

Automatic content-based georeferencing of historical topographic maps

Jonas Luft  | Jochen Schiewe

Lab for Geoinformatics and Geovisualization (g2lab), Hafencity Universität Hamburg, Hamburg, Germany

Correspondence

Jonas Luft, Hafencity Universität Hamburg, Henning-Voscherau-Platz 1, Hamburg 20457, Germany.
Email: jonas.luft@hcu-hamburg.de

Funding information

Freie und Hansestadt Hamburg; sharing.city.college

Abstract

Old maps are more than a cultural artefact: they are data. Data about the past still hold value for science and decision-making today. Libraries and archives have come a long way in digitising their inventories of thousands, sometimes millions, of historical maps using high-resolution scanning. Unfortunately, even with digital images the rich spatial and semantic information is inaccessible for people without a strong background in history and cartography. Only with georeferencing can historical maps be used in GIS and thus processed and compared with modern geospatial data. We introduce content-based image retrieval to automatically localise and georeference map images from topographic map series. We align the maps by extracting a subset of their symbols and cross-referencing them with online reference data from OpenStreetMap. We demonstrate our method with the Karte des Deutschen Reiches at 1: 100,000 scale, obtaining 96% correct location predictions and a median georeferencing error of 101 m.

1 | INTRODUCTION

Georeferencing is a prerequisite to making data in old maps usable in modern geographic information systems. It requires finding points in the map image that have equivalents in reference geodata. This recognition allows us to label points in the image with their position in the real world and calculate a mapping of the image into world coordinates. Suitable points for georeferencing have to be determining which points in the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Transactions in GIS* published by John Wiley & Sons Ltd.

map can be considered 'spatially salient', that is, carry information that allows us to match the points to a real location. A human reader might quickly identify spatially meaningful landmarks, but the dense multi-layered information in maps is hard to parse with automated procedures. Even with recent advances in computer vision (CV), separating these layers of information and determining reliable objects for predicting location remains a challenging task.

We introduce the use of content-based image retrieval (CBIR) methods to look for the spatial phenomena of a digitised map in big online geodatabases to predict the map's most likely location. Content-based image retrieval is an established method for different applications (Smeulders, Worring, Santini, Gupta, & Jain, 2000) but has so far not found much attention for use with maps (a notable exception being the work by Michelson, Goel, & Knoblock, 2008). However, using CBIR has great potential for processing maps because the availability of both digitised maps and more and more precise volunteered geographic information, most notably on OpenStreetMap (www.openstreetmap.org, OSM), has greatly increased (Dorn, Törnros, & Zipf, 2015).

1.1 | Motivation

The 'time- and place-independent provision of scientific sources' is an emerging key task of archives and libraries (Crom, 2016). However, manual georeferencing entails significant effort and requires expert knowledge of GIS and historical maps. Considering there are thousands, sometimes millions of maps in storage at national archives, new approaches are needed to make this trove of valuable data more accessible.

Augmenting the physical artefacts with digital spatial references not only allows libraries to index their archived maps by metadata but also makes the maps accessible through semantic or spatial queries (Chiang, 2015), for example by place name or spatial feature.

Moreover, georeferencing has implications for many other scientific and commercial endeavours as it allows any number of geospatial analyses on historical data. Many studies in the digital humanities have an interest in tracking change in human settlements, culture and landscape developments (Alvares-Sanches, Osborne, James, & Bahaj, 2020; Schröder, 2013). For built environments, a glimpse into the past can be invaluable, for example, when dealing with the disposal of unexploded ordnance, which can be anticipated with access to accurate contemporary mapping data (Kruse et al., 2017).

It can be difficult to compare the content of multiple maps when they are of different age and style. Georeferencing can help by making them spatially congruent and thus enabling researchers to compare information between very differently designed maps (Höhn & Schommer, 2017; Schlegel, 2019).

1.2 | Challenges

To determine a map's location, researchers cannot rely solely on the metadata provided in library catalogues. Historical maps sometimes lack metadata and often do not have clear designations of their location, much less their precise extent. Furthermore, the underlying spatial reference system used to create the map can be at best unclear and at worst flawed. Therefore, using a map's content instead of its metadata to determine the map's location and extent is advantageous. Determining a map's location from its content is possible because there are some invariant spatial phenomena that are depicted in the old maps.

Unfortunately, automatically extracting symbols of specific phenomena from digitised topographic maps is an unsolved problem. There are very clear rules on how to design and draw a readable map with regard to the arrangement of map symbols and the use of visual variables. Cartographers pick colours firstly to be representative of what they symbolise based on the reader's preconceptions (e.g., by the reader). It is significantly harder to do the reverse, separating all the symbols in a map image and deciphering their relationships.

Humans can rely on prior experience and Gestalt principles (Koffka, 1935) to recognise noisy map symbols, subconsciously tracing lines and connecting them semantically, even if the colour is not exactly the same or there are some gaps.

Machines, on the other hand, have trouble reading map symbols due to the complex digital representation of the abstract concept of 'colour'. To make matters worse, unexpected colour changes appear frequently in old maps due to age, degradation and archiving effects.

Despite advances in computer vision, historical maps still pose unique challenges. Powerful CV methods tend to rely on texture information of descriptive local patterns to detect and classify objects in photographic data (Hermann, Chen, & Kornblith, 2020). Maps, on the other hand, show highly repetitive, ambiguous local structures which are abstract and generalised. Generalisation is intended to reduce the information content and to make the map more easily readable by a human. The simplified local patterns resulting from generalisation do not carry enough information to be uniquely identified on their own. Rather, simplified symbols need to be considered within a larger context, which includes their geometric relationship to each other. After successful segmentation, we are left with a binary image. This exacerbates the problem of ambiguous patterns since this results in even less context and no texture. There have been many proposals to define new feature descriptors that accommodate non-local structure (Carmichael, Laganière, & Bose, 2010; Zhou, Rui, & Huang, 1999). But unfortunately, no convincing method has emerged so far to reliably detect semantic information in maps.

The top-down perspective of information retrieval, specifically CBIR, can provide the missing context. Typically, CBIR uses object classification and thus relies on delineation of regions of interest to describe the content of images. When applying CBIR to maps, we have to treat the whole image as the desired object.

Intuitively, this would call for global similarity methods (Wang, Bovik, Sheikh, & Simoncelli, 2004), but those are problematic because of possible defects in the image, such as the differing map margins, missing map symbols and noise. Furthermore, because maps might have been digitised in various settings, we have to account for different image resolutions and sizes, as well as the occasional scanning artefacts, such as cropped maps or visible colour correction cards.

1.3 | Related work

Many methods have been proposed for georeferencing increasingly available aerial imagery. Content-based algorithms are necessary for historical aerial images taken without positioning systems or for cases in which the spatial reference has been lost (Chen, Chen, & Tseng, 2016; Cléri, Pierrot-Deseilligny, & Vallet, 2014; Papakosta, Ioannidis, & Vassilaki, 2012). However, these methods do not easily apply to digitised maps, because of the previously mentioned difficulties. Instead, georeferencing of maps requires a laborious process of manually annotating ground control points (Fleet, Kowal, & Pridal, 2012; Janata & Cajthaml, 2020).

Fortunately, automatic georeferencing of historical topographic maps has slowly gained attention in recent years. Heitzler et al. (2018) have georeferenced sheets of the so-called Siegfried map, a large Swiss topographic survey series. They developed a method combining multiple techniques from computer vision and document analysis to first detect the map borders (neatlines) in the map image and then use optical character recognition to read the geographic coordinates on the map margins. With this combined information they could accurately georeference 40% of map sheets. The drawback of relying on the metadata on the map margins is that while this method might be suitable for a large series of uniform design such as the Siegfried map, it relies on a lot of prior knowledge of the layout of the map and margins and thus does not easily transfer to other map series. In a similar vein, Burt, White, Allord, Then, and Zhu, (2020) have proposed a system for a semi-automated user-guided system for map georeferencing. Their idea is to provide as much metadata about a map as possible, but require at minimum the map border coordinates and the graticule subdivision steps covering the map. Then the metadata are used to find as many graticule intersections as possible with template matching. With the detected graticule

intersections and their expected geographical coordinates, a polynomial transformation solution can be found with the least-squares method. Their method failed on about 12% of tested sheets, but can easily be corrected with user assistance and subsequently reach very high georeferencing accuracy.

