

Evaluating LevelEd AR: An Indoor Modelling Application for Serious Games Level Design

Lee Beever

Department of Computer Science
University of Chester
United Kingdom
l.beever@chester.ac.uk

Serban Pop

Department of Computer Science
University of Chester
United Kingdom
s.pop@chester.ac.uk

Nigel John

Department of Computer Science
University of Chester
United Kingdom
nigel.john@chester.ac.uk

Abstract— We developed an application that makes indoor modelling accessible by utilizing consumer grade technology in the form of Apple’s ARKit and a smartphone to assist with serious games level design. We compared our system to that of a tape measure and a system based on an infra-red depth sensor and application. We evaluated the accuracy and efficiency of each system over four different measuring tasks of increasing complexity. Our results suggest that our application is more accurate than the depth sensor system and as accurate and more time efficient as the tape measure over several tasks. Participants also showed a preference to our LevelEd AR application over the depth sensor system regarding usability.

Keywords— *augmented reality, level design, serious games.*

I. INTRODUCTION

When developing serious games, designers often create virtual worlds from scratch that facilitate the user and the intended experience. However, we believe a serious game virtual world, which employs a personal space or addresses prior user knowledge, will benefit from being based on a real-world location rather than imagination. Serious games such as a virtual reality (VR) wheelchair driving simulator [1], virtual evacuation training [2] and virtual fire safety training [3] would all benefit from allowing users to train in particular environments in which these events or users utilise. In all these cases, rigorous spacing and accurate depiction of distances and gaps play a very important role in the simulators’ efficiency and usability. There is also a potential benefit for entertainment and serious games applications that support passive haptics and substitutional reality [4].

To enable the complex task of indoor modeling a multitude of techniques are currently available. Static and mobile laser scanners are used to create complex point cloud virtual models which are commonly used in construction [5]. Mapping systems that utilize infra-red (IR) depth sensors [6] to model an indoor space are beginning to be used by interior designers and builders. Manual capture of measurements with a tape measure on a floor plan are often used for do-it-yourself (DIY) projects. These techniques, whilst effective can be time consuming and are not always accessible, due to cost and technical ability required for use. The resulting model/data produced is often not suitable for use in a serious game without significant adjustments.

This paper describes an augmented reality (AR) smartphone application called LevelEd AR built using Apple’s ARKit. This application allows users to capture a model of a real-world location that is suitable for use in a game engine, such as Unity to aid serious games level design. This paper contributes an evaluation of accuracy and usability of

our system compared to alternative low cost options as well as an example workflow for the system (Fig. 1).

II. RELATED WORK

To our knowledge, there are no academic works currently utilizing or evaluating Apple’s ARKit augmented reality framework as part of an indoor modelling system. Similar commercial systems now exist [7], but these focus mainly on the floor plan and not size and placement of objects within the space. There are, however, several academic works utilizing ARKit to develop systems in other domains [8][9]. Fusco and Coughlan [8] utilized ARKit to develop an indoor localization system for users with visual impairments. Whilst Dilek and Erol [9] produced an educational system for generating position-time graphs in real-time. Our work will contribute further to the results of these papers on the accuracy, usability and issues of ARKit and will benefit future systems.

The work on non-ARKit indoor modelling techniques is numerous. Systems built around laser scanning and photogrammetry have existed for many years. These systems have been used for building information modelling (BIM) [10]. This work has resulted in industrial laser scanning systems now commonly used in construction to create point cloud models of a site or indoor space [6]. However, these point cloud models are not suitable for use in serious games due to their complexity, lack of polygonal mesh data and inclusion of every object in that space present at the time of the scan. Turner and Zakhor [11], developed a system that initially generates 2D floor plans from complex point cloud data and then extrudes a simplified 3D model from the floor plan. This system is mostly effective at developing indoor models that are more suitable for serious game virtual environments than raw 3D point cloud data. However, it still requires the initial collection of 3D point cloud data from laser scanners, which are costly; and the system focuses only on recreating walls and not furniture or objects within the space. Some researchers have also explored systems that utilize infrared (IR) depth sensors to model indoor spaces in real-time. Kalantari and Nechifor [6], developed a custom application that utilises Occipital’s Structure Sensor attached to an iPad to model indoor spaces in real-time by scanning the area with an iPad. However, this often produced models that suffered from walls collapsing inwards when multiple walls were mapped. The system also requires additional or specific hardware to work that can be complex for users to learn and costly. LayoutNet [12] solves this problem by reconstructing a room layout in 3D based on a single RGB panoramic image using a convolutional neural network (CNN). The system is reasonably effective for standard shaped rooms but struggles with irregular rooms and only maps the wall.

