EG UK Computer Graphics & Visual Computing (2019)
G. K. L. Tam and J. C. Roberts (Editors)

Assisting Serious Games Level Design with an Augmented Reality Application and Workflow

Lee Beever[1], Serban Pop[1] and Nigel W. John[1]

[1]Department of Computer Science, University of Chester, United Kingdom

Abstract

With the rise in popularity of serious games there is an increasing demand for virtual environments based on real-world locations. Emergency evacuation or fire safety training are prime examples of serious games that would benefit from accurate location depiction together with any application involving personal space. However, creating digital indoor models of real-world spaces is a difficult task and the results obtained by applying current techniques are often not suitable for use in real-time virtual environments. To address this problem, we have developed an application called LevelEd AR that makes indoor modelling accessible by utilizing consumer grade technology in the form of Apple's ARKit and a smartphone. 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. Finally, we carried out a preliminary case study that demonstrates how LevelEd AR can be successfully used as part of current industry workflows for serious games level design.

CCS Concepts

• ***Human centered computing*** → Mixed / augmented reality; • ***Software and its engineering*** → Virtual worlds software;

1. 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.

One example is a wheelchair driving simulator used to train new powered wheelchair users in virtual reality (VR) [JPD*18]. This system may benefit from users training in a real-world location, such as their home or place of work. Using such spaces could help train users to navigate commonly visited locations and potentially reveal accessibility issues within these spaces. Serious games for virtual evacuation training [LGA*17] and virtual fire safety training [CR09] would also benefit from being based on a real-world location by allowing users to train in the actual location of a potential emergency evacuation. 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 benefit for entertainment and serious games applications that support passive haptics [Ins01] and substitutional reality [Sim15][EKL18]. These are forms of augmented virtuality [MK94] which blend full digital worlds with real-world locations. Users of these applications can move around a real world space and interact with walls and physical objects whilst the virtual reality headset displays a digital world on top of the space. Providing tactile feedback on top of visual feedback from a VR headset has shown to improve presence [Ins01][SS05]. With advancements in consumer VR technology improving, standalone 6 degrees of freedom (6DoF) headsets are starting to appear on the market and the need for accessible indoor modeling tools to support the above applications is likely to increase.

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 [VSS14]. Mapping systems that utilize infra-red (IR) depth sensors [KN16] 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 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

DOI: 10.2312/cgvc.20191253

that is suitable for use in a game engine, such as Unity and for serious games level design.

2. 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 [Sen19], 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 [FC18][DE18]. For example, Fusco and Coughlan [FC18] utilized ARKit to develop an indoor localization system for users with visual impairments, whilst Dilek and Erol [DE18] produced an educational system for generating position-time graphs in real-time using ARKit as a foundation. Both works demonstrate the effectiveness for ARKit to track movement of a device in a space, but also report issues with tracking inaccuracies (also known as drift), in spaces that have limited detail on surfaces or where the device was moved too quickly. Our work contributes further to this discussion on the accuracy and usability of ARKit and benefits future systems.

There are numerous works on non-ARKit indoor modelling techniques. Systems built around laser scanning and photogrammetry have existed for many years. These systems have been used for autonomous robot navigation [HKP90] as well as building information modelling (BIM) [Ara07]. 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 [VSS14]. 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.

With the development of laser scanning systems, various works explored generating simpler and cleaner models from the laser scanned point cloud data. Turner and Zakhor [TZ14], 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. Monszpart et al. [MMB*15], presents RAPter, which rebuilds a scene with a regular arrangement of planes from point cloud data. Their system analyses the point cloud data locally and generates planes where necessary. Ochmann et. al [OVW*16], analyses the point cloud data to create volumes that represent rooms and can detect wall depth by analyzing nearby volumes. This creates a model that shows relationships between walls to produce room segments. These systems are mostly effective at developing indoor models that are more suitable for serious game virtual environments than raw 3D point cloud data. However, they still require the initial collection of 3D point cloud data from laser scanners, which are costly; and these systems focus 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. Kintinuous [NDI*11] utilises the Microsoft Kinect IR depth sensor to allow for real-time unbounded modelling of spaces. A complex mesh is created that encompasses walls and objects in the space. Kalantari and Nechifor [KN16], 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. Meanwhile, Angladon et al. [AGC18] developed a system that utilises a Google Project Tango device (a tablet with a built-in RGB-D camera) that can scan a room and allows users to differentiate between walls and clutter to produce a 2D floor plan. ComNSense [ARD*18] utilises 3D point cloud data from RGB-D enabled devices and compresses the point cloud data locally using indoor grammars before uploading to a server. The uploaded data is then extracted into a 2D floor plan. All four systems demonstrate some issues. Kintinuous [NDI*11] produces complex and noisy mesh indoor models that would not be suitable for use in serious game virtual environments. The Structure Sensor application [KN16] often produced models that suffered from walls collapsing inwards when multiple walls were mapped. Whilst [AGC18] and [ARD*18] only produce 2D floor plans that would require additional steps to produce a virtual environment. All four systems require additional or specific hardware to function that can be costly and complex for users learn.

