

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Virtanen, Juho-Pekka; Kurkela, Matti; Turppa, Tuomas; Vaaja, Matti; Julin, Arttu; Kukko, Antero; Hyypä, Juha; Ahlavo, Marika; Edén von Numers, Jessica; Haggren, Henrik; Hyypä, Hannu

**Depth camera indoor mapping for 3D virtual radio play**

*Published in:*  
PHOTOGRAMMETRIC RECORD

*DOI:*  
[10.1111/phor.12239](https://doi.org/10.1111/phor.12239)

Published: 01/06/2018

*Document Version*  
Publisher's PDF, also known as Version of record

*Published under the following license:*  
CC BY

*Please cite the original version:*  
Virtanen, J-P., Kurkela, M., Turppa, T., Vaaja, M., Julin, A., Kukko, A., Hyypä, J., Ahlavo, M., Edén von Numers, J., Haggren, H., & Hyypä, H. (2018). Depth camera indoor mapping for 3D virtual radio play. *PHOTOGRAMMETRIC RECORD*, 33(162), 171-195. <https://doi.org/10.1111/phor.12239>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

## DEPTH CAMERA INDOOR MAPPING FOR 3D VIRTUAL RADIO PLAY

JUHO-PEKKA VIRTANEN\* (juho-pekka.virtanen@aalto.fi)

MATTI KURKELA (matti.kurkela@aalto.fi)

*Aalto University School of Engineering, Espoo, Finland*

TUOMAS TURPPA (tuomas.turppa@nls.fi)

*Finnish Geospatial Research Institute FGI, Masala, Finland*

MATTI T. VAAJA (matti.t.vaaja@aalto.fi)

ARTTU JULIN (arttu.julin@aalto.fi)

*Aalto University School of Engineering, Espoo, Finland*

ANTERO KUKKO (antero.kukko@nls.fi)

JUHA HYYPPÄ (juha.coelasr@gmail.com)

*Finnish Geospatial Research Institute FGI, Masala, Finland*

MARIKA AHLAVUO (marika.ahlavuo@aalto.fi)

*Aalto University School of Engineering, Espoo, Finland and  
HUMAK University of Applied Sciences, Helsinki, Finland*

JESSICA EDÉN VON NUMERS (jessica.eden@yle.fi)

*Swedish Yle, Helsinki, Finland*

HENRIK HAGGRÉN (henrik.haggren@aalto.fi)

HANNU HYYPPÄ (hannu.hyyppa@aalto.fi)

*Aalto University School of Engineering, Espoo, Finland*

\*Corresponding author

### *Abstract*

*By combining game engines with indoor mapping, it is possible to create interactive virtual environments that represent the real world. In this paper a virtual version of an audio installation in a historic building is produced, where the user freely explores the building and experiences a set of audio clips, creating a virtual radio play. A depth camera indoor mapping system, Matterport, captured a set of staged interiors. The aim was to evaluate the quality and usability of the indoor models and to demonstrate applying them in 3D application development. The quality*

*of the models was evaluated by comparison with laser scanning, revealing limitations with Matterport: increasing the size of the measured area lowered the level of detail and accuracy of volumetric models. The quality of textures was not affected by this limitation, maintaining the appearance of models. To obtain optimised models for mobile 3D applications, a manual revision process was applied.*

KEYWORDS: depth camera, game engine, game engine optimisation, indoor mapping, virtual environment

## INTRODUCTION

STORYTELLING AND EXPERIENCING NARRATIVES is a really fundamental cultural action, serving social, cognitive, emotional and expressive functions. While narratives are a mainstay in artistic genres, such as literature, theatre and cinema, they have also been explored in the context of interactive virtual systems (Roussou, 2001). Psychologically, the use of virtual environments in simulation relies on the assumption that the virtual environments triggering our senses (Allen and Madden, 1985) are perceived, and responded to, in the same manner as the real world (Arthur et al., 1997). Therefore, increasing the visual resemblance of the virtual environment should also increase the resemblance of the reaction triggered by the virtual experience. Consequently, the use of a virtual 3D environment as a pedagogical platform and interface between human sensory inputs and behavioural outputs may intensify both mental and physical learning processes (McGann, 2015). However, it has also been found that the presence and quality of a narrative plays a significant role in the depth of the virtual experience (Gorini et al., 2011). The immersion in virtual environments appears to be formed as a combination of sensory stimulation and narrative.

Game engines (for example, Trenholme and Smith, 2008) have been extensively used to create interactive virtual environments. Application areas include gaming and entertainment, urban planning (Herwig and Paar, 2002; Greenwood et al., 2009), landscape visualisation (Manyoky et al., 2014), construction (Shiratuddin and Thabet, 2011) and cultural heritage (Anderson et al., 2010). The term *serious gaming* has been used with applications containing game-like elements, such as tasked interaction (Bellotti et al., 2012); one of the most common applications being education (Arhippainen et al., 2011). Virtual 3D environments can mimic the real world, may be completely fictive or mix reality with fiction. Virtual environments used for simulation are typically realistic (for example, Heinrichs et al., 2008), while collaboration and education tasks have also been performed in fictive environments (for example, Arhippainen et al., 2011). In cultural heritage, virtual environments may combine current environments with historic reconstructions (Ogleby, 1999). Alavesa et al. (2014) also combine real and fictive elements in facilitating virtual storytelling.

Digital geometric models are needed for virtual environments. In the built environment, these may either be interiors or outdoor environments (Fritsch and Kada, 2004). In this area of interest, several methods have been developed to convert geographical information system (GIS) data (Virtanen et al., 2015) and building models (Bille et al., 2014) into formats compatible with game engines. This allows creation of a virtual environment for existing data sources. Digital geometric models can also be made specifically for the virtual application. Alatalo et al. (2016) present the creation of a photorealistic city model as a platform for multi-user interaction, made by manual modelling. The problem is that accurate manual modelling of large environments requires lots of resources (for example, Fassi et al., 2011).

An alternative approach is to employ 3D measuring methods and automated reconstruction algorithms (Zhu et al., 2011). Accurate 3D reconstructions help to increase the accessibility of poorly reachable sites by producing a digital counterpart easily distributed over the Internet (Remondino et al., 2010; Schmidt et al., 2011). Indoor 3D mapping can be accomplished using several different techniques, including photogrammetry (Shao et al., 2015), depth camera (Henry et al., 2012) and laser scanning (Lee et al., 2013). While applications of virtual environments have been reported for several disciplines, for example, cultural heritage (Ogleby, 1999; Portales et al., 2009; Anderson et al., 2010), architecture and construction (Herwig and Paar, 2002; Fassi et al., 2011) and medical training (Heinrichs et al., 2008), the 3D models used have typically been produced by combining manual modelling techniques with 3D data (Fassi et al., 2011) or 3D reconstruction has been applied for producing partial models of larger targets (Portales et al., 2009). Fewer papers exist for utilising automated 3D mapping for producing virtual environments from large indoor settings.

The aims of this article are: first, to evaluate the quality and usability of geometric indoor models produced with an indoor mapping system based on a depth camera; and second, to demonstrate a process for applying these models in mobile- and browser-based 3D application development.

### *Creating a Virtual Radio Play*

In this paper, interdisciplinary cooperation is presented between researchers and the staff of the radio theatre of Svenska Yle, the Swedish language unit of Yle, the Finnish Broadcasting Company. Yle's tasks as an organisation include the creation of media productions utilising the possibilities offered by new technology. Consequently, Yle is increasingly carrying over its content and media offer to the Internet. The changes in media platforms used by consumers, especially the growing prevalence of mobile devices, have improved Yle's opportunities for also reaching small specialist groups. The aim of the cooperation was to combine online 3D platforms and indoor mapping technology with radio theatre content. A planned audio installation was used as a prototype case.

The production, entitled "Sounds from a room of souls" (in Swedish "Röster ur själarnas rum"), was staged in a historic mental institution building in Helsinki (Finland). Being one of the first buildings in Europe designed specifically to operate as a mental hospital, the neoclassical edifice was designed in the 1830s by Carl Ludwig Engel, with the main building completed in 1841. The hospital operated in the premises for over 160 years, after which the building has been adopted for other uses. Yle had plans to realise a staged audio journey in the facilities during the *Night of Arts* event in Helsinki in 2016. A set of interiors was to be staged as historical mental hospital facilities and installed with a sound system to play short audio stories describing the history of the building. The scenes of the play were based on documentary material, such as the journals of the Chief Physician, Anders Thiodolf Saalan, from the 19th century, as well as other documents from various eras (Kajander-Maavuori, 2016; Karjalainen, 2016; Rothberg, 2016).

To increase the accessibility and temporal duration of the installation (the event was to be open for only two days), a virtual equivalent was also produced. This consisted of 3D models of the staged interiors with audio clips in a game engine. The main focus was in the design and creation of the audio content, with the 3D environment used to support exploration and experience of the sound clips. This allowed the user to experience the audio stories while virtually exploring the environment, creating a so-called "virtual radio play". For creating a virtual representation of the staged environment, indoor mapping was to be applied.

## ALTERNATIVES FOR INDOOR RECONSTRUCTION

*Image-based Reconstruction of Indoor Environments*

Photogrammetry can be used to reconstruct both indoor and outdoor environments. Manual photogrammetric reconstructions, such as the iWitness approach of Wendt and Fraser (2007), are labour intensive. Due to the amount of manual work involved, they are not feasible for large and complex environments. Recent studies (Furukawa et al., 2009b; Georgantas et al., 2012; Xiao and Furukawa, 2012) show that automatic techniques, such as dense image matching, are suitable provided that the environment supports automatic feature matching, for example, using the scale-invariant feature transform (SIFT; Lowe, 1999). However, these automated methods are sensitive to errors caused by reflective or textureless surfaces (Furukawa et al., 2009a; Jancosek and Pajdla, 2011; Lehtola et al., 2014), even if the camera is pre-calibrated in a laboratory (Fig. 1). In addition, geometric constraints, such as narrow doorways, reduce achievable geometric orientation accuracy of the rooms to the common frame.

