LabelAR, Immersive AR for Enhanced Spatial Exploration

Matteo Boglioni[1], Tommaso Di Mario[1], Andrea Sgobbi[1], Gabriel Tavernini[1],
Saxena Rupal[1,†] and Marcel Lancelle[2,†]
[1]*ETH Zürich* [2]*Magic Leap* *

{tdimario,mboglioni,asgobbi,gtavernini,rsaxena}@ethz.ch, mlancelle@magicleap.com

Abstract

LabelAR is a Mixed Reality application that enhances spatial exploration by overlaying virtual labels on real-world landmarks in Switzerland using the Magic Leap 2 headset. Built with Unity and leveraging Swiss geospatial datasets, LabelAR offers users an intuitive and shared AR experience in which each user can place labels on both the terrain and specific buildings. Through multiple user studies, the app demonstrated strong engagement and usability, with future improvements planned for GPS integration and automated alignment. LabelAR showcases the potential of MR technology for immersive outdoor exploration.

1. Introduction

Over the last decade, the implementation of Mixed Reality (MR) applications has experienced significant growth in both the business-applied and customer-commercial applications. We introduce LabelAR, a cutting-edge Mixed Reality application that allows the user to improve his visitation experience at various Zürich locations by simply wearing a headset. The application was developed using Unity and deployed on a Magic Leap 2 (ML2) device: a state-of-the-art headset designed to build AR apps at scale thanks to its high computing capabilities. Despite being designed specifically for indoor use, this project aims to extend and investigate ML2's capabilities and use-cases in outdoor environments: once the app is launched, the device employs the onboard spatial mapping and localization features to map and locate the user in the physical world using three RGB and one depth camera, allowing for a seamless alignment of the virtual scene. The labels are thus immediately displayed on relevant landmarks in the area around the user. LabelAR was designed to provide a shared experience, where each user can add/edit labels and share them with the community, enriching each other's experience.

*† Equal supervision

2. Related Works

Previous research on MR applications has been extensively developed for indoor environments, with examples including museum guide systems [7] and applications to semantically segment indoor scenes [5]. Many of the developed applications make use of the increased availability of building models. One of the main interests of our study, in contrast to previous works, is to explore whether the potential of AR headsets could also be exploited in outdoor settings, to highlight landmarks and the corresponding labels.

3. Methods

In this section, we present the methodologies that support LabelAR. We describe the essential components of our system, such as data preprocessing, architectures, and system integration, which ensure the app's functionality.

3.1. Swisstopo Models

The Federal Office of Topography Swisstopo is Switzerland's geoinformation centre. Their vast collection of data is freely available on their website [4].
More specifically, in our project we make use of two datasets:

- *swissBUILDINGS3D 2.0* is a vector-based collection that represents buildings as 3D models with roof geometry and overhangs. The great level of detail in all three dimensions, together with the high coverage and realistic representation of the building volumes, make this product a valuable starting point for our app. We use this data to display aligned overlaying meshes and enable long-range occlusion of landmarks.

- *swissALTI3D* is an extremely precise digital elevation model which describes the surface of Switzerland without vegetation and development. Due to the relatively limited on-device computing resources, we decided to down-sample the data to $10m$ pixels, allowing us to load larger portions of terrain. The altitude

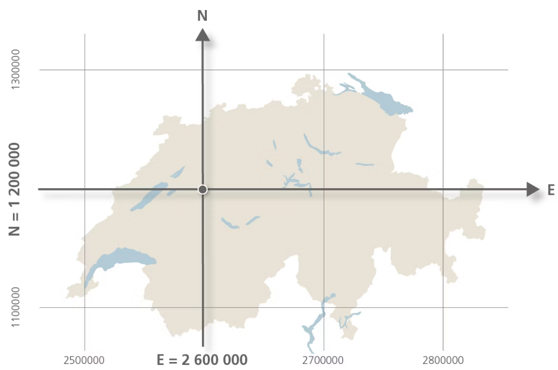

Figure 1. Zero point of the Swiss coordinate system

models allow insertion of landmarks on mountains surrounding the city of Zürich, and the resulting meshes also enable long-range occlusion through Unity.

