Synthesis Project (GEO1101)

Georeferencing Historic Map Series: An Automated Approach

Final Report

Delft, 16 November 2023

Group members

Eirini C. Tsipa | 5860881

Georgios Iliopoulos | 5842247

Oliver J. Post | 4677706

Rianne Aalders | 4593987

Supervisory Team

Ir. Edward Verbree

Ir. Jules Schoonman

Dr. Ir. Martijn Meijers

“It's impossible to map out a route to your destination if you don't know where you're starting from.”

-Suze Orman

Abstract

The present report is the end result of the project that was carried out as part of the Geomatics Synthesis Project in cooperation with AllMaps, an open-source platform dedicated to the viewing and georeferencing of historic maps. The main objective of the project was to automatically georeference historic map series curated and digitised by the Dutch National Archive. This was based on the corner coordinates of the map sheets. The first issue that had to be tackled was the reprojection of the original coordinates which were in Bonne projection to WGS84 coordinates. To determine the corners of the map content within the sheets two methods were implemented. The first one detects the lines based on HoughLines Probabilistic Transformation and the second one detects lines based on the distribution of black pixels in the rows and columns of the images. In addition to map sheets with corner coordinates, there are two other sets of images which were georeferenced utilising a convolution neural network that performs feature matching. The feature matching was performed by running the two sets of images against the georeferenced sheets with known corner coordinates. To minimise the search space for this process a geocoder was used to determine the approximate location of the image. The implemented methods appear to hold the potential for georeferencing old map series. It is worth noting that the developed algorithms, while effective in many cases, may encounter challenges when dealing with irregularities on map sheets caused by the passage of time, such as damage. Consequently, there is a great opportunity to further enhance the algorithms to ensure they can consistently and accurately georeference images, even when faced with such irregularities. This ongoing development will lead to improved georeferencing accuracy and user confidence.

Keywords: Automation, Georeferencing, Map Series, Cartography, Historic maps, Allmaps, Open data, AI, Feature matching, Corner detection, Bonnebladen, TMK, Bonne projection, TU Delft, Synthesis Project

Code: https://github.com/geor-tudelft/iiifmap

Table of Contents

Table of Contents

List of Figures

List of Tables

List of Acronyms

CRS - Coordinate reference system

GCP - Ground Control Point

IIIF - International Image Interoperability Framework

OCR - Optical Character Recognition

OSM - OpenStreetMap

NA - National Archives

TMK - Topografische en Militaire Kaart van het Koninkrijk der Nederlanden

WHG - World Historical Gazetteer

API - Application Programming Interface

WGS84 - World Geodetic System 1984

Preface

Synthesis - Etymology

Synthesis comes from the Latin synthesis, from Ancient Greek σύνθεσις (súnthesis, “a putting together; composition”), from συντίθημι (suntíthēmi, “put together, combine”), from συν- (sun-, “together”) + τίθημι (títhēmi, “set, place”)(*Synthesis - Wiktionary, the Free Dictionary*, n.d.).

The Synthesis Project is a fundamental component of our curriculum, offering students a unique opportunity to bridge theory and practice in the field of Geomatics. In this project, students collaborate with external partners, such as companies, governmental agencies, or TU Delft research groups, to address real-world geo-information challenges. The project aims to provide a holistic understanding of the entire geo-information process, from data collection to interpretation, visualisation, and application. It also emphasises project management skills and encourages effective communication and cooperation among diverse stakeholders. It is an experiential learning opportunity that bridges the gap between academic knowledge and practical application within the field of Geomatics. By working on real-world geo-information problems, students gain a comprehensive understanding of the entire process, apply their skills, develop project management expertise, and enhance their collaborative and communication abilities. This holistic approach ensures that graduates are well-prepared to tackle the complex challenges of the geo-information industry (*Course Browser Searcher*, n.d.).

The array of courses we've undertaken as part of our Geomatics program has been pivotal in not only broadening our knowledge base but also in honing the practical skills essential for the successful execution of the Synthesis Project. These courses, which span a diverse range of geospatial disciplines, have collectively formed the educational framework upon which we have built our project. They have provided us with a multifaceted skill set and a comprehensive understanding of geospatial technologies, positioning us at the forefront of addressing real-world geo-information challenges. Practically, the majority of the courses that we have taken during the first year of our studies contributed to the successful outcome of our project. Namely, GEO1000: Python Programming, GEO1002: GIS and Cartography, GEO1003: Positioning and Location Awareness, GEO1006: Geo Database Management Systems, GEO1007: Geoweb Technology, GEO1009: Geo-information Governance, GEO1015: Digital Terrain Modeling and GEO5017: Machine Learning for the Built Environment.

1. Introduction

In collaboration with Allmaps, an open-source platform dedicated to the viewing and georeferencing of historic maps, our team aims to accelerate the georeferencing process for historic map series curated by the Dutch National Archives. This project seeks to harness the power of automation to address a fundamental challenge in the realm of historical cartography - the labour-intensive and time-consuming nature of georeferencing.

Historical maps are invaluable treasures, encapsulating the geographical and cultural knowledge of bygone eras. They provide essential context for scholars, archaeologists, and urban planners, enabling a deeper understanding of how our landscapes, boundaries, and societies have evolved over time. Yet, georeferencing these maps, especially when dealing with extensive map series comprising thousands of individual sheets, is a task that requires expertise.

Our primary aim is to enhance the usability and accessibility of these historical maps by incorporating geospatial information. Georeferencing not only enriches the maps but also makes them more accessible to researchers, educators, and anyone with an interest in history. Moreover, by accelerating this process through automation, we hope to inspire other institutions to share their collections as open data, further enriching the global repository of historical knowledge.

Our approach draws upon our knowledge in cartography, geodesy, digital processing, and insights obtained from a comprehensive literature review. We are committed to developing a method that can expedite the georeferencing of the Dutch National Archives' map series, and potentially serve as a template for similar projects in the future.

Challenges abound in this endeavour, given that we are working with raster images of old maps. These challenges encompass transforming coordinates from the archaic reference systems used in the maps' creation, accurately stitching together map sheets that may include loose islands and overlapping areas, and mitigating errors introduced by manual steps in the process. Furthermore, the inherent complexities of automating processes for non-machine-readable data and dealing with damaged or distorted map sheets is expected to be encountered.

Historical maps, as timeless artefacts, offer us a portal to the past, and through this collaborative effort, we hope to unlock their potential in a modern, digital age. Our mission is twofold: to make the Dutch National Archive's georeferenced map series available on the AllMaps platform, and to create a scalable and adaptable method that can potentially benefit other institutions and projects seeking to bring historical maps to life through georeferencing and automation.

2. Problem definition

In this section we introduce the problem we try to solve as well as our motivation. The limitations that characterise the project and the needs to successfully complete it are also listed.

2.1 Project summary

In collaboration with AllMaps, an open-source platform for the viewing and georeferencing of historic maps, our team aims to accelerate the process of georeferencing historic map series from the Dutch National Archive, using automation. Georeferencing increases the usability and value of digitised maps by adding geospatial information. This process is labour-intensive and time-consuming when done manually, as map series can consist of hundreds to thousands of individual sheets. Accelerating this process increases the availability and usability of historical maps, for free and without restrictions, as a tool for research, education, and anyone with a general interest. Increasing the relevance of the digital maps can also incentivise archives to publish more of their collections online as open data. Our approach includes the use of our Cartography background, Geodesy knowledge, digital processing, and literature review. The final goals of this project and our collaboration are to make the supplied map series available on the Allmaps online platform and to develop a method adaptable to other institutions or projects.

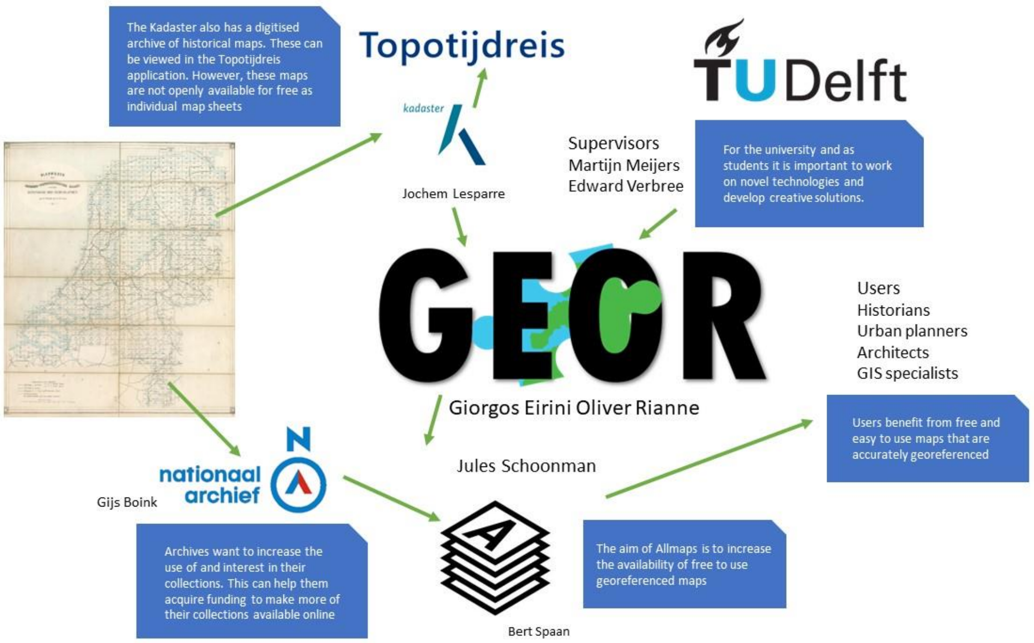


Figure 1The rich picture with an overview of actors and their interests.

2.2 Research question

Currently the process of georeferencing old map series requires a lot of manual work and it is very time consuming. To get a better understanding of how this process works, every scanned image of the archive is georeferenced individually by manually choosing points on the historic map and an already georeferenced web-map on Allmaps platform. But these archives consist of thousands of images since each map sheet is usually divided into smaller sheets. Our partner aims to georeference the whole collection of maps curated by the Dutch National

Archive. Therefore, accelerating this process is considered to be essential. By doing so, making those georeferenced maps available to the public with incorporated geographic information, is then facilitated. Our team will try to tackle the challenge of accelerating the process of georeferencing with the use of automation. Or more specifically, we will develop a method to accelerate the process of georeferencing of the Dutch National Archives' map series, using automation.

3. Theory and context

3.1 IIIF and the NA

IIIF is "a set of open standards for delivering high-quality, attributed digital objects online at scale" *Home*. (2023, July 10). IIIF. https://iiif.io/. It consists of several API's concerning the retrieval, display, access and metadata of images on the web. It is separated into several parts of which the Image API and Presentation API are most relevant for this research.

The Image API is about the retrieval of an image from its server and the Presentation API adds metadata, references multiple images together or in the right order. The Georeference extension is also part of the Presentation API. It describes a way to create a Georeference Annotation, where the metadata needed for georeferencing is stored. This annotation is in json format with a body of GeoJSON point coordinates. Therefore, the coordinates must be expressed in the WGS84 CRS (*Georeference extension*, n.d.).

The National Archives only utilise the image API. Instead of the presentation API they have an xml inventory detailing the metadata of the images. J. Schoonman has parsed this xml and put the relevant information in json format.

3.2 Allmaps' current workings

Allmaps consists of two major components. The Allmaps Editor and the Allmaps Viewer. In the Editor a user loads an image by requesting it from its IIIF server through a URL, manifest or collection. Then, the user adds a mask by clicking along the edge of the map area, the pink line in figure 2.

A manifest is the prime unit in IIIF which lists all the information that makes up a IIIF object. It communicates how to display your digital objects, and what information to display about them, including structure, to varying degrees of complexity as determined by the implementer. (For example, if the object is a book of illustrations, where each illustrated page is a canvas, and there is one specific order to the arrangement of those pages).

A collection is a set of manifests (or 'child' collections) that communicate hierarchy or gather related things (for example, a set of boxes that each have folders within them, and photographs within those folders)(*How It Works*, n.d.).

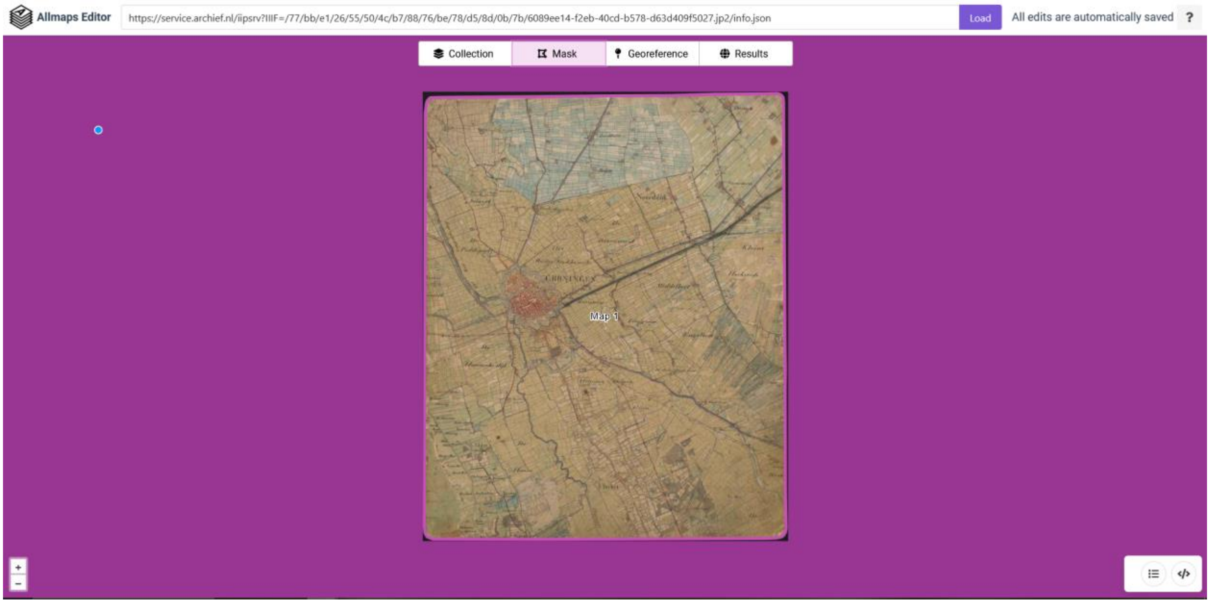


Figure 2 Creating a mask in the Allmaps Editor.

The next step is to georeference. The user clicks at a point on the map image (figure 3, left) and its corresponding point on OSM (figure 3, right), this pair is called a GCP (*Georeference extension*, n.d.). There need to be at least 3 point pairs to georeference the image. For the best result, these points should not lie on the same line and should be evenly distributed over the map area. Corners of man-made features that are as old as the map make the most accurate GCPs.

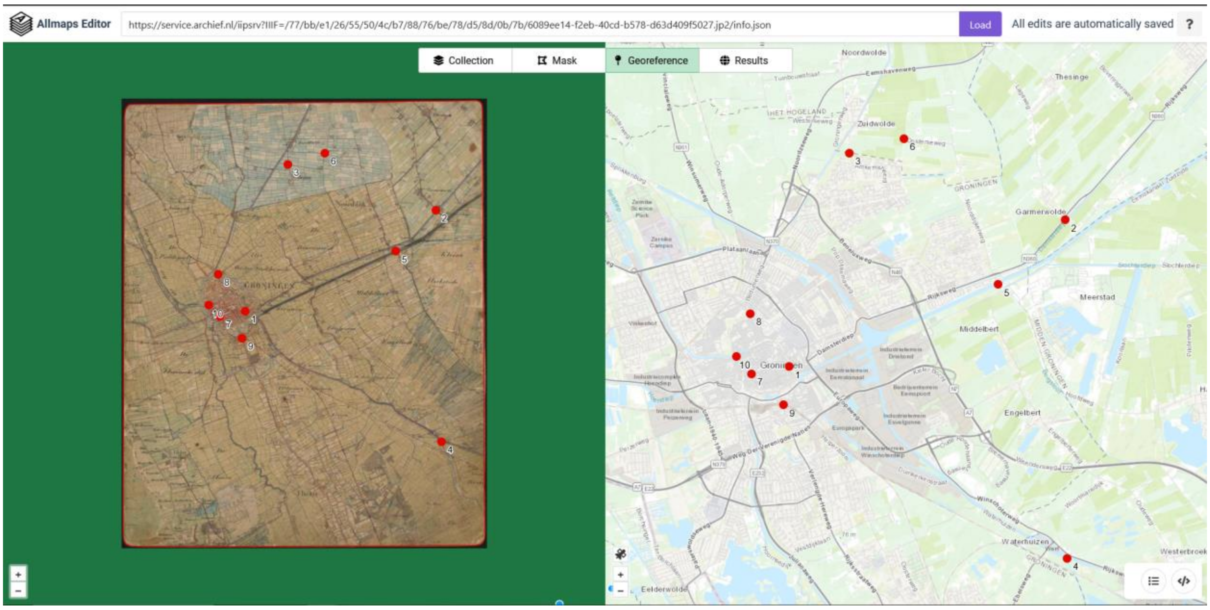


Figure 3 Picking point pairs in the Allmaps Editor.

The Allmaps Editor combines the mask and points into a georeference annotation page that can be processed and read by the Allmaps Viewer. To see the result of the georeferencing, the user is directed to the Allmaps Viewer (figure 4).

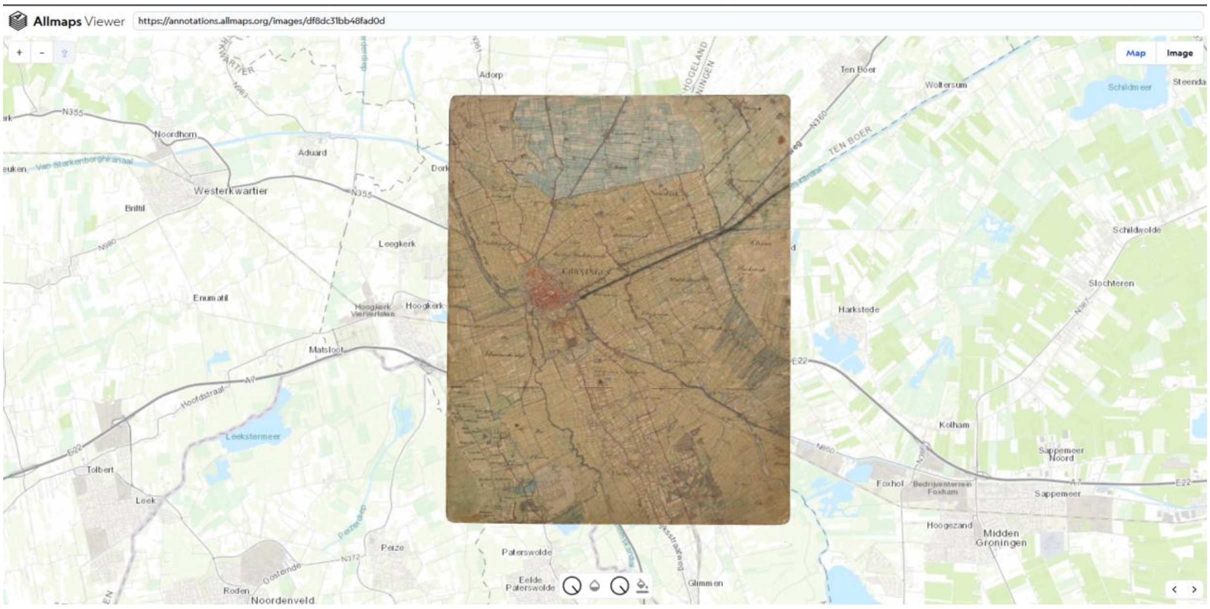


Figure 4 Displaying the result of georeferencing tin the Allmaps Viewer.

The Allmaps Viewer processes the provided annotation page by transforming the image to align the image to the base map. For this it uses either a first order polynomial or thin plate spline transformation. This is specified in the annotation page but can be switched with a keyboard shortcut.

3.3 Map history, projection, sheet index, characteristics

The Dutch National Archives houses a large collection of historical maps depicting places all over the globe. A selection of these are photo scanned without distortion at 300 dpi and published on their website (https://www.nationaalarchief.nl/) in compliance with IIIF standards.

In our research we will focus on three map series, the ‘veldminuten’ (fieldminutes) and the ‘nettekeningen’ (neat drawings) of the “Topographische en Militaire Kaart van het Koninkrijk der Nederlanden" (TMK) and the ‘minuutbladen’ (minute sheets) of the “Chromo-Topographische Kaart des Rijks in Bonneprojektie” (Bonnebladen). These map series originate from the 19th century and are part of the effort to create a cohesive map of the entire country of the Netherlands. The process of creating the map started with measuring angles in a network of triangles spanning the country, connecting to both Germany and France. These angle measurements were then used to calculate distances, this is the triangulation of Krayenhoff. With a projection and distances established, the map could be filled in with topographical features. Thas was done using a combination of copying previous local maps, especially from the cadastre, and in situ drawings. These copies and field drawings together are called the ‘veldminuten’, the first version recognisable as a map. In the next step, the collection of veldminuten is copied onto a grid of rectangular sheets which are the ‘nettekeningen’. Finally, the neat drawings are transferred to a printing medium to create the final copies to be distributed for use.

The archive has taken the sheet index into account in their naming of the files and structure of the archive. The last number in the file name is the sheet number.

3.3.1 TMK

The TKM was created between 1834 and 1944, both the veldminuten and nettekeningen are digitised by the NA. The TMK uses the ‘gewijzigde projectie van Flamsteed’ also called the ‘Fransche projectie’ This is an adapted version of the Sinusoidal, Sanson–Flamsteed or the Mercator equal-area projection where the main meridian passes over the Westertoren of Amsterdam instead of the prime meridian and the average parallel is at 51°30' latitude instead of the equator (Topographisch Bureau, 1861).

514

Nummers der bladen.	Hoekpunten.	Noorderbreedten.	Ooster- of Westerlengten.	Afstanden in ellen tot: de perpendiculair.	den hoofdmeridiaan.
4	*a*	53° 23′ 13″.75	0° 18′ 2″.34	210 000 N.	20 000 W.
	b	53° 23′ 13″.75	0° 18′ 2″.34	210 000 N.	20 000 O.
	c	53° 9′ 44″.91	0° 17′ 56″.76	185 000 N.	20 000 O.
	d	53° 9′ 44″.91	0° 17′ 56″.76	185 000 N.	20 000 W.
5	*a*	53° 23′ 13″.75	0° 18′ 2″.34	210 000 N.	20 000 O.
	b	53° 23′ 3″.11	0° 54′ 7″.08	210 000 N.	60 000 O.
	c	53° 9′ 34″.36	0° 53′ 50″.11	185 000 N.	60 000 O.
	d	53° 9′ 44″.91	0° 17′ 56″.76	185 000 N.	20 000 O.
6	*a*	53° 23′ 3″.11	0° 54′ 7″.08	210 000 N.	60 000 O.
	b	53° 22′ 41″.84	1° 30′ 11″.39	210 000 N.	100 000 O.
	c	53° 9′ 13″.21	1° 29′ 43″.07	185 000 N.	100 000 O.
	d	53° 9′ 34″.36	0° 53′ 50″.11	185 000 N.	60 000 O.
7	*a*	53° 22′ 41″.84	1° 30′ 11″.39	210 000 N.	100 000 O.
	b	53° 22′ 10″.14	2° 6′ 14″.86	210 000 N.	140 000 O.
	c	53° 8′ 41″.46	2° 5′ 35″.32	185 000 N.	140 000 O.
	d	53° 9′ 13″.21	1° 29′ 43″.07	185 000 N.	100 000 O.
8 (Half blad)	*a*	53° 22′ 10″.14	2° 6′ 14″.86	210 000 N.	140 000 O.
	b	53° 21′ 50″.24	2° 24′ 16″.20	210 000 N.	160 000 O.
	c	53° 8′ 21″.60	2° 23′ 30″.85	185 000 N.	160 000 O.
	d	53° 8′ 41″.46	2° 5′ 35″.32	185 000 N.	140 000 O.
9	*a*	53° 9′ 44″.91	0° 17′ 56″.76	185 000 N.	20 000 W.
	b	53° 9′ 44″.91	0° 17′ 56″.76	185 000 N.	20 000 O.
	c	52° 56′ 16″.07	0° 17′ 51″.23	160 000 N.	20 000 O.
	d	52° 56′ 16″.07	0° 17′ 51″.23	160 000 N.	20 000 W.

Figure 5 Example of a page with a corner coordinates table from The Meetkunstige beschrijving van Nederand (1861).

The Meetkunstige beschrijving van Nederand (1861) also has tables for the coordinates of all towns and cities measured and for the corners of the sheets (figure 5).

Figure 6 Examples of scans of the TMK veldminuten. Left: Non-rectangular shape of some scans. Right: shows an island without reference of where to place it along the mainland.

Figure 7 Example of a scan of the TMK veldminuten with a damaged bottom left corner.

Figure 8 Example of a scanned sheet of the TMK nettekeningen, showing the white border around the map sheet.

Figure 9 Example of a scanned sheet of the TMK nettekeningen, showing an irregular sheet.

3.3.2 Bonnebladen

The Bonnebladen are the successor to the Strookkaarten der Verdedigingsliniën des Rijks, and were made between 1865 and 1951. There are multiple versions of the same sheets as new maps were drawn at different points in time, figure 10. The earliest sheets are based on the veldminuten of the TMK and later sheets are based on newer reductions from cadastral maps. Linden (1973) calls the ‘gewijzigde projectie van Flamsteed’ the “Bonneprojectie” which places the Bonnebladen in the same CRS as the TMK.