Usually, indoor environments have been reconstructed using an image sequence taken with a camera with a wide-angle rectilinear (non-fisheye) lens. To avoid the problems of external orientation, spherical panoramic images can also be applied if the projection centre of the camera is located at the rotation centre of the camera mount (Haggrén et al., 2004; Kauhanen et al., 2016). It is also possible to generate dense point clouds and textured mesh models from spherical panoramic images (Pagani and Stricker, 2011; Gava and Stricker, 2016). Panoramic image sequences can also be used to produce photorealistic virtual environments like Google Street View (Anguelov et al., 2010). In such cases, geometric information is needed for solving the relative positions of the panoramic images. This approach does not allow completely free movement of the user, as the images are only available from a limited number of positions. Increasing the number of image acquisition points allows moving with finer steps, but also adds to capturing work.

*Laser Scanning of Indoor Environments*

Laser scanning is employed either from a static instrument (terrestrial laser scanning; TLS) or from a moving platform (mobile laser scanning; MLS) (Thomson et al., 2013;



FIG. 1. Segment of an example image from an individual room (left). The white walls proved to be too featureless for image matching (right).

Kaijaluoto and Hyyppä, 2015). In outdoor environments, most MLSs rely on global navigation satellite systems (GNSS). As these are not available indoors, other localisation methods are required (Lehtola et al., 2015). The concept of simultaneous localisation and mapping (SLAM) is increasingly applied in indoor MLS. Commercial indoor MLSs have emerged, ranging from light, highly portable devices to cart-mounted devices intended for highly efficient measuring of large indoor environments. Table I gives an overview of three different laser scanning systems, including a conventional TLS and two indoor MLSs.

Laser scanning systems typically produce dense 3D point clouds. As most systems operate only using a single laser wavelength, imaging sensors have to be applied if coloured point clouds are to be obtained. Several laser scanning systems include integrated cameras for this purpose. For producing textured mesh models from laser scanning point clouds, additional processing is required. Some commercial services have emerged offering mesh model production from laser scanning point clouds (Autodesk, 2017; Sequoia, 2017), but a dense, coloured point cloud remains the most common end product of laser scanning.

TABLE I. Properties of indoor mapping systems. The Faro Focus is a conventional TLS whereas the NavVis and ZEB-Revo are MLSs.

<i>System</i>	<i>Property</i>	<i>Description</i>
Faro Focus 3D S120	Sensors	Laser scanner, RGB camera, inclinometer, compass and height sensor
	Weight	5 kg, tripod also required
	Measuring speed & range	Up to 960 kHz, range up to 120 m
	Measuring accuracy	$\pm 2$ mm ranging error
NavVis M3	Sensors	3 Laser scanners, 6 cameras
	Weight	Approx. 40 kg (total weight of trolley)
	Measuring speed & range	Range 30 m, estimated capacity 2000 to 5000 m <sup>2</sup> /h
	Measuring accuracy	20 to 30 mm
ZEB-Revo	Sensors	Laser scanner, inertial measuring unit
	Weight	1 kg
	Measuring speed & range	Continuous measuring at 43 kHz, range 30 m
	Measuring accuracy	$\pm 0.1\%$

### *Depth Camera Indoor Mapping Systems*

Emergence of highly affordable consumer depth cameras, such as the Microsoft Kinect, stimulated a lot of research for applying depth cameras for 3D modelling of indoor environments, especially for robot navigation applications (Henry et al., 2012). Depth camera systems do not, ideally, operate outdoors as sunlight interferes with their sensors. The geometric quality of depth cameras has been studied and the results show that the random error of depth measurement ranges from a few millimetres up to about four centimetres (Khoshelham and Elberink, 2012). Following this, several commercial products utilising depth camera sensors have been available (DPI-8, 2017; Structure, 2017). In addition, depth camera sensors integrated into smartphones have been introduced (Tango, 2017). For a performance comparison of handheld systems ranging from consumer grade to professional instruments, see Kersten et al. (2016).

Due to their limited field of view, single depth sensors are not ideal for mapping large indoor environments (Diakité and Zlatanova, 2016). For a more comprehensive 3D acquisition, it is possible to combine data from several sensors (Kim et al., 2008).

*Matterport System.* As a highly integrated commercial indoor mapping system, Matterport combines a 3D measuring instrument with dedicated cloud-based services. One of the intended applications of the system is efficient production of indoor marketing material for real-estate applications. To accomplish this, the system combines three depth camera sensors with red/green/blue (RGB) cameras. Primary products of the system are panoramic image walkthroughs, produced with the RGB cameras. In addition, a 3D model of the environment is produced, facilitating localisation of panoramic imaging locations and rendering an overview of the environment.

In operation, the measuring speed of Matterport is approximately 30 seconds per scan; the recommended scanning location placement is every 1.5 to 2.5 m. The weight of the scanner is about 3 kg, excluding the obligatory tripod. The processing of scans in Matterport is performed with a high degree of automation utilising a cloud computing service associated with the instrument (Matterport, 2017a). An optimised alignment of the scans is solved and a textured mesh model created (Bell et al., 2013) once the scans have been uploaded to the service. The dimensional accuracy of the final model is estimated to be 1% (Matterport, 2017a). After processing, the 3D model with its texture images can be downloaded from the cloud.

While research exists comparing the performance of various depth-camera-based instruments (Kersten et al., 2016), the application of mobile devices with a depth camera for indoor mapping (Diakité and Zlatanova, 2016) and development of depth camera indoor mapping systems (Henry et al., 2012), very little research exists on the performance or application of the Matterport system.

## MATERIALS AND METHODS

### *Unity 5 Game Engine*

The Unity 5 game engine (Unity, 2017a) was used in the project for developing the final application. It offers a comprehensive environment for development of interactive 3D applications for several platforms, including browser-based applications with WebGL and mobile applications, for example, on Android platforms.

For 3D content, Unity utilises mesh models that can be imported into the Unity editor in a number of formats (Unity, 2017b). Extensions that allow Unity to render point clouds have been developed (Unity, 2017d), but mesh models remain its core 3D content format. Ideally, the mesh models should be of limited vertex count, have a metric scale, contain UV coordinates for texture mapping and with 2D texture images of limited resolution (Unity, 2017b).

For developing mobile or browser applications, more attention should be given to optimising the applications in terms of memory use and performance (Unity, 2017c). This further reduces the usability of high-resolution bitmaps and high-polygon-count meshes, as well as possible shader complexity in application development.

In the presented development case, a memory limit of 2048 Mb and the exclusion of advanced shader programs were imposed to allow sufficient performance on lower-end platforms. The texture files used were resampled to a resolution of 1024 × 1024 pixels, audio files were recompressed and converted to mono, and only diffuse shaders were used in the rendering.

### Indoor Mapping with Matterport

Characteristics of the historic building had to be taken into account when planning for the indoor mapping. First, no GNSS coverage was available inside the building. Nor was it possible to mount any permanent reference targets to the interior. As the interiors had to be measured as staged, the use of moveable targets was also considered problematic. This meant that the measuring method chosen had to operate using merely the features of the indoor environment for localisation. Second, the building contained a number of stairwells, some of which were quite narrow. A highly portable instrument was therefore needed. Third, the scanning was to be performed in a limited time frame, working in cooperation with the Yle scenographers. As it was impossible to empty the entire building for the survey at one time, the measuring was to be done in a number of sessions.

Out of available indoor mapping systems, the depth-camera-based Matterport, along with its associated cloud service, offered a high degree of automation for modelling combined with a portable instrument capable of mapping multi-room indoor environments. MLS and TLS were considered to offer too low a degree of automation in model processing compared with a depth camera. TLS was also considered too laborious for mapping large indoor environments with low occlusions. As for photogrammetry, the environment contained several featureless and poorly lit areas, making image-based acquisition difficult.

Indoor mapping of the building was performed with the Matterport instrument over the course of three days (Fig. 2). The scanned interiors included a number of smaller rooms (maximum size of  $\sim 20\text{ m}^2$ ), a large corridor complex with a length of tens of metres and a number of stairwells. Five individual rooms were scanned twice, alternating the number of scanning stations in the room on each occasion (Table II). In one particular space where a room was scanned twice, the position of a single bed in the room was shifted from one side of the room to the other between scans. This was done as the room was to be staged having two beds, but only one was available at the time. The scans were partially overlapping: some of the individual rooms were scanned together with the adjacent corridor and later on scanned separately. In a similar fashion, the scans of the stairwells also contained some segments of the corridors. This was done to obtain the necessary degree of overlapping detail, for example, from door frames, for combining the data from several rooms in further processing.



FIG. 2. Operating the Matterport system (left) and a segment of the completed textured model obtained from the cloud showing the end of the largest corridor (right).



TABLE II. Scanned interiors of the historic hospital building. Some rooms were scanned twice.

<i>Title</i>	<i>No. of scans</i>	<i>Polygon count</i>	<i>No. of texture maps</i>	<i>Projective top area (m<sup>2</sup>)</i>	<i>Mesh surface area (m<sup>2</sup>)</i>	<i>Bounding box size<sup>a</sup> (m)</i>
1st floor corridor system and rooms	82	49 828	101	480.5 <sup>b</sup>	2173.2	65.7
Hospital room 1	8	223 135	4	15.2	100.9	7.3
Hospital room 2	7	212 356	4	11.9	89.7	6.9
Hospital room 3	9	204 561	5	15.7	105.2	7.5
Hospital room 3	4	229 913	5	15.7	106.3	7.4
Hospital room 4	11	233 920	6	20.7	131.5	8.2
Hospital room 4	6	245 150	6	20.2	121.5	8.0
Hospital room 5	10	211 315	5	14.6	103.6	7.3
Hospital room 5	4	242 321	4	14.7	101.7	7.3
1st floor toilets	8	179 321	5	20.1	110.9	9.7
Staff room 1	12	175 570	4	12.2	86.7	7.0
Staff room 1	3	215 529	4	12.5	85.6	7.2
Staff room 2	12	220 463	6	20.8	125.6	8.2
Staff room 2	4	264 026	6	20.8	125.9	8.2
Isolation room	6	147 203	2	5.9	45.5	5.2
Ground floor corridor	17	79 886	9	47.3 <sup>b</sup>	225.2	18.0
Ground to 1st floor stairwell	31	112 012	20	87.3 <sup>b</sup>	462.2	23.8
Dining area	24	140 878	12	65.8 <sup>b</sup>	273.3	14.6
Utility room	29	163 871	16	51.4 <sup>b</sup>	389.9	17.4
Stairwell to attic	32	141 637	9	29.4	174.1	12.7
Attic	65	97 558	52	345.8 <sup>b</sup>	1183.3	40.9
TOTALS	384	3 790 453	285	1328.5	6321.8	

<sup>a</sup>Box size is corner to corner.