Howe, Weinman, Gouwar, and Shamji, (2019) implemented an approach to content-based georeferencing on small-scale political and thematic maps. They detect and geocode toponyms using convolutional neural networks trained on synthetic training data. The geocoded toponyms are used for initial location estimates. The map image is subsequently matched with vector data of state and county boundaries for fine alignment. We will compare their georeferencing results with ours in Section 4.2. Luft (2020) used a similar method of toponym detection and open geodata for geocoding. The text detection and recognition algorithms proved to be too error-prone on cluttered, highly detailed topographic maps and led to unsatisfactory georeferencing results, demonstrating the fragility of toponym-based methods and the sensitivity of georeferencing transform methods to noise and incorrect control point matches.

Using CBIR to exploit the wealth of publicly available digital maps was demonstrated by Michelson et al. (2008): they automatically mined images from the web and classified them as map or non-map. They queried images with similar content to already classified maps, based on textureless and colourless shape similarity, which proved to be both more accurate and more scalable than a comparable machine learning-based method.

Wolter, Blank, and Henrich (2017) have proven that rivers have a very unique topology and can be matched to textual descriptions of their course. This is a prime example of spatial reasoning over two very different feature spaces, i.e., the image space and text. The actual matching is done in an intermediary abstraction layer—the topology of the river represented as a graph. They demonstrate their method on a digital map image with little clutter that is easily segmented. Unfortunately, their method of river extraction does not transfer easily to historical map artefacts with dense multi-layered symbols. Nevertheless, they proved that rivers exhibit discriminable structure, which was an inspiration for the choice of water symbols for georeferencing in our work.

In conclusion, not much research focuses specifically on developing methods for automated georeferencing of maps. While research in cartography and GIS often targets specific use cases, we aim to develop a universal system independent of possible applications. Conversely, the transferability of the proposed method is primarily limited by the assumptions that the input maps allow.

1.4 | Contributions and assumptions

In this article we demonstrate our novel georeferencing method on a subset of sheets of the topographic survey map series *Karte des Deutschen Reiches* at 1: 100,000 scale. We extract coloured map symbols from digitised map images and use these symbols for content-based georeferencing by matching them to reference OSM data of possible map sheet locations.

Even though the same method could potentially be used with many map symbols, we use water symbols specifically. Water symbols are often distinctly coloured in survey maps and can be extracted with little specific knowledge of how feature classes are encoded in visual variables. Using water symbols is, of course, only possible for maps that show regions with bodies of water and colour water symbols distinctly. Maps without visible water symbols could possibly be georeferenced with a different choice of symbols, which demands more research into applicable segmentation and feature matching schemes.

In contrast to previous works, the proposed method does not rely on visible neatlines or a graticule on the map, but is strictly content-based. This enables successful georeferencing of incomplete maps, for example partially cropped maps, and maps with an irregular outline. However, some conditions need to be met in order to apply the proposed method on a set of maps. Approximate geographic coordinates of each map's boundaries need to be known. Those can usually be read from the map's margins. To match the map content to the reference data with linear transformations, the map's projection needs to be similar to the reference data (WGS84). Alternatively,

the reference data can be reprojected if the coordinate reference system of the maps is known and the maps are oriented approximately north-up. More importantly, some knowledge about the colouring of map symbols is necessary (e.g., rivers are blue) to extract them (Section 2.1) and represent them with OSM tags, as elaborated on in Section 2.2. High image resolution and high level of detail (depending on map scale) help to accurately extract these symbols.

2 | METHOD

2.1 | Segmentation

Image segmentation is concerned with partitioning image regions (i.e., sets of connected pixels) into different objects. The goal of *semantic* segmentation is to separate objects in the image not by object identity, but by class membership. More generally, semantic segmentation assigns each pixel in the image a class label. The simplest classification partitions the image pixels into only two classes, figure and ground (MacEachren & Mistrick, 1992).

In our case, we partition the scanned map image into pixels depicting water (figure) and all others (ground). The result is a binary image, a simplified map of bodies of water.

The segmentation method needs to be robust to different image resolutions and sizes and should not be affected by archival effects, such as paper degradation, effects of the digitisation process, colour bleed and colour space conversion.

There are different approaches to segmenting an image. We chose a purely bottom-up approach, using local pixel values only. Use of colour information is usually avoided in classic computer vision applications because it is highly susceptible to changes in illumination as well as technical properties of the camera used. For topographic maps, on the other hand, colour may be the most discriminable visual variable.

Segmentation is based on local or global thresholding of colour values. In order to classify objects by colour information the images have to have a comparable colour range. Even in professionally colour-corrected map scans, the colour ranges can differ significantly because of bleaching, discolorations or a different composition of inks used during printing. These effects make it hard to separate the different colours of the palette in RGB colour space, which is predominantly used in digital images.

A suitable alternative is $CIEL^*a^*b^*$, which represents colours in three dimensions: lightness (L^*), the green–red axis (a^*), and the yellow–blue axis (b^*). It was designed such that distance in the colour space is proportional to the visually perceived distance (CIE International Commission on Illumination, 2019).

This makes $L^*a^*b^*$ more suitable than RGB for distinguishing the limited set of print colours in maps (Sharma & Bala, 2003). After converting the images to $L^*a^*b^*$, we applied histogram stretching to enhance contrast and alleviate some colour degradation. After blurring to reduce high-frequency noise, thresholding is as simple as extracting all pixels with a negative b^* value. Some care has to be taken to set thresholding parameters for all dimensions because the pixel values are noisy and the segmentation is cleaner if thresholds are set more conservatively. Global thresholding is sufficient if map images were created with a professional flatbed scanner with uniform illumination properties. This threshold is robust to strong yellowing of the paper, since the desired blue colour will be at the opposite end of the b^* -axis from the background and thus easily separated. In our experiments, the $L^*a^*b^*$ colour space proved to be more suitable for thresholding the visually distinct blue water features than other colour spaces without needing to model inconsistent discolorations or to filter digitisation-induced noise.

The output of this simple segmentation procedure is a segmentation mask that remains very noisy even after some additional morphological operations (opening, closing). Figure 1 demonstrates the result of segmenting a map sheet: the segmentation shows a lot of white speckles, and lines are frequently interrupted. With this purely bottom-up approach, it is hard to robustly join or separate objects, such as lakes from their inflows and rivers from their branches, which would be necessary for topological or object-based matching methods.



FIGURE 1 Comparison of source and target image used for map matching. Left: the original map image. Middle: the source image, a segmentation mask of all pixels representing water symbols. Right: the target image, painted from OpenStreetMap data for the corresponding sheet's bounding box. Note that the target image does not contain map margins and is not perfectly identical to the source image because of segmentation uncertainty, noise and a possibly different subset of depicted water symbols. Map sheet no. 66 of the *Karte des Deutschen Reiches* © Staatsbibliothek zu Berlin. OpenStreetMap data © OpenStreetMap contributors

2.2 | Reference data

The CBIR approach to determining the location of a map is to find the most similar maps in a reference database and derive those maps' locations. The reference data can be official survey data or any other mapping data. Especially interesting are OpenStreetMap data, the largest and most complete open mapping data source. Because OSM data are available around the world our method is transferable to new locations. This would not be the case for official survey data, which are not openly available everywhere and are sometimes expensive or impossible to acquire.

The greatest challenge with OSM data is filtering for the desired phenomena by its system of 'tags': vector features are tagged with all manner of semantic information as key-value pairs. Unfortunately, the categorisation is sometimes inconsistent (e.g., `water=lake` but not `natural=water`) and categories are overlapping (Mooney & Corcoran, 2012).

