OCCLUSION SCREENING USING 3D CITY MODELS AS A REFERENCE DATABASE FOR MOBILE AR-APPLICATIONS

M. Alfakhori*, H. Dastageeri, S. Schneider, V. Coors

Centre for Geodesy and Geoinformatics, Stuttgart University of Applied Sciences (HFT Stuttgart), Stuttgart, Germany - (muhammad.alfakhori, habiburrahman.dastageeri, sven.schneider, volker.coors)@hft-stuttgart.de

Commission IV, WG IV/9

KEY WORDS: Augmented Reality, Mixed Reality, Occlusion, Urban Planning, e-Participation, HoloLens, 3D City Model, CityGML.

ABSTRACT:

Creating an immersive Augmented Reality (AR) experience requires aligning the digital content with the real environment so that the digital content appears and interacts in a similar way to a real object. For that, the virtual object must persist in its position across sessions and be occluded completely or partly, depending on whether a real-world object is in the line of sight. This research makes use of the cutting-edge technology available in the field of AR/MR, the Microsoft HoloLens 2.0, where the in-device Time of Flight (ToF) camera is used to scan the environment to create a Spatial Map. The CityGML LOD1 model enriches the Spatial Map and, in addition, is used as an occlusion mask via a custom rendering pipeline. Spatial Anchors can be tracked using the Simultaneous Localization And Mapping (SLAM) technique, used in conjunction with World Locking Tools to anchor the whole scene to the real-world coordinates. The developed occlusion effect is used in urban planning scenarios to introduce a new area design featuring a car-free environment. The proposed method is evaluated based on performance indicators, including frames per second (FPS) since it is highly correlated to user comfort. The findings of the users' study demonstrate that the occlusion effect achieves its purpose since most of the participants reported enhancements in the depth perception and overall experience by enabling the occlusion screening.

1. INTRODUCTION

Augmented Reality (AR) refers to the science and technology to overlay digital content over the real environment. In other words, AR allows users to view materials and objects that they would otherwise be unable to see (Kasperi et al., 2017). AR can be extended further to have a more interactive and immersive experience by introducing an interaction between digital content and the real-world environment to achieve a Mixed Reality (MR) and near-real experience. This necessitates a seamless transition between virtual and physical environments, as well as the ability for both to coexist and interact realistically. Simulating physical interactions like collisions, shadows, lighting, and occlusions is a difficult issue to overcome (Kasperi et al., 2017, Fortin and Hebert, 2006). For that, the digital content should align with the real environment including position, scale, and occlusion by other objects.

When building AR scenes, the purpose of occlusion is to maintain the laws of the line of sight. That is, any virtual item that is placed behind a real object should be occluded or hidden behind the actual object to provide the viewer with a realistic experience and improve the viewer's depth perception. As shown in Figure 1, when no occlusion effect is applied, the digital content seems to be floating above the scene; however, when the occlusion effect is applied and part of the objects is hidden, the scene appears more realistic, and the digital content blends with the real environment. To achieve occlusion, knowledge of the real environment is required through the presence of a three-dimensional representation of it, which in turn will act as a mask to hide the virtual objects. This real-world representation can be obtained in real-time using various 3D sensing techniques or can be obtained by using a generalized model in the form of a city model, which can be obtained in advance.

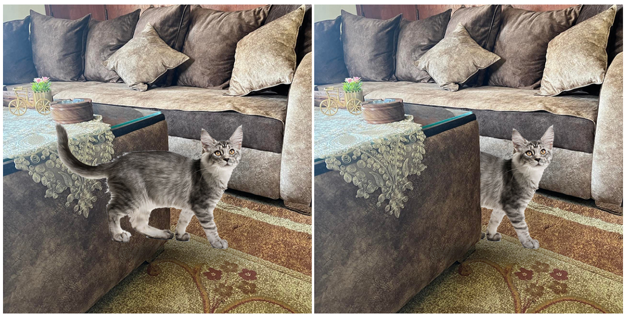

Figure 1. Demonstration of the occlusion effect. (*left*) Occlusion disabled. Occlusion enabled (*right*).

With the rapid growth of AR technologies and their numerous use cases including architecture and urban planning, it is essential that novel inventive methods be introduced, as well as improvements in performance metrics and their simple accessibility. Since urban planning plays an important role in our lives, new technologies such as AR have the potential to make significant contributions to this field of study. Increasing the use of qualitative AR solutions is essential to break free from their current limited range of applications, which are primarily focused on two-dimensional data visualization or marker-based systems (Cirulis and Brigmanis, 2013).

