull-down option available to the user at any point in time or location. Users (students) can interact with these browser windows from within the game with full browser functionality. This means that staff can set tasks, monitor progress, provide links to resources and generally provide online teaching support directly into the Situation Engine, using standard online teaching technologies. Staff need only tailor their online teaching resources to fundamentally alter the student experience of the particular situation, without ever having to modify the Situation Engine itself.

More generally, the cost of deployment of the Situation Engine is also falling. The cost of computer hardware required to run the system is difficult to estimate as CryENGINE®3 provides a range of native ports to a variety of platforms from Xbox 360™ and PlayStation®3 to standard PC laptops, workstations and remote access grid arrays. Whilst the licenses required to port to Xbox 360™ and PlayStation®3 are themselves prohibitively expensive, because they presume a commercial application, that possibility puts the purchase of entry-level technology at potentially about $300. Even cheaper and more powerful cloud computing is already revolutionizing gaming, and is only limited by the download/upload connection speeds. Associated functionality is also becoming very inexpensive. For example, Oculus VR™ are currently shipping development kits for a head-mounted display (the 'Rift') that provides exceptional field of view characteristics with very low-latency (see: http://www.oculusvr.com/). The Rift has an anticipated retail price of $300. We utilize the Microsoft Kinect™ to recognize gesture controls and track the physical movement of the user in their space to move virtually through the Situation Engine environment. This retails separately at around $140, or comes bundled with the Xbox 360™. When using the multi-player mode to collaborate with other users, we utilize the Ventrilo surround sound VoIP (Voice Over the Internet Protocols) group communication software (see: http://www.ventrilo.com/). This application is free. The current Situation Engine implementation, developed with Australian Government funding, is provided free of charge under a Creative Commons Attribution 3.0 Australia license (see: http://creativecommons.org/licenses/by/3.0/au/). The potential collective baseline cost of all enabling technologies and equipment per user, for the full scope of functionality, is therefore around $600. The full scope of potential functionality and various downloads can be reviewed/accessed at the Situation Engine website (see: http://www.be.unsw.edu.au/programs/situation-engines/).

5. CONCLUSIONS

Given a growing focus on employability and the need that such a focus lends to providing more substantive and more effective practice-based experiences to students, this paper describes the technical configuration and functionality for a new concept in simulation and learning technology – The Situation Engine. The Situation Engine provides for specific and managed practical building and construction experiences using advanced video game technologies. Each situation comprises a dynamic combination of environmental conditions, objects and properties, actors and behaviours and external data feeds. The current implementation of the Situation Engine is specific to residential/domestic construction technology, characteristic to Australia. It provides a hyper-immersive virtual reality simulation of a residential building project at various stages of construction, from excavation through foundations, timber framing and finishes to fitout.

The focus of this paper has been on the hyper-immersive nature of the Situation Engine: how immersion is defined and how it is realized. In particular, the distinction between immersion (as a quantitative measure of sensory fidelity) and presence (as a qualitative perception of 'being there') has been articulated and clarified. The paper has highlighted one of the various ways in which hyper-immersion is manifested in the Situation Engine: gestural control. Gestural control has been implemented using a Microsoft Kinect™ and proprietary gesture detection algorithms to monitor a range of gestures in parallel, including gestures that are context dependent.

An abiding issue with any hyper-immersive simulation technology, especially those developed for teaching and learning, is cost. A schematic of the development and deployment workflow for The Situation Engine is presented that supports particular resource efficiencies. Our estimate of the total time taken to develop the current Situation Engine implementation is in the order of 1-2,000 person hours, or somewhere around $50K. Such development costs are trivial in the broader context of commercial video game development, but may still be prohibitive. Two particular resourcing efficiencies are described that reduce the potential burden of development further: direct development of a situation by modification and indirect development by tailoring online teaching and learning resources. Finally, consideration is given to the baseline cost of deployment. A very near future baseline figure for all enabling technologies and equipment per user is in the order of $600.

Hyper-immersive virtual reality construction workplace simulations are now available at very reasonable development and deployment costs. The quality of the simulations is such that broader development offers significant potential to address the growing issue of how workplace experience can usefully and efficiently be provided to large cohorts of students. The Situation Engine marks genuine progress towards that potential.

6. ACKNOWLEDGEMENTS

Support for this research has been provided by the Office for Learning and Teaching (OLT), an initiative of the Australian Government Department of Education, Employment and Workplace Relations. The views expressed here do not necessarily reflect the views of the OLT.

7. REFERENCES

Billett, S. and Henderson, A. (eds) (2011) *Developing Learning Professionals: Integrating Experiences in University and Practice Settings*, New York: Springer.

Bradley, D., Noonan, P., Nugent, H. and Scales, B. (2008) *Review of Australian Higher Education: Final Report*, available: http://www.deewr.gov.au/he_review_finalreport/ [accessed on 30 January 2011].

Bystrom, K-E., Barfield, W. and Hendrix, C. (1999). A Conceptual Model of the Sense of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 8(2), 241-244.

Cleary, M., Flynn, R., Thomasson, S., Alexander, R. and McDonald, B. (2007) *Graduate Employability Skills*. available: http://aces.shu.ac.uk/employability/resources/ [accessed on 16 May 2013].

Goodwin, S., McDonald, R., Perkins, K., Wignall, L., Gale, G., O'Callaghan, K., Murray, F., McGraw, J., Redman, K. and Hopwood, J. (2012) *Employability Skills Framework Stage 1: Final Report*, available: http://deewr.gov.au/employability-skills [accessed on 16 May 2013].

Heeter, C. (1992). Being there: the subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, 1(2), 262-271.

Newton, S. (2012) A Situation Engine for Teaching and Learning Residential Construction Technology, in Sulbaran, T. (Ed), *Proceedings of 48th ASC Annual International Conference*, pp.9-17, available: http://ascpro0.ascweb.org/archives/cd/2012/paper/CEUE201002012.pdf [accessed on 23 March 2013].

Slater, M. and S. Wilbur (1997) A framework for immersive virtual environments (FIVE) : Speculations on the role of presence in virtual environments, *Presence-Teleoperators and Virtual Environments*, 6(6), 603-616.

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4), 73-93.

van den Hoogen, W. M., IJsselsteijn, W.A. and de Kort, Y.A.W. (2009) Effects of sensory immersion on behavioural indicators of player experience: movement synchrony and controller pressure. *Breaking New Ground: Innovation in Games, Play, Practice and Theory. Proceedings of DiGRA 2009*. London: Brunel University, pp.1-6.

Washbourn, P. (1996). Experiential learning: Is experience the best teacher?, *Liberal Education*, 82(3), 1–10.