This requires some fine-tuning of the Overpass¹ query used to download the desired OSM objects. The choice of selected tags and values is very important for comparability since not every phenomenon is marked equally in historical maps and the presence of symbols is dependent on map scale. The construction of a suitable Overpass query requires some knowledge of metadata (i.e., which features are drawn in the maps), and might have to be adjusted for different data sets. For example, in our data set (Section 3.1) highest-order streams are always drawn, but lower-order streams and agricultural ditches are coloured in the later three-colour editions only.

The reference geodata should be available as vector data (as is the case with OSM data). The advantage of vector data is that they are easily filterable and scale-independent. The challenge is that they are not immediately comparable to raster data, such as the scanned map images. Converting raster maps to vector data is not trivial and is a field of intensive study (e.g., Chiang, Leyk, & Knoblock, 2014).

Conversely, converting vector data to raster data is as simple as drawing rivers as lines and lakes as polygons in white colour on a black image. This creates a reference water map in the same image domain as the segmentation mask (see Figure 1) and both can be compared and matched with image-based methods, which are easier to implement and more robust under uncertainty than vector data-based methods such as subgraph or topology matching (Conte, Foggia, Sansone, & Vento, 2004). After rasterisation, the scale-independent arbitrary precision of vector data is lost. Hence, the resolution of the thus created raster reference images should approximately match the level of detail in the scanned map images that are to be georeferenced.

2.3 | Localisation method

The localisation process is guided by the possible map locations (Figure 2). Location and extent for each sheet in the series are represented as sheet bounding boxes. The complete set of bounding boxes was constructed in advance from the map series' metadata. For each of the bounding boxes, a rasterised reference map (target image) is constructed from OSM reference data with the aforementioned query. Predicting the location of a segmented map image (called the source image) works by finding the target image which covers the same area.

The source image and corresponding target image will never be exactly the same because the segmentation is noisy and the source image shows empty space at the map margins and thus might have different dimensions. Both images also might contain a different subset of water bodies either because the OSM query did not fit the rules of colourisation for the map or because the hydrology changed between the time of drawing of the map and the time of the OSM query, for example after canalisation of rivers or flooding of gravel pits.

Therefore, a similarity metric has to be chosen that is scale-invariant (because of a possible mismatch in scanning and rasterisation resolutions and differently cropped map margins of the input images), robust to noise and capable of exploiting the geometry and topology of the water maps. Being able to identify large-scale shapes (i.e., non-local relations, such as the course of a river) while allowing slight local changes in geometry is necessary to robustly match water maps from different sources and time periods.

We define the similarity between two images as the number of spatially consistent image feature matches between them. Therefore, two images are similar if many image features can be matched. Since the number of matches does not have an upper bound it cannot immediately make the correct location prediction. Instead, we

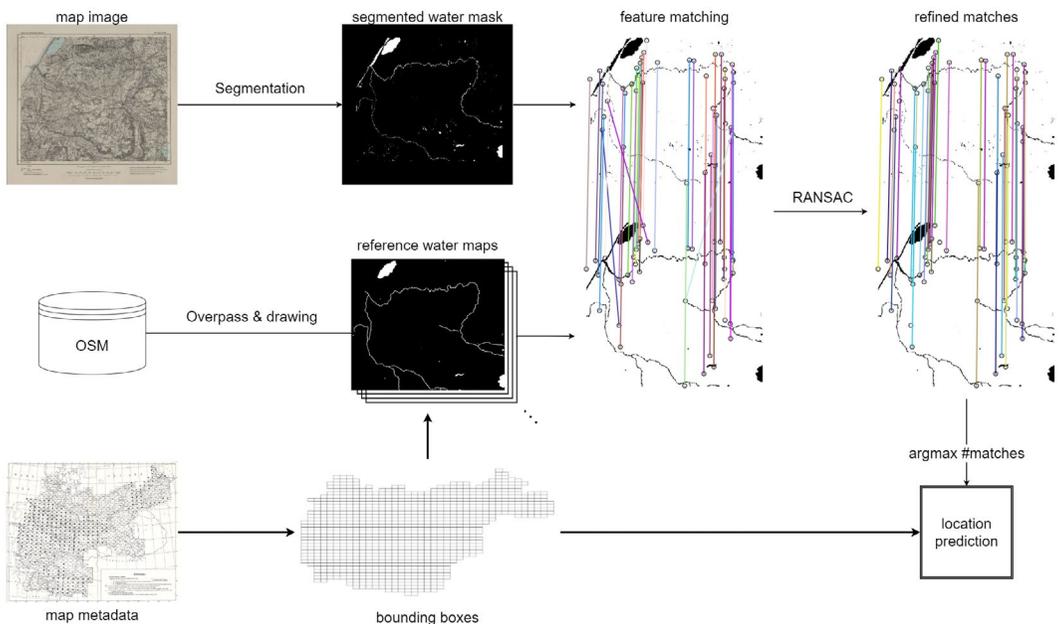


FIGURE 2 Overview of the localisation process. The input map image (top left) is segmented to extract all water symbols. The reference water maps are created by retrieving water objects from OpenStreetMap within the bounding boxes of each sheet in the map series. The segmented input image is compared with each of the reference maps by feature matching. The bounding box of the reference map that exhibits the most matching features after RANSAC refinement provides the location prediction for the input map image. Map © Staatsbibliothek zu Berlin. OpenStreetMap data © OpenStreetMap contributors

can create a total order of similarity of all reference images to a single segmented map image. The reference image with the highest number of matches is most likely to contain the same content as the scanned map.

We use 30×30 pixel patches as feature descriptors. Pixel patches are suitable for describing local geometry without much expected variance between matching features. Further advantages of patch features are the clearly defined similarity (the Euclidean distance between two patches' intensity values) and the minimal computational effort.

A set of image features can be used to describe two images and compare their content very efficiently. In contrast to global image similarity measures, feature-based comparison is robust to noise and deformations. However, in order to reach high accuracy with only few descriptors, it is important to extract feature descriptors at the most descriptive points in the image. Descriptive points are determined by detecting FAST corners (Rosten & Drummond, 2005). FAST corner detection is computationally inexpensive and works sufficiently well on binary images without texture or gradients. Corners are extracted only at the border between figure and ground regions, whereas uniform areas offer no discriminable features.

Patches are extracted from the source image (segmentation mask) and the best-fitting correspondences in the target image are found with cross-correlation template matching (Lewis, 2001) (Figures 2 and 3 show the feature matching step with the feature locations as circles and matching features connected with coloured lines).

Since many patches show recurring geometry, and there is no texture to distinguish between otherwise similar features, many conflicting matches are expected. Therefore, the conflicting matches are refined for spatial consistency by random sample consensus (RANSAC; Fischler & Bolles, 1981). RANSAC can estimate different transformation models to fit the observed matches. We are using affine transformation models to account for the expected deformations (Figures 2 and 3 show the refinement of feature matches; note the removal of diagonal matches).

Figure 3 shows the matched pixel patches between the source (left) and target (right) image (top) and the remaining matches after RANSAC estimation (bottom). The number of remaining feature matches after RANSAC gives a good indication of the quality of model fit, as can be seen in the example of a correct prediction. Estimating RANSAC transformation models and counting remaining feature matches is repeated for every possible map sheet location to find the sheet that provides the most matched features and thus the most likely sheet location prediction.

To recap, the localisation procedure (Figure 2) works as follows: (a) extract patches from the source image; (b) iterate over all possible map sheets; (c) create reference source images by downloading OSM data and drawing a target image for each sheet; and (d) try to find a RANSAC model for each sheet. Finally, the sheet with the highest number of remaining matches is selected as the location prediction.

2.4 | Registration

After the correct sheet location has been found, the sheet's known bounding coordinates are assigned to the map image. The coordinates refer not to the image corners, but to the intersections of the neatlines. The map images might contain an unknown amount of map margin, and might even be cropped so that the neatlines are no longer visible. Therefore, it is not always possible to use the map borders for georeferencing, as is proposed by Heitzler et al. (2018) and Burt et al. (2020).

