EVALUATION OF MICROSOFT HOLOLENS AUGMENTED REALITY TECHNOLOGY AS A CONSTRUCTION CHECKING TOOL

Loporcaro G., Bellamy L., McKenzie, P. & Riley, H.
University of Canterbury, Christchurch, New Zealand

Ralston, S.
Trimble Inc., Christchurch, New Zealand

ABSTRACT: *Increasing productivity is one of the most important objectives of the construction industry. Building Information Modelling (BIM) was introduced to facilitate collaboration and coordination in the life-cycle of buildings and infrastructures. More often, BIM models are preferred to traditional 2-dimensional drawings to communicate the design to the project stakeholders such as design team, contractors, and the client. Conversely, it is common to use traditional 2-D drawings on-site in the construction phase. Construction mistakes are often the results from errors in the construction drawings or from their misinterpretations. Over the last years, augmented and virtual reality has been introduced in the architecture, engineering and construction industry (AEC) as visualization and collaboration tools. This paper aims to evaluate Augmented Reality (AR), in particular, Microsoft HoloLens, as a construction checking tool. The paper compares different construction checking tools such as the traditional tape and measure, Trimble SX10 laser scanner and the Microsoft HoloLens using Trimble Connect as BIM software, to the Trimble SX10 Total Station, a current industry leader in surveying technology. Comparisons were based on accuracy, speed of testing, ease of use and efficiency. Testing was undertaken by comparing benchmark virtual models of the testing rooms, obtained from an initial survey, and the actual physical rooms. Results indicated that, overall, the HoloLens was the most efficient construction checking tool within a 15 mm to 50 mm accuracy range. It was not as accurate as a laser scanner or tape measure, which indicates further improvements are required before this technology can be recommended as a general construction checking tool.*

KEYWORDS: Augmented reality, Construction checking tools, Digital technologies, HoloLens

1. Introduction

In Augmented Reality (AR), 3-dimensional virtual objects are overlaid upon the physical world (Azuma, 1997). Although AR was initially developed in the 1990s (Milgram, Takemura, Utsumi, & Kishino, 1995), this technology has seen rapid market growth over the last decade, particularly through the popularisation of AR technology with handheld devices and smartphone usage. As this technology becomes increasingly common in many industries (medicine, education, marketing, gaming, etc) (Billinghurst, Clark, & Lee, 2014), its functionality could be extended to be used as a construction checking tool within the Architecture, Engineering and Construction (AEC) industry. The technology's potential could allow an accelerated construction checking process, enabling the user to recognize discrepancies within the built design, faster and more efficiently than current construction checking methods.

Microsoft HoloLens, shown in Figure 1, is a visual headwear device that uses a beta version of Trimble Connect software, a collaboration platform developed for the construction industry. HoloLens is equipped with a holographic display and Trimble Connect software which imports virtual models (e.g. SketchUp format) from cloud storage and overlays the model through the HoloLens headset onto the physical world. The result immerses the wearer into a blend of realities. The user can interact with the virtual model with intuitive hand gestures. Such technology is currently used for visualization purposes within the architectural industry, but more technical applications of this technology within the Civil Engineering discipline would see its commercial value rise (Blanco, Mullin, Pandya, & Sridhar, 2017).