There are also several works that utilize computer vision with monocular camera systems or static images. This solves the problem of requiring complex and costly additional equipment. Rent3D [CSK*15] reconstructs rental apartments from a set of monocular images and a 2D floor plan. The 2D floor plan is extruded into 3D and the system analyses the images to determine which parts of the space they encompass and add the images as textures to the correct walls. The system requires a 2D floor plan to begin with, which limits its effectiveness. Hedau et al. [HHF10], developed a system that analyses static images to detect the estimated size and shapes of walls along with the extents of large furniture in the room. The system is offline, and results are currently not stitched into a full scene. Wall and object scales are estimated and display some errors. LayoutNet [ZCS*18] can reconstruct 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 walls. These systems all show promise but as noted, they all suffer from different problems that mean they are not currently ideal for serious games level design.

Based on our review of the literature an indoor modelling system that is suitable for serious game 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 and thereby opening it up to a larger range of users.

We have addressed these issues with the development of our LevelEd AR indoor modelling application which is described in the next section.

3. System Overview

The system was developed for iOS devices that support ARKit (iPhone 6S and newer) using the game engine, Unity 2017.1 (Unity Technologies, San Francisco, CA). One of the aims for LevelEd

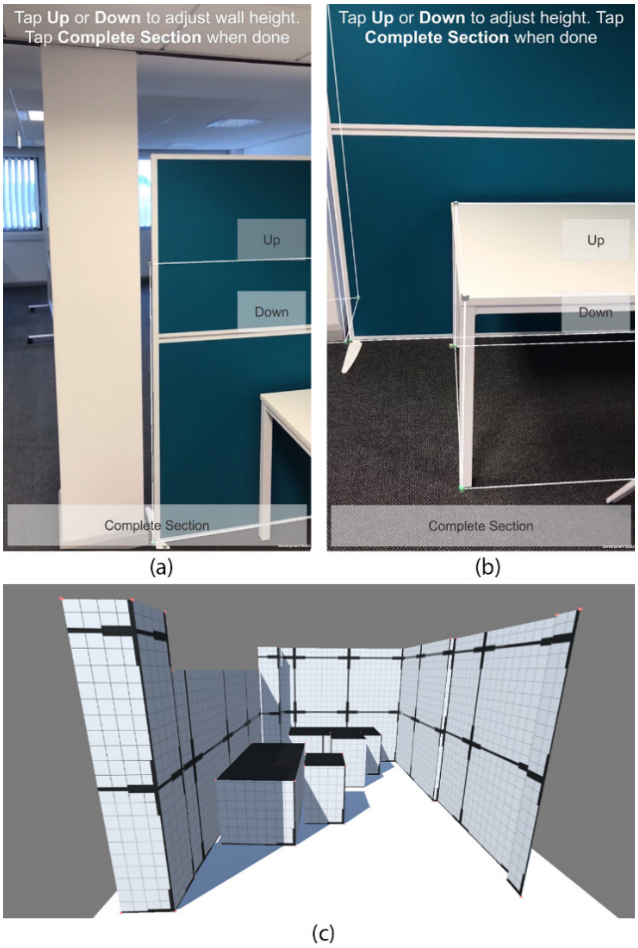


Figure 1: *LevelEd AR workflow from AR to Unity game engine. (a): User is mapping the dividing wall; the shape, scale and location will be captured. (b): User has completed the mapping of the shape, size and location of the table. (c): Model has been generated in Unity engine from the data uploaded by the user. Grid squares are 1m x 1m, with 10cm subsections.*

AR was to ensure it is widely accessible by making use of readily available consumer technology. Apple's ARKit 1.0 (released September 2017) was selected for this project due to the wide availability of 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 [Bol17]. There is also potential to port the application to Android devices. Our application is therefore accessible to many users without the need for the purchase or learning of specialist equipment.

LevelEd AR presents users with an AR view of the real-world location they are modelling. Users first confirm the floor location by pointing the device down, scanning the ground and then selecting the surface by tapping the screen. The user moves around the real-world location whilst aiming a yellow marker in the centre of their smartphone screen at key locations (such as room corners or around objects). By tapping the screen, the user can place an AR marker into the scene. This marker stays in place as the user continues to model the rest of the space. Users have two modes available to them:

Modelling walls: users place markers at intersections of walls within a room to map out the base of the walls. They can then raise up a second set of markers for the height of the walls (see Fig. 1a).

Modelling 3D objects: users place markers to surround the object. This can be an object of any number of sides but in the experiment, it was set to four. Once the base markers are in place, the user can raise up a second set of markers for the object height (see Fig. 1b).

The above modes enable users to model a space effectively, including not only the walls but also the obstacles and furniture (general shape for more complex furniture) within the space. The shape, location and scale of these are all captured by LevelEd AR. Users are also free to choose what will be included in the model. This avoids the modelling of temporary objects found in a space (such as trash or people), a problem which automated indoor modelling systems can suffer from.

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. 1c) 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. 5c).