The approach of this research is part of the project Mobility4iCity[1], which aims to contribute promoting sustainable mo-

* Corresponding author

[1] https://www.hft-stuttgart.de/forschung/projekte/aktuell/mobility4icity

bility, qualification of urban space, and protection against low-frequency noise in the Stuttgart region, as well as developing preliminary concepts for each of these goals. For the purposes of this study, the neighborhood of Leonhardsviertel in the city of Stuttgart, Germany is used as a study case, as it has relatively high urban density and a scarcity of outdoor green space. Where the AR application is used to show local residents how the new design of the area will look and get their feedback, which can then be used in the decision-making process at a later stage.

1.1 Motivation and Objective

The majority of occlusion screening research is focused on achieving pixel-perfect results. For this, studies either use real-time 3D sensing technology or a pre-existing model of the environment. Neither of the research was able to effectively integrate the two methodologies in its findings. Additionally, a prior study focused on the use of sensor-based tracking systems, which were found to be insufficiently accurate (Kasperi et al., 2017), which might have an impact on the entire AR experience. This research intends to make use of the most up-to-date state-of-the-art technology available in the field of AR/MR represented by the Microsoft HoloLens 2.0. Since the device does not have a GPS tracking system, it can only handle local coordinates, which implies that a computer-vision approach must be used to anchor the digital content to its intended location. Furthermore, the HoloLens 2.0 uses a Time of Flight (ToF) camera to scan the surrounding environment, which may be utilized to create an occlusion screening effect in some cases. The range of the ToF camera is restricted to five meters, which implies that although it performs well in indoor scenarios, it is unsuitable for outdoor application, particularly in a scenario such as urban planning. The Spatial Map, a 3D reconstruction of real-world surfaces in the environment that is scanned using the in-device ToF camera, would need to be further enhanced with a 3D city model to have a sufficient occlusion effect.

In response to creating a near real-world and immersive experience for the user, the main objective of this research is two-fold: Firstly, the occlusion issue is addressed, which occurs when virtual content is hidden behind a real object, by enriching the in-device scanned Spatial Map with a 3D building model. Secondly, placing digital content in the intended location and maintaining its position across multiple application sessions. And finally incorporating the above into an urban planning use-case such that an AR experience is created that reflects a car-free environment.

This research aims to answer the following questions:

- *How to anchor digital content to the real-world coordinates system in the study area?*
- *How to use in-device scanning capability for dynamic object occlusion?*
- *Is it possible to enrich the scanned and reconstructed spatial mesh map with a 3D building model to improve the outdoor occlusion effect?*

2. STATE OF THE ART

2.1 Occlusion Handling in AR

Occlusion is a highly effective 3D cue. It is, in fact, the most fundamental depth indicator (Kress, 2020). Therefore, to be able to achieve a seamless transition between virtual content and the real-world environment, physical simulation is needed (Fortin and Hebert, 2006). (Fortin and Hebert, 2006) presented two approaches for real-time occlusion handling. The first approach is model-based, where wrapping actual items in an approximately bounding box is the concept. For that reason, the tracking camera is used to locate the real object target. After that, a segmentation mask is generated by subtracting the background, which makes it suitable for a static viewpoint. In the second approach, the performance camera is replaced by a pair of stereo cameras to generate a depth map. Then, the depth is used for occlusion screening regarding the view-point. Additionally, the depth-based technique eliminates the need for a predefined model and enables the ability to occlude unknown objects.

(Kasperi et al., 2017) introduce a new method for occlusion screening using Open Street Map (OSM) data for outdoor usage, where they stressed on the need for 3D representation of the real-world environment. The model can be created manually, defined by the user, or reconstructed by using depth sensors. This method removes the target regions from the virtual building using masks constructed from OSM buildings' data. Due to missing height information in OSM building data and inaccuracy in the position due to GPS sensor errors, the obtained results were not pixel-perfect and full occlusion was not achieved.

More depth-sensor-oriented research highlighted the use of ToF cameras and LiDAR technology. (Fischer et al., 2007) use a ToF camera to generate a 2D depth map, which is used as an occlusion handler. However, Fisher's results highlighted some potential flaws since the data retrieved from the ToF camera often contains noise. (Behzadan and Kamat, 2010) use LiDAR technology to overcome incorrect occlusion handling.

2.2 Using the 3D City Model

The development of computational capabilities and the increase in the number of 3D city models facilitated their integration into AR technology. (Lee et al., 2012) provided an AR tracking testbed based on a 3D city model. Their research focused on overcoming the GPS sensor drift over a short period, especially in built-up areas. For that, a computer vision approach was followed, where Speeded-Up Robust Features (SURF) were used to detect distinct features. A transformation matrix can be calculated based on the coordinated 3D feature points extracted from the 3D city model and the 2D feature points from the live view video. Then, the virtual content can be overlaid onto the real-world environment. (Afrooz et al., 2018) investigated the usability of different LODs building models and virtual reality (VR) in order to determine the most appropriate one based on the intended usage.

