Automatic Georeferencing of Map Images Using Unsupervised Learning and Graph Analysis

Enrique J. Arriaga-Varela Stroly Inc. Kyoto, Japan enrique@stroly.jp Toru Takahashi Stroly Inc. Kyoto, Japan toru@stroly.jp

Abstract

We present a novel method for the automatic georeferencing of heterogeneous map images based on the analysis of the spatial relationships between their lines of text and the geographical locations they depict. Our approach differs from previous work in that the only input provided is the raster image, as it does not require additional hints or metadata. The method is also designed to be highly tolerant of maps with different art styles, scales, orientations, and cartographic projections. To accomplish this task, we leverage the power of modern OCR (Optical Character Recognition) and geocoding services to generate a series of candidate ground control points (GCP) and then discriminate between them using a combination of clustering algorithms and graph analysis. Experimental results for 359 map images demonstrate the viability of the proposed method. We achieved a precision ranging from 81.19% to 97.56% and a recall from 55.71% to 71.15%.

1 Introduction

The process of georeferencing, which means aligning digitized maps with their geographical coordinates, is a difficult and time-consuming task [10] [12] [13]. It is generally accepted that the level of understanding necessary to georeference an image correctly can be rather daunting [1] [4]. Over the last decades, there have been advances in the automatic georeferencing of map images and aerial photographs [5], [7], [12], [6]. However, there has been little discussion on developing methods that can deal with heterogeneous maps. For instance, some algorithms apply fixed image processing techniques to find features and match them with geographical databases [12] [5]. The drawback of these approaches is that such image processing may not work for other styles of images. Other techniques only work with specific kinds of images like street maps [7] or aerial photographs [12]. Finally, historical or contemporary illustrated maps may suffer from distortion or unusual perspectives or projections that can make the match with precise geographical information considerably harder to achieve [4] [8].

2 Our Method

The proposed method contains four modules, as shown in Figure 1. The only input by the user is the raster image, and the output is the georeference information or null if it could not be found.

2.1 Creation of Candidate Ground Control Points (GCP)

The OCR module obtains text annotations and sends them to the next module to be queried in a geocoding service that converts location names into geographic coordinates. The results of both are joined to create a series of ground control points (GCP). Each GCP is a match between the

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.



Figure 1: Overview of the architecture of the proposed method

coordinates (X, Y) of a point in the map image and the (Longitude, Latitude) coordinates of the location on earth the point is depicting.

2.2 Discrimination of Candidate GCP using unsupervised learning and graph analysis

The list of candidate GCP certainly contains spurious elements caused by the natural ambiguity of place and landmark names [11] and OCR or geocoding inaccuracies [14]. The third module analyzes the relationships between pairs and clusters of GCP to find uniform groups. The distance between each possible pair of GCP is expressed in two components: scale (meters per pixel) and the difference in orientation (angle they form). Agglomerative hierarchical clustering (AHC) ([9], [16]) is used to cluster these distances in groups with low variance in scale and orientation. To obtain a group of GCP from clusters of distances, a connection graph is created, and the maximum clique of this graph is calculated ([3], [2]). Each clique is a set of GCP in which all their distances have similar values of scale and orientation. It is expected that larger cliques and lower variances increase the chance of having found correct georeference information. The final module calculates the bounding box (NW and SE) and the orientation of the map using the best clique of GCP and the map dimensions in pixels. If no clique of a minimum size (defined as hyper-parameter) was found, the result is null.

3 Results

Minimum Size of the Selected Clique	3 GCP	4 GCP	5 GCP	6 GCP
Georeference Information Found	319	265	226	205
Correct Information (IoU >= 0.20)	259	242	219	200
Incorrect Information (IoU < 0.20)	60	23	7	5
Precision	81.19%	91.32%	96.90%	97.56%
Recall	71.15%	67.41%	61%	55.71%

Table 1: Results for 359 maps with different values of the minimum accepted sizes of GCP clique

A data set of 359 pictorial, historical, and hand-drawn maps provided by Stroly Inc. [15] was used for the computer experiments. The images were manually georeferenced by a human expert, and this bounding box is used as ground truth and compared with the prediction by the presented method. In order to calculate the Intersection over Union (IoU), both bounding boxes were converted into a projected coordinate system. The result is considered correct when the IoU with the ground truth is IoU >= 0.20. Table 1 shows the results of four runs of the method, each with a different value of the hyperparameter minimum size of the selected clique of GCP. In all cases, the only input was the raster image.