Figure 1 The Microsoft HoloLens device

1.1 A brief summary of AR as construction checking tools

Augmented Reality as construction monitoring tool has been investigated by a number of researchers over the recent years (Kwon, Park, & Lim, 2014; Meža, Turk, & Dolenc, 2015; Park, Lee, Kwon, & Wang, 2013; Yabuki & Li, 2007; Yeh, Tsai, & Kang, 2012). Yabuki and Li (2007) developed a prototype AR technology for checking steel reinforcing bars in bridge piers. Checking reinforcement layout based on 2D drawings is impractical, therefore such a system could reduce inspection time and costs. Park et al. (2013) created a framework to detect, classify and communicate defects on construction, while (Kwon et al., 2014) developed a management system to detect defects in reinforced concrete structures using BIM, image-matching and augmented reality. Meža et al. (2015) identified that the challenges associated with the use of AR in construction include: the virtual model not aligning well with the surrounding area; the small size and resolution of the hardware; and difficulties arising from obstructions within the field of view, such as a high constructions or fences, and healthy and safety risks associated with the use of the device within a environments . Another concern, made by Meža et al. (2015), was the potential safety implications of using AR as a wearable mobile device on a live construction site and how this may distract the user from hazards in the surrounding environment. Meža et al. (2015) concluded that AR can facilitate the understanding of project documentation, especially in the visualisation of 3D models within the field. However, AR remains technologically constrained by barriers such as indoor GPS, visual occlusion, frame-rates of virtual elements and general responsiveness. Previous research employed prototype AR technologies often no commercially available. On the other hand, little research has been conducted to check the feasibility of commercially available AR devices such as Microsoft HoloLens.

1.2 Research aim

The aim of this research is to investigate the feasibility of Microsoft HoloLens as a construction checking tool. In particular, Microsoft HoloLens is compared as a length measuring tool to the traditional tape and measure tool and Trimble SX Laser Scanner. Comparisons are made on the following performance indicators: accuracy, reliability, speed of testing, ease of use, and efficiency.

2. Methodology

2.1 Performance indicator

2.1.1 Accuracy

In order to compare the accuracy of each measuring tool (Taper and measure, HoloLens and the Trimble SX10 Laser scanner), wall lengths and heights within a room are measured with each tool and compared to the corresponding lengths and heights of a benchmark model of the room. The benchmark model, considered the most accurate model of the room, was created from measurements obtained using a Trimble SX10 Total Station (https://geospatial.trimble.com/products-and-solutions/trimble-sx10). Any discrepancies between measurements with each construction checking tool and benchmark measurements will yield an error. Thus the accuracy of each tool can be quantified based upon the size of these errors.

Additionally, spatial errors are evaluated via a coordinate system. Coordinate deviations between the benchmark corner positions and virtual model corners yield an error. Through these comparisons of accuracy, the reliability of these results was also ascertained.

2.1.2 Speed of testing

The speed of construction checking was timed for each different method, with times being averaged to give an average checking time for each method. This average measure mitigates the associated ‘learning curve’; an inherent part of any familiarisation process.

2.1.3 Ease of use

The complexity of a construction checking tool can be a barrier to its use in the construction industry. A subjective ranking of the complexity of the four construction checking techniques was given with a brief explanation as to its position.

2.1.4 Efficiency

Often quick construction checks are needed with results accurate to a desired level of accuracy. The efficiency of each tool was evaluated within target accuracy ranges (< 5 mm, 5 – 15 mm, 15 – 50 mm) based on the average time of testing.

2.2 Test procedure

2.2.1 Benchmark survey

Benchmark surveys were undertaken of five rooms selected on the University of Canterbury campus. These surveys were undertaken with the Trimble SX10 Total Station using its Direct Reflex (DR) capabilities. The survey was conducted by an experienced operator to ensure the reliability of the dataset. The Total Station was placed at the centre of each room in order to capture the coordinates of all the room corners and edges. Rooms with large architectural/structural features, such as curved and angled walls or large columns were discarded, as this would have increased the testing difficulty and might have affected the error analysis. Plain rectangular rooms were the ideal candidates.

During the benchmark survey, the Total Station (Figure 2) was set at a height where all, or most, of the vertices of the room were easily captured. A backsight was arbitrarily set and thus the x-axis and y-axis were arbitrarily set within the room. The z-axis was aligned to be perpendicular to the floor plane. The Total Station’s position was ‘zeroed’, giving its position within the room as (0,0,0) in x, y and z coordinates. Systematically, the corners of each room were mapped using DR. Each point mapped was noted with a number on an accompanying sketch for reference, as results would be issued in a CSV file. The CSV file was exported to SketchUp as a series of points in space, enabling a virtual 3D model to be created. Figure 3 shows a model created from the Total Station DR points.