Figure 10 Examples of the Amsterdam sheets from 1876 (top) and 1892-1902 (bottom)respectively. These sheets also have worn corners.

mapseries	TMK		Bonnebladen
name	Veldminuten	Nettekeningen	Minuutbladen
CRS	Bonne projection	Bonne projection	Bonne projection
scale	1:25000	1:50000	1:25000
number of scans	414	290	2991
number of sheets in index	62	62	776
characteristics	The ‘kartonnen’ (sub-sheets) are scanned individually. The pattern for combining the full sheet is highly irregular in shape. The sub-sheets have worn rounded corners and edges and have a black background. Sub-sheets have irregular shapes and sizes	Sub-sheets are scanned separately and scanned combined into full sheets. Each subsheet is a quarter of the size of the full sheet A full sheet has a white paper border and a black background	The sheets containing only water were not made. The sheets have worn rounded corners and edges and have a black background

Table 1 Overview of the map series

4. Methods

4.1 Research approach schema

Method: Map matcher

This method is specifically geared to the “veldminuten”, which are hard to position in an automated way. We first find the exact positions of the “nettekeningen” and then we place the “veldminuten” based on automatically matched features on the “nettekeningen”. Finally we apply the georeferencing to the individual “veldminuten”.

Strengths	Weaknesses
- If matching works well, this method can achieve **high accuracy** as the “veldminuten” will be placed on the “nettekeningen”, which would be georeferenced based on the original source coordinates.	- Matching **might not work for all sheets** or produce incorrect results. - Still would require **manual work**.

We developed a flowchart to visualise the whole pipeline of this method (Figure 2). During the creation of this flowchart, we determined that this method combines the strengths of method 2 with innovative automated processing.

We also saw opportunities for it to be widely usable. In our proposal, the fieldminutes are automatically placed on a highly similar reference map (the neat drawings of the same map). If the system can be made to rely on map features and not on a particular map, its usability could extend to many map series. As long as you have a georeferenced map that shares some similarities with your input map, it could perform the georeferencing in a highly automated way.

The main weakness of this method is the complexity of the matching. Depending on how successful we are in implementing this determines how usable the solution is. We are however confident that we can get it to work on the TMK field minutes as we have a highly similar reference map to perform the matching on. This means that the requirements of our client will be fulfilled almost certainly. The neat drawings of the TMK and the Bonnebladen are georeferenced using source coordinates, leading to very high accuracy. The fieldminutes of the TMK are also georeferenced with high accuracy when the matching algorithm performs adequately.

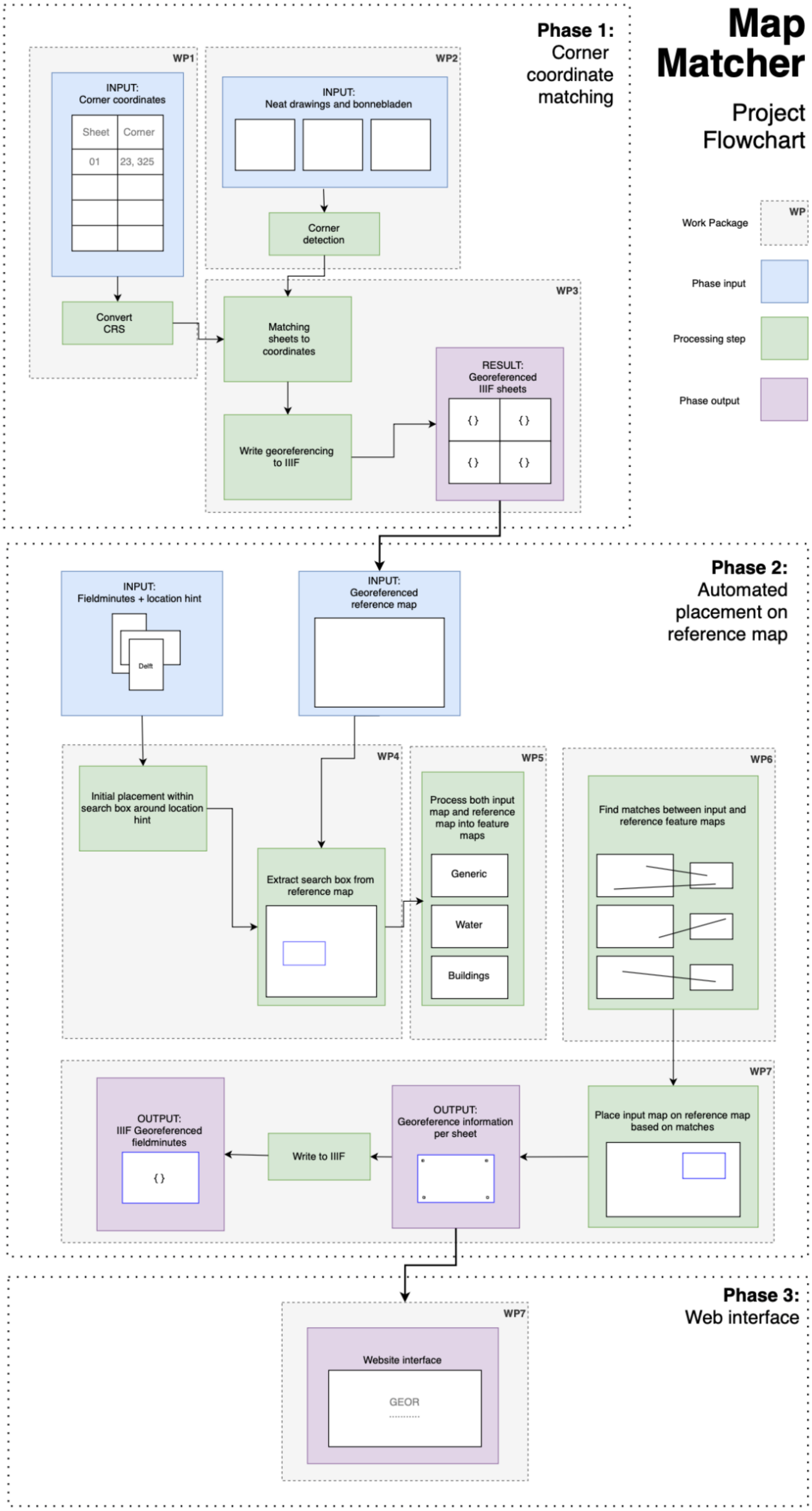


Figure 11 Sketch flowchart of the map matcher method. Starting from top left, following the arrows it goes through corner coordinate matching phase, an automated placement on reference map phase, and a web interface phase.

4.2 Scoping/ Managing expectations: MoSCoW

Must	• IIIF annotation of each mapsheet • Georeference mapsheets • Describe our project, process and results in a final report • Give a final presentation of the project on Geomatics day
Should	• Provide quality indices that can indicate the georeferencing quality • Make method also work for similar map series
Could	• Make a method work for many map series - Interoperability • Write a scientific paper about our project's findings and results • Prepare a presentation about our project for an event outside of the synthesis project course.
Won't	• Fully automate the whole process of georeferencing old map series • Georeferencing each map sheet automatically on an individual basis

Table 2 MoSCoW division of tasks

4.3 Research limitations

- Due to the nature of the Synthesis project course, we have limited time and a strict deadline to complete our research.
- Unavoidable inaccuracies in historical maps.
- Limited open source access to existing georeferenced historical map series.
- The accuracy of georeferencing depends on the quality of the historic maps and the reference data. Poor-quality or damaged maps may pose challenges.

4.4 List of needs

- High-quality scanned map sheets, preferably already in IIIF format.
- Sources with corner coordinates of map sheets.
- Existing software libraries or services for necessary steps like feature matching.
- Well-structured and annotated archive, with identifiers for each map sheet.
- Accurate transformation from the bonne projection to WGS84
- The manifest IIIF presentation API with information from the NA xml metadata

5. Workflow

5.1 Phase 1

5.1.1 WP1: Finding and processing the corner coordinates of both the TMK and the Bonnebladen

The goal of this work package was to deliver the corner coordinates of full TMK and Bonnebladen map sheets in the WGS84 CRS and in a GeoJSON format. These CRS and format were chosen because they are part of the IIIF standard that both the archive and Allmaps use.

The workflow for this package was to 1) find the corner coordinates of the map sheets in their original CRS and convert these to a machine readable format, 3) find and execute the transformation from the original CRS to WGS84, 4) output the WGS84 coordinates of the sheets to GeoJSON.

The sheet corner coordinates of the TMK are written down in the borders at the corners on the sheets themselves and in a table in the book '*Meetkunstige beschrijving van het Koningrijk der Nederlanden, bevattende de getallenwaarden, gebruikt bij de zamenstelling van de topografische en militaire kaart van het rijk'* (1861). The bonne projection coordinates in the table were manually copied from the pdf of the book to a csv file which could then be parsed.

Figure 12 The sheet index of the TMK.

After spending several weeks looking for a similar source for the Bonnebladen but not finding it, a different approach was used. Because the sheet index of the Bonnebladen follows a very regular grid structure, the corner coordinates for each sheet can be calculated when their position in the grid is known. For this, parts of the code of Schoonman (2023) and Meijers (2023) were used.

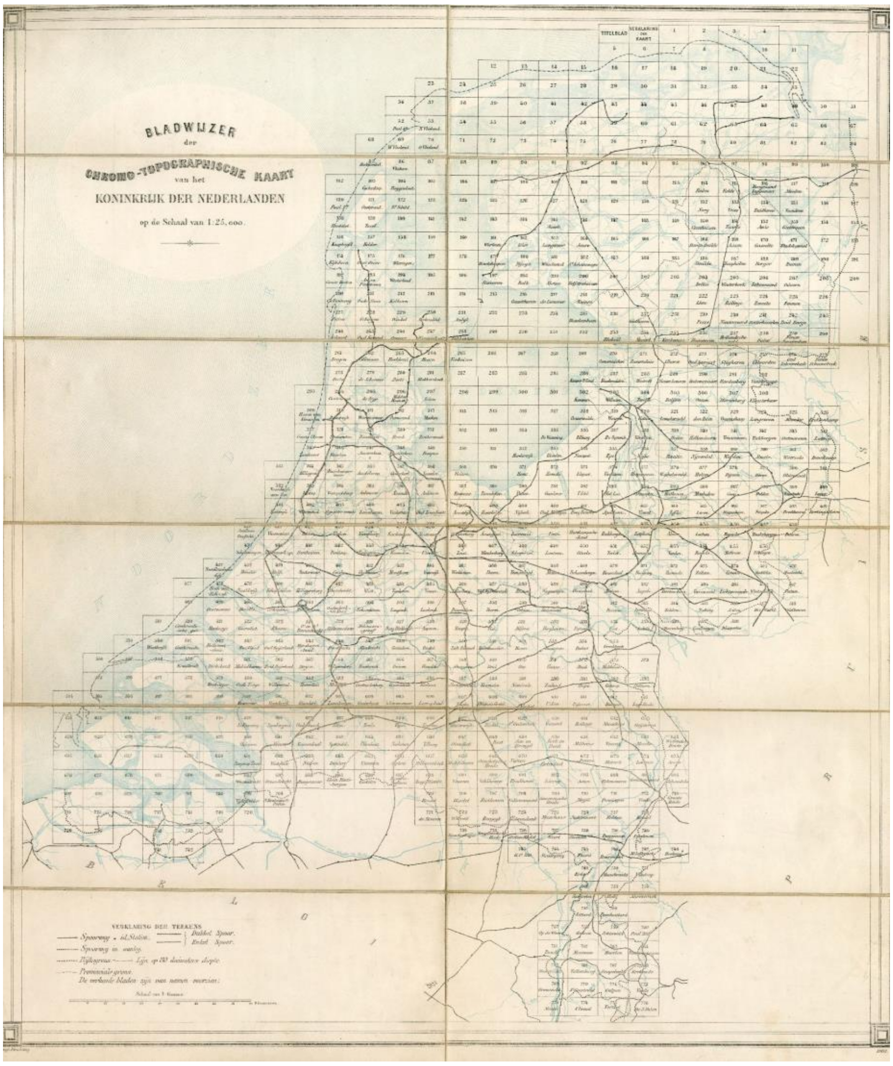
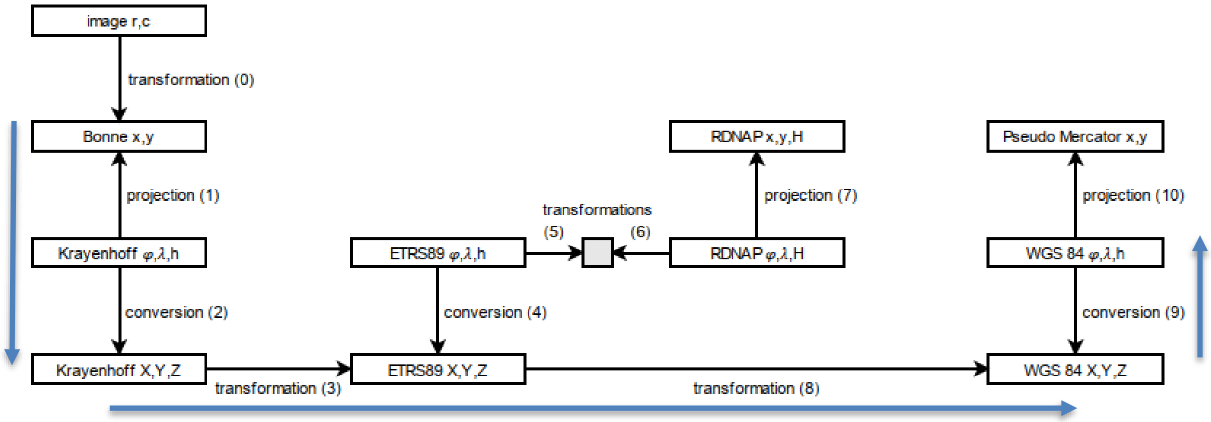


Figure 13 The sheet index of the Bonnebladen.

The first step in this process is to create a bounding grid measuring the maximum number of rows and columns in the sheet index. Next, the grid cells are assigned their sheet number. This was done by giving each row a description of how many 'empty' cells there are left of the map and how many map sheets are on the row. In the south of the Netherlands there are several rows where there are empty cells in between filled cells. These were included by adding them to the number of filled cells in the row but giving them the sheet number '0'. The other filled cells were numbered top to bottom left to right according to the sheet index. Finally, the corner coordinates were calculated based on the distance from the origin (the top left corner of sheet number 666 near Chaam), each cell is 40000 metres wide and 25000 metres high.

The next step for this work package was to transform these bonne coordinates to WGS84. For expertise on Dutch CRSs Jochem Lesparre from the Kadaster was contacted. He made and shared a nationwide PROJ pipeline to go from Bonne to RDNAP or WGS84 (Lesparre, 2023). Furthermore, he also shared information on the accuracy of the transformation. Van Riel (1924) published a direct transformation between bonne and RDNAP which has sub-meter accuracy for the first-order triangulation points. However, in one tested area it has a standard deviation of 74 metres for topographical features because of the inaccuracy of the original map.



+proj=pipeline +z=0

+step +proj=bonne +lat_1=51.5 +lon_0=0 +a=6376950.4 +rf=309.65 +pm=4.883882778 +inv

+step +proj=cart +a=6376950.4 +rf=309.65

+step +proj=helmert +convention=coordinate_frame +exact +x=932.9862 +y=86.2986 +z=-197.9356 +rx=2.276813 +ry=1.478043 +rz=4.673555 +s=50.09450

+step +proj=noop

+step +proj=cart +ellps=WGS84 +inv

Figure 14 PROJ Pipeline to transform from Bonne to RDNAP and WGS84, by Lesparre (2023)

The PROJ pipeline from Bonne to WGS84 consists of 5 steps. The third, transformation (3) in figure 15, is based on the table of Van Riel (1924). This step creates residuals of up to 6 meters relative to the table of Van Riel in a whirly pattern, typically associated with the inaccuracies of a triangulation.

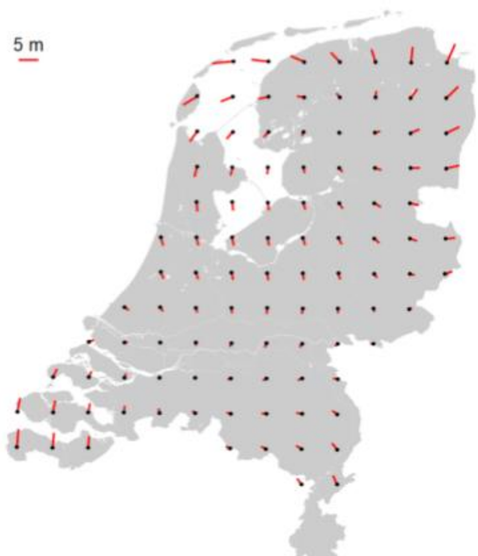


Figure 15 residuals from transformation (3) from Krayenhoff X,Y,Z to ETRS89 X,Y,Z relative to Van Riel. By Lesparre (2023).

The final step of this work package was to write both sets of sheet corner coordinates to GeoJSON. All the coordinates of a map series are written to one file where the four corners of a sheet are grouped in a MultiPoint feature with the sheet number as its id.

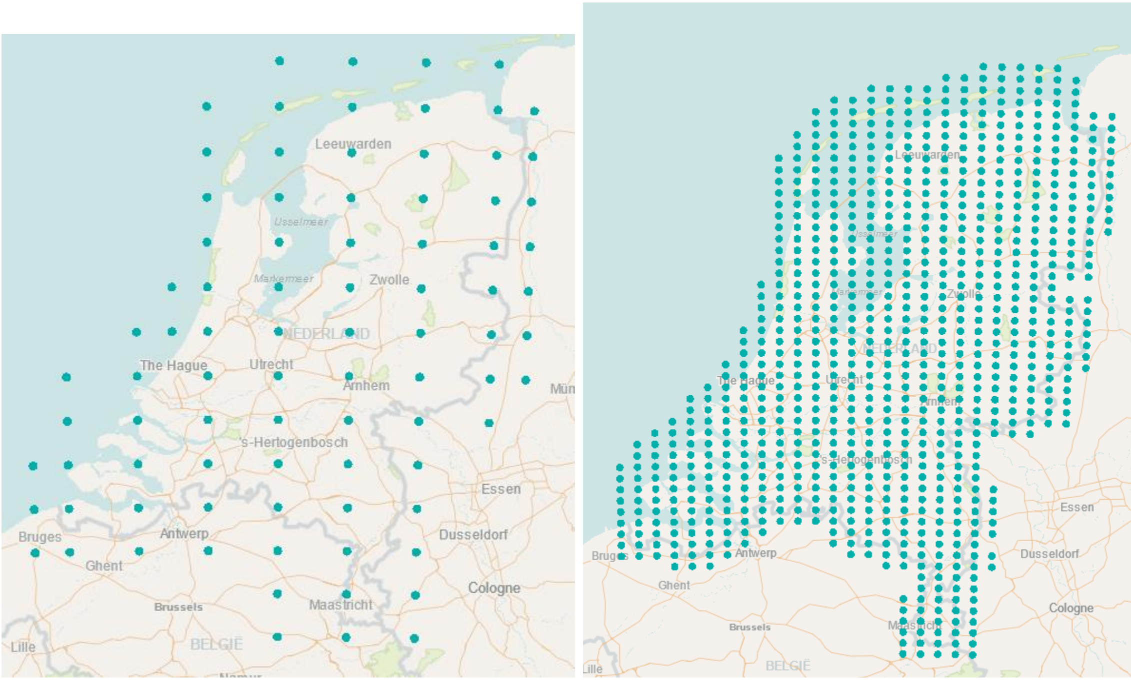


Figure 16 The resulting corner coordinates from the TMK (left) and Bonnebladen (right).

A weakness of this approach is that both of the methods to get the corner coordinates in the original CRS had steps of repetitive manual copying down of numbers which were susceptible to errors. These errors only became apparent after visualising the georeferenced map sheets but were then easy to spot, trace back, and correct.

In the case of the TMK it was discovered that map sheet number 24 is shown in the sheet index as a full size sheet and the corner coordinates of the full sheet are given in the '*Meetkunstige beschrijving van het Koningrijk der Nederlanden, bevattende de getallenwaarden, gebruikt bij de zamenstelling van de topografische en militaire kaart van het rijk'* (1861), however the map sheet in the archive is executed as a half sheet because most of it is empty sea. This was then rectified in the csv with the bonne projection coordinates of this sheet. A more recent sheet index (*Bladindeling Van De Nettekeningen*, n.d.) shows more irregularities, especially around the borders in the South. However, only the missing sheets, 50, 56 and 59 are an issue in the dataset from the National Archive.

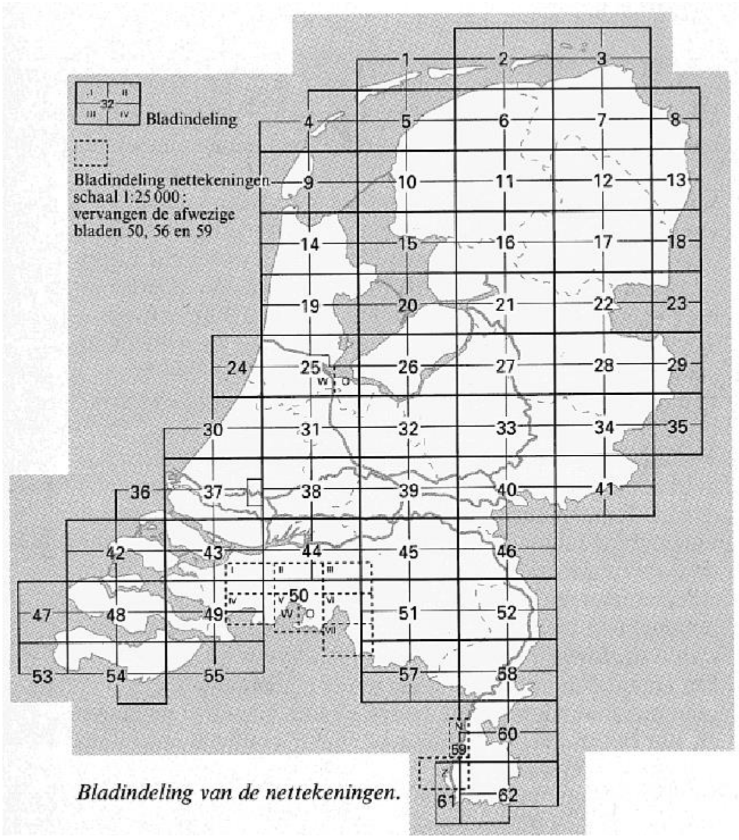


Figure 17 A modern sheet index of the TMK, by Kadaster.

Figure 18 Upside down zwolle 304 and right side up Zwolle 304A.

The Bonnebladen also have several exceptions to its regular grid. For example, sheet 10.304 Zwolle is scanned upside down in the archive. It might make more sense for the archive to correct the orientation of the image than to include an orientation check into the automated workflow. Another irregularity occurs around the border around Limburg. Here, sheets are not executed in the regular size given by the sheet index because most of the sheet area is not Dutch territory. If these different sizes are not taken into account the algorithm will stretch the image to fill the size of a full sheet.

Figure 19 Irregular sheet sizes in Limburg, by Schoonman (2023).

5.1.2 WP2: Corner detection

A vital part of automating map georeferencing is detecting the corners within which the map content exists for each sheet. Using the terminology of AllMaps, this step will create a mask of the map as illustrated in Figure 20 and Figure 21. Detecting those corners will enable assigning the coordinates, retrieved from the previous work package to the corners of the detected mask. As can be seen in the mentioned figures, the map sheet sets have different characteristics which impose handling them in differently. To do so, two approaches were developed, one based on the geometric characteristics to determine lines and one on the colours to remove the background of the images. The former applies to the 'nettekeningen' map sheets while the latter one was used for the 'veldminuten' and the 'bonnebladen'.

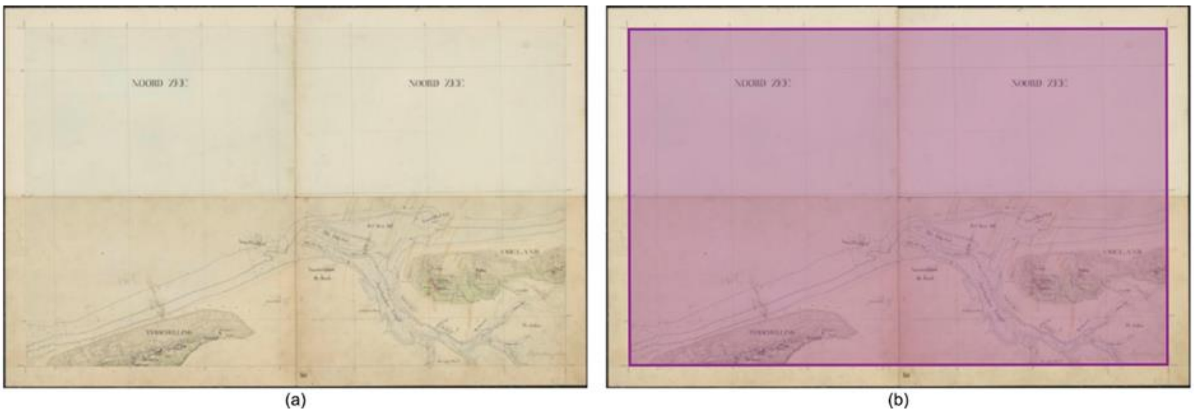


Figure 20 (a) Example of TMK nettekeningen map sheet. (b) Example of mask on the same map sheet. Source: https://service.archief.nl/iipsrv?IIIF=/ca/6c/10/d3/d1/cf/46/8f/bc/b6/c1/1d/6e/de/ec/9f/8c525345-e566-4796-8f1d-38f7187ce69c.jp2/full/full/0/default.jpg

Figure 21 (a) Example of TMK veldminuten map sheet. (b) Example of mask on the same map sheet. Source : https://www.nationaalarchief.nl/onderzoeken/archief/4.TOPO/invnr/7.1.1/file/NL-HaNA_4.TOPO_7.1.1?eadID=4.TOPO&unitID=7.1.1&query=

The goal of the first method is to identify the horizontal and vertical lines that enclose the region of the image with map content and from them extract the corners. The first step of the process is image preprocessing which includes cropping the image to discard the black background as well as reduce the white border. After that, the image is converted to grayscale and then a threshold is applied to convert the image into black and white (Figure 22 (a) and (b). Then, morphology operations are utilised to separate horizontal and vertical lines into two different images. More in detail, based on OpenCV library the horizontal and vertical lines are separated using erosion and dilation with kernels as structural elements. The size of the kernels is 21×1 for the horizontal lines and 1×15 for the vertical ones. Following this, HoughLines Probabilistic Transformation is applied to the images that only consist of horizontal and vertical lines separately (Figure 22 (c) and (d)). The detected lines are then filtered based on their bearing with a threshold of 4 degrees from the horizontal (180°) or vertical axis (90°) (Figure 22 (e) and (f)). Finally, from those filtered lines the ones with minimum and maximum x and y value are kept since it is assumed that those are the ones which are part of the outer border of the map. A mask to be valid has to be a minimum area equal to the 80% of the total area of the image, otherwise it is discarded and the mask becomes equal to the whole image. This technique has good results in regular sheets, which do not have damaged edges and the margin between the edge of the sheet and the map is wide. If the sheet has irregularities such as worn edges or the map's borders are the same with the sheet's edge then the algorithm performance is not accurate since it is very sensitive to them (Figure 23).