We additionally had to develop tools to import the swisstopo data formats into Unity since none were publicly available. Our conversion algorithm is based on previous work by Yannick Gassmann [6]. Furthermore, to improve the startup performance of the app we saved all converted models into Unity assets, leading to much faster load times.

3.2. Coordinates System

All the data provided by swisstopo uses the swiss coordinate system [3] which is based on the reference frame **LV95** (the 1995 national survey), with the point of origin (or zero point) located in Bern. The coordinates of this point are not in fact 0 m / 0 m as the name might imply, but rather E = 2,600,000 m (east) and N = 1,200,000 m (north).

3.3. PostGIS

PostGIS [2] extends the capabilities of the PostgreSQL relational database by adding support for storing, indexing, and querying geospatial data. PostGIS allows spatial indexing: quickly search and retrieve spatial data based on its location. It also provides a wide range of spatial functions that allow you to filter and analyze spatial data, measuring distances and areas, intersecting geometries, buffering, and more. We store all relevant landmarks as GPS points. We then make use of this to compute the distance from the user and determine the optimal text size and visibility for each label. Since Magic Leap 2 lacks GPS sensors, LabelAR can only be used in specific locations in Zürich. The user manually chooses the location from a menu.

3.4. Magic Leap Spaces

Spaces is an application that comes preinstalled on the Magic Leap 2. It is used to create maps called Spaces using SLAM (simultaneous localization and mapping). Each Space contains localization data, spatial anchors, and a 3D mesh representing a snapshot of the physical world when the environment is scanned. Each location was registered independently allowing the device to localize into them separately and store spatial anchors needed for the alignment.

3.5. Mesh Rendering

Once the user selects a location, the system begins loading the terrain and building meshes. The terrain mesh is prioritized and loaded first, as it's much bigger, it's then followed by an asynchronous process that loads up to 400 building meshes per frame.

Given the constrained on-device computing resources, building meshes are loaded only within a 2km radius of the origin. To make the experience as smooth as possible we further restrict the radius within which meshes are interactable (selectable through the controller's pointer) to only 1km. The sole exception is the terrain mesh, which remains interactable in its entirety.

3.6. Alignment

After the user selects a location all meshes and labels are loaded into the virtual Unity world using their position relative to the selected origin, which is stored in the DB. It is then crucial to align the virtual origin with its actual position in the real world, consequently aligning all the meshes and labels with their real-world counterparts.

When entering *alignment mode* all the world's mashes are rendered as semi-transparent making it easier for the user to notice misalignments and correct them. For the actual alignment, we implemented two modes:

- *Manual*: In this mode of alignment the user is presented with a 10x10x10cm cube to which the world origin is bound, meaning all labels, building and terrain meshes will move along with it. Interacting, grabbing, and moving the cube around is possible using the controller and its pointer. It's up to the user to move and rotate the cube into a position such that the various meshes align with the real world. To facilitate this process, we tried to set the real-world origin coordinates of the various locations close to easily identifiable building meshes (e.g. on Polyterrasse we chose one of the corners of ETH's entrance stairs).

- *Triangulated*: In this alternative mode a 4-step guided process allows the user to align the virtual world easily. During each of the first three steps, a preselected building is highlighted (its mesh is rendered with a distinct color) and a white sphere marks its vertex closest to the user. The world can then be rotated around its origin using the controller to align the building (particularly the vertex). Once the three angles between the

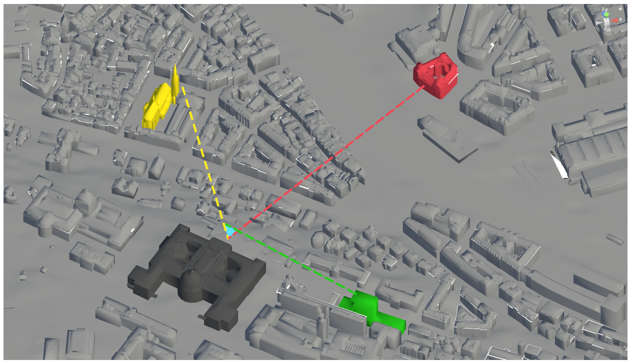

Figure 2. Computing the updated origin position during the triangulated alignment.

origin and the vertices have been collected a new origin position is computed using triangulation: the app calculates a user's position by evaluating three rays that go between the user's location and three visible landmarks. The user manually aligns the three points. The rays, which come from slightly various angles or perspectives, are extended into space to find their intersection points. Because the rays may not identify a single intersection point due to measurement noise or environmental conditions, their interceptions form a triangle. The algorithm determines the barycenter (centroid) of this triangle, which represents the user's estimated location. This method combines geometric principles with redundancy to create a reliable and accurate localization system.