<sup>b</sup>Maximum triangulation edge length 1 m, rather than the normal 0.1 m.

The projective top area of models was estimated by projecting all the mesh vertices onto the XY plane, performing 2D Delaunay triangulation and then calculating the area of the resulting mesh. The maximum edge length used in triangulation was 0.1 m, except for the models marked with an asterisk, where a length of 1 m was used, as the point clouds were too sparse to be triangulated with a shorter maximum edge length. It roughly resembles the floor area in single-storey interiors.

### *Terrestrial Laser Scanning*

TLS was performed with a Faro Focus3D S120 scanner with 905 nm wavelength and having a ranging error of  $\pm 2$  mm at 10 (90% reflectivity) and 25 m (10% reflectivity) (Fig. 3). The evaluation data was acquired with two resolution settings, providing point spacing of 6.3 and 12.3 mm at a distance of 10 m from the scanner. The scanner also captures panoramic images in order to provide colour for the 3D measurement points. In total, 45 scans were conducted. The principle for gathering the scanning data was that every room had one to three scans depending on the degree of occlusion. The rooms were also scanned from beyond the doorways to ensure that rooms and adjacent corridors could be registered together using sufficient mutual data overlap. The long corridors were measured by setting up a station every 5 m. The scans were matched together using visual alignment and an automatic cloud-to-cloud method provided by the Faro Scene software (Version 6.0).



FIG. 3. Preparing to scan the main corridor of the building (left) and a coloured point cloud resulting from several co-registered scans (right).

### *Evaluation of Geometric Deviations of Indoor Models*

Models produced with the Matterport instrument were compared with the TLS point clouds using CloudCompare (64 bit version 2.6.1). First, the point clouds produced by co-registering the individual scans were downsampled to produce a homogenous point density with a 1 cm point spacing and manually segmented to remove reflections from windows and other outlier points.

Matterport models were analysed by using the unedited mesh models from the cloud service. As the datasets did not completely overlap, the non-overlapping segments were removed. Data were preliminarily aligned manually and then using iterative closest point (ICP) for farthest-point removal. Three separate sets were then compared: one large corridor environment; one smaller corridor environment; and an individual small room. The comparison was performed by using the “Mesh to cloud” function of the software, in effect searching for pairs of closest mesh triangles and point cloud points and calculating their separation.

### *Manual Model Refinement*

The first stage of processing was performed automatically by the cloud service associated with the instrument. The pre-aligned scans were further registered, triangulated and textured automatically in the cloud service. The resulting mesh models and texture maps were then downloaded and 3D Studio Max was applied for refining the models.

The manual processing began by manual orientation of the different models, using overlapping areas, such as door frames, as a reference. A number of rooms were present in the large scan projects, but had also been scanned as separate projects. In these cases, the separately reconstructed rooms were preferred, as they offered a higher level of detail. The processing then progressed with refining the object structure of the automatically produced models. Issues in detail level and object structure are further illustrated in the Results section.

After this, the modelling efforts progressed room by room. Several steps were taken to improve the visual appearance and technical functionality of the models, using both manual and automatic tools. Non-manifold geometry was manually repaired. Holes left in the mesh were filled, distortions in surface geometry were smoothed out and unnecessary edge points were deleted. Difficulties were encountered in badly occluded areas, for which the geometry had to be manually crafted. This included highly reflective surfaces, such as windows, and

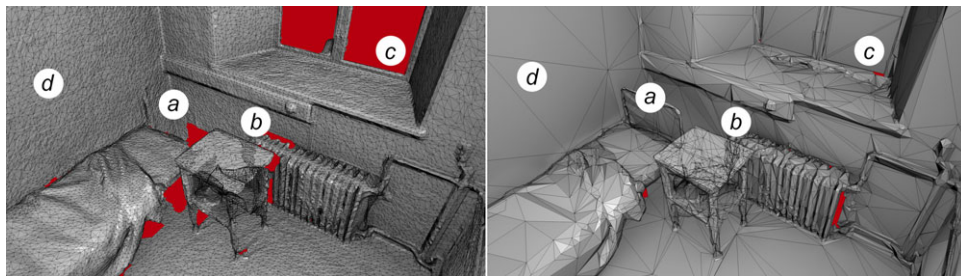


FIG. 4. Comparison of original mesh (left) and edited mesh (right) from a hospital room. (a) Manual reconstruction of missing furniture details. (b) Manual filling of holes. (c) Addition of planar surfaces to windows. (d) Significantly reduced polygon count on planar surfaces.

geometrically complicated details, like radiators and railings, both of which were poorly reconstructed in the scans. The aim was to accomplish as low a polygon count as possible, whilst maintaining the essential geometry and appearance of the model (Fig. 4). The texture maps were manually edited to remove historically non-correct details, such as post-it notes accidentally left on the walls.

The final model used in the application was assembled by combining several separately processed models, combining the individual scans of rooms and corridors into a continuous indoor environment. Some alterations were made to the floor plan of the building to facilitate the storytelling. For example, the isolation room, which was in reality located in the basement, was moved to the first floor. This was done as there were no other spaces in the basement required for the story, and the route to there would have been very complex and required expanding the scanning campaign to a number of stairwells and corridors. In total, the final model consisted of 292 000 polygons.

#### *Audio Production and Set-up*

In total 42 audio clips, resulting in a total play length of approximately 50 minutes, were produced for the virtual radio play. In the completed clips, 23 actors played a total of 46 scripted roles. In addition, a piano was used as a musical instrument.

For triggering the audio content in the application, the location of the user is utilised: the audio clips are played when the user appears within a defined boundary. In addition, maximum and minimum volume boundaries were defined. The minimum volume boundary defines where the volume begins to gradually increase towards the maximum boundary. Within the maximum boundary, the volume remains constant. For a number of doors that the user can pass through, both location and orientation are used: the user enters when standing next to a door and facing towards it. Audio content was also associated with the user opening (or failing to open, if locked) doors. These audio effects were not triggered by the user location, but by the user clicking, or trying to operate, on the defined objects. The final audio set-up in the virtual radio play, therefore, consisted of a set of audio clips varying in intensity, triggered by the user's presence, and another set of clips triggered by interacting with an object (Fig. 5).

The extents of the triggering boundaries for the audio clips varied greatly in size. In practice, some of the sounds were atmospheric in nature (for example, a piano playing in the background) and some very intensive, located dialogue. Therefore, it was possible for the user to hear several clips at once, depending on the position.

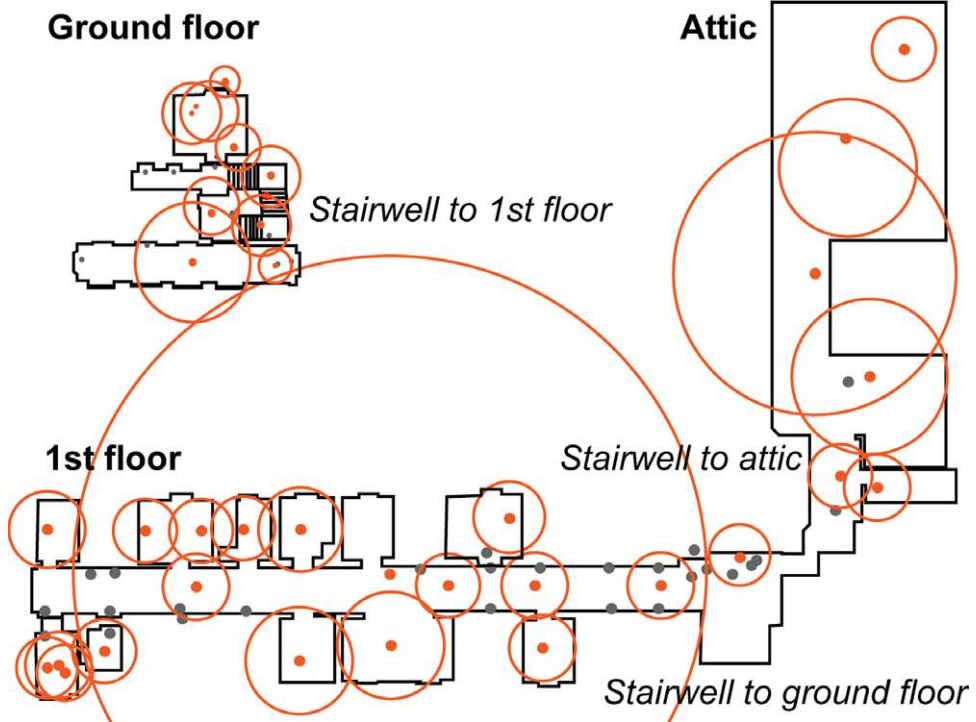


FIG. 5. The audio set-up showing presence-triggered sounds (orange dots) along with their trigger boundaries and clips triggered by interaction (grey dots).