4 Conclusion

We developed a novel solution to georeference heterogeneous map images with no need for additional input from the user, hand-crafted image processing, nor access to special geographical databases. Our experiments have proven that even with small but carefully calculated sets of GCP, the precision achieved is greater than 80%, and it can reach almost 98% for larger sets.

The method showed limitations in cases where the text found on the map is sparse or difficult to be recognized by OCR. Similar shortcomings were encountered on maps where the text content or position within the image does not match the geographical locations depicted.

Broader Impact

This section will describe a couple of scenarios in which we believe the proposed method could be applied.

In the first one, an implementation could be executed in batch for large collections of map images without georeference information. Such map collections are usually found in museums, libraries, or universities. In this scenario, it would be possible to obtain important statistics about the maps in the collection. These statistics include the scale and orientation of the maps, as well as the countries, regions, or territories they depict. This information would increase the cultural value of each map and of the collection as a whole. Additionally, as we have described in this work, the batch processing could be achieved with no additional input from the user and in a time and resource-efficient manner.

In the second scenario, the proposed method could be executed individually in the context of a web system similar to the one described in [15]. In such systems, users can upload their own maps and manually perform the georeferencing using web tools. The inclusion of this tool in the workflow could dramatically improve the user experience with the automatic or semi-automatic creation of ground control points. This is especially true because a great percentage of the users of such websites are not experts in geographical information tools or systems. Additionally, if the use case requires professional-level georeferencing, the GCP and the geographical bounding box returned by this tool would be a starting point to significantly speed up the process.

References

- BAJCSY, P., AND ALUMBAUGH, T. J. Georeferencing maps with contours. In *Journal Proceedings of 7th World Multiconference on Systemics, Cybernetics, and Informatics. Vol X* (2003), pp. 27–30.
- [2] BOPPANA, R., AND HALLDÓRSSON, M. M. Approximating maximum independent sets by excluding subgraphs. *BIT Numerical Mathematics* 32, 2 (1992), 180–196.
- [3] BRON, C., AND KERBOSCH, J. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM 16*, 9 (1973), 575–577.
- [4] CAJTHAML, J. Methods of georeferencing old maps on the example of czech early maps. In *Proceedings of the 25th International Cartographic Conference. Paris: International Cartographic Association* (2011).
- [5] CHEN, C.-C., KNOBLOCK, C. A., SHAHABI, C., CHIANG, Y.-Y., AND THAKKAR, S. Automatically and accurately conflating orthoimagery and street maps. In *Proceedings of the* 12th annual ACM international workshop on Geographic information systems (2004), ACM, pp. 47–56.
- [6] CLÉRY, I., PIERROT-DESEILLIGNY, M., AND VALLET, B. Automatic georeferencing of a heritage of old analog aerial photographs. *Photogrammetric Computer Vision 2*, 3 (2014), 33.
- [7] DESAI, S., KNOBLOCK, C. A., CHIANG, Y.-Y., DESAI, K., AND CHEN, C.-C. Automatically identifying and georeferencing street maps on the web. In *Proceedings of the 2005 workshop* on *Geographic information retrieval* (2005), ACM, pp. 35–38.
- [8] FIORI, S. R. Touristic maps: the challenge of using art in the digital era. *XXII International Cartographic* (2005).
- [9] FISHER, W. D. On grouping for maximum homogeneity. *Journal of the American statistical Association 53*, 284 (1958), 789–798.
- [10] FLEET, C., KOWAL, K. C., AND PRIDAL, P. Georeferencer: Crowdsourced georeferencing for map library collections. *D-Lib magazine 18*, 11/12 (2012).
- [11] GRITTA, M., PILEHVAR, M., AND COLLIER, N. Which melbourne? augmenting geocoding with maps. ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers) 1 (2018), 1285–1296.