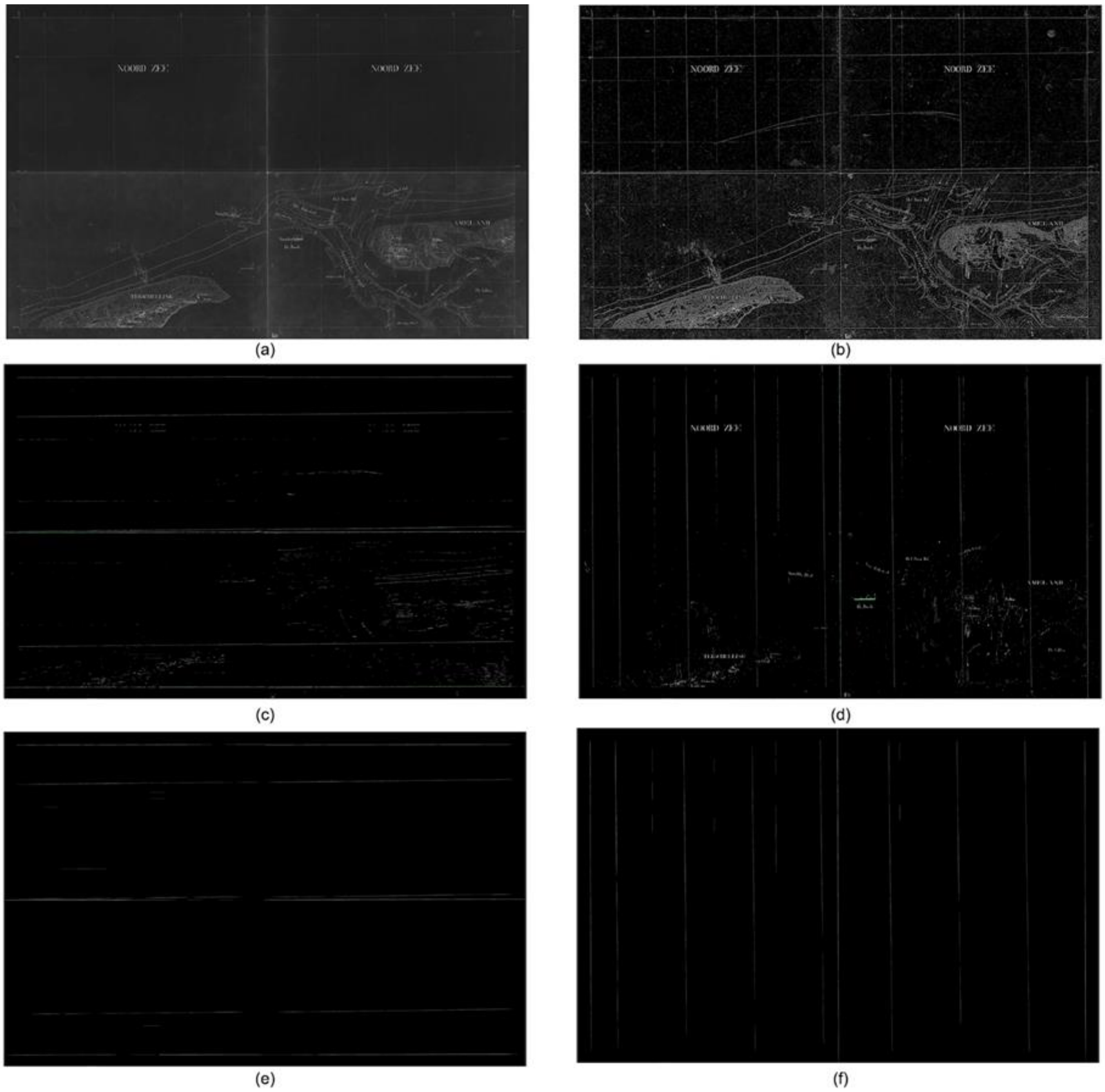


Figure 22 Steps of image processing to detect lines. (a) bitwise image (b) black and white image (c) image after applying horizontal structural element with overlaid green lines as detected from HoughLines Probabilistic Transformation (d) image after applying vertical structural element with overlaid green lines as detected from HoughLines Probabilistic Transformation (e) final horizontal lines after filtering based on angle (f) final vertical lines after filtering based on angle.

To evaluate the accuracy of that method the automatically generated masks were compared with manually generated masks and the Euclidean distance of corresponding points (e.g. top left with top left, etc.) was calculated. This process was carried out for a set of 31 images out of a total of 59 images. As mentioned before the algorithm has high accuracy in regular sheets scoring for example in some cases a Euclidean distance average of 3.83 and 7,83 pixels. The table with the comparison results can be found in Table 9 (page 82). The unit of measurement used in the table is pixels and the distance 0 to 4 is the distance of points on the 4 corners starting from top left and clockwise. It is important to mention here that due to some code bug that was solved by the time writing the present report the mask generator and the quality check algorithm have some strange behaviour that is not systematic (i.e. not the same in every image) and therefore it was difficult to debug.

Figure 23 Example of damaged map sheet.

The second method for the *'veldminuten'* and the *'bonnebladen'* map sheets is based on a completely different concept. The main idea behind it is that different columns and rows of an image have different numbers of black pixels. When there is a column or row with significantly more black pixels this is a good indication that there lies a line. Following the initial steps of the previous method the image in question is converted to grayscale and a threshold is applied to have an array with 0's and 1's. After that, the number of black pixels (1) is counted for each row and column of the image. In addition to that, the standard average number of black pixels and standard deviation are calculated. Based on the total number of pixels in each column or row the local maxima is calculated. This is done by comparing the number of black pixels of a row or column with each two immediate neighbours (previous and next one). After acquiring the indices with more black pixels than their neighbours, they have to be filtered. This is done by comparing the number of black pixels with a value. The value to identify a peak is defined by the average number of black pixels and a multiple of the standard deviation as shown in below:

$$\bar{x_i} + \sigma_i \cdot \; a \; < peak \,\forall\, i \;\in\; Image_{rows}, where, \bar{x_i} \; the \; average, \sigma_i \; the \; deviation, and \; a \; is \; constant$$

$$\bar{x_j} + \sigma_j \cdot \; a \; < peak \,\forall\, j \;\in\; Image_{columns}, where, \bar{x_j} \; the \; average, \sigma_j \; the \; deviation, and \; a \; is \; constant$$

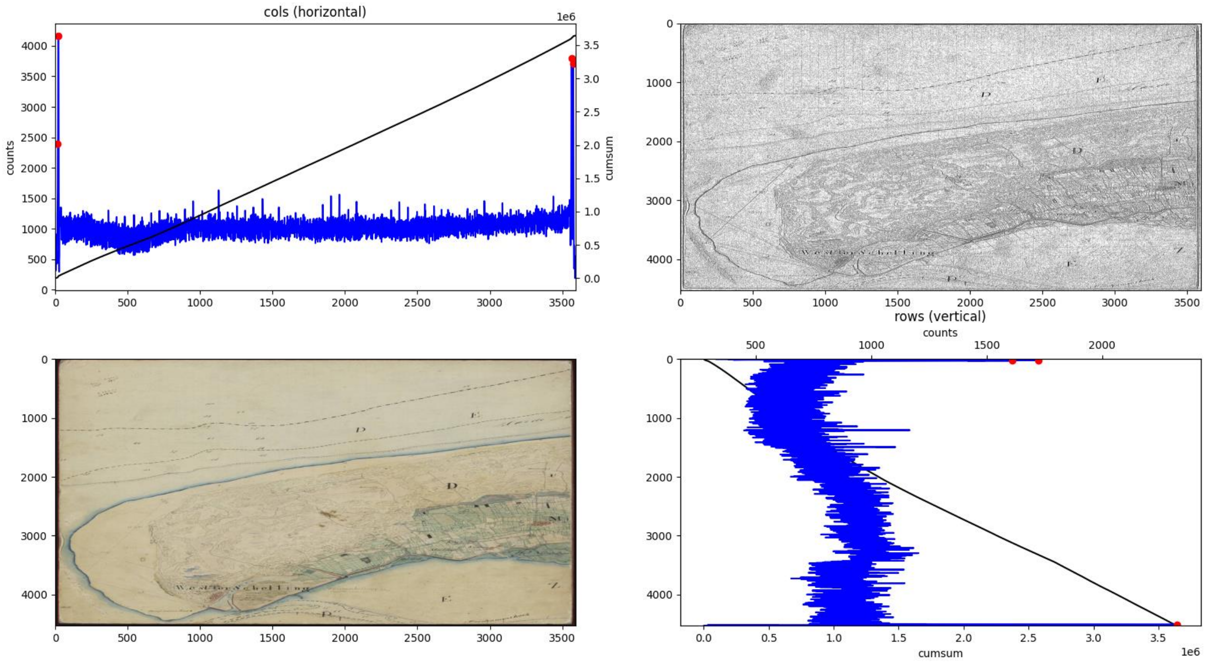


Figure 24 Example of identifying peaks of black pixels. The constant used is a=4. On top left is the number of black pixels per column. On top right the thresholded image. On the bottom left is the original image. On bottom right is the number of black pixels per row.

Since it is assumed based on visual check that by setting the constant a equal to 4, the peaks are close to the edges, from the final peaks the ones chosen to define the mask are those which have the greater value. To define the corners' points, the minimum and maximum x and y values of the selected edges are then extracted. A disadvantage that comes with that method is that there is no metric to quantify the accuracy of the mask detection except for visual checking. This is because there were not manually created masks to test against the automatically generated ones.

5.1.3 WP3: Automated writing of IIIF georeference annotation pages for the TMK and Bonnebladen

This part of the workflow combines the results of WP1 and WP2 into annotation pages that can display the images in the Allmaps Viewer. The package processes one map series at a time.

First, for each sheet the sheet number is extracted from the metadata and the pixel corner coordinates are extracted from the mask in the annotation page created in WP2. Next, the sheet number is used to look for the corresponding geographic corner coordinates in the GeoJSON generated by WP1. The corresponding pixel and geographic coordinates are stored as 'ControlPoints' that are written to an updated annotation page.

These annotation pages can then be loaded individually or as a series into the Allmaps viewer.

5.2 Phase 2

5.2.1 WP4: Search box determination and processing

For the second phase of the project, the results of the first phase are required. We aim to georeference the field minutes of the TMK using the georeferenced neat drawings from Phase 1, put together as a whole and then used as a base map for the process.

Work Package 4, the first work package of the second phase, aims to determine a search box that defines the borders of the reference map used for the feature matching of Work Package 5.

A search box in our case, is a box, defining the borders of the search area of the feature-matching tool that follows up in our pipeline. It is used as a reference map so that the position of the map sheet that is being georeferenced- in the case of the TMK, each fieldminute- is detected on the base map. In that way the feature matcher no longer needs to search for matches on the whole base map, but only around the area that the search box determines that it should. Thus, the process of feature matching is a lot faster and we make sure that the detection of the corresponding features is done in the correct area. That way we ensure that the georeferencing is not based on invalid GCPs detected in random areas around the map, a potential weakness of the deep learning model.

For our specific project, we can use the neat drawings of the TMK as base map for georeferencing the field minutes. Due to the structure and metadata of the Dutch NA, it is easy for us to retrieve which field minute should be on which neat drawing. Therefore, for our project, it is not required to use the search box to limit the area, since we have a limited area in the neat drawing. Also, since the original neat drawing is full quality and not warped, the georeferencing from feature matching is expected to be of higher quality.

However, for the sake of reproducibility and interoperability, we still wanted to implement a way that allows any kind of map series to be georeferenced using our method with any type of base map. That way, even if the sheet structure is different between maps, or no sheet division is used at all, the pipeline can still work. For that, we came up with the idea of the search box determination step.

We decided the search box would be created from a centre point, and an offset on the x and y axis from this point. In our implementation, we ask the user of the library to give this centre point as a coordinate in WGS84. In our study case, the Dutch NA has included as metadata the title of the map sheet that the field minute belongs to. This title corresponds to a geographic location in the Netherlands. Both for us to use, and any other user of the library we implemented a geocode function to produce the search box centre point from that toponym.

For our study case and because we are dealing with historical maps, we chose to work with the World Historical Gazetteer, which is a digital platform that works with open data principles and compiles historical place names and associated data. A strong point of this platform is that it does not matter if the user gives it an older version of the toponym because it is aware of the historical toponyms that are included in its database.

We make use of the WHG in the “geocoder” file of our source code that performs the actual geocoding of a toponym. More specifically, it takes the toponym and a list of ISO country codes as input. It validates the input to ensure the list of country codes is not empty and that each code is valid. Then, it constructs a URL to query the WHG API with the toponym and the provided country codes. It sends an HTTP GET request to the API and processes the response. If a single match is found, it extracts the WGS84 coordinates and returns them in the form of a GeoJSON. If more than one or none are found it returns None. During tests, we noticed that one toponym can correspond to more than one country. That is the reason why the user is also asked to give the code of the country at the start.

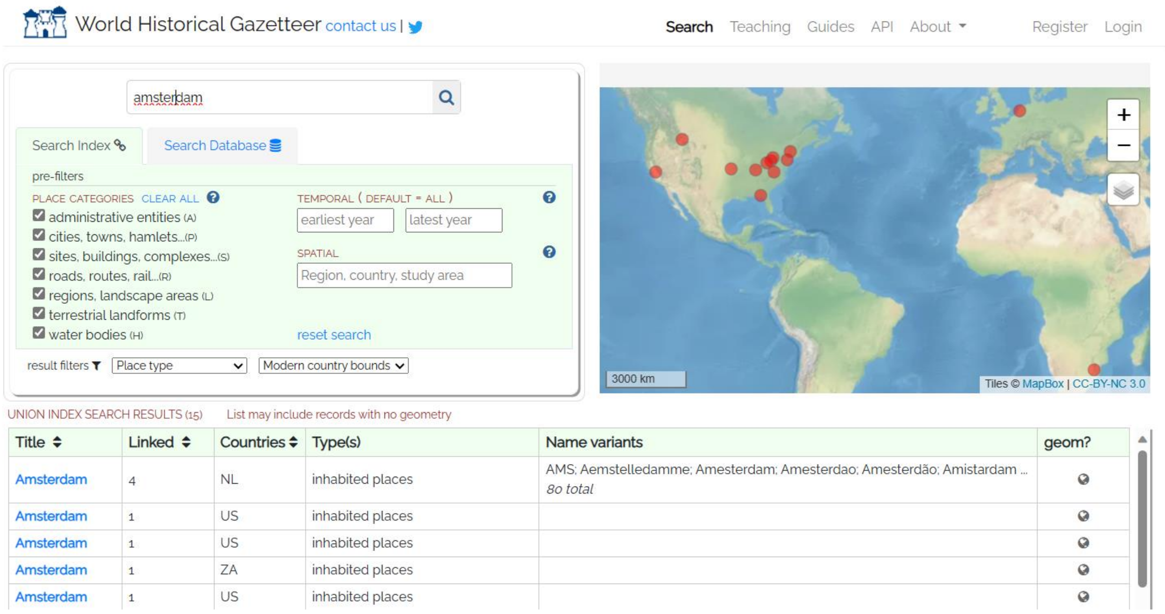


Figure 25 Searching for "Amsterdam" on the WHG gives back 3 countries ISOs(NL, ZA, US).

To determine the x and y offset of the searchbox, project-specific information is needed. If the user gives a centre point close to the geographic centre of the map sheet they are georeferencing, a small search box could be used. In our case, the toponym in the metadata corresponds to the toponym of the much bigger neat drawing that the field minute is part of. We inspected the nettekeningen to determine if the toponym is located in the centre or more unpredictably distributed.

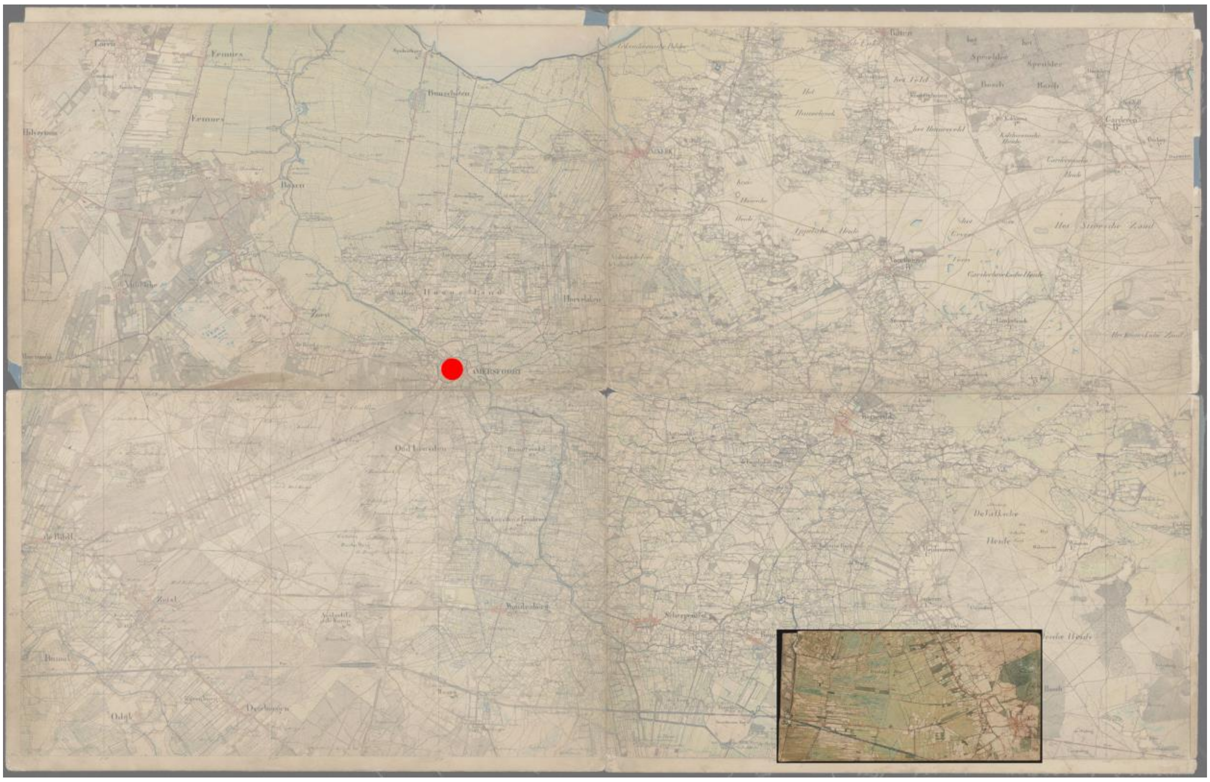


Figure 26 With a red point: Amersfoort on the neat drawing of the TMK shown with opacity. On the bottom right, a field minute corresponding to Amersfoort's neat drawing.

From this, we concluded that the toponym's location on the neat drawing map sheets is irregularly distributed.

We conclude that, both from the field minute's toponym belonging to a much larger neat drawing map sheet and the actual location of this toponym on the neat drawing being irregular, we need to use a searchbox much larger than the field minute and even substantially larger than a neat drawing.

Based on the size of the neat drawing area and the distribution of the toponyms, while still minimising the search box area to improve performance, we ended up with an x offset of 30 km on both sides of the centre point and a y offset of 20 km above and below the centre point. The example in Figure 27 shows a searchbox(red) created from the toponym of sheet "Ameland", where Ameland is not located in the centre of the map sheet(blue). Because of the selected offset, still the entire map sheet lies inside the search box. As visible by the remaining space around the sheet, even if the toponym lies further from the centre, the search box will span the entire sheet.

Figure 27 The selected search box(red) for the fieldminute of Ameland(blue). The red point represents the centre point of the search box, where Ameland was geocoded.

After determining the search box area, the next step is extracting the searchbox as an image from the reference map. We thought of multiple methods for achieving this, including a WMS server or writing our own code to extract from a IIIF manifest. Due to the complexity of creating our own WMS server or developing our own exaction code that could work across multiple individual IIIF images, we decided to go for a ready-to-be-used solution our client informed us about. Allmaps offers a XYZ tile server[1] solution that can produce image tiles from a supplied IIIF georeference annotation. We implemented code that requests all tiles within the search box area, and combines them into one composite image.

Our client informed us about the downsides of this solution, including the negative impacts of CRS warping and image quality degradation on the georeferencing result. Also, implementing our own image extraction could be much faster.

This whole process has been implemented in the SearchBox and ReferenceMap classes, which work packages 5 and 6 use to continue with the pipeline.

[1] XYZ map tiles for IIIF maps georeferenced with Allmaps / Allmaps | Observable (observablehq.com)

5.2.2 WP5 & WP6: Finding feature matches between map and reference map

In these work packages, the map-to-be-georeferenced is compared with an already georeferenced map using feature matching. The georeferenced map "reference map" comes from work package 4. In our project the maps-to-be-georeferenced "input map" are the veldminuten of the TMK.

Originally the plan was to split the input map into several different images that would describe different features on the map. For example a roads map, a water map, and a buildings map. This idea was based on the paper *Automatic content-based georeferencing of historical topographic maps* (Luft & Schiewe, 2021). In this paper, the water features from a historical map are extracted and, using a feature matching algorithm, compared with a modern day map of the same area.

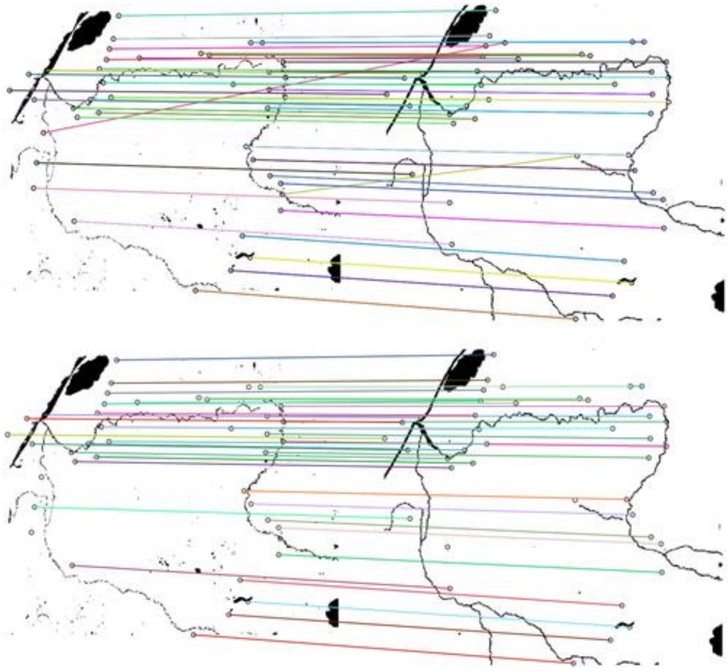


Figure 28 Result from Luft & Schiewe: Matches of detected features between source image (i.e., the segmented input map image, left) and the target image (i.e., the reference OpenStreetMap data, right). (Luft & Schiewe, 2021).

We considered using this exact approach, and then also adding masks for buildings and road features.

Our segmentation attempts ran into a dead end however. It turns out on the TMK veldminuten water features are often much less easily distinguishable than on Luft & Schiewe's map sheets of Germany. Within the TMK fieldminutes series the colour used for water varies significantly, making it very challenging to automatically extract the water features. Initial tests to extract water features were therefore unsuccessful.

Figure 29 Water features on the Delft TMK fieldminute sheet. The water features are easy to distinguish due to the blue colour use.

Figure 30 Water features on the Utrecht TMK fieldminute sheet. Water features are almost impossible to distinguish based on colour.

For the segmentation of buildings, colour would also be an appropriate filtering method as red is used to denote buildings on the map. On closer inspection, we found that even within the same map series (TMK veldminuten and TMK nettekeningen) the method for drawing buildings can differ a lot.

Figure 31 Amersfoort on the TMK fieldminutes. Buildings are drawn with high detail.

Figure 32 Amersfoort on the TMK neat drawings. Blocks are drawn with an outline.

Figure 33 Delft on the TMK neat drawings. Blocks are drawn as solid red blocks.