4. Experiment Methodology

In this section we describe the experiment that was conducted to evaluate the LevelEd AR system. The experiment was primarily developed to evaluate the accuracy of LevelEd AR for indoor modelling of real world locations. However, the underlying technology, ARKit, and its ability to accurately track the movement of a user and maintain the position of objects in a space over time was also evaluated as a consequence. The experiment was also designed to explore the usability of AR as a means for enabling indoor modelling on consumer devices.

4.1 Experiment Tasks

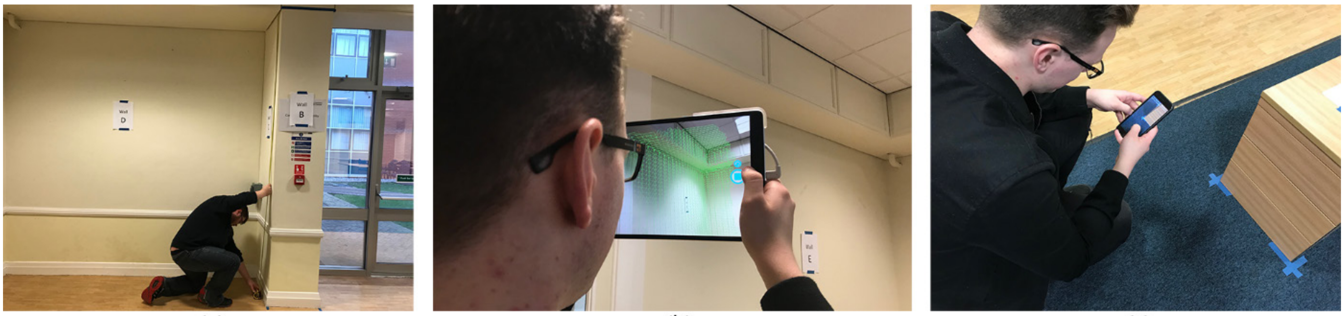
To evaluate the accuracy and usability, participants were asked to complete four separate measuring tasks (see Fig. 2). These tasks were as follows:

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 4 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).



(a) (b) (c)

Figure 2: *(a): Participant completing Task 4 with measuring tape and paper. (b): Participant completing Task 4 with the Room Capture software with the Structure Sensor. (c): Participant completing Task 2 with our LevelEd AR application.*

The above tasks allowed for the evaluation of the accuracy of the different measuring techniques under controlled circumstances. The first three tasks check the accuracy of completing measurements of singular common objects whilst the fourth task was included as a way to replicate a more complex task such as modelling a whole room including furniture. The time taken to capture the measurements for each of the above tasks was also recorded.

4.2 Measuring Instruments

Participants completed the measurement of the four tasks using three different measuring instruments (see Fig. 2). 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. The time taken was recorded from when the participant started until the final measurement was captured on the sheet of paper.

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. The time taken was recorded from when the scan was started until the scan was completed. Once the participant was happy with the scan they were then asked to use the Room Capture built in measurement tool to acquire the requested measurements from the scanned model.

LevelEd AR application: users used an iPhone 7 Plus and the LevelEd AR application to model the tasks in AR. The time was taken from when the users started the application and until they had modeled the task and uploaded the file to the webserver.

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.

4.3 Participants

The experiment was completed by 18 participants, 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.

5. 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. Considering the measurements required, the fourth task incorporated three different values and has been split into Tasks 4.1, 4.2 and 4.3.

The significance was tested by employing a two-way repeated 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.

5.1 Measurements

The results show that the instruments significantly differ from each other in terms of performance overall ($F(2,34) = 73.89$, $p<.001$, $\eta^2p = .813$). The same effect was observed for the tasks in all cases, which suggests the tasks vary significantly in complexity ($F(1,17) = 4533.90$, $p<.001$, $\eta^2p = .996$). 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 ($F(1,17) = 20.14$, $p<.001$, $\eta^2p = .542$), 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.

The first task (Fig. 3a) required an area measurement, and using the Tape as the ground truth, it is shown that using Tape or LevelEd produces no significant difference. Planned contrasts showed that for Task 1, the Tape value was significantly larger than Structure's measurement ($F(1,17)=88.47$, $p<0.001$, $\eta^2p = 0.839$). 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 [KN16].