The first three steps are visualized in Fig. 2, while in the fourth and final step the user can fine-tune the world's rotation and altitude before confirming.

The position and rotation of the aligned origin are then persisted to storage as a spatial anchor in the Magic Leap Space, making the alignment procedure a one-time operation. Every following restart of the application will be able to localize in the space and automatically align the world using the spatial anchor.

3.7. User Interface

The User Interface (UI) of LabelAR was developed with usability and simplicity in mind, to ensure an intuitive and enjoyable user experience. This section describes the main interfaces of the UI, namely: Scene Selection, View Settings (Main Menu), Edit Labels and Alignment. All of the UI elements are showcased in the Appendix (8).

3.7.1 Scene Selection

This interface serves as the entry point of the application, the user is prompted to choose among the available spaces. As seen in Fig. 6, this menu contains a scroll view with both the name and an image of each location. Once the user selects the desired location, the corresponding buildings and labels are loaded and aligned using the corresponding Magic Leap Space (3.4) and the saved spatial anchor.

3.7.2 Main Menu

This is the interface that the user is shown once the Menu button is clicked on the controller. From this interface the user can choose the transparency of the non-labeled meshes in their view range. This can be done using three intuitive buttons with descriptive text and icons; the meshes can be set to be either fully transparent ($\alpha = 0$), semi-transparent ($\alpha = 0.5$), or solid ($\alpha = 1$). This menu also contains two toggles:

- *Short-Range Occlusion.* The onboard depth sensor is used to generate a real-time mesh representation of the user's surroundings, enabling for example hand occlusion.
- *Real Visibility.* Leverage weather APIs to hide distant labels on low-visibility days.

Lastly, the Edit Labels and the Alignment interfaces can be opened using the corresponding buttons.

3.7.3 Edit Labels

Once the user presses the Edit Labels button, a scrollable list of labels visible from their current location is displayed. Next to each label, two buttons allow the user to either edit or delete the label.

3.7.4 Alignment

This interface guides users, with clear instructions, through the process of correcting the meshes' alignment using the methods described in Section 3.6. Users can use the Up and Down buttons to change the altitude and the bumper button on the controller to precisely correct the orientation. Step-by-step instructions are provided for the simpler *Triangulated* approach, with each of the three selected buildings getting highlighted when it has to be aligned.

4. User Study

This section will describe the user studies we conducted to evaluate the performance, usability, and design of LabelAR. The two-phase study is aimed at identifying areas for improvement and guiding design decisions, particularly with respect to UI choices. The two phases are aimed at balancing both qualitative and quantitative insights, ensuring a comprehensive evaluation.

4.1. Qualitative Assessment with a 5-Second Test

This phase began testing once we had finalized all main features of the application, a couple of weeks before the demo, and allowed us to identify and address design weaknesses to produce a more polished final product. The main goal of this phase was to gather qualitative feedback from a user pool ranging widely in age, professional background, and familiarity with AR applications.
To do so, we simply asked bystanders to try the application in two different locations in Zürich: the ETH Polyterrasse and Lindenhof. We tried to provide little guidance and let the users experiment with the application over a short time span (generally up to a minute, although some users had less time available), providing them with an already aligned environment.
We then collected feedback, focusing on these four aspects:

- **Enjoyment:** We asked users whether they found the experience engaging and fun, gauging their overall satisfaction with the application.
- **Usefulness:** Participants were asked to evaluate whether the application added value to their experience by providing helpful information.
- **Ease of Use:** We investigated whether users were able to intuitively pick up and use the interface without requiring prior instructions.
- **Intrusiveness:** We asked users whether the application hindered their enjoyment of the scenery, aiming to assess whether the AR interface detracted from the physical environment.