Figure 2 Benchmark survey using the Trimble SX10

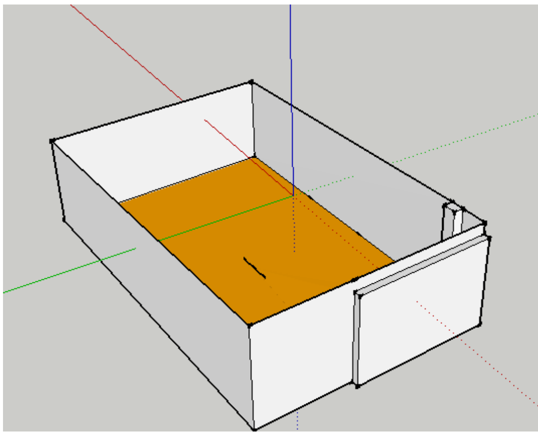

Figure 3 Benchmark virtual model in SketchUp

Within the virtual benchmark model, lines from diagonal corners were drawn, creating a large "X" across the floorplan of the model. At the intersection of these lines, a one-metre line was drawn back towards the mid-span of one of the walls. This line served as a reference line. The same line was recreated in the physical room whereupon virtual models from the HoloLens tests would be anchored. The purpose of the benchmark models was to enable a comparison for lengths within the model to the corresponding lengths found from the construction checking tools methods.

2.2.2 Microsoft HoloLens testing

A blind-testing method was developed. Three virtual models of each room were created in SketchUp by reducing the dimensions of the benchmark models. One person created the reduced models, while the other engaged in the HoloLens testing.

The benchmark CSV points were taken and reduced by the same arbitrary value in the x and y direction and a different arbitrary value in the z-direction. These points now represented a reduced version of the benchmark model. Three different sizes reduced models of the same room were created, with reductions in the x and y directions being anywhere between 0.1 metres to 1.5 metres. Height (z) reductions were typically no greater than 0.5 metres. Figure 4 shows the original benchmark model in white and a reduced model in yellow. The virtual model was uploaded on the cloud and the retrieved on HoloLens using the Trimble Connect HoloLens (TCH) application. Initially, the room had to be scanned by the HoloLens. The TCH application formed a mesh lining of the room (Figure 5). This process took between five and ten minutes, depending on the size of the room being scanned.

After the mesh had been fully formed, one of the three reduced models for that particular room was loaded by the TCH application and appropriately scaled to a 1:1 size. The reduced model was orientated and anchored so that the reference line on the model was aligned upon the physical reference line in the room.

The test was conducted by taking measurements at each corner perpendicular from the physical room to the reduced model as viewed by the HoloLens user. Measurements were taken using a physical tape measure, with the user standing directly above the tape to minimise parallax error and with the built-in, real-to-virtual measurement tool on the TCH application. Both sets of measurements were recorded. These measurement methods can be seen in

Figure 6 and Figure 7, respectively.