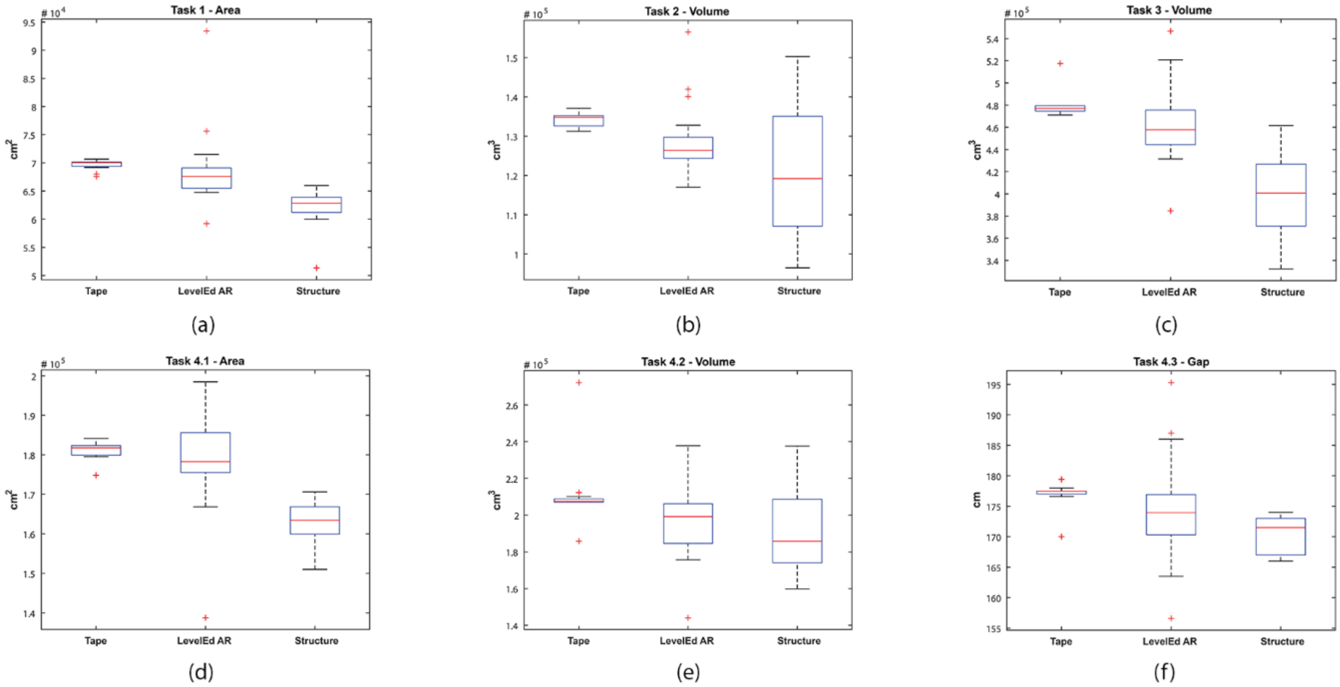


Figure 3: *(a): Task 1 – Area measurement of single wall, (b): Task 2 – Volume measurement of small box, (c): Task 3 – Volume measurement of larger box against a wall, (d): Task 4.1 – Area measurement of 4 continuous walls, (e): Task 4.2 – Volume measurement of small box, (f): Task 4.3 – Gap measurement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

Unlike the first task, the second task showed no difference between LevelEd and Structure, however, both were significantly separated by the ground truth. Hence, for Task 2 (Fig. 3b), the Tape measurement is significantly different from both LevelEd ($F(1,17)=7.06$, $p=0.017$, $\eta^2p = 0.293$) and Structure's outputs ($F(1,17)=13.94$, $p=0.002$, $\eta^2p = 0.450$). 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.

The next two tasks were of a higher complexity than the previous ones and in both, the Structure sensor showed a significant loss in accuracy in comparison with the other two instruments (Fig. 3c & 3d), hence, the Structure's accuracy was significantly lower than both Tape and LevelEd. The Tape instrument had larger values than Structure's (Task 3 - $F=86.01$, $p<0.001$, $\eta^2p =0.835$; Task 4.1 - $F=212.13$,$p<0.001$, $\eta^2p =0.926$); and LevelEd's values followed the same pattern (Task 3 - $F=36.59$, $p<0.001$, $\eta^2p = 0.683$; Task 4.1 - $F=37.70$, $p<0.001$, $\eta^2p = 0.689$). Although Task 3 is similar to Task 2 due to both measuring the volume of a box, Task 3 featured a much larger box situated next to a wall. This increased the complexity of the task and decisions required from the participant. Task 4 required participants to move the iPad more significantly whilst completing the task with Structure than previous tasks. This often resulted in walls shortening and collapsing in on themselves in the scanned model [KN16], as reported above for Task 1. This was not as pronounced with LevelEd, although there were some over and under recording of distances.

Results recorded for Task 4.2, a volume measurement, showed the same pattern as for Task 2 where the same type of measurement was required (Fig. 3e). Hence, the Tape measurement significantly differs to both LevelEd ($F(1,17)=5.66$, $p=0.029$, $\eta^2p = 0.250$) and Structure ($F(1,17)=4.74$, $p=0.044$, $\eta^2p = 0.218$). 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 (a box similar in shape and size to Task 2), 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, only a simple straightforward gap measurement was required (between the starting wall of Task 4.1 and the box in Task 4.2). As expected the Tape measurement was larger than Structure's ($F=85.63$, $p<0.001$, $\eta^2p =0.834$) with LevelEd being no different than the ground truth (see Figure 3f). 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 [FC18][DE18].

5.2 Time

Similarly, for the time (measured in seconds), significant differences were observed throughout the test between the choices of the instrument ($F(2,34) = 116.99$, $p<.001$, $\eta^2p = .873$). 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 ($F(1,17) = 249.58$, $p<.001$, $\eta^2p = .936$). We observed that each instrument performs significantly different with each completed task ($F(1,17) = 119.36$, $p<.001$, $\eta^2p = .875$).