4.2. Quantitative Assessment and A/B Testing

The second phase was conducted once the application had been finalized for the demo. The selected participants had prior experience with MR interfaces, but none had previously used LabelAR. We assessed the following:

A/B Testing of the Alignment Interface.
We asked the users to perform the initial alignment using the two proposed interfaces: *Manual* and *Triangulated*. We performed the two alignments in succession, alternating the order between users to ensure fairness. Throughout the experiment, we tracked the users' *Frustration* and *Mental Fatigue* levels. We then asked them to compare the two interfaces based on their **Accuracy** and **Ease of Use**, as well as provide an overall preference between them.

Unsupervised Task Completion.
To obtain quantitative data, we asked the users to complete three fundamental tasks without any external guidance. Starting from the Scene Selection interface, we timed the users while performing the following tasks:

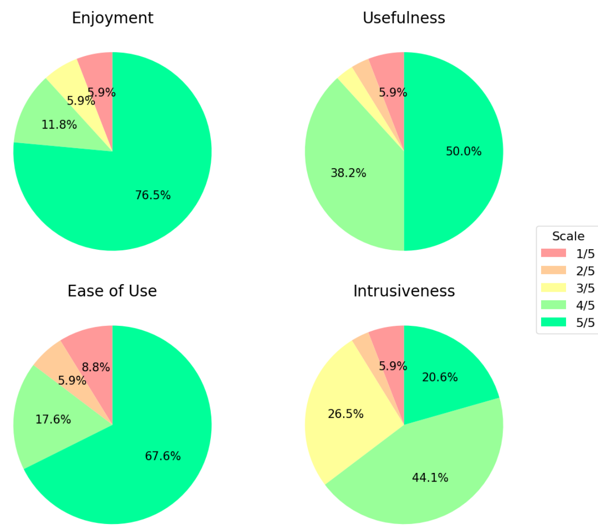

Figure 3. *Results of the 5-Second test.* While the majority of users enjoyed the experience and quickly grasped the interface, some found wearing a MR headset intrusive when admiring the scenery.

1. *Creating a new Label.* The users were expected to notice how to interact with objects through the pointer.
2. *Changing the Building Opacity to Opaque.* This task required the users to interact with the Menu and understand the purpose of the various buttons.
3. *Deleting the new Label.* The final task was more open-ended, as it could be completed either by using the Edit Menu or by interacting with the new label.

5. Results

This section will analyze the results of the user studies we previously outlined, highlighting how these have helped guide the development of LabelAR.

5.1. Results of the 5-Second Test

We conducted the test on 34 users split between the two locations. The users represent a diverse sample of experience levels with VR/MR technology, as shown in Table 1:

	$\# users$
No Experience	14
Experienced VR	19
Experienced MR	6

Table 1. Experience of 5-Second test users with VR/MR headsets.

We plot user ratings of the four main aspects we described in the previous section within Fig. 3. In terms of *Enjoyment* and *Usability* we consider the results satisfactory, while for *Usefulness* and *Intrusiveness* some

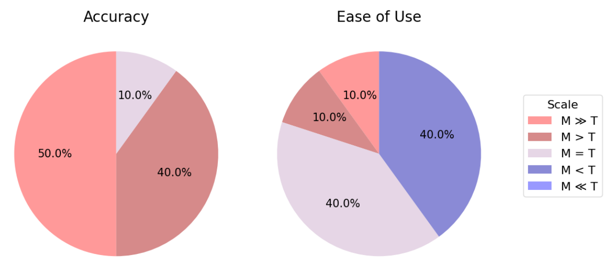

Figure 4. *Results of the A/B test.* Nearly all users found Manual alignment (M) to be more accurate, although quite a few considered the Triangulation (T) interface simpler to use.