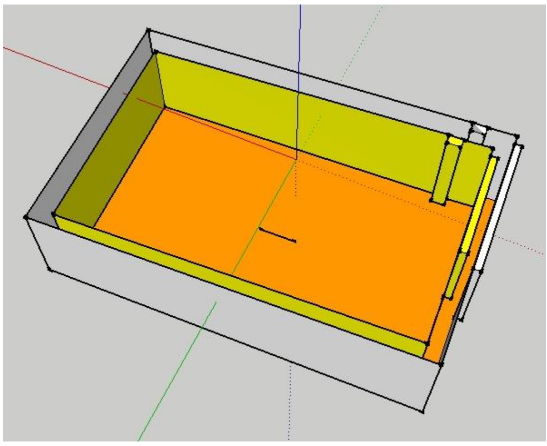

Figure 4 Reduced models fit inside the benchmark model

Figure 5 Trimble Connect HoloLens creates a mesh of the room

Figure 6 Physical HoloLens measurements

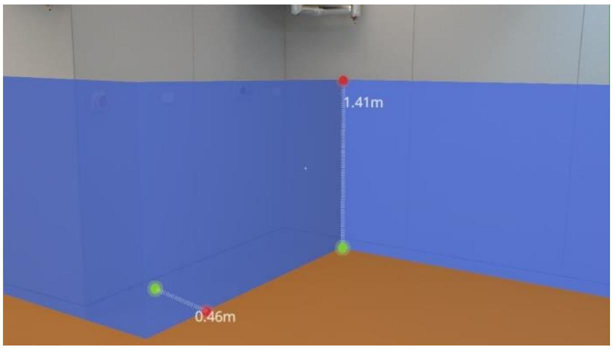

Figure 7 Real-to-virtual HoloLens measurements

Heights from the physical room floor to the top of the reduced model's walls were taken with both the physical and the virtual tape. Several heights were taken across the room, the number of which varied with the number of different roof/wall heights present.

Unlike the other testing methods, the lengths from the HoloLens testing had to be "assembled". Only the four external lengths and heights within each test could be used for comparison. An example of how one of these lengths was assembled is seen in Figure 8. Note that the lengths 'a_1' and 'a_2' are the perpendicular length measured from the physical wall to the virtual model.

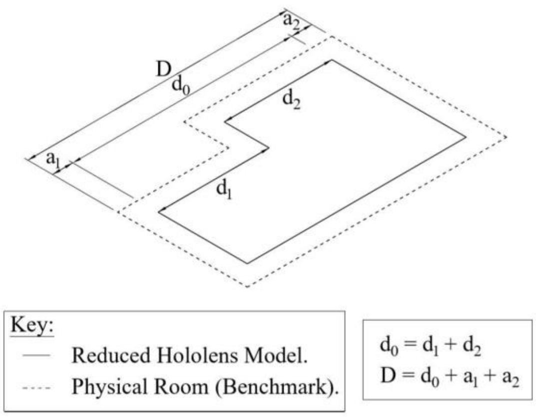

Figure 8 Assembly of HoloLens measured lengths

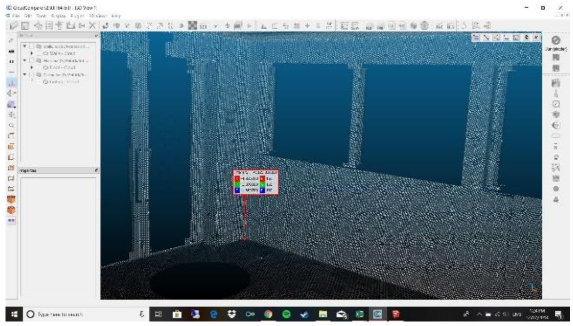

Figure 9 CloudCompare vertices selection

2.2.3 Trimble SX10 Laser Scan testing

A laser scan using Trimble SX10 (the same device used for the benchmark test) was undertaken. The laser scanner was set up to capture a 360° view of the room in a high-density 3D scan. The laser scanner captures thousands of measurement points within a room as a point cloud. The laser scan data files were opened using CloudCompare, with vertices being manually selected and recorded into a CSV file. Figure 9 shows how data points were selected from CloudCompare. Similarly to the benchmark models, the CSV file was exported to SketchUp and a virtual model of the 3-D scan was created. Each length within the model was given a number which corresponded to the same number and thus, the same length in the benchmark model.

2.2.4 Tape and measure

Floorplan lengths within a room were measured with a 30-metre steel tape by two people. All measurements were taken to the nearest five millimetres. Two vertical heights were taken, generally, on diagonally opposite corners of the room. All floorplan lengths and taken heights were modelled on SketchUp with the assumption that all vertices were exactly perpendicular (i.e. each wall was 90o to the adjacent wall).

2.3 Results and discussion

2.3.1 Accuracy: length measurements

The benchmark lengths were subtracted from the corresponding lengths found from each construction checking tool methodology to give error values. Errors that were positive showed that the particular construction checking tool used underestimated the benchmark lengths, while negative errors indicated an overestimation of measured lengths. Figure 10 shows this phenomenon. All the errors found across all five rooms for all construction checking techniques were used to create a distribution through statistical bootstrapping of errors (Figure 11). The normalised frequency plotted in Figure 11 accounts for different sizes of datasets for each construction checking tool. Table 1 shows the average errors and standard deviations. Outliers were removed from the datasets before the bootstrapping occurred.