2.3 AR for Urban Planning

To increase the degree of immersion in the urban planning solution and to enable urban planning professionals to travel about city streets while projecting virtual 3D buildings, enabling them to view both the actual city and the virtual structures concurrently, different AR applications have been implemented. (Allen et al., 2011) have shown that interactive solutions can greatly improve public participation in urban planning processes. An app for smartphones shows the proposed architecture design overlaid on the existing real-world environment. It allowed users to view the proposed 3D architectural model and provide feedback. As a result, the Graphical User Interface (GUI) needed to be as simple and user-friendly as possible. Smartphone familiarity directly influenced the users' experience.

The City 3D-AR pilot project uses GPS longitude and latitude data to place 3D objects in the real world. Its main challenges are representing large 3D objects in an outdoor environment and detecting building locations based on participant distance and viewing angle. The traditional fiducial marker technique is only suitable for indoor solutions over short distances (Cirulis and Brigmanis, 2013). A processing and rendering laptop connected to a USB GPS sensor was used for this. The output was viewed on a Vuzix Wrap 920AR AR Head-Mounted Display (HMD). The viewer location is calculated continuously to support viewer movement and automated 3D object transformations, rotations, and scaling in response to a view's changed angle and distance. To create an interactive AR experience, a database of 3D buildings is added, allowing the viewer to change architectural models. CityViewAR was created after the 2011 Christchurch earthquakes to visualize panorama photographs of the city (Lee et al., 2012). The app was made to aid in the rebuilding of destroyed structures and future city development. A GPS-enabled smartphone is required to navigate the panoramas. The panoramic images rotate as the viewer rotates the device's gyroscope sensor.

3. METHOD

This research puts forward the idea of using the 3D building model as a custom shader to enrich the Spatial Map and create the occlusion effect. The geometry of the scanned Spatial Map which was enriched further by the geometry of the 3D building model, was used as an occlusion mask as shown in Figure 2.

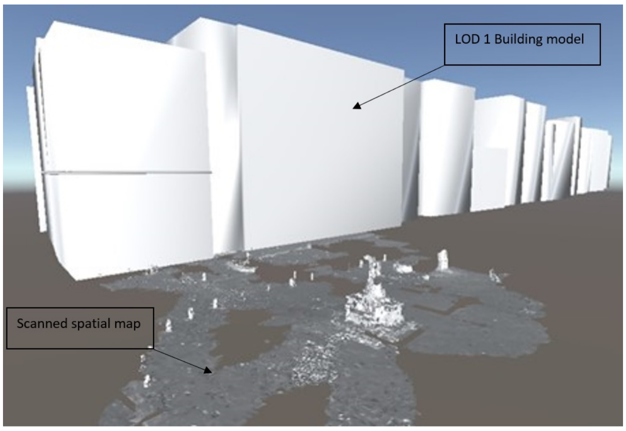

Figure 2. Occlusion mask based on the geometry of the scanned Spatial Map and 3D building model.

3.1 Used Resources

The AR application was developed using Unity, which is a cross-platform game engine capable of developing 3D and 2D games, as well as interactive simulations and other experiences [2]. Mixed Reality Toolkit (MRTK) provided a cross-platform input system, foundational components, and common building blocks for spatial interactions, for which it was used as an AR Software Development Kit (SDK). MRTK also leverages access to different platform capabilities, including hand tracking and gaze. MRTK provided direct access to Spatial Anchor and World Looking Tool, both of which were used to place and maintain the location of the augmented content. MRTK version 2.7.2 was used to develop the application.

Microsoft HoloLens 2.0 illustrated in Figure 3 is a wearable AR HMD device that shows interactive 3D holograms. The HoloLens has six front cameras: four grayscale and two stereo pairs. The ToF camera has two modes: long-throw for mapping the environment (five meters) and short-throw for hand tracking (one meter). It captures images and videos. Because the HoloLens lacks a GPS or GNSS sensor, it relies solely on computer vision to comprehend its surroundings and determine its location.

Figure 3. HoloLens 2.0 camera configuration. (*yellow*) Grayscale tracking cameras. (*Red*) ToF camera. (*Blue*) RGB camera.