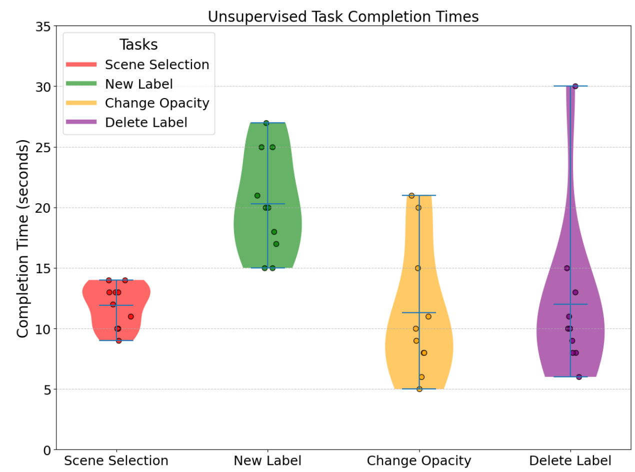

Figure 5. *Unsupervised Task Completion Times.* We plot completion times for the 3 tasks described, as well as the time elapsed from the appearance of the Scene Selection UI to the loading of the final building.

users expressed criticism towards MR interfaces; a user complained about the hardware being cumbersome, while another did not see the point of such landmark labels. Nonetheless, the overall reception of the application was very enthusiastic, a sentiment also shared by users at an internal Magic Leap demo conducted a couple of weeks prior.

An important addition made to the LabelAR interface due to the 5-Second test was the controller overlay displaying the actions available: since the initial tests required a lot of supervision to perform most simple tasks, we decided to make the control information always available to the user.

On top of this, we also recorded bugs and issues encountered during the tests. The most prevalent were three:

1. *Slight Misalignment*, experienced by 4 users, could be easily fixed in Edit Mode.
2. *Tracking Lost*, experienced by 2 users, likely caused by passing the headset around between multiple people (this is not an expected use case of the device).
3. *Imprecise Controls*, experienced by 6 users, complaining that at times, pressing the trigger to select landmarks may unpredictably shift the pointer ray. This issue has remained unresolved as we suspect it may be an issue with the Unity SDK itself.

5.2. Results of the Alignment A/B Test

We conducted the second phase of the test in Polyterrasse on 10 users, and showcase the main results of the alignment A/B test in Fig. 4.
Although some users appreciated the simpler Triangulation-based interface, nearly everyone reported better accuracy using Manual alignment. Considering the alignment is a one-time operation, ideally performed by an expert, and that its accuracy will directly affect all subsequent use of the application, we concluded that manually aligning the origin is the most fitting choice, although triangulation can help as an initial refinement step.

Finally, we also display the changes in *Frustration* and *Mental Fatigue* levels after each alignment in Table 2. Most users found the alignment procedure to be slightly frustrating, with both options performing similarly, however Fatigue levels remained extremely close throughout the experiments, most likely thanks to the short duration.

	Neither	*M*	*T*	*Both*
Frustration	1	3	4	2
Mental Fatigue	7	1	2	0

Table 2. Number of users whose Frustration/Mental Fatigue levels increased during the alignment. *Neither* means no increase, *M/T* means the level increased for only one of the interfaces and *Both* means both caused an increase.

5.3. Results of Unsupervised Task Completion

The Task Completion test was conducted on the same 10 users doing the A/B testing. We plot completion times in Fig. 5. Only a single user required external assistance: this was because the user created the label on the terrain, which cannot be directly selected with the pointer and instead requires going through the Edit Menu.

1. *Creating a new Label.* This task had the longest average execution times since it requires using the keyboard to enter the new name.
2. *Changing the Building Opacity to Opaque.* Most users found this quite intuitive and just needed to understand how to open the Menu.

3. *Deleting the new Label.* Out of the 10 users, 8 deleted the label by selecting it with the pointer, and only 1 went through the Edit Menu without supervision, showcasing how direct interaction is much more intuitive than menu navigation.

For this test, the controller overlay was fundamental as the users had to figure out the controls on their own. Although not immediately, all users found it and used it to successfully navigate the interface.

6. Discussion and Future Work

As highlighted in our user study, LabelAR has proven to be an intuitive and immersive experience for a numerous and varied user group. One of the most notable takeaways is the effectiveness of Magic Leap's mapping and localization subsystems in outdoor environments, something that is not considered supported out of the box. Many users commented on how well the alignment held up, even when extensively navigating the space and reaching outside the initially mapped area. Nonetheless, the controller's reliance on IR LEDs for tracking still caused issues when used in really bright sunlight, and the absence of GPS-capabilities required a lot of work-arounds, meaning that the hardware still has a ways to go for outdoor use.