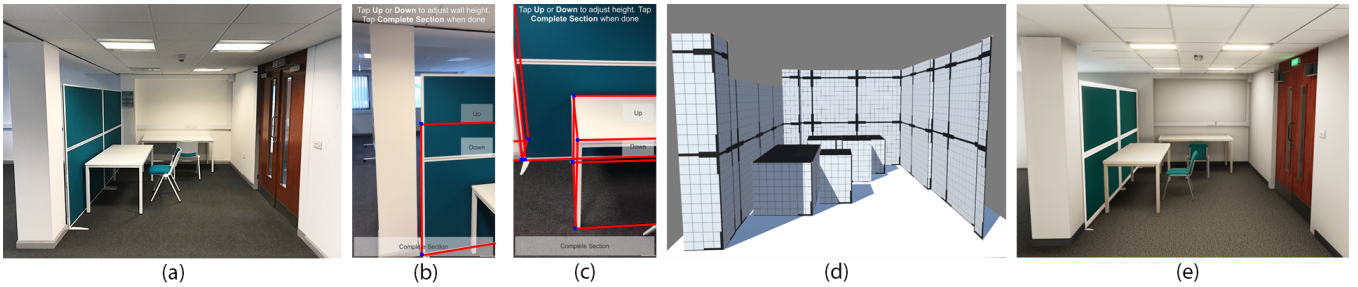


Figure 1. LevelEd AR workflow from LevelEd AR to Unity game engine. (a): Users selects a room to model (b): User is mapping the dividing wall; the shape, scale and location will be captured. (c): User has completed the mapping of the shape, size and location of the table. (d): The data has been downloaded and the model generated in Unity. (e): A users has used the model as a guide to develop a 3D virtual environment for a serious game with correct scale.

Based on our review of the literature an indoor modelling system that is suitable for serious games level design has not been developed or investigated. Such a system should: produce a simplified mesh model that can be used directly in a serious game virtual environment or as a guide during development; model both walls and objects within the space, accurately placed with respect to their location, where necessary, including irregular rooms; give the user the choice as to what parts of a space are modelled or not modelled; be accessible to users without the requirement of additional costly and complex hardware. We have addressed these issues with the development of our LevelEd AR indoor modelling application which is described in the next section.

III. SYSTEM OVERVIEW

One of the aims for LevelEd AR was to ensure it is widely accessible by making use of readily available consumer technology. LevelEd AR was built using Unity 2017.1, and Apple’s ARKit 1.0 was selected for this project due to the wide availability of existing compatible devices. As of July 2017, there were an estimated 380 million ARKit compatible smartphone devices. This is expected to grow to 850 million by 2020 [13]. There is also potential to port the application to Google’s ARCore for Android devices to further improve availability.

With LevelEd AR, users can model the scale, location and general shape of walls and objects in a real-world location using an AR view. Users can model walls by placing AR markers at intersections of walls within a room to map out the base of the walls (Fig. 1b). Users can also model 3D objects by placing markers to surround the object (Fig. 1c). This can be an object of any number of sides but in the experiment, it was set to four to aid usability for new users. Once the base markers are in place for a wall or object, the user can raise up a second set of markers and connecting edges for the height of the wall/object. This results in a wireframe model of the mapped objects.

LevelEd AR makes use of several key ARKit functions, such as the ability to detect horizontal planes and key points of interest. The system works by casting a ray into the scene from the centre of the screen (filled with the AR camera view). A marker object tracks the raycast hit location and can then be anchored in place with a tap of the screen. The marker locations are used to create data in the form of wall objects (a series of planes) or 3D objects (of any number of sides). The data is serialized to a file and then uploaded to a webserver once complete. In Unity, the data can be downloaded and a model of the environment generated (Fig. 1d) from the data to be used as part of the level design process as a guide or in some cases such as walls, used in the final version of the level/virtual environment (Fig. 1e).