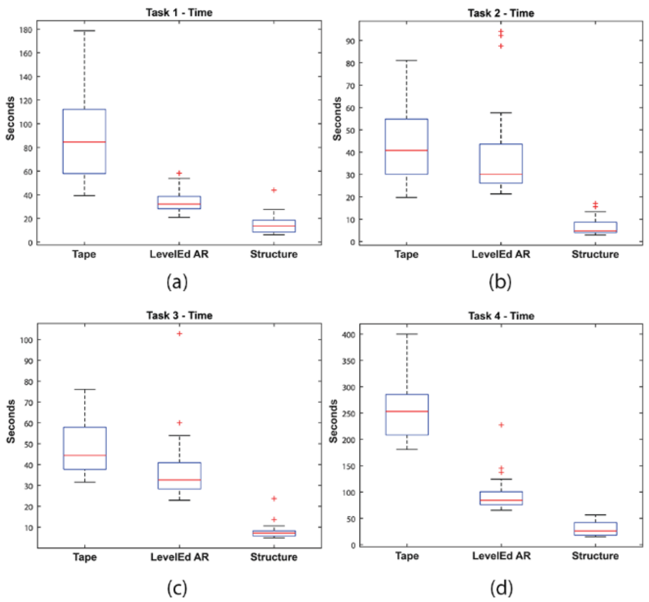


Figure 4: *(a): Task 1 – Time to measure single wall, (b): Task 2 – Time to measure small box, (c): Task 3 – Time to measure larger box against a wall, (d): Task 4 – Time to measure 4 continuous walls, small box, and gap between box and starting wall.*

Following the same pattern, we observed that for the first task (Fig. 4a) 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 ($F(1,17)=46.75$, $p<0.001$, $\eta^2p = 0.733$) and Structure ($F(1,17)=80.23$, $p<0.001$, $\eta^2p = 0.825$). Moreover the time using LevelEd was subsequently higher than using the Structure ($F(1,17)=40.44$, $p<0.001$, $\eta^2p = 0.704$). The results of Task 2 (Fig. 4b) showed that the time spent using Structure was significantly lower than both other instruments (vs. Tape - $F(1,17)=84.41$, $p<0.001$, $\eta^2p = 0.832$); vs. LevelEd - $F(1,17)=31.15$, $p<0.001$, $\eta^2p = 0.647$)). The same pattern was observed for Task 3 (Fig. 5c), where using the Structure proved to be the most efficient time-wise (vs. Tape - $F(1,17)=93.36$, $p<0.001$, $\eta^2p = 0.846$); vs. LevelEd – $F(1,17)=21.18$, $p<0.001$, $\eta^2p = 0.587$). 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. 4d). Planned contrasts showed that for Task 4, the time participants spent using the Tape instrument was significantly larger than using LevelEd ($F(1,17)=103.81$, $p<0.001$, $\eta2p = 0.859$) and Structure ($F(1,17)=286.01$, $p<0.001$, $\eta2p = 0.944$), Moreover the time using Structure was overall lower than all the others (vs. LevelEd - $F(1,17)=78.78$, $p<0.001$, $\eta2p = 0.823$). As the tasks increased in complexity, the gap between Tape and the other two instruments appears to have increased.

6. Virtual Environment Workflow Preliminary Study

Whilst the experiment evaluated LevelEd AR for accuracy, time efficiency and usability of the application; we also carried out a preliminary case study to understand how LevelEd AR can be used effectively as part of a workflow to create detailed virtual environments. We worked with three participants who developed two virtual environments which were both based on a real world location. All three participants had prior skills in 3D modelling. A real world location was chosen that was more complex than the tasks used in the experiment in order to effectively test the workflow. The real world location was captured by the authors using the LevelEd AR application and took a total of 3 minutes and 41 seconds to capture and upload to the webserver. The indoor model captured was downloaded in Unity and saved to a scene. Participants were provided with the Unity project with the indoor model ready in the scene (see Fig. 5b) along with one photograph of the original real world location (see Fig. 5a). Participants were not given access to the location.