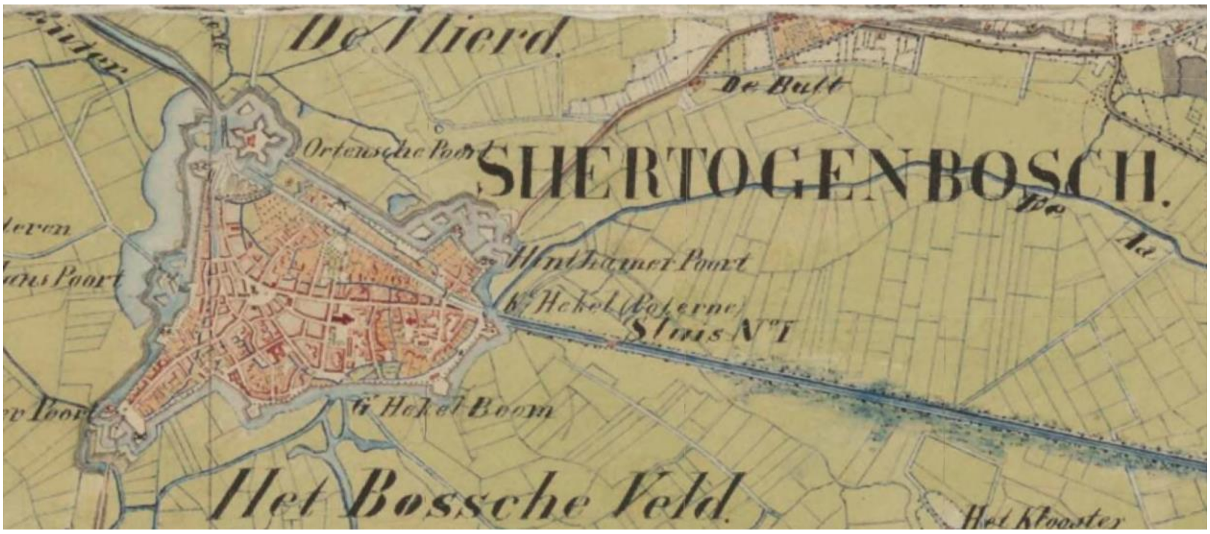


Figure 34 ‘s Hertogenbosch on the TMK neat drawings. Blocks are drawn with a red drop shadow.

Finally, the sources we found for methods to extract road networks all required the maps to be already georeferenced for that specific method to work (Uhl, 2022) or require retraining of a deep learning model (Jiao, 2022).

The combination of these conclusions led to the decision that segmentation of historical maps that would also be reusable for other map series was not feasible within the time and resources of this project.

Following this decision, we performed a range of experiments. Firstly, we tried out algorithms for template matching (OpenCV: Template matching, n.d), which is the process of moving a sample image across a target image and, using statistical analysis, determining the most likely position of the sample image on the target image. The results from this method for our specific project were promising, when input images were placed in approximately the correct location on the reference map. However, due to the nature of this algorithm, it would completely fail on even minor rotation and scale changes in either the input image or the reference image. This would mean it could definitely not work on a reference map in another CRS. Also, based on manual overlaying, it turned out that the fieldminutes do not overlay exactly with the neat drawings. Based on these two conclusions, we decided a more advanced method of placing the fieldminutes was needed that could account for all these challenges. Feature matching algorithms seemed the obvious choice for this, as we could transform the image based on the found matches.

We tested multiple feature matching implementations, including ORB (Oriented FAST and BRIEF) (Rublee et al., 2011) and SIFT (Scale Invariant Feature Transform) (Lowe, 1999). We experimented with these feature matching techniques both on the raw image, and on a pre-processed image. We tested the following preprocessing techniques: Thresholding, Canny edge detect, Sobel edge detect, erode, and dilate. All with a combination with Gaussian blur.

We both inspected the raw results and clustered results using DBSCAN (Ester, 1996). Upon inspecting the results, none of the tested methods produced promising results. Even when testing feature matching between two identical sheets but with slightly different processing (to simulate differences in map scale and style) barely any actual matches were found.

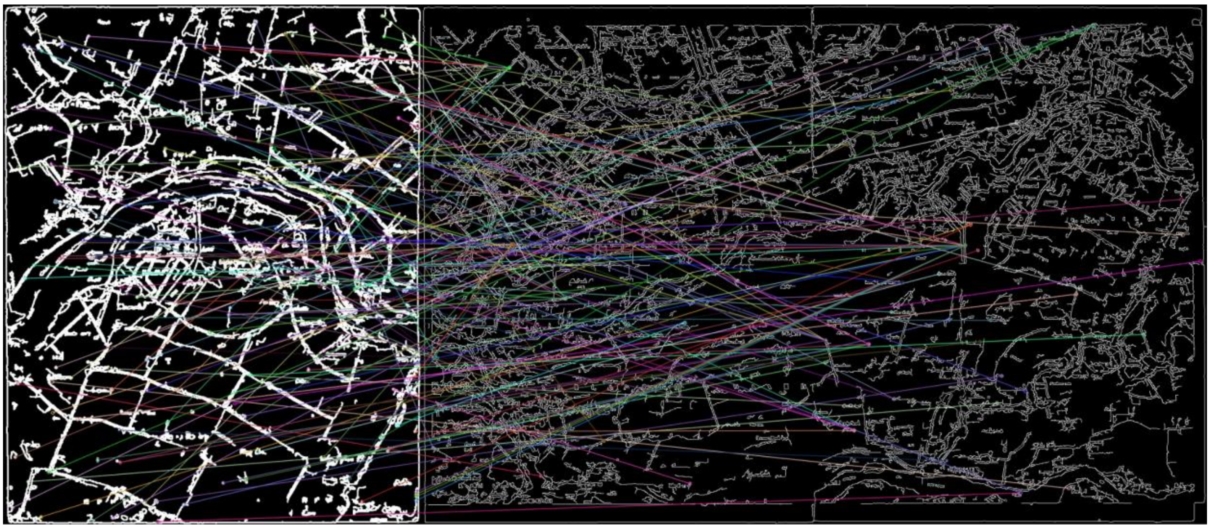


Figure 35 Subset from result from ORB based feature matching on edge masks of a fieldminute sheet (left) and a neatdrawing sheet (right).

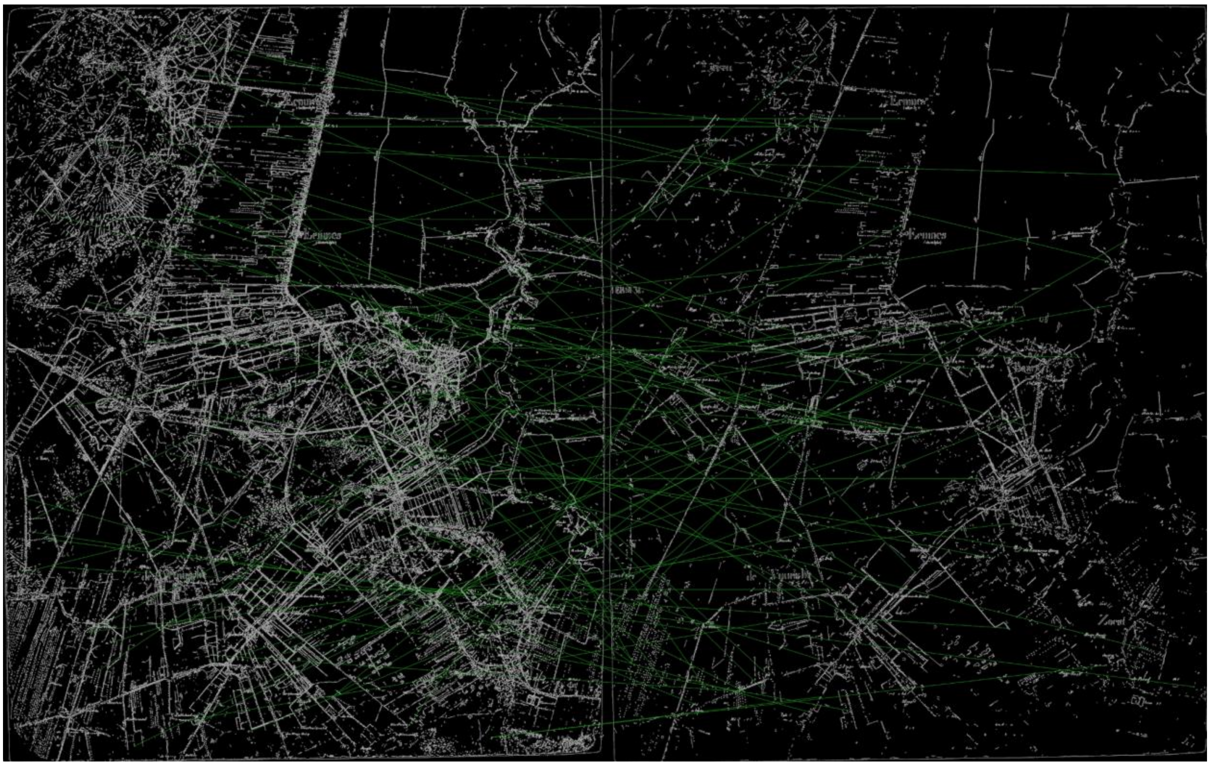


Figure 36 Subset of ORB feature detection match results on the same fieldminute (left and right) with altered edge detection parameters. Manual inspection revealed only one correct match.

Finally, we found the paper and codebase *Deep learning algorithm for feature matching of cross modality remote sensing images* (Lan et al., 2021). This deep learning model promised to produce much more similar descriptors for the same features on two maps that were significantly different (i.e. matching between a colour satellite image and an infrared satellite image, two satellite images at different moments in time, and even between OpenStreetMap and a satellite image).

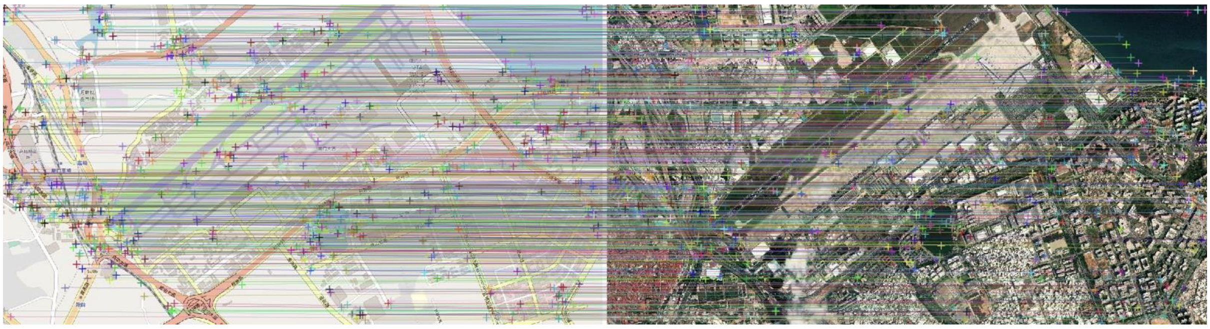


Figure 37 Image from GitHub repository of Lan et al., showing found features between OpenStreetMap(left) and satellite image (right). (Lan et al., 2021)

Already when using the model and supplied pipeline out of the box, the results were very promising. Many matches were found between the input image and the reference map, and outliers were effectively filtered out by the use of DBSCAN (Ester, 1996) in the code of Lan et al..

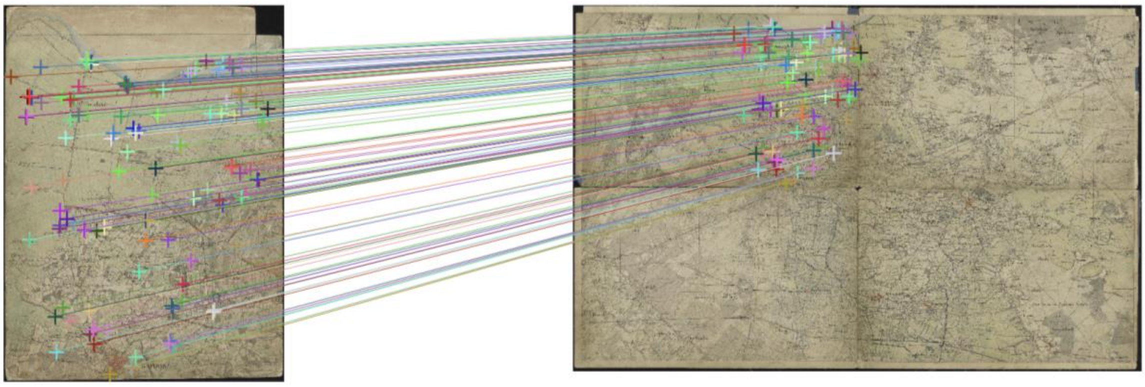


Figure 38 Plot from result of running the code from Lan et al. on a sample TMK fieldminute (left) and corresponding neatdrawing (right).

We adapted part of the code of Lan et al. to work with both directly georeferencing from the corresponding neatdrawing and by using the reference map derived from workpackage 4 combined with a location hint from geocoding the title of the fieldminute. The retrieved matches are passed onto workpackage 7 to perform further filtering and turn them into a georeference.

5.2.3 WP7: Georeferencing based on matches and output to IIIF

We transform the found matches that are below an arbitrary residual threshold that is linked to the resolution of the image using an affine transformation. After this transformation, we keep a specified number of matches with the lowest residuals (different values used, but often keeping 4 or 5 for the georeferencing). These matches are then georeferenced by associating the pixel coordinates on the reference map to coordinates in the WGS84 CRS. In the case of the reference map retrieved from the XYZ server inaccuracies are introduced here. In the case of using our IIIF image annotations of the neat drawings as reference, we can use their georeferencing directly to interpolate the pixel coordinates.

As in workpackage 3, the georeferencing is exported to GeoJSON and included in the reference annotation pages. Workpackage 7 mostly got absorbed by the other workpackages.

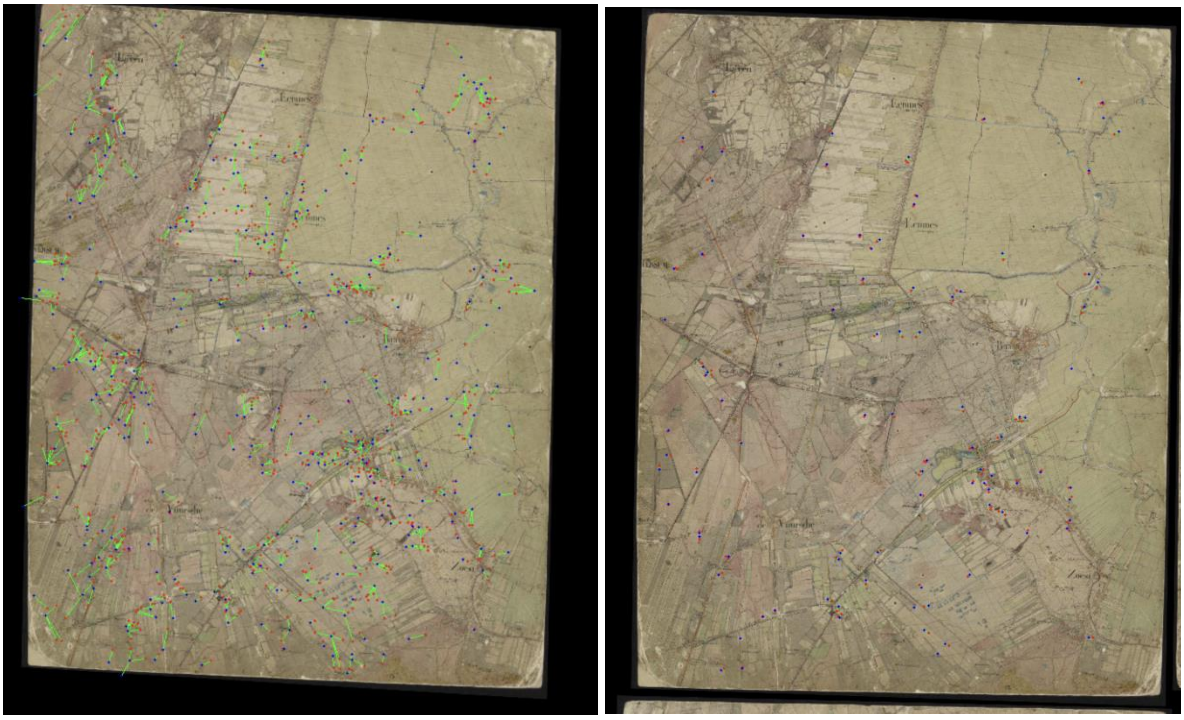


Figure 39 Matches before and after filtering based on residual length.

6. Results

6.1 Results of Phase 1

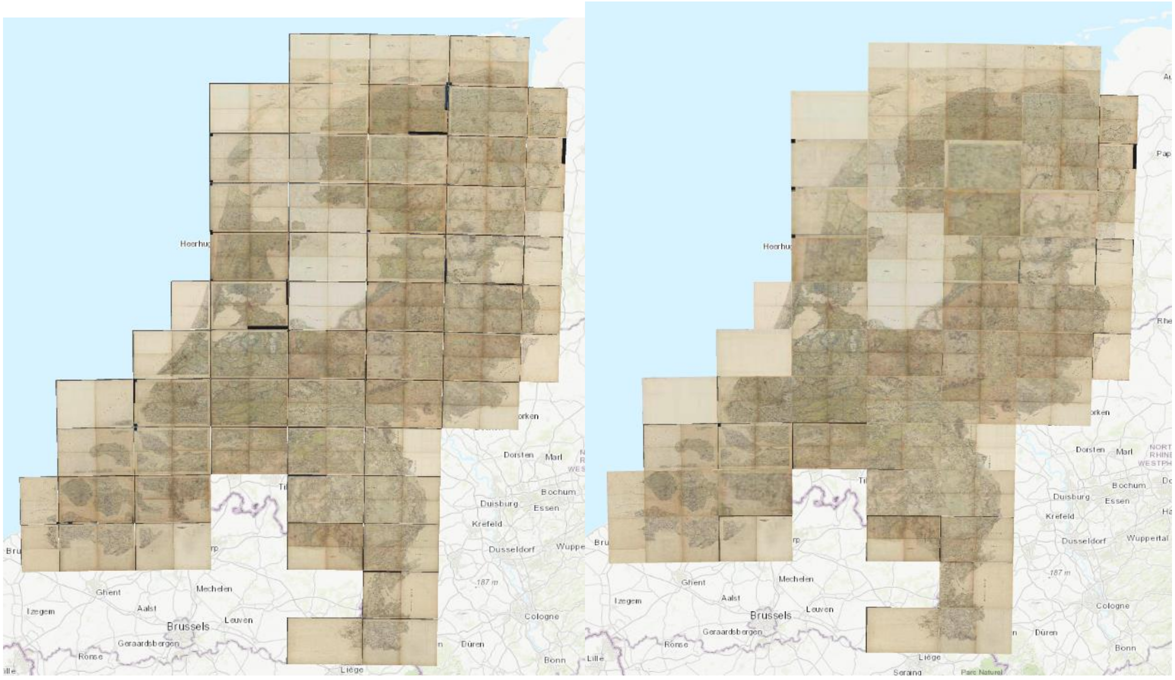


Figure 40 TMK neat drawings without mask (left) and with mask (right).

All correct edges	8
Three correct edges	15
Two correct edges	14
Mask at quarter sheet	11
No mask	11
Total	59

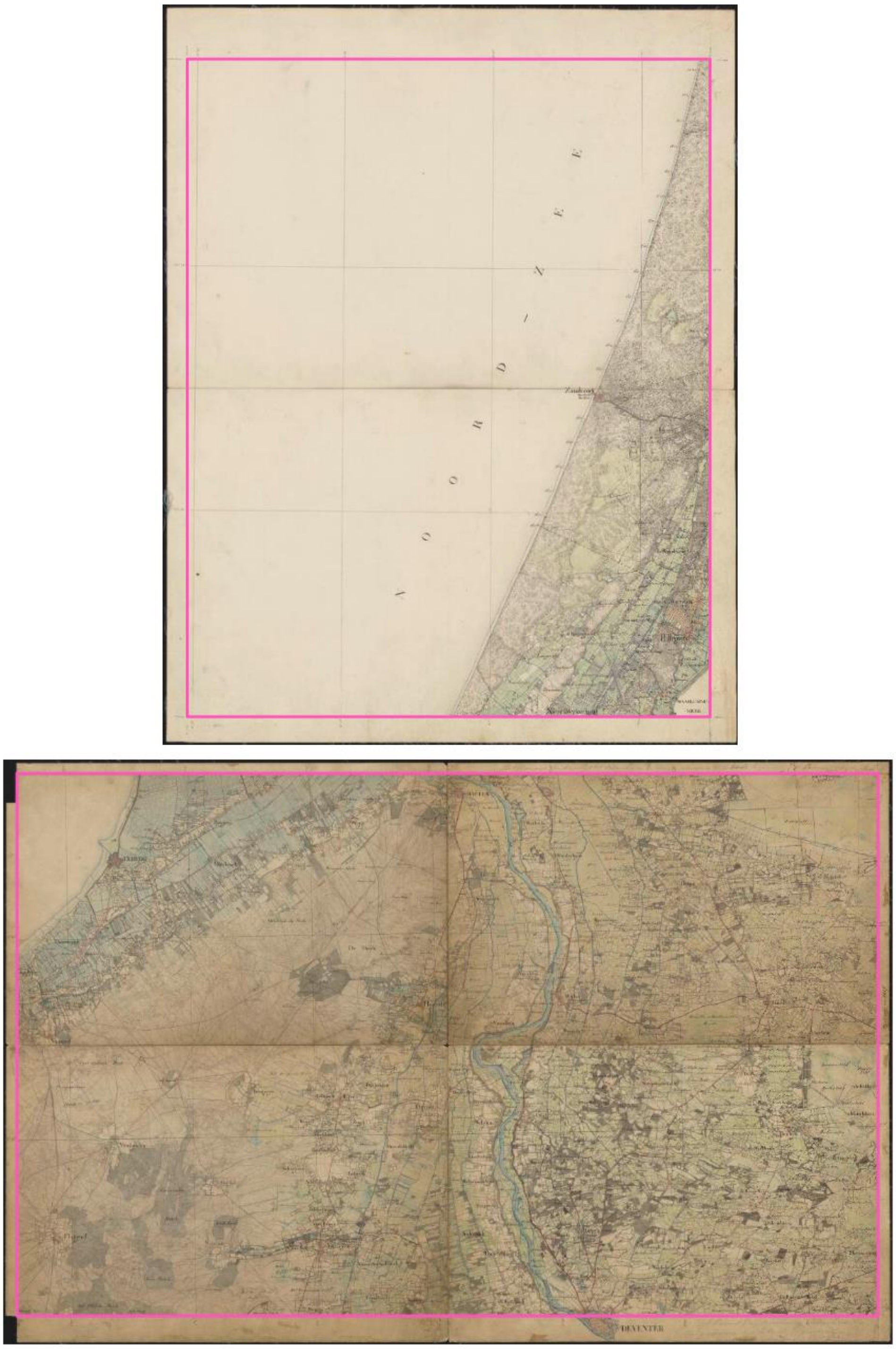


Figure 41 Examples of a correct mask.

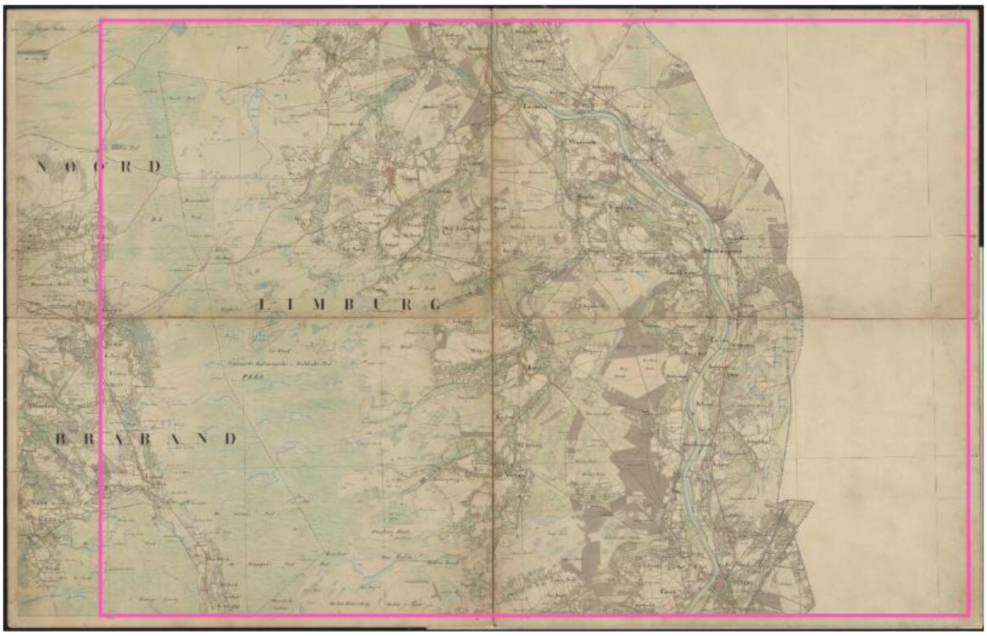


Figure 42 Example of one wrong edge.