IV. EXPERIMENT METHODOLOGY

To evaluate the accuracy and usability of LevelEd AR, participants were asked to complete four separate measuring tasks with the time taken to complete each task also recorded. These tasks were as follows: **Task 1:** Measuring a single wall – the length and height were captured. **Task 2:** Measuring a small cupboard with open space around all sides – the length, depth and height of the small cupboard were captured. **Task 3:** Measuring a large cupboard against a wall – the length, depth and height of the large cupboard were captured. **Task 4:** Measuring four consecutive walls and a small filing cabinet – measurements recorded were the length and height of each wall (Task 4.1), the length, depth and height of the small filing cabinet (Task 4.2) and the distance of the cabinet from the first wall (Task 4.3).

Participants completed the measurement of the four tasks using three different measuring instruments. These instruments were selected based on their similarity in cost and accessibility to the proposed system. The instruments were: **Measuring tape and paper:** users manually measured the tasks using a tape measure and recorded the measurements on a sheet of paper provided. **Room Capture application and Structure Sensor:** users used an iPad Pro 10.5" with a Structure Sensor attachment along with the Occipital Room Capture software to scan the task locations and then gather the specified measurements. **LevelEd AR application:** users used an iPhone 7 Plus and the LevelEd AR application to model the tasks in AR. Participants utilized all three measuring instruments to complete all four measuring tasks. A randomized crossover design was used for both the order of measuring instruments utilized and the order of measuring tasks completed.

The experiment was completed by 18 participants recruited from students and academic/support staff from across the university. They consisted of 3 females and 15 males ranging from 18-59 years of age. 27.8% were between the ages of 50-59, 5.6% between the ages of 40-49, 22.2% between the ages of 30-39 and 44.4% between the ages of 18-29. Prior experience of AR was mixed with 16.7% having no prior experience, 38.9% rating themselves as novices, 27.8% rating themselves as intermediate and 16.7% rating themselves as advanced.

V. EXPERIMENT RESULTS

In our analysis, the measuring techniques used are called instruments and denoted with “Tape” for measuring tape and paper (which was also used as the ground truth), “LevelEd” for our AR application and “Structure” for the Structure Sensor and Occipital Room Capture application. The significance was tested by employing a two-way repeated