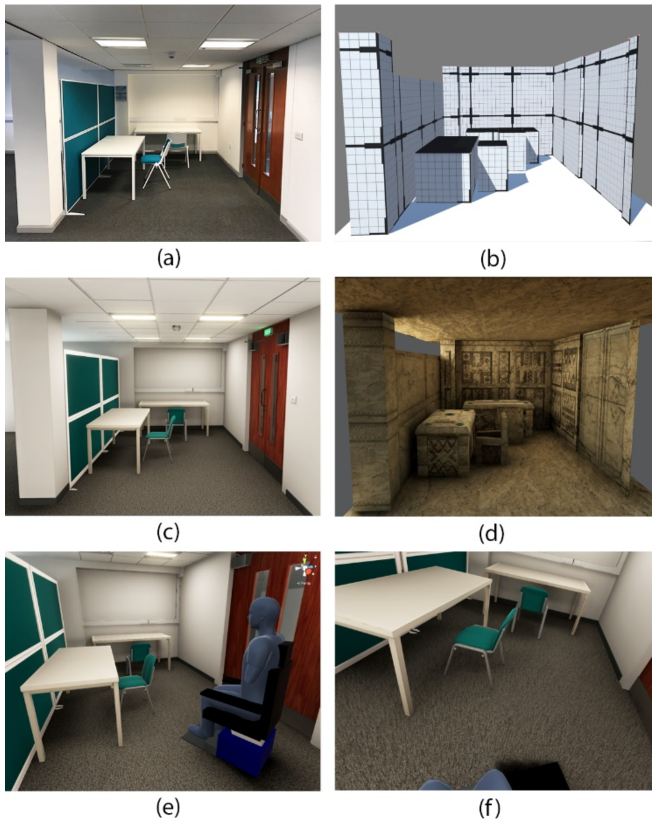


Figure 5: *(a): Photo of the real world location. (b): Model of the location captured by LevelEd AR and downloaded to Unity. (c): Realistic recreation of the location as a serious games virtual environment. (d): Egyptian themed virtual environment created for potential substitutional reality [Sim15]. (e): Realistic scene integrated into a powered-wheelchair VR training simulator. (f): View from VR headset when in the realistic scene in the powered-wheelchair VR training simulator.*

One participant was asked to recreate the space virtually as close as possible to the original scene so it could be used as part of a serious game virtual environment (see Fig. 5c). The other two participants were asked to create an Egyptian tomb themed virtual environment that would fit within the confines of the space and map to the objects included in the area (even if this will break the context somewhat). This virtual environment could potentially be used as part of a substitutional reality [Sim15] experience (see Fig 5d). Participants spent between 1-3 days creating their highly detailed and game ready environments from scratch. In the case of the realistic environment, subsequent environments could be created much quicker since the models can be reused.

To complete the preliminary workflow study, the serious game scene (Fig. 5c) was integrated into a powered wheelchair [JPD*18] VR serious games application (Fig. 5e and Fig. 5f). The scale of the scene was realistic and correct with the scene requiring no additional work as it was already built to scale to match a real-world location. This scene could now be reused and integrated into any serious games application that utilises real world scales. We intend to explore this integration further in future work, paying attention to the workflow and the potential for improved presence and learning from using a real-world location.

Although no formal evaluation of the resulting virtual environments was carried out, participants demonstrated the potential for LevelEd AR to be used to capture spaces and to provide a guide from which to develop a virtual environment that is sufficiently spatially accurate for serious games. Participants also completed a questionnaire and provided valuable feedback on the workflow. Participants noted that having a captured model to work from in Unity meant they did not need to estimate distances or sizes from the image provided. Instead the image was used only to gauge colour and texture. Participants suggested the ability to export objects from the captured model that could be imported into 3DS Max and in-scene object measurements would have sped up the workflow. This is something that could be added in future iterations of the workflow.

7. Discussion

The results of the experiment show that LevelEd AR application measurements are closer to the Tape measure (which we consider to be the ground truth) 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) [Bro96] questionnaire after completing the tasks with each instrument. All three instruments met the SUS usability threshold of 68, with Structure being the least favoured and Tape and LevelEd being closely favoured (see Fig. 6). 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 Tape instrument.

Participants in the 40-59 age group spent a total of between 159 and 477 seconds completing all 4 tasks whereas users in the 18-39 age group spent a total of between 143 and 205 seconds. It was observed during the experiment that this increase in time was due to some of those in the 40-59 age group struggling to control the marker on screen and with the perception of dealing with a 3D object being presented on camera feed on a 2D screen. No significance statistical test was performed on this data due to the low number of participants in this age group, however one explanation can be the users' perception towards AR in particular

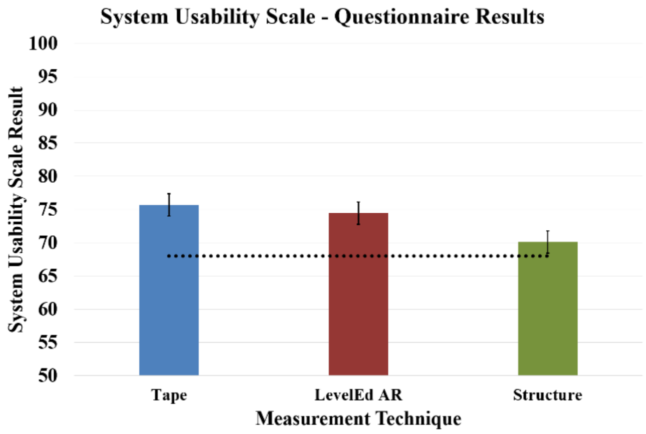


Figure 6: *Results from SUS [Bro96] questionnaire (Mean, standard error (whiskers) and Usability threshold (dotted line)).*

and mobile technology in general. The lack of useful non-games applications (at the time of the test) which used AR might have contributed to this perception. Additional training and tutorials for the application may help alleviate these issues. Our work therefore adds some insights and help towards proper understanding of the phenomenon.