Instead, we use the map content for fine alignment of the map image. We use the established *enhanced correlation coefficient maximization* (ECC) method for image registration (Evangelidis & Psarakis, 2008). With ECC alignment, an affine transformation model that aligns the source image (segmentation mask of the input image with margins; Figure 1, middle) to the target image (OSM reference water map without margins; Figure 1, right) is found through iterative optimisation. The algorithm terminates after having found a maximum matching score, the correlation coefficient. The transform model estimated by RANSAC can be used as a starting iteration for the ECC registration in order to reduce the time to convergence and possibly improve overall accuracy. The advantage of using ECC for registration in addition to the transform already found by RANSAC is that it can exploit information from all pixels of the image and not only from the sparse points of interest.

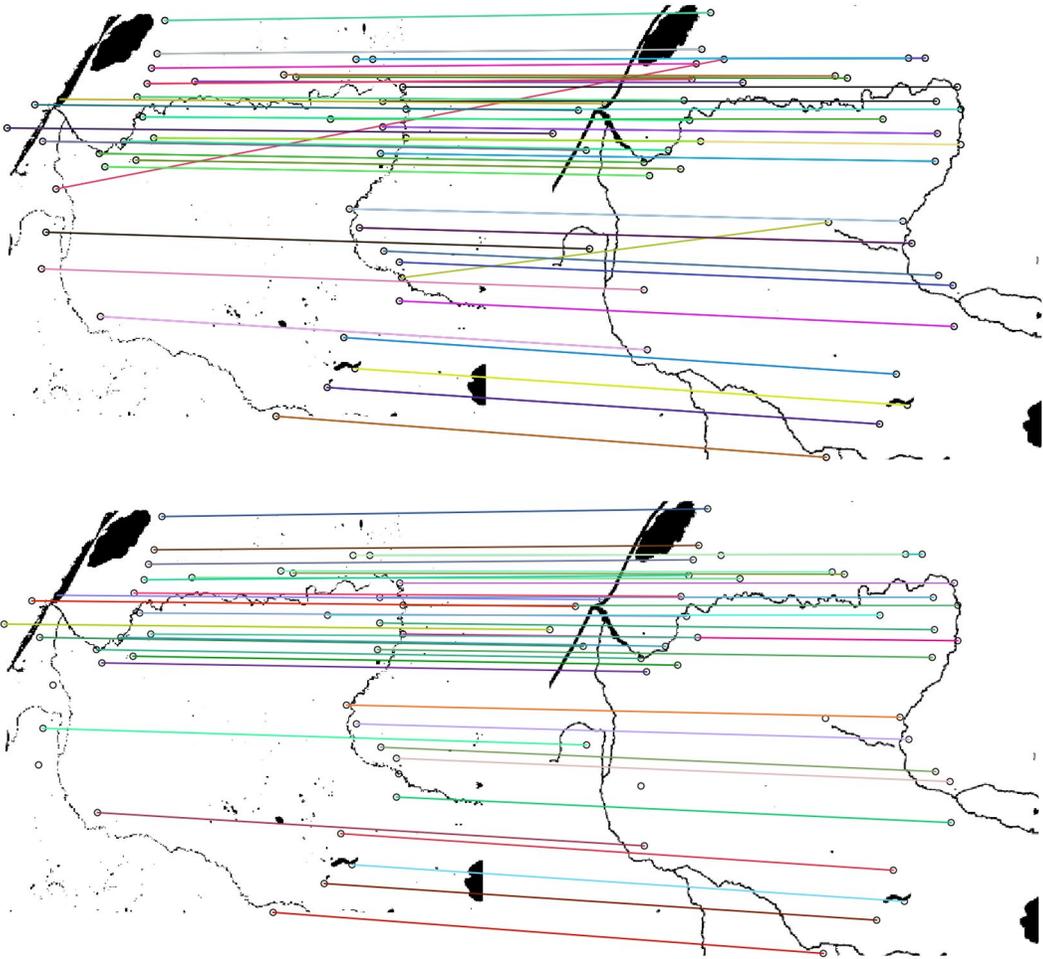


FIGURE 3 Matches of detected features for sheet 66 between the source image (i.e., the segmented input map image, left) and the target image (i.e., the reference OpenStreetMap data, right). Black circles show detected interest points, coloured lines are matches of interest points between images. Top: all detected matches. Bottom: refined matches after RANSAC with affine transformation model

After the transform model has been found, we can crop the source image to the extent of the target image to remove the map margins. This creates an output image containing only map content for seamless stitching of georeferenced map sheets. In the end, the actual spatial reference is set by using OGR/GDAL's (www.gdal.org) translate and warp tools to assign the sheet's bounding box coordinates to the cropped output image and reproject it as desired.

3 | EXPERIMENTS

3.1 | Data set

We evaluated the proposed method on a set of sheets of the 1: 100,000 scale Karte des Deutschen Reiches (abbreviated as KDR100). The KDR100 was a topographic survey map series jointly commissioned and produced in the late nineteenth century by the German states of Prussia, Bavaria, Saxony and Württemberg. An additional set of sheets was later included, adding the Ostmark and Polish territories (Reichsamt für Landesaufnahme, 1931).

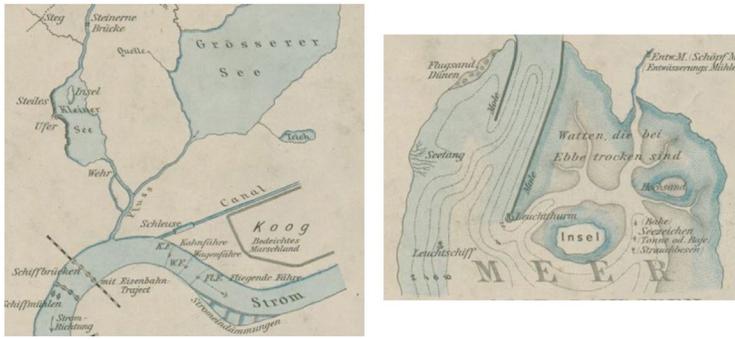


FIGURE 4 Different representation of water bodies in the Karte des Deutschen Reiches: on the left inland water with solid blue colouring, on the right open sea with only coastal areas in blue. Map key of the Karte des Deutschen Reiches © Staatsbibliothek zu Berlin

In total, the map series encompasses 674 original and 236 additional sheets in central and eastern Europe, each sheet spanning 30' longitude by 15' latitude. For this article, we used a subset of 56 sheets, downloaded from Wikimedia in 400 ppi resolution.² These sheets are sourced from different editions of the map. The sheets from the first edition are in black print only, with coloured rivers and lakes and coloured administrative borders. The maps of the later edition are printed in three colours: blue for rivers, lakes and swamps; brown for topography; and black for all other symbols and labels.

In the different map editions, bodies of water are coloured differently. Where the earlier hand-coloured maps only show larger rivers traced in blue, the later colour-printed maps also shows smaller brooks, ditches and swamps in blue colour. The North Sea and Baltic Sea are not coloured completely, but only show a slight gradient of blue colour along the coastline (see an excerpt from the map key in Figure 4).

The spatial reference system of the KDR100 is based on the ellipsoid of Bessel (1841) and the Rauenberg datum. The geography is projected to the map plane using the *Preußische Polyederprojektion*, where the Earth is separated into regular polyhedral segments of the desired map size. The trapezoid map sheets thus formed have straight neatlines and minimal flattening deformation of only 0.16 mm at scale 1 : 100, 000 (Reichsamt für Landesaufnahme, 1931).

The maps are north-oriented (i.e., have no rotation), but come in different sizes and with an unknown position of the map content, because in some cases they have been positioned and cropped differently during digitisation. The affine transform model is suitable to address the geometric distortions of the digitisation process, but cannot fully rectify the trapezoid shape of map sheets, which leads to an expected minimal error of about 10–15 pixels (px) at the map corners.

3.2 | Results

3.2.1 | Evaluation pipeline

Corresponding to the two-stage georeferencing process, there are two relevant results to evaluate: first, the number of successfully predicted sheet identities; and second, the accuracy of the subsequent fine alignment. In this case, the automatic evaluation of correct localisation is straightforward because the ground truth sheet name is known for each image, either from the file names or manually annotated. We look up the predicted bounding box in the sheets overview to get the sheet name. Then we compare the predicted sheet name with the ground truth.