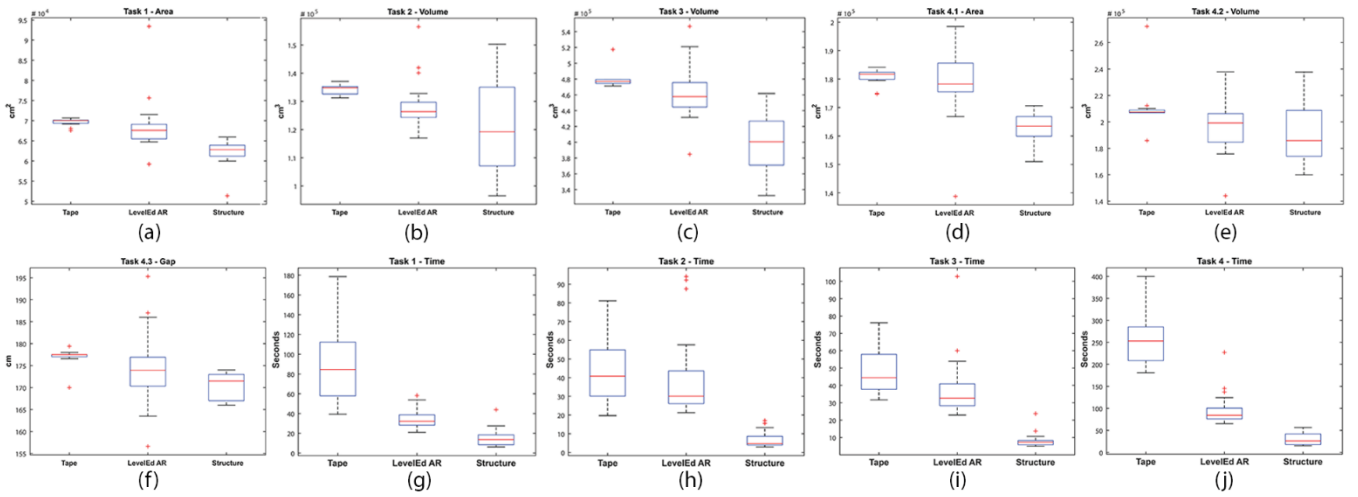


Figure 2. (a): Task 1 – Area measurement of single wall, (b): Task 2 – Volume measurement of small box, (c): Task 3 – Volume measurement of larger box against a wall, (d): Task 4.1 – Area measurement of 4 continuous walls, (e): Task 4.2 – Volume measurement of small box, (f): Task 4.3 – Gap measurement between Task 4.2 box and starting wall of Task 4.1.

measures ANOVA for both measurements and time, a method supported by the very large effect sizes observed throughout. The degrees of freedom were adjusted to the lower bound estimate according to the result of the sphericity test.

A. Measurements

The results show that the instruments significantly differ from each other in terms of performance overall ($F(2,34) = 73.89, p < .001, \eta^2 p = .813$). The same effect was observed for the tasks in all cases, which suggests the tasks vary significantly in complexity ($F(1,17) = 4533.90, p < .001, \eta^2 p = .996$). Moreover, with respect to the interaction between instruments and tasks, we observed that each instrument performs significantly stronger on some of the tasks but weakly on others ($F(1,17) = 20.14, p < .001, \eta^2 p = .542$), an important result which needs to be investigated further. We followed up the significant interaction with six separate one-way ANOVAs. The results were plotted in order to identify and visualize significant trends, which will help characterize better the interaction between instrument and task. Hence, planned contrasts showed that for Task 1, the Tape value was significantly larger than Structure’s measurement ($F(1,17) = 88.47, p < .001, \eta^2 p = 0.839$). The lack of complexity in this task brought no difference regarding traditional point-by-point measurements, however Structure’s under reported measurements could be due to collapsing walls shortening the distances recorded as previously experienced by others [6].

Unlike the first task, the second task showed no difference between LevelEd and Structure, however, both were significantly separated by the ground truth ($F(1,17) = 7.06, p = 0.017, \eta^2 p = 0.293$; $F(1,17) = 13.94, p = 0.002, \eta^2 p = 0.450$). In this task, the complexity increased, and the LevelEd results supported by several large outliers were not significantly different to the large variation in the Structure measurements. In the next two, more complex tasks (Fig. 2c & 2d), the Structure sensor showed a significant loss in accuracy in comparison with the other two instruments (Tape: Task 3 - $F = 86.01, p < 0.001, \eta^2 p = 0.835$; Task 4.1 - $F = 212.13, p < 0.001, \eta^2 p = 0.926$ and LevelEd: Task 3 - $F = 36.59, p < 0.001, \eta^2 p = 0.683$; Task 4.1 - $F = 37.70, p < 0.001, \eta^2 p = 0.689$). Task 3 featured a much larger box situated next to a wall, which increased the complexity of the task. Task 4 required participants to move the iPad more significantly whilst

completing the task with Structure. This often resulted in walls shortening in the scanned model [6], as reported above for Task 1. This was not as pronounced with LevelEd.

Results recorded for Task 4.2, showed the same pattern as for Task 2 where the same type of measurement was required (Fig. 2e). At this task, both instruments employed failed to show differences, providing in the process a loss in accuracy and larger variations over Tape. Some of the factors responsible for this result were the task’s limited complexity, the order completed within Task 4, and subsequent exhaustion of the participants. Another aspect for LevelEd with Task 4.2 is the potential for drift (tracking inaccuracies) to occur, increasing over time. Finally, at Task 4.3, as expected the Tape measurement was larger than Structure’s ($F = 85.63, p < 0.001, \eta^2 p = 0.834$) with LevelEd being no different than the ground truth (see Fig. 2f). However, the larger variation in the measurements of LevelEd may be explained due to the potential for drift to occur more frequently over time with markerless AR [8][9].