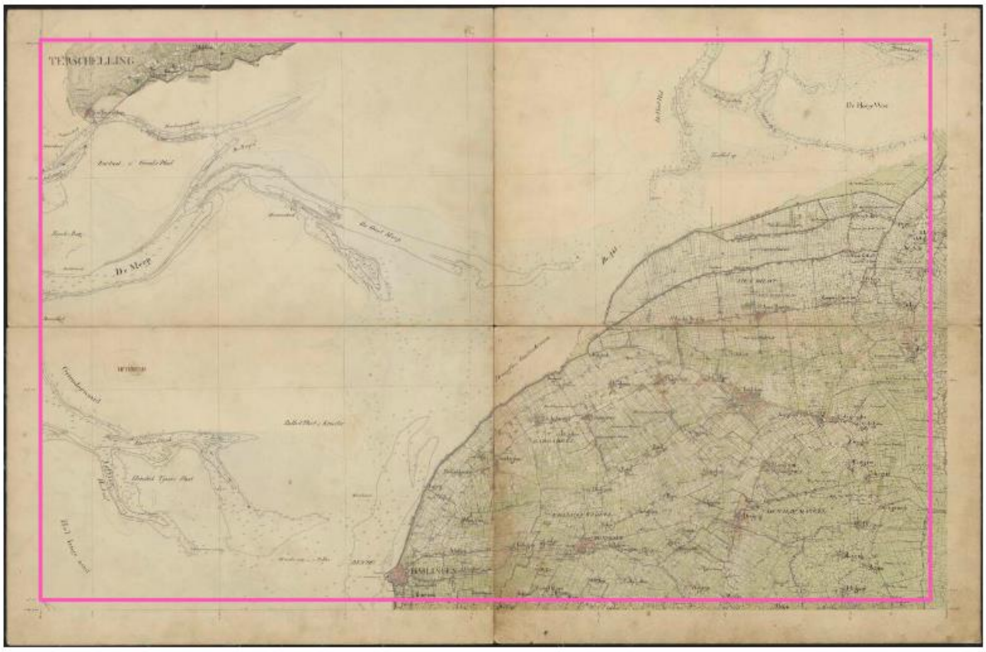


Figure 43 Example of two wrong edges.

Figure 44 Example of no mask.

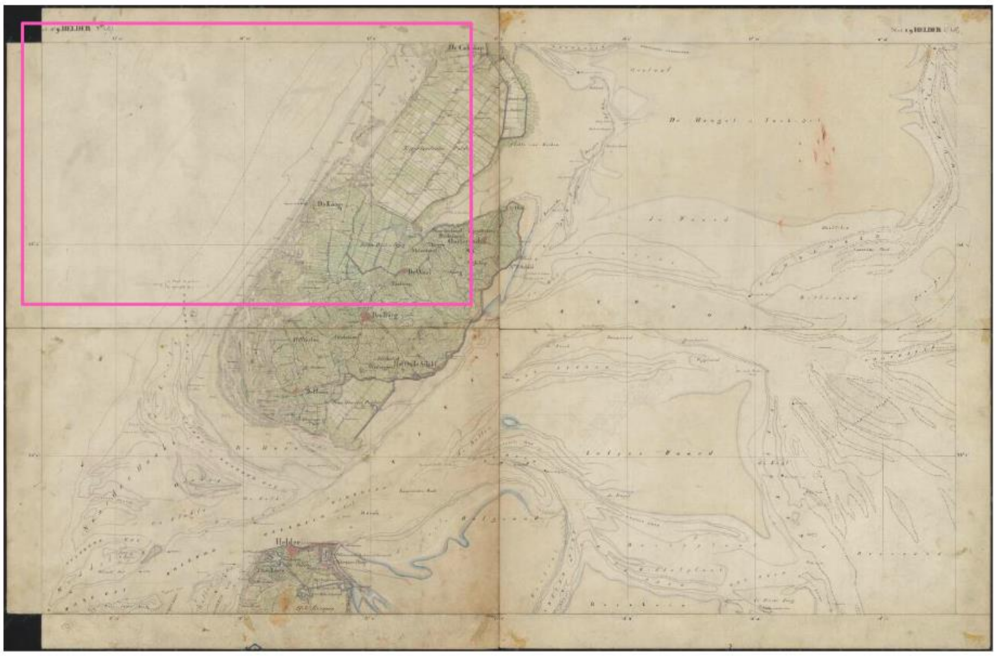


Figure 45 Example of a quarter mask.

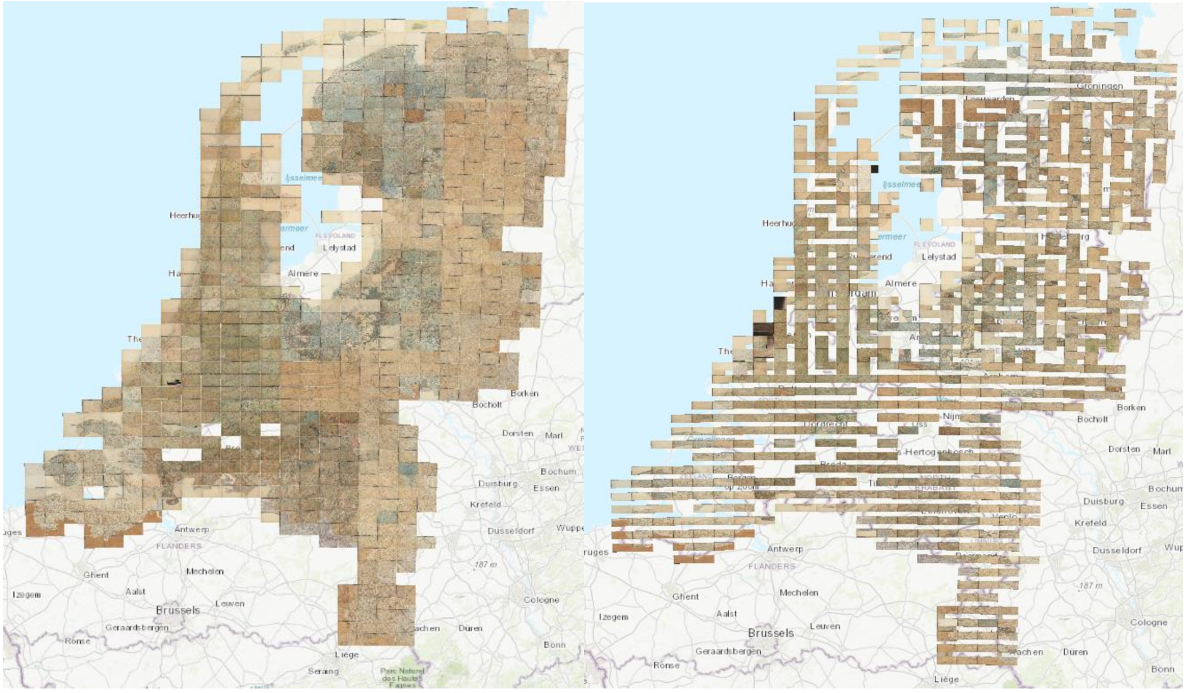


Figure 46 Bonnebladen without mask (left) and with mask (right).

Figure 47 Close-up of the Bonnebladen with mask.

Something goes wrong in the combination of the Bonnebladen pixel and geographic coordinates. Some images are put in a square mask and rotated, while others are compressed horizontally and some of the right is clipped off.

Scale inaccuracy is the percentage difference between the scale the map should be and the scale the georeferenced final map is. So if the map is exactly right the scale inaccuracy is 0% and if it's twice as big as it should be its inaccuracy is 100%. The following plots and statistics show the scale inaccuracy of the final georeferenced maps.

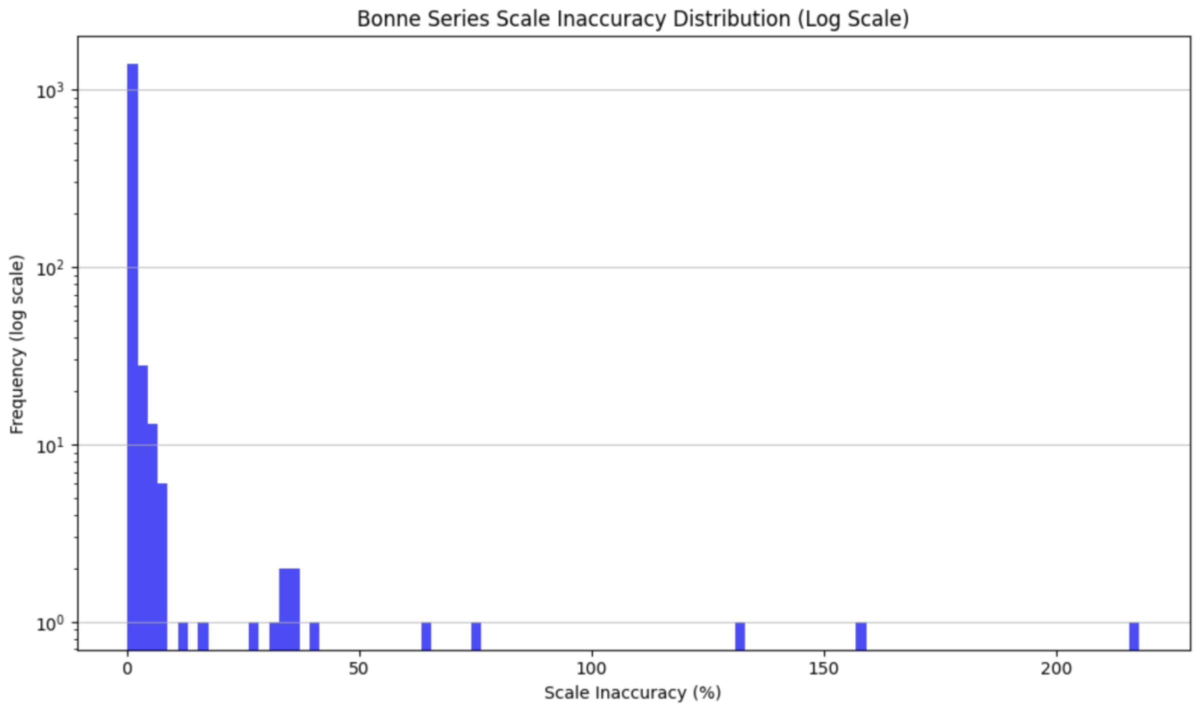


Figure 48 Scale inaccuracy statistics for Bonnebladen Series.

Scale inaccuracy statistics for Bonnebladen Series:	
Minimum	0.14%
Maximum	217.97%
Median	1.02%
Mean	1.76%
Standard Deviation	8.58%

Table 3 Scale inaccuracy statistics for Bonnebladen Series.

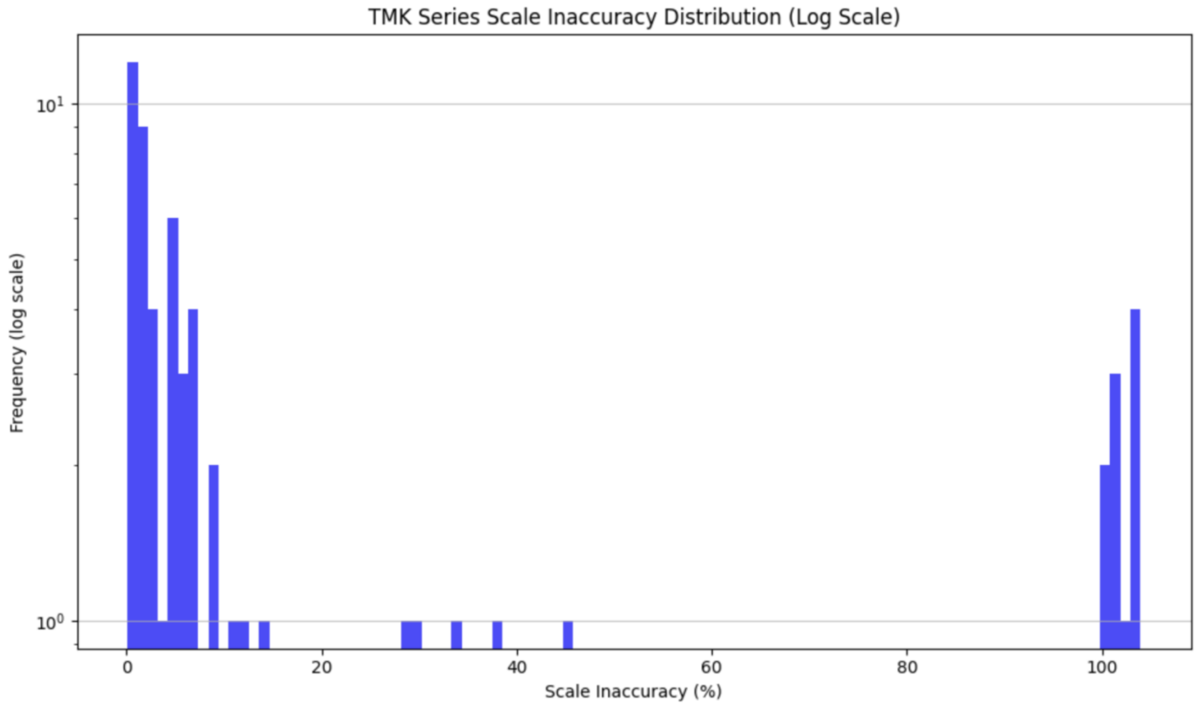


Figure 49 Scale inaccuracy statistics for TMK Series.

Scale inaccuracy statistics for TMK Series	
Minimum	0.11%
Maximum	103.94%
Median	4.78%
Mean	23.05%
Standard Deviation	36.97%

Table 4 Scale inaccuracy statistics for TMK Series

In order to calculate residuals, we manually georeferenced corners of church towers between the historic map sheets and google maps. These coordinates were then compared to coordinates interpolated from our georeferencing to calculate residuals. The results are visible in table 5 and 6.

The residuals of the TMK are not consisted due to the wrongfully applied masks. The residuals of the Bonnebladen are more consistent and accurate except for Maastricht. This is due to the over-stretching of the maps sheet because of its irregular size.

Church tower	Pixel x	Pixel y	Real lat	Real lon	Interpolated lat	Interpolated lon	Residual (m)
Maastricht	4121	2090	50.848819	5.687873	50.842892	5.642607	3245
Oirschot	1213	1237	51.503519	5.307014	51.547399	5.196825	9050
Aalten	3965	2809	51.926727	6.580846	51.927126	6.554467	1809
Vlissingen	2406	3925	51.442753	3.574768	51.443634	3.572856	164
Delft	2971	555	52.012597	4.361457	52.032046	4.183996	12333
Utrecht	4475	2462	52.09101	5.122854	52.179730	4.859829	20483
Amsterdam	5139	3338	52.374088	4.892436	52.373609	4.889481	207
Zwolle	2760	2888	52.511853	6.092773	52.625137	5.927926	16816
Medemblik	4242	2301	52.773155	5.103315	52.772959	5.103593	28
Groningen	3173	4433	53.216666	6.562291	53.217730	6.562010	119
Leeuwarden	741	4959	53.204921	5.804712	53.204678	5.805407	53

Table 5 TMK residuals in WGS84

Church tower	Pixel x	Pixel y	Real lat	Real lon	Interpolated lat	Interpolated lon	Residual (m)
Maastricht	2058	1606	50.848819	5.687873	50.848739	5.673089	1037
Oirschot	4431	2719	51.503519	5.307014	51.503879	5.306263	65
Aalten	3191	555	51.926727	6.580846	51.926365	6.580554	44
Vlissingen	4249	2605	51.442753	3.574768	51.443320	3.573878	88
Delft	1977	2527	52.012597	4.361457	52.012821	4.361661	28
Utrecht	3033	1410	52.09101	5.122854	52.090796	5.122702	26
Amsterdam	292	1296	52.374088	4.892436	52.374016	4.892574	12
Zwolle	998	2590	52.511853	6.092773	52.512155	6.092863	34
Medemblik	2325	977	52.773155	5.103315	52.772966	5.104307	69
Groningen	1013	685	53.216666	6.562291	53.216397	6.562455	31
Leeuwarden	732	1712	53.204921	5.804712	53.204992	5.804586	11

Table 6 Bonnebladen residuals in WGS84

6.2 Results of Phase 2

The results from phase 2 are of varying quality. We implemented the georeferencing of the TMK fieldminute sheets both using the corresponding neatdrawing as a reference map and using the searchbox extracted image from the whole TMK neatdrawing map (the result of workpackage 4).

The results from the pipeline are unpredictable. Running the same configuration with the same mapsheet on the same reference sheet can give highly varying results. The following images show the same mapsheet, georeferenced using the pipeline with the exact same configuration.

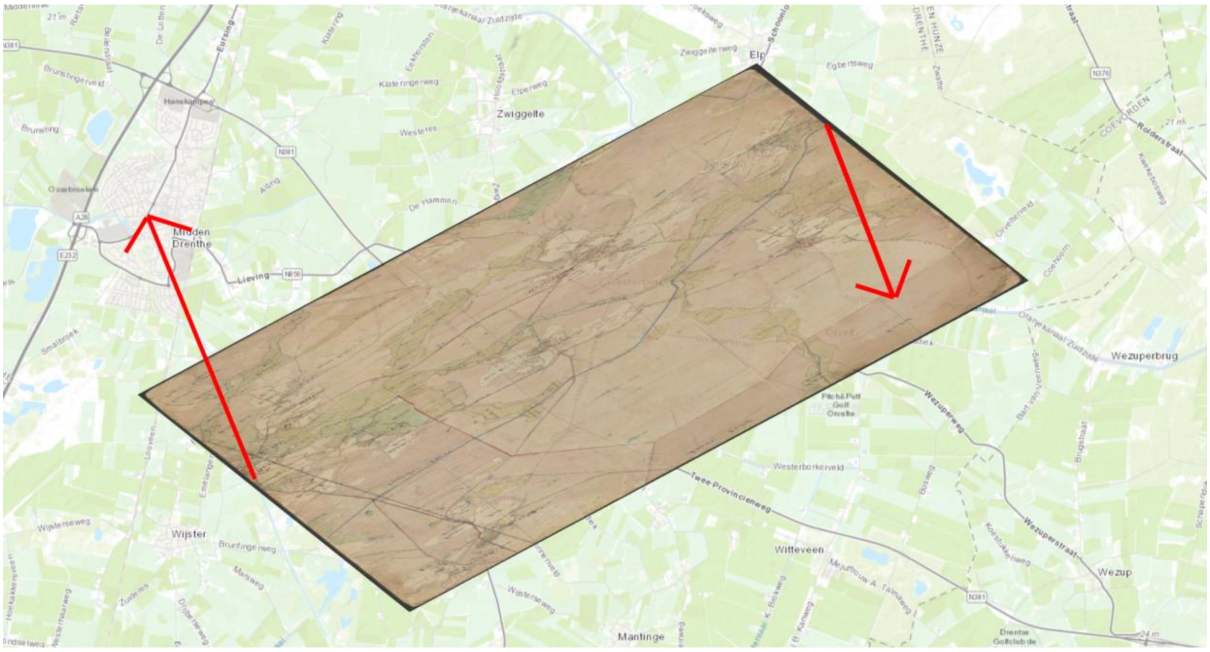


Figure 50 Result of georeferencing is in the correct location, the correct scale but highly skewed and squashed. Red arrows show where the map locations should have been georeferenced to.

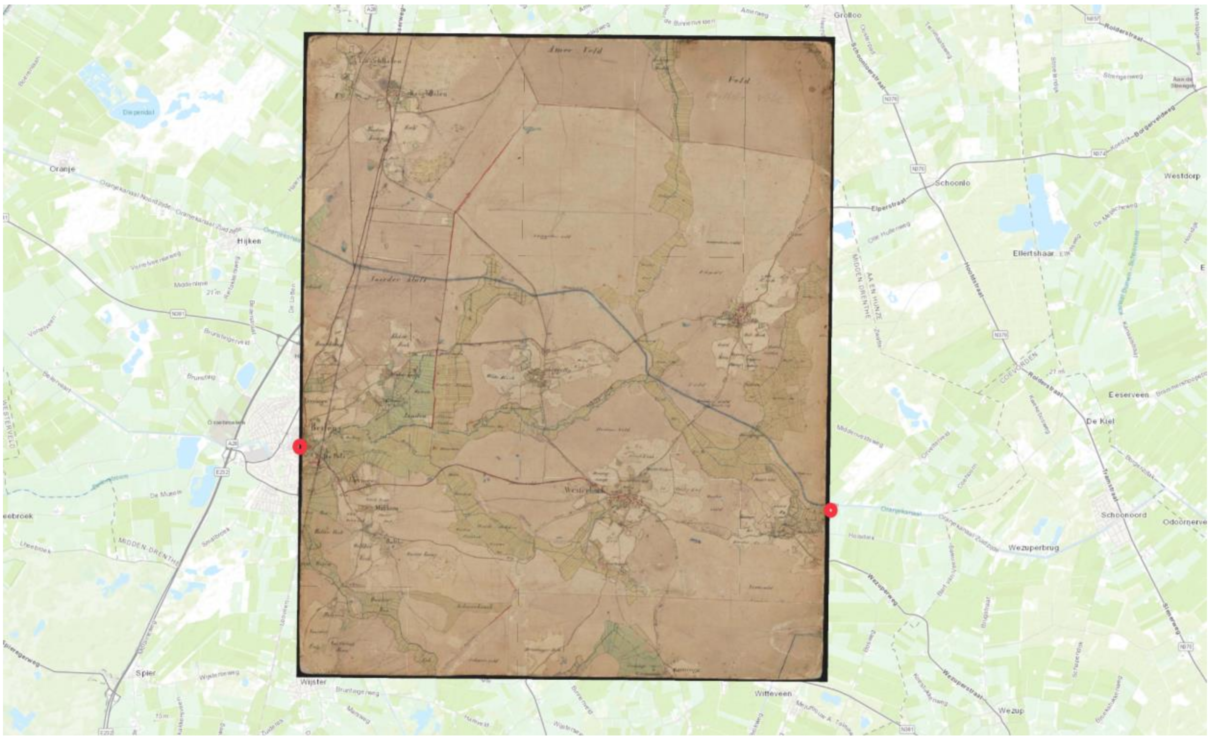


Figure 51 Same map as in previous figure, but now with correct georeferencing. Red dots mark the locations of the arrows from the previous image.

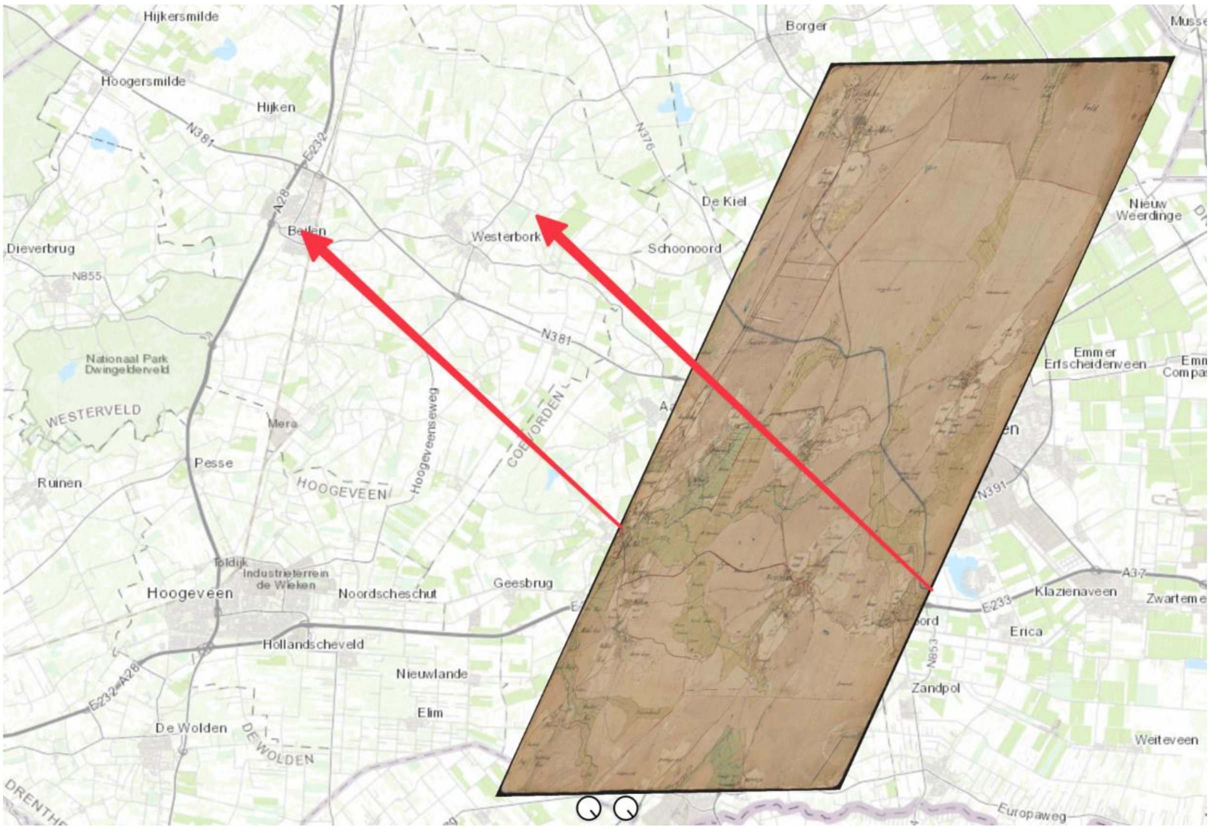


Figure 52 Same map as previous figures, now in the incorrect location, with incorrect scale, and highly skewed. Red arrows mark the same locations as previous figures.

When visualising the matches, the erratic nature of the found matches also becomes apparent, even if the affine transformation based on the collection of all matches still is quite accurate.