Evaluation of the georeferencing quality, on the other hand, is a little more involved (see Figure 5). For line-preserving transformations, any transformation error will interpolate linearly across the image. Consequently, the points of highest

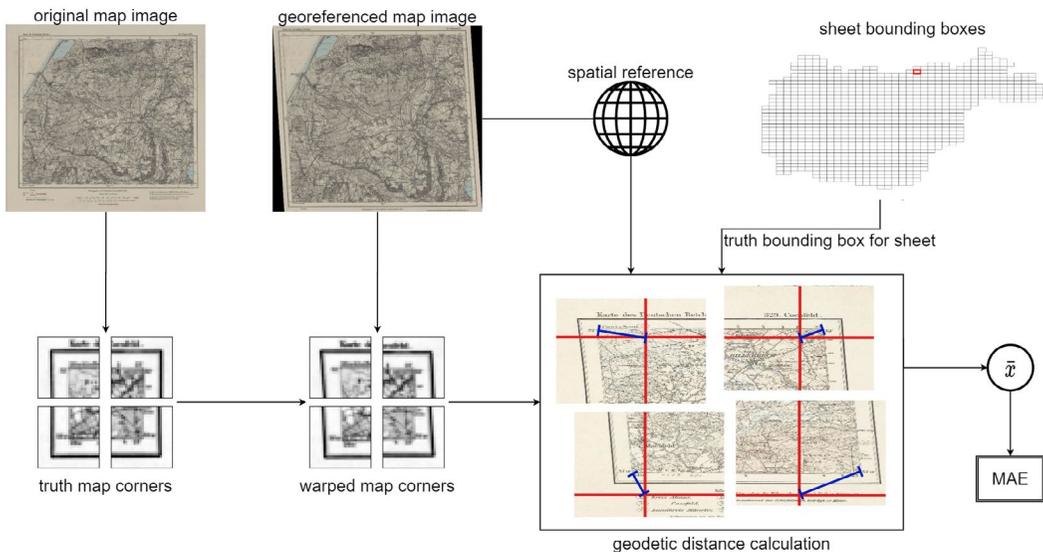


FIGURE 5 Evaluation pipeline. The manually annotated ground truth map corners are extracted from the original input map images. The map corners are then detected in the georeferenced and possibly warped output map image with template matching. With the embedded spatial reference in the georeferenced image, the distance (blue bars) of the detected corners to the ground truth bounding box corner (red line intersections) can be calculated. Averaging the four corner distances yields a singular error value per map sheet. Map © Staatsbibliothek zu Berlin

error will always be on one of the map corners. Therefore, it is sufficient to analyse the displacement of the transformed map corners from their expected positions in the series' sheet layout. Conveniently, the alignment of corners also directly determines the ability to seamlessly join neighbouring transformed map sheets, which is a major concern for map users.

Using only the edges, specifically the corners, of the map content for error calculations avoids observing errors that stem from the cartographic process (surveying and drawing). Cartographic errors can lead to symbols being placed at the wrong location on the map. Because neatlines are the first thing constructed and have the least projection error, they can be assumed to be drawn at the 'correct' place.

We have manually annotated the map corners of the original, full-resolution, unwarped map images. This provides a uniform and comparable baseline for evaluating georeferencing accuracy. Assuming the distortion (shearing and scaling) is not too strong, the map corners will still look very similar after warping. This enables us to use template matching to find the same corners in the warped output images.

Having found the map corners of the warped output images, we can interpolate their geographic coordinates from the assigned spatial reference. The said geographic coordinates of the warped map are now compared to the ground truth corner coordinates of the sheet boundaries. To obtain values that are comparable across all maps and are not influenced by projection, we reproject output coordinates and ground truth coordinates to WGS84. To calculate a singular error value for each map sheet we measure the geodetic distance between ground truth and detected map corners.

3.2.2 | Evaluation of location predictor

The proposed localisation method predicted the correct bounding box (and thus sheet number) for 53 of the 55 input maps.

The robustness of the similarity function based on selecting the transform model with the number of template matches can be more closely evaluated by verifying the statistical distance between the chosen highest number of matches and the distribution of match count incidences for all other tested sheets. The statistical distance between an observation x and a distribution described by mean μ and standard deviation σ is calculated with the (simplified) one-dimensional case of the Mahalanobis distance $D = (x - \mu)/\sigma$.

When the sheet with the highest number of template matches (observation x) has a large distance to the remaining incidences of templates matched for all other sheets (defined by mean and deviation), that is, D is above a certain threshold, x is less likely to be explained by the distribution of 'noise' observations describing incorrect sheet matchings, but instead is generated by a different process and consequently the correct prediction.

Using the Mahalanobis distance assumes a normal distribution of match counts for the noise, which is an acceptable approximation for high distances. However, this heuristic loses precision in cases with low values of observations x and a small distance of the highest match count, where observed match counts are more uniformly distributed in a small interval.

Even though calculating standard deviations is biased for small sample sizes, the statistical distance gives reliable result as a 'correctness' classifier with few observations, too (i.e., only testing against 10–20 sheet locations).

In our experiments, the incorrect location predictions had a distance of below 3σ , that is, the best number of matched features was less than three standard deviations away from the mean number of matched features over all sheets. Incorrect predictions are thus clearly distinguishable from correct predictions with larger distances, from 5σ to 105σ .

Figure 6 justifies the use of statistical distances of location predictions by comparison to Lowe's test ratio. Lowe's ratio is a common method in feature matching to estimate whether the maximal matching score is in fact from a reliable match or was more likely observed due to chance. The split is commonly made at a score factor of .7 (Lowe, 2004). Using this ratio supports the classification by statistical distance, albeit with a smaller safety margin.

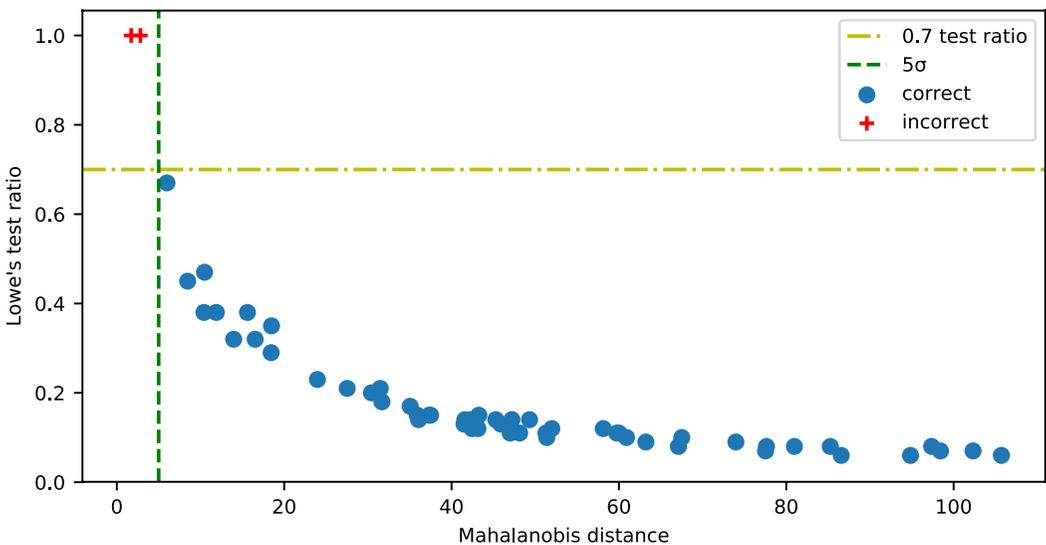


FIGURE 6 Evaluation of uncertainty measures to determine correct location predictions. Statistical distances of each sheet's location prediction in relation to Lowe's test ratio. Five standard deviations (yellow horizontal dashes) is the highest possible threshold to pick only correct location predictions (blue circles) and discard all incorrect location predictions (red crosses)