3.2 Data Preparation

The available real-world representation consists of a 3D building model in the CityGML format. CityGML is an Open Geospatial Consortium (OGC) open-source standard format for storing and exchanging 3D city models. In CityGML version 2.0, the buildings are divided into multi-scale semantic representations based on Level of Detail (LOD), which refers to the model's resemblance to its real-world version and has consequences for the model's usability (Biljecki et al., 2016). The LOD1 model is provided for the study area and is sufficient to function as an occlusion mask since the observer is looking from the ground and no roof information is required (Kasperi et al., 2017, Li and Fan, 2014). The deployed model includes the whole city of Stuttgart and is provided in CityGML2.0 as XML format, thus requiring further data processing to use inside the game engine. The FME workbench was used to extract the buildings in the study area using an attribute filter based on the block number, and then an OBJ writer was used to generate a mesh output. The building model is constructed using UTM projected coordinates, with the Z value representing the height of the building. Intending to use the building model in the game engine, the OBJ model was rotated and moved to a local coordinate system with an origin at 0,0,0 as seen in Figure 4.

3.3 Occlusion Shader

The occlusion shader calculates and determines how each pixel is rendered on the screen. The shader is wrapped in a material that can be assigned to the 3D building model to act as an

[2] https://docs.unity3d.com/Manual/index.html

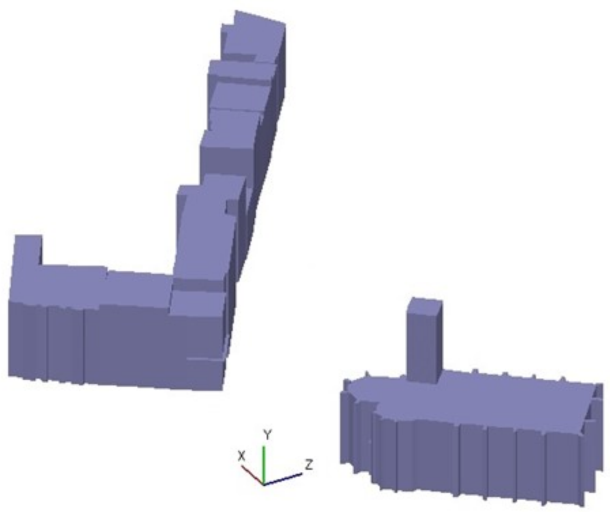

Figure 4. 3D mesh building model of the study area.

occlusion mask. To render the occlusion mask before all other meshes in the scene, a Queue value is set to one unit lower than the geometry value where the geometry value is always 2000 and represents the priority to render solid objects. ZWrite is enabled to write the pixels from the occlusion mask to the depth buffer. The ZTest is set to render the geometry at the front and not draw any pixels behind it from other objects. In other words, all objects behind the occlusion mask will not be rendered. Moreover, the ColorMask is set to zero, enabling the shader to modify all three RGB values and the alpha channel as well, which represents transparency.

```
1 Shader "Occlusion"
2 {
3    SubShader
4    {
5        Tags { "Queue"="Geometry-1" }
6        ZWrite On
7        ZTest LEqual
8        ColorMask 0
```

Lines 15-31 define the needed structure to hold the input and output position information. The POSITION parameter is used to pass the 3D coordinates of each pixel, which are stored in 32bit, whereas SV POSITION stores the object screen coordinates in addition to the Z value returned by the ZTest.

```
15          CGPROGRAM
16          #pragma vertex vert
17          #pragma fragment frag
18
19          #include "UnityCG.cginc"
20
21          struct appdata
22          {
23              float4 vertex : POSITION;
24              UNITY_VERTEX_INPUT_INSTANCE_ID
25          };
26
27          struct v2f
28          {
29              float4 position : SV_POSITION;
30              UNITY_VERTEX_OUTPUT_STEREO
31          };
```

Lines 33-38 represent the main function used to convert vertex object-space 3D coordinates to screen 2D coordinates, then the function fixed4 in lines 40-42 returns a black color value for each vertex when it's rendered on screen. The HoloLens uses additive color mixing. As a result, the color black appears transparent. However, the geometry of the building model which acts as an occlusion mask is drawn in the scene, any object hidden behind it will be occluded.

```
33          v2f vert (appdata input)
34              {v2f output;
35              UNITY_SETUP_INSTANCE_ID(input);
36
        UNITY_INITIALIZE_VERTEX_OUTPUT_STEREO(output);
37              output.position =
        UnityObjectToClipPos(input.vertex);
38              return output;}
39
40          fixed4 frag (v2f input) : SV_Target
41          {return fixed4(0.0, 0.0, 0.0, 0.0);}
42          ENDCG
43      }
44  }
```

To control the occlusion effect based on the Spatial Map of the 3D building model, a user menu was added, where the user can enable and disable the occlusion screening as well as visualize the occlusion mask in visible colors.

3.4 Anchoring

To render the 3D content to its desired location, Spatial Anchors which is defined as a keypoint in the environment where the system tracks over time, were used and automatically distributed in real-time. For that, World Locking Tools were added to the Unity scene and the main camera game object was stored in it. The World Locking Tools script initiates the World Locking Manager, which in turn takes control of placing the spatial anchors and performing the camera transformation needed. Scale inaccuracy is another important factor.