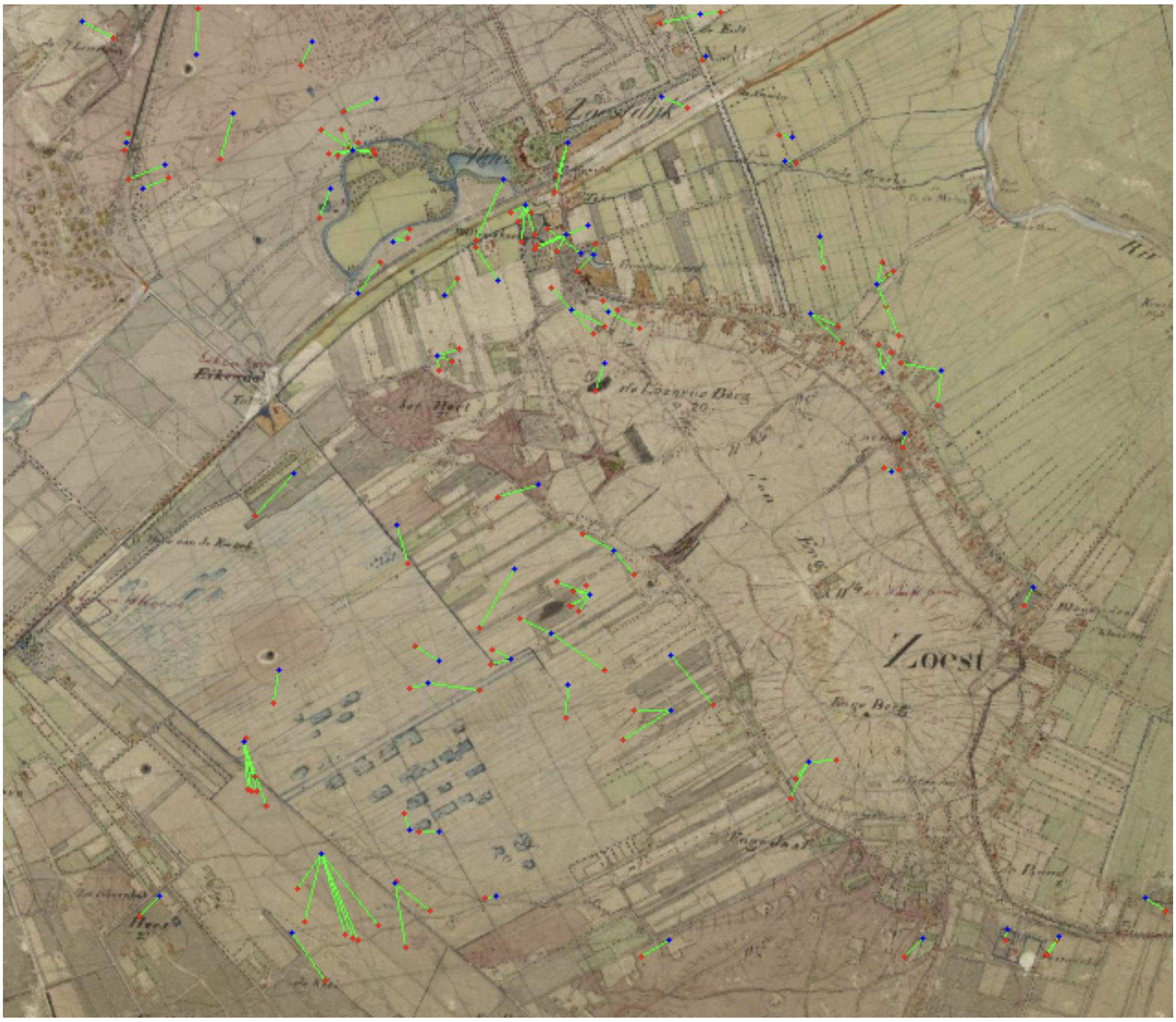


Figure 53 Erratic pattern of found matches. Blue point shows match location on the input image (shown) and red the found match on the reference image (not shown).

Optimally we could fix this unpredictability in the pipeline, but in the given time we were not able to resolve the issue. Therefore, we implemented quality metrics for the georeferencing, and made the pipeline simply retry the georeferencing until the quality metrics were within a certain threshold or a maximum amount of retries was reached. We implemented this for the pipeline that feature matches against the original neat drawings, but due to time constraints not on the pipeline that matches against the search box extracted reference map.

On the reference map, this effect can sometimes lead to even more extreme results because it has a bigger background image to find matches in. Below is a good and a bad example from feature matching against the reference map. The shearing is possible in the first place because we want to support reference maps in a different CRS than the input map. Therefore restricting the matches to a simple transformation of rotation translation scale is not possible.

Figure 54 Good result from georeferencing against the reference map.

Figure 55 Bad result from georeferencing against the reference map.

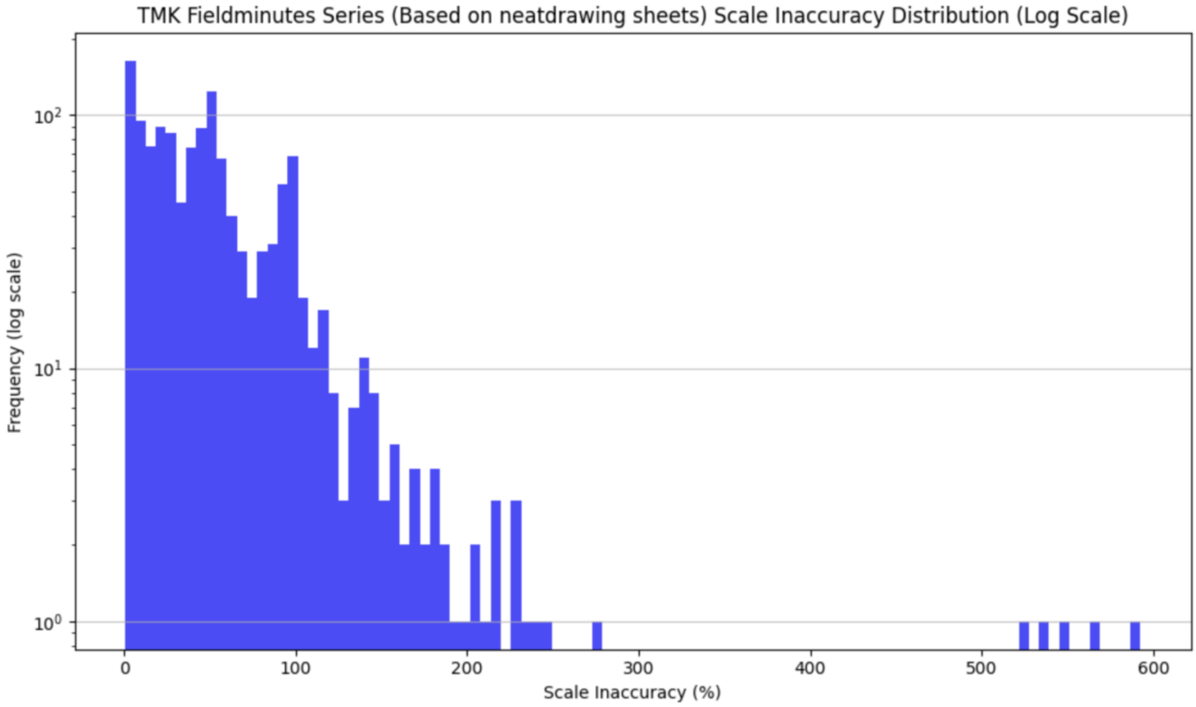


Figure 56 Scale inaccuracy of the TMK fieldminutes georeferenced based on individual neatdrawing sheets.

Minimum	0.78%
Maximum	592.57%
Median	44.37%
Mean	52.33%
Standard Deviation	52.31%

Table 7 Statistics for Field Series.

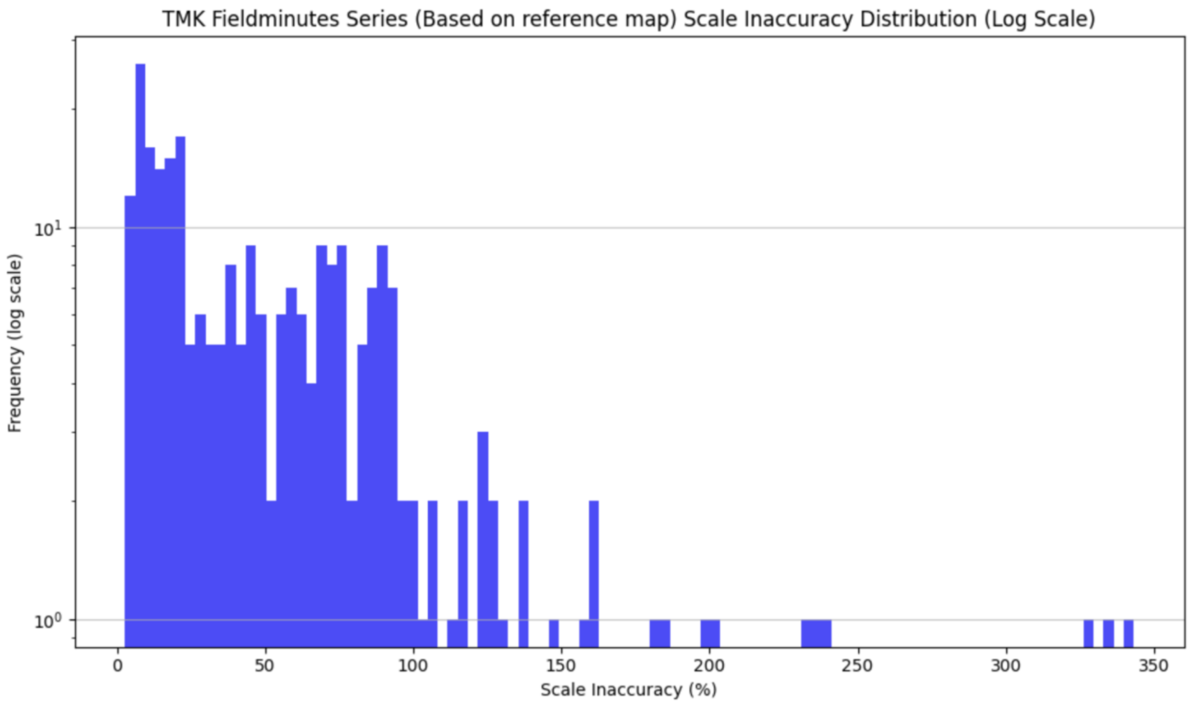


Figure 57 Scale inaccuracy of the TMK fieldminutes georeferenced based on searchbox and reference map.

Minimum	2.73%
Maximum	343.23%
Median	42.21%
Mean	54.58%
Standard Deviation	53.96%

Table 8 Statistics for Reference Map Field Series.

After adding a shear threshold of 0.02 and scale inaccuracy threshold of 10%, and making the pipeline retry up to 10 times, we get the following resulting map based on neat drawing sheets. Please note that the implementation of the pipeline is slow, especially with 10 retries. We ran the implementation for more than 12 hours and in that time it reached 60%, so the final result would be more complete.

In total, from the 355 field minute sheets, 219 were processed, 53 were within the threshold, and 166 sheets failed even after 10 retries. This gives an accepted rate of 24%.

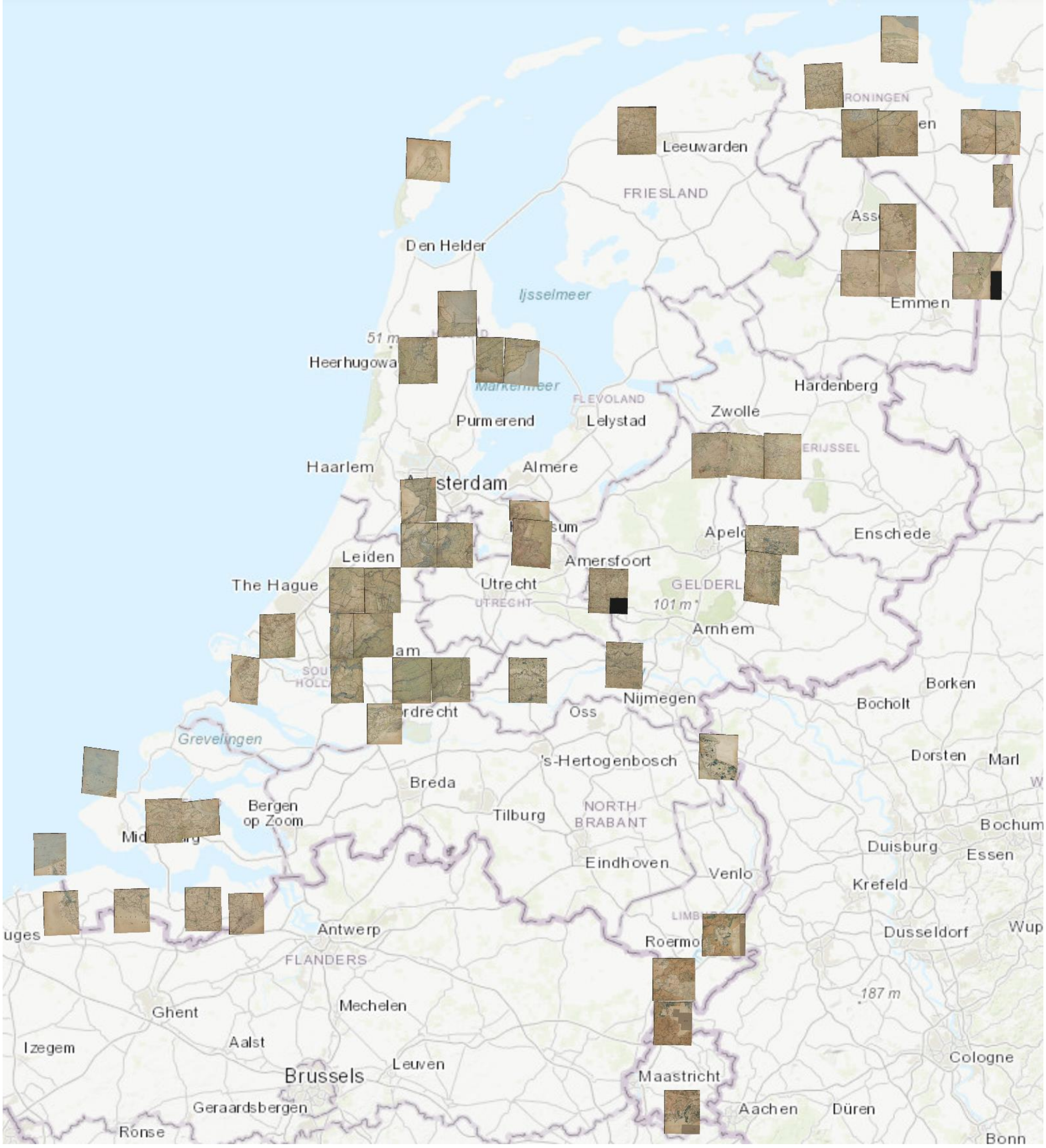


Figure 58 Resulting map after filtering by shear and scale threshold georeferenced based on the TMK neatdrawing sheets.

Then for the georeferencing based on the searchbox on a reference map (We used the IIIF manifest of the georeferenced TMK neat drawings also as reference map) we reached the following result after filtering based on the threshold. Also this pipeline took a long time to run, but since we didn't use any retries it was completed after a few hours. However, due to the geocoder sometimes not finding matches or finding multiple matches, still not all mapsheets could be processed. It processed 263 of the 355 sheets, and after thresholding 38 sheets were accepted. This gives an accepted rate of 14%.

Figure 59 Resulting map after filtering by shear and scale threshold georeferenced based on the searchbox extracted reference map of the full IIIF TMK neatdrawings.

Attempts were made to calculate residuals for the result, but due to a combination of time constraints, most sheets not being accepted purely based on our thresholds, and the stacking of inaccuracies by using the TMK fieldminute map from phase 1 as a basemap we decided the residuals would not give a good indication of the quality.

Visual inspection shows that a part of the accepted sheets got georeferenced accurately, see the figures below. However, many sheets that were within the quality thresholds are still significantly off. The accuracy of the sheets georeferenced based on the individual neatdrawing sheets is significantly higher than the ones georeferenced based on the searchbox and reference map.

Figure 60 In Allmaps.org viewer, a georeferenced fieldminute with background removed to show the accuracy relating to the background map.

Figure 61 Georeferenced results forming an interconnected sequence that lines up accurately.

The georeference annotations are available here:

- Based on neatdrawing sheets: all sheets including retries, sheets within threshold
- Based on reference map: all sheets, sheets within threshold

7. Discussion and Recommendations

7.1 Phase 1

For the process of corner detection it is important to develop an algorithm to handle the irregularities of the map sheets. More specifically it would be very helpful to detect the shape of the image and missing parts, since in many cases there are parts of sheets that are completely cut off. In addition, the applied approach only detects rectangular masks which is not always the case due to the reason explained previously. Also, it is believed that a combination of the two methods for detecting lines could be used to verify that lines really exist on the places they were detected as well as for the determination of the corners. Finally to improve even more the accuracy of the detected corners and lines it is recommended to combine with the already mentioned approaches a content aware method so that the algorithm can take into account where map content exists and try to fit the mask around it.

To improve the accuracy of georeferencing based on the sheet index, more attention has to be spent on the exceptions and irregularities. Especially concerning the dimensions of the map areas. This would avoid inappropriate stretching of images to fit the regular grid.

Although it is not the largest source of deviations, the accuracy of the transformation from Bonne to WGS84 could be improved by the introduction of a correction grid (red arrow in Figure 62). This new transformation would remove the whirly pattern of residuals in figure 62.

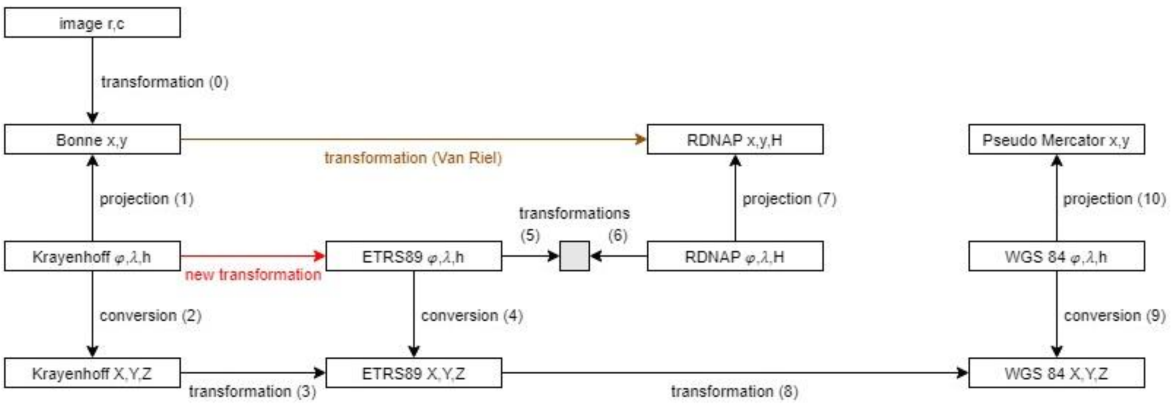


Figure 62 Proposed improvement for the transformation from Bonne by Jochem Lesparre.

The current approach georeferences the images of the combined neatdrawings, this leaves the images of individual sub-sheets without georeference. To include them, it should be possible for each of them to inherit a corner coordinate from the full sheet and the other corners can then be calculated based on the size and scale of the image.

In work package 3, the sheet number is extracted from the metadata of the annotation page. However, the use of the sheet number in the metadata is not IIIF compliant. The sheet number has to be stored as a string and in the metadata of the manifest. The algorithm should either pull the sheet number from this manifest metadata or the workflow needs an internal bookkeeping system to associate the sheet number with the right annotation page, pixel coordinates and geographic coordinates.

It could be useful to store the position of the sheet in the sheet index by its row, col coordinates in the grid. This would allow for easier retrieval and analysis of neighbouring sheets

7.2 Phase 2

Regarding work package 4 of our workflow, we still believe that if the time was more and therefore there was more space to implement and try out other ideas and approaches, the results could have potentially been different. Through experimentation and further investigation, the research could be enhanced in terms of quality. However it seems quite intriguing thinking that our findings could work as a way to ignite further ideas and be used in future research. Therefore, we would like to provide some of our ideas and approaches that we did not have the time to implement ourselves, in the hope that they might be useful to others.

First of all, based on our client's comments, the process of retrieving an image from the reference map for matching could be done using a different approach in order to avoid using the warped images as a base map for the Georeferencing and instead use the original unwarped ones. More specifically, it would involve determining which sheet contains the bounding box and finding the appropriate annotation. Secondly, transforming the bounding box coordinates to pixel coordinates using the Allmaps CLI would be needed. Then, the transformation command would be executed, which returns the pixel coordinates. The next step would be the construction of the IIIF Image request. This step involves setting the region and size parameters in the format "[x,y,w,h]/[w,]/0/default.jpg." The constructed URL for the request includes the service's address and the transformed pixel coordinates. The resulting URL can be used for matching, with the option to adjust the size by changing the "1000" parameter. Please note that if the bounding box is spread across multiple sheets, this method becomes more complex, as it may require performing these steps for each relevant sheet to piece together the complete image. So, we believe that this approach opens new possibilities for potentially more accurate results if someone tries to implement it in future research.

Secondly, another way that this package's results have to do with the geocoding of the toponyms. More specifically, the use of the gazetteer could as well be replaced with more complex and advanced ways of detection. For example, our original thought was to use Optical Character Recognition. Due to the openness limitations of Google Cloud's Vision API that we tried out and was the one that gave back the most usable results, the idea was not implemented as part of our pipeline. Additionally, the filtering of duplicates appearing in the geocoder, which right now in our implementation does not give back a result, based on geographic distance could enhance the quality of the end result and is worth improving.

Finally, regarding the process of retrieving an image from the reference map for matching, the current use of XYZ could be further extended by implementing even more methods, like the image extraction code described above, but also WMS in order to give more options and freedom to the user. Diversifying the methods for retrieving reference map images not only provides more options but also enhances the quality and scope of geospatial research. It empowers users to select the most suitable method for their specific needs, whether it is the XYZ approach or the more versatile WMS integration, ultimately contributing to a richer and more productive user experience.

The results from the full georeferencing pipeline in phase 2 are interesting but definitely not fully satisfactory. The rate of successfully georeferenced sheets is low, and the quality from even those deemed within the thresholds is varying. A lot of improvements can be made.

When it comes to the georeferencing based on the matches, a lot further research can be done in both finding more accurate matches. Our attempts at utilising traditional feature matching algorithms (i.e. ORB and FAST) failed, but successfully implementing this could lead to a lot more predictable results. The deep learning approach followed by Lan et al. and used by us seems to be very powerful, but also difficult to predict and reproduce consistently.

Part of this workflow could still be a segmentation of the image into feature maps, as discussed in work package 5. Optimally only the features that are most likely to be the same between time periods and different map style and scale would be extracted, like building footprints. The big challenge is how to extract these in a uniform way across many different maps, with different styles, drawing conventions, and colours (if not black and white). With more time, resources, and expertise this could be achieved and might lead to much better results.

Also more research can be done in the clustering of the results from the matching phase. Our project simply picks the biggest cluster based on affine transformed residuals, but more elaborate filtering could reveal more consistently usable results from the returned matches. We did not manage to reach any significant conclusions in the patterns of the matches, or a good way to filter them.

When it comes to specifically using the model from Lan et al., more research could be done on the impact of image resolution on the quality of the results. Our research showed more consistent results when lower resolution was used, but higher quality although inconsistent results with higher resolution.

Our specific implementation in code could be made a lot more efficient. Currently the neat drawings get separately downloaded and feature extraction using the deep learning model runs separately for every field minute. This is a lot of redundant processing. Either with caching or better code structure this could be prevented.

7.3 Visualisation

At the end of the project, we expected to have a standalone application that can be used for georeferencing automatically single or collections of images. The implementation of this would be based on the IIIF standard. To serve the purpose of ease-of-use the ultimate goal was to offer the application with a Graphic User Interface (GUI). This visual graphic environment was planned to be the last component of the project and the last Work Package. Its purpose is to give access to users through a graphical environment so that programming knowledge is not required. The application was thought to be hosted on a web server and the users would be able to access it through a web page. On this page, the users would be prompted to upload an image or a set of images they want to be georeferenced automatically using either IIIF manifests or to upload actual files. The result of the georeferencing process would be presented overlayed with OSM Basemap so that the users could have an overview of it before downloading the data. The data (i.e., the control points of the georeference) would be available in various formats such as IIIF annotation and text format in case the users wants to see them with a different application (e.g., QGIS). However, due to time limitations this idea was not implemented. We still believe that the user experience would be enhanced and that it is worth trying as a future project

8. Conclusions

This project sets out to implement an ambitious pipeline in order to complete the assignment given to us by our client. Phase 1 of the pipeline shows that the combination of historical sources and accurate transformations created by experts and historically recorded corner coordinates are highly suitable as a starting point for georeferencing a map series. Considering the corner detection and masking the content of the map, this phase shown that it is a highly challenging task. Big variations in map sheets make this process even more challenging. In the case of the Bonnebladen due to its predictable black background, corner detection produces consistent results, and the final georeferenced map has good and consistent residuals. The TMK neatdrawings however have high variation in shape, border, and background when it comes to its map edges. Our approach of generating masks based on the detected lines from the Hough Line Transformation proved to be very sensitive to noise. This resulted in accurate corner detection on some sheets, and inaccurate corner detection on others. The final georeferenced TMK neatdrawings therefore have inconsistent residuals.

Phase 2 proves to be an interesting proof of concept for automatically georeferencing historical map sheets when corner coordinates are not available. The implementation in this project is not yet suitable to be used for consistent and accurate georeferencing, but initial results do show that there is potential for its use case. The question is however if the accuracy could ever be as high as manually selected matches.

In conclusion, this project navigates the complexities of automating the georeferencing of historical maps by implementing multiple methods. The results are not perfect, but can either be used as a starting point for manually georeferencing or the automation could be further improved in future research.

9. Challenges and Limitations

During the project multiple challenges arose that had to be overcome. Most of them relate to the map sheets themselves. First of all, the different coordinate system used for the creation of the map series had to be reprojected to WGS84. For this reason a PROJ pipeline was created to acquire new coordinates. Also, the map series are entirely consistent. For example it is more difficult to stitch together sheets than contain islands and have gaps. In addition, since the map sheets are old, some of them have been damaged which made corner detection a problem way more difficult than initially estimated. Considering the limitations of the current project, the main one was the limited amount of time available to carry on all the steps. Furthermore, it was the first time for all the group members working with old maps and within the limited time it was not possible to acquire all the necessary knowledge to tackle the issues that arose more efficiently.

