Technische Universität Wien
Fakultät für Mathematik und Geoinformation
Department für Geodäsie und Geoinformation

Diplomarbeit

Elicitation and Evaluation of Mid-Air Hand Gestures for Global Earth Observation Data presented on Large Public Displays

zur Erlangung des akademischen Grades
Diplom-Ingenieur

im Rahmen des Studiums
Geodäsie und Geoinformation

eingereicht von
Tobias Stachl, BSc
Matrikelnummer 01425144

ausgeführt am
Department für Geodäsie und Geoinformation
der **Fakultät für Mathematik und Geoinformation**
der **Technischen Universität Wien**

unter der Betreuung von
Univ.Prof. Dr.sc. Ioannis Giannopoulos und
Univ.Prof. Dr.techn. Wolfgang Wagner
und Mitwirkung von
Univ.Ass. Dr.phil. Markus Kattenbeck

Wien, im Juli 2021

Unterschrift Verfasser

Unterschrift Betreuer

Technische Universität Wien
Faculty of Mathematics and Geoinformation
Department of Geodesy and Geoinformation

MASTERTHESIS

Elicitation and Evaluation of Mid-Air Hand Gestures for Global Earth Observation Data presented on Large Public Displays

submitted in fulfillment of the requirements for the degree of
Diplom-Ingenieur

as part of the Master's program
Geodesy and Geoinformation

by
Tobias Stachl, BSc
Matriculation number 01425144

at the
Department of Geodesy and Geoinformation
of the **Faculty of Mathematics and Geoinformation**
of the **Technische Universität Wien**

under the supervision of
Univ.Prof. Dr.sc. Ioannis Giannopoulos and
Univ.Prof. Dr.techn. Wolfgang Wagner
and
Univ.Ass. Dr.phil. Markus Kattenbeck

Vienna, in July 2021

Signature author

Signature supervisor

Declaration

I hereby declare that I have written this thesis by myself without any help or assistance of others. External literature used to clarify the content or provided data sources are fully cited. All mentioned information is in accordance with fact or truth up to my knowledge.

Vienna, in July 2021

Acknowledgements

The accomplishment of this work would not have been possible without the support of several persons.

At first, I would like to express my gratitude to my supervisor Dr. Markus Kattenbeck. In our first meeting, he said, that for him, a Master thesis is a collaboration between the student and the supervisors. The continuous support from the beginning until the end, valuable advises, numerous meetings and discussions proved this without any doubt.

I would like to thank Prof. Ioannis Giannopoulus and Prof. Wolfgang Wagner, who made it possible to conduct my master thesis as a collaborative work between the two Research Units Geoinformation and Microwave Remote Sensing.

Many thanks go to my colleagues from the Department of Geodesy and Geoinformation, whose suggestions and feedback were very helpful throughout every stage of this work.

Over the last years, many fellow students have become true friends while we had been studying and working together in productive groups or meeting off-campus. May those friendships continue beyond our college days.

Furthermore, I would like to take this opportunity to thank my family for their continuous encouragement and support, especially for always having an understanding ear and time for my concerns as well as for their financial support during my education.

And last but not least, I want to devote special thanks and my appreciation to my girlfriend Sarah, who also finished her Master studies recently. Without her love, patience and continuous support in all situations the last years would certainly not have gone so well.

Abstract

Global Earth observation data and therefrom derived information are essential for understanding the current state, natural processes and dynamics of the Earth system. However, the characteristics of these complex spatial data sets make accessibility to a lay audience challenging. Therefore, a prototype of a map viewer was developed as an interactive visualization tool to increase the approachability of these data sets. The prototype is designed to, on the one hand, display global satellite-based radar backscatter data and, on the other hand, be controllable by touchless mid-air hand gestures.

The definition of an agreed upon set of touchless mid-air hand gestures is still an open question, especially in the context of interacting with complex data sets. To determine the base functions and features of such a novel system, a focus group discussion was conducted. In total, seven participants including the moderator discussed about existing online web maps and their advantages and disadvantages. Further major topics were how satellite-based radar backscatter data could be explained and presented to novices as well as how touchless hand gestures are currently perceived and could improve the interaction with such data.

Based on the results of the focus group discussion, a gesture elicitation study with 30 participants was conducted. With individually held online meetings, test persons proposed gestures for twelve different tasks. In total, **462** gesture proposals as well as subjective ratings and qualitative statements of the mental creation process were gathered. From this elicitation touchless mid-air hand gestures and common narratives among the test persons were obtained. Quantitatively, for navigational map tasks which includes zooming and panning, the consensus could be clearly defined with the calculated agreement rates. For all other referents, the qualitative analysis which also took the second and even third proposals into account, gave suggestions for common narratives.

Then, the consensus set of gestures and narratives as well as the desirable features obtained from the focus group discussion, were basis of the prototype implementation. The Leap Motion Controller (LMC) was used to determine the position and characteristics of the end-user's hands within its field of view. The satellite-based radar backscatter data derived by Bauer-Marschallinger et al. [5] was used as the base map.

The emerging system was planned to be located in a public space in the premises of the Department of Geodesy and Geoinformation at the TU Wien. Therefore, the usability and the user experience of the prototype was tested with ten participants. With an average score of 75.75% of the System Usability Scale (SUS), the prototype already reached a positive overall tendency of the usability. Furthermore, the User Experience Questionnaire (UEQ) also indicated a positive tendency especially for the scales *Attractiveness*, *Stimulation* and *Novelty*. Whereas, the lower values for *Perspicuity*, *Efficiency* and *Dependability* suggested that improvements of the prototype were needed.

As a last step, the discovered issues were countered with solutions of the implementation, such as redesigning the tutorial or a rethought realization of the semantics of features.

Kurzfassung

Globale Erdbeobachtungsdaten und daraus abgeleitete Informationen sind essenziell für das Verständnis des aktuellen Zustands, der natürlichen Prozesse und der Dynamik des Systems Erde. Die Eigenschaften dieser komplexen, räumlichen Datensätze machen die Zugänglichkeit für ein Laienpublikum jedoch zu einer Herausforderung. Daher wurde ein Prototyp einer Web-Applikation die kartenähnliche Inhalte darstellt als interaktives Werkzeug zur Visualisierung von räumlichen Daten entwickelt, um die Ansprechbarkeit dieser Datensätze zu erhöhen. Der Prototyp ist so konzipiert, dass er einerseits satellitengestützte Radardaten anzeigt und andererseits durch berührungslose Handgesten steuerbar ist.

Die Definition von berührungslosen Handgesten ist weiterhin eine offene Frage, insbesondere im Kontext mit der Interaktion mit komplexen Datensätzen. Um die grundlegenden Funktionen und Eigenschaften eines solchen Visualisierungs-Systems zu ermitteln, wurde eine Fokusgruppendiskussion durchgeführt. Insgesamt sieben Teilnehmer inklusive des Moderators diskutierten über bestehende Online-Karten und deren Vor- und Nachteile. Weitere Hauptthemen waren, wie satellitengestützte Radardaten erklärt und präsentiert werden können. Außerdem wurde diskutiert, wie berührungslose Handgesten derzeit wahrgenommen werden und wie sie die Interaktion mit solchen Daten verbessern könnten.

Basierend auf den Ergebnissen der Fokusgruppendiskussion wurde eine Gestenerhebungsstudie mit 30 Teilnehmern durchgeführt. In individuell durchgeführten Online-Meetings schlugen die Probanden Gesten für zwölf verschiedene Aufgaben vor. Insgesamt wurden **462** Gestenvorschläge sowie deren subjektive Bewertungen und qualitative Aussagen über den mentalen Entstehungsprozess erhoben. Aus dieser Erhebung wurden berührungslose Handgesten und übereinstimmende Schilderungen der Probanden gewonnen. Quantitativ konnte für navigatorische Kartenaufgaben, wie Zoomen und Verschieben, der Konsens mit den berechneten Grad der Übereinstimmung klar definiert werden. Für alle anderen Aufgaben ergab die qualitative Analyse, die auch den zweiten und sogar dritten Gestenvorschlag berücksichtigte, Vorschläge für übereinstimmende Beschreibungen.

Die übereinstimmenden Gesten und Narrative sowie die aus der Fokusgruppendiskussion gewonnenen Eigenschaften waren schließlich die Grundlage für die Umsetzung des Prototyps. Der Leap Motion Controller (LMC) wurde verwendet, um die Position und Charakteristika der Hände des Benutzers zu bestimmen. Die von Bauer-Marschallinger et al. [5] erstellten satellitengestützten Radardaten wurden als Basiskarte für die Web-Applikation verwendet.

Der entstehnde Prototyp sollte in einem öffentlichen Raum im Bereich des Departments für Geodäsie und Geoinformation der TU Wien aufgestellt werden. Daher wurde die Benutzerfreundlichkeit und Bedienbarkeit des Prototyps mit weiteren zehn Teilnehmern getestet. Mit einer durchschnittlichen Punktzahl von 75, 75% des System Usability Scales (SUS) erreichte der Prototyp bereits eine positive Bewertung der Benutzerfreundlichkeit. Auch die erreichten Werte des User Experience Questionnaire (UEQ) zeigten insbesondere für *Attraktivität*, *Anreiz* und *Neuartigkeit* eine positive Tendenz. Die niedrigeren Werte für *Verständlichkeit*, *Effizienz* und *Zuverlässigkeit* deuten hingegen darauf hin, dass Verbesserungen des Prototyps erforderlich sind.

In einem letzten Schritt wurden die Bereiche des Prototyps, in denen Probleme entdeckt wurden, überarbeitet und verbessert.

Contents

List of Figures

Chapter 1

Introduction

For the last two decades, interest in touchless and gestural interaction with visual interfaces has been steadily growing. Increasing availability of low-cost devices, such as Microsoft Kinect or Leap Motion Controller, helped this technology to be studied and adopted in many areas. Furthermore, touchless interaction brings several advantages compared to other modalities such as touch-based or mouse-keyboard input. For example, users are not required to come close to the screen to interact. This can be useful, if the display is placed in a non-touchable or non-reachable area. Additionally, the user can stay at a comfortable distance and does not have to touch the display due to possible hygienic reasons [66]. Especially with the spread of Covid-19, the need of touchless interaction technologies has increased significantly [32]. Furthermore, recent work indicated that mid-air gestural interaction can increase the usability and accessibility of applications [7]. Mid-air hand gestures can be used to manipulate or interact with digital content or remote devices without touching any surfaces or devices. Based on sensor tracking of hand movements, postures and gestures are used to trigger actions and operations on a computer system, which resembles to a interaction style of human-computer interaction (HCI). LaViola Jr et al. [38] stated that *„HCI is the process of communication between users and computers (or interactive technologies in general)"*.

As a style of HCI modality, touchless mid-air hand gestures were studied and used in several implementations such as in museums in combination with large displays (e.g. [55]) or as an additional modality for popular gaming platforms like the Nintendo Wii or Microsoft Xbox (e.g. [22]). However, there is no established mid-air gesture vocabulary or guideline to follow in order to implement such an interaction system. Also Nielsen et al. [49] pointed out that *"there is no such thing as a universal gesture vocabulary for every application"*. Hence, the identification of an appropriate set of touchless mid-air hand gestures needs to be done for each use case in order to meet the expected performance of the implemented system [2].

In this work, the Leap Motion Controller (LMC) developed by Ultraleap [60] was used to track and perceive positions and movements of the user's hands. The LMC provides information of detected hands within its field of view such as hand palm position and direction, fingertips position, pinching strength and other relevant aspects in order to work with mid-air hand gestures. As a Human Interface Device (HID), LMC allows a computer to be controlled by hand and finger movements. LMC's have successfully been used to recognize hand gestures in previous work (e.g. [54], [40], [8]). Its suitability for deriving mid-air hand positions and gestures was also reported by Bachmann et al. [2]. A recent study by Martins and Notargiacomo [42] showed that the LMC can be implemented successfully for two-dimensional game environments. Additionally, the sensor was used in proprietary systems such as the *Touchless Pedestal* by Ideum [30].

Viewing and interacting from a greater distance will help to gain a more holistic picture of the presented data. In the field of HCI public displays have become a common used technology enabling people to interact with dynamic information on large screens (e.g. [21], [53]). They can be found in airports, universities, shopping malls and others.

The Department of Geodesy and Geoinformation at TU Wien, including the research unit Microwave Remote Sensing is producing complex earth observation data sets based on satellite measurements on a global scale. For presentation and exploratory reasons, those data sets should be available for a wider audience, experts as well as non-experts, in a public environment. This would not only increase the popularity of such data sets, but could also let them benefit from quality curation by volunteers.

Global earth observation data sets are likely not suitable for display on a single screen due to their intrinsic characteristics such as global extent and spatial resolution. Hence, the user faces a situation in which a distinction between important and irrelevant information is necessary. Efficient representation and robust interaction are of key interest in order to interpret complex information [19]. Existing public display systems often show content which is specifically designed for the display and environment for which it was planned for. Therefore, the presented data can be adjusted to the screen's extent as well as to the expected target groups or interactions.

The data which is used in this work is not adjustable for different ways of presentation. Hence, to overcome the trade-offs between characteristics of data and displays, HCI interfaces often allow the user to move the displayed content (panning) as well as vary the scale at which the data is shown (zooming) [14]. Those mechanisms are started and controlled by distinct interactions of the user with the visual interface.

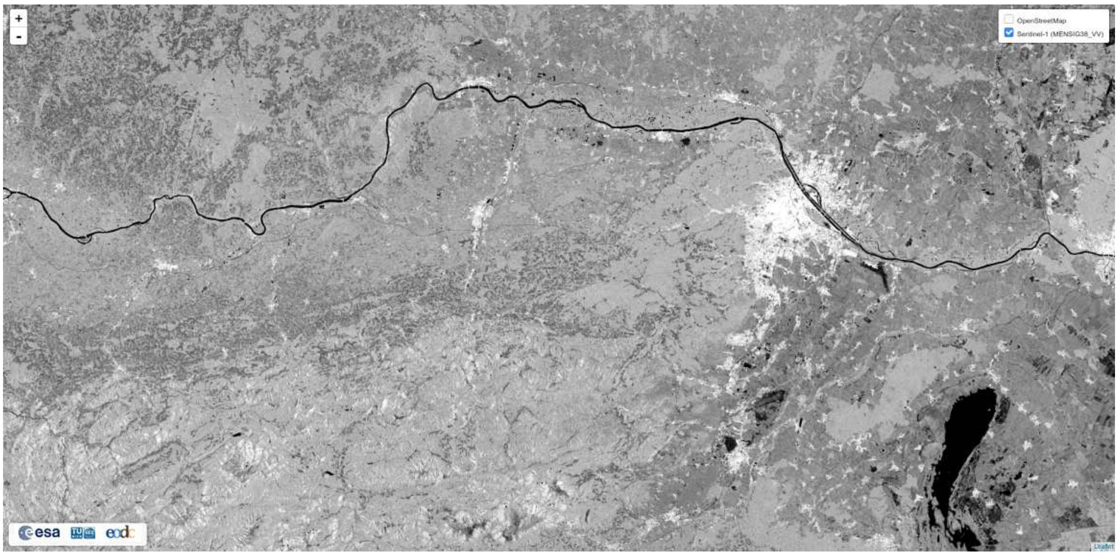


Figure 1.1: Sentinel-1 Global Backscatter Model (S1-GBM)

The visualized data set, an example of which is shown in figure 1.1, is based on synthetic aperture radar (SAR) backscatter data, remotely sensed by the satellite mission Sentinel-1, consisting of the platforms Sentinel-1A and -1B. It contains pixel-based backscatter values, which were normalized for the influence of the observation geometry and temporally averaged for the years 2016/2017. The research unit Microwave Remote Sensing produced this data set (S1-GBM) in collaboration with the Earth Observation Data Center for Water Resources Management (EODC) [5]. Both, the co-polarized (VV) and cross-polarized (VH) backscatter data were used for presentation. Due to visualization performance reasons, the data set was transformed to tiled image files of greyscale values and saved as image pyramids.

The main objective of this work is the implementation of a data viewer which visualizes these S1-GBM data layers and is controlled by touchless mid-air hand gestures.

As a first step, a literature review tries to outline the state of the art regarding interaction systems which use touchless mid-air hand gestures as input modality at the time of writing.

Due to the novelty of the emerging system, it was of special interest to understand which features and operations would be of high importance for the end-user. Hence, a focus group discussion was conducted to gain further insights.

To understand if there are preferred touchless mid-air hand gestures to trigger these actions, a gesture elicitation study (GES) was carried out. This study empirically identified a set of mid-air hand gestures to interact with the S1-GBM data on a large public display. The resulting user-defined consensus set was then implemented in a working prototype.

The emerging system is planned to be placed in a public environment, which raised the question of the usability and experience potential end-user would have. Therefore, a usability study was conducted in order to reveal areas of possible confusion as well uncover opportunities to improve the overall user experience and usability of the prototype. Improvements suggested by the usability study were implemented.

Consequently, the data viewer should offer an easy-to-use possibility to present and interact with recent global data sets on public events such as conferences, keynotes or internal team meetings.

Chapter 2

Related work

This work can be positioned within the body of work on designing touchless gestural interfaces for public displays in order to increase the availability of maps based on global Earth observation (EO) data to remote sensing professionals and lay people alike.

Global EO data allows scientists to efficiently monitor natural processes and dynamics of the Earth system. Vast amounts of EO data exploit opportunities to produce valuable information products. However, especially the intuitive visualization to a diverse community of these complex data sets in order to increase the approachability is topic of current research:

Due to the volume, variety and complexity of global EO data sets and therefore needed computational infrastructure, the accessibility to the general public is limited. For example, the work of Bucur et al. [11] described the Earth Observation Data Center for Water Resources Management (EODC) as a public-private partnership founded in May 2014, in Austria. Like other big data infrastructures [65], the EODC as a dedicated data center offers a framework to access global EO data, process them and extract results via a cloud platform.

Furthermore, the work of Shrestha et al. [58] described how ESRI technology can support data management, analysis and visualization through the suite of ArcGIS web apps. Huntington et al. [29] proposed a free web-based application called *Climate Engine*, which is based on Google's parallel cloud-computing platform *Google Earth Engine* and is able to process, visualize, download and share global EO data.

Similarly to the aforementioned contributions, Hodam et al. [27] emphasizes the potential of global EO data to „open the minds“ to new perspectives on the Earth by bringing this complex topic into classrooms with digital integrated learning environments.

Private entities also introduced several map data viewers, which showcase the available satellite-based products and data collections of the companies. Examples of them are EO Browser[1] or the map viewer by Descartes Labs[2].

Furthermore, map viewer systems have been created by public entities to present study results as part of national or international scientific projects. One example is GISportal[3] which also served as basis for the data viewer originated from the project DRIDANUBE to present map-based data regarding drought risk in the Danube region[4].

Touchless gestural interfaces have been studied already for a long time. As reviewed by Cai et al. [12], a majority of the proposed solutions were based on depth sensors or cameras such as the Leap Motion Controller or Microsoft Kinect.

However, only few previous work has examined how users define gestures interacting with an application presenting map-based data sets such as global satellite-based radar backscatter data. For instance, Bellucci et al. [6] explored the use of Nintendo Wii Remote controller (Wiimote) as a touchless input modality for placing annotations on

[1] https://www.sentinel-hub.com/explore/eobrowser/
[2] https://maps.descarteslabs.com
[3] https://github.com/pmlrsg/GISportal
[4] https://www.droughtwatch.eu/

a map in large display environments. Their prototype is able to collaboratively place multimodal annotations on a map and help operators of a crisis management unit to coordinate activities of different forces. But they did not focus on a user-defined gesture set, which is the main objective of the current work.

The work of Sathiyanarayanan and Mulling [56] investigated hand gesture recognition and map navigation using the, meanwhile discontinued, *MYO* armband on Apple Maps. The authors developed a prototype implementing already existing gestures by the manufacturer of the MYO armband to interact with Apple Maps and conducted a study with 20 participants. The System Usability Scale (SUS), which provides measures about the usability of the prototype and a second questionnaire, aiming at the comprehension of the ergonomic aspects of gestural interaction were carried out after the study. Conclusively, their results showed that participants felt comfortable using touchless interaction, but the observed low accuracy of gesture execution by the MYO armband resulted in distrust and frustration among the users.

Since the definition of a convenient set of touchless mid-air hand gestures remains still an open question, special focus was given on gesture elicitation studies (GES). A systematic review by Villarreal-Narvaez et al. [63] investigated 216 studies which performed a GES. The basis of this way of obtaining a user-defined gesture set was provided by the work of Wobbrock et al. [69]. They designed a method to collect test person's preferred symbolic input and calculate agreement among them, which was first applied by Wobbrock et al. [70]. Although, they used a touch-sensitive device to derive a user-defined gesture set for predefined tasks.

The study of Morris et al. [45] compared a user-created gesture set for interactive surfaces with one created by three HCI design experts. The 22 participants of their study preferred the user-created gesture set. This confirms user-centered system design approaches in comparison to approaches defined by HCI researches since they are seen as more physically and conceptually complex.

The work of Vatavu and Zaiti [62] presented insights from a GES in the context of interacting with touchless mid-air hand gestures with a TV set. Eighteen participants contributed and rated free-hand gestures for 21 distinct TV tasks. The tested tasks included functions which are common for television watching, but also used in web user interfaces such as *Open menu* or range inputs like *increase* or *decrease volume* which are often visualized as sliders. The authors compared agreement rates between experienced users versus novice users, between task categories and correlation between agreement and average thinking time. Additionally, the authors use the LMC as tracking device. Therefore, their work approaches several aspects of the rationale of the current study.

By the time of writing of the work of Wittorf and Jakobsen [68], the main body of literature was dedicated to explore what was technically possible. Thus, they presented results of a GES where 20 participants performed gestures for 25 actions on a three-meter wide display, which resembles the context in the present study. After the gesture creation process, they classified the elicited gestures using a defined taxonomy and provide comments from post-study interviews.

A comprehensive literature review by Vogiatzidakis and Koutsabasis [64] put its focus on gesture elicitation studies for mid-air interaction. As one conclusion of their work, they emphasized the necessity of assessments of the content of design suggestions for particular contexts of use. Furthermore, they also suggested that further work need to be done to *„further validate the results of gesture elicitation with technical tests about measured usability and fatigue“*. The findings of the usability test in the end of the current work, can contribute to the latter.

The work of Hoffmann et al. [28] investigated differences in user preference for voice, touch and mid-air gestures as input modalities to control smart homes. The authors conducted an elicitation study in which 13 participants were asked to propose commands for eleven tasks for each modality. During the command creation process, the participants suggested mainly user interface elements such as buttons and sliders for the touch modality. Since the prototype in the current study also includes user interface elements, the results of their consensus interaction sets can be used for comparison. The results of their quantitative and qualitative analysis clearly showed a preference for voice and touch control over mid-air gestures. Nevertheless, they pointed out that further analysis of elicited mid-air hand gestures need to be done in order to measure their usability and fatigue which is also in line with Vogiatzidakis and Koutsabasis [64].

Finally, Gentile et al. [23] focused on the elicitation and evaluation of zoom gestures. At first, they used a GES to identify a consensus set of zoom gestures while interacting with desktop displays. In a second study, they gained further insight into perceived workload, usability and effectiveness of the elicited gestures.

Focus group discussions were emerging as a research method in the 1950's, as open ended interviews were expanded to group discussions [59]. Focus groups produce mainly qualitative information about the scope of the investigated topic by interactive and directed discussions. The method is widely used in market research, product planning and in system usability studies (e.g. [13], [43], [50]). The developers of interactive systems are eager to discover what users want from the system. Through the modality of the moderated discussion, participants provide spontaneous reactions and ideas which would never emerge in a usability test [36]. In this study, a focus group discussion is used to determine the desirable features and expectations of the proposed data viewer based on touchless mid-air hand gestures.

Chapter 3

Focus group discussion

Focus groups are widely used in market research, product planning and system usability studies. For instance, during the design phase of a product or service, they can help gathering information about expectations of potential users. In this study, a focus group discussion was conducted in order to determine desirable features and operations of the emerging data viewer system.

In this chapter, the procedure of the discussion, recruitment process of the participants, the preliminary survey and the gathered results are described.

3.1 Procedure

Due to the Covid-19 pandemic and consequently enacted restrictions for accessing the premises of TU Wien, the focus group discussion was carried out online via the video conferencing tool Microsoft Teams. Prior the discussion, all participants were provided an explanation of the study procedure and completed a signed consent form as well as the privacy policy for the Research Unit Geoinformation, Department of Geodesy and Geoinformation. Additionally, they were asked to fill out a survey which covered demographic information and questions about possible prior knowledge in data viewer systems, satellite-based radar backscatter data or touchless interaction methods.

Using open-ended questions, also partly based on the given answers of the preliminary survey, a moderator facilitated the focus group discussion. The moderator wrote down notes on a shared screen, which all participants could see. In addition, a typist who was connected without a microphone, wrote down notes on paper. The discussion was audio and video taped. The participants were encouraged to discuss and answer the given questions in an open manner.

3.2 Recruitment

The recruitment of the six participants of the focus group discussion was done by asking potential people in person. Using an online appointment scheduling software, a suitable date was found. The focus group was composed of six persons which were recruited among undergraduate students, colleagues and external persons within the personal area of the study designer. Those participants were divided in two categories:

The first category represented by two persons were experts, who have been working as research scientists within the Research Unit Microwave Remote Sensing at the Department of Geodesy and Geoinformation. Both were using satellite-based radar backscatter data as well as map-based data viewers, which are showing global data sets, on a daily basis. Furthermore, both are also working as (assistant) lecturers.

The second category was representing the target group of the emerging system. This category was subdivided in two subcategories, namely (1) students and (2) external persons. The first subcategory was represented by two students, who were in their first year. Those were also indicated as potential users of the emerging system. Since they have chosen the bachelor study program of Geodesy and Geoinformatics, it was assumed that they have a certain degree of interest in learning about and working with geographical data sets. This fact brought moderately expert knowledge to the group discussion, which was desired. The second subcategory was represented by two external persons who have not been, according to the answers given in the preliminary survey, familiar with satellite-based radar backscatter data sets as well as data viewer systems presenting complex global data sets. An exception to those data viewers were online web mapping applications and routing services such as Google Maps or OpenStreetMap, which offer free access to global street data or optical satellite images to a broad audience.

3.3 Preliminary survey

Two days before the focus group discussion, all participants received a link to a preliminary online survey. The survey covered three closed-ended demographic questions in order to gather basic information about the participants which included age category (e.g. 26-35 years), employment status (e.g. employed, full or part time) and gender. Additionally, nine closed-ended dichotomous (with a third answer "not sure") and one open-ended questions regarding prior knowledge in the topics covered in the focus group discussion were asked. The first five questions covered prior knowledge in using web mapping services or geographical information system (GIS) software. The last five questions covered prior knowledge in using touchless mid-air hand gestures as an input modality. The answers of the survey helped the study designer to form suitable open-ended questions for the focus group discussion. Further, the degree of explanation to the topics could be thereof adapted.

3.3.1 Survey results

Demographic information

Within the six participants, one person was female and their age ranged from 18 to 35 years. As described in section 3.2, the participants were selected according to predefined categories (experts and target group). This was verified by asking about the participant's employment status: Three persons, including the experts and one external person were employed, full or part time and the remaining three persons were students, with a secondary employment.

Prior knowledge and experience

All participants stated that they are using web mapping services such as Google Maps or OpenStreetMap to find locations and for navigational purposes. Hence, everyone was familiar with web mapping services, know what those are and how they can be used. No one was already adding or adapting map features on OpenStreetMap. This was asked, because a positive answer would indicate a high interest and skill in using web mapping services. Three persons stated that they have used GIS Software such as ESRI's ArcGIS or QuantumGIS to work with and visualize geographical data, whereas one person was not sure. The question if a participant is working and creating geographical data on a daily basis within their studies or work, two persons stated yes, two no and two were not sure. This result indicates exactly the spread of the participants between the three predefined categories, experts, students and external persons. Lastly, no one stated that they have never used a web mapping service or GIS software.

The second part covered prior knowledge in using touchless mid-air hand gestures for interacting with computer systems. All participants stated that they had controlled actions while gaming with Microsoft Kinect or Nintendo Wii. Three persons have already used touchless mid-air hand gestures while wearing an Augmented- or Virtual-Reality system. All participants denied that they have used touchless mid-air hand gestures with web mapping services, whereas two persons already used touchless mid-air hand gestures while interacting with a public display system (e.g. in museums).

The last question was open-ended and covered the aspect if the participants had been interacting with other systems using touchless mid-air hand gestures. Only one person answered, that he or she played with the musical instrument Theremin for short period of time, which is controlled without any physical contact.

3.3.2 Conclusions

The answers helped the moderator to form open-ended questions to ask in the focus group discussion. The preliminary survey showed that all participants had a certain degree of prior knowledge in using data viewer systems such as web mapping services. Hence, the discussion could be started at a certain level without the necessity of explaining every detail of data viewer systems. To discuss the features of a data viewer, Google Maps was presented as an example.

Differently, prior knowledge in touchless mid-air hand gestures in context with complex data are not spread among all participants. Therefore, a video example showing touchless mid-air hand gestures was presented.

3.4 Data analysis

The audio recordings were transcribed verbatim and translated from German to English. The combination of the transcript with the moderator's and typist's written notes was the basis of the subsequent analysis. In an analytical process, pieces of this data were constantly and successively compared with other parts of the data to derive similarities and differences. This is referred to the method of constant comparison [16] within the strategy of grounded theory [24]. It advances coding, inductive categorization and conceptualization and resulted in the following subsections:

1. Features of a data viewer
2. Understanding and using satellite-based radar backscatter data
3. Interaction based on touchless mid-air hand gestures

After the categorization, the process of axial coding [16] gave suggestions for relationships between concepts within the main themes. To accomplish that, the data was analyzed multiple times to capture both general properties and dimensional variation ob the relationships. The results of the axial coding process is presented in section 3.4.4.

As a last step, conclusions of the data analysis were drawn to define desirable functionalities and features for the emerging data viewer system.

3.4.1 Features of a data viewer

One goal of the focus group discussion was to get a deeper insight, why and how the participants use web mapping services, which provide map-based data such as street maps or satellite imagery. Using Google Maps as an example (see figure 3.1), the participants were encouraged to list actions they perform frequently and which of them are considered as must have features (emphasized as **bold** text).

- **Navigation and routing**
- **Localization of points of interest (POI)** such as restaurants, companies or sights
- **Additional information to those POIs** such as opening hours or contact information
- **Overlay optical satellite images**
- Google StreetView
- Derive coordinates (latitude, longitude) of desired locations
- Tilting the map to get topographic information
- **Availability of global data**
- **Rummaging around and explore globally** (vacation planning)
- Saving POIs and sharing them with other users
- **Live traffic information**
- Intuitive and easy to use interface

One particular action or feature was especially emphasized by all participants, namely the possibility of exploring and rummaging around globally without a specific intention. Also in combination with the localization of POIs, the participants highlighted the opportunity that there is no need of knowing a specific address or name of a premise.

Another aspect which was raised by a participant was the intuitive user interface of the presented example. Constant visual appearance and location of interface elements such as search field or buttons decrease cognitive load and increase learning effects, which lead to a satisfactory and more frequent use.

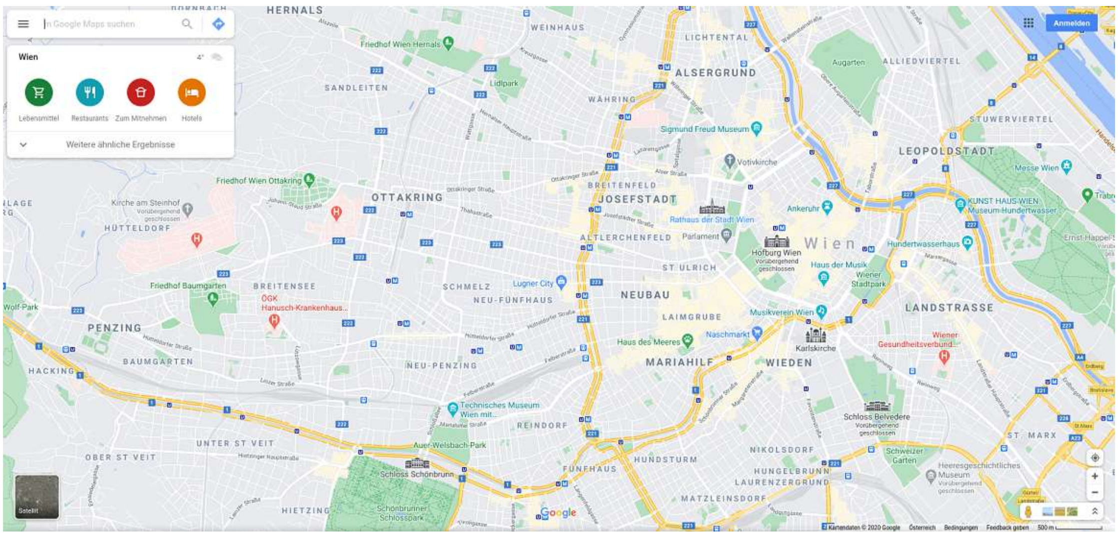


Figure 3.1: Google Maps was used as an example for discussing features of a data viewer.

3.4.2 Understanding and using satellite-based radar backscatter data

After presenting the participants a data viewer which showed satellite-based radar backscatter data (see figure 3.2), the following main questions were raised:

- Where am I?
- What are the bright or dark areas representing?
- Where are specific types of land cover (e.g. mountains, hills, forest, or cities)?

In summary, the participants except the experts were overwhelmed with the presented data. The main aspects were, on the one hand, the lack of orientation, and on the other hand, missing additional information about the presented data. This brought the discussion immediately to an ideation process emerging several ideas.

Firstly, the overlay of names of certain areas such as big cities, lakes or mountains as well as information about the topography could help users to orientate. In order to further support orientation, the current scale should be shown.

Secondly, the participants pointed out the need of further explanations of the presented data. Presenting differences in land cover with certain examples (e.g. forest, cities or water) could help the users' interpretation. To emphasize also an educational use case, those examples could be specific, geological important events such as floods or intense farming areas. From the experts' perspective a translation of the shown grayscale values to actual observed values, also the value range and temporal information is important.

Another idea was to provide the opportunity to swap between time stamps in order to detect how areas or values change over time. Also the possibility to swap between satellite-based radar backscatter and optical satellite data could help both for orientation and interpretation.

Like already discussed in section 3.4.1, the possibility of exploration without certain intentions, was raised again in combination with satellite-based radar backscatter data. Museums (e.g. explanation and exploration), national parks (e.g. exploration of the surrounding area) and educational purposes (e.g. awareness-raising) were listed as possible use cases of such a system.

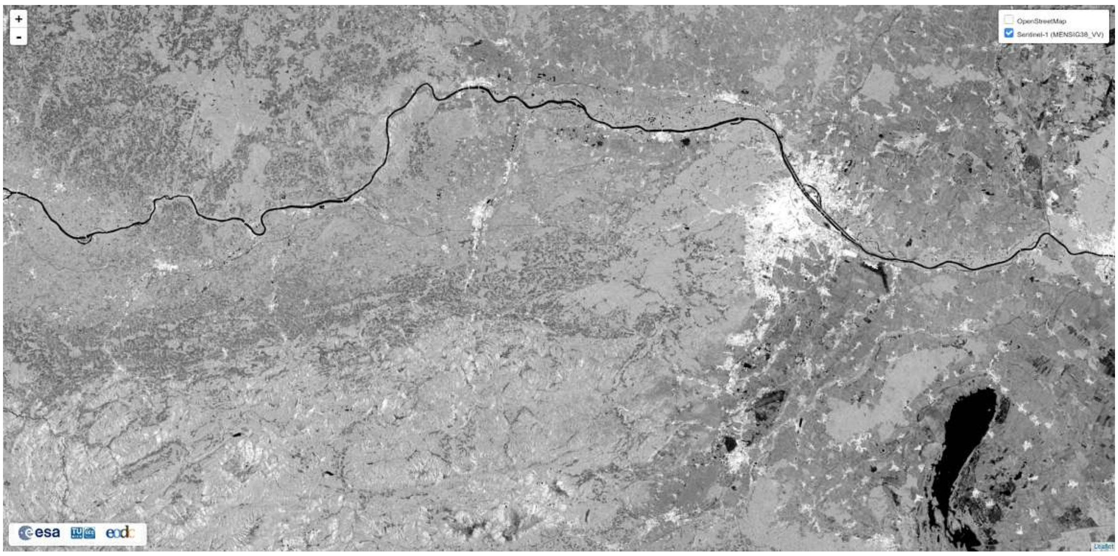


Figure 3.2: Satellite-based radar backscatter data was used as an example to discuss the shown data.

3.4.3 Interaction based on touchless mid-air hand gestures

The third main theme of the focus group discussion was about touchless mid-air hand gestures used for interaction with computer systems in general. As a starting point, the YouTube video[1] of a touchless mid-air hand gesture interface in the movie „Minority Report“ was shown as an example (see figure 3.3).

Several positive and negative aspects of interaction systems based on touchless mid-air hand gestures were listed. One person reported, that such systems are imagined and experienced to be entertaining both for the user and other people watching, due to extraordinary approach of interaction. Another participant described, that contrarily to mouse input, one is metaphorically „moving through the data“, which increases the entertaining factor. Since the Covid-19 pandemic was present at the time of the focus group discussion, also the hygienic aspect was raised as advantageous because the user would not have to touch any surfaces.

On the opposite side, participants were concerned about reachable accuracy in comparison with a mouse-keyboard input system. However, they agreed upon that a comparable accuracy is not needed for the proposed use cases. Also, the learning process of touchless mid-air hand gestures was expected to be significantly longer than using a mouse input method.

The positive and negative statements resulted into the following aspects:

- Few, but easy-to-use, clear and distinguishable gestures
- Large display is necessary
- Short introduction to implemented gestures

Most participants stated that the implementation of the touchless mid-air hand gestures should be comparable to a mouse input. The reasons for that statement was to reduce the time of the learning phase for the new input modality and decrease the complexity of possible gestures. More precisely, the user should be able to pan (left, right, up and down), zoom (in and out) and select. In addition, a person stated the need of a function, hence gesture, to get back to the initial view, if the user got lost while exploring through the data.

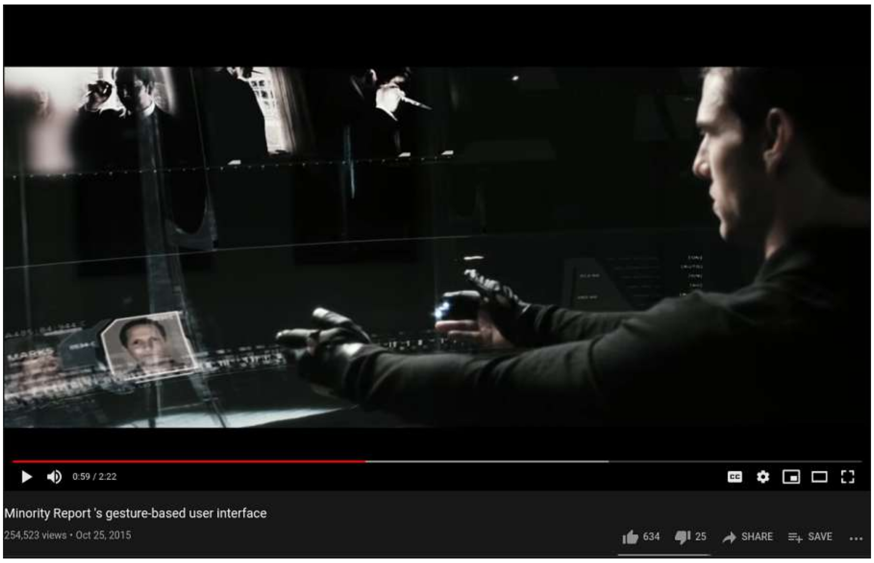


Figure 3.3: The YouTube video of a touchless mid-air hand gesture interface in the movie „Minority Report“ was used as an example.

[1] www.youtube.com/watch?v=PJqbivkm0Ms (Last accessed on 2021-06-14)

3.4.4 Relationships

Relationships of concepts between the three main themes were found through axial coding.

All participants agreed upon that the concept of **intuitive interaction** is a positive feature of the presented web mapping tool Google Maps. In addition, the same concept was reported as an important requirement for systems utilizing touchless mid-air hand gestures as an input system. The gestures should be easy-to-use, clear and distinguishable in order to reduce cognitive load while using. To further reduce ambiguities, an introduction to the hand gestures should be implemented.

The concept of **rummaging around and exploring data** was reported as an use case of Google Maps as well for a data viewer showing satellite-based radar backscatter data. Most of the participants would use the emerging system to explore interesting areas such as their residence or distinctive and prominent regions (e.g. Amazon or big cities).

In combination with the latter, **the necessity of additional information** took a major part of the whole discussion. All participants agreed that additional information about the presented data would be needed for orientation and explanation reasons. They suggested both a distinct hand gesture for superimposing additional data (e.g. naming of distinctive regions) and adjusting the contrast of the satellite-based radar backscatter data.

3.4.5 Conclusion

The goal of the focus group discussion was to determine desirable features and functionalities for a data viewer system showing satellite-based radar backscatter data and using touchless mid-air hand gestures as an input system. From the analysis, three themes emerged and clear areas of agreement between the participants were identifiable. In addition, concepts drawn within a theme were found in other categories as well which increased the importance of those. Most participants were overwhelmed after presenting satellite-based radar backscatter data without any explanations. Therefore, additional information at different levels is necessary to orient the user and explain the shown data. Hence, functions for showing and hiding ancillary data are needed.

The concept of rummaging around and exploring global data as well as educational reasons were reported as important use cases for the emerging system.

All participants expressed the importance of an intuitive interaction system, both in existing applications and future implementations. This indicated, users expect that they can perform actions similar to a mouse or smartphone input system.

Conclusively, the following features and corresponding touchless mid-air hand gestures will be tested in the subsequent gesture elicitation study:

- Selecting
- Panning (left, right, up, down)
- Zooming (in and out)
- Contrast adjustment
- Back to initial view
- Superimpose and hide additional information

3.4.6 Discussion

Considering the emerged features based on the focus group discussion, results are in line with the work of Cockburn et al. [14]. They reviewed zooming, overview + detail and focus + context techniques and the empirical work that evaluated them. Especially the work of Guiard and Beaudouin-Lafon [25] is of special interest, which suggested that panning and zooming the view is used to bring the target in view. This is in line with the given statements of the participants who suggested zooming and panning as a natural way of interacting with the presented map.

Comparable to the current study process, the work of Morris [44] conducted a study with 25 participants, interviewing them to gain insight into scenarios in which they would like to use a web browser on their living room TV controlled with touchless hand gestures. In their post-study questionnaire, the participants indicated enthusiasm for a variety of tasks which can be compared to the here found concept of rummaging around and exploring such as *looking up facts and trivia*, *finding photos online* or *shopping online*. Those tasks were the basis of their subsequent gesture elicitation study.

Chapter 4

Gesture elicitation study (GES)

In the previous chapter, the process and results of the focus group discussion were presented. Based on the findings of the latter, a gesture elicitation study (GES) was designed and performed in order to obtain a consensus set of touchless mid-air hand gestures.

4.1 Gesture elicitation process

Designing an interactive system, which is based on touchless mid-air hand gestures, is intricate. The high degree of variability of input gestures, results in an indefinite number of possible gesture implementations. Hence, a GES, which is a widely used method in human-computer interaction research [69], was performed to identify a user-defined consensus set of touchless mid-air hand gestures.

A GES is an unified approach to gather useful information during the development process of interaction systems. Study participants are presented with a number of different *referents*, which are described as effects of an action. Then, the participants are individually asked to perform a gesture, which matches best or is most intuitive to bring that effect about. The collected gestures are then categorized into a set of taxonomies. Qualitative and quantitative analysis of the collected data results in a user-derived gesture set.

The crucial part of the GES is the user-centered design process by putting potential users at the center of gesture design. This brings up the advantage, that the technique is not limited to current sensing technologies, which enables the designers to focus on the participant's intentions and desires, instead of limiting the outcome to what is currently technically convenient.

4.1.1 Legacy Bias

The avoidance of constraints during a gesture elicitation process should reveal the user's most intuitive way of interacting with a referent. But due to possible prior knowledge and experience with other interfaces and technologies, such as WIMP (windows, icons, menus and pointing) or touch-based mobile interfaces, their gesture proposals are most likely biased [46]. This is called *legacy bias*.

As stated by Morris et al. [46], *„Legacy bias limits the potential of user-elicitation methodologies for producing interactions that take full advantage of emerging application domains, form factors, and sensing capabilities.“* To reduce the impact of legacy bias, Morris et al. [46] proposed several solutions like priming users or letting the users propose more than one gesture per referent. The latter is referred as *production*.

On the opposite, Köpsel and Bubalo [34] explained why interaction designers could benefit from legacy bias. As a paradigm of user-centered design, the user's comfort and satisfaction with a system is in the main focus, in order to avoid repelling them.

Gestures produced with legacy bias assures that the participants of a GES were not primed or strictly forced to propose one or more gestures. Hence, they intentionally or unintentionally take recourse to movements they already know from prior knowledge or experience with other devices or input modalities. This leads to the advantage, that the proposed gestures tend to be more simplistic than in a GES which used techniques to reduce the legacy bias.

The work of Vogiatzidakis and Koutsabasis [64] reviewed 47 studies for touchless mid-air interaction. From them, 15 studies (31.9%) adopted at least one technique to reduce legacy bias, while *production* was the most frequent one. Considering the explanations of Köpsel and Bubalo [34], the remaining 32 studies (68.1%) did not utilize a technique to reduce legacy bias.

In this study, *production* was used to get the best of both worlds. The goal was the evaluation and implementation of an interaction system based on a set of user-defined touchless mid-air hand gestures. Especially when implementing such system in a public domain, the initial hurdle should be minimal in order not to discourage first time users. Different people, either university personnel, students or external persons are inherently invited to approach the system. Therefore, the simplicity of the gestures was of major interest to provide an easy-to-use and easy-to-learn interaction system.

Additionally, the experimenter encouraged the participants to „see outside the box“ and propose as many gestures as they wanted to. By asking to define more than one gesture, the participants were forced to go through the situation described by the referent repeatedly. Hence, also non biased motions were proposed.

4.2 Study procedure

After the recruitment process, every participant of the GES got an e-mail message in which the informed consent form and study description was enclosed. Additionally, they received a link to a mandatory preliminary online survey and to a platform to schedule an appointment for the study [1]. After the participants provided their signed informed consent and privacy policy for the Research Unit Geoinformation, Department of Geodesy and Geoinformation, the experimenter set up a corresponding online meeting and sent the invitation link to the test person. The GES was conducted per participant as an online meeting using the tool *Microsoft Teams*. The whole study process was audio and video recorded using the free and open source software *OBS Studio* [2].

4.3 Participants

Thirty participants were recruited among undergraduate students, colleagues and other persons within the personal area of the experimenter. The average age was 27.97 years ($SD = 5.88$).

Since the emerging data viewer should present geographical data on a map-based web application, the question arose, if there is a difference in the participant's mental models based on their prior knowledge in handling and/or analyzing such data. To gain further insight into a possible difference, the pool of participants was split into a *expert* and a *non-expert* group. Their characteristics are explained by the following fictional user profiles, also referred as *personas*[15].

[1] `www.termino.gv.at` (Last accessed on 2021-04-10)
[2] `https://obsproject.com/` (Last accessed on 2021-04-10)

4.3.1 Personas

Non-expert

Paul is 32 years old and teaches Geography and Mathematics in a High School in Vienna, Austria. Due to his profession, he is an expert in creating presentations and learning documents for his 14 to 18 years old students. It is easy for him to explain to them the basics of regional as well as local phenomena for instance the global water cycle or a national economic system. In his free time, he is regularly searching for geocaches with his friends. For that, he owns a high-precision GPS-receiver as well as a smartphone to assist finding unknown locations.

Expert

Claire is 38 years old and has been working for the last ten years with the Austrian agency for meteorology and geodynamics. She studied the Bachelor and Master program Geodesy and Geoinformation at the University of Technology in Vienna. During her studies, she gained expert knowledge about how geographical data is derived from earth observation satellite observations. This includes processes such as the reprojection to two-dimensional image arrays and time series analysis. In her thesis she wrote about continuous interpolation of discrete station observations of surface soil moisture. Using geographical information software (GIS) *QuantumGIS* at expert level and algorithms written in *Python*, she is able to plot and analyze different characteristics of high-resolution image data. After work, she likes to go to restaurants with her boyfriend. To find new locations or to get navigational information, she regularly uses *Google Maps* on her smartphone.

4.3.2 Preliminary survey

Demographic information was gathered by two closed-ended questions about the participant's age and gender. Nine Likert scale questions with a five-point scale split into two subsections and two open-ended questions were asked to gain a deeper insight into prior knowledge and experience relevant for the GES.

The first sub-category covered questions about prior knowledge with map-based data viewers. Examples of them are online map services such as *Google Maps* or *OpenStreetMap* as well as geographical information system software such as *Quantum GIS* or *ESRI ArcGIS*. Additionally, the participants were asked regarding their preferred device for performing search queries to map services. The second sub-category covered questions about prior experience with computer systems based on touchless interaction.

It was found, that all participants use an online map service to find an unknown location or, less frequently though, rummage around interesting places like the next holiday destination. Also, nearly all participants use mainly a touchscreen-based devices such as smartphones or tablets for the latter. Only one participant reported never or very rarely using a touchscreen-based device. In comparison, the usage of a desktop-based systems is less common. A possible explanation to the latter is that people are usually on the go when they are in need of navigational or location information. The pre-selection of the participants based on the defined personas was also reflected in the survey: 14 participants never or very rarely use GIS software.

The second sub-category gathered information about prior knowledge and experience with systems based on touchless hand gestures. 21 participants at least once played video games on hardware like *Microsoft Kinect* or *Nintendo Wii*, which are based on motion sensing input devices. Five participants performed touchless hand gestures while using an Augmented- (AR) or Virtual-Reality (VR) system. Not surprisingly, no participant has ever used touchless hand gestures controlling a map-based data viewer. However, 12 participants stated that they have at least once interacted with a touchless hand gesture system in combination with a public display such as installations in museums.

Furthermore, six participants additionally stated in the open-question, that they have performed touchless hand gestures either with a Theremin, gesture recognition on a smartphone, smart TV or controlling music in a car.

4.4 Referents

Each referent was a compound of two pictures (three for *Back to start*) representing a *before-after* situation, respectively. Every picture had a headline describing the referent and its current state (*before* or *after*). Since all recruited participants spoke German, the study was carried out in German to avoid language barriers. Figure 4.1 shows the referent *Zoom in* with its before and after image. For reference, the complete set of referents is saved as image files in the external directory, which is denoted in the appendix section.

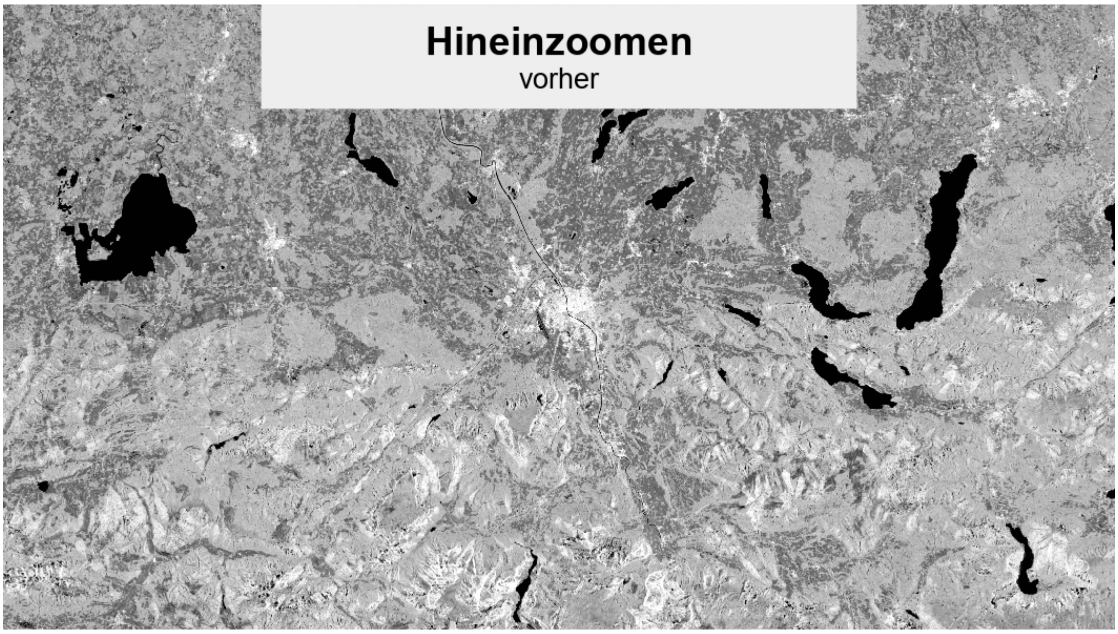
Based on the results of the focus group discussion, twelve tasks for the elicitation study were selected:

1. Zoom in
2. Zoom out
3. Pan
4. Increase contrast
5. Decrease contrast
6. Select rectangular area
7. Show satellite image
8. Show country borders
9. Combination of zoom and pan
10. Show menu
11. Select a button in the menu
12. Back to start

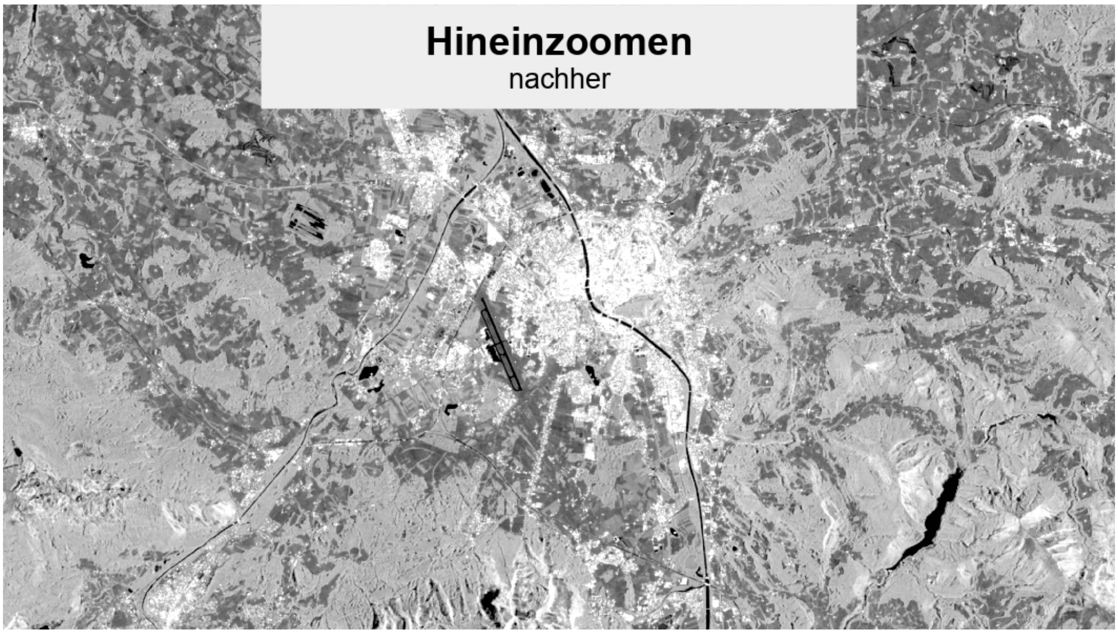
The referents were designed to optimally reflect the situations as discussed in the focus group. Therefore, the satellite-based radar backscatter map presented in the existing data viewer (see figure 3.2) was the base map of each referent. This data is available as a *XYZ Layer* based on image pyramids, which can be loaded from the internal server infrastructure of the Department of Geodesy and Geoinformation.

Each referent was created using the GIS software *Quantum GIS*(*QGIS*). QGIS allowed us to load map-based data layers as well as additional layers such as country borders or optical satellite imagery onto the canvas. Furthermore, the contrast adjustments were done by basic image processing tools provided by QGIS. The *Menu Design Guidelines* [3] by Ultraleap [60] were taken into account to design the buttons and menus shown in the respective referents.

[3] https://developer-archive.leapmotion.com/documentation/python/practices/Leap_Menu_Design_Guidelines.html (Last accessed on 2021-04-10)



(a) Before



(b) After

Figure 4.1: The referent *Zoom in* with its before and after image.

4.5 Experimental setup

After starting the online meeting, the experimenter welcomed the participant and made sure, that the video and audio was working flawlessly. A prepared script (see attached document in the appendix - `ges_study_procedure_text.pdf`) explaining the study and each referent was used, to increase the consistency of the study. This script was presented equally to each participant. Before the start of the actual study, an exemplary iteration with one referent (*Pan*) was carried out. In order to achieve a situation which is resembling the planned setup of the emerging prototype as good as possible, the participant was asked to stand up and perform the hand gestures in the area in front of the upper body. Subsequently, the *before-after* structure of the referents were explained and potential questions of the test person were answered.

The *Combination of zoom and pan* referent was tested last, because it is a combination of the two previous tested *Zoom* and *Pan* referents. The *Select a button in the menu* referent is based on the the *Show menu* referent. Therefore, the two were considered as a single task and always tested in the same order.

In order to counterbalance immediate sequential effects, a 10×10 Latin square design of the referents was used [9].

After finishing the exemplary iteration, the main part of the gesture study started. After each gesture proposal for the given referent, the experimenter asked the participant to rate the proposed motion using a five-point Likert scale on the following questions:

- How difficult was it to come up with the gesture? (Q1)
- How well do the gesture and the triggered action match? (Q2)
- How difficult was the gesture to perform? (Q3)

Subsequently, the participants were asked to describe the mental creation process and if a similar motion was already used in another context or with another device.

As soon as the gesture elicitation process ended, the following final questions were asked. Those were expected to help concluding the study and gaining further insights about the mental model of the participants.

- Were there any interactions for which it was very difficult to find a gesture?
- If yes, would you have liked to see more elements (e.g. menus, buttons, etc.) to perform the interaction?
- Do you have additional comments or thoughts?

Figure 4.2 shows a screenshot during a video of a gesture elicitation process with a participant. The left part is displaying the shared screen of the experimenter which showed the images and questions of the referents to the test person. The right part is the video recording. The participant is currently performing a hand gesture for the *zoom in* referent.

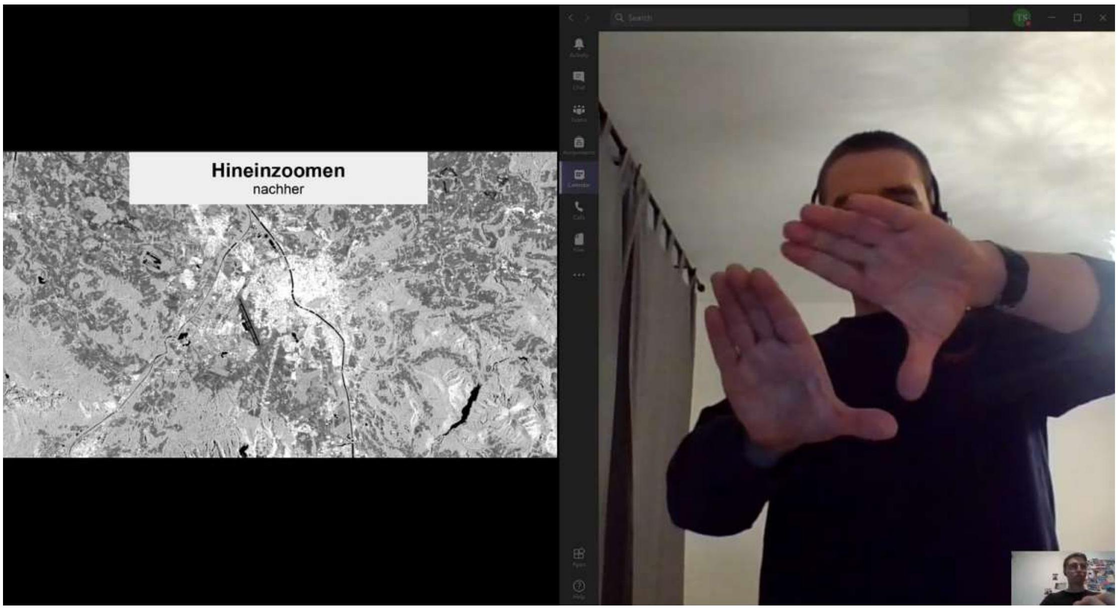


Figure 4.2: A screenshot during a video of a gesture elicitation study process with a participant. The person pictured has given consent to the use of this image.

4.6 Analysis

During the gesture elicitation process, the experimenter wrote down the quantitative subjective ratings as well as qualitative statements of the gesture creation process. After collecting all the gestures, the proposals were sorted by referent and order of appearance. The first and second gesture proposals were subsequently classified in taxonomies explained in section 4.6.1.

For the qualitative analysis, each video and audio recording was reviewed in order to find overlapping narratives among participants. The test person's mental creation process description during and after each gesture proposal was written down in key words. By combining the gestural motions with the corresponding narratives of the participant, further categorization and classification of the proposals could be reached.

This was particularly necessary for referents which produced *non-responses*. In the GES, a non-response was declared if a test person did not produce a gesture as a first guess for a specific referent.

4.6.1 Mid-air Gesture Classification

The main classification dimensions were originally defined by Wobbrock et al. [70]. Based on this work, Hoffmann et al. [28] adopted their taxonomies for their study which compared voice, touch and mid-air commands. Their taxonomies for mid-air commands were used for this study. However, their taxonomies did not cover static gestures where the user does not move the hand(s) in any direction. Hence, the *dimension* taxonomy was slightly refined by adding *static pose* to also cover gestures without any movement involved. The final gesture taxonomy categories are listed in table 4.1.

The *nature* dimension helps to distinguish gestures which are either *symbolic*, *physical*, *metaphorical* or *abstract*. An example for a symbolic gesture is showing the hand palm in direction to the screen to signalize to *stop* an action. A *physical* gesture implicates a direct manipulation of a visible object on the screen. An example is moving the flat hand in a desired direction to pan the visible map. The flat hand symbolizes the center of the map, like moving a printed map on a table. An example of a *metaphorical* gesture is placing the flat hand on the chest denoting a metaphor like „come back to me“in order to reset the map extent to the initial state, swiping as if to push away the first layer of a stack, or pushing an imaginary button with a flat hand. Finally, gestures which did not fit in any other *nature* categories were defined as *abstract*.

Responses of the system *after* or *during* the gesture are summarized in the *flow* dimension. An example for a *discrete* gesture would be a selection of a button - the feature which is controlled by the button will be started *after* the selection. Whereas a *continuous* gesture would be bringing together both hands in order to zoom out of the map.

The *context* dimension describes if the gesture can be performed independently or requires a specific context shown on the screen. For example, moving a flat hand vertically to increase or decrease contrast is *in-context*, whereas double-tapping with an index-finger to decrease the contrast, or clapping twice to reset the map extent is defined as *no-context*.

The *interaction* dimension compares if the gesture is performed with one or both hands.

The *dimension* describes how many axes were involved in the movement. A *static pose* does not have any movement in any direction such as posing the flat hand against the screen. For example, just rotating the wrist is classified as *single-axis*, the translation of a flat hand or drawing a circle with an index-finger is happening along *tri-axis* and a combination of the last two is defined as *six-axis*.

The *position* dimension defines the state of the hand and fingers when the gesture is started. The categories are *open* or *closed* hand as well as *single* or *multiple* fingers.

The *movement* dimension describes if the finger position changes. For example, opening a closed hand (fist) is referred as *movement*.

The *complexity* dimension discerns *simple* and *compound* gestures.

Nature	Symbolic	Gesture visually depicts a symbol
	Physical	Gesture imitates a physical action
	Metaphorical	Gesture indicates a metaphor
	Abstract	Gesture is arbitrary
Flow	Discrete	Response occurs *after* the gesture
	Continuous	Response occurs *during* the gesture
Context	In-context	Gesture requires specific context
	No-context	Gesture does not require specific context
Interaction	Unimanual	Gesture performed with one hand
	Bimanual	Gesture performed with both hands
Dimension	Static pose	No motion along any axis
	Single-Axis	Motion around a single axis
	Tri-Axis	Translational hand motion or wrist rotation
	Six-Axis	Translational hand motion and wrist rotation
Position	Open hand	Gesture started with a open hand
	Closed hand	Gesture started with a closed hand (fist)
	Single finger	Gesture started with one stretched finger
	Multiple fingers	Gesture started with one or more stretched fingers
Movement	No movement	No change in finger position
	Movement	Change in finger position
Complexity	Simple	Gesture consists of a single gesture
	Compound	Gesture can be decomposed into simple gestures

Table 4.1: Mid-air hand gesture taxonomy.

4.6.2 Agreement rate

The definition of the agreement rate (eq. 4.1) was first introduced by Wobbrock et al. [69] and later refined by Vatavu and Wobbrock [61]. According to Villarreal-Narvaez et al. [63], the agreement rate has been widely used in gesture elicitation studies. It is a numerical measure quantifying the agreement among gestures proposed by several participants. $AR(r)$ values ranges between 0 and 1, with 0 denoting total disagreement and 1 absolute agreement. $AR(r)$ is defined as follows:

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} \left(\frac{|P_i|}{|P|} \right)^2 - \frac{1}{|P|-1} \tag{4.1}$$

where $|P|$ denotes the total size of the set (30 gestures from 30 participants), P is the set of all proposals for referent r, and P_i subsets of identical gesture proposals from P.

Vatavu and Wobbrock [61] computed probability distribution functions of AR for various numbers of participants from 10 to 50. They showed that by increasing the number of participants, the peak of the probability distribution functions is shifting toward lower AR values. Hence, referents for which participants declared non-responses, were excluded for the determination of the agreement rates since they would not be comparable to referents without non-responses.

4.7 Results

With 30 participants and ten elicited tasks, a total of **462** gesture proposals were collected. **273** of them were first, **153** second and **36** third proposals. The first proposals were classified into **90** different gestures. In figure 4.3 the proportions of all first gestures (in percentage %) in each category of the taxonomy are shown.

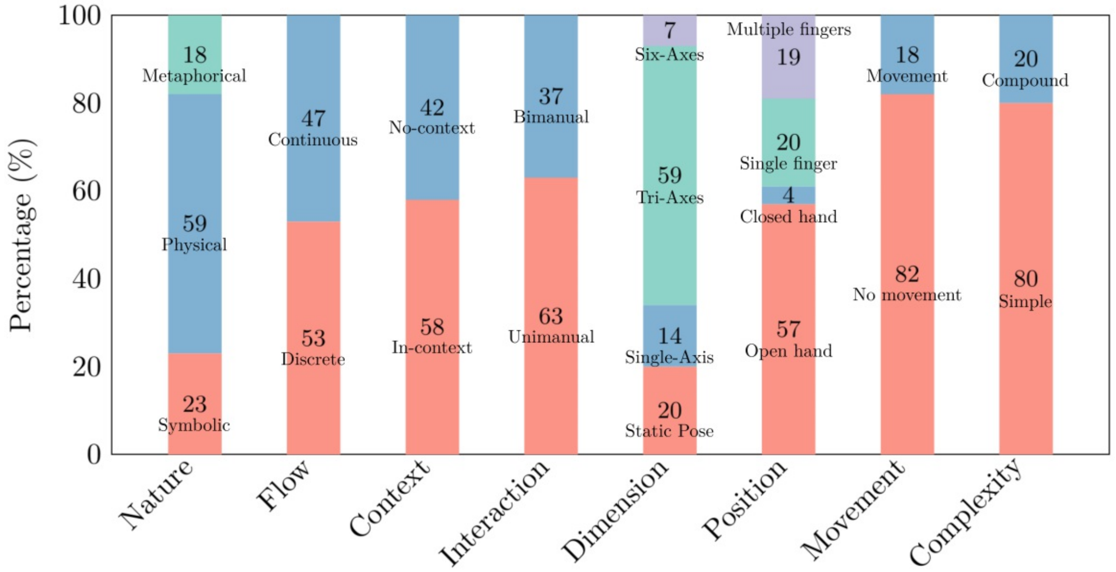


Figure 4.3: The proportion of gestures in each category of the taxonomy.

As shown in figure 4.3, most first gestures were physical in nature (59%), meaning that test persons were implicating that they were manipulating a physical object. Furthermore, all gestures could be classified in either symbolic, physical or metaphorical nature which resulted in zero abstract gestures.

The flow taxonomy is quite balanced with 53% classified as discrete and 47% as continuous gestures. Further analysis showed, that this is highly referent dependent. For instance, all captured gestures for the task *show country borders* were classified as discrete, whereas all *zoom* gestures were continuous.

Most of the gestures were classified as in-context (58%), meaning that the user required specific context presented on the screen such as moving a flat hand vertically to adjust the contrast.

The majority of the proposed gestures were performed using only one hand (63%). But also further analysis showed that this is highly dependent on the referent. For example, 24 of the 30 participants suggested a two-handed gesture for *zooming*.

Considering the movement of the gesture, which is covered in the dimension taxonomy, a majority of the suggested gestures (59%) was performed involving tri-axes. The second most frequent dimension was static pose with 20%.

A majority of 57% were started with an open hand, whereas only 4% were started with a fist. 20% of the gestures were started with a stretched single finger and 19% with multiple fingers. An example of the last would be stretched index-finger with a thumb depicting pliers.

Most of the gestures (82%) did not have any movement in the state of the hand or fingers involved. Also, 80% were classified as simple gestures, meaning that they could not be further simplified.

4.7.1 Consensus set

Based on the video and audio recordings as well as subjective analysis of each of the proposals a user-defined gesture set for touchless mid-air hand gestures was derived. For each individual referent which did not produced non-responses (*Zoom in*, *Zoom out*, *Pan*, *Show menu*, *Select a button in the menu* and *Select rectangular area*) the agreement rates ($AR(r)$) according to section 4.7.1 were calculated to measure the consensus between participants.

In table 4.2 the corresponding agreement rates and the subjective ratings as weighted means for the first gesture proposals are shown. As described in equation 4.1, P denotes the number of different gestures classified for the respective referent.

Referent	P	**AR**	**Q1**	**Q2**	**Q3**
Zoom in	6	0.545	4.70	4.76	4.76
Zoom out	6	0.545	4.70	4.76	4.76
Pan	5	0.398	4.94	4.94	4.87
Show menu	9	0.280	4.76	4.57	4.87
Select a button in the menu	5	0.333	4.73	4.60	4.77
Select rectangular area	9	0.161	4.43	4.45	4.43

Table 4.2: Agreement rates and weighted means of subjective ratings of referents without non-responses.

The referents with the highest agreement rates were *zoom in* and *zoom out* with $AR(r) = 0.545$, respectively and *pan* with $AR(r) = 0.398$. *Select rectangular area* received the lowest agreement rate with $AR(r) = 0.161$ and nine different proposals among all 30 participants. The subjective ratings, which were described in section 4.5, are in general very high with 5 denoting the best and 1 the worst possible value.

After classifying the first gesture proposals and calculating the agreement rate for referents without non-responses, focus on the second or even third gesture proposals was given. Considering additionally the participant's qualitative statements agreement could also be analyzed for referents with non-responses.

Referent	**# Non-responses**	**# Different gestures**	**Q1**	**Q2**	**Q3**
Increase contrast	11	8	3.68	3.79	4.69
Decrease contrast	12	8	3.75	3.70	4.76
Show country borders	16	9	3.21	3.57	4.64
Show satellite image	15	11	3.72	3.83	4.68
Back to start	10	14	3.80	3.45	4.75

Table 4.3: Number of non-responses, number of different gestures of the remaining first proposals and weighted means of subjective ratings for referents with non-responses.

Table 4.3 shows the number of non-responses, number of different gestures of the remaining first proposals and the weighted means of subjective ratings for each referent with non-responses. In comparison with the referents without non-responses, the lower means of the subjective ratings show an overall higher complexity of the tasks. If the participants proposed a first gesture, they reported a more difficult creation process due to the complexity of the task. Often, they were not satisfied with their own proposals which was then reflected in the subjective ratings.

In table 4.4 the agreement rates and weighted means of subjective ratings of the second gesture proposals are shown. Additionally to table 4.2, N indicates the number of given proposals. Note that the agreement rates in this table are calculated on basis of the number of given proposals and are therefore not comparable to the agreement rates presented in table 4.2.

Referent	P	N	**AR**	**Q1**	**Q2**	**Q3**
Zoom in	8	17	0.221	4.35	4.35	4.59
Zoom out	5	13	0.218	4.15	4.38	4.69
Pan	4	6	0.200	4.17	4.67	4.83
Increase contrast	8	11	0.073	3.00	2.91	4.46
Decrease contrast	9	12	0.061	3.42	3.83	4.42
Show menu	9	16	0.108	4.13	4.25	4.81
Select a button in the menu	4	11	0.200	4.36	4.00	4.82
Back to start	12	19	0.058	3.26	3.42	4.58
Show country borders	9	14	0.077	2.93	3.43	4.57
Select rectangular area	9	20	0.126	4.20	4.05	4.35
Show satellite image	11	14	0.044	3.29	3.57	4.29

Table 4.4: Agreement rates and weighted means of subjective ratings of the second gesture proposals. N indicates the number of given proposals.

In the following, the consensus analysis is described separately for each referent. For reference, a full list of gestures proposed by the participants is saved in an external text document (`ges_gesture_classification.ods`) specified in the appendix.

Zoom

In the current GES, *zoom* was tested separately for *zoom in* and *zoom out*. Since both representing a semantically equivalent action, no significant difference in the gesture classification was found. Hence, the AR as well as the qualitative ratings are the same for both tested referents. The most commonly suggested gesture (22 participants) for *Zoom in* was both hands starting together with subsequently moving apart. The reversed gesture, both hands with some distance between brought together is translated as *Zoom out*. A majority of the second gesture proposals (8 out of 17) suggested the same aforementioned gestures. In total, six different first gestures were proposed.

Pan

A similar high agreement was reached for the *Pan* interaction. 18 participants suggested panning with an open hand, whereas additional five used the same movement with a pointing index-finger. Based on the taxonomy described in section 4.6.1, those two were classified as different gestures. But qualitative analysis showed, that participants had the same narrative in mind, which further strengthens the overall consensus. In total, five different first gestures were proposed.

Show menu

For activating or showing the menu, a „burger“-menu button was shown on the referent as an additional user interaction element. Compared to the latter referents, the agreement rate is dropping to 0.280, due to the high number of nine different proposals. A majority of 15 participants suggested a simple button tap with the index-finger as their favorable gesture. The second most commonly suggested gesture (six) was holding a flat hand above the button for a short period of time. Participants stated that they would expect the location of their hand/finger shown on the map.

Activate legend button

As already mentioned above, the referent of activating a button was always tested after the *Show menu* referent. Again, the most commonly suggested gesture was simply tap the button with the index-finger (15 participants). Nine participants used a flat hand to tap the button. In contrast to the *show menu* results, several participants mentioned that it is not important that the exact position of their hand/finger is printed on the screen. Instead, they would expect, that a button is highlighted according to the horizontal position of their hand/finger. The gesture of holding three stretched fingers to the screen was suggested by four participants as a first gesture and four as a second gesture, since the highlighted button was the third counted from the top. This can be counted as the third most commonly suggested gesture.

Increase and decrease contrast

Adjusting the contrast of the map can be seen as a dichotomous action with increasing and decreasing a value. Therefore, like *zooming*, this task was tested separately for *increase contrast* and *decrease contrast*. Due to the dichotomous form of those two referents, similar gestures were proposed. The most commonly suggested gesture (eight participants) was moving a flat hand either from top to bottom (*decrease*) or vice versa (*increase*). But a majority of twelve participants clearly stated that they would expect a certain type of user interface element as their first proposal. The remaining ten participants proposed seven different gestures. Overall, 20 test persons mentioned at either during testing or qualitative statements, that they would expect a vertical slider. Hence, a clear qualitative consensus was found between all participants, resulting in a combination of presenting the user an additional interaction element such as a vertical slider and the proposed gesture of moving a hand from top to bottom and vice versa.

Show country borders

Due to the high number of 16 non-responses, main focus was given on the participant's explanations and narratives to find possible agreement. Those stated that they would prefer tapping a button on the screen or through a menu. If no additional element would be present on the map, they would try to symbolize a border such as drawing a circle with an index-finger, symbolizing a circle with the thumb and index-finger of both hands or draw a zig-zag line. Additionally, six participants (three as a first, and three as a second proposal) suggested a narrative of putting an additional layer on top of the map. The remaining 14 participants suggested nine different first gestures. Consequently, from a gesture point of view, no significant agreement was found. But qualitative analysis suggested that participants agreed upon a similar narrative of symbolizing a border. This finding can be used to enhance the visual appearance of the interaction elements or gestures as supporting shortcuts.

Show satellite image

Similar to the latter referent, gesture proposals were diverse for the *Show satellite image* referent. 15 participants firstly suggested to use an interaction element (button or menu), which was classified as non-responses. The remaining 15 participants proposed eleven different gestures. Among them, seven suggested a swipe gesture with an open hand. Qualitative statements and the analysis of the second proposals showed that the consensus narrative was *changing* or *adding a map layer*. In total, 19 proposals suggested a layer change with different gestures such as the aforementioned swipe with an open hand or drawing a circle imitating a rotary button. Consequently, the participants did not agree upon a single gesture, but on a narrative.

Select rectangular area

The calculated *AR* with 0.161 is relatively low, due to the fact that two different gesture proposals were dominant. The most commonly (ten) suggested gesture was symbolizing a rectangle with thumb and index-finger of both hands to start the action and subsequently moving apart to adjust the size of the rectangle. The second most commonly (six) suggested gesture was choosing two diagonal corner-points of the rectangle with the index-finger. It is important to note, that the participants suggested to activate this function by tapping on a corresponding button. Compared to the other referents with non-responses, this case was not counted as non-responses because participants suggested a combination of a gesture with using a button or menu. The common narrative is to draw a rectangle, but participants were not consistent in how to start the action. Therefore, selecting the function through a menu or button and a subsequent placing and adjusting of the rectangle with the most commonly suggested gesture seemed to be reasonable.

Back to start

Gesture proposals for the referent *Back to start* were very varying with 14 different gestures by 20 participants. The remaining proposals were classified as non-responses. Qualitative analysis suggested three different narratives: 1) *rewind*, *undo* or *one step back*; 2) *home*, *homebutton* and 3) *termination*. Compared to the other referents, participants motions were more static and symbolic. They symbolized a roof with both hands, a “X„with both arms or showed their flat hand against the screen. Consequently, no significant quantitative consensus was found. Qualitatively, participants referred to a *home*- or *rewind*-button, which seemed to be the favorable solution.

Combination of Zoom and Pan

The referent of *combination of Zoom and Pan* was presented lastly in every round. As expected, all test persons did, in fact, a combination of their elicited gestures for *Zoom* and *Pan*. At first, participants zoomed out, then panned and subsequently zoomed in with their already suggested gestures.

4.7.2 Subjective final evaluations

After each gesture elicitation process, three final questions were asked to conclude the participant's study session. A majority of 26 of the 30 participants stated that they would expect more interaction elements on the user interface such as a menu or buttons. Particularly for referents which produced non-responses, participants asked for further elements of the user interface. The rationale behind this statement is that on the one hand, one would not even know that a specific feature is implemented and therefore would not propose a gesture to utilize the latter. On the other hand, for complex features such as showing country borders or adjust contrast, one would not expect that a gesture was even implemented and immediately search for a menu. For adjusting the contrast, participants would look for a slider, since they are often visually implemented in image processing software or even in smartphone applications like *Instagram*.

Furthermore, five participants suggested the implementation of a tutorial. This should help the end-user to familiarize with the input modality of touchless mid-air hand gestures and features of the emerging data viewer.

Shneiderman et al. [57] stated that feedback of a user interface is vital to signalize that a user's input has been successfully issued. Also in the current study, the majority of the participants would prefer to have visual feedback on the position of their hands or fingers.

4.7.3 Mental modal observations

Especially for the first gesture proposals participants often referred to prior experience with existing interaction systems based on either touch or mouse input (legacy bias). Their movements were transferred from (multi-)touch gestures used on touchscreen-based devices with their fingers to *bigger* movements with the whole hand or both hands in order to fit their mental model of a *bigger* screen in front of them.

4.7.4 Experts versus Non-experts

The numbers of proposed gestures, either for the first or thereafter gestures, did not differ significantly. For the first proposals, 51.7% of a total of 273 gestures were made by experts, 48.3% by non-experts. Similarly, 50.3% of a total of 153 second proposals were made by experts, 49.7% by non-experts. Differently, a majority of 63.9% of a total of 36 third proposals were made by non-experts, 36.1% by experts.

4.8 Discussion

During the gesture elicitation process, it was emphasized that the test person should imagine that the data viewer is presented on a large screen in front of them. As stated by Ostkamp and Kray [52], public displays always have a certain *situatedness* which can only be simulated partially. However, the work of Magrofuoco and Vanderdonckt [41] showed that gesture elicitation studies can be held distributed in time and space using a cloud platform. Hence, conducting this GES virtually can be seen as a workable trade-off between the enacted access restrictions during the Covid-19 pandemic and limitations regarding situatedness.

Wittorf and Jakobsen [68] derived a user-defined gesture set where 20 participants performed mid-air gestures on a three-meter wide display. Just as the results of the present study, their resulting gestures were largely influenced by surface interaction as well as a tendency of being larger than gestures elicited for smaller displays. Particularly for the referents *zoom* and *pan* participants of the current study tend to use this notion to simply enlarge their legacy biased motions from a touchscreen input to a touchless input with a large screen. Additionally, they found that their majority of the proposed gestures (55%) were physical in nature, which is similar to the present result (59%).

The work of Gentile et al. [23] focused on touchless mid-air hand gestures for zooming. The gestures for the *zoom* referents elicited in the current study, reached higher agreement rates compared to their results. Comparing the interaction taxonomy, 25 of the 30 participants of the current GES performed two-handed first gestures, which is inline with their work.

After a comprehensive analysis process of the GES, no significant differences in the gesture making process between experts and non-experts was found. This finding is also reflected in the similar proportions of the number of proposed gestures. Subjective assessment by the experimenter saw a slightly more creative or dynamic gesture making process for non-experts. Comparatively, experts more often described a mental model of a „toolbox“based on image processing or GIS software, when they subjectively evaluated their gestures or non-responses. Summarized by Villarreal-Narvaez et al. [63], existing literature to date did not differ between expert levels or prior experience during the recruiting phase. Therefore, no study was found to compare the results.

For the first gestures, participants clearly tend to propose easier gestures than for the second or third proposals which is also reflected in the dropping subjective ratings for gestures after the first proposal (see table 4.4). Compared to Morris et al. [46], who's pilot study showed that on average the user is rather satisfied with the third proposal than with the first, no preference for the second or third proposal was declared. Interestingly, participants seemed to establish a self-made bias during the study process. Participants discarded ideas, because they did not want them to overlap with preceding commands.

The subjective final evaluations of the test persons clearly suggested the implementation of interaction elements on the user interface such as buttons or menus. In the study of Köpsel and Bubalo [34], participants had to learn abstract gestures for interacting on a multitouch display and execute them for four sessions. The users showed increased performance from session to session and also subjective task demand and frustration decreased. However, the authors did not test UI elements specifically, their findings coincides with the results of the GES described here. Taking advantage of legacy bias helps to reduce the cognitive load learning a new gesture. This, suggests to support the user with UI elements and easy to repeat selection gestures with, for instance, a single tap with a finger.

Considering the participant's qualitative statements for the *back to start* referent, some participants stated that they would avoid dynamic gestures in order to prevent panning the map. This supports the implementation of a button controlling the feature of resetting the map view.

Chapter 5

Prototype

After analyzing the focus group discussion and the gesture elicitation study, the gained insights were the basis of the implementation of the prototype. In a nutshell, the emerging system should provide map data of global earth observation data sets and be controlled with touchless mid-air hand gestures. The implemented functions and features were as follows:

- Map pan
- Map zoom
- Back to start
- Rectangle zoom
- Map layer change
- Adjust contrast
- Tutorial

The need of presenting additional information, which was denoted in the focus group discussion, was implemented by adding additional map layers, which can be changed.

5.1 Setup

The prototype was implemented as a web map application that can be run on every computer system which meets the following requirements.

5.1.1 Hardware and software requirements

The web application was created based on *Hyper Text Markup Language (HTML)*, *Cascading Style Sheets (CSS)* and *JavaScript*. A local web server instance was necessary in order to run the prototype on a local machine. While developing, a local *SimpleHTTPServer* instance based on *Python* was installed on a Linux based host computer. On a host computer which uses Microsoft Windows as operating system, the open-source software *XAMPP* creates a local web server based on Apache. Furthermore, a working internet connection was needed to retrieve map data from remote web servers.

For tracking and perceiving mid-air hand postures and gestures, the *Leap Motion Controller (LMC)* (see section 5.1.2) was used. The manufacturer provided a software development kit (SDK) inclusively the needed driver for the sensor. Both the SDK (*Leap Motion Orion v3.2.1 SDK*) and the driver needed to be installed on the host computer in order to be able to retrieve tracking data. The LMC was connected with the host computer via an USB cable.

To access the tracking data in a web application, the Leap Motion JavaScript application programming interface (API) was used which was provided as a standard JavaScript library called *LeapJS*. LeapJS is an open-source project and distributed separately from the main Leap SDK via a GitHub repository[1]. To use LeapJS, the `leap.js` file must

[1] `https://github.com/leapmotion/leapjs`

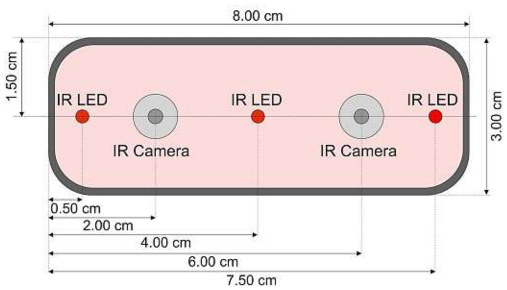
be included within the head of the HTML script of the web application. After the installation of the Leap Motion SDK, the LMC sends tracking data through a WebSocket server connection. LeapJS managed the Websocket communication and evaluated the data into proper JavaScript objects.

The design of the prototype was optimized to work with any display which supports a resolution of 1920x1080 pixels.

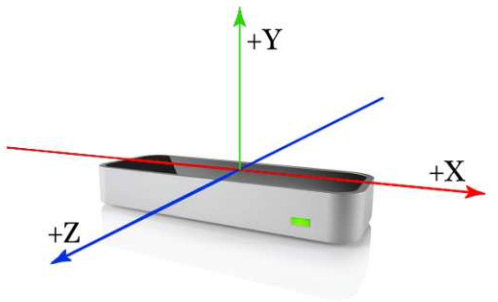
5.1.2 Leap Motion Controller (LMC)

The LMC is a consumer-grade optical hand tracking device which captures movements of user's hands and fingers developed by Ultraleap [60]. It was primarily designed to provide accurate three-dimensional coordinates of the hands and finger joints positions for interactive software applications. The sensor uses three infrared emitters and two CCD cameras. The LMC employs a right-handed Cartesian coordinate system with its origin on the top center of the device. The LMC's view of hands, a schematic overview and a rendered image with its coordinate system is shown in figure 5.1.

(a) The LMC's view of hands [60].



(b) A schematic overview of the LMC [67].



(c) A rendered image of the LMC with its coordinate system [60].

Figure 5.1: The Leap Motion Controller (LMC).

The sensor's field of view is an inverted pyramid of about 150° with an effective range from approximately 25 to 600 millimeters above the device. The tracking information is provided as a set of data for a single moment in time (*frame*) with a speed of up to 200 frames per second [26].

For the derivation of the information about identity, position and other characteristics of the detected hands in its field of view, the Leap Motion software uses an internal model of a human hand. Therefore, it is able to provide predictive tracking even when parts of a hand are not visible. Despite the fact that the LMC can detect more than two hands or hand-like objects, it is recommended keeping at most two hands in the field of view. All calculations are performed on the host computer using a proprietary algorithm. Values are reported in units of real world millimeters.

5.2 User interface

5.2.1 Tutorial

As stated by Limerick [39], two concerns are emphasized especially when designing for public displays: 1) catching the attention of passer-bys, and 2) conveying interactivity with the public system.

In order to overcome these issues and help a new potential user to get familiar with the novel interaction type of touchless mid-air hand gestures, a tutorial was introduced. By (animated) images the possible gestural interactions were presented.

Furthermore, the new user needed to complete the tutorial by selecting the buttons on each corner. Accomplishing that, the user should gain a feeling in how sensitive and in which physical range the hand motions needed to be in order to successfully use the prototype. Figure 5.2 shows the tutorial.

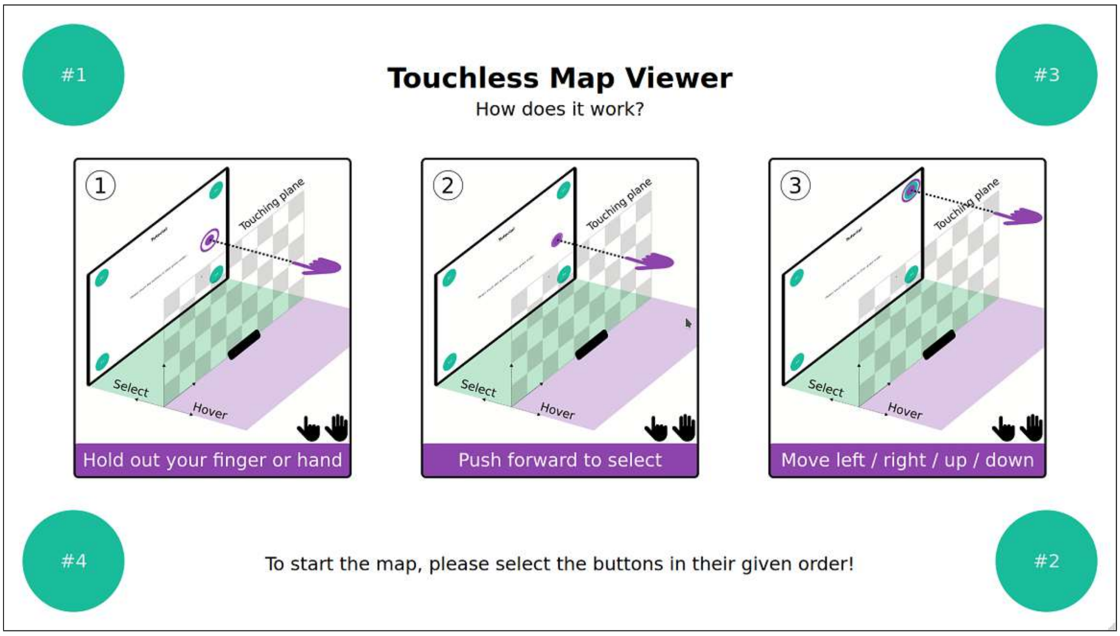


Figure 5.2: The tutorial presented on the TV screen in front of the participant. Figure (2) and (3) were implemented as animated GIFs showing the test person how to select a button and move the cursor.

5.2.2 Touchless Map Viewer

The user interface consisted of a set of buttons and the map layer as the central object of the prototype. Figure 5.3 shows the starting screen which was presented to the user after completing the tutorial. Starting from the upper left corner, the buttons are described as follows: 1) Homebutton, 2) Rectangle-zoom-button, 3) Layer-button, 4) Contrast-button and 5) Exit-button.

The colors of all UI elements were derived from *Flat UI Palette V1* accessible through the website by Flat UI Colors [20]. Additionally, Ultraleap [60] provided a comprehensive documentation[2] on usability tips, which helped to design the appearance and behavior of UI elements.

Figure 5.3: The starting screen, after completing the initial tutorial.

The icons on the buttons were created using the open-source vector graphics editor Inkscape [31].

The qualitative analysis of the GES described in section 4.7.1 resulted in consensual narratives among participants for specific tasks. For instance, the metaphor of a house representing a *homebutton* like it is implemented in common web browsers was therefore used also for the here implemented homebutton. Furthermore, some participants of the GES proposed a static gesture showing a circle with one or both hands. Other test persons were describing their proposed motions as *enclosing* or *bordering*, which can also be metaphorically referred to a circle. Therefore, the icon of the Layer-button will be enclosed with a circle, when the country borders map layer is activated.

[2]https://docs.ultraleap.com/touchless-interfaces/usability-tips/

Map layers

The open-source JavaScript library *LeafletJS* [1] was used as basis for the map implementation and its operations. With its powerful API, Leaflet provides numerous interaction events and map operations. Most importantly, it supports an easy integration of *tile layers* as map layers.

The following map layers were included in the prototype:

1. Global Backscatter Model
   (a) Sentinel-1 VV
   (b) Sentinel-1 VH
2. OpenStreetMap by OpenStreetMap contributors [51]
3. Satellite and aerial imagery in natural colors by Esri [18]
4. Country borders by Natural Earth [47]

The map layers *OpenStreetMap*, *satellite* and *country borders* are denoted as additional information, which was described as desirable feature in the focus group discussion.

The map layers are accessible as *XYZ tile layers*. *XYZ tile layers* are comprised of tiled image pyramids which are saved in the *Portable Network Graphics* (PNG) raster format. The tile layers are made available through web servers using a specific URL format (`https://serveradress.tld/{z}/{x}/{y}.png`).

The *country borders* layer could be overlapped on each map layer. It was provided as a local GeoJSON file which contains polygons for all the world's countries [47].

5.2.3 User interaction

The user interaction was based on touchless mid-air hand gestures. The positions and movements of the user's hands were tracked by the LMC and provided as JavaScript objects through LeapJS. As already mentioned, the data was provided in frames. Each frame contained the needed tracking data and was therefore the main JavaScript object to work with.

As a starting point, a `Controller()` object needed to be created which set up the LMC and its WebSocket connection. The Controller class was the main interface to the LMC. With a callback function, the main function `onFrame()` was invoked per frame.

Function onFrame()

The function `onFrame()` was the main algorithm which was called on each frame provided by the LMC. An overview of the structure is shown in figure 5.4.

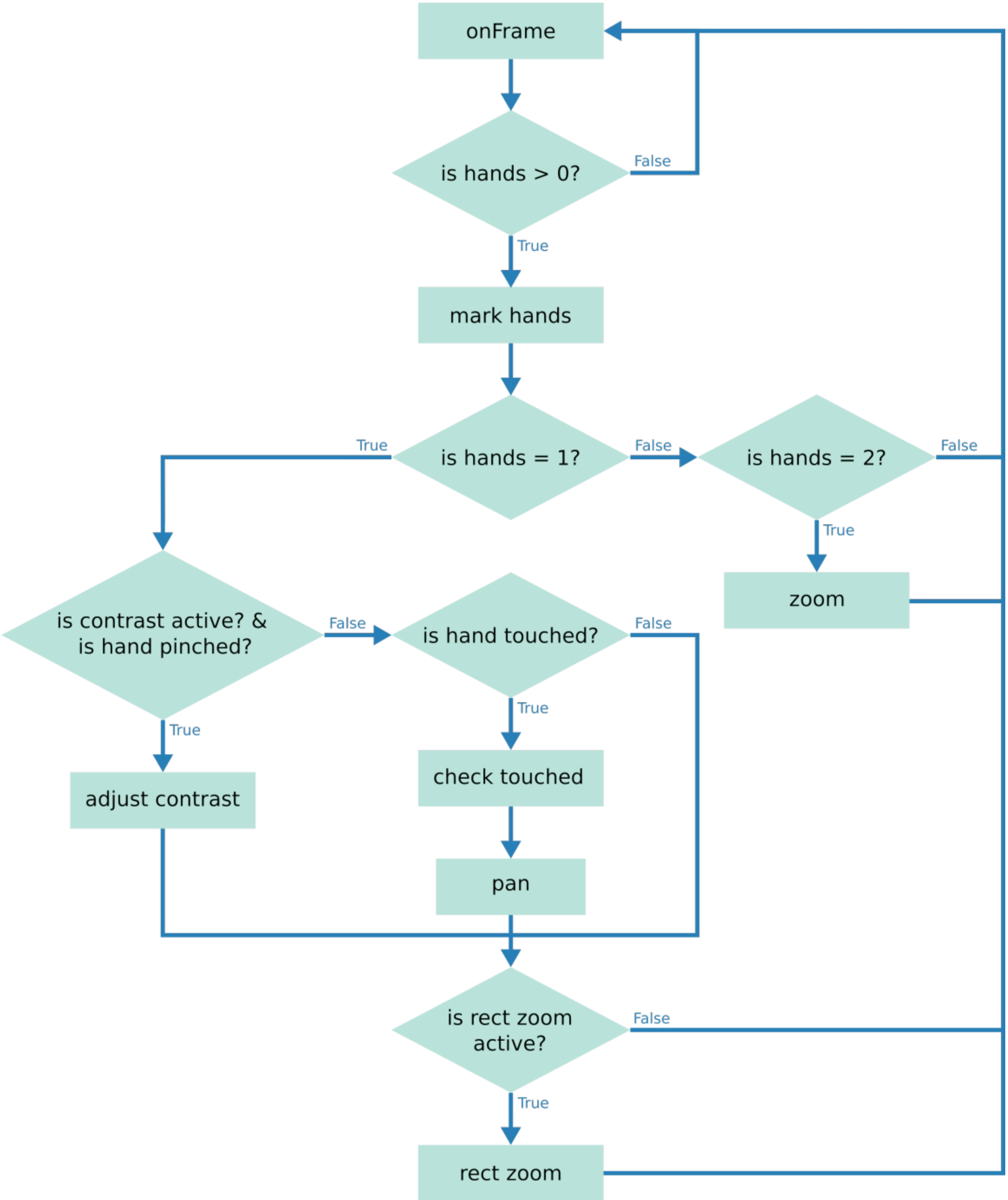


Figure 5.4: Flowchart of the function `onFrame()`.

Deriving the hand's position

To correctly derive the position of the user's hands, the LeapJS class `InteractionBox` was used. The `InteractionBox` class represents a box-shaped three-dimensional region completely within the field of view of the LMC (see figure 5.5). Each frame object has its own `InteractionBox` which was used to normalize Leap Motion position coordinates from physical millimeters to dimensionless coordinates with a value range from 0 to 1.

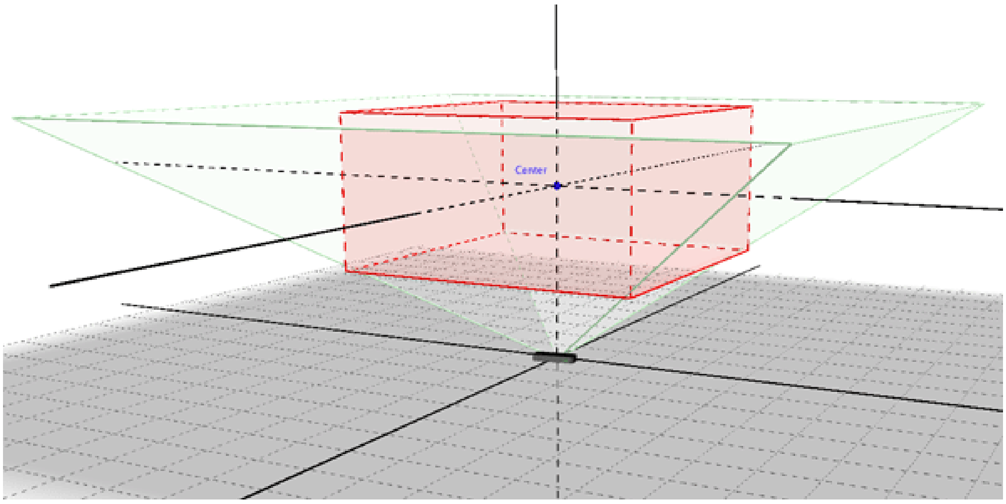


Figure 5.5: The `InteractionBox` defines a rectangular region with the field of view of LMC.

Additionally, besides exact position coordinates per frame, LeapJS was also providing a stabilized palm position of each hand within the field of view. Smoothing and stabilization is performed to create a more stable position which is more suitable for interaction with a two-dimensional user interface.

Furthermore, right and left hands were mapped differently. The left hand's origin is shifted to the right and the right hand's origin to the left. This resulted into, for instance, a less cumbersome movement of the right hand to reach the left corner of the full screen web application and vice versa.

While developing, different options to derive the position of the user's hands were tested such as tip positions of different fingers, stabilized or unstabilized hand palm positions. Finally, the stabilized hand palm position resulted as the most robust option. Hence, tracking the position of a hand was also possible with different hand forms such as flat or open hands or a hand with pointing finger(s).

Marking of hand's position and selection mechanism

As soon as a hand was within the field of view of the LMC, its position was marked on the screen. The marker had two different states: 1) hovering (purple) and 2) selecting (green).

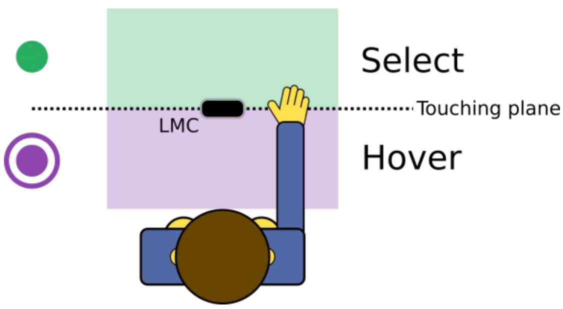


Figure 5.6: Top view of how the *hover*- and *selection*-mechanism.

The x- and y-axes span out a plane which was defined as the *touching plane*. If the user crossed this plane in direction of the z-axis (toward and away from the user), a select gesture was invoked. Hence, the touching plane was splitting the interaction area of the LMC into a hovering and selection zone. In figure 5.6, the difference of the visual appearance of the marker as well as the distinction between *hover* and *select* is shown.

Panning

In order to pan the map, the user needed to cross the touching plane to fix the current position of the map. At this very moment, the x- and y-coordinates of the center point and the selected point were saved. With a method provided by LeafletJS, the pixel coordinates of the selected point were transformed to the corresponding geographical coordinates in latitude and longitude. As soon as the user's hand was moving in selection mode, the difference between the new position and the original selected point was calculated. This difference was added to the coordinates of the map center. Subsequently, the map was panned to the new map center.

Rectangle zooming

Rectangular zooming can be very helpful to zoom in to a specific area. At first, the user needed to activate this function by selecting the corresponding button. A short help text was temporarily shown to the user in order to signalize that the function had been activated. The consensus derived from the GES described in section 4.7.1, suggested a selection of two corner points with a single hand after activating the feature. Therefore, the user needed to select the first corner point. Then, a dark blue rectangle was drawn from the selected point to the current position of the hand (see figure 5.7). To prevent the algorithm to immediately zoom in while processing the succeeding frame, a minimum area of 100 pixels for both sides was defined. To select the second point, the user needed to pull back the hand into the hovering zone first. After selecting the second point, the map was zooming to the desired rectangular area.

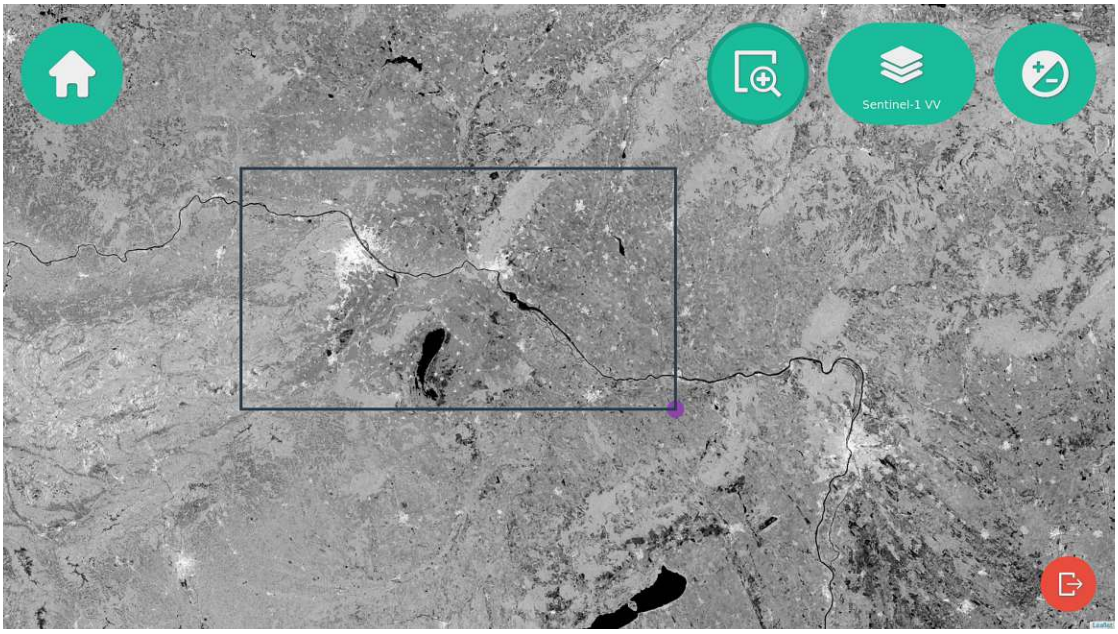


Figure 5.7: A dark blue rectangle is shown after the user selected the first corner point.

Changing the map layer

In order to switch between different map layers, the user needed to select the corresponding layer button. Afterwards, a list of the implemented map layers was shown. The user could decide which layer to load or close the menu by selecting the corresponding button. Figure 5.8 shows the list of the implemented map layers in the prototype. As long as the list is shown, panning and zooming was not available.

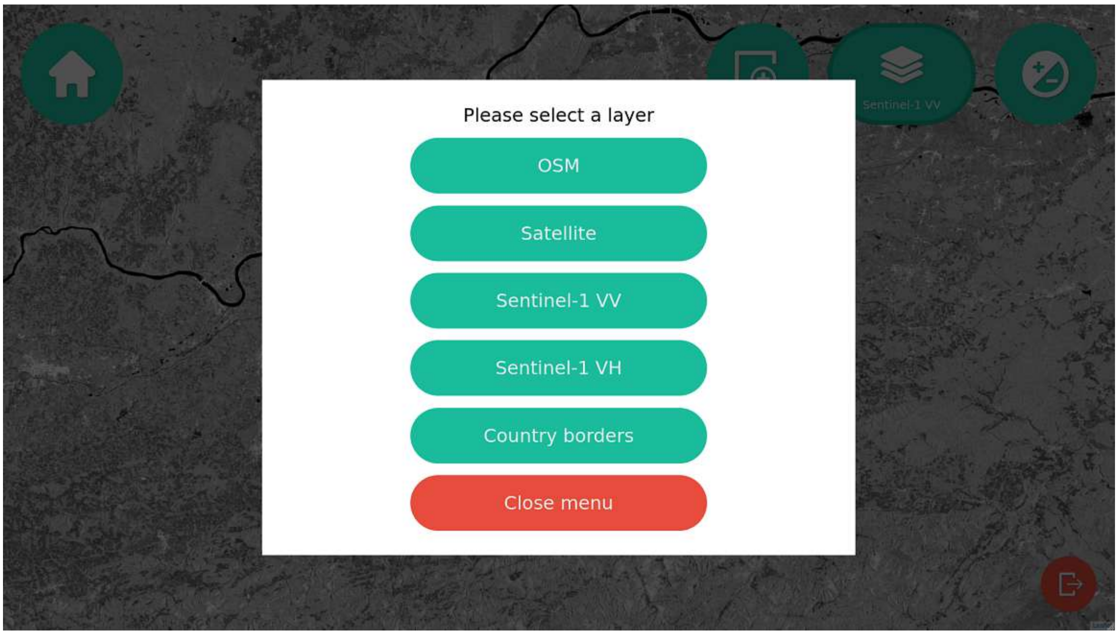


Figure 5.8: The list of the implemented map layers.

Adjusting the contrast

After the selection of the contrast button, a vertical slider was shown on the right edge of the UI. By pinching, which is described as pressing a finger against the thumb, and a subsequent vertical motion, the user was able to adjust the value of the slider, hence, the contrast. In the JavaScript code, the CSS function `contrast()` was used for the CSS key `filter`. It takes a percentage value to adjust the contrast, where a value under 100% decreases, while a value over 100% increases it. A value range from 0% to 200% was implemented.

The GES suggested to implement the adjustment of the contrast by vertical motion of a flat hand. Although, to differentiate the motion from panning and selecting, *pinching* was used to fix the thumb of slider. Note that the LMC is able to detect a pinching gesture with any finger. Furthermore, the slider is only a visual help and is not implemented as an input element. By holding the pinching gesture and moving the hand up and down, the value of the contrast is adjusted anywhere on the map.

While developing, an explanatory tutorial was added which is shown in figure 5.9. This is presented to the user as an animated image for ten seconds when selecting the feature for the first time.

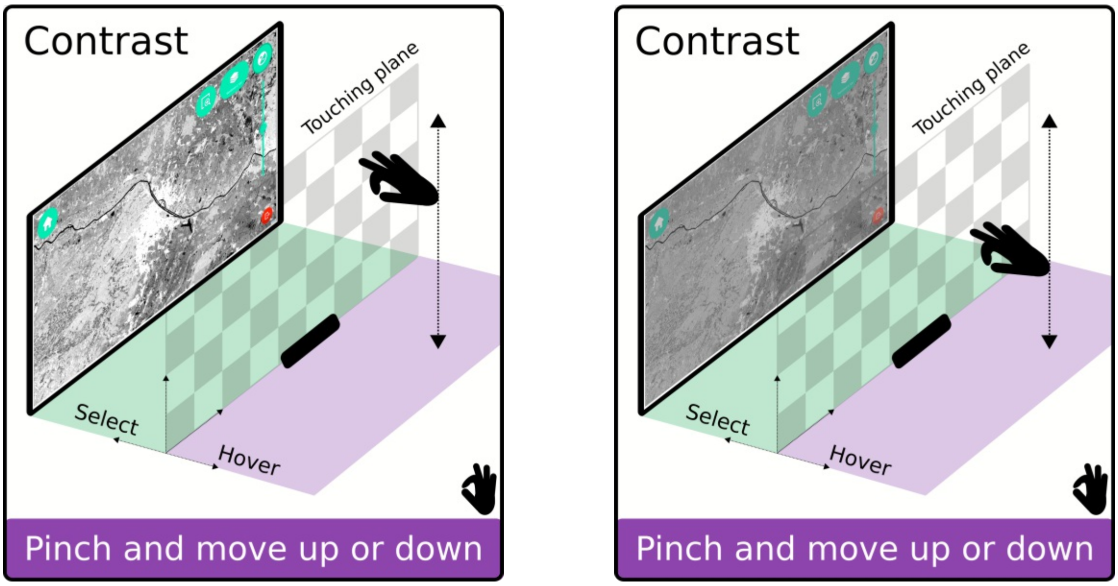


Figure 5.9: The tutorial which is presented to the user when selecting the feature for the first time. Two states of the animated image are shown.

Zooming with both hands

As soon as two hands are within the field of view, two markers are shown. For zooming the user needs to bring both hands into the selection zone. Zooming in is performed by moving both hands away from each other, hence, increasing the distance between the two hand positions. Vice versa, bringing the hands further together, the map is zoomed out.

The magnitude of the zooming is calculated with a static factor and the alteration of distance between the two hands (*separation*) with the following equation 5.1. In equation 5.2 *separationStart* is the initial distance of both hands when both hands brought into the selection zone. *sepDiff* is the ratio between the current *separation* and *separationStart*. In equation 5.3 the new zoom level *newZoom* is calculated by subtracting *sepDiff* with a pre-defined *zoomScale*. *zoomScale* is used to adjust the sensitivity of the zooming process. Furthermore, the center point between both hands is derived in order to allow the user to focus on a specific location while zooming.

$$separation = \sqrt{(P_{1,x} - P_{2,x})^2 + (P_{1,y} - P_{2,y})^2} \tag{5.1}$$

$$sepDiff = \frac{(separationStart - separation)}{separationStart} \tag{5.2}$$

$$newZoom = curZoom - sepDiff * zoomScale; \tag{5.3}$$

Back to start

To return to the initial view, the user needs to select the corresponding homebutton. After selecting, the map resets the center position and zoom level to the initial state. Note that the map layer remains unchanged.

5.3 Limitations

The following limitations were discovered during the developing process of the prototype.

The LMC provides tracking data per frame with a very high frame rate. For each frame the whole script needs to be processed by the host computer. Furthermore, map tiles need to be downloaded and updated while changing the map state. In order to decrease the processing power consumption, several boolean switches were introduced. For instance, the code for the feature rectangle zoom is only processed when the corresponding button was activated. Testing the prototype with different host computers showed that processing power had a major impact on the fluency of the interactions. Therefore, it is recommended to run the web application on a current host computer.

Additionally, several variables needed to be handled as global variables in order to save their state independently from frames. An example are the variables used saving the current zoom level while zooming.

The handling of UI elements is also different compared to developing websites with a mouse/keyboard input modality. In web browsers, mouse and keyboard inputs trigger events on UI elements such as a mouse click on a button. In contrast, no event triggering is possible in this prototype because the input modality are touchless mid-air hand gestures. To interact with UI elements, it is necessary to constantly derive (per frame) the exact hand's positions and check if any UI element is currently hovered or selected. Additionally, changing the color of UI elements needs also be performed per frame.

Chapter 6

Usability Study

6.1 Usability study process

Usability testing is a popular research methodology in the area of software design which helps to reveal areas of possible confusion as well as uncover opportunities to improve the overall user experience and usability of a system [4]. The implemented prototype of the *Touchless Map Viewer* is designed to be located in a public environment. Hence, different people, either university personnel, students or external persons are inherently invited to approach the system. Therefore, this study is essential to evaluate the usability and suggest improvements to adapt the prototype optimally for potential end-users. Furthermore, this study also helped to understand the initial hurdle and elaborate possibilities to reduce it to a minimum level.

6.2 Study procedure

The emerging system was planned to be finally placed at a public location in a hallway within the premises of TU Wien. Hence, the usability study was carried out in person per participant at that very place to achieve a situation as close to reality as possible. The final study plan was discussed with and approved by Dr. Marjo Rauhala who is the research ethics coordinator at TU Wien.

At the beginning of each test, the experimenter welcomed the test person and asked to fill out and sign a written informed consent form as well as the privacy policy for the Research Unit Geoinformation, Department of Geodesy and Geoinformation. Subsequently, the test person was asked to fill out a survey which covered demographic information such as age and gender, as well as prior experience with map-based data viewers and touchless hand gesture control. The same survey as applied for the GES was used. Every participant was informed that audio and video were recorded during the study process. Then, the experimenter asked the participant to start the test by reading the printed instructions to the tutorial and subsequent tasks which were placed on the desk.

As already described in section 5, the prototype was designed to encourage passers-by to interact with the system. Therefore, the test was carried out in such way that the participants were interacting solely with the application without interference of the experimenter. Only in situations in which the test person got overchallenged or lost control, the experimenter took corrective or declarative action.

After completing the tutorial and the tasks, the test person was given the opportunity to continue using the system and further elaborate thoughts if desired. Additionally, the participant was asked to optimally adjust the height and distance of the Leap Motion Controller, which was subsequently measured and noted by the experimenter. Finally, the participant was asked to fill out a System Usability Scale (SUS) [10] and User Experience Questionnaire (UEQ) [37].

6.2.1 Tutorial

Figure 5.2 shows the tutorial presented on the TV screen to the participant. In the printed instructions, the test person was asked to complete the tutorial in order to start with the subsequent tasks.

6.2.2 Tasks

To test the usability and user experience of the implemented features, the following six tasks were defined:

1. **Pan** - Pan the map to a location of your choice using one hand.
2. **Zoom out** - Zoom out to the whole mainland of Europe using both hands.
3. **Rectangle Zoom** - Activate the rectangular zoom mode by selecting the corresponding button. Choose the size and position of the rectangle by selecting two corner points.
4. **Adjust contrast** - Adjust the contrast to your linking by selecting the corresponding button. The presented is a visual help, which indicates the level of contrast.
5. **Layer change** - To change the presented layer, select the corresponding button. A list of layers will be shown. Please choose the layer *OpenStreetMap (OSM)* by selecting the corresponding button.
6. **Back to start** - Get back to the initial view (Vienna) by selecting the corresponding button.

6.3 Participants

Due to the pandemic during the time of the study, access restrictions to the buildings of TU Wien were in place. Therefore, the participants were recruited among colleagues of the experimenter, who had exceptional access. Those can also be defined to be part of the target group, since the emerging system of this work is planned to be located within the premises of the university. Furthermore, they have had no involvement in the preceding focus group discussion, GES or any other development process of the system. In comparison to the recruiting for the focus group discussion and GES, no emphasis on the level of prior experience with GIS software was given.

According to Nielsen and Landauer [48], five participants are enough to find almost as many usability problems as one would find using more test subjects, taking the cost-benefit ratio into account. For this study, ten participants were recruited to gain insights in possible usability problems as well as provide quantitative measures of usability (SUS) and user experience (UEQ). Note, that the low number of ten participants decreases the stability of statistical interpretation of the quantitative results. However, the SUS and UEQ provide an overall tendency of their measured aspects, which . Seven participants were male and three were female. The average age was 28.9 years ($SD = 2.96$).

6.3.1 Preliminary survey

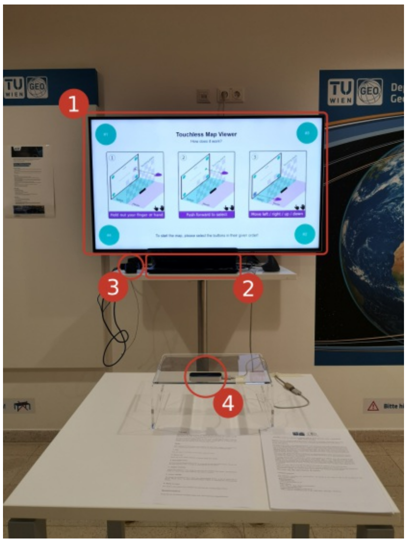
The preliminary survey was already described in detail in section 4.3.2. Not surprisingly, the answers were very similar to the answers given in the preliminary survey for the GES.

All participants stated that they use always or often an online map services to find an unknown location or, less frequently, rummage around interesting places. Touchscreen-based devices such as smartphones or tablets were used more often than desktop-based systems. The results regarding the usage of GIS-Software such as *Quantum GIS* or *ESRI ArcGIS* were diverging: three participants stated that they use it always or often, three stated rarely or never. The remaining four test persons affirmed an occasional use.

Six participants stated that they have had controlled actions using Microsoft Kinect or Nintendo Wii on TV screen at least once (two often, one occasionally and three rarely). Four persons used touchless hand gestures while wearing an augmented or virtual reality system to control actions. From them, one explained that he/she does that on a regular basis for scientific purposes. Not surprisingly, no participant ever used touchless hand gestures controlling a map-based data viewer. However, five participants stated that they have at least once interacted with a touchless hand gesture system in combination with a public display.

6.4 Apparatus

The system used for the usability test and showed in figure 6.1 consisted of 1) a Panasonic TX-55EXW604 55 inches LED TV screen, 2) a Dell G5 15 5587 laptop with an Intel Core i7-8750H Hexa-Core CPU, 3) a Logitech HD Webcam which was used for audio and video recording and 4) a Leap Motion Controller was connected to the laptop via a USB extension cable. The web application was designed for a screen resolution of 1920x1080 pixels.



(a) Experimental setup. 1) LED TV screen, 2) Dell laptop, 3) Logitech Webcam and 4) Leap Motion Controller on a acrylic computer monitor stand.

(b) The experimental setup at a final stage of the study planning.

Figure 6.1: Experimental setup for the usability study.

The LMC determines the orientation and position of the tracking area and touching plane. Hence, a change in its position is crucial in terms of the user's interaction. Therefore, the distance from the screen and height of the LMC could be adjusted by the experimenter to the test person's liking. The height adjustments were done with different underlayments such as a Styropor pad or acrylic computer monitor stand.

6.5 Analysis and results

As a first step, the audio and video recordings were reviewed in order to gather issues the participant encountered while performing the tutorial and tasks. For each issue the following were recorded:

1. The task the user was attempting to complete
2. The exact problem
3. If and how the experimenter interfered
4. User's comments

Then, the emerging data was grouped per task in order to identify repeating problems and recurring issues. Conclusively, a prioritized list of issues and problems to solve was created.

6.5.1 Tutorial

The tutorial should provide first instructions how to use the system in a intuitive and easy way. An easy to understand tutorial is key for a minimum initial hurdle. Four buttons on each corner of the TV screen (see figure 5.2) should show the user on the one hand, how to select an user interface element and on the other hand to distinguish the borders of the sensor's interaction box.

The usability test showed, that every test person had issues completing the tutorial, especially with the correct selection of a button. Although, they understood that they need to push forward to select a button, they did the latter while the cursor was already green, hence, in selection mode. As described in the focus group discussion in section 3.4.3, selection is one of the most important parts of user interaction. Thus, this problem was declared with the highest priority. The implemented (animated) images were described as too overloaded. Therefore, those need a reconsideration and redesign to explain how hovering and selecting of UI elements works in a more intuitive and easy way.

Five participants had problems to select the buttons in the bottom row. The reason was the sensors' reduced capability of detecting hands, formed to a pointing index-finger gesture, on the edges of its interaction box. Therefore, the experimenter encouraged the participants to try a flat, open hand for selecting the buttons in the bottom row, which increased the detecting performance of the sensor. A possibility to overcome this issue is to decrease the available range of physical motion, which is mapped to the displayed area. Hence, the displayed area will not cross the sensors' field of view and provide accurate tracking until the edges.

6.5.2 Tasks

Pan

After the participants figured out, either with or without the experimenters' interference, how hovering and selecting works, no major issues for the panning task were found. However, one test person used a *grabbing* gesture (closing hand to a fist) to fix the map position and subsequently panned the map. While grabbing, the participants' hand also moved forward and unintentionally crossed the touching plane. Hence, the system recognized a selection and therefore fixed the position and panned the map.

Zoom out

Only two test persons did not have any issues with the task of zooming out and in the map. One test person started to move both hands back and forth to initiate a zooming, which resulted in an explanatory interference of the experimenter.

Five participants stated, that for them it was difficult to have the motions of both hands and the corresponding displayed cursors under control. The cognitive load of zooming was generally described as very high. To reduce the latter, a participant suggested, the change the implementation of the cursors in such way that they will not move during zooming. Therefore, the user will only have to focus on the movement of its' hands.

One test person unintentionally zoomed in that far, that the map extent only covered open water which was represented as a completely black surface. To provide the user more control of the maps' position, a small map overview could be implemented.

Two participants confused the buttons of adjusting the contrast and rectangle zoom with zooming in or out. A more clear set of icons will help to reduce possible confusion, also a short help text showing when hovering over the corresponding button.

Rectangular zoom

Exactly the half of the participants combined, after selecting the rectangle zoom button, a selecting and dragging gesture to draw the rectangle. However, the results of the GES showed a clear tendency to a subsequent selection of two diagonal corner points. Therefore, and also affirmed by a suggestion of a participant, the implemented tutorial for this feature is too sparsely highlighted. Emphasizing the necessity of selecting two single corner points will help to overcome this issue.

Adjust contrast

The most common issue raised by five participants was that the visual representation of the thumb of the contrast slider was too similar to a button. This made the users to try to select the slider in order to change the contrast level, which resulted in panning the map.

After selecting the contrast button for the first time, an animated tutorial explaining how adjusting the contrast works, was shown for 10 seconds. The five participants who raised the issue seemed to take no notice of this tutorial, since they got immediate visual feedback of the map while unintentionally panning. It appeared the participants were overchallenged and tried to select the thumb of the contrast button multiple times.

Also here, an easier to understand tutorial should help to emphasize the pinching form of the hand. Nevertheless, as soon as the participants figured out how to adjust the contrast, the feedback was throughout positive.

Layer change

At this point of the usability study, the participants already gained a certain amount of experience in the usage of the system. Therefore, changing the presented layer did not raised any issues.

Back to start

Also for the task of going back to start, no participant faced any problems.

6.5.3 System usability scale (SUS)

The system usability scale (SUS), proposed by Brooke [10], is an easy and fast way to quantitatively evaluate any kind of system. It uses ten Likert scale items with five response options from *Strongly agree* to *Strongly disagree* and helps to measure and quantify the perception of usability.

The average usability score is 75.75% with a maximum value of 95.0% and a minimum value of 32.5% (SD = 17.48%). Considering the findings of Bangor et al. [3], the current implementation can be categorized between *Good* and *Excellent*. Although, the minimum SUS score of 32.5% indicates that major issues occurred, the overall tendency of the usability of the prototype is positive.

6.5.4 User experience Questionnaire (UEQ)

The User experience Questionnaire (UEQ) is a fast and reliable method to quantify the user experience (UX) of interactive systems. Introduced by Laugwitz et al. [37], the UEQ covers usability and user experience aspects such as attractiveness, perspicuity, efficiency, dependability, stimulation and novelty. The questionnaire comprises 26 items with each two conflicting terms. The test person need to decide as spontaneously as possible which of the terms better describes the tested system. A provided analysis tool helped to extract information of the questionnaires.

The mean and variance of each item over all participants were calculated. Then, the items were grouped to the UEQ scales and again the mean and variance calculated (see table 6.1). According to Laugwitz et al. [37], the value range is between -3 (horribly bad) and +3 (extremely good). Values between -0.8 and 0.8 represent a *neutral evaluation*, values higher than 0.8 represent a *positive evaluation* and values lower than 0.8 a negative evaluation.

Scale	Mean	Variance
Attractiveness	2.283	0.36
Perspicuity	1.400	1.00
Efficiency	1.600	0.49
Dependability	1.375	0.45
Stimulation	2.450	0.18
Novelty	2.075	0.71

Table 6.1: UEQ scales and their corresponding mean and variance.

The UEQ scale *Attractiveness* represents the overall impression of the product, which reached a high value of 2.283. Also, *Stimulation* and *Novelty* describing excitement and innovation as hedonic quality aspects have reached high values above 2.0.

The UEQ scale *Perspicuity* has a high variance of 1.00. This scale consists of the items *clear/confusing*, *complicated/easy*, *easy/difficult to learn* and *not understandable/understandable*. The latter three have individually high variances between 1.6 and 2.4, which indicates a diverging perception.

Efficiency tries to measure how users could solve their tasks without unnecessary effort and reached a mean value of 1.6. *Dependability* describes if the user felt in control of the interaction and reached the lowest mean, but still positive evaluation of 1.375.

6.5.5 Height and distance of LMC

During the usability test, the participants were encouraged to adjust the position of the Leap Motion Controller to their preference. Since the LMC determines the orientation of the touching plane, a change in its position is crucial in terms of the user's interaction. After the test persons felt comfortable with the setting, the height and distance to the screen of the sensor's position was measured (see table 6.2).

Height [*cm*]	**# Participants**
85	4
97	6

Distance [*cm*]	**# Participants**
120	1
125	3
130	2
135	4

Table 6.2: Height and distance to the TV screen of the Leap Motion Controllers' position.

6.5.6 Discovered issues

Conclusively, the following table 6.3 of discovered issues and corresponding possible solutions, ordered by importance, was the basis for further improvements of the prototype:

	Issue	**Solution**
1	User tried to select UI element while cursor was already in selection mode	Redesign and simplify tutorial
2	Difficult to focus on both cursors while zooming	Fixed position of cursors while zooming
3	User dragged to draw a rectangle	Different handling of corner point selection
4	User tried to select button of contrast slider	Redesign tutorial and slider
5	User lost cursor in bottom area	Reconsider LMC's interaction box

Table 6.3: Discovered issues and possible solutions.

6.6 Discussion

The qualitative and quantitative findings of the usability study suggest that the users would be receptive to the ability of controlling a map viewer with touchless mid-air hand gestures. The positive tendency of the system usability scale (75.75%) and of the UEQ scale *Attractiveness* (2.283) indicates a high level of user satisfaction already in an early stage of the prototype development phase.

A similar average SUS score was found by Gentile et al. [23], who tested the difference in usability of two mid-air zoom gestures. Their usability assessment resulted in 69.13 for *Zoom #1* and 75.63 for *Zoom #2*.

Considering this positive results, presenting global Earth observation data using this system, can help to increase the accessability of these complex data sets to a broader audience including novices. As pointed out by Hodam et al. [27], also in higher education, as well as lifelong learning, „EO data are meeting an ever-growing audience that is eager to learn about the subject". Hence, an attractive system which enables the user to playfully rummage around complex EO data will foster the understanding of the Earth system.

The work of Di Geronimo et al. [17] presented a system, called *MyoShare*, which allows content to be shared among devices using mid-air gestures. After a GES, they assessed the usability of MyoShare by conducting two user studies comparing to different input modalities such as *speech*, *shortcuts* and *menu selection*. Their results showed that mid-air gestures were enjoyed most by their participants, but were less efficient than *shortcuts*. This is in line with the usability evaluation of the current study, showing a high degree of *Attractiveness*, but lower values for *Efficiency* or *Dependability*.

Considering the measured values for distance and height of the LMC, no clear preference among all participants could be found. Due to nature of touchless mid-air hand gestures, the test person's body height and arm length plays an important role as well as the subjective experience during motions. Therefore, a height adjustable and movable stand on which the sensor is mounted, could further improve the user experience.

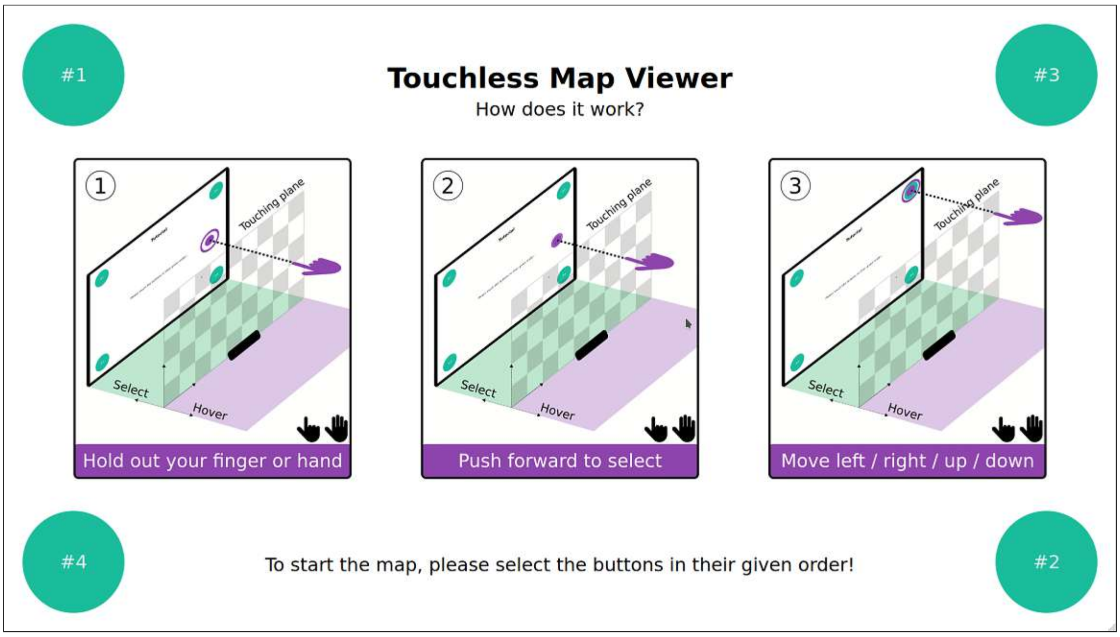
The work of Vogiatzidakis and Koutsabasis [64] performed a literature review about gesture elicitations studies for mid-air interaction. They concluded, that further work need to be done in order to validate results of gesture elicitation with technical tests about measured usability. The results of the current usability study can provide further insights into the latter.

Although the results of the SUS and UEQ suggest good overall usability, valuable insights in how the prototype could be improved further was gained. In the following, the reported discovered issues and their corresponding implemented solutions are presented.

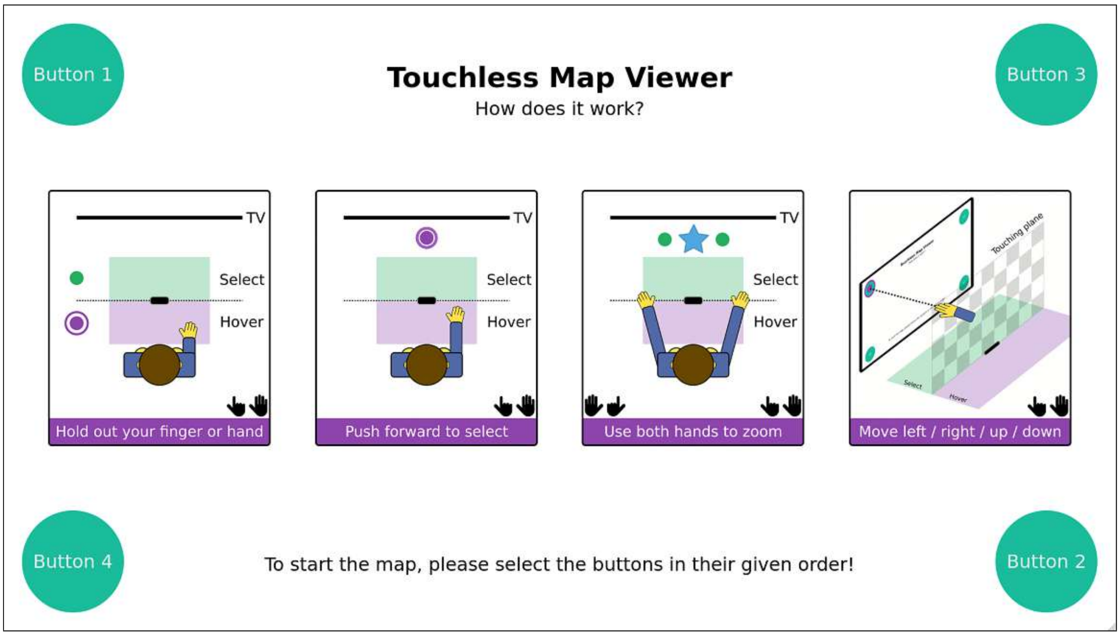
6.7 Prototype adjustments

6.7.1 Redesign of tutorial

During redesigning the tutorial, special focus on the highlighting of the difference between hovering and selecting was given. Another explanatory image was added which specifically emphasized the difference in visual appearance of the two cursor states, hover and select. Figure 6.2 show the before and after visualization of the tutorial. Note, that also in the new tutorial all images except the first are animated.



(a) Old tutorial.



(b) New tutorial.

Figure 6.2: Redesign of the tutorial.

To support the distinction between hover and selection mode, a physical nudge could further increase the usability. For example, a thin metal wire which borders the interaction box on the left and right side, could give the user an indication about the physical dimension of the touching plane. Of course, the possible danger of injuries would be needed to take into account.

6.7.2 Zooming with both hands

As suggested by a participant of the usability study, the cursors should not move during a two-handed zoom in order to decrease the cognitive load. The marking of both hands was changed in such way that they are fixed during zooming. Additionally, the scale factors for adjusting the sensitivity and speed of zooming in and out were adapted.

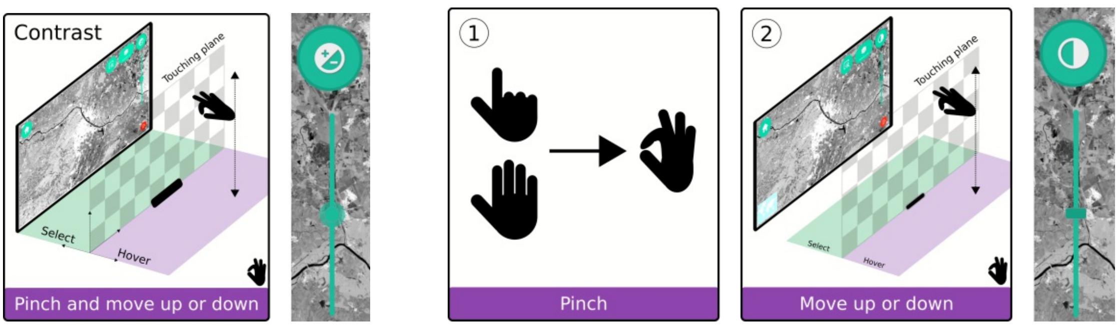
Consequently, the end-user is able to fixate the hand positions one by one by pushing each hand forward into the selection zone.

6.7.3 Reconsideration of rectangle zoom

The originally implemented minimum area of the rectangle which prevented the map accidentally zoom in at the succeeding frame, was deleted. Instead, a boolean switch is used to determine if the hand was pulled back to the hovering zone after selecting the first corner point. If yes, the user is able to select the second corner point. With that improvement, the feature is more robust and less susceptible to errors.

6.7.4 Redesign of contrast tutorial

The redesign of the contrast tutorial emphasizes the necessity of changing the hand form from an open hand or pointing finger to a pinching gesture. In the second animated image, the vertical movement of the pinched hand is shown. Also the threshold for determination if the user is pinching was increased in order to achieve a more stringent distinction. Furthermore, the icon of the contrast button was changed and the thumb of the slider was redesigned to prevent a possible confusion with a button. Figure 6.3 shows the redesign of the tutorial, button icon and the slider compared to the original implementation.



(a) Old (b) New

Figure 6.3: Redesign of the contrast tutorial, button icon and slider.

6.7.5 Reconsider LMC's interaction box

The prototype used the class `InteractionBox` to normalize coordinates of the hand position to a value range from 0 to 1. Hence, the minimum value of the interaction box maps to 0 whereas the maximum value maps to 1. Those normalized coordinates were then multiplied by the application's width for the x pixel coordinate and multiplied by the application's height for the y pixel coordinate. In order to increase the robustness of hand detection in the bottom area of the web application, a factor of 100 pixels was added to the application's height. Therefore, the normalized y coordinates reached its minimum value already at a greater distance from the sensor. Hence, also the distance to the edges of the field of view was increased.

6.7.6 Further improvements

Reconsidering the findings of the focus group discussion and comments stated by the participants during the usability study, the following improvements were added to the prototype. Figure 6.4 shows the new added legend (1), scale bar (2) and map overview (3).

Figure 6.4: As further improvements a legend (1), scale bar (2) and map overview (3) were implemented.

Legend (1)

Legends tell the user of the map the meaning or value range of the presented data. Considering the complexity of satellite-based radar backscatter data, an expert stated in the focus group discussion that a legend would help to understand the presented map. The legend is shown for the radar backscatter maps Sentinel-1 VV and Sentinel-1 VH with individual value ranges, respectively.

Scale bar (2)

Scale bars provide a visual indication of distance on the map and are an essential part of a map [35]. The implemented scale bar is associated with the shown map. Hence, if the map scale changes, also the scale bar is updated to remain correct.

Map overview (3)

The map overview displays the current centered map extent on a small world map. This helps the user to stay in control while zooming and panning the map.

Chapter 7

Conclusion and Outlook

In this work, a novel prototype of a map viewer which, on the one hand, displays satellite-based radar backscatter data as a complex Earth observation (EO) data set and, on the other hand, is controllable with touchless mid-air hand gestures was implemented. Desirable features and functions of such a system were derived by conducting a focus group discussion. Based on those results, a gesture elicitation study was conducted which obtained touchless mid-air hand gestures and common narratives among the test persons. But, the high degree of diversity of the proposed gestures showed that the determination of a agreed-upon gesture set is still an open question.

The user experience and usability evaluation of the emerging prototype provide insights that the input modality of touchless mid-air hand gestures can increase the enjoyment while using a public display environment. But it also showed that designing such a system is still intricate and asks for iterative development processes.

However, the implemented map viewer showed high attractiveness to first-time users which was tested in the usability study. As stated by Kapur et al. [33], it is important to create systems which provide hands-on engagement with real data in order to „allow users at all levels to remain up-to-date with EO technologies". Therefore, the presented map viewer can contribute to foster understanding and advertising complex EO data sets by being accessible as a public display system controlled by touchless mid-air hand gestures. The map viewer combined with the novel input modality offers the possibility to browse and discover complex map-based data sets in a playful way and therefore be seen as a complement to existing map-based data viewers.

Furthermore, the combination of a gesture elicitation study with a subsequent usability assessment can provide suggestions for prototype improvements and increasing the user experience with touchless input modalities.

A limitation of the current study were the access restrictions during the Covid-19 pandemic, which made it necessary to conduct the focus group discussion and gesture elicitation study (GES) online. Therefore, a comparable GES which is conducted in person using the existing prototype could give deeper insight into the mental models of the participants and further refine the consensus gesture set. Future work could also include the conduction of new tests, obtaining the usability and user experience of the implemented solutions.

Appendix

Folder contents

```
/
├── 01_Focus_Group_Discussion
│   ├── fg_consent_form_signed
│   │   ├── Datenschutzerklärung_<NAME>.pdf
│   │   └── Einverständniserklärung_<NAME>.pdf
│   ├── fg_full_transcript.pdf
│   ├── fg_informed_consent_form.pdf
│   ├── fg_livenotes.pdf
│   ├── fg_preliminary_survey_responses.pdf
│   └── fg_recording.mp4
├── 02_Gesture_Elicitation_Study
│   ├── ges_informed_consent_signed
│   │   └── Consent_<NAME>.pdf
│   ├── ges_recordings
│   │   └── <ID>.mkv
│   ├── ges_referents_images
│   │   └── <REFERENT>.png
│   ├── ges_analysis_per_referent.ods
│   ├── ges_gesture_classification.ods
│   ├── ges_informed_consent_form.pdf
│   ├── ges_notes_per_participant.ods
│   ├── ges_participants.ods
│   ├── ges_preliminary_survey_responses.ods
│   ├── ges_study_procedure_text.pdf
│   └── gesture_study_referents.odp
├── 03_Usability_Study
│   ├── us_recordings
│   │   └── <ID>.mkv
│   ├── us_consent_form.pdf
│   ├── us_preliminary_survey_responses.ods
│   ├── us_setup.png
│   ├── us_SUS_analysis.ods
│   ├── us_tasks.pdf
│   └── us_UEQ_analysis.xlsx
├── 04_Prototype
│   └── leap-map.zip
└── 05_Masterthesis_Tobias_Stachl.pdf
```

References

[1] Open-source javascript library for interactive maps - leaflet js. https://leafletjs.com/. Accessed on May 31, 2021.

[2] Daniel Bachmann, Frank Weichert, and Gerhard Rinkenauer. Review of three-dimensional human-computer interaction with focus on the leap motion controller. *Sensors*, 18(7):2194, 2018.

[3] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3): 114–123, 2009.

[4] Carol M Barnum. *Usability testing essentials: ready, set... test!* Morgan Kaufmann, 2020.

[5] B Bauer-Marschallinger, S Cao, C Navacchi, V Freeman, F Reuß, Geudtner D, B Rommen, F Ceba Vega, P Snoeij, E Attema, C Reimer, and W Wagner. The normalised sentinel-1 global backscatter model – mapping earth's land surface with c-band microwaves. *Scientific Data*, 2021. in Review.

[6] Andrea Bellucci, Alessio Malizia, Paloma Diaz, and Ignacio Aedo. Don't touch me: multi-user annotations on a map in large display environments. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 391–392, 2010.

[7] Moniruzzaman Bhuiyan, Rich Picking, et al. A gesture controlled user interface for inclusive design and evaluative study of its usability. *Journal of software engineering and applications*, 4(09):513, 2011.

[8] Assam Boudjelthia, Sofeem Nasim, Janne Eskola, Joshua Muyiwa Adeegbe, Oula Hourula, Simon Klakegg, and Denzil Ferreira. Enabling mid-air browser interaction with leap motion. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 335–338, 2018.

[9] James V Bradley. Complete counterbalancing of immediate sequential effects in a latin square design. *Journal of the American Statistical Association*, 53(282): 525–528, 1958.

[10] John Brooke. Sus: a "quick and dirty" usability scale. *Usability Evaluation In Industry*, 189, 1996.

[11] Andreea Bucur, Wolfgang Wagner, Stefano Elefante, Vahid Naeimi, and Christian Briese. Development of an earth observation cloud platform in support to water resources monitoring. *Earth Observation Open Science and Innovation*, 15:275, 2018.

[12] Ziyun Cai, Jungong Han, Li Liu, and Ling Shao. Rgb-d datasets using microsoft kinect or similar sensors: a survey. *Multimedia Tools and Applications*, 76(3):4313–4355, 2017.

[13] Bobby J Calder. Focus groups and the nature of qualitative marketing research. *Journal of Marketing research*, 14(3):353–364, 1977.

[14] Andy Cockburn, Amy Karlson, and Benjamin B Bederson. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR)*, 41(1):1–31, 2009.

[15] Alan Cooper, Robert Reimann, and David Cronin. *About face 3: the essentials of interaction design*. John Wiley & Sons, 2007.

[16] Juliet Corbin and Anselm Strauss. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications, 2014.

[17] Linda Di Geronimo, Marica Bertarini, Julia Badertscher, Maria Husmann, and Moira C Norrie. Exploiting mid-air gestures to share data among devices. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–11, 2017.

[18] Esri. World imagery [basemap]. scale not given. "world imagery", May 19, 2021. URL `https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9`. Accessed on May 31, 2021.

[19] Vinícius Fagundes, Raul Fernandes, Carlos Santos, and Tatiana Tavares. Visualization of climate data from user perspective: Evaluating user experience in graphical user interfaces and immersive interfaces. In *International Conference on Human Interface and the Management of Information*, pages 55–70. Springer, 2017.

[20] Flat UI Colors. Flat ui palette v1 — flat ui colors - 280 handpicked colors ready for copy and paste. https://flatuicolors.com/palette/defo. Accessed on May 31, 2021.

[21] Marcus Foth, Martin Tomitsch, Laura Forlano, M Hank Haeusler, and Christine Satchell. Citizens breaking out of filter bubbles: urban screens as civic media. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays*, pages 140–147, 2016.

[22] Rita Francese, Ignazio Passero, and Genoveffa Tortora. Wiimote and kinect: gestural user interfaces add a natural third dimension to hci. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 116–123, 2012.

[23] Vito Gentile, Daniele Fundarò, and Salvatore Sorce. Elicitation and evaluation of zoom gestures for touchless interaction with desktop displays. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays*, pages 1–7, 2019.

[24] Barney G Glaser, Anselm L Strauss, and Elizabeth Strutzel. The discovery of grounded theory; strategies for qualitative research. *Nursing research*, 17(4):364, 1968.

[25] Yves Guiard and Michel Beaudouin-Lafon. Target acquisition in multiscale electronic worlds. *International Journal of Human-Computer Studies*, 61(6):875–905, 2004.

[26] Jihyun Han and NE Gold. Lessons learned in exploring the leap motion sensor for gesture-based instrument design. Goldsmiths University of London, 2014.

[27] Henryk Hodam, Andreas Rienow, and Carsten Jürgens. Bringing earth observation to schools with digital integrated learning environments. *Remote Sensing*, 12(3): 345, 2020.

[28] Fabian Hoffmann, Miriam-Ida Tyroller, Felix Wende, and Niels Henze. User-defined interaction for smart homes: voice, touch, or mid-air gestures? In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, pages 1–7, 2019.

[29] Justin L Huntington, Katherine C Hegewisch, Britta Daudert, Charles G Morton, John T Abatzoglou, Daniel J McEvoy, and Tyler Erickson. Climate engine: Cloud computing and visualization of climate and remote sensing data for advanced

natural resource monitoring and process understanding. *Bulletin of the American Meteorological Society*, 98(11):2397–2410, 2017.

[30] Ideum. Touchless pedestal - zero-touch interaction for large displays, 2020. URL `https://ideum.com/products/touchless/touchless-pedestal`. Accessed on May 31, 2021.

[31] Inkscape. Inkscape - free and open source vector graphics editor. https://inkscape.org/. Accessed on May 31, 2021.

[32] Muhammad Zahid Iqbal and Abraham Campbell. The emerging need for touchless interaction technologies. *Interactions*, 27(4):51–52, 2020.

[33] Ravi Kapur, Val Byfield, Fabio Del Frate, Mark Higgins, and Sheila Jagannathan. The digital transformation of education. *Earth observation open science and innovation [Internet]. ISSI Scientific Report Series*, 15:25–41, 2018.

[34] Anne Köpsel and Nikola Bubalo. Benefiting from legacy bias. *interactions*, 22(5): 44–47, 2015.

[35] MJ Kraak and A Brown. Web cartography: developments and prospects. 2001.

[36] Richard A Krueger. *Focus groups: A practical guide for applied research*. Sage publications, 2014.

[37] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pages 63–76. Springer, 2008.

[38] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[39] Hannah Limerick. Call to interact: Communicating interactivity and affordances for contactless gesture controlled public displays. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays*, pages 63–70, 2020.

[40] Wei Lu, Zheng Tong, and Jinghui Chu. Dynamic hand gesture recognition with leap motion controller. *IEEE Signal Processing Letters*, 23(9):1188–1192, 2016.

[41] Nathan Magrofuoco and Jean Vanderdonckt. Gelicit: A cloud platform for distributed gesture elicitation studies. *Proceedings of the ACM on Human-Computer Interaction*, 3(EICS):1–41, 2019.

[42] Rafael Martins and Pollyana Notargiacomo. Evaluation of leap motion controller effectiveness on 2d game environments using usability heuristics. *Multimedia Tools and Applications*, 80(4):5539–5557, 2021.

[43] David L. Morgan. Focus groups. *Annual Review of Sociology*, 22(1):129–152, 1996. doi: 10.1146/annurev.soc.22.1.129.

[44] Meredith Ringel Morris. Web on the wall: insights from a multimodal interaction elicitation study. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pages 95–104, 2012.

[45] Meredith Ringel Morris, Jacob O Wobbrock, and Andrew D Wilson. Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010*, pages 261–268. 2010.

[46] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, MC Schraefel, and Jacob O Wobbrock. Reducing legacy bias in gesture elicitation studies. *interactions*, 21(3):40–45, 2014.

[47] Natural Earth. World borders in geojson format - data comes from natural earth. https://datahub.io/core/geo-countries. Accessed on May 31, 2021.

[48] Jakob Nielsen and Thomas K Landauer. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 206–213, 1993.

[49] Michael Nielsen, Moritz Störring, Thomas B Moeslund, and Erik Granum. A procedure for developing intuitive and ergonomic gesture interfaces for hci. In *International gesture workshop*, pages 409–420. Springer, 2003.

[50] Tobias O. Nyumba, Kerrie Wilson, Christina J Derrick, and Nibedita Mukherjee. The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and evolution*, 9(1):20–32, 2018.

[51] OpenStreetMap contributors. OpenStreetMap's Standard tile layer retrieved from https://(s).tile.openstreetmap.org/(z)/(x)/(y).png. `https://www.openstreetmap.org`, 2021.

[52] Morin Ostkamp and Christian Kray. Supporting design, prototyping, and evaluation of public display systems. In *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems*, pages 263–272, 2014.

[53] Callum Parker, Martin Tomitsch, Nigel Davies, Nina Valkanova, and Judy Kay. Foundations for designing public interactive displays that provide value to users. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[54] Leigh Ellen Potter, Jake Araullo, and Lewis Carter. The leap motion controller: a view on sign language. In *Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration*, pages 175–178, 2013.

[55] Jessica Roberts, Amartya Banerjee, Annette Hong, Steven McGee, Michael Horn, and Matt Matcuk. Digital exhibit labels in museums: promoting visitor engagement with cultural artifacts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[56] Mithileysh Sathiyanarayanan and Tobias Mulling. Map navigation using hand gesture recognition: A case study using myo connector on apple maps. *Procedia computer science*, 58(2015):50–57, 2015.

[57] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Pearson, 6th edition, 2016. ISBN 013438038X.

[58] Sudhir Raj Shrestha, Matthew Tisdale, Steve Kopp, and Brett Rose. Earth observation and geospatial implementation: Fueling innovation in a changing world. In *Earth Observation Open Science and Innovation*, pages 301–310. Springer, Cham, 2018.

[59] Jane Farley Templeton. *The focus group: A strategic guide to organizing, conducting and analyzing the focus group interview.* Probus Publishing Company, 1994.

[60] Ultraleap. Leap motion controller - optical hand tracking module. `https://www.ultraleap.com/product/leap-motion-controller/`. Accessed: July 27, 2020.

[61] Radu-Daniel Vatavu and Jacob O Wobbrock. Formalizing agreement analysis for

elicitation studies: new measures, significance test, and toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1325–1334, 2015.

[62] Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. Leap gestures for tv: insights from an elicitation study. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, pages 131–138, 2014.

[63] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O Wobbrock. A systematic review of gesture elicitation studies: What can we learn from 216 studies? In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pages 855–872, 2020.

[64] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technologies and Interaction*, 2(4): 65, 2018.

[65] Wolfgang Wagner. Big data infrastructures for processing sentinel data. In *Photogrammetric week*, volume 15, pages 93–104, 2015.

[66] Robert Walter, Gilles Bailly, and Jörg Müller. Strikeapose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 841–850, 2013.

[67] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5): 6380–6393, 2013.

[68] Markus L Wittorf and Mikkel R Jakobsen. Eliciting mid-air gestures for wall-display interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, pages 1–4, 2016.

[69] Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*, pages 1869–1872, 2005.

[70] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1083–1092, 2009.