Despite the overall positive accuracy and usability of LevelEd, there were some tasks that either demonstrated minor inaccuracies (Fig. 3b – Task 2) or variations (Fig. 3f – Task 4.3) 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 is a common issue with markerless techniques and in particular with ARKit, as was previously reported by both [FC18] and [DE18]. Our research further confirms this problem.

Common reasons for drift are numerous. It could be attributed to the user moving the device too quickly, causing the camera image to blur and computer vision systems becoming unable to track key points frame to frame. It could be due to some surfaces lacking sufficient points of interest which again causes the computer vision system issues when trying to track key points between frames.

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. 3f – Task 4.3 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. Currently ARKit maintains only a partial map of the scanned area and so this makes recalibration difficult over longer more complex tasks. However, ARKit changes are not within our remit and instead improvements could be addressed through the user experience of the application. We suggest the system could warn users when they are moving too fast or when drift has potentially happened, prompting users to try again. This is an area that would benefit from further research including tasks requiring longer and more complex measurements, more akin to mapping a full room to see the full effect of drift and potential solutions in action.

Another reason that could explain drift is the participants' fatigue. Users seem to reach exhaustion faster when moving around the room with the attention and focus oriented towards making measurements through the AR device. Others have noted the "gorilla arm" effect of holding a device out for long periods of time affecting the use of AR [GLZ*17]. This is easily observable in longer tasks as a comparison with the time spent by participants using the Structure app where their immersion and the physical requirement is limited. As reaching physical and mental exhaustion is a process that builds up over time, together with the technical aspects of AR, this can be a reasonable explanation for an increase in drift over time as users rush to finish the task. Fatigue and exhaustion in AR are some of the topics that will need further research.

8. 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 and entertainment games that could also support passive haptics [Ins01] and substitutional reality [Sim15][EKL18]. The system was developed with consumer technology in the form of a smartphone and Apple's ARKit framework. Users were presented with an AR view of the world whereby they could place markers for walls or objects to capture their location, size and then to define the height. The data captured was then available to download within the Unity engine to generate an indoor model suitable for use directly in a serious game, to be used as a guide whilst developing a virtual environment or converted to collision meshes once more complex assets have been added.

At the end of section 2 we outlined some key features that the system must address. These have all been met as the system produces a simplified mesh suitable for direct use in a serious game virtual environment, it allows users to model both walls and objects within the space including irregular rooms, it gives the users a choice as to what should be modelled and finally, it is accessible to users due to the utilisation of ARKit and the support for a large number of pre-existing smartphone devices.

Whilst the prototype has been successful at meeting the above aims, there are still minor issues that need to be addressed and researched further. The problem of drift and its potential causes need further research along with the part fatigue and exhaustion play in the use of AR systems. Further exploration of these issues would help to develop a better understanding of the problems and inform potential solutions that researchers and developers can utilise to improve the accuracy and usability of future systems.

Finally, we also intend to explore and validate the workflow demonstrated in the paper further by integrating virtual environments built using captured data from LevelEd AR into serious games applications and potentially those that support passive haptics and substitutional reality. This is part of a larger project investigating different level design workflows using AR and VR.

9. Acknowledgements

We would like to thank Dr. Marian Andrei Stanciu for his assistance with the statistics in this paper. We would also like to thank William Sowden, Chloe Harrison and Jack Partridge for their participation in the preliminary workflow case study.

References

[ARD*18] ABDELAAL M., REICHELT D., DÜRR F., ROTHERMEL K., RUNCEANU L., BECKER S., FRITSCH D.: ComNSense: Grammar-Driven Crowd-Sourcing of Point Clouds for Automatic Indoor Mapping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (March 2018), 1–26. doi: 10.1145/3191733 2

[AGC18] ANGLADON V., GASPARIN I S., CHARVILLAT V.: Room Floor Plan Generation on a Project Tango Device. International Conference on Multimedia Modeling (February. 2018), 226-238. doi:10.1007/978-3-319-73600-6_20. 2

[Ara07] ARAYICI Y.: An approach for real world data modelling with the 3D terrestrial laser scanner for built environment. *Automation in Construction 16*, 6 (September. 2007), 816–829. doi:10.1016/j.autcon.2007.02.008. 2

[Bol17] BOLAND M.: ARtillry Intelligence: 380 Million iPhones Are Compatible with ARKit. (July. 2017), http://artillry.co/2017/07/26/artillry-intelligence-arkits-installed-base-is-380-million-iphones/ (accessed June 26, 2018). 3

[Bro96] BROOKE J.: SUS - A quick and dirty usability scale. Usability evaluation in industry 189, 194 (1996), 4–7. URL: https://hell.meiert.org/core/pdf/sus.pdf (accessed June 26, 2018). 7

[CR09] CHITTARO L., RANON R.: Serious Games for Training Occupants of a Building in Personal Fire Safety Skills. *Conference in Games and Virtual Worlds for Serious Applications,* (March. 2009), 76–83. doi:10.1109/vs-games.2009.8. 1

[CSK*15] CHENXI L., SCHWING A G., KUNDU K., URTASUN R., FIDLER S.: Rent3D: Floor-plan priors for monocular layout estimation. *EEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June. 2015), 3413-3421. doi:10.1109/cvpr.2015.7298963. 2