The main issue with the current iteration of LabelAR is the setup required to work in new areas: the location must be added to the PostGIS DB with an approximate GPS position, the local Space must be mapped with the headset and then the buildings need to be aligned. Future work may simplify this in various ways:

- *GPS Positioning.* By pairing up the headset with a GPS-capable device, for example via a phone app streaming GPS data to the device, one could completely do away with Space selection. Provided the building and terrain data is loaded, LabelAR could then directly work from anywhere in Switzerland.

- *Automatic Alignment.* To further facilitate the setup and improve usability we envision leveraging computer vision techniques to automate the process of aligning meshes to their real-world counterparts. Streaming data from the headset's sensors and cameras to the backend could potentially enable us to develop an algorithm able to continuously compute and refine the alignment of the real world and the 3D models.

- *Automatic Labeling.* Currently, each Space is manually populated with an initial set of labels, which can then be expanded by the users. One could imagine streamlining this by automatically extracting landmarks, for example through Google's Cloud Vision API [1], and using them as suggestions to the user. Given a detection box from the model, one could extract the relevant building meshes via ray casting in Unity and temporarily show them, pending user confirmation.

7. Conclusion

LabelAR has been shown to be an intuitive and enjoyable exploration experience leveraging the latest MR technology and the large-scale geospatial datasets made available by the Swiss government. Although some users were not fully convinced of its usefulness, most were impressed by how it seamlessly overlaid on top of the real world. We also described how the simple addition of a GPS-enabled device could allow LabelAR to transform into a more market-ready application, allowing users to highlight and share any landmarks of their choice, from the churches of Zürich to the peak of the Matterhorn.

References

[1] Cloud vision api, 2024. 6

[2] Postgis, 2024. 2

[3] The swiss coordinates system, 2024. 2

[4] Swisstopo, 2024. 1

[5] Agrawal Dhruv, Lobsiger Janik, Bo Yi, Kaufmann Véronique, and Iro Armeni. Hololabel: Augmented reality user-in-the-loop online annotation tool for as-is building information. 07 2022. 1

[6] Yannick Gassmann. Swisstopo datenimporter für unity, 2021. 2

[7] Sabine Khalil, Andreas Kallmuenzer, and Sascha Kraus. Visiting museums via augmented reality: an experience fast-tracking the digital transformation of the tourism industry. *European Journal of Innovation Management*, 27, 02 2023. 1