Acknowledgments

We would like to express our heartfelt gratitude to the individuals and organisations who have played an integral role in the successful completion of our project. Their continuous support, guidance, and contributions have been invaluable throughout this journey.

First and foremost, we extend our sincere appreciation to our esteemed supervisors Edward Verbree and Martijn Meijers, for their continuous guidance and mentorship. Their insights and expertise have been instrumental in shaping our project and ensuring its success.

We are also deeply grateful to our external partner, Jules Schoonman, and his colleague Bert Spaan from Allmaps, whose collaboration and expertise enriched our project with valuable perspectives from the field. Their dedication and willingness to work with us have been greatly appreciated.

We extend our thanks to Jochem Lesparre and his colleague, Daniël te Winkel, for their expertise and assistance with the CRS transformation. Their contributions have been essential in bringing our ideas to fruition.

We would like to acknowledge Leon van Wissen and the Globalise team for providing the script that generated the files crucial to our project. We are thankful for their open-source contribution and the accessibility it provided.

Special thanks go to Gees Boink for providing us with an enlightening tour of the national archive and sharing insights into the historical map series. His expertise and support have enriched our understanding of the subject matter.

Our gratitude extends to professors Bastiaan van Loenen, Hugo Ledoux and Edward Verbree for their coordination of the course, which laid the foundation for our project's success.

We also appreciate the professors of our Master's program for imparting their knowledge and expertise, which formed the basis for the work we have undertaken. Their dedication to education and research has been a source of inspiration.

We would like to acknowledge the Dutch Kadaster, Allmaps.org, the TU Delft Library, and the Dutch National Archive for their valuable resources and support, which were instrumental in our research.

This project would not have been possible without the collective efforts and contributions of the individuals and organisations mentioned above. We are truly grateful for their support and guidance throughout this endeavour.

References

Barget, M. (2022, January 2). *Reading historical maps with optical character recognition (OCR)*. INSULAE. https://insulae.hypotheses.org/485

Bladindeling van de nettekeningen. (n.d.). Kadaster. https://www.kadaster.nl/documents/1953498/2097376/Bladwijzer+Nettekeningen.pdf/9b552779-8791-5031-38b9-a9b805949b9e?t=1576079760786

Brovelli, M. A., Minghini, M., Giori, G., & Beretta, M. (2012). Web Geoservices and Ancient Cadastral Maps: The Web C.A.R.T.E. Project. *Transactions in Gis*, *16*(2), 125–142. https://doi.org/10.1111/j.1467-9671.2012.01311.x

Cantaloupe Image Server :: Home. (n.d.). https://cantaloupe-project.github.io/

Course browser searcher. (n.d.). https://studiegids.tudelft.nl/a101_displayCourse.do?course_id=64114

Dardavesis, I. (2022). *Indoor localisation and location tracking in semi-public buildings based on LiDAR point clouds and images of the ceilings*. TU Delft Repositories. http://resolver.tudelft.nl/uuid:40b823ed-9334-4a35-a69b-83c7ab393e5c

Dyouri, A. (2022). How To Use a PostgreSQL Database in a Flask Application. *DigitalOcean*. https://www.digitalocean.com/community/tutorials/how-to-use-a-postgresql-database-in-a-flask-application

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., & others. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise.* In kdd (Vol. 96, pp. 226–231).

gdaltransform — GDAL documentation. (n.d.). https://gdal.org/programs/gdaltransform.html

gdalwarp — GDAL documentation. (n.d.). https://gdal.org/programs/gdalwarp.html

Gede, M., & Varga, L. (2021). Automatic georeferencing of topographic map sheets using OpenCV and Tesseract. *Proceedings of the ICA*, *4*, 1–4. https://doi.org/10.5194/ica-proc-4-38-2021

Georeference extension. (n.d.). https://iiif.io/api/extension/georef/

Gold, C. (n.d.). SIMPLE TOPOLOGY GENERATION FROM SCANNED MAPS. *Cartogis.org*, 337–346. https://cartogis.org/docs/proceedings/archive/auto-carto-13/pdf/simple-topology-generation-from-scanned-maps.pdf

Gold, C. M., Nantel, J., & Yang, W. (1996). Outside-in: an alternative approach to forest map digitizing. *International Journal of Geographical Information Systems*, *10*(3), 291–310. https://doi.org/10.1080/02693799608902080

Hoitink, Y. (2023, June 9). *Using AI for transcriptions: VerledenTekst*. Dutch Genealogy. https://www.dutchgenealogy.nl/using-ai-for-transcriptions-verledentekst/

Home. (2023, July 10). IIIF. https://iiif.io/

How it works. (n.d.). https://iiif.io/get-started/how-iiif-works/

Jiao, C., Heitzler, M., & Hurni, L. (2022). *A fast and effective deep learning approach for road extraction from historical maps by automatically generating training data with symbol reconstruction.* International journal of applied earth observation and geoinformation, 113, 102980. https://doi.org/10.1016/j.jag.2022.102980

Király, G., Walz, U., Podobnikar, T., & Kokalj, Ž. (2008). Georeferencing of historical maps -- methods and experiences. *ResearchGate*. https://www.researchgate.net/publication/256307982_Georeferencing_of_historical_maps_--_methods_and_experiences

LAN Chaozhen, LU Wanjie, YU Junming, XU Qing. (2021) *Deep learning algorithm for feature matching of cross modality remote sensing images*. Acta Geodaetica et Cartographica Sinica, 2021, 50(2): 189-202.

Lesparre, J. (2023). Approximate transformation of geographic coordinates from Krayenhoff's triangulation and coordinates in the Bonne projection used for historical map series of the Netherlands. Kadaster, Netherlands Partnership Geodetic Infrastructure (NSGI).

Linden, J. A. van der. (1973). *Topographische en militaire kaart van het koningrijk der nederlanden*. Fibula-Van Dishoeck.

Lowe, D. G. (1999). *Object recognition from local scale-invariant features.* In Proceedings of the seventh IEEE international conference on computer vision (Vol. 2, pp. 1150–1157).

Luft, J., & Schiewe, J. (2021). Automatic content-based georeferencing of historical topographic maps. *Transactions in Gis*, *25*(6), 2888–2906. https://doi.org/10.1111/tgis.12794

Machines reading maps. (n.d.). Machines Reading Maps. https://machines-reading-maps.github.io/

Meijers, M. (2023) synthesis 2023. http://gist.bk.tudelft.nl/~bmmeijers/volatile/2023/synthesis/

Niemeijer, F. (2021). Godcreated the earth but the dutch made their own country: The saying unravelled. *Vakblad Vitruvius*.

NYPL Internal Automated Cropping with Computer Vision. (n.d.). Slides. https://slides.com/ericshows/automated-cropping-with-computer-vision#/1

OpenCV: Introduction to SIFT (Scale-Invariant Feature Transform). (n.d.). https://docs.opencv.org/4.x/da/df5/tutorial_py_sift_intro.html

OpenCV: Template Matching. (n.d.). https://docs.opencv.org/4.x/d4/dc6/tutorial_py_template_matching.html

OpenCV: ORB (Oriented FAST and Rotated BRIEF). (n.d.). https://docs.opencv.org/3.4/d1/d89/tutorial_py_orb.html

PROJ — PROJ 9.3.0 documentation. (n.d.). https://proj.org/en/9.3/

Schoonman, J. (2023, October 25). Bonnebladen. Observable. https://observablehq.com/@allmaps/bonnebladen

Rublee, E., Rabaud, V., Konolige, K. & Bradski, G. R. (2011). *ORB: An efficient alternative to SIFT or SURF.*. In D. N. Metaxas, L. Quan, A. Sanfeliu & L. V. Gool (eds.), ICCV (p./pp. 2564-2571), : IEEE Computer Society. ISBN: 978-1-4577-1101-5

Spaan, B. (2020, September 14). *gdaltransform*. Observable. https://observablehq.com/@bertspaan/gdaltransformSpaan, B. (2023, May 15). *Georeferencing a map by transcribing toponyms*. Observable. https://observablehq.com/@allmaps/georeferencing-a-map-by-transcribing-toponyms

synthesis - Wiktionary, the free dictionary. (n.d.). https://en.wiktionary.org/wiki/synthesis

Topographisch Bureau. (1861). *Meetkunstige beschrijving van het Koningrijk der Nederlanden, bevattende de getallenwaarden, gebruikt bij de zamenstelling van de topografische en militaire kaart van het rijk*. https://play.google.com/store/books/details?id=CTxNAAAAcAAJ&rdid=book-CTxNAAAAcAAJ&rdot=1

Touya, G., Coupé, A., Jollec, J. L., Dorie, O., & Fuchs, F. (2013). Conflation optimized by least squares to maintain geographic shapes. *ISPRS International Journal of Geo-information*, *2*(3), 621–644. https://doi.org/10.3390/ijgi2030621

Uhl, J., Leyk, S., Chiang, Y., & Knoblock, C. A. (2022). *Towards the automated large-scale reconstruction of past road networks from historical maps.* Computers, Environment and Urban Systems, 94, 101794. https://doi.org/10.1016/j.compenvurbsys.2022.101794

Van Riel, H. F. (1924). Topografische kaart en Rijksdriehoeksmeting. -, 48–56.

Welcome to Flask — Flask Documentation (2.3.x). (n.d.).

https://flask.palletsprojects.com/en/2.3.x/

Appendices

Appendix 1: Project organisation

This section includes the description of the organisation of the project, as well as the responsibilities, division of tasks, planning and time management.

Responsibilities within the project

Role	Task	Team member
Coordinator	Responsible for communication with external partners. Organisation of meetings. Time management and planning.	Eirini Tsipa
Data manager	Oversees the available data as well the produced ones. Ensures effective data management and storage.	Rianne Aalders
Technical manager	Organises the structure of the project code-wise. Allocates code responsibilities between the members.	Oliver Post
Quality manager	Ensures quality of deliverables: - reports - code	Giorgos Iliopoulos

Table 2: The team division of main responsibilities

Major contribution and responsibilities of each supervisor to the project.

Role	Name of supervisor	Task
Supervisor 1	Edward Verbree	Provides academic oversight and coaching throughout the project. Offers real-world context or resources relevant to the project.
Supervisor 2	Martijn Meijers	Provides academic oversight and coaching throughout the project. Offers real-world context or resources relevant to the project.

External Client	Jules Schoonman	Defines the project scope, provides guidance, and supports the project to align with his needs and objectives. Offers real-world context, data, or resources relevant to the project.

Table 3: Main responsibilities of supervisors

Structure and Team dynamics

Reaching the endpoint of our project we can say that time and deeds have made clear how our team's dynamics were demonstrated in our everyday work and interaction. Our team consisted of 4 highly motivated and hardworking team players. Our strong points determined the quality of our work. Balance was a characteristic of our team, positively affecting our collaboration. The abundance of work types, points of view, and characters made our collaboration intriguing and the final result successful.

As a team, we decided to work in a hybrid manner, combining both in-person and remote work. Most of the time, we had team meetings and worked in a common space at Geolab, utilizing tools such as big screens and whiteboards for effective teamwork. However, if the work allocation and project requirements allowed for it, we would work remotely. This provided us with the flexibility to focus on personal growth and responsibilities while contributing to the team's common goal. During our physical team meetings, we engaged in planning, work allocation, brainstorming, in-depth discussions of ideas and approaches, and problem-solving. Decisions were made collectively, with each team member contributing their perspective based on their assigned role. This ensured clear division of responsibilities. Our main objective was to apply our Geomatics knowledge and synthesize it into a successful final result that benefited everyone's development. Therefore, all team members actively participated in all tasks to the best of their abilities. Our collaboration with our supervisors and external partner was characterized by effective communication and guidance. From the beginning, we had a clear understanding of the partner's requirements and the datasets we were working with. We received a workshop on the Allmaps platform, which provided valuable insights. Additionally, we had the opportunity to visit the National Archive in The Hague, which enhanced our understanding of the data and the structure of the old maps we were working with. In addition to our team meetings, we held weekly physical meetings with our supervisors and the external partner. These meetings were highly valuable, and we took notes of the feedback and remarks provided for future reference and improvement. Email communication proved to be an efficient channel for scheduling meetings, seeking guidance, and receiving papers and links related to our work. The feedback and scientific work shared by our supervisors and partner inspired us and helped structure our workflow. Organization was a key factor in ensuring a smooth workflow for our team. From the beginning, we utilized various tools to facilitate collaborative work and provide common access to all relevant materials. Google Drive was used for data storage, deliverables, papers, and diagrams. Individual tasks were distributed and tracked using Google Keep. Diagrams were created using diagrams.net, and reports were composed on Google Docs. This approach ensured an efficient and structured workflow. We continuously gained experience in using these tools and explored features and extensions to enhance the quality of our work. To keep track of our meetings, work progress, and previously implemented methods and ideas, we maintained a project diary. This allowed us to reflect on past discussions and ensure continuity in our work.

Diary			
Date	**Topics**	**Meeting**	**w/Supervisors**
04/09	Introduction to the course, Meeting with the external partner, brainstorming on the project.		✓
05/09	Composed the 150-word summary of our project, and creation of the Project's thumbnail and asked for feedback from our supervisors.		✗
06/09	Introductory workshop on the Allmaps platform by Jules Schoonman.		✓
08/009	Received feedback from our supervisors and finalised our summary.	-	-
12/09	Listed questions to ask Gijs Boink from the National Archive, and brainstormed 3 initial ideas on approaching our topic.		✗
13/09	Visit to the National Archive in the Hague, Meeting with Supervisors about our ideas and feedback. The map-matching approach idea was conceived that day. Finding a balance between the expectations of the external partner and the research potential.		✓
14/09	Worked on the composition of our PID and created our team's logo.		✗
18/09	Finalised and submitted our PID and logo. Assigned roles to our team members.		✗
20/09	Feedback on PID and suggestions for our mid-term report.		✓
22/09	Brainstormed on the different approaches. Put our ideas into diagrams. Reached a final decision on the approach that we will follow. Assigned tasks regarding the Mid-term report.		✗
25/09	Assigned Work Packages to the team members. Worked on the Mid-term report.		✗
26/09	• Handed in the midterm report, • Team meeting: created the ppt for the midterm presentation.		✗
27/09	• Midterm presentation, Meeting with the supervisors: Feedback on our		✓

	presentation. ● Handed in our individual reflections.		
28/09	Sent email to Jochem Lesparre		✗
04/10	Meeting with the Supervisors: focus on IIIF, Jochem's mail Team meeting: specifics about the data input and output of our work packages and the quality checks		✓✗
18/10	Worked on corner detection and feature matching		✗
23/10	Meeting on IIIF and Proj pipeline		✓
24/10	Meeting about corner detection and proj		✗
25/10	Meeting about Phase 1 and Phase 2 progress		✓
31/10	Team meeting: updates on our progress, final touches		✗
02/11	Final meeting with supervisors		✓
03/11	Work on the final report and submission		✗

Diary of meetings and progress

Appendix 2: Initial proposals and methodologies

We developed four methodologies to accelerate georeferencing using automation. During our orientation phase, we investigated the potential and feasibility of each method. Below, each method is briefly described and a summary is given on why it was accepted/rejected.

Method: Map stitcher (Rejected)

Develop an algorithm that can stitch together the individual map sheets into one big map. To decrease computation time, toponyms are used for an initial guess of the map sheet position. When the whole map is stitched together, it is manually georeferenced. Finally, the georeferencing of the individual map sheets is calculated automatically based on their positions.

Strengths

- Requires **no additional input** besides the map series itself
- **Automates many steps** of the georeferencing process, only requiring manual selection of three control points
- Can be applied to **any map series**

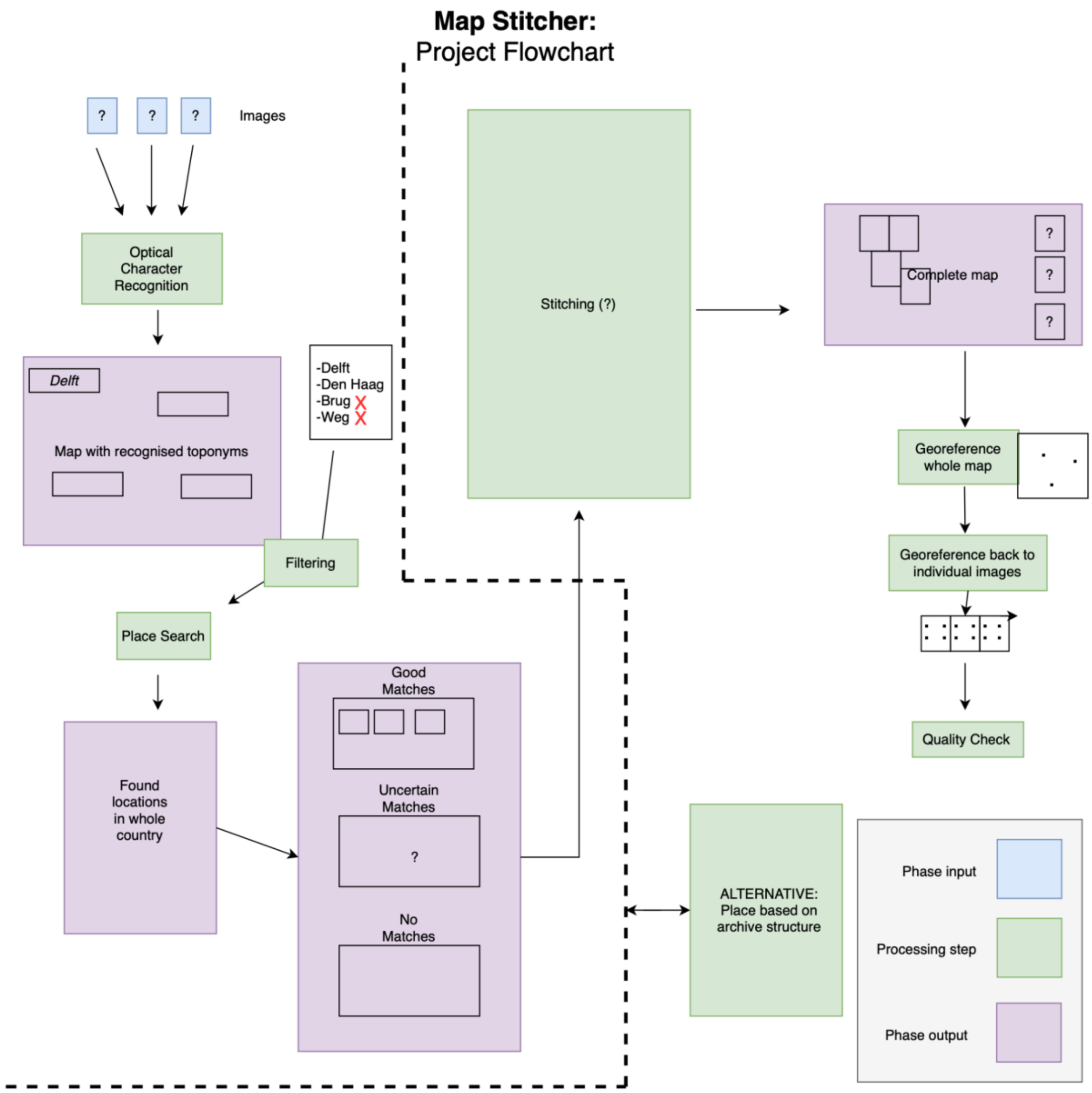
Weaknesses

- Many **difficult** steps to implement
- Requires a lot of **processing power**
- Does not work for multiple **different scales or projections**
- Steps are **highly dependent** on each other, and multiple steps involve a lot of **threats**

The prototype geocoder required an advanced optical character recognition system. Most OCR modules would not detect any usable toponyms in the scan. The Google Cloud Vision API yielded on average between zero and five usable toponyms per scan. This formed the first major weakness of this approach, as many scans would be impossible to place with any level of certainty with no or only one usable match. An alternative would be to use the archive structure, as it has placename information per sheet, but this would eliminate the main strength of this approach (that it needs no additional input).

Another major weakness of this approach is the stitching phase. Traditional image stitching algorithms require overlap, which is not present in most map series scans. Stitching based on edge matching becomes extra complicated due to edge damage on the scans. We predict that due to the nature of the TMK fieldminutes, with gaps, overlaps, and dangling island pieces, stitching could never fully work for all scans.

Due to these two major weaknesses, we decided to reject this method.



Sketch flowchart of the map stitcher method. Starting from the top left, following the arrows it goes through a toponym searching phase (with as an alternative an archive structure method), a stitching phase, and a georeference phase.

Method: Exact corner coordinates (Rejected)

Using historical sources, we attempt to find the exact coordinates of the corners of each map sheet. Then, we transform these coordinates from the historical CRS to WGS84. We match the individual sheets to these coordinates and apply the transformation. For the “veldminuten” we will either have to manually place them, or find a way to automate their placement.

Strengths	Weaknesses
- Presumably the **most accurate** method, due to using the original coordinates	- Corner coordinate **data might be missing** for many map series - Needs manual processing for each map series, **not very automated**

We concluded that this method can achieve the requirements of the stakeholders for the neat drawings of the TMK and for the Bonnebladen, but not for the fieldminutes of the TMK as there are no corner coordinates available for these scans. Instructions are provided to manually place the fieldminutes, but this would involve large amounts of manual work. Furthermore, we determined that this method does not involve significant Geomatics research value, as the only automation used is very straightforward and not innovative.

For these reasons, we rejected this method as a standalone solution, but we used it as a starting point for method “Map matcher”.

Method: Toponym georeferencer (Rejected)

We match toponyms on the individual sheets with their real-world coordinates. Based on the matches on all sheets, we figure out an average transformation for the whole map. Optionally we can look into offset correction that would be caused by toponyms consistently being placed with an offset from their real location.

Strengths	Weaknesses
- **No additional input** is needed. - Applicable to **many map series**	- **Not very accurate** where the linked points are in relation to their toponym. - **Difficult to recognize sufficient toponyms** on each sheet? Due to language, font, and the amount of names in uninhabited areas.

Based on our preliminary testing of a toponym extraction algorithm for method “Map stitcher”, we determined that there are not enough matches on each sheet to use this method. Also, the fact that the method is by nature not very accurate means it will not satisfy the requirements of the client. Therefore, we rejected this method.

Appendix 3: Work package division

(For method Map Matcher)

Phase 1

- **WP0: Feasibility study**
  - Lead team member: Oliver
  - Team members involved: All
  - Necessary input: None (sources)
  - Identified threats:
  - Tasks:
    - Literature review
    - Evaluate possible solutions
- **WP1: Finding and processing the corner coordinates of both the TMK and the Bonnebladen**
  - Lead team member: Rianne
  - Team members involved: All
  - Necessary input: None (sources)
  - Identified threats:
    - Unable to find CRS transformation or incorrect transformation
    - Unable to find corner coordinates of Bonnebladen
    - Manual transcribing errors
  - Tasks:
    - Find source for Bonnebladen corner coordinates. (in collaboration with Coordinator for contact with outside sources). No sources found so far except Edwards mention of Jochem
    - Decide on how to store coordinates digitally (in collaboration with Data Manager)
    - Divide different sources across all team members for manual digitizing
    - Write a simple quality check for format mistakes while digitizing (in collaboration with Quality Manager)
- **WP2: Corner detection**
  - Lead team member: Rianne
  - Team members involved: Rianne, Giorgos
  - Necessary input: TMK and Bonnebladen scans
  - Identified threats:
    - Non-rectangular sheets

        - Rounded corners
        - Different types of scans, black background, white border, etc…
    - Tasks:
        - Decide on output of detected corners. Will it be stored in image, will it be an annotation file, will it immediately go to the next part of the code and not be stored (in collaboration with Technical Manager and Data Manager)
        - Decide if tool will also do masking or cropping
        - Research existing methods. Are they libraries that can do it out of the box? Are there papers about methods?
        - Determine how to check the quality of the results (in collaboration with Quality Manager)
- **WP3: Automated writing of georeferencing of TMK and Bonnebladen to IIIF**
    - Lead team member: Giorgos
    - Team members involved: Giorgos, Rianne
    - Necessary input: Results of WP1 and WP2. To prevent blocking, the leads of WP1 and WP2 can make example outputs
    - Identified threats:
        - Incorrect input from WP1 or WP2
    - Tasks:
        - Contact Jules about exact expectations of IIIF writing (in collaboration with Coordinator). Also, check if Gijs wants the coordinates in another format too for future scanning?
        - Research into the IIIF standard and the IIIF georeferencing extension
        - Decide on how to implement. How does this fit into the pipeline? (in collaboration with Technical Manager)
        - Decide on how to store/visualise the outputs (in collaboration with Data Manager)
        - Write the code
        - Think of quality check to see if the full set of images form a complete map. Are there missing parts? Overlaps? Are there mistakes? (in collaboration with Quality Manager)
        - Does this output correspond with the stakeholder's expectations? (in collaboration with Coordinator)

Phase 2

- **WP4: Search box determination and processing**
    - Lead team member: Eirini
    - Team members involved: Eirini, Rianne
    - Necessary input: Fieldminute scans and archive structure and a georeferenced reference map (preferably output of WP3)
    - Identified threats:
        - Placenames that occur multiple times in the Netherlands
        - Old way of writing placenames
        - Reference map in different CRS than WGS84
    - Tasks:
        - Decide on how the input of this part of the pipeline has to be structured. How does the user input their scans with location hint? How does the georeferenced base map have to be supplied? What if the images are online only? Is this what the stakeholders would want? (in collaboration with Technical Manager and Data Manager and Coordinator)
        - Get the required images (or image links) from the archive and their location hint into the decided on input format
        - Determine what the search box requirements are
        - Decide on a strategy to match the location hints to a real world location (how to limit the search so it doesn't find hits in other countries?) (What if the map spans an area that is now multiple countries?
        - Decide on a metric to determine how confident we are the location we matched to the location hint is accurate (in collaboration with Quality Manager)
        - Write code to place the search box and extract the search box from the reference map
        - Decide on what the output of this work package is? Is it an intermediary file? Will it go straight to the next part of the code? (in collaboration with Technical Manager and Data Manager)
- **WP5: Splitting of map into feature maps (like roads, water, etc.)**
    - Lead team member: Oliver
    - Team members involved: Rianne, Oliver, Eirini
    - Necessary input: Output of WP4
    - Identified threats:
        - Many different map styles

        - Maps might be damaged
        - Different color coding
        - Many possibilities for incorrect processing
        - If the output of the splitting of one map style is very different to that of another map style, the matching will not work
        - Can be computationally expensive
    - Tasks:
        - Research together with WP6 lead if this is indeed the best strategy. Should we split them up? Should it be vectors or raster images?
        - Research if there are libraries/papers out there for doing this
        - Collaborate with WP6 lead about what these maps should look like/contain for the best matching performance. What categories should be used? Both from the perspective of how well they can be extracted and how useful/suitable they are for matching
        - Since this is the most critical part of the 2nd phase, together with WP6 lead discuss how it could be scheduled and implemented so we will have a sufficiently working version for sure, but also have space to make an excellent version (in collaboration with Coordinator)
        - Think of an automated way to check the quality of the output (in collaboration with Quality Manager)
        - Decide on what the output of this work package is? Is it an intermediary file? Will it go straight to the next part of the code? (in collaboration with Technical Manager and Data Manager)
        - Determine how openstreetmaps could be split into feature maps of the same nature

- **WP6: Finding matches between input map and reference map (from feature maps)**
    - Lead team member: Oliver
    - Team members involved: Oliver, Giorgos, Eirini
    - Necessary input: Output of WP5
    - Identified threats:
        - Incorrect matches
        - Too few correct matches
        - Can be computationally expensive
    - Tasks:
        - Research into libaries/papers on how this matching can be performed.