3.5 Evaluation Criteria

The visual effect of the occlusion screening was shown via screenshots, and the occlusion effect's impact was evaluated through user feedback. Furthermore, the suggested method was assessed using a set of performance indicators. Since pixel-perfect occlusion is not necessary to acquire the intended occlusion effect (Kasperi et al., 2017), no statistical analysis was performed.

3.5.1 Performance: The proposed method's performance was evaluated using the application's FPS. The greater the FPS, the less frequent the user blinks, which reduces stress. In low FPS rendering, judder can cause screen tearing when irregular frames are produced in the same view. The tracking system

Figure 5. Alignment of the 3D building model occlusion mask with real environment. Rendered in blue color for better visualization.

was assessed on its ability to consistently and reliably localize the user's location and maintain the position of the augmented content independent of the user's movement. However, a drift of augmented content location might happen if it is placed outside of the mapped environment or far away from the Spatial Anchor. While the whole scene is anchored and not an individual object, this problem is not expected to happen. Jitter can be seen by users as a high-frequency movement of augmented content, which may occur as environment tracking fails. The HoloLens has a built-in hologram stability pipeline that gives good results in maintaining the stability of the augmented content (Vassallo et al., 2017). An eye calibration is necessary to avoid the swim effect where the whole augmented content appears floating while the user moves his head.

3.5.2 Users' Study: A user study was conducted to gather feedback regarding the augmented content. A demo session for users begins with a brief introduction to the research and its overall goal so that the users are aware of what they are looking for in terms of evaluation and may prepare accordingly. Following that, the participants were free to move around the study area, where they can enable and disable various occlusion options as they pleased. The users were given a questionnaire that consisted of five sections and a total of twelve questions. Each questions has five options based on(Likert, 1932), were the answers are ranging on a scale from 1 to 5, with 1 indicating "Strongly Disagree" and 5 representing "Strongly Agree".

4. RESULTS

4.1 Tracking System and Anchoring

Using World Locking Tools does not require a pre-scan as it simultaneously maps the environment and localizes the device using the in-device Spatial Map. Additionally, the area was not bounded, as new Spatial Anchors were added as the user moved. The localization of the device position and the 3D scene coordinate transformation was done in real-time. The linear deviation represents the distance between the currently detected Spatial Anchors in the current session and the reference points corresponding to them. A lower value indicates finer alignment. The angular deviation indicates the angle difference between the user location in comparison to the Spatial Anchors in both current and reference session (see Table 1).

Table 1. Comparison of deviation across sessions using different amount of Spatial Anchors.

Test Session	Session 1	Session 2	Session 3
No. Spatial Anchors	63	127	271
Max Linear Deviation	34 mm	33 mm	22 mm
Max Angular Deviation	0.6830 °	0.3323 °	0.2807 °

4.2 Occlusion Screening

The developed occlusion screening's visual effect based on the in-device scanned Spatial Map is shown in Figure 6 , wherein the left side shows how the 3D design appears when the occlusion effect was disabled. The grass and the grass hedge appear in front of the statue and the tree even though they are supposed to be in the line-of-sight between the observer and the augmented content. In comparison to the right side where the occlusion effect was enabled, the 3D design was occluded based on the shape of the scanned real-environment objects. Nevertheless,

Figure 6. Occlusion screening based on the Spatial Map. (*left*) Disabled. (*right*) Enabled.

the occlusion mask based on the Spatial Map had flaws where it did not align perfectly around the edges. Also, some parts were missing or were not scanned, which led to an incomplete occlusion effect. Some noises resulting from the incomplete scanning of the environment have been observed, particularly while pedestrians pass in front of the observer. Even after the pedestrians were no longer visible in the scene, it took time for the Spatial Map to update and exclude their masks.

Figure 7 illustrates the 3D building model that was used as an occlusion mask. The left side shows how the 3D design appeared floating over the real building when the occlusion is disabled. While the right side represents how the scene looked when the occlusion screening was enabled. The occlusion mask aligns with the real-life equivalent and the augmented content occluded along the edge of the building. The occlusion mask appeared sharp around the edges since the geometry of the buildings was represented with straight meshes.

Figure 7. Occlusion screening based on the 3D building model. (*left*) Disabled. (*right*) Enabled.

Figure 8 provides a comparison of the occlusion screening effect based on the Spatial Map scanned by the device on the left and by using the 3D building model as an occlusion mask on the right side. It's noteworthy that HoloLens could not scan the glass facade of the building, resulting in an incomplete occlusion effect. Furthermore, if the observer stands outside of the scanning range, which is five meters, the occlusion based on the Spatial Map will not be effective. In opposition, using the Spatial Map as an occlusion mask, the augmented content was occluded on top of the tables beside the building.

