Unlocking Textual Content from Historical Maps - Potentials & Applications, Trends, and Outlooks

Yao-Yi Chiang

Spatial Sciences Institute,
University of Southern California
Los Angeles, California 90089, USA
yaoyic@usc.edu

Abstract. Digital map processing has been an interest in the image processing and pattern recognition community since the early 80s. With the exponential growth of available map scans in the archives and on the internet, a variety of disciplines in the natural and social sciences grow interests in using historical maps as a primary source of geographical and political information in their studies. Today, many organizations such as the United States Geological Survey, David Rumsey Map Collection, OldMapsOnline.org, and National Library of Scotland, store numerous historical maps in either paper or scanned format. Only a small portion of these historical maps is georeferenced, and even fewer of them have machine-readable content or comprehensive metadata. The lack of a searchable textual content including the spatial and temporal information prevents researchers from efficiently finding relevant maps for their research and using the map content in their studies. These challenges present a tremendous collaboration opportunity for the image processing and pattern recognition community to build advance map processing technologies for transforming the natural and social science studies that use historical maps. This paper presents the potentials of using historical maps in scientific research, describes the current trends and challenges in extracting and recognizing text content from historical maps, and discusses the future outlook.

Keywords: digital map processing, text recognition, optical character recognition, historical maps, geographic information system, natural science, social science, biology, spatial humanity

1. Introduction

Historical maps are an irreplaceable primary source of geographical and political information in the past (e.g., historical place names, landmarks, natural features, transportation networks, and war, trade, and diplomacy networks). The image processing and pattern recognition community started to develop computational methods for the extraction and recognition of the content from archived images of maps since the early 80s (Chiang, Leyk, & Knoblock, 2014). With the exponential growth of available map scans in the archives and on the internet, a variety of

disciplines in the natural and social sciences grow interests in using historical maps in their studies. For example, the Mappa Mundi by Fra Mauro (ca. 1450) (Figure 1) contains not only place names but also provides "*natural philosophy, description of places and people, commercial geography, history, navigation and direction of expansion, and, finally, on what we can nowadays call methodological issues. In addition, Fra Mauro's world map also includes hundreds of images, representing cities, temples, funerary monuments, streets, and ships, as well as a scene in the lower left corner representing Earthly Paradise*" (Nanetti et al., 2015).

Figure 1. East Asia Mainland of the Mappa Mundi (ca. 1450), Fra Mauro

In many cases, historical maps are the only source that provides professionally surveyed historical geographic data. Map archives such as the U.S. Geological Survey (USGS) National Geologic Map Database,[1] USGS Topographic Maps,[2] David

---

Rumsey Map Collection,[3] OldMapsOnline.org,[4] and the National Library of Scotland,[5] together store millions of this type of historical map in either paper or scanned format. For example, between 1884 and 2006, the USGS has created over 200,000 topographic maps. According to the USGS, in the United States these topographic maps "*portray both natural and manmade features. These maps show and name works of nature including mountains, valleys, plains, lakes, rivers, and vegetation. They also identify the principal works of man, such as roads, boundaries, transmission lines, and major buildings.*" The USGS National Geospatial Program has scanned these historical paper maps. Collectively, these publicly available scanned maps portray the evolution of the American landscape over a 125-year period.[6] Similar map series exist in many countries, e.g.; the Ordnance Survey maps in the U.K archived by the National Library of Scotland.

In the case of more recent historical maps produced using modern geospatial survey technologies (e.g., the USGS Topographic Map series, Ordnance Survey six-inch series, and other national agency series dated from the early 1800), the detailed map data on the states of landscapes in the past are essential for understanding the causes and consequences of environmental change and support a variety of natural and social studies on topics such as cancer and environmental epidemiology, urbanization, biodiversity, landscape changes, and history. (See Gregory et al., 2015 for more examples and methodologies in historical geographic information systems.) Many of these historical maps are not georeferenced, and almost all of the maps have content that is not machine-readable. Existing map processing technologies are still limited in making a large number of historical maps fully searchable by their content because the archived documents often suffer from bleaching, blurring, and false coloring (e.g., Khotanzad & Zink, 2003; Leyk & Boesch, 2009; Leyk et al., 2006). The reader is referred to Chiang, Leyk, & Knoblock (2014) and Chiang et al. (2016) for detailed reviews and case studies on map processing techniques and systems.

Today, a researcher can spend a great deal of time and effort searching and cross-referencing data sources to find relevant maps. Then they need to digitize the map for converting the map content to a machine-readable format (e.g., Godfrey & Eveleth, 2015; Nanetti et al., 2015). The researcher may need to search in various publication repositories, map repositories, search engines, and then they will often not find the historical map that they are looking for and work without it. In many cases, these historical maps exist, and it just requires too much effort to locate and digitize them. The result is that researchers waste time and resources and do not get as far as

---

[2] USGS topoView. (2016). [Website]. Retrieved from http://ngmdb.usgs.gov/maps/TopoView/
[3] David Rumsey. (2016). [Website]. Retrieved from http://www.davidrumsey.com/
[4] OldMapsOnline. (2016). [Website]. Retrieved from http://www.oldmapsonline.org/
[5] NLS. (2016). [Website]. Retrieved from http://maps.nls.uk/
[6] USGS topoView. (2016). [Website]. Retrieved from http://ngmdb.usgs.gov/maps/TopoView/

they could have in their work because the relevant information is not discoverable or takes too long to prepare for scientific analysis.