B. Time

Similarly, for the time (measured in seconds), significant differences were observed throughout the test between the choices of the instrument ($F(2,34) = 116.99, p < .001, \eta^2 p = .873$). The same effect was observed for the individual tasks, hence their significant difference in complexity was preserved. Moreover, in this case, there were only four tasks as the duration of the sub-tasks of Task 4 were summed up ($F(1,17) = 249.58, p < .001, \eta^2 p = .936$). We observed that each instrument performs significantly different with each completed task ($F(1,17) = 119.36, p < .001, \eta^2 p = .875$).

Following the same pattern, we observed that for the first task (Fig. 2g) the time required to perform the measurements differed significantly between all three instruments, with the Structure instrument being the fastest. Planned contrasts (four separate one-way ANOVAs) showed that for Task 1, the time participants spent using the Tape instrument was significantly larger than using LevelEd ($F(1,17) = 46.75, p < 0.001, \eta^2 p = 0.733$) and Structure ($F(1,17) = 80.23, p < 0.001, \eta^2 p = 0.825$). Moreover the time using LevelEd was subsequently higher than using the Structure ($F(1,17) = 40.44, p < 0.001, \eta^2 p = 0.704$). The results of Task 2 (Fig. 2h) showed that the time spent using Structure was significantly lower than both other

instruments (vs. Tape - $F(1,17)=84.41$, $p<0.001$, $\eta^2_p = 0.832$); vs. LevelEd - $F(1,17)=31.15$, $p<0.001$, $\eta^2_p = 0.647$). The same pattern was observed for Task 3 (Fig. 2i), where using the Structure proved to be the most efficient time-wise (vs. Tape - $F(1,17)=93.36$, $p<0.001$, $\eta^2_p = 0.846$); vs. LevelEd - $F(1,17)=21.18$, $p<0.001$, $\eta^2_p = 0.587$). The similar time between LevelEd and Tape could be due to the boxes being sufficiently small enough for the participant to easily capture the three measurements required to calculate the volume with Tape.

A similar behaviour to Task 1 was registered for the last task, the most time consuming one (Fig. 2j). Planned contrasts showed that for Task 4, the time participants spent using the Tape instrument was significantly larger than using LevelEd ($F(1,17)=103.81$, $p<0.001$, $\eta^2_p = 0.859$) and Structure ($F(1,17)=286.01$, $p<0.001$, $\eta^2_p = 0.944$). Moreover the time using Structure was overall lower than all the others (vs. LevelEd - $F(1,17)=78.78$, $p<0.001$, $\eta^2_p = 0.823$). As the tasks increased in complexity, the gap between Tape and the other two instruments appears to have increased.

VI. DISCUSSION

The results of the experiment show that the LevelEd AR application measurements are closer to the Tape measure than the Structure sensor and Room Capture application in most tasks. For many tasks, especially the ones of increased complexity (such as Task 4), our AR application proves to be more accurate than the Structure sensor and requires less time than the tape measure. This is a major usability and accessibility benefit which enables the users to acquire fast and reliable geometrical information of their environment using consumer technology.

Participants were asked to complete a System Usability Scale (SUS) [14] questionnaire after completing the tasks with each instrument. All three instruments met the SUS usability threshold of 68, with Structure (70) being the least favoured and Tape (76) and LevelEd (74) being closely favoured. This is a positive result and suggests that LevelEd is more accessible than the Structure system for indoor modelling and participants were almost as comfortable using LevelEd as the traditional Tape instrument.

Despite the overall positive accuracy and usability of LevelEd, there were some tasks that either demonstrated minor inaccuracies (Fig. 2b) or variations (Fig. 2f) with measurements. We believe there are several factors that could have caused these inaccuracies/variations. One major factor in the accuracy of LevelEd AR is the problem of drift. Drift occurs when the device loses track of its position in the real world and becomes out of sync with the virtual world. This can cause measurements to become inaccurate as the system may under or over report the distance it has moved since the last marker was placed. This was also noted by [8] and [9] and our work further confirms this. The fourth task was intentionally designed to study the above effect and we believe that the longer the amount of time that is spent on a mapping task then the larger the potential for drift to impact the accuracy, as drift and inaccuracies accumulate over time. This is evident in Fig. 2f which shows larger variations in measurements for LevelEd. This was the longest and most complex task as the box was measured after the four walls thereby increasing opportunities for drift. Potential solutions to the issue of drift could come from improvements to the computer vision algorithm used in ARKit. However, ARKit

changes are not within our remit and instead improvements could be addressed through user experience. We suggest the system could warn users when they are moving too fast or when drift may have happened, prompting users to try again. This is an area that would benefit from further research including tasks requiring longer and more complex measurements, like mapping a full room to see the full effect of drift and potential solutions in action.