RESULTS

The overall data acquisition and modelling process is summarised in Fig.6. Three working days were used for staging and 3D scanning of the interiors. In the data acquisition campaign, all of the encountered indoor environments were successfully mapped with the

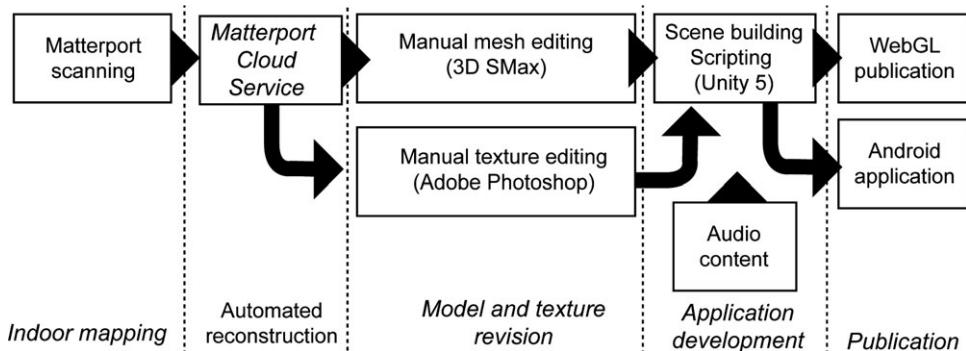


FIG. 6. The final applied process combining automated reconstruction in the Matterport cloud with manual model revision.

Matterport system. This was the expected result as the environments did not contain problem features such as large open interiors, areas with only repetitive features, outdoor or brightly sunlit areas, or highly mirroring environments, all of which are reported to potentially cause issues (Matterport, 2017a). As such, the measuring system was efficient, with scanning times for an individual scan being shorter than for most TLS instruments (for example). In smaller spaces, such as individually scanned rooms smaller than 20 m<sup>2</sup>, the limited range of the instrument when compared with laser scanners did not necessitate more measuring positions either. In longer corridors, more measuring positions were needed by Matterport than a typical TLS would have required.

The automated reconstruction in the cloud service was initiated immediately following measurement. Accurate processing times were not recorded, but the automatically produced models were in all cases available in the morning following the scanning day. The editing phase for the models consisted of editing the meshes and the texture images separately, and formed the most laborious part of the project, requiring several weeks of working time from an experienced modeller. A model revision phase was required for combining separately processed models and optimising the models for mobile 3D. Afterwards, the Unity 5 game engine was used to program the final application, published both as online content and mobile software. In practise the application development and model editing partially overlapped, in total requiring approximately seven weeks. The model editing work was estimated to have taken roughly 40% of the working time during this stage, with the remaining time spent on programming the application (about 30%) and debugging and revising the application according to feedback from the project group (about 30%).

In the application development phase, a user interface was designed and implemented for the application. The main focus was in maintaining a simple user interface and not to obstruct the actual content. Movement and camera control were realised following the conventions of mobile- and browser-based 3D games. The user moves through the building in a first-person view (Fig. 7). A 2D map function was also implemented, allowing the user to jump directly to a specified space or location (Fig. 8). The resulting application was published online both as a WebGL version and as a downloadable Android application for mobile users.

In total, during the two-day event, approximately 3000 people visited the installation on site. The produced game engine version received 813 unique visitors during the first two weeks following its publication. During the following six months, the number of unique visits had increased to 1001, with a total of 1663 downloads.

### *Geometric Deviations and Detail Level of Models*

When the Matterport models were compared with the TLS point clouds, several deviations were observed (Fig. 9). In general, it appears that Matterport produces models that are slightly too large, with this being most visible in the largest models. This produces significant mismatches in larger datasets, when the models can no longer be fully aligned.

Most significant local deviations were found in small details visible in the TLS clouds but not reproduced by the mesh generation algorithm applied in the Matterport automated processing including, for example, houseplants, small objects and radiator installations (Fig. 9). There was also a minor corner rounding effect visible in models.

The aforementioned issue was significantly more severe when the models covered a larger area. When the measured target zone is smaller, the Matterport system produced more detailed meshes. On the other hand, with larger entities the mesh quality decreases. This is clearly visible when comparing the reconstruction of details in large target areas (such as a corridor with several associated rooms) and a scan of an individual small room (Fig. 10).

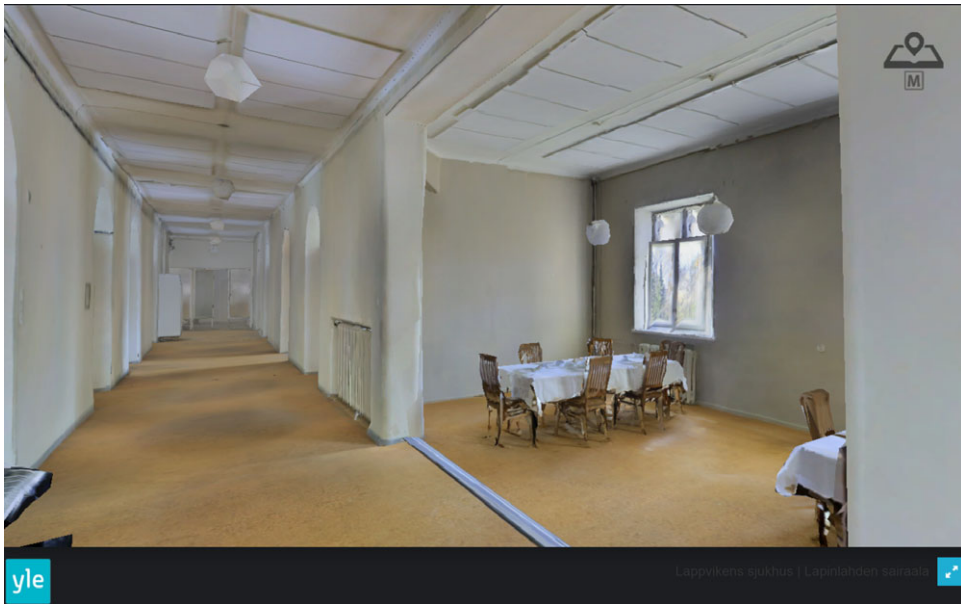


FIG. 7. Running the application, where the user is exploring the space.

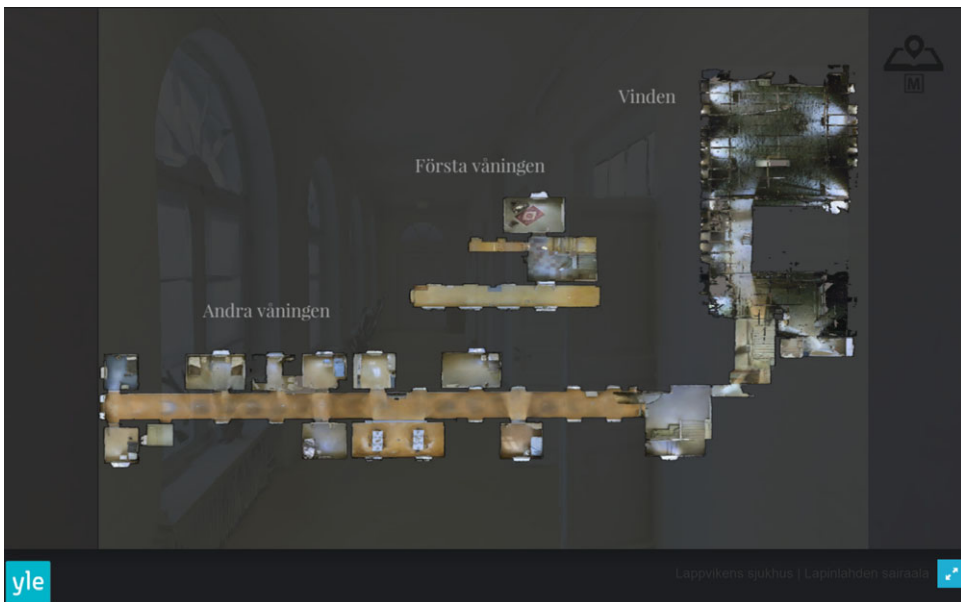


FIG. 8. Running the application in map view.

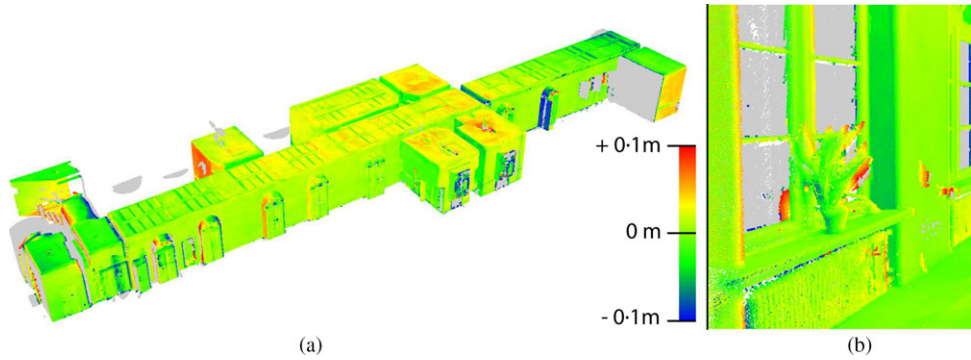


FIG. 9. Comparison of a Matterport mesh with a TLS point cloud. (a) Model coloured by deviation, depicting a large corridor scene. (b) A segment from a smaller corridor, showing several plants and other details.

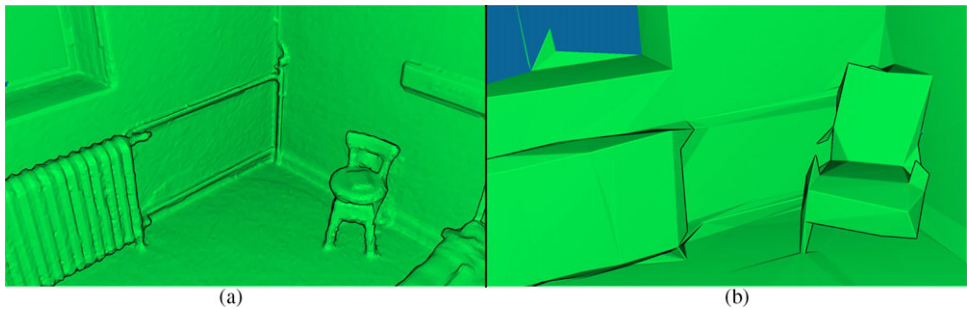


FIG. 10. Comparison of Matterport reconstruction in scans covering only a small room (a) and a large corridor scene with several adjacent rooms (b). The degree of detail in reconstruction varies significantly.