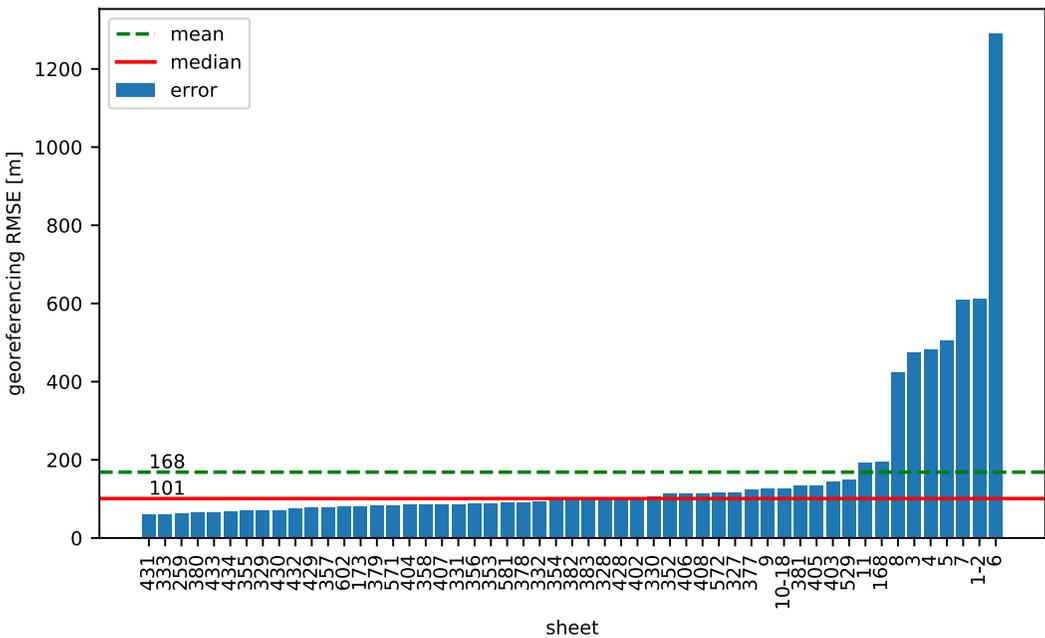
3.2.3 | Evaluation of georeferencing accuracy

Figure 7 shows errors for all correctly localised sheets of the Wikimedia data set. Well-registered maps show root-mean-square error (RMSE) less than 200 m (ca. 2 mm on map). The best fit was measured for sheets 431 and 333 with an RMSE of 59 m (ca. 0.6 mm on map). Some maps are distorted more strongly, usually because one half of the mapped area has more rivers and the remaining part is poorly aligned, which can lead to an error greater than 1,000 m (ca. 1 cm on map).

When removing the unsuccessfully located maps, sheets 12 and 79a, the Wikimedia data set's mean georeferencing error is 168 m. Of the remaining map sheets, nine have an error above the mean and 44 sheets have errors below the mean. The median of the error distribution is at 101 m. This shows most of the maps have a similarly low error and high georeferencing quality, with less than one-sixth of sheets showing subpar georeferencing quality.

A closer inspection of the maps with an above-median error highlights two main sources of inaccuracy. First, some map images suffer from missing information. Older editions of the KDR100 are coloured by hand and inconsistent on which rivers are coloured. Some older maps depict rivers which have no colour. In this case, our method cannot be used for localisation or for registration. When the remaining blue colour is either unevenly distributed over the span of the map or too sparse to be clearly distinguished from noise, the registration will show error rates in excess of 1,000 m.

Second, changes in the hydrological landscape, usually by human intervention through canalisation of meandering rivers, construction of harbours or newly formed lakes from flooding of former quarries lead to new shapes and patterns in the reference OSM data. Its iterative nature makes the registration algorithm prone to converge to local minima. For example, former river runs align with current oxbow-lake boundaries, leading to erroneously high correlation scores. In some cases, these interferences are strong enough to overrule the remaining 'matchable' structure, leading to subpar overall registration quality.



4 | DISCUSSION

4.1 | Challenges

A closer look at the failure cases suggests that the complications resulting in bad georeferencing can also hinder correct localisation. One extreme example is the incorrectly predicted sheet 12, an early edition print (no blue print, but manually coloured) in which none of the rivers are coloured and only tiny blue ponds are visible. Rivers are present in the reference image, but ponds were not retrieved by the Overpass query. Consequently, the two images show very low similarity and could not be matched.

A second extreme example, sheet 79a, is a unique case as it only shows the tiny island of Helgoland and otherwise has little map content. The only blue area is the small stretch of coastline surrounding the island. Coastlines are imprecise features on their own because their gradient colour is not segmented very clearly. Furthermore, Helgoland underwent significant artificial deformations in the early twentieth century and its coastlines from 1893 and today bear almost no resemblance. Unfortunately, our data set was too small to make a general argument about the probability of correctly predicting coastal map sheets.

Using full affine transformations introduces some shearing errors, even for otherwise well-aligned maps. Affine transformations are suboptimal for our data set since the map sheets are almost rectangular. Implementing a simpler transform model, however, limits the transferability, since handling affine deformations may very well be necessary for other map series (e.g., maps at a smaller scale have a less rectangular shape). On the other hand, more complex line-preserving transformations, such as the homography transform, would be able to compensate trapezoid map shapes but would further increase the risk of the estimation converging to local minima. Figure 8 illustrates how using the homography transform can improve the georeferencing accuracy of some images while reducing average performance.

4.2 | Comparison with the state of the art

Automatic georeferencing of maps, particularly historical topographic maps, is novel research. We are aware of only few recent proposals for geolocating sheets of a map series. Heitzler et al. (2018) use marginal information

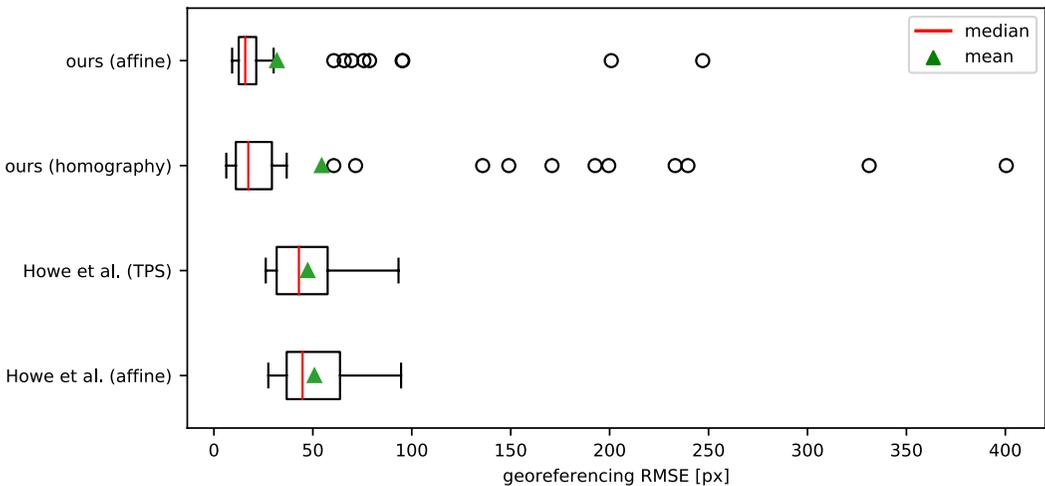


FIGURE 8 Comparison of error distributions of our method with Howe et al. (2019). The upper two distributions show the RMSE over four map corners for our method on 53 maps, using affine and homography transformation models. The lower two distributions show their RMSE over 20 maps with a varying number of control points, using affine and thin-plate spline (TPS) transformation models

instead of map content and were able to locate 40% (of 20) sheets correctly. In comparison, we located 96% (54 of 56) sheets correctly. This suggests that their multi-stage approach to detect and comprehend metadata on the map margins is viable, albeit less robust than using the actual map content.

Heitzler et al. (2018) do not give numerical error values of georeferencing accuracy comparable to our RMSE of map corners. Burt et al. (2020), on the other hand, show that using neatline corners and graticule intersections can deliver almost perfect accuracy of 1–4 px RMSE. This high accuracy reflects the detection accuracy of graticule intersections on the map image and is hardly influenced by the georeferencing process itself.