Figure 8. Comparison between occlusion screening based on (*left*) the Spatial Map and (*right*) the 3D building model.

4.3 Performance Evaluation

Two distinct 3D designs were used in the performance test, which focused on determining the FPS throughout five minutes of continuous application use. The first design represents the final implementation and consists of around 1.84 million triangles, most of which come from the detailed design of the parking garage building and trees. The second design replaces the grass texture shader with a 3D detailed grass model, which adds another 3.8 million triangles, bringing the total to 5.24 million triangles in total. The occlusion mask based on the geometry of the building model added around a thousand triangles, which makes them negligible in comparison to the total number. Figure 9 shows the number of rendered frames over the test duration. Using the first design, the application was able to render at 60 FPS on average with 1-2 drop frames. On the other hand, using the second design, the application is rendered at 34.2 FPS with a maximum of 45 FPS and a minimum of 20 FPS. As the dropping of the frames occurred in conjunction with the movement of the user or the rotation of his head and the change of the rendered scene.

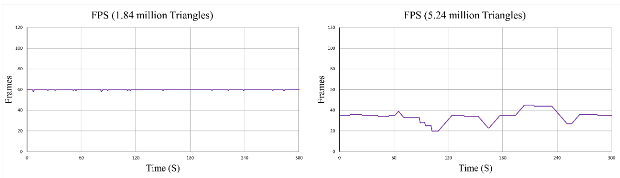

Figure 9. Relation between the number of triangles and rendered FPS.

4.4 Users' Study Feedback

The users' opinions were gathered through a demo session followed by a questionnaire, which is summarized in Figure 10. The experiment enrolled 14 participants from a variety of backgrounds and age groups. The first section elicited responses regarding participants' prior knowledge, with three responding with "excellent" knowledge of AR and two responding with "very good" knowledge. While most participants responded with "sufficient" prior knowledge, the remainder of the responses fell into the "fair" and "poor" categories. The average response time for this question was 3.14, and the median response was 3, which in turn reflects the diversity of the sample of the study. When asked how familiar they were with see-through HMDs, only two participants responded "excellent", while the remaining responses were evenly distributed among the other options, resulting in an average of 2.86 and a median of 3.

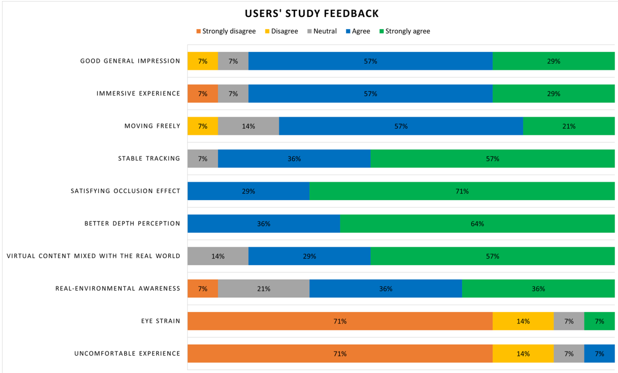

Figure 10. Summary of the users' study feedback.

Following that,Then the participants were asked about their overall impression of the developed application. Four participants "strongly agreed" that it was a positive impression, four participants "agreed," and only one participant were "neutral" and one "disagree" as well. The average and median answers to this statement were, respectively, 4.07 and 4, which means the majority of participants were positive about the developed application. When participants were asked if the augmented reality experience was immersive, four responded "strongly agree" and eight responded "agree". Only two respondents indicated that they were "neutral" or "disagree". Both the average and median answers to this statement are 4. The next statement inquired whether the participants were free to move during the experience; three responded "strongly agree" and eight responded "agree". Only two of the participants were "neutral", and one was "disagree". The average answer for this statement was 3.93 and the median was 4. The following statement was about the tracking system, in which participants were asked if the augmented content remained stable in its position throughout the experience. Eight participants "strongly agree" and five participants "agree", but only one participant was "neutral". The average and median responses to this statement were respectively 4.5 and 5. The occlusion effect was the main topic of the fourth section of the questionnaire. To begin, participants were required to state that the occlusion effect was sufficient for this application. The response to this statement was remarkable, with ten "strongly agree" and the remaining "agree", resulting in an average of 4.71 and a median of 5. Then they asked if the occlusion effect enabled them to have a better depth perception were nine answered "strongly agree" and five answered "agree". The average and median answers were 4.64 and 5 respectively. The final section was related to the participants' safety, with the first statement inquiring whether they were aware of the real environment during the experience. Five participants responded "strongly agree", five participants responded "agree", three participants responded "neutral," and only one participant responded "strongly disagree". Both the average and median answers for this statement were 4. When participants were asked if they experienced any eye strain during or after the experience, ten responded "strongly disagree", two responded "disagree", and only two responded "neutral" or "strongly agree". The average response was 1.57, while the median response was 1. The last statement was concerned if the participants had any uncomfortable feelings, ten responded "strongly disagree", two responded "disagree", and only two responded "neutral" or "agree". The average response was 1.50, while the median response was 1.