In a large corridor combined dataset, there is also a noticeable difference between the component datasets. For example, in Fig. 9 observe the deviations on the ceiling surface, comparing the middle of the corridor with its ends. The corridor appears slightly bent vertically. This was investigated further by segmenting the floor surface from both TLS and Matterport, and performing a planar fitting with consecutive distance evaluation. From the results (Fig. 11) it can be observed that, for the long corridor dataset, the TLS point cloud reproduces the floor in a significantly more planar form than that found in Matterport data.

To study this issue further, the planarity of the floor surface was evaluated from an individual TLS scan. This was to eliminate the possibility of a cumulative co-registration error in the TLS. In the Matterport cloud, a fairly smooth descent of approximately 5 cm can be seen from one end of the corridor to the other (area marked in Fig 11(a)). When viewed as a single scan, some variations in the floor can be observed (Fig. 11(b)), but they generally fall in the  $\pm 2$  cm range and do not display the same trend.

### *Effects Induced by Model Processing*

The model optimisation process described earlier had a significant impact on the computational size of the generated models, as intended. Fig. 12 provides the vertex and

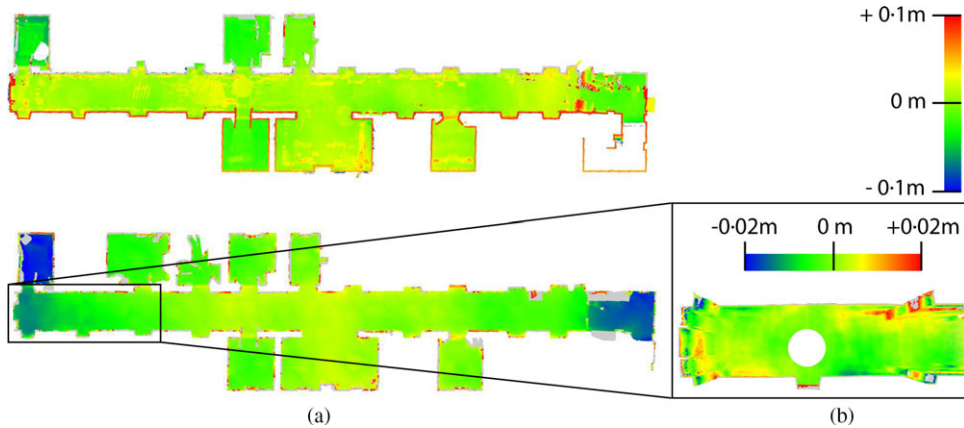


FIG. 11. Comparison of floor surface deviations from a plane for TLS (a, above) and Matterport (a, below). For one end of the corridor, analysis was also performed for a single TLS scan (b); note the different scale.

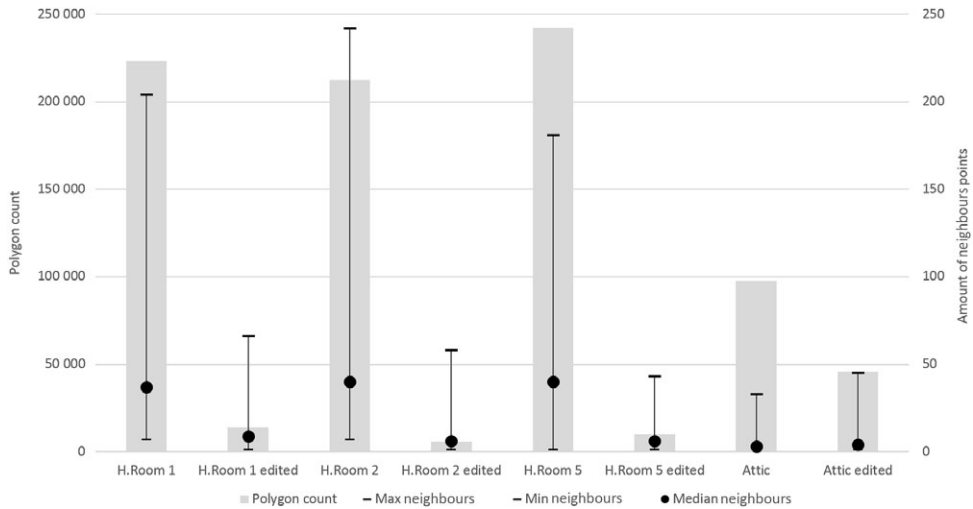


FIG. 12. Polygon counts (grey) and neighbouring point counts (black lines; medians as black circles) for a set of sub-models for four rooms (two adjacent columns for each room; left unedited; right edited).

polygon counts for a set of four models in their original state (as downloaded from the cloud) and after optimisation/editing (before being transferred to the game engine). In addition, the number of neighbouring points has been calculated for the vertices, using a search radius of 0.1 m. The maximum and minimum values are given, along with the median for the mesh.

The impact of processing is clearly visible in the smaller rooms. Fig. 13 shows the mesh vertices of a room coloured according to the number of neighbouring points. In the original model, the density remains relatively homogenous over large planar surfaces, such as walls. In the processing, the vast majority of vertexes are removed from these planar surfaces. The vertices that remain relate to edges and areas with complex curvature, such as furniture details.



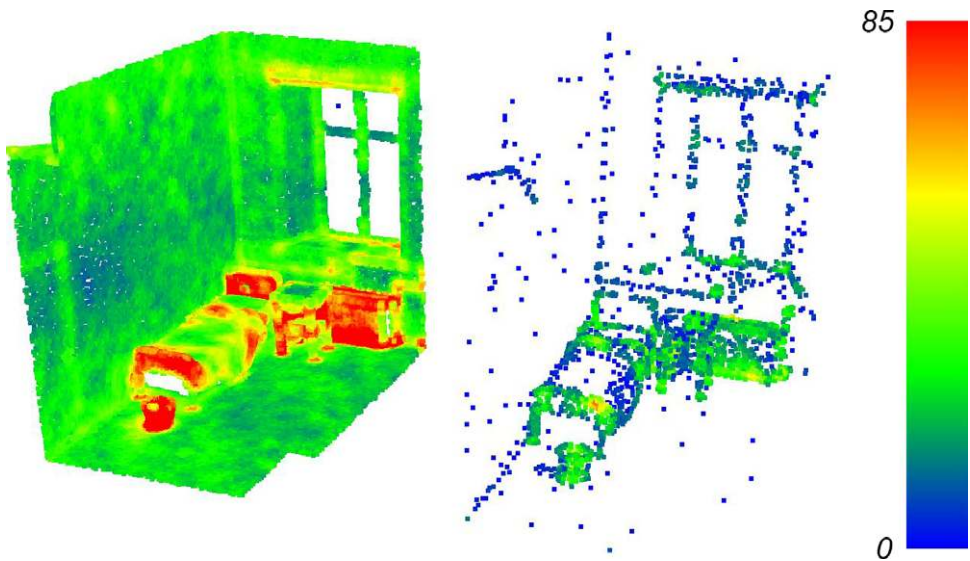


FIG. 13. Vertex points of an individual hospital room (data for the ceiling and two walls removed for illustration purposes), coloured according to the number of neighbouring points before (left) and after (right) processing.

Manual editing also affected the object structure of models. The automated reconstruction created several texture maps for each project and divides the mesh into corresponding objects. Using this structure directly would have created a scene with a large number of separate objects, resulting in a large number of draw calls for the graphics processing unit (GPU), thereby reducing performance. Additionally, in larger spaces, some of the automatically built meshes were large, which would have caused issues for optimisation by occlusion culling and frustum culling (removing models beyond the viewing area), as bounding-box dimensions are used in determining whether objects are rendered. The object structure of models was revised, combining meshes from individual rooms and splitting meshes from large corridor spaces (Fig. 14). The goal was to maintain a sufficiently

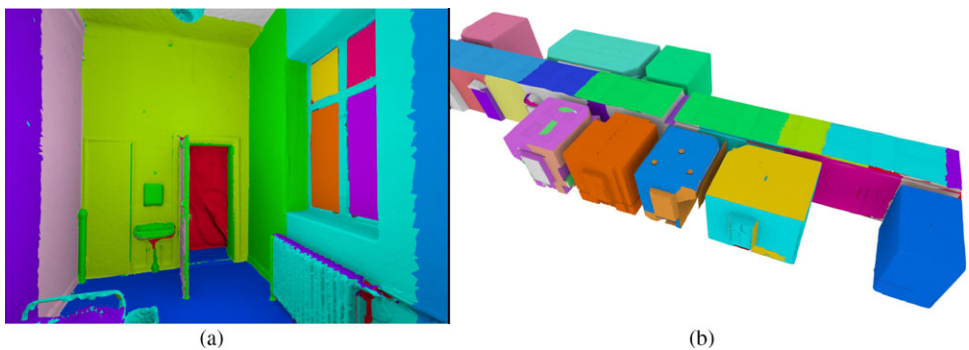


FIG. 14. Left: the object structure of an individual room created by Matterport system. Right: the revised object structure. Colours for objects have been assigned randomly.

low number of meshes to reduce draw calls, but simultaneously keep the extents of individual meshes limited, so that they are properly removed from rendering by occlusion culling and frustum culling.

### DISCUSSION

The ability of the measuring system to automatically produce textured mesh models of indoor environments was a critical factor in the work presented and also the reason for applying the specific system. While the mesh models were, technically, directly applicable in the game engine content production, they did require a significant amount of manual editing.

The amount of work in the manual editing exceeded, by far, the hours required for performing the measuring or developing the application, undermining the benefits of automated reconstruction offered. The need to optimise the final applications for smooth operation on low-end platforms increased the need for manual optimisation of the polygon geometry and object structure. Had the end product only been intended for high-end PCs, the amount of manual work would have been reduced. The estimation of working times was further complicated by the number of changes to the final application that arose from the project team during the application development and model revision phase.