When the graticule cannot be extracted from a map and we have to rely on map content instead, significantly larger errors are expected. This is because the map content introduces more uncertainty than the presumably perfectly accurate neatlines (which we therefore use for evaluation). To evaluate our georeferencing accuracy, we compare our results to the only other content-based georeferencing method, developed by Howe et al. (2019). They use maps in different scales (1: 443,529 up to 1: 7,500,000), which makes the comparison subject to rescaling effects. Therefore, we compare error values in the image domain (pixels) instead of geodesics (metres). For that, we converted our geodetic distances to pixel distances by multiplying by the scanning resolution. We use only four points to calculate the mean error, whereas their number of tie points depends on the number of detected objects. Averaging over a higher number of points makes their error values less susceptible to outliers but does not make a great difference in general for line-preserving transformations, such as the affine transformation, when they are calculated by minimising RMSE.

Our mean RMSE georeferencing distance with affine transformation is 31.84 px and thus an improvement over their mean RMSE of 50.8 px using thin-plate spline transformation. Notably, our median of 15.92 px is significantly better than their 46.1 px, indicating a higher possible accuracy for the majority of sheets with our method. The distribution of map errors is illustrated in Figure 8 for direct comparison.

It is noteworthy that Howe et al. (2019) state significant ground truth errors, namely the RMSE of an affine transform with manually provided control points ranging from 16.9 to 84.2 px. This high baseline error suggests that either their choice of transformation model might be unsuitable, or the chosen features (toponyms) are not the best choice, because they are too imprecise for the selected maps. In our use case, the choice of the affine transform is suboptimal as well, since it cannot compensate for the trapezoid map geometry, thus also leading to a minimal error of 7.5 px in the optimal case of only one misplaced corner. Our advantage is that we use sheets of a map series with very similar cartographic design. Conversely, the maps used by Howe et al. (2019) show greater variation in design, even though they appear to be specifically selected to look sufficiently alike.

The main difference between their method and ours is the choice of phenomena and the consequential divergence of methods for extraction and identification of said phenomena. This makes either method applicable to different maps. Toponym recognition promises satisfactory results on small-scale maps with low density of information. However, optical character recognition performs badly on topographic maps because of the large amount of clutter, overlapping features and irregular typesetting (Luft, 2020). Our method, on the other hand, is not greatly affected by clutter or unusual map designs. Instead, we require consistently and distinctly coloured map symbols. We argue that our method transfers more easily to a wider range of maps. Nonetheless, toponym recognition as implemented by Howe et al. (2019) is viable for a number of maps to which our method is not applicable.

4.3 | Conclusion

It is hard to make a general statement about what georeferencing accuracy is necessary for maps to be used in practical applications. We argue the median precision of 101 m (about 1 mm on a 1 : 100,000 scale paper map) allows the use of the output maps for all applications except those with the highest precision requirements. The precision is satisfactory for spatial indexing and spatial or semantic queries in map-based applications. A prominent advantage of our method is the ability to predict whether the results are correct or not. With a reliable measure of

uncertainty, human users (e.g. librarians and historians) can focus their attention on georeferencing the incorrect predictions manually and save time on manual error checking.

Our proposed georeferencing approach suggests good transferability to other topographic map series. Apart from black-and-white maps, water is coloured in almost all topographic maps. Thus, water symbols should be extractable in many maps as easily as in our experiments. Experiments with the earlier hand-coloured sheets show that manuscript maps with uneven colour might work just as well as full-colour prints. The earliest map in the set dates back to 1861 (sheet 10–18) whereas the latest map is from 1913 (sheet 407). As can be seen in Figure 7, their error values are not far apart, at 86 m and 126 m, respectively. The most degraded sheet in the data set (sheet 1–2) shows a rather large error of 610 m, but this sheet was especially challenging because it contains almost no water features other than a large stretch of tide-affected shoreline. The robustness of the segmentation method appears to be promising with the data set at hand. However, more experiments are needed to determine the suitability for significantly older maps, where the variety of papers and dyes poses unique challenges.

Concerning transferability to other map types, we have to distinguish between the specific implementation of our method using hydrological phenomena and the more general idea, using the same localisation and registration workflow but segmenting other phenomena. Since we do not explicitly model what we match, we can theoretically use any other phenomena that can be segmented and are unique. The maps in question only need to display some spatial phenomena in geographical precision and geometric design that is sufficiently similar to current maps.

When using hydrological phenomena specifically, only maps with coloured bodies of water can be used. In particular, rivers are necessary for precise alignment since they provide the most discriminability at medium scale. For small-scale maps, coastlines might be descriptive as well. In any case, the registration precision suffers when matchable features are not distributed evenly over the map.

The choice of suitable phenomena is a question of robustness to historical changes. Changes in the landscape are one of the main reasons why some of our maps show a large registration error. It is noteworthy, however, that this method can match rivers which have changed significantly during the last 100 years without explicitly modelling changes in hydrology.

5 | FUTURE WORK

The performance of our method improves on the state-of-the-art. However, to achieve more consistent quality, more research is needed into which map properties have the strongest influence on georeferencing error. Understanding these map properties is especially important for applying the proposed method to new maps of different style, scale and geodetic accuracy. Many experiments with diverse maps are needed to determine how well the method generalises.

The use of pixel patches implies the assumption of using north-up oriented maps (other than slightly skewed scans). Rotation-invariant feature descriptors allow the processing of maps with rotation. The robustness of our matching method has to be evaluated for the thus greatly extended search space.

Especially when looking at different map types and scales, detecting other symbols and using additional phenomena for more content information are within reach. The discovered differences in performance between our method and the state-of-the-art support that using different sources of information can make alignment more robust and more accurate. The main difficulty with the use of multiple sources of spatial information, however, is the lack of clarity concerning how to combine different sets of control points in a single transformation solution, especially when they have differing uncertainty. It becomes apparent that the classic transformation and warping methods used in geoinformation systems are not sufficient to handle raster georeferencing under uncertainty, such as noisy ground control points.

Iterative image registration, such as ECC (as described in Section 2.4), is susceptible to premature convergence to local minima. Better, more sophisticated segmentation might give cleaner images that converge better during

registration. The currently used segmentation method is quite naive; it worked sufficiently well for our data set but is far from state-of-the-art in computer vision. When working on maps with more colours, especially with half-toning, more sophisticated segmentation methods, such as superpixel-based methods which have been successfully applied to satellite imagery; see, (Seppke, Dreschler-Fischer, & Wilms, 2016) are promising. Some of our experiments used older editions of the KDR100 with irregular colouring or degraded paper. These experiments showed that a more sophisticated method, using more local contextual knowledge, will be helpful to avoid hand-picked segmentation thresholds for different data sources.

In this article we applied relatively simple methods with promising results. To widen the applicability of our approach, we focus our ongoing research on relaxing the need for prior knowledge, most prominently the map sheets' coordinates. To truly exploit the CBIR approach, viable sheet bounding boxes need to be generated automatically. Possible approaches for generalisation are either training a machine learning model on randomised synthetic data or spatial tree-like look-up structures. Traversing the immensely larger search space of possible locations and bounding box dimensions is only feasible with pre-computed lookup structures of the reference data, employing cutting-edge CBIR techniques for high-performance databases and efficient look-up.

For that, it is necessary to implement a more efficient but still accurate and robust feature matching method. First experiments with classic CV descriptors (e.g., SURF; Bay et al., 2008) were not promising for the use in topographic maps; instead, water-filling features (Zhou et al., 1999) have been proposed by Michelson et al. (2008) for map classification. Still, a hierarchical matching method, possibly in scale space, is necessary for an efficient coarse-to-fine search through a large number of possible locations.

ACKNOWLEDGMENTS

Open Access funding enabled and organized by Projekt DEAL.

ORCID

Jonas Luft  <https://orcid.org/0000-0003-1380-182X>

ENDNOTES

¹ The API for querying OSM data; www.overpass-api.de.