5. DISCUSSION

The results show that by using Spatial Anchors in conjunction with World Locking Tools, it is possible to anchor the augmented content in the intended location, which would then answer the research question. *How to anchor digital content to the real-world coordinate system in the study area?* The results also indicate the reliability of the visual tracking system and Inertial Measurement Unit (IMU) sensors in maintaining a stable tracking of the user position throughout the usage of the application, which is almost indistinguishable from the results of (Hübner et al., 2020). The responses from the small study group also support this. The maximum linear deviation does not change when comparing different test scenarios with different numbers of Spatial Anchors. The angular deviation increases with the number of Spatial Anchors used. Two factors contribute to the speed and robustness of localization using Spatial Anchors: the amount of data used from different sensors and the advantage of using the dedicated processor for the SLAM algorithm, which is aligned with previous findings (Cirulis et al., 2020). Due to the tracking system's reliance on natural features, which may become unavailable due to changes in the scene, such as snow covering the ground or falling leaves, environmental changes may impact the tracking process.

Using a ToF sensing camera in real-time to create an occlusion mask for small features in the environment, such as statues and trees, confirms previous results (Fischer et al., 2007). Due to the generated noise, the Spatial Map can only partially occlude moving objects, such as cars. This answers the second research question: *How to use in-device scanning capability for dynamic object occlusion?* However, the scan range limitation necessitates a 3D building model.

The occlusion effect based on the 3D building model was achieved and the visual effect can be seen from different observation points based on the line-of-sight between the observer and the digital content. Also, the augmented content was fully visible when there was no building on the line-of-sight. According to these findings, the study's occlusion handler is performing as intended; it mimics the impact that would have occurred if the virtual object had been a real object placed at the same position and answer the third research question: *Is it possible to enrich the scanned and reconstructed spatial mesh map with a 3D building model to improve the outdoor occlusion effect?* However, there were some issues with aligning the 3D building model with the actual building in the real environment, particularly around the edges, which indicate that the occlusion effect was not pixel-perfect. It also appears that the established occlusion screening effect gave the users a better understanding of depth, where the virtual objects seemed to mix with the real objects.

Because the occlusion effect uses simple geometry that doesn't add up to the total number of polygons, the performance evaluation found no differences when the occlusion effect was enabled. Thus, the FPS remained consistent and did not cause eye strain for the participants in the users' study, which was conducted after the conclusion of the trial. However, using a more detailed building model by employing various methods, such as structure from motion, which contains more polygons, could result in poor performance because it has been discovered that there is a direct relationship between the number of polygons and the performance of the model, which is in complete agreement with the findings (Tepper et al., 2017).

By comparing the results of the proposed solution with the previous studies, the integration of in-device scanned data and the 3D building model provided a more realistic occlusion effect, which was confirmed by the results of the users' study. All previous studies used only one method, either 3D sensing or the use of 3D environmental representation, and did not combine the two methods. Also, there were no problems with the height of the buildings because the height of the buildings is known to be the opposite of what (Kasperi et al., 2017) had to deal with when they generalized the height of the OSM data. Furthermore, the proposed method does not require any additional devices and can run in real-time as well. In contrast to prior research employing VR headsets (Afrooz et al., 2018), the majority of participants did not experience dizziness when using the AR headset.

One of the challenges in selecting a design was the restricted computer capability of the HoloLens, which prevented the incorporation of complex designs. As a result, the potential to create more realistic designs was restricted. An additional constraint was generated by the idea of the device's screen, which is a see-through screen with a limited brightness, resulting in decreased visibility in bright environments. However, in general, users' satisfaction with the application prototype was positive.

The user menu could be used to visualize multiple design concepts so that the user could see the differences and compare them. Later in the design process, urban planners and designers can assess each design's public acceptability. Seeing the planned changes will also encourage residents and increase their willingness to embrace the neighborhood redesign.