When discussing the quality of models produced with the depth camera system, it should be noted that the primary intended use of the system was the production of panoramic image walkthroughs, with the 3D model being somewhat of a side product. Further, the experiments were carried out with the first production version of Matterport. The manufacturer has since released a new version of the instrument (Matterport, 2017b) and made new features available as a beta version in the cloud service (Matterport, 2017a).

In this work, analysis of the automatically produced models revealed several issues. First, the models have significantly varying polygon counts and density. If one compares the polygon densities (polygons/mesh surface area), it can be seen that it is low (23 polygons/m<sup>2</sup>) for the large corridor dataset, improved in smaller corridors on the ground floor (355 polygons/m<sup>2</sup>) and high for small rooms (for example, 2211 polygons/m<sup>2</sup> in hospital room 1) (Table II). This seems to correlate with the size of the measured area (Fig. 15). This

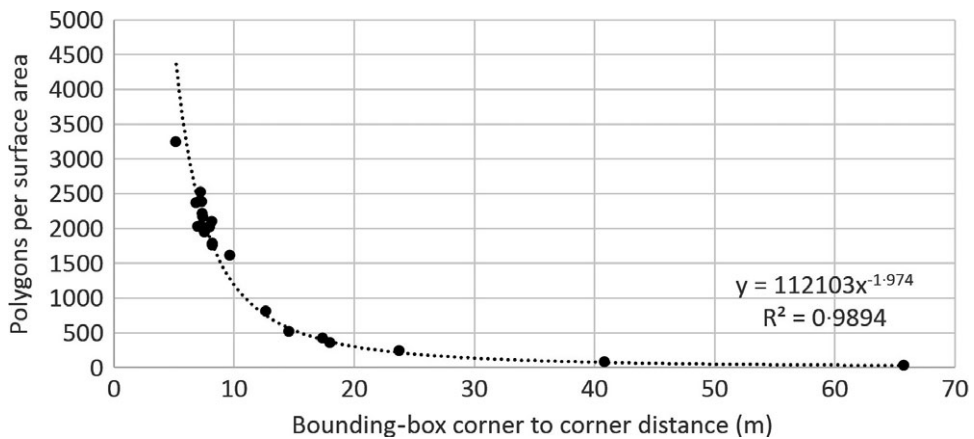


FIG. 15. The correlation between the bounding-box size and mesh polygon density for all models in the project.

is clearly visible in the level of model detail (Fig. 10). In large datasets, the system's ability to represent small geometric details was greatly reduced. Adding measuring positions did not have a significant influence on the polygon count of the model produced (in hospital room 5, a model with four scans actually had more polygons than when using 10 scan positions). This would imply that, for obtaining detailed models, the size of the measuring area should be kept as small as possible to preserve as high a level of detail as possible. Since the model fits almost identically to a mathematical model of the form  $y=1/x^2$  (Fig. 15), the limitation is obviously coming from Matterport memory allocation in the cloud service for a given object volume. The phenomenon is assumed to be more dramatic if the studied object has large variability in three dimensions, not just in two dimensions.

The limitations of the system are also acknowledged by the manufacturer, stating a relative 1% accuracy for the mesh models. In this project, for the smaller hospital rooms (bounding-box corner-to-corner distance below 10 m), a 10 cm accuracy could be expected. For the large corridor dataset (corner-to-corner distance 65.7 m), this would imply an accuracy of 0.65 m. The observed mismatches between the TLS data and Matterport fell within these specified limitations (Fig. 11).

As the models used in game engine content production also commonly feature image textures, the geometric properties of the models are not the only factors affecting their appearance. The Matterport system automatically acquires photographic textures for the models and maps them to the mesh model. As seen in Table II, the models contained a varying number of texture maps, having a size of 2048 by 2048 pixels. Each texture map therefore contains over 4 megapixels of image data (as there are some gaps in texture maps; the actual amount of information is smaller). The number of texture pixels per model surface area for the large corridor was actually higher (194 831 pixels/m<sup>2</sup>) than for the smaller corridor (167 623 pixels/m<sup>2</sup>) or for the smallest room (166 276 pixels/m<sup>2</sup>). Even though the degree of detail decreased for the model geometry, the texture resolution did not decrease with area size.

For content production, these observations imply that Matterport is suited for producing 3D overviews of indoor environments with a high degree of automation. For well-optimised game engine models, further manual steps are required. More detailed models, that display smaller features, are only produced when the measured areas are small enough. However, these details remain in the texturing. The benefits and disadvantages of measuring and the automated reconstruction process applied are summarised in Table III.

Alternative model production techniques for this case might have been the application of surface reconstruction algorithms from dense point clouds obtained with TLS (for example, Lorensen and Cline, 1987; Kazhdan and Hoppe, 2013). Some commercial services offering textured mesh model generation from point cloud datasets are already available (Autodesk, 2017; Sequoia, 2017). By applying these, the production of game engine compatible mesh models from TLS data might have been possible. In this respect, one of the limitations of the applied system is the unavailability of raw data. In Matterport,

TABLE III. Summary of benefits and disadvantages of the system applied for indoor measurements.

<i>Benefits</i>	<i>Disadvantages</i>
Measuring work possible by single operator	Models had significant variation in density
By automated reconstruction the initial models were attained with very little manual work	No access to raw measuring data for user
Results of automated reconstruction were compatible with game engine content production workflow	Extensive manual refinement of models was still required

downloading anything other than the automatically produced mesh models has so far not been possible. This feature is currently in the beta version (Matterport, 2017a). The availability of co-registered point clouds would significantly increase the utility of the instrument with other workflows and potentially help reduce the issues encountered with automatically generated mesh models.

An additional benefit of applying the Matterport system is that it also produces a panoramic image set automatically. While this was not used in the presented case, it nevertheless is an easily comprehensible presentation of the spaces as they were at the time of imaging. As such, the system can be used to efficiently document temporary indoor setups, such as the staged interiors in this case.

### *Impact of Virtual Radio Play*

Using a game engine as a platform, a virtual version of the audio installation was produced. The application was published online alongside the unveiling of the installation. There were some differences between the virtual version and the actual installation. The lighting conditions of the virtual version matched those of the scanning, which was performed in daytime. As the actual event occurred late in the evening, the atmosphere of the building was quite different (Fig. 16). The situation was also different since other visitors were present in the space during the event. In the virtual version, the user is alone in the building.

The presented case also provides an example of applying new technical possibilities in cultural productions. The end result contains elements from computer gaming, installation and radio theatre. A virtual environment can be used to produce spatial experiences and multisensory works. This has an influence on the traditional concepts of installations and



FIG. 16. An individual hospital room (left) and the ground floor corridor (right) shown in the virtual version (top) and during the actual exhibition of the installation (bottom).

spatial art, where the primary mode of experience has been a person's physical presence. By utilising a virtual environment, more freedom is gained compared to physical installations in existing buildings. At the same time, the possibility of physical presence is lost. Understanding the impact of digital mediation on the experience would benefit from in-depth psychological studies and extensive user tests.

### *Virtual and Augmented Reality Technologies*

Virtual reality (VR) offers, by definition, any kind of experience that enables an effective immersion for its user in a responsive virtual world (Brooks, 1999). It is a medium with the essential goal of assuring you that you are somewhere else (Parisi, 2015). The user can interact with the virtual environment analogously to the real world (Mazuryk and Gervautz, 1996). Thus, VR has an enormous potential to revolutionise the way people experience virtual spaces. The crucial enabling technologies for VR are: (1) 3D displays; (2) motion tracking; (3) input devices; and (4) software frameworks and development tools that allow real-time graphics rendering of 3D virtual worlds (Brooks, 1999; Parisi, 2015). VR has numerous applications in gaming and entertainment, virtual worlds, tourism, architecture and real estate, design and planning, education and training, telepresence and cooperative working, among many others (Mazuryk and Gervautz, 1996; Parisi, 2015). Developments of browser-based VR, computer gaming hardware and headsets utilising smartphones are making VR available to the masses. A second significant development direction is the integration of VR and augmented reality (AR) into mixed reality, allowing immersive presentation of virtual elements on top of the real world (Boland and McGill, 2015).

For creating a virtual experience, utilising VR technology would have surely enabled a deeper immersion. As the adoption of VR hardware is still low, this would have limited the accessibility approach, central to producing a virtual version in this case. Utilising VR would have required the development of a user interface allowing moving and interacting with the space, as the current keyboard and mouse/touchscreen user interface is not ideally suited for VR. The sufficiency of models for head-mounted VR displays also remains an open question.

AR mixes virtual elements with the real world. It has been successfully applied in cultural heritage by bringing additional "layers" to historically significant sites (Portalés et al., 2009). In the presented case, location-based technologies and augmented reality could have enabled bringing interactive content to the physical installation. However, this would not have allowed studying the content outside the site.

### *Digital 3D Models as a Platform*

In the literature, digital geometric models are often suggested as a platform for various uses (for example, Alatalo et al., 2016). However, in the project presented here, the focus shifted from mapping and modelling to storytelling. The user does not only explore a model, but an entity composed of literary content (stories) presented as audio, an interaction system (moving, location-based triggers) and the environment (model). The model became subordinate to the story. Further, the storytelling aspect of the project conflicted with the measured model of the interior: some elements had to be removed from the model to maintain the credibility of the narrative and some of the spaces were relocated to make discovery of the story easier.

Furthermore, the model used in application development is not only a model of a space, but may contain other elements, such as certain furnishings. In this case, the interiors

were measured as staged, fitted out with a set of furniture and with other items in them. If the application had been built using an empty model of the building, all the staging would have had to be done virtually. The idea of a “general purpose model” that can be applied to a variety of different cases becomes problematic. Even if a pre-existing interior model of the building had been available, it would have required extensive editing to be applicable for this use.