The HoloLens with the physical tape gave the least accurate results with an average error of -87.7 mm. This indicated that the measured lengths were smaller than the benchmark measurements. Using the HoloLens accompanied with the virtual tape gave an average error of 22.1 mm smaller than the benchmark measurements. The error averages for both HoloLens data sets fall at the peak value (the median) of their distributions and shows that the HoloLens tended to overestimate spatial dimensions. However, the error averages for the laser scanner and tape & measure fall to the left of their largest peaks and result in a negative average error. The peak values for the laser scanner and the tape & measure are 3.9 mm and 10.1 mm respectively. It can be said that both the tape & measure and the laser scanner have a tendency to underestimate spatial dimensions.

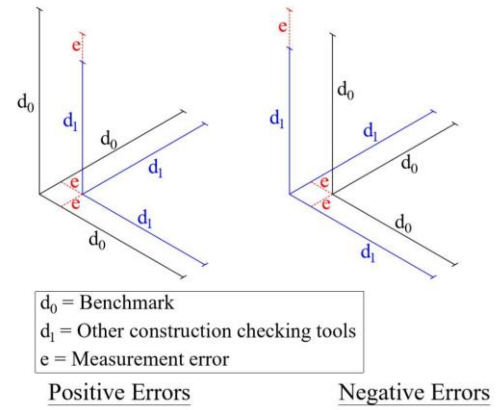

Figure 10 Errors detection using length-based comparisons

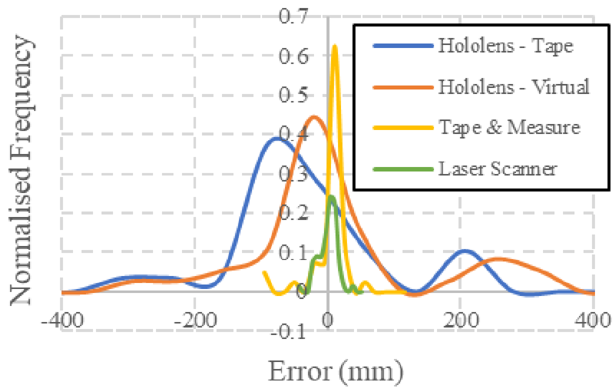

Figure 11 Distribution of errors associated with the lengths of each measurement

Table 1 Average errors and standard deviation

Method	Average error (mm)	Standard deviation (mm)
HoloLens (Tape)	-87.7	146.8
HoloLens (Built-in virtual tape)	-22.1	136.0
Tape & Measure	-5.0	30.2
Laser Scanner	-2.9	13.5

Theoretically, both sets of HoloLens results should show the same average error and display the same standard deviation. However, this is not true as the differences between these results can be attributed to a number of different factors. One of the main factors is the positioning of the mesh with regards to the wall. Here, the mesh was often not aligned upon the wall face exactly, being up to 50 mm either side of it. Real-to-virtual measurements are more accurately, mesh-to-virtual measurements as the HoloLens picks up the walls through the placement of the mesh and tries adopt them into the virtual space. As the mesh does not correctly line the wall, a systematic error was introduced into each virtual measurement taken. This error is one of the causes for the difference between the HoloLens two methods of measurements. The physical measurements taken with the HoloLens were taken from the physical wall to the virtual model. Errors could occur as a result of incorrect viewing of the tape (i.e. parallax error). Additionally, the HoloLens has a minimal scaling error associated with it. This is due to both HoloLens error results being significantly negative. This scaling error is more prominent in the HoloLens tape results than the HoloLens virtual measurements. While both sets of measurements are similar, the HoloLens tape measurements harboured a scaling error based on the perception of the HoloLens user. Predictably, the Laser Scanner and Tape & Measure results showed the greatest levels of accuracy. Ultimately, all results proved reliable as the distinctive peaks of the error distributions in Figure 11 would indicate.