6. CONCLUSION

This study proposed the use of a 3D city model to enrich the in-device scanned Spatial Map where the geometry of both was used as an occlusion mask. Achieving an immersive and representative experience by anchoring the digital objects to their true position in the real-world environment and hiding the augmented content when it is located behind a real object. Enabling or disabling occlusion is entirely dependent on the intended application and goal of the experience. As disabling occlusion enables users to discover hidden objects or portions of buildings. The proposed method, wrapped in an application, is being used in the field of urban planning to enable the public to participate in the decision-making process, where feedback from the local community can be incorporated into the planning of future scenarios. Concerning assessing the proposed occlusion screening effect, a study was conducted on users from different backgrounds, where 14 users tested the application and evaluated the different occlusion options. Most of the participants reported that activating the occlusion option improved their depth perception and provided a more immersive experience.

7. FUTURE WORKS

Future work will focus on improving in-device spatial mapping to detect and reduce moving object noise. Also, developing a semantic scene understanding to automatically complete missing surfaces and fill the holes in the 3D objects, as well as simplification of the reconstructed surfaces to have straight and sharp edges, especially around the building corners, is a topic of further research. Furthermore, to enhance the participation experience, a virtual assistant might help users through the experiment and collect feedback. The same 3D building model might be used to simulate shadowing, air movement, and collisions in the future. Due to the rapid evolution of AR technologies and SDKs, it may be able to develop applications that provide a more realistic and intricate experience. Keeping up with rapid technological advancements is another problem for any research effort of this nature.

ACKNOWLEDGEMENTS

This work has been jointly developed in the project Mobility4iCity (Funding number: 13FH9I05IA) and INSPIRER (Funding number: 16SV8744). Both projects are supported by the German Federal Ministry of Education and Research (BMBF). The authors are responsible for the content of this publication.

The authors would like to thank Mr. Philipp Willkomm and Mr. Luca Casagrande from M.O.S.S Computer Grafik-Systeme GmbH for their support. Further the authors thank the anonymous reviewers who have helped to increase the quality of the paper.

REFERENCES

Afrooz, A., Lowe, R., Leao, S., Pettit, C., 2018. *Chapter 9: 3D and virtual reality for supporting redevelopment assessment*. 162–185.

Allen, M., Regenbrecht, H., Abbott, M., 2011. Smartphone augmented reality for public participation in urban planning. ACM Press, 11–20.

Behzadan, A. H., Kamat, V. R., 2010. Scalable Algorithm for Resolving Incorrect Occlusion in Dynamic Augmented Reality Engineering Environments. *Comput. Aided Civ. Infrastructure Eng.*, 25, 3-19.

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, 59, 25-37.

Cirulis, A., Brigmanis, K. B., 2013. 3D Outdoor Augmented Reality for Architecture and Urban Planning. *Procedia Computer Science*, 25, 71-79.

Cirulis, A., Brigmanis, K., Zvejnieks, G., 2020. Analysis of Suitable Natural Feature Computer Vision Algorithms for Augmented Reality Services. *Baltic Journal of Modern Computing*, 8.

Fischer, J., Huhle, B., Schilling, A., 2007. Using Time-of-Flight Range Data for Occlusion Handling in Augmented Reality. *IPT-EGVE 2007 - 13th Eurographics Symposium on Virtual Environments, 10th Immersive Projection Technology Workshop*, 109-116.

Fortin, P., Hebert, P., 2006. Handling occlusions in real-time augmented reality : Dealing with movable real and virtual objects. IEEE, 54–54.

Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., Wursthorn, S., 2020. Evaluation of HoloLens Tracking and Depth Sensing for Indoor Mapping Applications. *Sensors*, 20, 1021. https://www.mdpi.com/1424-8220/20/4/1021.

Kasperi, J., Edwardsson, M. P., Romero, M., 2017. Occlusion in outdoor augmented reality using geospatial building data. Part F131944, ACM, 1–10.

Kress, B. C., 2020. *Optical Architectures for Augmented-, Virtual-, and Mixed-Reality Headsets*. SPIE.

Lee, G. A., Dunser, A., Kim, S., Billinghurst, M., 2012. Cityviewar: A mobile outdoor ar application for city visualization. IEEE, 57–64.

Li, J., Fan, X., 2014. Outdoor augmented reality tracking using 3d city models and game engine. IEEE, 104–108.

Likert, R., 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 22 140, 55. https://legacy.voteview.com/pdf/Likert$_1$932.*pdf*.

Tepper, O. M., Rudy, H. L., Lefkowitz, A., Weimer, K. A., Marks, S. M., Stern, C. S., Garfein, E. S., 2017. Mixed Reality with HoloLens. *Plastic and Reconstructive Surgery*, 140, 1066-1070. http://journals.lww.com/00006534-201711000-00034.

Vassallo, R., Rankin, A., Chen, E. C. S., Peters, T. M., 2017. Hologram stability evaluation for microsoft hololens. 10136, SPIE, 1013614.