        - Discuss with WP5 lead on how to approach the matching and what output from WP5 is necessary
        - Already think if this method could also work with maps that are vastly different/modern openstreetmaps
        - If matching is working, determine a way to filter out only the correct matches
        - Decide how this code fits into the pipeline (in collaboration with Technical Manager)
        - Since this is the most critical part of the 2nd phase, together with WP5 lead discuss how it could be scheduled and implemented so we will have a sufficiently working version for sure, but also have space to make an excellent version (in collaboration with Coordinator)
        - Determine a metric that indicates the confidence that the map is correctly placed (in collaboration with Quality Manager)
        - Perhaps manually make a test dataset to see how well it’s performing
        - If time allows, test if it can be made to work with openstreetmaps as reference map

- **WP7: Georeferencing based on matches and output to IIIF**
    - Lead team member: Eirini
    - Team members involved: Eirini, Giorgos, Oliver
    - Necessary input: Output of WP6
    - Identified threats:
        - None so far
    - Tasks:
        - Decide on how to go from matches to georeferencing. Will all matches be stored as georeferencing points? Will they be converted to corner points?e how this code fits into the pipeline (in collaboration with Technical Manager)
        - Decide on how to store/visualise the outputs (in collaboration with Data Manager)
        - Decide on final quality check (in collaboration with Quality Manager)
        - Does this output correspond with the stakeholder’s expectations? (in collaboration with Coordinator)

Phase 3

- **WP8: Web interface**
    - Lead team member: Giorgos
    - Team members involved: Giorgos, Oliver
    - Necessary input: Code skeleton to base buttons and visuals on
    - Identified threats:
        - Storage/performance needs on the server
        - How to make it user friendly
        - Might take a lot of time to make or slow down development because research is needed to make it web compatible
    - Tasks:
        - Should this be discussed with Jules if they want to offer it on their website? (in collaboration with Coordinator)
        - Determine how the code should work to be able to run on a website
        - Collaborate with Technical Manager and Data Manager to make sure all formats and code decisions will work on the web
        - Decide on a framework?
        - What do the stakeholders expect from the web interface? (in collaboration with Coordinator)
        - Design the website interface (with feedback from Quality Manager)

Appendix 4: Overview of relevant theory/concepts

This subsection introduces the theoretical background that is relevant to the project. The literature and other sources are organised based on the work package they relate to, in other words the work package they will be used in, either as literature reference or material for code-wise implementation. To help the reader, the subsection is divided according to the work packages (i.e., WP1 - WP8). In addition, the resources which were studied at the initiation phase of the project are included under "Sources not directly related to the WPs", while information about the map series is introduced at the end of the section.

WP1: Finding and processing the corner coordinates of both the TMK and the Bonnebladen

(Topographisch Bureau, 1861) Book by the makers of the TMK about their calculations, contains tables with coordinates of place names and map corners

Linden (1973) Book about the making of the TMK by a historian

There is FOSS software to track and transform CRS among them Proj (*PROJ — PROJ 9.3.0 Documentation*, n.d.) and GDAL. GDAL has the option to both transform (*Gdaltransform — GDAL Documentation*, n.d.) and warp (*Gdalwarp — GDAL Documentation*, n.d.). Spaan (2020) describes how gdaltransform can be used in the browser to georeference a historic map.

WP2: Corner detection

Gede and Varga (2021) Describes a pipeline for detecting the corners of map content, recognising the sheet identifier and then using the sheet corners as GCP to georeference the map sheets.

WP3: Automated writing of georeferencing of TMK and Bonnebladen to IIIF

IIIF stands for International Image Interoperability Framework and "is a set of open standards for delivering high-quality, attributed digital objects online at scale. It's also an international community developing and implementing the IIIF APIs. IIIF is backed by a consortium of leading cultural institutions." (*Home*, 2023). An introduction and tutorial to working with IIIF can be found at https://training.iiif.io/intro-to-iiif/

"Cantaloupe is an open-source dynamic image server for on-demand generation of derivatives of high-resolution source images. With available operations including cropping, scaling, and rotation, it can support deep-zooming image viewers, as well as on-the-fly thumbnail generation." (*Cantaloupe Image Server :: Home*, n.d.) It is compliant with the IIIF image API.

Allmaps links to the map images on an IIIF server and on the client side computes the warping and placement of the map image on OpenStreetMap from reference points. This transformation is then stored as a Georeference Annotation (GA), in GeoJSON format, to the original image.

The GA expresses the image coordinates in WGS84 which in the case of many historic maps requires a transformation from the old CRS. There is no space in the GA to include information on the original CRS or projection. Instead this information can be recorded in the metadata of the IIIF Manifest or “in a machine readable format referenced through the seeAlso property.” (*Georeference Extension*, n.d.).

WP4: Search box determination and processing

Work package 4 requires more information on the IIIF standard and the data structure we are going to use.

WP5: Splitting of map into feature maps (like roads, water, etc.)

Gold (n.d.) proposes a method for improved line extraction to generate simple topology from scanned maps. “Instead of concentrating on the black pixels (linework) one may emphasise the white pixels (polygonal areas) and build the relationships from these. The process has three steps. Firstly a set of white pixels are selected, at a user-specified distance from the black linework, and these are given a polygon label on the basis of a flood-fill algorithm that scans all connected white pixels. Secondly these selected pixels are used as data points to generate the standard Euclidean point Voronoi diagram. Thirdly, this structure is scanned to extract only those Voronoi boundaries between pixels having different polygon labels. The result of this operation is a set of vector chains or arcs that are guaranteed to separate regions of white space, and are guaranteed to connect precisely at nodes.” There was a preprocessing step where “The original input was retraced to eliminate roads, contour lines, etc. that had been superimposed on the same map”.

C. M. Gold et al. (1996) Also uses Voronoi to extract boundaries. It is the precursor to the paper above.

Luft and Schiewe (2021) Automated historic map georeferencing through the extraction of water features and matching these with OSM. This method might be inaccurate for Dutch maps because of the amount of (man-made) changes in water features over the past centuries, (Niemeijer, 2021).

WP6: Finding matches between input map and reference map (from feature maps)

(*NYPL Internal Automated Cropping With Computer Vision*, n.d.) uses computer vision techniques “to compare a lower resolution cropped image with the uncropped master tif file” Their method detects features using SIFT and FLANN to match. Create a list of good matched features then draw lines between each set of matched features. Also uses SIFT and FLANN to detect boundary of image and crop. Compares difference between detected corner points of image and the image bounding box to adjust rotation. RMSE of pixel by pixel differences to test accuracy.

Dardavesis (2022) Has used different feature description, detection and matching techniques (ORB, SIFT, FLANN) to match individual non overlapping images to a reference image. “PhotoMatch is an open-source software that delves into automatic feature matching, enabling the implementation and evaluation of different feature detection, description and matching techniques. Aside from these, it also includes image enhancement methods to improve the quality of an image dataset.” Images were changed to grayscale before processing.

(*OpenCV: Introduction to SIFT (Scale-Invariant Feature Transform)*, n.d.) Describes the workings of the SIFT algorithm and how to use it.

WP7: Georeferencing based on matches and output to IIIF

For work package 7, part of the sources and code can be reused to write georeferencing annotations to IIIF format.

WP8: Web interface

Flask is a tool to work on web applications with Python (*Welcome to Flask — Flask Documentation (2.3.x)*, n.d.). Their website has information on how to use it. Dyouri (2022) writes about how to combine Flask with a postgreSQL database.

Sources not directly related to the WPs

Touya et al. (2013) proposes a method to warp geographic data while preserving shape features using least squares."Conflation can generally be divided into three steps: data matching, geometry deformations and attribute merging. Each step is a challenging task, but the paper only focuses on the geometry deformation task, once data matching has succeeded." This paper mentions several previous studies focussing on data matching: "Data matching can be achieved by theme specific methods. For instance, as coastlines are very sinuous, similarity, and one-to-one matching, can be achieved using Fréchet distance [9]. In networks, like road or river networks, topology, i.e., how network elements are connected to each other, allows matching networks with quite different level of details [10]""efficient generic matching methods have been proposed in the literature. The method proposed by Walter and Fritsch [5] examines the similarity of features in each pair, and statistical considerations globally determine the best set of matching pairs. The data matching process described by Samal et al. [11] proposes to use a large number of similarity measures and averages the similarity scores to find the best match. As data matching relies on several different criteria, multiple criteria analysis techniques can be applied to decide if features can be matched [12]: geometry, attribute, and topology criteria can be mixed to match features with a very high accuracy. If the matches are considered globally, finding a matching that maximizes the similarities can also be considered as an optimization problem [13]". Their new method for warping: "[this paper] focuses on the development of an alternative to rubber sheeting that allows preserving the shapes of conflated features, whether the features are man-made with straight shapes and right angles, or natural with smooth curved shapes." This method utilises constraints that are turned into equations which are part of a system that is solved using least squares. There are constraints to preserve shape, conflate data and maintain data consistency.

Spaan (2023) describes a workflow to georeference an old map image using manually chosen placename points and a historical geocoder.

Barget (2022) Describes the use of the optical character recognition service "Transkribus Lite" to extract text from historic maps.

Hoitink (2023) Describes using the Loghi AI model to extract text from scans of historic dutch records. The model is trained on Dutch manuscripts from the 1600s and 1700s.

(*Machines Reading Maps*, n.d.) Is a project to extract machine readable text from (historic) maps. More information on how to use the developed mapKurator can be found here: https://knowledge-computing.github.io/mapkurator-doc/#/docs/introduction

Brovelli et al. (2012) Worked on georeferencing and publishing Italian map series with free open source software. They used GIMP to mosaic maps spanning several sheets into a single image. GCP and CP pairs with both a polynomial “a global approximate interpolator which links ground coordinates (*X*,*Y*) to image coordinates (*x*,*y*), for a generic order *m”* and a thin plate spline “a local exact technique that minimises the interpolating surface curvature” method to warp the image with a combination of QGIS and PCI Geomatica OrthoEngine (http://www.pcigeomatics.com). Thin plate spline causes distortions in areas away from GCPs. They then compared the RMSE from different maps and orders of polynomials to determine accuracy.

Reproducibility

The Association of Geographic Information Laboratories in Europe (AGILE) has guidelines for published research to increase reproducibility in the GIS domain. These can be found here: https://osf.io/numa5

Other

Király et al. (2008) Did a comparison of different georeferencing methods for different datasets.

Appendix 5: Tasks and Work Allocation

Deliverables

Table Deliverables list

Overview of the deliverables and their deadlines

No.	Deliverable Title	WP	Lead Team Member	Dissemination level	Due date
D1	PID	WP	Rianne Aalders	Private	18/09/23
D2	Midterm presentation	WP	Eirini Tsipa	Private	27/09/23
D3	Final report	WP	Oliver Post	Public	03/11/23
D4	Presentation	WP	Giorgos Ilioopoulos	Public	09/11/23

Gantt chart

The work packages of the selected approach (i.e., Map Matcher) as introduced in section 2 are organised in the following Gantt Chart so that the working group can keep track of the progress and set priorities. To keep the Gantt Chart short only the work packages are included. To see the tasks of each work package, the reader may refer to the Appendix section following below.

			Week number										
		Lead team member	35	36	37	38	39	40	41	42	43	44	45
WP0	**Feasibility study**	**Oliver**											
T0.1	Research into the feasibility of the different methods						25 Sep						
WP1	**Finding and processing the corner coordinates of both the TMK and the Bonnebladen**	**Rianne**											
WP2	**Corner detection**	**Rianne**											
WP3	**Automated writing of georeferencing of TMK and Bonnebladen to IIIF**	**Giorgos**											
WP4	**Search box determination and processing**	**Eirini**											
WP5	**Splitting of map into feature maps (like roads, water, etc.)**	**Oliver**											
										MVP			
WP6	**Finding matches between input map and reference map (from feature maps)**	**Oliver**											
										MVP			
WP7	**Georeferencing based on matches and output to IIIF**	**Eirini**											
WP8	**Web interface**	**Giorgos**											
D7.1	Reporting						26 Sep					3 Nov	9 Nov
T7.2	Meetings												

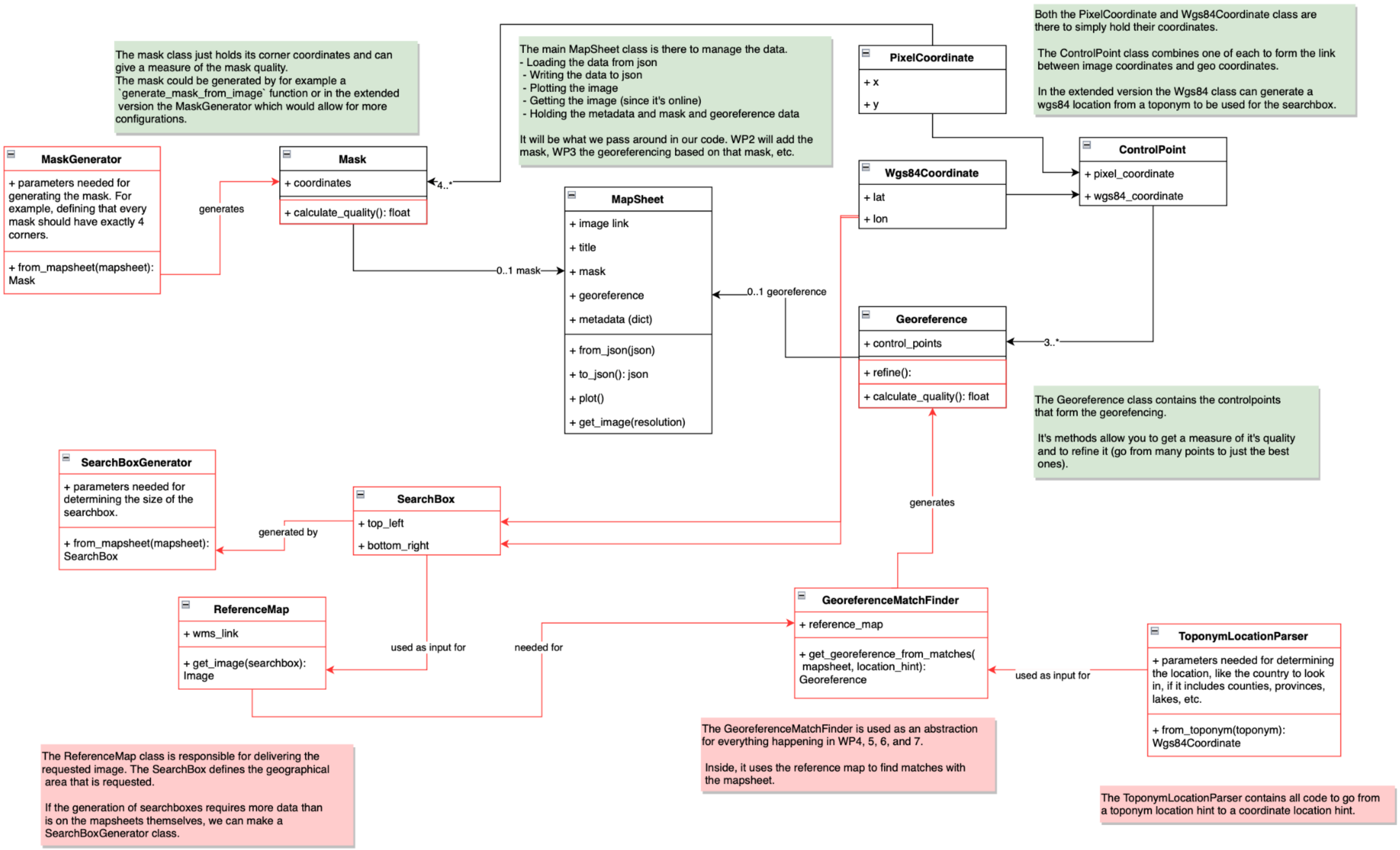
Deliverables

Activity

Internal Deadline

Work Package (WP) / Deliverable (D)/ Task (T) / Minimum Viable Product (MVP)

Appendix 6: Initial UML diagram



Code structure has largely stayed the same, but methods and attributes have changed since this UML diagram was created.

Appendix 7: Dissemination and communication

During the first phases, we will focus on the project itself. If in the final phases, it becomes apparent that we achieved the interoperability we were aiming for it becomes essential to share our results for further research. Possible channels for sharing our results are:

- De Hollandse Cirkel - A Dutch foundation for promoting information about the field of geodesy.
- Publishing a paper in Tijdschrift Geo Info - A Dutch magazine about a broad range of topics within the geoinformation sphere.
- Present for “Gebruikersgroep BRT Basisregistratie Topografie” - A meeting that’s held four times a year for users of the Dutch BRT maps.

Appendix 8: Table of corner detection accuracy measurement

sheetnumber	distance0	distance1	distance2	distance3
5a4584a2-e824-4972-8ec1-a73c61e73710.jp2	17	23	391	409
3ed5eef4-6549-4d2f-b2f4-2eb9d9dc802e.jp2	224	4876	5786	3073
72e97ade-83a4-4f8c-a217-4627b985759c.jp2	171	210	164	94
c482c317-cf42-4241-9692-bfba3a9ac543.jp2	9	16	25	14
8cc76d2a-2dce-446c-be0a-9d75e1ee5e01.jp2	0	9	1	5
2a0d2d73-d7ac-49c6-8058-79f4fac23e55.jp2	281	5002	5928	3199
5bfee1c4-dc5f-4e9c-9dd6-3e3d41b41451.jp2	11	6	4	10
84f5d964-4181-4947-9891-1838e604db0f.jp2	39	100	155	90
1e5ce292-60de-4f43-9beb-5fa993756128.jp2	310	4985	5916	3223
3d7b92b3-a7c6-4c59-9ba8-7a1a4988c3cc.jp2	387	15	13	388
8c525345-e566-4796-8f1d-38f7187ce69c.jp2	21	152	187	109
ab03e43c-367d-41c8-9c70-a766cf170afa.jp2	224	4875	5783	3121
dd147ce8-444c-4f2f-8d4c-fd95a85c0517.jp2	293	4972	5906	3203
0ce527c8-ddae-4d05-a6e4-51bbee6f55d3.jp2	10	11	15	3
5108cb71-a467-484a-adba-2c03f3ab0265.jp2	28	160	190	106
168e31d0-9c8b-4f91-ba04-29dff5aea8f6.jp2	15	131	158	84
3c28f382-0100-4d55-a68d-8ec281065c06.jp2	42	27	70	67
002de85c-eb86-4e0e-add2-5f05929072fe.jp2	284	4976	5888	3190
7043f56a-a342-4988-8dba-5a56369bdeef.jp2	31	168	187	91
d1b6298e-2cea-4bfe-97d6-3a436e81726a.jp2	242	4847	5758	3131
e58660fc-fc4e-4da1-80b6-2f9cdebda259.jp2	46	192	241	154
3a4b5a98-c11d-4a74-91de-94c66fa7ee88.jp2	168	4872	5786	3125
8f3c3f6f-5035-4a82-9c51-0ff5565d9170.jp2	25	701	669	40
259cce76-6d9f-4d4a-820f-c23fc38c6ee5.jp2	354	173	166	369
30f342d1-26d4-4218-97f7-9e0eece7920e.jp2	537	386	378	539
8219f4aa-b252-47e7-8a43-e79e3c006001.jp2	26	150	175	96
ab705dbf-d182-49ad-b8b8-e43ccabb5d03.jp2	5	11	17	7
e325a9cf-6306-49ca-bab4-7535dedd80a7.jp2	114	212	184	40
953ecb14-f7cc-40fa-8e7d-6458ba5e94a4.jp2	7	72	4	79
ce1c0d3b-ea56-4f86-9c84-b3d25802bb39.jp2	42	117	166	92
7f79b72f-0e1d-4fc0-ab40-e37db8a4844a.jp2	57	1078	1714	1330
Average	**130**	**1404**	**1678**	**951**
Standard deviation	**145**	**2122**	**2516**	**1347**

Table 9 Euclidean distance between corresponding corner of automatically and manually generated masks.

The calculated distances are in pixels. Distance0 refer to the distance of the automatically generated point to the manually generated one on the top left corner of the image. Distance1, 2 and 3 refer to the corners in clockwise direction starting from top-left (distance0).

Contributors

Authors

Giorgos Iliopoulos | G.Iliopoulos@student.tudelft.nl

I am 26 years old, and I am Greek. My first degree is an integrated engineering diploma in Spatial Planning and Development during which it became apparent to me the value of geographic information. The true potential of this information can only be explored when it is digitally enabled with state-of-the-art tools. Leveraging it can benefit decision-making in the public and private sector and have a great economic and societal impact. This comprehension made me pursue my second degree in Geomatics because I wanted to not only be able to use relevant software and tools but also take a look behind the scenes of how those things are built and function. The current project is taking place within the framework of this Master program. My expectations are to apply my previously acquired knowledge to a real-world problem – that of automating georeferencing of old map series – but also improve my problem solving skills since reality is more complex and has more factors to consider and raising issues to tackle than an assignment. In addition to these, I expect to improve my cooperation and communication skills with the other team members, supervisors, and external partners.

Eirini C. Tsipa | e.c.tsipa@student.tudelft.nl

I come from Greece and I hold a MEng in Rural and Surveying Engineering, with a focus on Cadastre, Photogrammetry, and Cartography. Currently, I am working towards my MSc in Geomatics at TU Delft. I appreciate the science of Cartography and working on this project adds extra value to my studies. In my eyes, maps are powerful tools and exploring their potential even further, by using programming to automate the process, seems exciting. In this project, I hope to strengthen my teamwork and decision-making abilities and improve my coding skills in relation to Cartography. Additionally, I am enthusiastic about the opportunity to support the mission of the national archive and collaborate with Allmaps on their valuable work.

Rianne Aalders | R.M.Aalders@student.tudelft.nl

I am from the Netherlands and moved from Groningen to Delft to do my bachelor in Architecture and the Built Environment at TU Delft. During my bachelor's, I enjoyed the analysis of maps and building plans and the relationships between large scale processes the most. Which resulted in a minor in environmental sciences and my choice for the master Geomatics. During this project I hope to learn about the history of map making and about programming, specifically with regard to computer vision and web integration. For my personal development I want to work on my time management and group communication skills.

Oliver Post | O.J.Post@student.tudelft.nl

I'm from the Netherlands. Before Geomatics, I completed the architecture bachelor at TU Delft and did part of the architecture masters. During the COVID lockdown, I developed a passion for coding which brought me to the Geomatics masters. In this project, I hope to become a better team player and to learn more Geomatics specific coding skills on a project with usability in the real world. Also, my interest in general history makes the process of working with historic maps all the more interesting.

Supervisors

Edward Verbree | E.Verbree@tudelft.nl

works at the TU Delft, in the Faculty of Architecture and the Built Environment (Bouwkunde). His expertise is in positioning, location awareness and point clouds. In the master program Geomatics, he is involved in the courses positioning and location awareness, sensing technologies and the synthesis project.

Martijn Meijers | B.M.Meijers@tudelft.nl

works at the TU Delft, in the Faculty of Architecture and the Built Environment (Bouwkunde). His expertise is in GIS, geo-information modelling, map generalisation, geo-database systems and cartography and geo-visualisation. In the master program Geomatics, he is involved in the courses python programming for Geomatics and Geoweb technology.

External Client

Jules Schoonman | J.A.Schoonman@tudelft.nl
is an architect and researcher working as a digital curator at the Delft University of Technology Library, where he is responsible for the presentation of the digital special collections and the initiation of digital scholarship projects. Together with Bert Spaan, he founded Allmaps in 2020, a platform for curating, georeferencing and exploring digitised maps that are available through the specifications of the International Image Interoperability Framework (IIIF).