2.3.2 Spatial comparisons

While lengths are typically the first form of measurement in construction practice, the HoloLens introduced some unique situations where the reduced virtual model had been properly anchored but showed a ‘skewing’ of the model. This set of data analysis focused on the positioning of the model within the physical space.

To examine the spatial accuracy of the HoloLens, the same testing data was used as the length comparisons. However, the data was used in its 3-dimensional form via a coordinate system to compare discrepancies between the virtual model and what is displayed through the HoloLens. The virtual models are not the same as what is viewed through the HoloLens and these discrepancies are ultimately the spatial error of the HoloLens. Using

Pythagoras' Theorem the spatial error can be calculated as shown in Figure 12.

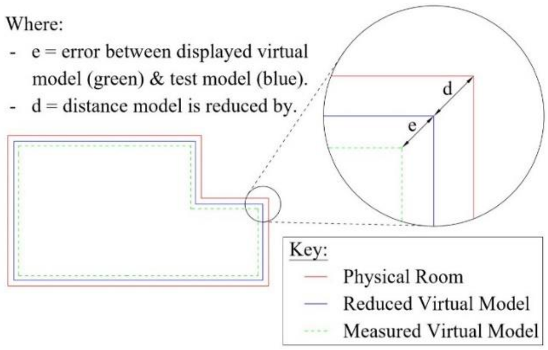

Figure 12 Errors detection using spatial-based comparisons

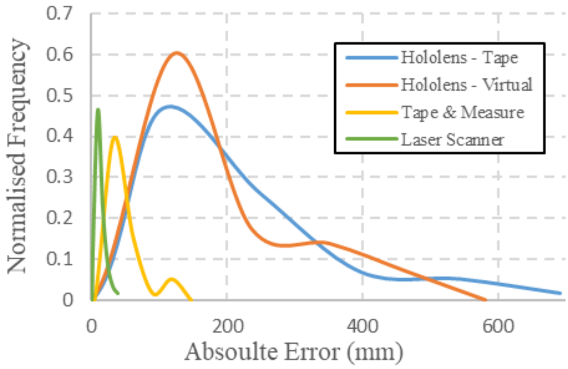

Figure 13 Distribution of errors associated with the spatial coordinate comparison

Note that Figure 12 depicts a perfect scenario whereby each model lines up perfectly within the next. However actual practice showed a skewing and slight rotation error between the physical room and reduced model. With no set coordinate system, or axis aligned, perpendicular or parallel to any of the physical walls, an absolute error using Pythagoras' Theorem seemed the most appropriate way to measure these errors. Laser scan and tape & measure errors were simply the same absolute distance between the benchmark and their respective scan and tape & measure models. Figure 13 shows all errors found across all five rooms for all construction checking techniques. Error distributions were created for each construction checking tool through a statistical bootstrapping protocol.

The average error and the standard deviations for these distributions are shown in Table 2. Results show that the most accurate construction checking tool was the laser scanner, while HoloLens, using the virtual measurement tool, gave the most inaccurate result with an average error of 121.4 mm. These large errors could be attributed to different factors such as scaling of the virtual model, inaccurate anchoring or measurements, or skewing of the model. As these errors are taken as absolute errors, the individual contribution of each of these factors cannot be ascertained. Instead, the HoloLens system's error is just evaluated as a whole.

Through further data analysis of these errors, it was found that the percentage error; the quotient of the error, e, over the distance the model was reduced by, d (see Figure 12), decreased rapidly with increasing distance. Figure 14 illustrates this phenomenon. This shows that smaller distances that were measured had larger percentage errors compared to the larger distances that were measured. This implies that as the measurements became larger, the size of the error that occurs is relatively consistent.