## CONCLUSIONS

A game engine is suited for producing interactive virtual environments for various purposes. In the presented case, a virtual environment for exploring audio content was created using a game engine and an indoor mapping system, allowing the “virtualisation” of a real environment. In the final product, the user freely explores a building and experiences a set of audio clips creating a narrative. The final application was published as a web-based production and as a downloadable application for Android smartphones and tablets.

For applications that take place in the real world, digital geometric models of the environment are needed. Building these models manually is time consuming. 3D measuring methods and automated reconstruction algorithms offer an alternative to this manual process. The authors applied a combination of depth camera indoor mapping and manual model revision for producing indoor models of a historic hospital building, which were used to develop a game engine application. The indoor mapping system was found to be efficient, exceeding the measuring speed of a TLS in confined interiors.

The automatically reconstructed models produced with the system were found to be usable as a starting point for the game engine content production. However, the quality of the automated reconstruction left plenty of room for improvement. The models were extensively refined by manual processing to produce a model of a complex interior measured in several stages and to achieve sufficient optimisation for mobile 3D applications. The amount of work for optimising the models for game engine use exceeded the effort needed for the indoor mapping or application development.

The applied indoor mapping system, Matterport, was best suited for producing geometrically detailed models of smaller areas. In larger areas, the detail level and accuracy of models decreased, potentially due to fixed memory allocation in the cloud service. The quality of the textures, however, was not affected by this, partially aiding to maintain the appearance of models. In addition, some warping was observed in the model of a long corridor environment. Observed errors remained well within the tolerance stated by the manufacturer. The nature and intended use of the system should be taken into account when discussing these limitations. Objects having large volumetric variation in all three dimensions cannot always be mapped with the required accuracy, even though the number of scans is increased.

Despite the issues found, the applied system, together with the editing process introduced, proved successful in documenting a large indoor environment and producing models suited for mobile 3D applications.

## ACKNOWLEDGEMENTS

The authors are grateful to the Academy of Finland for their support of the Center of Excellence in Laser Scanning Research (CoE-LaSR; 272195) and project COMBAT (293389). In addition, the support from the European Regional Development Fund (ERFD) for projects AKAI (301130) and SOLUDUS is acknowledged. The Finnish Funding Agency for Innovation is acknowledged for support in project VARPU (7031/31/2016).

## REFERENCES

- ALATALO, T., KOSKELA, T., POUKE, M., ALAVESA, P. and OJALA, T., 2016. VirtualOulu: collaborative, immersive and extensible 3D city model on the web. *21st International Conference on 3D Web Technology*, Anaheim, California, USA. Pages 95–103.
- ALAVESA, P., OJALA, T. and ZANNI, D., 2014. PROPS: 3D-game-like mediator for improvisational storytelling. *Entertainment Computing*, 5(4): 381–390.
- ALLEN, C. T. and MADDEN, T. J., 1985. A closer look at classical conditioning. *Journal of Consumer Research*, 12(3): 301–315.
- ANDERSON, E. F., MCLOUGHLIN, L., LIAROKAPIS, F., PETERS, C., PETRIDIS, P. and DE FREITAS, S., 2010. Developing serious games for cultural heritage: a state-of-the-art review. *Virtual Reality*, 14(4): 255–275.
- ANGUELOV, D., DULONG, C., FILIP, D., FRUEH, C., LAFON, S., LYON, R., OGALE, A., VINCENT, L. and WEAVER, J., 2010. Google Street View: capturing the world at street level. *Computer*, 43(6): 32–38.
- ARHIPAINEN, L., PAKANEN, M., HICKEY, S. and MATTILA, P., 2011. User experiences of 3D virtual learning environment. *15th International Academic MindTrek Conference: Envisioning Future Media Environments*, Tampere, Finland. Pages 222–227.
- ARTHUR, E. J., HANCOCK, P. A. and CHRYSLER, S. T., 1997. The perception of spatial layout in real and virtual worlds. *Ergonomics*, 40(1): 69–77.
- AUTODESK, 2017. <https://www.autodesk.com/products/recap/overview> [Accessed: 10th July 2017].
- BELL, M., GAUSEBECH, D. and BEEBE, M., 2013. Capturing and aligning three-dimensional scenes. *U.S. Patent Application 13/776,688*.
- BELLOTTI, F., BERTA, R., DE GLORIA, A., D'URSI, A. and FIORE, V., 2012. A serious game model for cultural heritage. *Journal on Computing and Cultural Heritage*, 5(4): Article 17.
- BILLE, R., SMITH, S. P., MAUND, K. and BREWER, G., 2014. Extending building information models into game engines. *2014 Conference on Interactive Entertainment*, Newcastle, Australia. Pages 1–8.
- BOLAND, D. and MCGILL, M., 2015. Lost in the rift: engaging with mixed reality. *XRDS: Crossroads, The ACM Magazine for Students*, 22(1): 40–45.
- BROOKS, F. P., 1999. What's real about virtual reality? *IEEE Computer Graphics and Applications*, 19(6): 16–27.
- DIAKITÉ, A. A. and ZLATANOVA, S., 2016. First experiments with the Tango tablet for indoor scanning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3(4): 67–72.
- DPI-8, 2017. <http://trl.trimble.com/docushare/dsweb/Get/Document-748341/> [Accessed: 10th July 2017].
- FASSI, F., ACHILLE, C. and FREGONESE, L., 2011. Surveying and modelling the main spire of Milan Cathedral using multiple data sources. *Photogrammetric Record*, 26(136): 462–487.
- FRITSCH, D. and KADA, M., 2004. Visualisation using game engines. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(B5): 621–625.
- FURUKAWA, Y., CURLESS, B., SEITZ, S. M. and SZELISKI, R., 2009a. Manhattan-world stereo. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, Florida, USA. Pages 1422–1429.
- FURUKAWA, Y., CURLESS, B., SEITZ, S. M. and SZELISKI, R., 2009b. Reconstructing building interiors from images. *12th IEEE International Conference on Computer Vision*, Kyoto, Japan. Pages 80–87.
- GAVA, C. C. and STRICKER, D., 2016. A generalized structure from motion framework for central projection cameras. *Computer Vision, Imaging and Computer Graphics Theory and Applications*, Berlin, Germany. Pages 256–273.
- GEORGANTAS, A., BRÉDIF, M. and PIERROT-DESSEILLIGNY, M., 2012. An accuracy assessment of automated photogrammetric techniques for 3D modeling of complex interiors. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(B3): 23–28.
- GORINI, A., CAPEVILLE, C. S., DE LEO, G., MANTOVANI, F. and RIVA, G., 2011. The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking*, 14(3): 99–105.
- GREENWOOD, P., SAGO, J., RICHMOND, S. and CHAU, V., 2009. Using game engine technology to create real-time interactive environments to assist in planning and visual assessment for infrastructure. *18th World IMACS/MODSIM Congress*, Cairns, Australia. Pages 2229–2235.
- HAGGRÉN, H., HYYPÄ, H., JOKINEN, O., KUKKO, A., NUIKKA, M., PITKÄNEN, T., PÖNTINEN, P. and RÖNNHOLM, P., 2004. Photogrammetric application of spherical imaging. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5/W16): 22–32.
- HEINRICH, W. L., YOUNGBLOOD, P., HARTER, P. M. and DEV, P., 2008. Simulation for team training and assessment: case studies of online training with virtual worlds. *World Journal of Surgery*, 32(2): 161–170.

- HENRY, P., KRAININ, M., HERBST, E., REN, X. and FOX, D., 2012. RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments. *International Journal of Robotics Research*, 31(5): 647–663.
- HERWIG, A. and PAAR, P., 2002. Game engines: tools for landscape visualization and planning? *Trends in GIS and Virtualization in Environmental Planning and Design: Proceedings at Anhalt University of Applied Sciences 2002* (Eds. E. Buhmann, U. Nothhelfer and M. Piersch). Wichmann, Heidelberg, Germany. Pages 172–182.
- JANCOSEK, M. and PAJDLA, T., 2011. Multi-view reconstruction preserving weakly-supported surfaces. *IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, Colorado, USA. Pages 3121–3128.
- KAJALUOTO, R. and HYYPÄ, A., 2015. Precise indoor localization for mobile laser scanner. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(4/W5): 1–6.
- KAJANDER-MAAVUORI, H., 2016. Särkyneen mielen ääni kuultavissa Lapinlahdessa. <http://kaupunkikanava.fi/sarkyneen-mielen-aani-kuultavissa-lapinlahdessa/> [Accessed: 25th January 2017].
- KARJALAINEN, P., 2016. Lapinlahden entinen mielisairaala herää eloon – kuuntele, millaista siellä on. <http://yle.fi/uutiset/3-9113064> [Accessed: 25th January 2017].
- KAUHANEN, H., RÖNNHOLM, P. and LEHTOLA, V. V., 2016. Motorized panoramic camera mount-calibration and image capture. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3(5): 89–96.
- KAZHDAN, M. and HOPPE, H., 2013. Screened Poisson surface reconstruction. *ACM Transactions on Graphics*, 32(3): 29/1–29/13.
- KERSTEN, T. P., PRZYBILLA, H.-J., LINDSTAEDT, M., TSCHIRSCHWITZ, F. and MISGAISKI-HASS, M., 2016. Comparative geometrical investigations of hand-held scanning systems. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41(B5): 507–514.
- KHOSHELHAM, K. and ELBERINK, S. O., 2012. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors*, 12(2): 1437–1454.
- KIM, Y. M., CHAN, D., THEOBALT, C. and THRUN, S., 2008. Design and calibration of a multi-view TOF sensor fusion system. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Pages 1–7.
- LEE, S. Y., MAJID, Z. and SETAN, H., 2013. 3D data acquisition for indoor assets using terrestrial laser scanning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(2/W1): 221–226.
- LEHTOLA, V. V., KURKELA, M. and HYYPÄ, H., 2014. Automated image-based reconstruction of building interiors – a case study. *Photogrammetric Journal of Finland*, 24(1): 1–13.
- LEHTOLA, V. V., VIRTANEN, J.-P., KUKKO, A., KAARTINEN, H. and HYYPÄ, H., 2015. Localization of mobile laser scanner using classical mechanics. *ISPRS Journal of Photogrammetry and Remote Sensing*, 99: 25–29.
- LORENSEN, W. E. and CLINE, H. E., 1987. Marching cubes: a high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4): 163–169.
- LOWE, D. G., 1999. Object recognition from local scale-invariant features. *Seventh IEEE International Conference on Computer Vision*, 2: 1150–1157.
- MANYOKY, M., WISSEN HAYEK, U., HEUTSCHI, K., PIEREN, R. and GRÊT-REGAMEY, A., 2014. Developing a GIS-based visual-acoustic 3D simulation for wind farm assessment. *ISPRS International Journal of Geo-Information*, 3(1): 29–48.
- MATTERPORT, 2017a. [https://support.matterport.com/hc/en-us/article\\_attachments/115010723247/Guide\\_-\\_Matterport\\_for\\_Construction\\_Documentation.pdf](https://support.matterport.com/hc/en-us/article_attachments/115010723247/Guide_-_Matterport_for_Construction_Documentation.pdf) [Accessed: 12th July 2017].
- MATTERPORT, 2017b. <https://matterport.com/pro2-3d-camera/> [Accessed: 17th July 2017].
- MAZURYK, T. and GERVAUTZ, M., 1996. *Virtual Reality – History, Applications, Technology and Future*. Technical report, Technische Universität Wien, Vienna, Austria. 72 pages.
- MCGANN, J. P., 2015. Associative learning and sensory neuroplasticity: how does it happen and what is it good for? *Learning & Memory*, 22(11): 567–576.
- OGLEBY, C. L., 1999. From rubble to virtual reality: photogrammetry and the virtual world of ancient Ayutthaya. Thailand. *Photogrammetric Record*, 16(94): 651–670.
- PAGANI, A. and STRICKER, D., 2011. Structure from motion using full spherical panoramic cameras. *IEEE International Conference on Computer Vision Workshops*, Barcelona, Spain. Pages 375–382.
- PARISI, T., 2015. *Learning Virtual Reality: Developing Immersive Experiences and Applications for Desktop, Web, and Mobile*. O'Reilly Media, Sebastopol, California, USA. 172 pages.
- PORTALÉS, C., LERMA, J. L. and PÉREZ, C., 2009. Photogrammetry and augmented reality for cultural heritage applications. *Photogrammetric Record*, 24(128): 316–331.