The challenges in working with historical maps present an enormous collaboration opportunity for the image processing and pattern recognition community to build advance map processing technologies for transforming the scientific studies that currently use textual content in historical maps. Therefore, it is important to understand the current landscape in the broad applications of historical maps. This paper first describes the potentials and current applications of historical maps in a variety of studies, including topics in natural science (Section 2) and social science (Section 3). Next, the paper describes the current trend in extracting and recognizing textual content from historical maps (Section 4). Finally, the paper discusses the future outlook in text recognition technologies in map processing (Section 5).

2. Potentials and Applications of Historical Maps in Natural Science

Historical data archives (e.g., museum and herbaria collections, digital photography and newspaper archives) support a variety of scientific studies in natural science on topics such as biodiversity (e.g., Hill et al., 2009), evolutionary biology (e.g., Lavoie, 2013), human disease (e.g., Yoshida et al., 2014), plant biology (Davis et al., 2015; Vellend et al., 2013), and ecology (e.g., Newbold, 2010; Pyke & Ehrlich, 2010), but geolocating the historical localities mentioned in archives (e.g., Calflora Observation Database[7] and the Global Biodiversity Information Facility; Samy et al., 2013) is challenging and very often a tedious manual process using historical maps. Murphey et al. (2004) reviewed the problems in georeferencing museum collections. They compared a number of geoparsing tools including the GEOLocate (Rios & Bart, 2010) and BioGeomancer (Guralnick et al., 2006). Since then, a variety of advanced algorithms for geoparsing has been proposed (e.g., Leidner & Lieberman, 2011) and open-source software packages (e.g., CLAVIN,[8] CLIFF (D'Ignazio, Bhargava, & Zuckerman, 2014)), and the Edinburgh Geoparser (Alex et al., 2015) are available. These algorithms and tools are widely used in geolocating places in the unstructured text and also used in spatial humanities research (e.g., Gregory et al., 2015). However, these tools need a "gold data" gazetteer to provide the location information of recognized place names, and the lack of historical reference gazetteers remains a challenge. The result is that even if the geoparsing software can correctly identify a historical name as a geolocation reference in the unstructured text, the geocoordinates of the historical name is still unknown if the place name no longer exists. To locate

the place names that no longer exist in contemporary data sources, a researcher needs to search and cross-reference a variety of data sources such as archives of historical maps, newspapers, and photography.

For example, a data record in an online database of California herbarium specimens describes an August 16th, 1902 observation of *Artemisia douglasiana* (California mugwort) at the location "near Mesmer" in Los Angeles. The place name Mesmer near or within both the City and County of Los Angeles no longer exists in the contemporary geographic data sources, including authoritative sources like the U.S. Census[9] and USGS GNIS (the United States Geological Survey Geographic Names Information System)[10] and open sources, such as GeoNames,[11] OpenStreetMap,[12] and Wikipedia. Searching "Mesmer" in the GeoNames gazetteer results in an airport "Mesmer Airport" in New York and a street "Rue Mesmer" in Haiti. Neither of the results helps to geolocate the observation of California mugwort in 1902. A Google search with the keywords "Mesmer" and "Los Angeles" reveals a few interesting facts that could be helpful for geolocating Mesmer. First, the search results include a person, Louis Mesmer (1829 – 1900), who was a prominent businessman and the owner of the famous United States Hotel in Los Angeles. Because it was common to name locations after well-known families (e.g., Wilshire, Hancock, and Doheny in Southern California), Mesmer could be a place name in the Los Angeles area in the past. Second, the search results contain a link to a map in the Los Angeles Public Library collections showing a proposed development plan in 1924 for the "Mesmer City" in Los Angeles (Figure 2). At the time, Mesmer City was advertised as "*In the direct path of the Los Angeles' growth toward the ocean.*".[13] This map further narrows down the search space for Mesmer to somewhere nearby Culver City and Baldwin Hills. Together, the time and location information from the search results points to the USGS topographic map that contains the Mesmer in 1901 (Figure 3). In this case, Mesmer is geolocated, but the entire process cannot be scaled to handle thousands of records in an efficient manner.