Table 2 Average errors and standard deviation

Method	**Average error (mm)**	**Standard deviation (mm)**
HoloLens (Tape)	102.7	147.2
HoloLens (Built-in virtual tape)	121.4	114.8
Tape & Measure	32.3	28.6
Laser Scanner	9.4	7.4

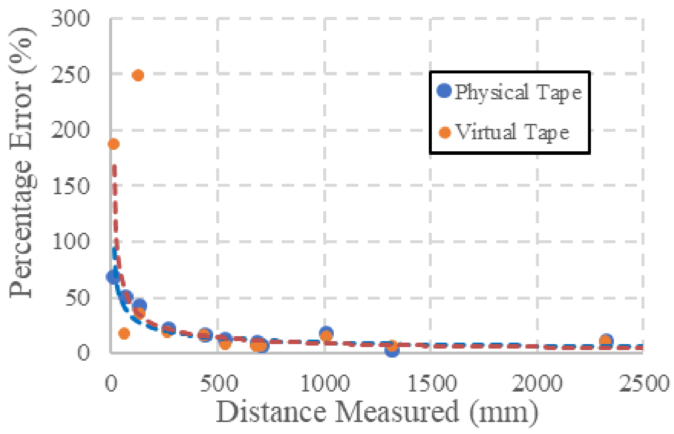

Figure 14 Percentage error with increasing distance

2.3.3 Speed of testing

Each test for every construction checking method was taken from the beginning of the test until the end of testing, including any initial setup time. Testing times progressively became faster as familiarisation with the technology increased. Increasingly faster times emphasises the inherent 'learning curve' associated with dealing with new technology. However, the average time for testing was deemed more appropriate to comment upon as these times have more value for field-testing. Table 3 shows the average time for each test across the five rooms surveyed.

Table 3 Average testing time across all rooms

Method	Average time (min' sec")
Tape & Measure	7'35"
Laser Scanner	12'52"
HoloLens (Tape)	5'54"
HoloLens (Built-in virtual tape)	9'13"
Total station (benchmark)	15'09"

2.3.4 Ease of use

These rankings are based on the subjective opinions of the authors of the tests. Table 4 ranks the ease of use of each construction checking tool. Measuring with a tape and measure device is an intuitive process since it is common practise, therefore it was deemed the easiest. Laser Scanner operation was of moderate difficulty, in part, due to the use of unfamiliar technology. However, once the Trimble SX10 Laser Scanner had been set up and levelled properly, the digital interface for the scan was intuitive and ultimately as simple as selecting a "start scan" button.

HoloLens operation was slow to begin with but became increasingly easier to use once the initial familiarisation with the technology had occurred. The HoloLens's intuitive hand gestures and vocal commands contributed greatly to its ease of use. Physical measurements taken were significantly harder than virtual measurements taken with the TCH application. This was due to the difficulty of seeing the physical tape over the HoloLens virtual display.

Table 4 Ranking of construction checking tolls from easiest to hardest

Method	Rank
Tape & Measure	1 (Easiest)

Laser Scanner	2
HoloLens (Built-in virtual tape)	3
HoloLens (Tape)	4 (Hardest)

2.3.5 Efficiency

The relative efficiency of each construction checking tool can be found by comparing the average time of testing to the average accuracy. Length based measurement errors have been used to evaluate the accuracy of each construction checking tool. Table 5 shows the most efficient construction checking tool for each accuracy range. Note that the times for testing across all five rooms were usually no less than 5.5 minutes and no greater than 16 minutes. This means that given a very large room (i.e. greater than 50 square metres) it would be outside the range of room sizes tested for this project and thus the most efficient construction checking tool for each accuracy range may change. Additionally, the efficiency ranking in Table 5 does not account for other technological capabilities, but for using the tools for surveying alone.

Table 5 Ranking of construction checking tolls from easiest to hardest

Desired accuracy range	Most efficient method
< 5 mm	Laser scanning
5 – 15 mm	Tape & Measure
15 – 50 mm	HoloLens (Built-in virtual tape)

3. Recommendations

This section focuses on addressing issues encountered during HoloLens testing and if possible providing recommendations. As the TCH software used by the HoloLens was still in its Beta version, issues/inconsistencies were always a probability. The following subsections categorise these issues.