[DE18] DILEK U., EROL M.: Detecting position using ARKit II: generating position-time graphs in real-time and further information on limitations of ARKit. *Physics Education* 53, 3 (May 2018), 1–6. doi:10.1088/1361-6552/aaadd9. 2, 5, 7

[EKL18] ECKSTEIN, B., KRAPP, E. & LUGRIM, B., Towards Serious Games and Applications in Smart Substitutional Reality. *10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, (2018). doi: 10.1109/vs-games.2018.8493444. 1, 8

[FC18] FUSCO G., COUGHLAN M. J.: Indoor Localization using Computer Vision and Visual - Inertial Odometry. *International Conference on Computers Helping People with Special Needs (2018).* 2, 5, 7

[GLZ*17] GRUBERT J., LANGLOTZ T., ZOLLMAN S., REGENBRECHT H.: Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization* *and Computer Graphics* 23, 6 (June 2017), 1706–1724. doi:10.1109/tvcg.2016.2543720. 7

[HHF10] HEDAU V., HOIEM D., FORSYTH D.: Thinking Inside the Box: Using Appearance Models and Context Based on Room Geometry. *European Conference on Computer Vision* (September. 2010), 224-237. doi:10.1007/978-3-642-15567-3_17. 2

[HKP90] HOPPEN P., KNEIERIEMEN T., VON PUTTKAMER E.: Laser-Radar based Mapping and Navigation for an Autonomous Mobile Robot. *IEEE International Conference on Robotics and Automation,* (May. 1990), 948–953. doi:10.1109/robot.1990.126113. 2

[Ins01] INSKO B.: Passive Haptics Significantly Enhances Virtual Environments, PhD, University of North Carolina at Chapel Hill, 2001.

[JPD*18] JOHN W. J., POP R. S., DAY W. T., RITSOS D. P., HEADLEAND J. C.: The Implementation and Validation of a Virtual Environment for Training Powered Wheelchair Manoeuvres. *IEEE Transactions on Visualization and Computer Graphics 24*, 5 (March. 2018), 1867–1878. doi: 10.1109/tvcg.2017.2700273. 1, 6

[KN16] KALANTARI M., NECHIFOR M.: Accuracy and utility of the Structure Sensor for collecting 3D indoor information. *Geo-Spatial Information Science* (October. 2016), 202-209. doi:10.1080/10095020.2016.1235817. 1, 2, 5

[LGA*17] LOVREGLIO R., GONZALEZ V., AMOR R., SPEARPOINT M., THOMAS J., TROTTER M., SACKS R.: 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,* (July. 2017), 381–389. doi:10.24928/jc3-2017/0058. 1

[MK94] MILGRAM P., KISHINO F.: A taxonomy of mixed reality visual displays, *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321-1329. 1

[MMB*15] MONSZPART A., MELLADO N., BROSTOW G. J., MITRA N. J.: RAPTER: Rebuilding Man-made Scenes with Regular Arrangements of Planes. *ACM Transactions on Graphics 34*, 4 (July. 2015), 103:1–103:12. doi:10.1145/2766995. 2

[NDI*11] NEWCOMBE R. A., DAVISON A. J., IZADI S., KOHLI P., HILLIGES O., SHOTTON J., MOLYNEAUX D., HODGES S., KIM D., FITZGIBBON A.: KinectFusion: Real-Time Dense Surface Mapping and Tracking. *IEEE International Symposium on Mixed and Augmented Reality* (2011), 127-136. doi: 10.1109/ISMAR.2011.6092378. 2

[OVW*16] OCHMANN S., VOCK R., WESSEL R., KLEIN R.: Automatic reconstruction of parametric building models from indoor point clouds. *Computers & Graphics 54*, (February. 2016), 94–103. doi:10.1016/j.cag.2015.07.008. 2

[SS05] SANCHEZ-VIVES M. V., SLATER M.: From presence to consciousness through virtual reality, *Nature Reviews Neuroscience* 6, 4 (April 2015). 332–339. doi:10.1038/nrn1651 1

[Sen19] Sensopia, magicplan. [Application]. Montreal, Canada: Sensopia Inc., 2019. 2

[Sim15] SIMEONE, A. L., Substitutional reality: Towards a research agenda. *IEEE 1st Workshop on Everyday Virtual Reality (WEVR)*, (2015) 19-22. doi: 10.1109/wevr.2015.7151690. 1, 6, 8

[TZ14] TURNER E., ZAKHOR A.: 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 - Volume 1: GRAPP, (VISIGRAPP 2014).* 22–33. doi: 10.5220/0004680300220033. 2

[VSS14] VOLK R., STENGEL J., SCHULTMANN F.: Building Information Modeling (BIM) for existing buildings - Literature review and future needs. *Automation in Construction 38*, (March. 2014), 109–127. doi:10.1016/j.autcon.2013.10.023. 1, 2

[ZCS*18] ZOU C., COLBURN A., SHAN Q., HOIEM D.: LayoutNet: Reconstructing the 3D Room Layout from a Single RGB Image. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (March. 2018), 2051-2059. 2