² Available in the public domain at [de.wikipedia.org/wiki/Karte_des_Deutschen_Reiches_\(Generalstabskarte\)](http://de.wikipedia.org/wiki/Karte_des_Deutschen_Reiches_(Generalstabskarte))

REFERENCES

- Alvares-Sanches, T., Osborne, P. E., James, P., & Bahaj, A. S. (2020). Tracking a city's center of gravity over 500 years of growth from a time series of georectified historical maps. *Cartography and Geographic Information Science*, 47, 524–536.
- Bay, H., Ess, A., Tuytelaars, T., & van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346–359.
- Bessel, F. W. (1841). Über einen Fehler in der Berechnung der französischen Gradmessung und seinen Einfluß auf die Bestimmung der Figur der Erde. *Astronomische Nachrichten*, 19(438), 97–116.
- Burt, J. E., White, J., Allord, G., Then, K., & Zhu, A.-X. (2020). Automated and semi-automated map georeferencing. *Cartography and Geographic Information Science*, 47(1), 40–66.
- Carmichael, G., Laganière, R., & Bose, P. (2010). Global context descriptors for SURF and MSER feature descriptors. In *Proceedings of the 2010 Canadian Conference on Computer and Robot Vision*, Ottawa, Ontario, Canada (pp. 309–316). Piscataway, NJ: IEEE.
- Chen, I.-W., Chen, H.-R., & Tseng, Y.-H. (2016). Automatic image matching and georeferencing of digitized historical aerial photographs. In *Proceedings of the 37th Asian Conference on Remote Sensing*, Colombo, Sri Lanka (pp. 1100–1108).
- Chiang, Y.-Y. (2015). Querying historical maps as a unified, structured, and linked spatiotemporal source: Vision paper. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, WA (pp. 1–14). New York, NY: ACM.

- Chiang, Y.-Y., Leyk, S., & Knoblock, C. A. (2014). A survey of digital map processing techniques. *ACM Computing Surveys*, 47(1), 1.
- CIE International Commission on Illumination. (2019). *Colorimetry—Part 4: CIE 1976 L*a*b* colour space* (No. ISO/CIE 11664-4:2019 in Colours and measurement of light). Geneva, Switzerland: International Organization for Standardization.
- Cléri, I., Pierrot-Deseilligny, M., & Vallet, B. (2014). Automatic georeferencing of a heritage of old analog aerial photographs. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3), 33–40.
- Conte, D., Foggia, P., Sansone, C., & Vento, M. (2004). Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(8), 265–298.
- Crom, W. (2016). Kartendigitalisierung—buntes Bild oder Mehrwert. *KN-Journal of Cartography and Geographic Information*, 66(5), 243–248.
- Dorn, H., Törnros, T., & Zipf, A. (2015). Quality evaluation of VGI using authoritative data—A comparison with land use data in southern Germany. *ISPRS International Journal of Geo-Information*, 4(3), 1657–1671.
- Evangelidis, G., & Psarakis, E. (2008). Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10), 1858–1865.
- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.
- Fleet, C., Kowal, K. C., & Pridal, P. (2012). Georeferencer: Crowdsourced georeferencing for map library collections. *D-Lib Magazine*, 18(11/12). Retrieved from <http://www.dlib.org/dlib/november12/fleet/11fleet.html>
- Heitzler, M., Gkonos, C., Tzorlini, A., & Hurni, L. (2018). A modular process to improve the georeferencing of the Siegfried map. *e-Perimtron*, 13, 85–100.
- Hermann, K. L., Chen, T., & Kornblith, S. (2020). *The origins and prevalence of texture bias in convolutional neural networks*. Preprint, arXiv:1911.09071.
- Howe, N. R., Weinman, J., Gouwar, J., & Shamji, A. (2019). Deformable part models for automatically georeferencing historical map images. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, IL (pp. 540–543). New York, NY: ACM.
- Höhn, W., & Schommer, C. (2017). Georeferencing of place markers in digitized early maps by using similar maps as data source. *Digital Humanities 2017 Conference Abstracts*, Montréal, Canada.
- Janata, T., & Cajthaml, J. (2020). Georeferencing of multi-sheet maps based on least squares with constraints: First military mapping survey maps in the area of Czechia. *Applied Sciences*, 11, 299.
- Koffka, K. (1935). *Principles of gestalt psychology*. London, UK: Lund Humphries.
- Kruse, C., Neuberger, H., Rottensteiner, F., Hoberg, T., Ziem, M., Huth, J., & Heipke, C. (2017). Automatische Detektion von Bombenkratern in Kriegsflugbildern mittels markierter Punktprozesse. *Publikationen der DGPF*, 26, 245–261.
- Lewis, J. P. (2001). *Fast normalized cross-correlation*. Retrieved from <http://www.scribblethink.org/Work/nvisionInterface/nip.pdf>
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
- Luft, J. (2020). Automatic georeferencing of historical maps by geocoding. In *Proceedings of the Automatic Vectorisation of Historical Maps: International Workshop Organized by the ICA Commission on Cartographic Heritage into the Digital*, Budapest, Hungary (pp. 77–89).
- MacEachren, A. M., & Mistrick, T. A. (1992). The role of brightness differences in figure-ground: Is darker figure? *Cartographic Journal*, 29(2), 91–100.
- Michelson, M., Goel, A., & Knoblock, C. A. (2008). Identifying maps on the world wide web. In T. J. Cova, H. J. Miller, K. Beard, A. U. Frank, & M. F. Goodchild (Eds.), *Geographic information science: GIScience 2008* (Lecture Notes in Computer Science, Vol. 5266, pp. 249–260). Berlin, Germany: Springer.
- Mooney, P., & Corcoran, P. (2012). The annotation process in OpenStreetMap. *Transactions in GIS*, 16(4), 561–579.
- Papakosta, C., Ioannidis, C., & Vassilaki, D. (2012). The use of linear features as ground control information for the georeferencing of old aerial photos. In *Proceedings of FIG Working Week 2012*, Rome, Italy (pp. 1–12).
- Reichsamt für Landesaufnahme. (1931). *Das Reichsamt für Landesaufnahme und seine Kartenwerke* (Chap. XIX: Die Karte des Deutschen Reiches 1:100000). Berlin, Germany: Verlag des Reichsamtes für Landesaufnahme
- Rosten, E., & Drummond, T. (2005). Fusing points and lines for high performance tracking. In *Proceedings of the 10th IEEE International Conference on Computer Vision*, Beijing, China (Vol. 1, pp. 1508–1515). Piscataway, NJ: IEEE.
- Schlegel, I. (2019). Empirical study for a deployment of a methodology for improving the comparability between historical and current maps. *KN-Journal of Cartography and Geographic Information*, 69(2), 121–130.
- Schröder, J. (2013). Georeferenzierung und Entzerrung der brandenburgischen Kartenblätter des Schmettauschen Kartenwerks. *Vermessung Brandenburg*, 18, 14–23.

- Seppke, B., Dreschler-Fischer, L., & Wilms, C. (2016). A robust concurrent approach for road extraction and urbanization monitoring based on superpixels acquired from spectral remote sensing images. In *Proceedings of the DLT DGPF Symposium*, Stuttgart, Germany (pp. 1–5).
- Sharma, G., & Bala, R. (2003). Uniform color spaces and color differences. In G. Sharma & R. Bala (Eds.), *Digital Color Imaging Handbook* (pp. 28–39). Boca Raton, USA: CRC Press.
- Smeulders, A., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349–1380.
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(14), 600–612.
- Wolter, D., Blank, D., & Henrich, A. (2017). Georeferencing river networks using spatial reasoning. In *Proceedings of the 11th Workshop on Geographic Information Retrieval*, Heidelberg, Germany (pp. 1–2). New York, NY: ACM Press.
- Zhou, X. S., Rui, Y., & Huang, T. (1999). Water-filling: A novel way for image structural feature extraction. In *Proceedings of the 1999 International Conference on Image Processing*, Kobe, Japan (Vol. 2, pp. 570–574). Piscataway, NJ: IEEE.

How to cite this article: Luft, J., & Schiewe, J. (2021). Automatic content-based georeferencing of historical topographic maps. *Transactions in GIS*, 25, 2888–2906. <https://doi.org/10.1111/tgis.12794>