- REMONDINO, F., RIZZI, A., GIRARDI, S., PETTI, F. M. and AVANZINI, M., 2010. 3D Ichnology – recovering digital 3D models of dinosaur footprints. *Photogrammetric Record*, 25(131): 266–282.
- ROTHBERG, I., 2016. Hör Lappvikens roster. <https://www.hbl.fi/artikel/hor-lappvikens-roster-2/> [Accessed: 25th January 2017].
- ROUSSOU, M., 2001. The interplay between *form*, *story*, and *history*: the use of narrative in cultural and educational virtual reality. *International Conference on Virtual Storytelling*, Avignon, France. Pages 181–190.
- SCHMIDT, N., SCHÜTZE, R. and BOOCHS, F., 2011. 3D-Sutra – interactive analysis tool for a web atlas of scanned Sutra inscriptions in China. *Photogrammetric Record*, 26(136): 488–505.
- SEQUOIA, 2017. <http://sequoia.thinkboxsoftware.com/> [Accessed: 10th July 2017].
- SHAO, Z., LI, C., ZHONG, S., LIU, B., JIANG, H. and WEN, X., 2015. 3D reconstruction of cultural tourism attractions from indoor to outdoor based on portable four-camera stereo vision system. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(4/W5): 193–197.
- SHIRATUDDIN, M. F. and THABET, W., 2011. Utilizing a 3D game engine to develop a virtual design review system. *Journal of Information Technology in Construction*, 16: 39–68.
- STRUCTURE, 2017. <https://structure.io/> [Accessed: 4th July 2017].
- TANGO, 2017. <https://get.google.com/tango/> [Accessed: 4th July 2017].
- THOMSON, C., APOSTOLOPOULOS, G., BACKES, D. and BOEHM, J., 2013. Mobile laser scanning for indoor modelling. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(5/W2): 289–293.
- TRENHOLME, D. and SMITH, S. P., 2008. Computer game engines for developing first-person virtual environments. *Virtual Reality*, 12(3): 181–187.
- UNITY, 2017a. *Unity*. <https://unity3d.com/> [Accessed: 10th July 2017].
- UNITY, 2017b. *Unity Documentation. Art Asset Best Practice Guide*. <https://docs.unity3d.com/Manual/HOWTO-ArtAssetBestPracticeGuide.html> [Accessed: 10th July 2017].
- UNITY, 2017c. *Unity Documentation. Practical Guide to Optimization for Mobiles*. <https://docs.unity3d.com/Manual/MobileOptimizationPracticalGuide.html> [Accessed: 10th July 2017].
- UNITY, 2017d. *Unity. Point Cloud Viewer and Tools*. <https://www.assetstore.unity3d.com/en#!/content/16019> [Accessed: 17th July 2017].
- VIRTANEN, J.-P., HYYPPÄ, H., KÄMÄRÄINEN, A., HOLLSTRÖM, T., VASTARANTA, M. and HYYPPÄ, J., 2015. Intelligent open data 3D maps in a collaborative virtual world. *ISPRS International Journal of Geo-Information*, 4(2): 837–857.
- WENDT, A. and FRASER, C., 2007. Image-based modelling of architectural sites: the *iWitness* approach. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5/W47): 5 pages.
- XIAO, J. and FURUKAWA, Y., 2012. Reconstructing the world's museums. *Twelfth European Conference on Computer Vision, Florence, Italy*, 1: 668–681.
- ZHU, L., HYYPPÄ, J., KUKKO, A., KAARTINEN, H. and CHEN, R., 2011. Photorealistic building reconstruction from mobile laser scanning data. *Remote Sensing*, 3(7): 1406–1426.

## Résumé

En combinant les moteurs de jeu avec la cartographie d'intérieur, il est possible de créer des environnements virtuels interactifs qui représentent le monde réel. Dans cet article, une version virtuelle d'une installation audio dans un bâtiment historique est produite, où l'utilisateur explore librement le bâtiment et peut écouter un ensemble de clips audio, créant une pièce radiophonique virtuelle. Un système de cartographie d'intérieur basé sur une caméra de profondeur, Matterport, a capturé un ensemble de scènes d'intérieur. Le but était d'évaluer la qualité et la facilité d'utilisation des modèles d'intérieur et de montrer qu'on pouvait les utiliser dans le développement d'applications 3D. La qualité des modèles a été évaluée par comparaison avec un balayage laser, mettant en évidence les limites du système Matterport: l'augmentation de l'étendue de la zone relevée dégrade le niveau de détail et la précision des modèles volumétriques. La qualité des textures n'est pas affectée par cette limitation ce qui permet de préserver l'apparence des modèles. Afin d'obtenir des modèles optimisés pour les applications 3D mobiles, une révision manuelle a été mise en œuvre.

## Zusammenfassung

Durch die Kombination von Spielekonsolen mit Indoor-Mapping, wird es möglich, interaktive virtuelle Welten zu erstellen, die die reale Welt repräsentieren. In diesem Beitrag wird eine virtuelle Version einer Audio-Installation in einem historischen Gebäude erstellt, die es dem Nutzer erlaubt, das Gebäude frei zu erkunden und einen Satz von Audio-Clips zu hören. Mit dem kamerabasierten Indoor-Mapping System Matterport wurde ein Satz von Innenräumen erfasst. Das Ziel war es, die Qualität der Modelle der Innenräume nachzuweisen und ihre Anwendbarkeit für die Entwicklung der 3D Anwendungen aufzuzeigen. Die Qualität der Raummodelle wurde durch einen Vergleich mit Laserscanning evaluiert, wobei die Grenzen von Matterport aufgezeigt wurden: bei der Aufnahme von größeren Messbereichen wird der Detaillierungsgrad und die Genauigkeit der volumetrischen Modelle reduziert. Die Qualität der Texturen wurde jedoch nicht beeinflusst. Für optimierte Modelle für mobile 3D Anwendungen ist eine manuelle Bearbeitung unerlässlich.

## Resumen

Combinando los motores de juegos con técnicas de cartografía de interior, es posible crear entornos virtuales interactivos que representen el mundo real. En este documento se produce la versión virtual de una instalación de audio en un edificio histórico, donde el usuario explora libremente el edificio y experimenta un conjunto de clips de audio, creando una reproducción virtual de radio. Un sistema de cartografía de interior con una cámara de profundidad, Matterport, captó un conjunto de interiores por etapas. El objetivo fue evaluar la calidad y la usabilidad de los modelos de interior y demostrar su aplicación en el desarrollo de aplicaciones 3D. La calidad de los modelos se evaluó en comparación con el escaneo láser, revelando las limitaciones del sistema Matterport: al aumentar el tamaño del área medida se reduce el nivel de detalle y la precisión de los modelos volumétricos. La calidad de las texturas no se vio afectada por esta limitación, manteniendo la apariencia de los modelos. En la optimización de modelos para aplicaciones móviles en 3D, los modelos se revisaron manualmente.

## 摘要

将游戏引擎与室内测图相结合,可以创建代表真实世界的交互式虚拟环境。本文制作了一个历史建筑的音频设备的虚拟版本,使用者可以自由地探索该建筑并体验一组音频,创建一个虚拟的音频播放环境。采用一个基于深度相机的室内测图系统, Matterport, 来获取一系列的阶段式内部信息。其目的是为评估由此方式生成的室内模型之质量和可用性,并展示在三维案例中的应用。通过与经由激光扫描获取的模型进行比较,显示由 Matterport 获取的模型质量的有限性。当测量区域增大时,三维模型的细节和精度随之降低。但是模型纹理的质量不受此限制的影响,依然保持模型的外观。为了获得移动三维应用的优化模型,本研究对模型进行了手工编辑。