- [12] KIM, J. S., MILLER, C. C., AND BETHEL, J. Automated georeferencing of historic aerial photography. *Journal of Terrestrial Observation* 2, 1 (2010), 6.
- [13] KIRÁLY, G., WALZ, U., PODOBNIKAR, T., CZIMBER, K., NEUBERT, M., AND KOKALJ, Ž. Georeferencing of historical maps-methods and experiences. Spatial Information Systems for Transnational Environmental Management of Protected Areas and Regions in the Central European Space. Selected Results and Outputs of the Interreg IIIB Project SISTEMaPARC, Rhombos-Verlag, Berlin (2008), 53–63.
- [14] POULIQUEN, B., KIMLER, M., STEINBERGER, R., IGNAT, C., OELLINGER, T., BLACKLER, K., FUART, F., ZAGHOUANI, W., WIDIGER, A., FORSLUND, A.-C., ET AL. Geocoding multilingual texts: Recognition, disambiguation and visualisation. arXiv preprint cs/0609065 (2006).
- [15] VERMEULEN, H., TAKAHASHI, T., TAKAHASHI, M., OHTSUKA, K., NAKAGAWA, T., AND UEDA, H. Stroly: A historic and illustrated maps platform. In *Culture and Computing (Culture Computing)*, 2011 Second International Conference on (2011), IEEE, pp. 195–196.
- [16] WARD JR, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 58, 301 (1963), 236–244.

A A visualization of the results and comparison with the ground truth

This section presents some examples of successful and unsuccessful results obtained by executing the proposed method for the data set of 359 map images discussed in the results section.

Figure 2 shows the visualization of the predictions obtained for maps with different styles and how this prediction compares with the ground truth. In each of the four cases, the input image is shown on the left side, and the ground truth and results are plotted over a web map on the right side. The predicted bounding box is plotted on blue, while the ground truth (georeferenced by a human expert) is plotted on red. The GCP selected in the best clique are numbered and drawn on both sides in a white square with a red outline. On the left side each GCP is located in the (X, Y) coordinates of the text annotation on the raster image. On the right side, they are drawn on the (*Longitude*, *Latitude*) returned by the geocoding service. We can see how the match (or *IoU*) between the prediction and the ground truth is better for larger cliques of GCP.



Figure 2: Examples of successful results (IoU with ground truth >= 0.20) of the proposed method. Top-left: JTB, Top-right: Globe-trotter (CC-BY-SA3.0), Bottom-left: Suiko, Hagiwara Yoshiki Office, Bottom-right: Maizuru City.

Figure 3 shows some examples of map images for which the method could not find a bounding box. In most cases, this is because there is not enough text, or it is difficult to be recognized by current OCR methods. In other instances, the text may not include the current name of real locations on earth, or it may not be placed on the (X, Y) coordinates of the location it is depicting.



Figure 3: Examples of non-successful results (no bounding box predicted) of the proposed method. From left to right: Sendai Station East Area Management Council, Ritsumeikan University, mapping prun design workshop B-team, public domain.

B Implementation Details

B.1 Technological considerations

In this work, we outline a method with a series of steps that are independent of the underlying technology used to achieve them. For instance, the OCR step can be performed by any program that is able to return the precise coordinates (X, Y) of each text annotation. This includes open-source alternatives such as Tesseract OCR¹ or commercial ones like Google Vision API². The geocoding step can also be done with open-source and open-data alternatives, such as OpenStreetMap's Nominatim³ or its commercial alternative, MapQuest⁴. In the same manner, the clustering and graph analysis can be easily performed using scientific open-source packages.

B.2 Execution time

The time necessary to execute the method for a single map heavily depends on the characteristics of the image and the underlying technologies used to achieve each step. For instance, while the OCR step usually takes just a few seconds, the geocoding of all the text annotation can take several minutes. This is especially true if the used geocoding service does not allow querying all the annotations at the same time. To overcome this possible problem, we have developed a fork⁵ of OpenStreetMap's Nominatim to allow the batch geocoding of any number of lines of text with a single HTTP request.

Using batch geocoding and other optimizations outside of the scope of the current document, we have created an implementation in which each map image takes an average of 6 seconds, with a maximum of 31 seconds and a minimum of 0.9 seconds. The system was tested on an Amazon AWS lambda⁶ serverless architecture.

¹https://github.com/tesseract-ocr/tesseract

²https://cloud.google.com/vision

³https://nominatim.org/ (Note: bulk geocoding on OpenStreetMap's instance is against the usage policy) ⁴https://developer.mapquest.com/

[&]quot;https://developer.mapquest.com/

⁵https://github.com/StrolyCom/Nominatim/tree/high_performance_batch

⁶https://aws.amazon.com/lambda/