VII. CONCLUSIONS

In this paper we have presented a prototype application that enables users to capture indoor models of real-world locations for use in serious games level designs and the application meets the requirements we outlined at the end of section 2. We intend to explore the use of AR and level design further as part of a larger project investigating different level design workflows using AR and VR.

REFERENCES

- [1] N. W. John, S. R. Pop, T. W. Day, P. D. Ritsos, and C. J. Headland, "The Implementation and Validation of a Virtual Environment for Training Powered Wheelchair Manoeuvres," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 5, pp. 1867–1878, May 2018. doi:10.1109/tvcg.2017.2700273
- [2] R. Lovreglio, V. Gonzalez, R. Amor, M. Spearpoint, J. Thomas, M. Trotter, and R. Sacks, "The Need for Enhancing Earthquake Evacuee Safety by Using Virtual Reality Serious Games," *Lean and Computing in Construction Congress - Volume 1: Proceedings of the Joint Conference on Computing in Construction*, Jul. 2017. doi:10.24928/jc3-2017/0058
- [3] L. Chittaro and R. Ranon, "Serious Games for Training Occupants of a Building in Personal Fire Safety Skills," *2009 Conference in Games and Virtual Worlds for Serious Applications*, Mar. 2009. doi:10.1109/vs-games.2009.8
- [4] B. Insko, "Passive Haptics Significantly Enhances Virtual Environments," University of North Carolina at Chapel Hill, 2001.
- [5] R. Volk, J. Stengel, and F. Schultmann, "Building Information Modeling (BIM) for existing buildings — Literature review and future needs," *Automation in Construction*, vol. 38, pp. 109–127, Mar. 2014. doi:10.1016/j.autcon.2013.10.023
- [6] M. Kalantari and M. Nechifor, "Accuracy and utility of the Structure Sensor for collecting 3D indoor information," *Geo-spatial Information Science*, vol. 19, no. 3, pp. 202–209, Jul. 2016. doi:10.1080/10095020.2016.1235817
- [7] Sensopia, *magicplan*. [Application]. Montreal, Canada: Sensopia Inc., 2019.
- [8] G. Fusco and J. M. Coughlan, "Indoor Localization Using Computer Vision and Visual-Inertial Odometry," *Computers Helping People with Special Needs*, pp. 86–93, 2018. doi:10.1007/978-3-319-94274-2_13
- [9] U. Dilek and M. Erol, "Detecting position using ARKit II: generating position-time graphs in real-time and further information on limitations of ARKit," *Physics Education*, vol. 53, no. 3, p. 035020, Mar. 2018. doi:10.1088/1361-6552/aaadd9
- [10] Y. Arayici, "An approach for real world data modelling with the 3D terrestrial laser scanner for built environment," *Automation in Construction*, vol. 16, no. 6, pp. 816–829, Sep. 2007. doi:10.1016/j.autcon.2007.02.008
- [11] E. Turner and A. Zakhor, "Floor Plan Generation and Room Labeling of Indoor Environments from Laser Range Data," *Proceedings of the 9th International Conference on Computer Graphics Theory and Applications*, 2014. doi:10.5220/0004680300220033
- [12] C. Zou, A. Colburn, Q. Shan and D. Hoiem, "LayoutNet: Reconstructing the 3D Room Layout from a Single RGB Image," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2051–2059, March 2018.
- [13] M. Boland, "Artillery Intelligence: 380 Million iPhones Are Compatible with ARKit", Retrieved September 09, 2018 from <http://artillery.co/2017/07/26/artillery-intelligence-arkits-installed-base-is-380-million-iphones/>
- [14] J. Brooke, "SUS - A quick and dirty usability scale," in *Usability evaluation in industry*, Edited by: P. W. Jordan, B. Thomas, B. A. Weerdmeester and A. L. McClelland, London: Taylor & Francis, 1996, pp. 189–194.