8. Appendix

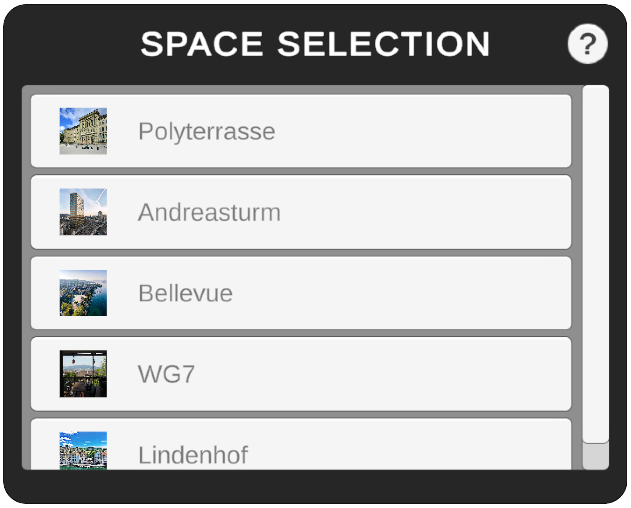

Figure 6. Select the location in which to localize TODO

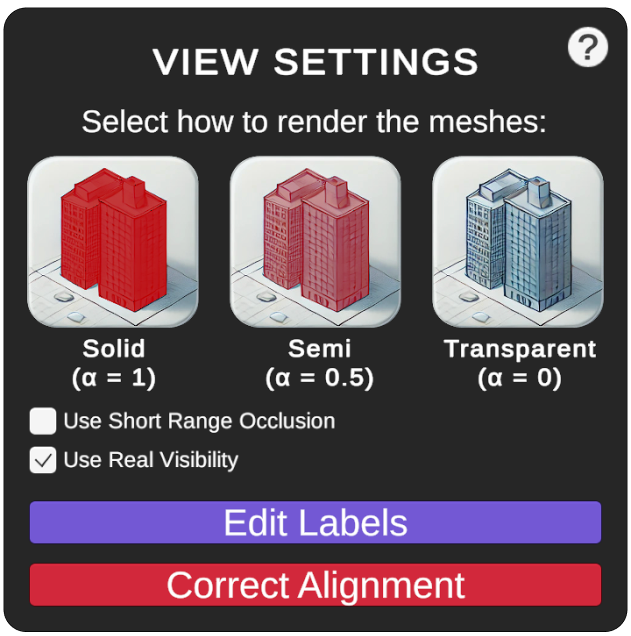

Figure 7. Main menu TODO

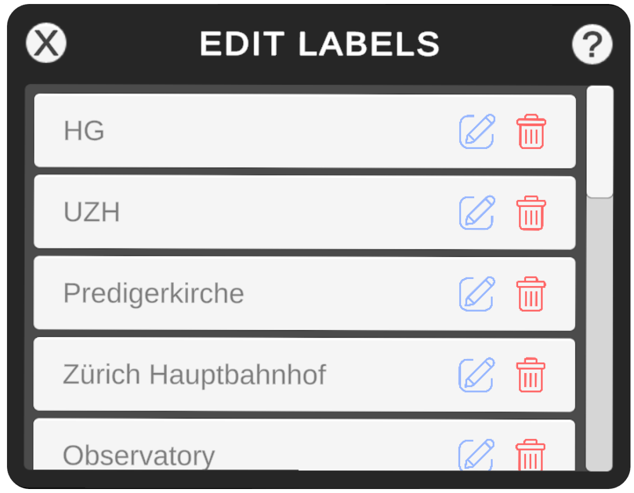

Figure 8. Edit labels menu TODO

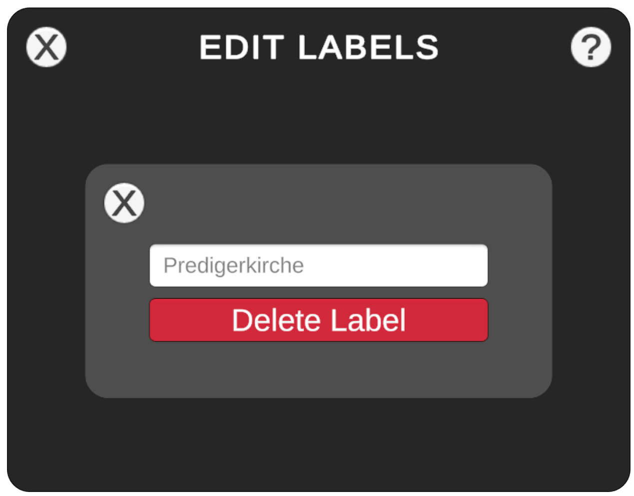

Figure 9. Editing a label TODO

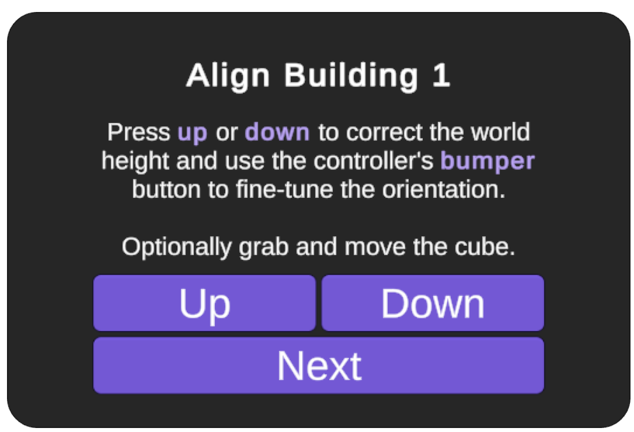

Figure 10. UI guiding the user through the alignment TODO