One of the test rooms selected had two sections of wall that were matte black. Both, the HoloLens and the SX10 Laser Scanner could not mesh or pick up this surface and thus only partial data was used from this particular room. One major issue that was found during the testing was an apparent "parallax error". This was noticed when testing the HoloLens with the physical tape. With little or no head movements, the virtual model was perceived to have moved relative to the room. This was also apparent whenever the model was being anchored and whenever the anchoring of the reference line was being double-checked from any point that was not directly above the reference line. It was as if the HoloLens adjusted itself to its correct location the closer you moved towards the said location and in part explains the aforementioned scaling issue.

It would be significantly faster and more accurate if the model was able to directly "snap to the ground" or to the mesh upon the ground. Instead, the transformation tool was used as a "fine adjustment" to get the model floor plane lined up with the physical floor. This was hard to do accurately and became relatively time-consuming. This was a potential source of error for all z-axis measurements, as the judgement was used to subjectively place the virtual model on the ground whereas it may have been several millimetres above or below the actual ground level.

During two tests, the entire model misaligned along the horizontal plane and skewed itself upwards on an angle of 30° approximately. This may be attributed to the sloped ceiling of the room, which appeared to be of a similar angle. Collected results for these tests were aborted and testing restarted.

4. Conclusions

Through a series of tests, the accuracy of the HoloLens using the TCH application has been compared to other construction checking tools. The first test was a length based comparison where the HoloLens (Tape) was found to have the highest average error of 87.7mm. The laser scanner was the most accurate construction checking tool with an average error of 2.9mm. These results were reinforced through spatial comparisons where the HoloLens, had the highest average error of 121.4 mm. The laser scanner was the most accurate with an average error of 9.4 mm.

When considering the time that the different tools took to test, the HoloLens was the most time-efficient construction checking tools. However, even with its desirable time efficiency, the HoloLens is not accurate enough to be a viable construction checking tool at its current level of accuracy.

Further developments of HoloLens software and hardware will propel this tool towards becoming a viable construction checking tool.

5. REFERENCES

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355-385. https://doi.org/10.1162/pres.1997.6.4.355

Billinghurst, M., Clark, A., & Lee, G. (2014). A survey of augmented reality. *Foundations and Trends in Human-Computer Interaction, 8*(2-3), 73-272. https://doi.org/10.1561/1100000049

Blanco, J. L., Mullin, A., Pandya, K., & Sridhar, M. (2017). The new age of engineering and construction technology. *Capital Projects & Infrastructure.* Retrieved from https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/the-new-age-of-engineering-and-construction-technology

Kwon, O.-S., Park, C.-S., & Lim, C.-R. (2014, 2014/10/01/). A defect management system for reinforced concrete work utilizing BIM, image-matching and augmented reality. *Automation in Construction, 46*, 74-81. https://doi.org/https://doi.org/10.1016/j.autcon.2014.05.005

Meža, S., Turk, Ž., & Dolenc, M. (2015). Measuring the potential of augmented reality in civil engineering. *Advances in Engineering Software, 90*, 1-10. https://doi.org/10.1016/j.advengsoft.2015.06.005

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). *Augmented reality: a class of displays on the reality-virtuality continuum* (Vol. 2351): SPIE.

Park, C.-S., Lee, D.-Y., Kwon, O.-S., & Wang, X. (2013, 2013/08/01/). A framework for proactive construction defect management using BIM, augmented reality and ontology-based data collection template. *Automation in Construction, 33*, 61-71. https://doi.org/https://doi.org/10.1016/j.autcon.2012.09.010

Yabuki, N., & Li, Z. (2007). *Cooperative Reinforcing Bar Arrangement and Checking by Using Augmented Reality*, Berlin, Heidelberg.

Yeh, K.-C., Tsai, M.-H., & Kang, S.-C. (2012). On-Site Building Information Retrieval by Using Projection-Based Augmented Reality. *Journal of Computing in Civil Engineering, 26*(3), 342-355. https://doi.org/doi:10.1061/(ASCE)CP.1943-5487.0000156