[9] U.S. Census Gazetteer. (2016). [Data set]. Retrieved from https://www.census.gov/geo/maps-data/data/gazetteer.html

[10] USGS GNIS. (2016). [Data set]. Retrieved from http://geonames.usgs.gov/

[11] GeoNames. (2016). [Data set]. Retrieved from http://www.geonames.org/

[12] OpenStreetMap. (2016). [Website]. Retrieved from https://www.openstreetmap.org/

[13] Los Angeles Public Library Map Collection. (2016). [Website]. Retrieved from https://www.lapl.org/collections-resources/visual-collections/map-collection

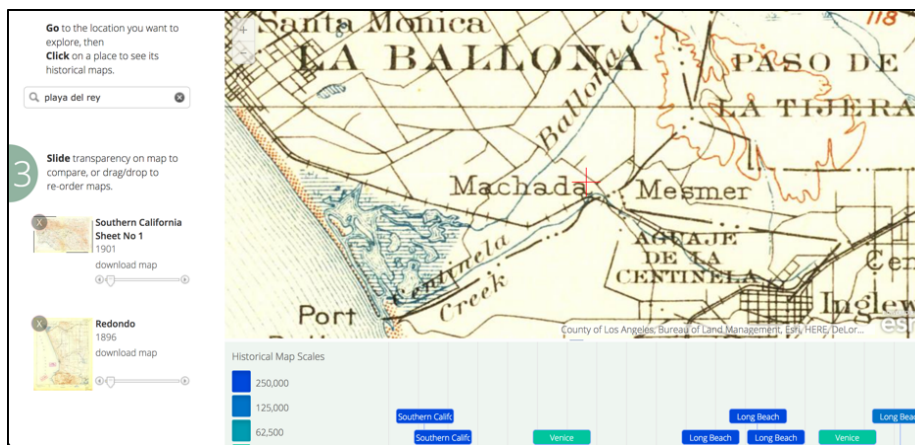**Figure 2.** Map of the Mesmer City development

Figure 3. The USGS Historical Topographic Map shows the location of Mesmer. (Southern California Sheet No. 1, circa 1901)

Historical GIS (Geographic Information System) (Gregory & Ell, 2007) could alleviate the problem of geolocating historical locality references by providing a platform for collecting datasets of historical place names, but the datasets are rarely available. Even when historical gazetteers are available, their spatiotemporal coverage is often sparse. For example, the U.S. Census only provides post-2010 and also 2000 and 1990 census gazetteer files. NHGIS (the National Historical Geographic Information System at the Minnesota Population Center)[14] provides historical demography data down to the census tract level but only a few place names. The Ramsay Place Names File from the State Historical Society of Missouri provides a historical gazetteer covering locations in the State of Missouri from 1928 to 1945 (Adams, 1928). The website "A Vision of Britain through Time" from the GB

---

[14] NHGIS. (2016). [Website]. Retrieved from https://www.nhgis.org/

Historical GIS at the University of Portsmouth[15] provides historical place names in the Great Britain dated back in the early 19th century.

As shown in the examples of natural science studies in this section, the ability to automatically use the textual content in historical maps as the locality reference source will be able to transform historical data records in documents and collections into georeferenced datasets. This ability will enable natural science researchers to efficiently find, query, and analyze a variety of historical records by location.

3. Potentials and Applications of Historical Maps in Social Science

Historical maps also play an important role in social science studies. Kurashige (2013)[16] used historical census data, voting records, and precinct numbers and boundaries extracted from a 1920 map to study "who" (e.g., occupations and political parties) in Los Angeles voted for the 1920 California Alien Land Law that discriminates against Japanese (Figure 4).
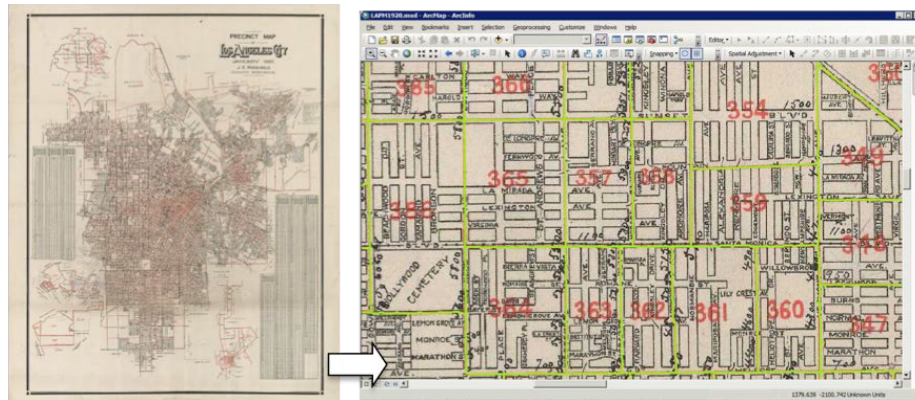

Figure 4. Automatically unlocking precinct boundaries in a historical map for analyzing historical voting records with demographic datasets.

Ngo, Swift, & Chiang (2015) used historical maps and land records to build an interactive visualization of land reclamation in Hong Kong (Figure 5). This web tool[17] is among the top hits when searching Hong Kong land reclamation on Google.

---

[15] A Vision of Britain through Time. (2016). [Website] Retrieved from http://www.visionofbritain.org.uk/

[16] Dr. