



Enabling post-recording deep georeferencing of walkthrough videos: an interactive approach

by

Samuel Navas Medrano

A thesis submitted in partial fulfillment for the
Master degree of Science in Geospatial Technologies

in the
Institute for Geoinformatiks

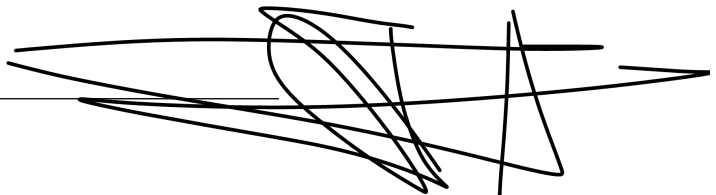
February 2016

Declaration of Authorship

I, SAMUEL NAVAS MEDRANO, declare that this thesis titled, 'ENABLING POST-RECORDING DEEP GEOREFERENCING OF WALKTHROUGH VIDEOS: AN INTERACTIVE APPROACH' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a Master degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Samuel Navas Medrano



Date: 26th of February, 2016

"I'll sleep well tonight."

Henry Ford



Abstract

Institute for Geoinformatiks

MSc in Science of Geospatial Technologies

by Samuel Navas Medrano

The usage of large scale databases of georeferenced video stream has an infinite number of applications in industry and research since efficient storing in geodatabases to allowing the performing of spatial queries.

Due to the fact that the video capturing devices have become ubiquitous, a good source for the acquisition of a lot of video contents is the crowdsourced approach of Social Media. However, these social apps usually do not support geo metadata or it is very limited to a single location on Earth. In other cases, the regular user usually does not have the required hardware and software to capture video footage with a deep georeference (position and orientation in time). There is a clear lack of methods for the extraction of that spatial component in video footage.

This study proposes and evaluates a new method for the manual capture and extraction of the spatial geo-reference in the post production phase of video content. The proposed framework is based on a map-based user interface synchronized with the video stream. The efficiency and usability of the resulting framework were evaluated performing a user study, in addition, the resulting geo-metadata of the manual extracted georeference has been compared with the one previously captured by hardware in order to evaluate the goodness of the method.

Acknowledgements

First of all, I would like to express my greatest gratitude to my supervisors Prof. Christian Kray, Prof. Pedro Cabral, and Holger Fritze for their guidance, critique, feedback and as well as all the time they have deserved to my person. I would also like to thank Dr. Carl Schultz for being my advisor. I really appreciate their patience and continuous support.

I would like to thank all my classmates for their endless friendship and encouragement. We have shared the joy and suffering of moving to an unknown country. These memories and adventures are ones that I will always treasure.

A special mention goes to Anita, Mehrnaz, and the others for supporting me during my thesis with great kindness. Finally, I need to thank Joanna for encouraging me in everything that I do, and Fani for reminding me to have bits of fun during my thesis.

I thank my uncle Miguel and their family, who have helped me move to Castellón abruptly and gave me the pleasure of their company during my weekends there. Finally, I thank my brother and parents, who love me unconditionally and have supported me for as long as I can remember.

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	iv
List of Figures	vii
List of Tables	viii
Abbreviations	ix
Symbols	x
1 Introduction	1
1.1 Motivation	1
1.2 Objective and aims	1
1.3 Requirements	2
1.4 Outline	3
2 Literature review	4
2.1 Georeferencing and Geotagging	4
2.2 Geographic video databases	5
2.3 User study	6
2.4 Map based UI and visualisation	7
2.5 Disaster scenario and volunteer driven approach	8
2.6 Routing	8
3 Approach, methods and tools	9
3.1 Approach	9
3.2 Methods	9
3.3 Tools	10
3.4 Other resources	11
4 Software Prototype	12
4.1 Concept	12

4.2	Implementation	12
4.2.1	Layout	12
4.2.2	Interactive map	13
4.2.3	Geocoding, reverse geocoding and routing	14
4.3	Way-points positioning and orientation	15
4.4	Interpolation	16
5	User Study	17
5.1	Experimental design	17
5.2	Participants	18
5.3	Stimuli	18
5.4	Procedure	18
6	Results	20
6.1	Software Prototype	20
6.2	User Study	20
6.2.1	Qualitative questions	20
6.2.2	Task Completion Time	21
6.2.3	Usability	21
6.2.4	Accuracy	23
6.2.5	Empirical observations	25
7	Discussion and evaluation	27
7.1	Discussion	27
7.2	Evaluation	29
8	Conclusions	31
8.1	Contributions	31
8.2	Limitations	32
8.3	Future work	32
A	User study document	35
B	Results visual comparison	40
B.1	Route 2 (complex)	40
B.2	Route 4 (complex)	40
B.3	Route 5 (simple)	41
B.4	Route 6 (simple)	41
	Bibliography	46

List of Figures

4.1	User Interface layout.	13
4.2	Reconstructed FoV model.	15
6.1	Screenshot of the resulting software prototype.	21
6.2	Group means and confidence intervals.	22
6.3	Boxplot of the different times for completing the georeferencing.	22
6.4	SUS score boxplot.	22
6.5	SUS scores histogram.	22
6.6	Position RMSE boxplot.	24
6.7	Position RMSE confidence intervals.	24
6.8	Orientation RMSE boxplot.	24
6.9	Orientation RMSE confidence intervals.	24
7.1	Participant 10 georeference and ground truth comparison.	30

List of Tables

1.1	The core and optimal project requirements.	2
4.1	Main Route Services comparison.	15
5.1	Route complexities.	17
6.1	SUS scores summary by questions.	23
6.2	Wilcoxon signed rank test of the position RMSE.	25
6.3	Wilcoxon signed rank test of the orientation RMSE.	25
6.4	Results of geocoding the video names by the Google API.	26

Abbreviations

AJAX	A synchronous J avascript A nd X ML
API	A pplication P rogram I nterface
CSS	C ascade S tyle S heets
Cv	C oefficient of v ariation
FoV	F ield of V iew
GIS	G eographic I nformation S ystem(s)
GPS	G lobal P ositioning S ystem
HTML	H yper T ext M arkup L anguage
JS	J ava S cript
Lerp	L inear I nter p olation
MSE	M ean S quared E rror
NGO	N on- G overnmental O rganization
OCG	O pen G eospatial C onsortium
OSM	O pen S treet M ap
SLerp	S pherical L inear I nter p olation
SUS	S ystem U sability S cale
UAV	U nmanned A erial V ehicles
UI	U ser I nterface
VGI	V olunteer G eographic I nformation

Symbols

φ	latitude	decimal degrees
λ	longitude	decimal degrees
σ	standard deviation	na
μ	population mean	na
T	Kendall rank correlation coefficient	na

Chapter 1

Introduction

1.1 Motivation

Georeferenced multimedia can be beneficial for visually navigating in space or performing spatiotemporal queries in large databases of multimedia content[1]. Despite this, the wide range of available video and photos on the Internet usually do not include geographic metadata.

The possible reasons could vary from the possibility that user does not have access for the necessary means to perform a hardware geotag, privacy issues that may arise[2], or that they are simply oblivious to the possibility of georeferencing content. There could also be a lack of awareness of the potential benefits of the information.

Furthermore, most of the content that we find geotagged in social media by conventional approaches are usually limited to just one geographical point[3] or is not accurate enough[4]. To perform spatial queries in georeferenced media to harness the benefits, it is a minimum requirement to capture and store information about the position and field of view of the scene in time[5].

1.2 Objective and aims

This thesis aims to propose and evaluate a new method for the manual capture and extraction of the spatial geo-reference in the post production phase of video content. The suggested framework for achieving this goal is based on a map-based user interface synchronized with video stream. The efficiency and usability of the resulting framework were evaluated by performing a user study. In addition, the resulting georeference

Core	Optimal
Only taking in account terrestrial walkthrough videos.	Only taking in account terrestrial walkthrough videos.
Extract georeference: position in time.	Extract deep georeference: position and orientation in time
	Create paper UI prototypes and ask users for feedback
Retrieve geospatial information from the metadata and suggest the user a possible location.	Retrieve geospatial information from the metadata and suggest the user a possible location.
	Detect breakpoints in the same video and split the original video in separate videos (hard cuts, transitions, etc)
	Be able to detect static video sequences (no significative movement)
	Be able to detect very changes video optical flow and suggest the user a change of the orientation.
Perform an user study	Perform an online user study
	Generate routes given different points.
	Integrate front-end and back-end (be able to process video from the user-side web app)

TABLE 1.1: The core and optimal project requirements.

of the manual extracted georeference has been compared with the hardware extracted georeference in order to evaluate the correctness of the method.

The time frame for the completion of the project has a length of twenty weeks. The first task of the project was state of the art research. One aspect was concerning previous work done in the field of georeferencing data in general, with multimedia content in particular. Another aspect was concerning the variety of ways to integrate a map-based and video UI. The next task was to develop the web-based software prototype for allowing the user to capture this georeference by using our method. Finally, a user study was performed to evaluate the usability of the proposed method and to collect the user' data and compare it to the ground truth for evaluating the method's accuracy and correctness.

1.3 Requirements

The core and optimal requirements are described in the table. 1.1

1.4 Outline

The structure of the remainder of the document is as follows. Section 2 describes the related work previously done in the field. Section 3 presents the methods, tools and basic setup that has been used. Section 4 contains the details regarding the implementation of the software prototype. Section 5 details the methods conducted in the user study with the obtained results. Section 6 discusses the conclusions and results obtained in the thesis, then provides suggestions for further research in the area.

Chapter 2

Literature review

2.1 Georeferencing and Geotagging

With the latest increase in the popularity of GIS, a greater necessity to collect and access geospatial data has emerged. One medium of that information is multimedia content, but it has not been considered very often. The usage of the geotagged and georeferenced multimedia content has been proven to be very useful for industry and academia, one being a recent apparition of Large-scale video collections[6] where it is possible to retrieve content by performing advanced spatial operations.

Some uses of well-managed geospatial multimedia content could bring are to provide assistance in a natural disaster or another time-critical situation[7, 8]. This data can be compiled into a descriptive database for different types of analysis. Setting up a good procedure may help for the georeferencing of more multimedia with similar characteristics[9].

Traditionally, the georeferencing process was classified into two different types: by integrated hardware (automatic) and by software solutions (manual). Chippendale P[10].

Nowadays, most of the recording devices provide a set of sensors which allow some kind of geotagging to recorded media. A simple GPS receiver, which is a standard on every smartphone or camera can provide a geographic position to every capture. Alternatively, if the device is connected to a Wi-Fi or a GSM/CDMA network, there are many hybrid methods for estimating a user's position[11].

In addition, these devices usually contain a series of sensors such as a compass and a set of accelerometers which can be used for estimating the orientation of the camera. This information, combined with the focal length of the camera can be used for estimating the

orientation where the camera is facing and the inclination of the device which is sufficient information for constructing a very realistic field of view for that content. Wang, Yin, Seo, Zimmermann and Shen 2013[12] provide a new method for correcting the data collected by phone sensors (compass and accelerometers) which reduce the captured noise, estimates the orientation by performing an optical flow approach in order to get more accurate results in the process of hardware georeferencing extraction in the process of video capturing.

If the user has not activated, do not have access to these technological capabilities, or do not wish to georeference their multimedia on the fly, there is an abundance of software solutions that provide the possibility of manual georeferencing or geotagging such as GoogleEarth¹, Flickr² and Panoramio³.

Other mixed approaches have also emerged. The recent phenomena of crowdsourcing and social media bring an excellent opportunity to fetch and georeference multimedia content. Google has attempted to make the georeferencing process more accessible to the user in order to make the process more efficient. They suggest an automatic approach to identify landmarks and correlate them to already georeferenced photos in a database[13].

In a similar way, the Im2gps project[14] attempts to solve the same problem by creating a database from already georeferenced photos and comparing new sets with geographic keywords and image features (eg. histogram, lines, geometry, etc) by using computer-vision techniques.

Despite their promising results with well-known areas and landmarks, these methods are still in a very early phase of development and not suitable for universal use. That is the reason that proves the necessity of a new method for allowing a deep manual georeferencing which in addition to serving as a substitute of the mentioned above methods in the cases where those methods are not working, could complements and help those methods to generate better results.

2.2 Geographic video databases

As mentioned above, there are numerous applications for retrieving geospatial data from video content but the most extensive are the collection of geographic video information in databases. Arslan Ay et al.[5, 15] defines how to build a 'viewable scene' model of all the information capture by a camera in a video stream by indicating the camera location

¹www.google.com/earth

²www.flickr.com

³www.panoramio.com

and scene field of view, and how to display it in a comprehensive human readable way. Kim et al.[16] provides a more flexible method for specifying the FoV by using a vector approximation in order to take it as accurate as a circular sector FoV representation. Even though it is efficient for querying in terms of time efficiency, it is not flexible enough for performing video search.

Some recent research even combines the geographic based information with other types of metadata like text annotations or visual information retrieved from landmarks[17–19]. Another approach is to use spatial information for performing spatiotemporal queries is to use it next to the one where geographic information is captured by hardware, leading to a more accurate identification of landmarks.

2.3 User study

Once the method has been defined and the prototype implemented, it is necessary to evaluate it. For this purpose, a user study is performed which among other parameters, will calculate the method's usability.

There is a large amount of defined and scientifically proven usability surveys which have become standard in industry and academia. A lot of research and comparison between the different possibilities has been done as the ones carried out by Bangor[20] or Tullis and Stetson[21] where the System Usability Scale (SUS; Brooke[22]) has provided very reliable results despite its simplicity compared to another such as the Questionnaire for User Interface Satisfaction (QUIS; Chin et al.[23]) and the Computer System Usability Scale (CSQU; Lewis[24]).

Even with the antiquity of those questionnaires, they are still currently being used for testing the usability in any range of modern products even though newer systems were not contemplated when those questionnaires were written. Bangor et al.[20] and Sauro[25] have determined that SUS can be applied to any interface in spite of the used technology.

It should also be considered that Bangor et al. 2009 perceived not a strong, but significant correlation between age and SUS score which might suggest that bigger ages do could negatively affect the usability score achieved by a system but they were not able to find a significant difference in gender. The research made by Tullis and Stetson's[21] has also shown that a very valid reliable SUS score can be obtained despite a very small population sample (8-12 user), enabling the retrieving of measures on how people perceive the usability of a system or product.

It has been suggested by Lewis and Sauro[26] that the measures from SUS questionnaire can be divided into two factors. The items 4 and 10 would measure the learnability and the other 8 items will be related to the overall usability of the system. Borsci et al. 2009 also confirmed the existence of this two different factors on the SUS questionnaire but also proved that are correlated. Even if the ISO 9241, part 11 (ISO, 1998[27]) define that learnability as one of the aspects of usability the existence of this two factor could be meaningful for provided more detailed information about the user study.

Finally, Bangor et al.[20, 28] provided a framework for the comparison of SUS scores between different kind of interfaces by correlating the 0 to 100 SUS Score to the letter grades system employed by major universities and to their own adjective scale rate (F to A). Facilitating, in this case, to acknowledge the meaning in terms of usability of individual SUS scores.

2.4 Map based UI and visualisation

Elwood[29] take the concept of GeoWeb previously defined by Haklay et al.[30] and expose how the visualization techniques for geographic data has being adapted from the classical approach in order to fit the new modern phenomena as the volunteered geographic information (VGI) took from crowdsourcing, in this way at the same time that the web has been transformed and become more social and participative the techniques for displaying geographic information has also evolved by using the possibilities that new technologies as HTML5 provide making possible for the users to generate their own content.

For the development the software prototyped used during this dissertation for allowing users to generate their own georeference content it has been taken as an example the GeoVid hosting system for georeferenced video presented by Seo et al.[3], where the videos and their geographic metadata are synchronously visualised in a map-based web app, suggesting a solution for interactive video browsing and querying.

In addition, the user study performed by Çöltekin et al.[31] has been taken as a reference in order to design an efficient and usable layout for the software prototype. In this study, it evaluated two different kinds of typical map-based web user interfaces by combining usability methods with eye tracking records in order to support and identify design problems and lead to better web map UI designs [32].

2.5 Disaster scenario and volunteer driven approach

The concept of GeoWeb mentioned above has had a big impact supporting crisis management, the possibility of the users of generated their own content has encouraged the volunteered crowdsourcing. Zook et al.[33] had analyzed how VGI has been with web-based mapping services during the last Haiti disaster where people could assist governmental organizations and NGOs with relief efforts remotely.

This volunteered and crowdsourced drive approach can be complemented by other techniques which take advantage of georeferenced video for in the topic of disaster management and relieve. Even though this thesis does not aim to manage aerial video footage, it is necessary to mention that the current tendency in the acquisition of georeferenced video is by using UAVs. This kind of footage could be used for mitigating the effects of disaster scenarios. The Duct Fire-fly project[34] proposes a system for effective usage of georeferenced video data recorded by a UAV distributing the information in real time by a disaster management system to all the levels of the fire brigade organization.

2.6 Routing

Nallur et al.[35] provides a comprehensive overview of the free and commercial possibilities in route services and engines that are currently available in the context of being used for smart cities. In their research, they also analyze some of the geographic open data sources that are necessary for running your own route engine. Finally, they show a modification of the GraphHopper engine for using different types of smart city sensor data, which provides other innovative routing criteria such as noise or pollution avoidance.

Chapter 3

Approach, methods and tools

3.1 Approach

The overall method followed during this dissertation was initiated by researching the state of the art of the different techniques and methods for georeferencing multimedia content, dynamic visualization of geographic content and allowing interactivity and user-driver content generation. It was followed by performing several iterations of prototyping until it became a functional web prototype which allowed a successful georeferencing process. Finally, an experiment consisting of a test user study and a definitive user study where each participant had to perform two georeferencing task and fulfill a questionnaire. The resulting outcome has been statistically analyzed in order to estimate how suitable is the proposed method for their initial purpose of being a reliable alternative for the previously existing georeferencing methods.

3.2 Methods

The project has followed an incremental development building model with some iterative elements for the development of the map-based software prototype. This is a software development model which defines a life cycle of a grouped task which is repeating in small iterative steps. This way, it is possible to get feedback at an early stage, which helps improve and adapt the system to the new necessities that are discovered during the development process. In the beginning of the project, an initial set of requirements had been defined then redefined during the development phase according to newly discovered necessities during the regular meetings with the coordinators.

For the evaluation of the suggested framework, a user study has been performed in order to obtain feedback in terms of usability and accuracy. In terms of the usability measurement, the System Usability Scale (SUS) questionnaire approach[22, 36] has been followed.

For analyzing the accuracy, it has been statistically compared the resulting georeference from the user study participants and the georeference extracted by the device hardware in the time of the recording with the ground truth. Currently, there is no simple way of estimating a ground truth. In this case, the same person who recorded the video was the one who did a manual georeference with no time or waypoint limitations to provide the ground truth. However, this approach has its limitations that have been discussed in Section 8.2.

3.3 Tools

The software prototype has been developed and run during the user study under an instance of Apache[37] 2.4.7 (Ubuntu) installed on Linux Mint[38] 17.2 Cinnamon 64-bit.

HTML5, JavaScript, and CSS3 have been the main programming languages used for the development of the software prototype. During the implementation process, several libraries and frameworks have been used. Bootstrap[39] is an open source front-end web development framework which makes more powerful the web development and it is commonly used to grant a responsive design in the resulting web app, which is advantageous for the software to run on any screen size and resolution. For the map-based programming of the web prototype, Leaflet [40] has been used as the main open-source JavaScript library for web interactive maps. Additional front-end web development libraries used were jQuery[41], which provides easy methods to perform AJAX request and simplify some JavaScript operations. VictorJS[42] was used to be able to work with 2D vectors, a functionality which is not included in the core Javascript. IntroJS[43] was used for implementing the UI introductory tutorial.

In order statistically analyze the results, R programming[44] has been used with the intent of extracting valid conclusions. For visualizing the results of the user study, the results were plotted on ArcGIS Desktop[45]

In relation to the hardware used for the user study, all of the software was running under a Lenovo Ideapad Z580¹ laptop connected to a Dell P2414H Monitor² of 24 inches, Dell KB212³ QWERTZ keyboard and a Logitech M185 mouse⁴.

3.4 Other resources

All of the data used in the software prototype, the map tiles and features of the Münsterland region in Germany, have been retrieved from Open Street Map⁵.

¹<http://shop.lenovo.com/us/en/laptops/ideapad/z-series/z580/>

²<http://accessories.us.dell.com/sna/productdetail.aspx?c=usl=encs=19sku=320-9794>

³<http://accessories.euro.dell.com/sna/productdetail.aspx?c=ukl=ens=dhscs=ukdhs1sku=580-17639>

⁴<http://www.logitech.com/en-roeu/product/wireless-mouse-m185>

⁵<http://download.geofabrik.de/europe/germany/nordrhein-westfalen/muenster-regbez.html>

Chapter 4

Software Prototype

4.1 Concept

The main purpose of the software prototype is to provide a method for the users to reproduce the video in a comfortable way while they interact with the map. One way of creating an effortless experience is to assist them in the process of matching the locations where the video has been recorded with actual places on the map. Every time the user identifies a location, the interface will introduce it to the system by placing the position and orientation of a video key-frame in the map in the form of a waypoint which will then be linked to that temporal moment in the video stream. The system will calculate the most probable route between different waypoints introduced by the users and then calculate the intermediate values for position and orientation for the keyframes between the ones that have been georeferenced by a waypoint. This allows the user to preview the calculated result before submitting it.

4.2 Implementation

4.2.1 Layout

For the implementation of the software prototype, the Bootstrap framework was employed to fulfill a responsive design approach. Even though the prototype is not designed to be used on mobile devices such as smartphones or tablets, it was considered important to be adapted to any kind of screen resolution, aspect ratio, or browser zoom level.

The chosen layout for implementing the map-based UI with synchronous video can be seen in the Figure 4.1. Bootstrap divides the device screen in a using grid system of 12

columns that adapt itself to any device viewport. The illustrated size combination was designed for providing a comfortable way to visualize data and interact with the UI.

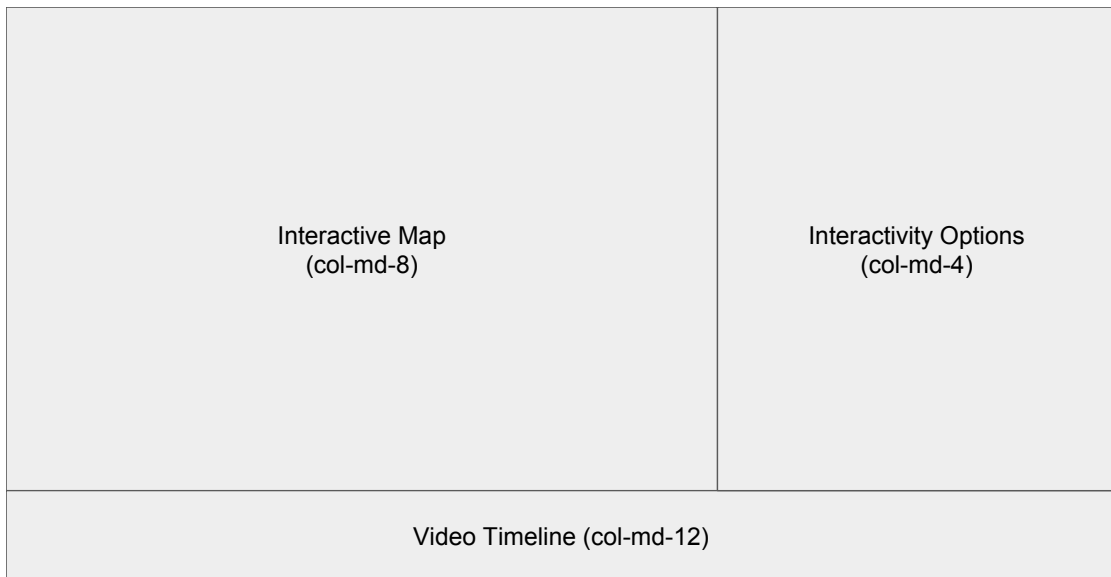


FIGURE 4.1: User Interface layout.

4.2.2 Interactive map

One of the first implementation decisions were to choose the services to retrieve the data necessary for building a web-based user interface. After considering a variety of commercial and open source options, the decision to keep the software prototype free of commercial use as much as possible.

Leaflet was the chosen library for providing map visualization and interaction, due to their open-source nature, light weight, simplicity and general acceptance in the developing community. Leaflet has well adapted to our purpose and it was very simple to extend it to our custom necessities. Regarding to the coordinates reference system used it has been left the Leaflet default, the WGS84 Web Mercator (EPSG:3857) with a Spherical Mercator projection. Which is the most common configuration for online maps, and it used by all free (as OSM) or commercial (Google Maps) tile services.

Leaflet does not include any tiles. The one that has been chosen are the OSM map tiles. This was assuming that there would be a greater amount of detail in order to enhance spatial orientation of the users, compared to another service such as Google Maps. Because of the large community support and the constant updates, OSM data was chosen for the project.

OSM data is free for everybody to use¹, although OSM tile servers are not. As long as the use of the prototype is not too heavy and proper attribution is displayed, then it is not a problem. Other alternatives to this configuration are to run a private tile server using OSM data or select another free tile server based on OSM data².

4.2.3 Geocoding, reverse geocoding and routing

With the intent of making the process of georeferencing as simple as possible for the user, the software prototype performs a geocoding operation with the metadata of the video, such as the name of the location. This geocoding operation assigns a location (in geographic coordinates) to a postal address, thus providing a list of possible locations to the user.

Even though the non-commercial software policy was followed during the project, it is worth mentioning that after testing different web geocoding services, it was concluded that the Google Maps API was returning more accurate results than other open-source APIs such as the OSM Nominative.

The software prototype also performs the opposite operation called reverse geocoding, which attempts to translate the first point of the video in a postal address in order to store it as part of the metadata of the resulting georeference. In this case, both APIs were returning similar outcomes. Therefore, It was decided to use OSM Nominative because their non-commercial nature.

The most important service used in the project was the one used to calculate the route in between the way-points introduced in the system by the user. Due to the importance of calculating a good route from that intermediate points and the fact that this calculated route will be explicit part of the resulting georeference, a more extensive view of the available technologies has been provided in Table 4.1

Finally, GraphHoper was used to run a local process of the code in the same local³ machine where the web prototype was executed. In this case, we reduce the latency of downloading the result and are able to guarantee that the route service is always available.

¹<http://www.openstreetmap.org/copyright>

²http://wiki.openstreetmap.org/wiki/Tile_usage_policyAlternativeOpenStreetMapTileProviders

³As GraphHoper is open source code, a local instance has been downloaded and implemented with OSM from Münster region.

	Google Maps API	Open Route Service
N° Request	2.500 day	Unlimited
N° Waypoints	8	0*
Transport	Foot, Car, Bike and Public Transport	Foot, Car and Bike
Data Availability	Global	Global
Data Source	GoogleTM	OSM
	GraphHopper	GraphHopper Local¹
N° Request	50.000 day	Unlimited
N° Waypoints	10	Unlimited
Transport	Foot, Car and Bike	Foot, Car and Bike
Data Availability	Global	Münster Region ¹
Data Source	OSM	OSM

TABLE 4.1: Main Route Services comparison.

4.3 Way-points positioning and orientation

For the user to place waypoints on the map, they simply click the locations on the map layout. Leaflet automatically converts the pixel coordinates into the geographic coordinates.

In the process of reconstructing an artificial Field of View (FoVScene) of the video scene in order to fulfill the requirement established by Arslan Ay et al.[5], a circular sector is usually built. However, some methods often lean towards for simplifying the geometry of the scene FoV by approximating it to a minimum bounding rectangle[16]. In this case, it has been opted to follow the OGC 05-115 specification for Video Web Services[46] where are suggested methods of representing a video field of view in a web environment. Following that specification, a mixed approach for representing the scene FoV by using polygon containing three geographic points. The result is a cone shape, where the first point is the video position, and the second and third points are situated 30 meters north from the origin in an angle of 60 degrees.

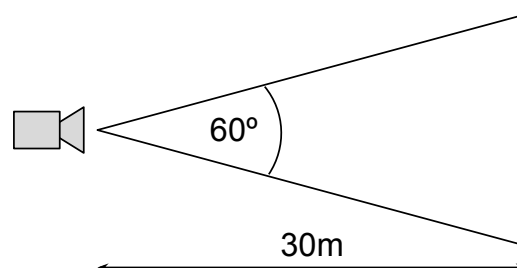


FIGURE 4.2: Reconstructed FoV model.

If the video metadata contains information about the focal length, a more accurate field of view cone could be simulated. This default configuration suggested by the OCG

adjust perfectly to our purpose due to the urban environment where the videos were recorded. This, however, would not be completely accurate for another environment such as countryside, where the lack of vertical obstacles would allow the camera record at a larger distance.

Once the FoV polygon is constructed, the user can rotate it until they have the desirable angle which fits accurately represents the video. For performing the rotation, the following formula is used:

$$\begin{aligned}\varphi' &= \varphi_c + \cos(\alpha) * (\varphi_0 - \varphi_c) - \sin(\alpha) * (\lambda_0 - \lambda_c) * \cos(\varphi_c) \\ \lambda' &= \lambda_c + \sin(\alpha) * (\varphi_0 - \varphi_c) / \cos(\varphi_c) + \cos(\alpha) * (\lambda_0 - \lambda_c)\end{aligned}$$

Where a geographic coordinate (φ, λ) is rotated around a centre represented by another geographic coordinate (φ_c, λ_c) by a given α value in degrees.

4.4 Interpolation

The waypoints that the users introduce to the system may not contain sufficient information for getting a complete georeference. It is necessary to calculate the intermediate values that the video could follow between those waypoints. As long as the system has information about in which time has been placed, each waypoint, and the distance between those waypoint routes, it is possible to apply interpolation algorithms for estimating the position and orientation for those intermediate time values.

In the case of the position it has performed a linear interpolation (LERP), as it is the most simple interpolation method and is fast to compute. Although it does not offer the most accurate results, it is still suitable for our purpose because the generated routes contain points that are very close to distance; even the haversine distance formula is not linear, it behaves like so. For performing LERP, it is necessary to consider the total time, total distance (which is provided by our routing service), and the individual distance between each point(as mentioned above). For the same reasons, Spherical Interpolation algorithm (SLERP) has been performed for interpolating orientation.

Chapter 5

User Study

The purpose of this dissertation is to suggest a new method for georeferencing and evaluate how suitable is for achieving its purpose. The performed experiment aims to measure the efficiency and usability of the resulting framework, in addition to retrieving output from the resulting geo-metadata of the manual extracted georeference in order to evaluate the correctness of the method. The participants were asked to perform a series of video georeference processes while their input and time was recorded.

5.1 Experimental design

Without considering the training task, the resulting routes were designed in order to have different orders of complexities[47]. Four different routes have been recorded: two simple and two complex, as it is shown in Table 5.1.

The routes have been randomly assigned to each participant. The participants were given the less time for the simple routes and more time the complex ones. The dependent variables are the completion time for measuring the efficiency and the accuracy of the result to measure the quality.

	Route 2	Route 4	Route 5	Route 6
Complexity	Complex	Complex	Simple	Simple
Distance (m)	673	654	342	339
Streets	6	5	3	3
Intersection	5	3	2	2
Direction changes	4	4	1	1
Video duration (s)	476	498	256	271

TABLE 5.1: Route complexities.

The SUS approach was chosen for measuring the usability of the proposed method was because of its simplicity, reliability and effectiveness even with small population samples, such as the one in the user study.

5.2 Participants

Thirteen people with an average age of 27 years (SD: 2.5, Range: 22-31 years) and an equal balance of gender (7 males and 7 females) participated in the experiment. All of them had university-level training related to geoinformatics. Not all of participants had English as their mother tongue but were all fluent English speakers. It also important to mention that all of the participants work or study in the building around the videos were recorded. Therefore, all of the participants answered that they were familiar with the environment. For the user study, no monetary compensation was offered to the participants.

5.3 Stimuli

For the user study, an experiment document has been provided to the participants (see Appendix A). The document consists of a brief introduction to the experiment, a series of three qualitative preference questions, the SUS questionnaire, and again the three qualitative preference questions in a negative order, in order to avoid acquiescence bias. This is a phenomenon in which the user tends to always agree with the question and detect random answers. This is not crucial with the SUS questionnaire because it is already mixed positive and negative questions in their schema.

The experiment has been performed on a Linux workstation running Google Chrome Internet Browser. The hardware provided to the user to perform it was a 24' monitor, a keyboard, and an optical mouse.

5.4 Procedure

After welcoming the participants, they were requested to read the introductions of the experiment. They were then asked to perform the UI tutorial. The participants were then requested to complete the georeferencing of the training video, with the interviewer providing verbal instructions for them to be able to complete the georeferencing of the training video in no more time than 5 minutes. During the training, the participants

were allowed to ask unlimited questions about the usage of the UI to the interviewer, but not about the georeferencing method (e.g. how often place a new waypoint). The purpose of the training is to enable the participants to become confident with the user interface. The 5-minute training limit is imposed for not allowing some participants to gain more experience with the system which could lead to some participants performing better than others in the real user study.

Afterward, they were asked to complete the georeferencing independently on the simple video and the complex video within 5 minutes each. During the training session and each georeferencing task, the participants were informed every time a minute had passed. After completing the georeferencing tasks, participants were asked to fill the SUS and NASA-TLX questionnaire and responded to three additional qualitative preference questions. After completing the questionnaire, participants were asked to provide some additional feedback to see if they considered it appropriate the study. The participants were dismissed then thanked for their participation.

Chapter 6

Results

6.1 Software Prototype

In Figure 6.1, it is shown how the final software prototype has allowed the manual georeferencing of videos. The final layout and interactive elements were adjusted based on the feedback received by several pilot user studies, where it was revealed the necessity of eliminating the video controls integrated by HTML5 and substituted by a custom ones. It was also necessary to add new controls for modifying the speed of the video, allowing the user to speed up instances of the video where there were no significant events.

A slightly different version was used for performing the user study, in which the only difference was the removal of the new video submission element and the addition of an input field for annotating the participant number in each experiment.

6.2 User Study

6.2.1 Qualitative questions

A correlation analysis was conducted to know how well the answers to the positive questions matched the ones of the negative questions. The results found for the items Q1 and Q4 related to the environment knowledge were highly significant ($T=-0.72 > \pm 0.7$ and $p=0.003 < 0.05$) where the distributions were: 5 (50%), 4 (3.7%), 3 (14.3%), 2 (0%) and 1 (0%). On this scale, 5 meant the maximum knowledge of the environment and 0 the minimum knowledge. Meaning that most of the user have a good knowledge of the environment where the videos were recorded.

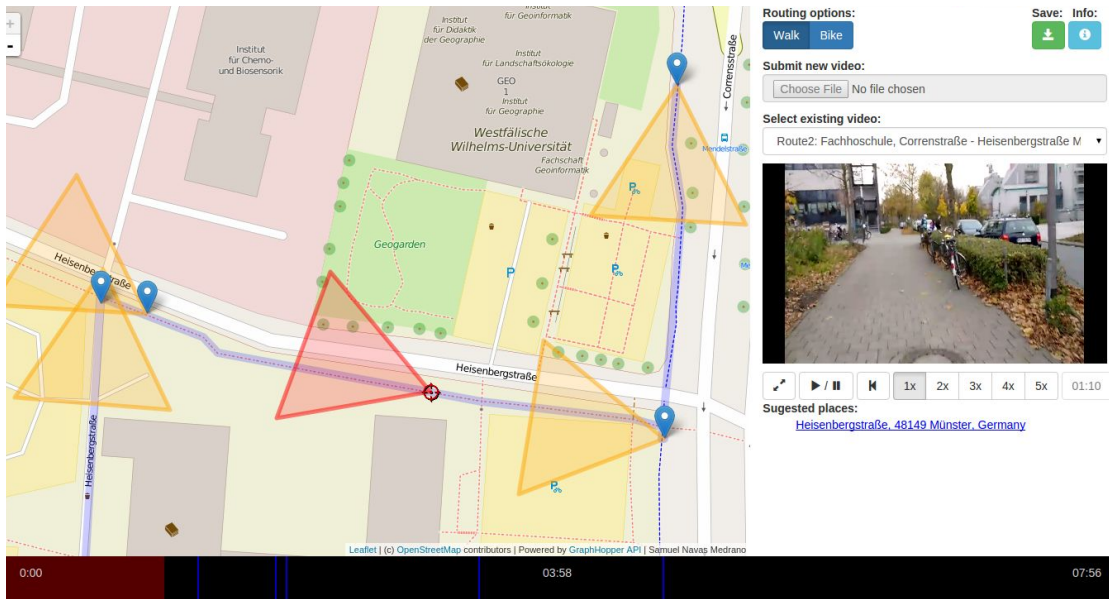


FIGURE 6.1: Screenshot of the resulting software prototype.

However, items Q2, Q3, Q5 and Q6 has shown no correlation between them ($T=-0.13 < \pm 0.7$ $p = 0.59 > 0.05$ and $T=-0.15 < \pm 0.7$ $p = 0.52 > 0.05$). Due to this fact, those items were not taken in consideration. That result could be given because most of the user did not have temporal awareness while they were georeferencing the videos, despite the verbal announcements of the interviewer each minute.

6.2.2 Task Completion Time

As it was written in the experimental design, the participants had a time limit of 5 minutes for georeferencing each video. The simple routes were approximately 4:30 minutes while the complex ones were 8 minutes.

As it is shown in the Figure 6.2, by the Kruskal Wallis test (chi-squared=4.79, p-value=0.1872) on average, the participants had spent the 5 minutes time limit in trying to complete the georeference in all the routes. It should be noticed that the variability is increased in the simpler routes (R5 and R6), where it is possible to find a wider range of completion times as we can see in the Figure 6.3.

6.2.3 Usability

After performing all of the tasks of the user study, the participants were asked to answer the items of the SUS questionnaire.

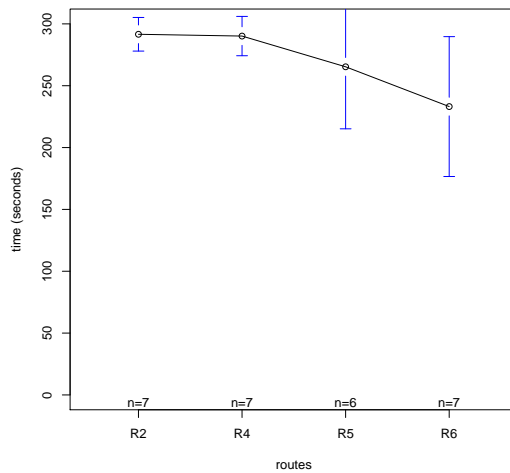


FIGURE 6.2: Group means and confidence intervals.

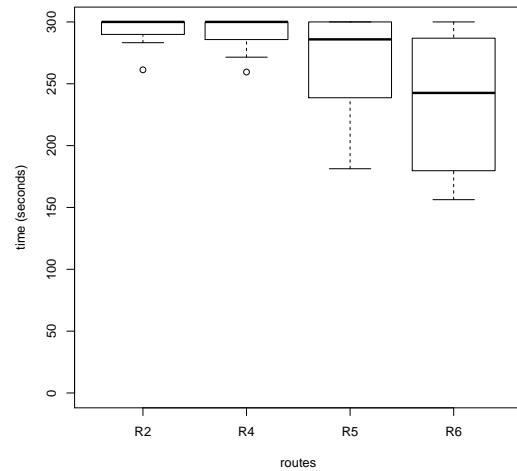


FIGURE 6.3: Boxplot of the different times for completing the georeferencing.

Only one observation was removed from the sample because it was placed several standards deviation far from the mean as it is shown the Figure 6.4. All of the resulting data has shown a homogeneous distribution around the SUS score 80. Even the individual analysis of the SUS items has shown results with an average under 0.3 as it is displayed on the Table 6.1 with the exception of item 4 which has an average a bit higher (0.32), this item next to the item 10 are the ones which measure the learnability of the system.

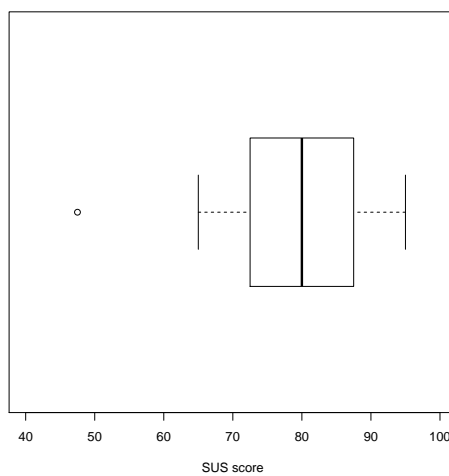


FIGURE 6.4: SUS score boxplot.

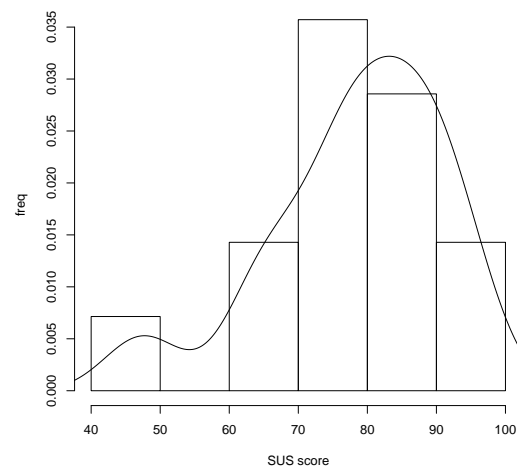


FIGURE 6.5: SUS scores histogram.

SUS Question From strongly disagree (0) to strongly agree (4)	Mean	SD	VAR
S1. I think that i would like to use this system frequently	3.36	0.74	0.22
S2. I found the system unnecessarily complex.	3.28	0.85	0.25
S3. I thought the system was easy to use	3.14	0.66	0.21
S4. I think that I would need the support of a technical person to be able to use this system	3	0.96	0.32
S5. I found the various functions in this system were well integrated	2.85	0.66	0.23
S6. I thought there was too much inconsistency in this system	3.28	0.61	0.18
S7. I would imagine that most people would learn to use this system very quickly.	3.14	0.66	0.21
S8. I found the system very awkward to use	3.57	0.51	0.14
S9. I felt very confident using the system	3.07	0.61	0.20
S10. I needed to learn a lot of things before I could get going with this system	0.28	0.72	0.22
TOTAL	80.7	9.6	0.12

TABLE 6.1: SUS scores summary by questions.

6.2.4 Accuracy

For estimating the accuracy of the participant results, an error between the data generated by the participants and ground truth was estimated. There were two different variables to compare: the position and the orientation. It has been calculated the root mean squared error (RMSE) of the two variables by using the following formula:

$$Position_{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^n \overline{haversine(ParticipantPosition_{ti}, GroundTruthPosition_{ti})}^2}$$

$$Orientation_{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^n (ParticipantOrientation_{ti} - GroundTruthOrientation_{ti})^2}$$

As illustrated in the Figure 6.6, the participant sample of the position does not grow much higher than 100 meters with the exception of some outliers. However, the variation (σ) of the data is too high, which shows a heterogeneous nature of the sample as we can see in the Figure 6.7. The same kind of behavior is observed for the data obtained from the orientation of the georeference, where the samples neither goes higher than 100 degrees of error from the ground truth, as we can see in the Figure 6.8 and Figure 6.9.

One way of knowing the quality of the obtained results is to compare them to the one obtained by the hardware extraction. The wide range of confidence intervals caused by the heterogeneity of the participants data make difficult to compare with the data extracted from the hardware georeference process. Just to compare the population mean of the participants data with the hardware georeferencing would not be enough because that procedure will produce a distorted result.

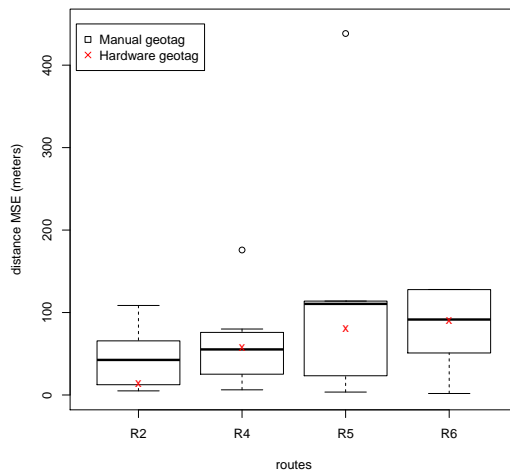


FIGURE 6.6: Position RMSE box-plot.

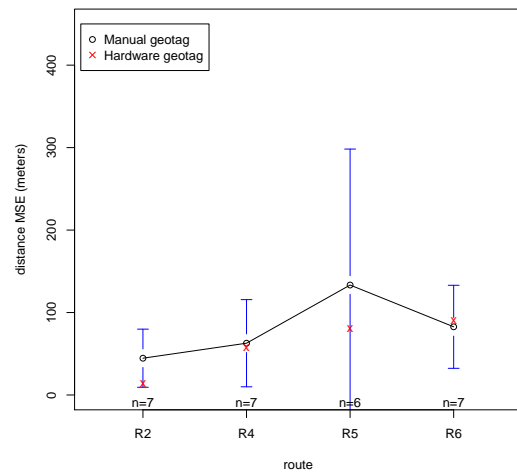


FIGURE 6.7: Position RMSE confidence intervals.

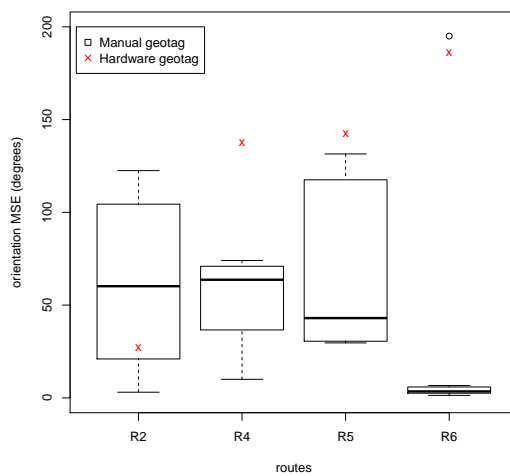


FIGURE 6.8: Orientation RMSE box-plot.

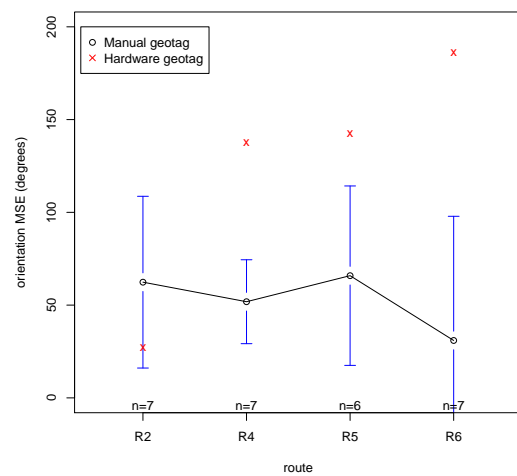


FIGURE 6.9: Orientation RMSE confidence intervals.

For getting a reliable comparison between the average of the two kinds of georeferences, it is necessary to perform a non-parametrical analysis of the variance. In our case, it is the Wilcoxon single-rank test. Another alternative method would be the paired Students' t-test for matched pairs. However, due to the high variability of the population which is not normally distributed, Wilcoxon method is much more reliable.

As it is showed in the Table 6.2, the mean of the manual extraction of georeference performed by the participants is not significantly different from the georeference extracted by hardware at the moment of recording the route videos.

Routes	p-value	Proved hypothesis
R2	0.1563	RMSE Manual extraction = RMSE Hardware extraction
R4	0.9375	RMSE Manual extraction = RMSE Hardware extraction
R5	0.8438	RMSE Manual extraction = RMSE Hardware extraction
R6	1	RMSE Manual extraction = RMSE Hardware extraction

TABLE 6.2: Wilcoxon signed rank test of the position RMSE.

Routes	p-value	Proved hypothesis
R2	0.1563	RMSE Manual extraction = RMSE Hardware extraction
R4	0.0078	RMSE Manual extraction < RMSE Hardware extraction
R5	0.0156	RMSE Manual extraction < RMSE Hardware extraction
R6	0.0156	RMSE Manual extraction < RMSE Hardware extraction

TABLE 6.3: Wilcoxon signed rank test of the orientation RMSE.

However, we find a slightly different result at the Table 6.3, where the mean of the manually extracted data shows an improvement from the one extracted by hardware. Meaning that on average (p-value < 0.05), the participants have performed better than the GPS and accelerometers sensors of the phone.

6.2.5 Empirical observations

Due to performing different iterations of the user study there are some behaviors observed on multiple occasions in some of the participants:

In general, the participants showed problems for identify the position of the video in a large straight road. This problem could be related to the loss of depth information in 2D video and the lack of skills for the human eye to identify it if there are not enough landmarks or if the participant does not pay attention to landmarks. This phenomenon has been responsible for some participants turning in the wrong crossroad sometimes.

Some participants also had problems identifying which direction the video was turning, especially when the direction of the video was different than North-South. This phenomenon also explains why on Route 6, which is the one oriented in a south to north direction shows such a big difference in the orientation error respect to the other routes. As it is explained above, this effect would be highly mitigated with the possibility of rotating the map, allowing the user to always be oriented in a down to the top perspective of movement.

Video title	Zoom level	Scale
Route2: Fachhochschule, Correnstraße - Heisenbergstraße Münster	17	1:4,000
Route4: Neubau Pharmazieinstitut, Corrensstraße Münster	16	1:8,000
Route5: Institut für Anorganische and Analytische Chemie, Corrensstraße Münster	16	1:8,000
Route6: Correstrasse - Mendelstraße Münster	17	1:4,000

TABLE 6.4: Results of geocoding the video names by the Google API.

Another noticeable piece of feedback from the participants is their lack of experience with the OpenStreetMap¹ map layers and claimed to be more familiar the Google Maps² and suggest that the task would be easier to orientated in a Google Maps map.

However, as OSM map layers show much more landmark information than Google Maps, it is reasonable to say that after the user gets more experience with the OSM representation of spatial features, the user orientates themselves more easily which will provide more accurate results.

Smalls levels of initial zoom after performing the reverse geocoding from the video metadata made it difficult for the participants to find the video starting position. Some participants were unable to find the starting position correctly after the 5 minutes time-frame provided for a task, even if the starting point was the building where they usually work or study. This event influenced the result of accuracy, because of the people who spent more time finding the starting point had been rushed for being able to complete the georeference in time. On Table 6.4, we can see the levels of zoom obtained after performing geocoding of the video title. Although the zoom level does not show a statistic correlation (p-values highers than 0.05) with the accuracy of the results due to the heterogeneity of the data (perhaps caused by the different spatial cognition skills of each participant), it is important to mention this phenomenon.

¹www.openstreetmap.org

²maps.google.com

Chapter 7

Discussion and evaluation

7.1 Discussion

The positive and negative items in the user study questionnaire answered by the participants revealed a correlation in the knowledge of the environment fact but not on the spent time fact. This odd phenomenon could have several interpretations. All of the participants were working or studying in the same building, which is the common starting point for all the routes, therefore, most of them had shown knowledge of the environment where the videos were recorded. As they had shown no acquiescence bias, it is hard to prove that they answered the next item randomly. Perhaps the reason why the participants did not show a correlation in the time spent on the question is because they were too focused on the completion of the referencing task that they were not aware of how much time they had spent on it.

Regarding the average task completion time, not much of a conclusion can be extracted, contrary to the expected most of the participant spent the full time limit to complete the georeference of the simple and complex routes. However, it can be appreciated that the graphic results showed more variance in the completion time. This is because even if the average of the participants reached the time limit, more participants did not in the simple routes than in the complex.

The average SUS score has evaluated the system to be quite usable by the participants. SUS determines usability by providing a score from 0 (negative) to 100 (positive). However, the SUS score can not be measured as a percentage. While it is technically correct that a SUS score of 70 out of 100 represents 70% of the possible maximum score, it suggests the score is at the 70th percentile.¹ Unfortunately, 80 points of mean (SD=9.6)

¹<http://www.measuringu.com/sus.php>

means nothing on their own and needs to be compared with the SUS score of other products and systems. Bangor[20] proved that by performing a one-way analysis of variance that the SUS score of different products varies significantly by the specific kind of user interface design being tested ($\alpha = .05$, $p < .001$). They separated the different types of UI into cellphone equipment, customer premise equipment, graphical user interface for OS-based computer interfaces, interactive voice response phone systems and Internet-based web pages and applications. In the case of our method, it fits in the last category of the web-based UI, in where was measure the SUS score of 1180 different systems getting a mean SUS score of 68 (SD=21), which is significantly lower than our method result. According to Bangor[28], the resulting 80 SUS score would represent an excellent score in their proposed adjective ranking and a B in a university score rank. This means that the usability is within an acceptable range.

During the design of the user study, it was assumed that due to the common background in geoinformatics of the potential participants, all of them will yield similar experiences using map interfaces and have spatial cognition and map orientation skills. However, the accuracy of the performed results has resulted in a very heterogeneous sample. In addition, this phenomenon had been also observed during the user study. The participants showed different levels of spatial cognition skills and usage of map based user interfaces. These two facts could have had a major effect on the accuracy of the resulting sample. However, as there is no immediate way to test the level of spatial cognition and map orientation skills in each person and there were no questions about the participants' experience using mapping interfaces, it was not possible to correlate to the results. Nevertheless, this could be applied to future research on the topic.

Initially, it was predicted that the difficult routes would yield less accuracy but the results did not reflect this. This could be because of the route difficulty classification method. For example, results from Route 2 (complex) show a similar orientation error to the resulting from Route 5 (simple) while the Route 4 (complex) and Route 6 (simple) shows a decreasing of the orientation mean error. This means that the estimation of the routes complexity have not done that well and there are other variables that could have influenced the results.

One reason could be the orientation of the direction changes. During the experiment, users often got confused when there was a turn and the video is not facing south to north orientation. As Route 5 had many of these turns, it may explain why there was less accuracy.

Overall, the mean position error of the manual georeference is significantly similar to the hardware georeference. In the case of the mean orientation error, the manual georeference is even better than the hardware one. This result does not mean that the method

proposed in this dissertation is always better than performing a hardware georeference because there are many factors that can differ in the results. However, it is proved that is a reliable alternative when the videos have been recorded without a deep georeference.

7.2 Evaluation

The most prominent conclusion extracted from the user study results is that the proposed method is in fact, usable. Most of the users showed a favorable reaction to it, making clear that the framework is acceptable for the task of the georeference extraction.

Regarding the effectiveness of the method, the discussion is more complex. Due to the small size of the sample and the heterogeneous nature discussed above, it is more difficult to extract any apriori conclusions. The most probable reason that causes that spread in the obtained data could be the difference in the spatial cognition skill in the participants' sample. However after performing several Wilcoxon Tests, it is statistically correct to say that the spatiotemporal accuracy of the manually georeferenced data extracted by the proposed methods is equally accurate as the one hardware extracted in case of the position and even better in the case of the orientation. However, it is necessary to mention the presence of techniques for mitigating the hardware extracted georeference error in orientation as the proposed by Wang et al.[12] for correct orientation sensor-extracted data.

In the best case, the participants have generated a manual georeference that correspond quite well with reality in terms of temporal position and orientation. In the worst case where the process of georeferencing has been not completed or the participant has been spatially disoriented and provided inaccurate results, this information can be taken to extract some useful insight. For example in the case of the Participant 10, the georeferenced route was turned in the wrong crossroad however the resulting georeference is able to maintain the shape of the original route, as we can see in the Figure 7.1.

The overall results demonstrate that the proposed method is very suitable for retrieving manual deep georeference in the post recording phase of editing video content. The technique is also compatible with other existing geotagging or georeferencing approaches.

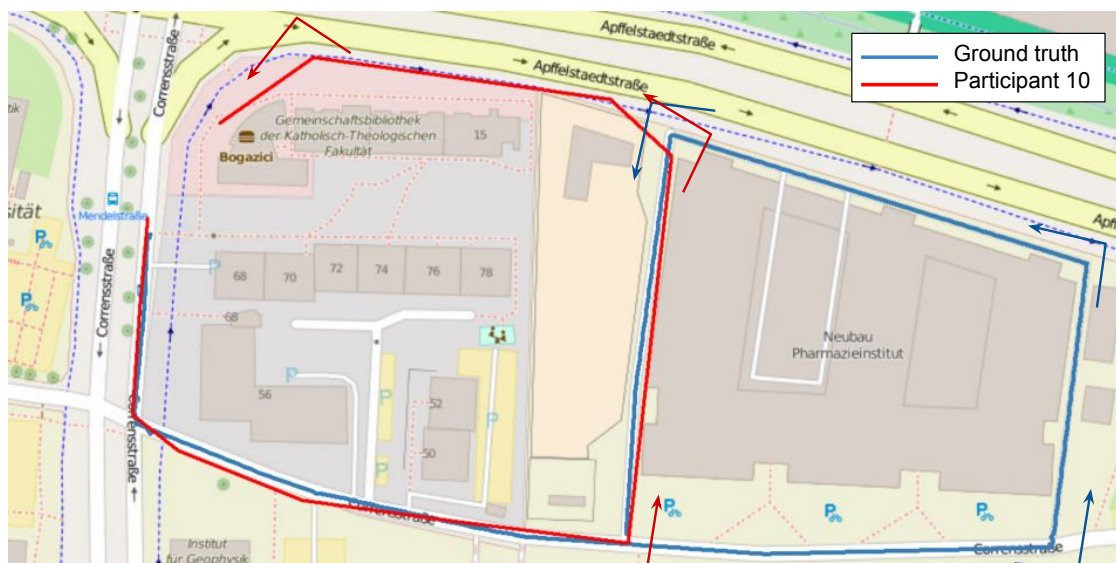


FIGURE 7.1: Participant 10 georeference and ground truth comparison.

Chapter 8

Conclusions

8.1 Contributions

This thesis contributes to the area of GIS, by providing a new method and perspective to the field of georeferencing spatial information in videos. It is motivated by the lack of possibilities for extracting deep enough geographic metadata (position and orientation in time) for fully taking advantage of the new applications of georeferenced multimedia content. This dissertation aims to test the hypothesis that this method is suitable for the purpose of georeferencing multimedia content in terms of:

- *Usability*: the method is not discouraged from continued use
- *Accuracy*: the generated geographic metadata is close enough to the ground truth

The first result of the performed user study implies a high grade of usability of the method on a scale which has been scientifically proved by previous studies. The second result provides strong evidence in favor of the correctness of the hypothesis. The user was able to perform successful georeferences in limited time to obtain a similar accuracy to the hardware extracted ones.

The creation of this method was driven by the intent of it being used in a volunteer-driven approach where the user can use this method in order to provide user-generated geographic metadata to the community eg.in a disaster scenario. These cases would provide an opportunity where the suggested approach would be useful, due to the fact that the spatial information has probably changed by external factors and the previous data is not longer correct.

8.2 Limitations

The clearest limitation of the implemented research is the small size of the sample in consequence of the small number of the user study participants due to the volunteer and non-rewarded nature of the experiment. This fact, in addition to the different variety of spatial cognition profiles, causes a big variability in some of the user study results as is the competition time and the accuracy. This limit influenced the extracted conclusions.

Due to the nature of the experiment, all of the videos have been recorded (and geographically tagged) using the same non-professional hardware. This fact makes the accuracy comparison between the manual and the hardware extraction less reliable. It is virtually impossible to extract for a single video different hardware georeferences. To prepare for the user studies, it would have been possible to employ a difference device for recording each route. However, this possibility would only be added another independent variable to the experiment making the resulting results even more variability, which has proven to be problematic.

During the performance of the dissertation, some non-realistic assumptions had been made. The most obvious is the acquisition of the ground truth data. Since the GPS data was not accurate enough to be considered the ground truth, manual extraction had been done by the same person who recorded the videos to accurately analyze the results. This fact could have lead to a mitigation of the manual georeference error.

The last limitation of this thesis is imposed by the flexibility of the data used. In this case, the method relies on the correctness of the data provided by OSM, but this data is more or less complete depending on the specific place. In the case, of the made user study, some of the paths where the video was recorded, like the main streets, were represented by the OSM with a pedestrian sidewalk separated from the traffic circulation, but some more secondary streets are not represented with traffic-segregated sidewalks. This affected the approach with a loss of accuracy. The routing service provided by GraphHoper and the way that it is implemented on the software prototype is limited if the video goes trough areas where there is not defined way, as a park or big garden. However, these kinds of limitations could be easily avoided in future work.

8.3 Future work

Due to the innovative nature of the proposed method, there is a wide range of possibilities leading to improvements and future research. The most obvious is to extend the method for fitting in not only terrestrial footage but also consider aerial, this would be a great

improvement due to the popular increase of the usage of UAVs for recording video especially in critical disaster situations.

In this project, it was assumed that we would not have the information about the recording device focal length. In this case, it has been opted to reconstruct the field of view artificially following the OCG specification. However, letting the user choose their own focal length, or even detect it automatically from the metadata of the video would lead to a more realistic field of view. In addition to considering the changes of vertical orientation in the recording devices, because extremely low and high values of vertical orientation will lead to a shorter field of view.

The most intermediate advance is to make the method more robust to sudden orientation and speed changes. Improving the interpolation or maybe offer different interpolation algorithms (Bilinear, Bicubic, Bezier, etc) and settings to allow the user decide which one fits better for each segment of the route between waypoints as the animation software like Maya[48] or Blender[49] does.

Another option for avoiding odd and unrealistic inputs by the user is to detect odd changes of orientation and speeds and inform to it to the user. Calculating the speed between each segment and the overall speed of the route it could be possible to inform the user of unprovable situations.

A computer vision approach to the method it could be also very interesting. From the possibility of determining the optical flow of the scene[50] the system could be able to calculate the changes in the horizontal orientation and even the vertical inclination of the recording device, and in this way assist the user in the process of georeferencing. Even to performing landmark identification, it could help the user to have better spatial cognition and mitigate the confusion effect in the situation mentioned above in this document.

Following the line of preventing the situation where has been discovered that the user gets spatially disoriented. The possibility of rotating the basemap could solve the doubt of some user when the video is turning left or right in an intersection. Leaflet does not implement this kind of basemap rotation as a core feature and it is not planed to add it in future updates[51]. Another possibility is to implement an option for letting the user perform its own geographic search by a geocoding operation could also help to avoid the situation where the suggested places tool was not showing an accurate enough location and the user get lost looking for the video starting point.

The resulting software prototype has only made with the purpose of being effective in the user study, but making it usable for a general public would allow the potential user to georeference their own kind of video, uploading it directly to the system or just

retrieving it from social media. For this objective, it would be very recommendable to integrate a back-end feature for analyzing the input video in order to detect hardcuts and transitions, being able to divide any clip into subclips and georeference it separately.

Once the system has been generalized for any kind of public usage, applying to it a gamification[52] would motivate user and volunteer communities to put more effort into the georeferencing process, thus obtaining more georeferences with the potential to be more accurate.

Last but not least, the possibility of applying advanced spatiotemporal analysis on manual georeferences results from the user study could provide a more scientific overview about why some of the complex routes yielded better results than the simpler ones. This could lead to identifying which kind of places and/or situations are sources for less accurate georeferences.

Appendix A

User study document

“Enabling post-recording deep georeferencing of walkthrough videos: an interactive approach”

- User Study -

Age:	Gender:	# Participant:
------	---------	----------------

Note: This experiment has an estimated time duration of 30 minutes, you can interrupt and quit the study at any time.

Introduction

Nowadays, it can be observed an increase in the Geographic Information System (GIS) popularity. Almost all the information that has been captured has a geographic or spatial component which is considered essential for many organizations with different purposes, being useful for research and industry. Therefore having access to accurate data has become crucial for analyzing phenomena. Some of that geospatial data are, in fact, multimedia content such as pictures or video footage. Having a large-scale database of georeferenced video stream can be really useful for many applications. Due to the fact that the video capturing devices have become ubiquitous, a good source for the acquisition of a large amount of video contents is the crowdsourced approach of Social Media.

However these social apps usually don't support geo metadata or it is very limited to a single location in Earth. In other cases, the regular user usually doesn't have the required hardware and software to capture video footage with a deep geo-reference (position and orientation for the whole video duration). There is a clear lack of methods for the extraction of that spatial component in video footage.

This study proposes and evaluate a new method for the manual capture and extraction of the spatial geo-reference in the post production phase of video content. The proposed framework is based on a map-based user interface synchronized with the video stream. The objective of the user will be to link fragments of the video stream with their correspondent spatial position on the map for that time, placing waypoints on the map each time the user identifies a known location in the video footage.

1. I am familiar with the environment of the video

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

2. I had enough time to complete the geotagging of the first video

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

3. I had enough time to complete the geotagging of the second video

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

1. I think that i would like to use this system frequently

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

2. I found the system unnecessarily complex.

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

3. I thought the system was easy to use

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

4. I think that I would need the support of a technical person to be able to use this system

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

5. I found the various functions in this system were well integrated

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

6. I thought there was too much inconsistency in this system

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

7. I would imagine that most people would learn to use this system very quickly.

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

8. I found the system very awkward to use

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

9. I felt very confident using the system

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

10. I needed to learn a lot of things before I could get going with this system

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

1. I am not familiar with the environment of the video

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

2. I felt rushed to complete the geotagging of the first video

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

3. I felt rushed to complete the geotagging of the second video

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

Appendix B

Results visual comparison

This appendix is to serve as an auxiliary overview of the user study results. It contains a map of the participants' manual georeference (in semitransparent blue) in comparison to the ground truth (green) and the GPS data (red crosses). To have visual simplicity, only the position are compared in those maps and not the orientation. In that map is easy to appreciate where the different manual geotags concentrates due to the high intensity of the blue highlighter and also how some participants have diverged from the correct route by the light intensity of the blue traces.

B.1 Route 2 (complex)

It can be appreciated how the majority of the users georeferenced the video through the sidewalk, where the blue color are more concentrated, with the exception of some users who chosen the road instead. It could be also appreciated the points where some users had to stop the geotagging process because they reached the time limit.

B.2 Route 4 (complex)

One route 4, it could be appreciated how one error in the geotagging process was intensified by the effect of the route engine. In the process of surrounding the Pharmazeutische Institute building by its footpath, the user would select the street next to it (Apffelstaedtstraße), as those two ways are not connected the route server has considered that for reaching that point the most optimums route is going back all the way around. As it is represented by the shade of blue in Apffelstaedtstraße, this error has been performed by several participants even though the paths that follow the ground truth has a more

intense shade of blue which means that the majority of participants has followed that path.

B.3 Route 5 (simple)

Due to the simplicity of this route, the only appreciable detail is how one of the participants has gone away from the correct path by following the road instead of the sidewalk.

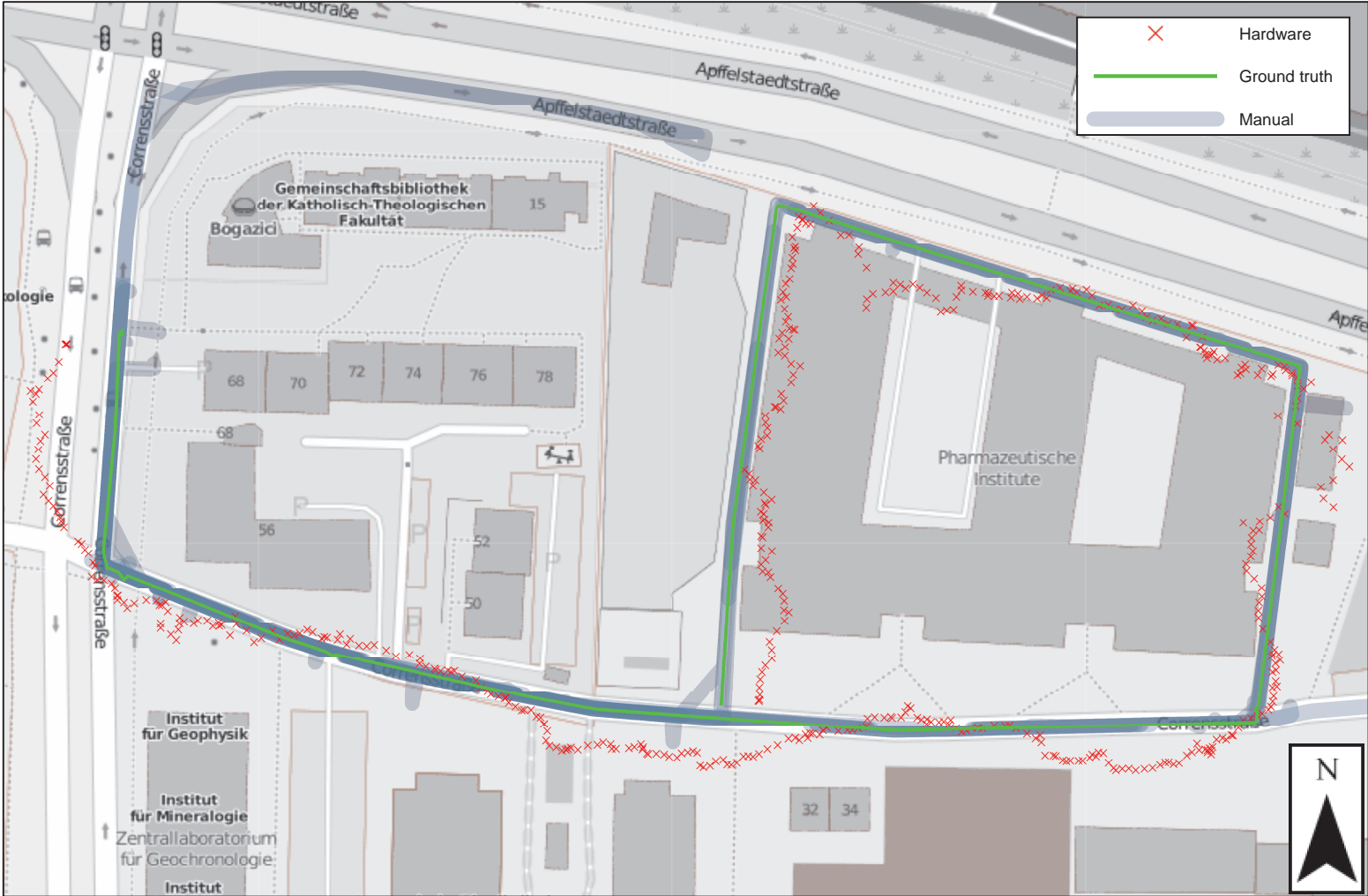
B.4 Route 6 (simple)

This route is situated at a very large crossroad which contains many secondary residential footways which have made the geotagging process difficult for some of the participants. This could explain why the Route 6 is one which has a higher position error but less orientation error.

Route 2 (Complex)



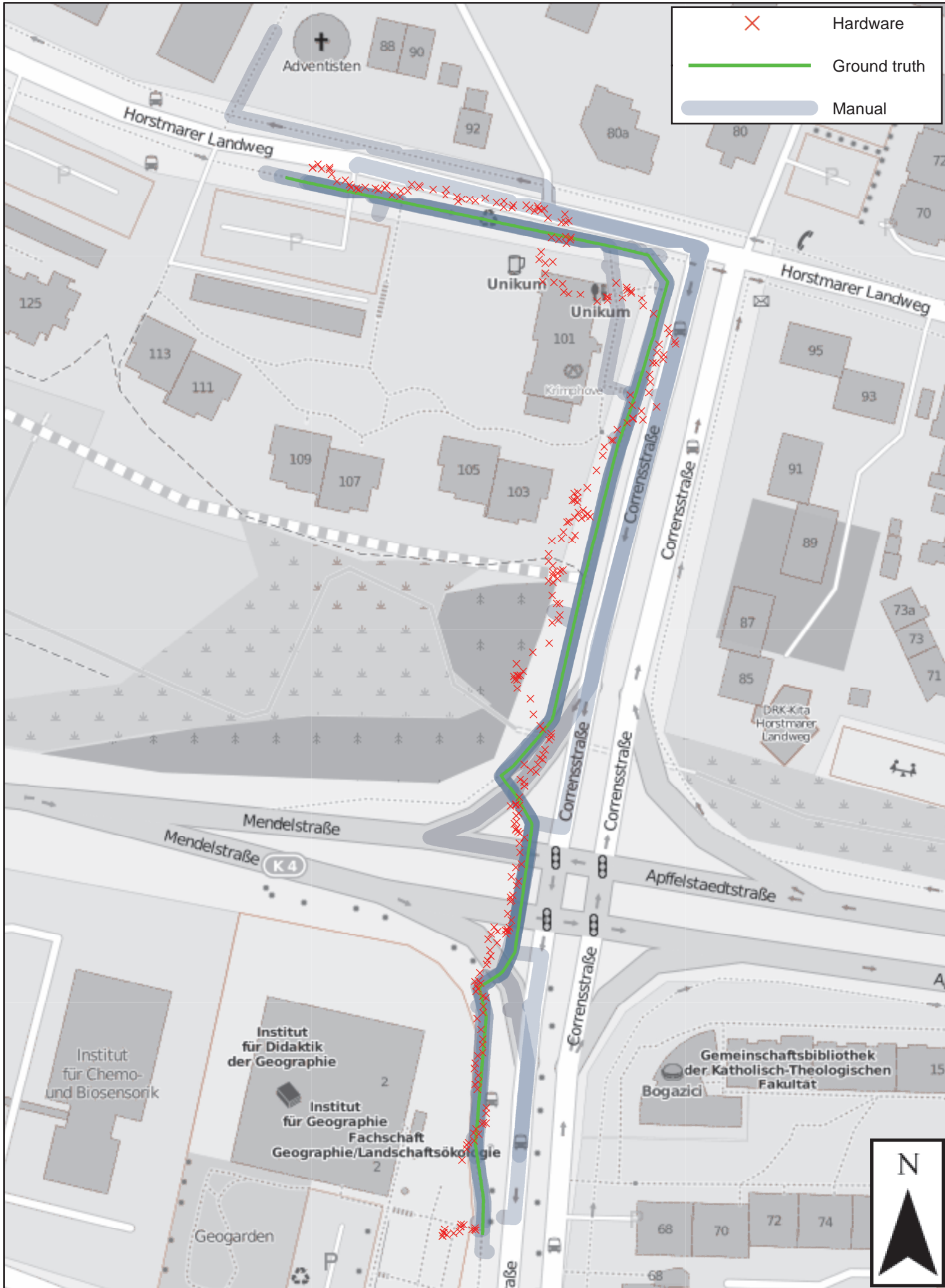
Route 4 (Complex)



Route 5 (Simple)



Route 6 (Simple)



Bibliography

- [1] J. W. Mills, A. Curtis, B. Kennedy, S. W. Kennedy, and J. D. Edwards, “Geospatial video for field data collection,” *Applied Geography*, vol. 30, no. 4, pp. 533–547, 2010.
- [2] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair, “Over-exposed?: privacy patterns and considerations in online and mobile photo sharing,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2007, pp. 357–366.
- [3] B. Seo, J. Hao, and G. Wang, “Sensor-rich video exploration on a map interface,” in *Proceedings of the 19th ACM international conference on Multimedia - MM 11*. Association for Computing Machinery (ACM), 2011.
- [4] C. Hauff, “A study on the accuracy of flickrs geotag data,” in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval - SIGIR 13*. Association for Computing Machinery (ACM), 2013.
- [5] S. A. Ay, R. Zimmermann, and S. H. Kim, “Viewable scene modeling for geospatial video search,” in *Proceeding of the 16th ACM international conference on Multimedia - MM 08*. Association for Computing Machinery (ACM), 2008.
- [6] S. A. Ay, L. Zhang, S. H. Kim, M. He, and R. Zimmermann, “GRVS,” in *Proceedings of the seventeen ACM international conference on Multimedia - MM 09*. Association for Computing Machinery (ACM), 2009.
- [7] A. Curtis and J. W. Mills, “Spatial video data collection in a post-disaster landscape: The tuscaloosa tornado of april 27th 2011,” *Applied Geography*, vol. 32, no. 2, pp. 393–400, 2012.
- [8] L. Montoya, “Geo-data acquisition through mobile GIS and digital video: an urban disaster management perspective73,” *Environmental Modelling & Software*, vol. 18, no. 10, pp. 869–876, 2003.

- [9] Y.-T. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T.-S. Chua, and H. Neven, "Tour the world: building a web-scale landmark recognition engine," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1085–1092.
- [10] P. Chippendale, M. Zanin, and C. Andreatta, "Collective photography," in *2009 Conference for Visual Media Production*. Institute of Electrical & Electronics Engineers (IEEE), 2009.
- [11] M. Y. Chen, T. Sohn, D. Chmelev, D. Haehnel, J. Hightower, J. Hughes, A. LaMarca, F. Potter, I. Smith, and A. Varshavsky, "Practical metropolitan-scale positioning for GSM phones," in *Lecture Notes in Computer Science*. Springer Science Business Media, 2006, pp. 225–242.
- [12] G. Wang, Y. Yin, B. Seo, R. Zimmermann, and Z. Shen, "Orientation data correction with georeferenced mobile videos," in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL13*. Association for Computing Machinery (ACM), 2013.
- [13] Y.-T. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T.-S. Chua, and H. Neven, "Tour the world: Building a web-scale landmark recognition engine," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Institute of Electrical & Electronics Engineers (IEEE), jun 2009.
- [14] J. Hays, A. Efros *et al.*, "Im2gps: estimating geographic information from a single image," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.
- [15] S. A. Ay, L. Zhang, S. H. Kim, M. He, and R. Zimmermann, "GRVS," in *Proceedings of the seventeen ACM international conference on Multimedia - MM 09*. Association for Computing Machinery (ACM), 2009.
- [16] S. H. Kim, S. A. Ay, B. Yu, and R. Zimmermann, "Vector model in support of versatile georeferenced video search," in *Proceedings of the first annual ACM SIGMM conference on Multimedia systems - MMSys 10*. Association for Computing Machinery (ACM), 2010.
- [17] Y. Yin, B. Seo, and R. Zimmermann, "Content vs. context," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 3, pp. 1–21, feb 2015.
- [18] Z. Shen, S. A. Ay, S. H. Kim, and R. Zimmermann, "Automatic tag generation and ranking for sensor-rich outdoor videos," in *Proceedings of the 19th ACM international conference on Multimedia - MM 11*. Association for Computing Machinery (ACM), 2011.

- [19] S. A. Ay, S. H. Kim, and R. Zimmermann, "Generating synthetic meta-data for georeferenced video management," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS 10*. Association for Computing Machinery (ACM), 2010.
- [20] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *International Journal of Human-Computer Interaction*, vol. 24, no. 6, pp. 574–594, jul 2008. [Online]. Available: <http://dx.doi.org/10.1080/10447310802205776>
- [21] T. S. Tullis and J. N. Stetson, "A comparison of questionnaires for assessing website usability," in *Usability Professional Association Conference*, 2004, pp. 1–12.
- [22] J. Brooke, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.
- [23] K. Norman, B. Shneiderman, B. Harper, and L. Slaughter, "Questionnaire for user-interface satisfaction," 1998.
- [24] J. R. Lewis, "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use," *International Journal of Human-Computer Interaction*, vol. 7, no. 1, pp. 57–78, 1995.
- [25] J. Sauro, "Measuring usability with the system usability scale (sus)," 2011.
- [26] J. R. Lewis and J. Sauro, "The factor structure of the system usability scale," in *Human Centered Design*. Springer, 2009, pp. 94–103.
- [27] W. ISO, "9241-11. ergonomic requirements for office work with visual display terminals (vdts)," *The international organization for standardization*, 1998.
- [28] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009.
- [29] S. Elwood, "Geographic information science: Visualization, visual methods, and the geoweb," *Progress in Human Geography*, vol. 35, no. 3, pp. 401–408, jul 2010.
- [30] M. Haklay, A. Singleton, and C. Parker, "Web mapping 2.0: The neogeography of the GeoWeb," *Geography Compass*, vol. 2, no. 6, pp. 2011–2039, nov 2008.
- [31] A. Çöltekin, B. Heil, S. Garlandini, and S. I. Fabrikant, "Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis," *Cartography and Geographic Information Science*, vol. 36, no. 1, pp. 5–17, 2009.

- [32] S. Roche, E. Propeck-Zimmermann, and B. Mericskay, "GeoWeb and crisis management: issues and perspectives of volunteered geographic information," *GeoJournal*, vol. 78, no. 1, pp. 21–40, jun 2011.
- [33] M. Zook, M. Graham, T. Shelton, and S. Gorman, "Volunteered geographic information and crowdsourcing disaster relief: A case study of the haitian earthquake," *SSRN Electronic Journal*, 2010.
- [34] M. van Persie, M. C. van Sijl, E. Wisse, J. B. Tjoe-Awie, A. J. de Jong, and W. Bakker, "Integration of real-time UAV video into the fire brigades crisis management system," in *Intelligent Systems for Crisis Management*. Springer Science Business Media, 2012, pp. 327–339.
- [35] V. Nallur, A. Elgammal, and S. Clarke, "Smart route planning using open data and participatory sensing," in *Open Source Systems: Adoption and Impact*. Springer Science Business Media, 2015, pp. 91–100.
- [36] J. Brooke, "Sus: a retrospective," *Journal of Usability Studies*, vol. 8, no. 2, pp. 29–40, 2013.
- [37] Apache, *version 2.4.7*. Apache Software Foundation. [Online]. Available: www.apache.org
- [38] L. Mint, *version 17.2*. Linux Mark Institute. [Online]. Available: www.linuxmint.com
- [39] Bootstrap, *version 3.3.6*. M. Otto, J. Thornton, and Bootstrap contributors. [Online]. Available: <http://getbootstrap.com>
- [40] Leaflet, *version 0.7.7*. V. Agafonkin. [Online]. Available: <http://leafletjs.com/>
- [41] jQuery, *version 1.11.3*. The jQuery Foundation. [Online]. Available: <https://jquery.com/>
- [42] Victor.js, *version 1.1.0*. M. Kueng, G. Crabtree. [Online]. Available: <http://victorjs.org>
- [43] Intro.js, *version 1.1.1*. A. Mehrabani. [Online]. Available: <http://usablica.github.io/intro.js/>
- [44] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org>
- [45] ArcGIS, *version 10.2.2*. Environmental Systems Research Institute (ESRI). [Online]. Available: <https://www.arcgis.com>

-
- [46] J. Lewis, “GeoVideo Web Service.” Open Geospatial Consortium Inc., 2006.
- [47] G. Wallet, H. Sauzéon, J. Rodrigues, and B. N’Kaoua, “Transfer of spatial knowledge from a virtual environment to reality: Impact of route complexity and subject’s strategy on the exploration mode,” *Journal of Virtual Reality and Broadcasting*, vol. 6, no. 4, pp. 572–574, 2009.
- [48] A. Maya. Autodesk, Inc. [Online]. Available: <http://www.autodesk.com/products/maya>
- [49] Blender Foundation, *version 3.3.6*. [Online]. Available: <https://www.blender.org/>
- [50] W. H. Warren and D. J. Hannon, “Direction of self-motion is perceived from optical flow,” *Nature*, vol. 336, no. 6195, pp. 162–163, 1988.
- [51] Rotation of map and contents to x degrees. · issue #268 · leaflet/leaflet. [Online]. Available: <https://github.com/Leaflet/Leaflet/issues/268>
- [52] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, “From game design elements to gamefulness: defining gamification,” in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. ACM, 2011, pp. 9–15.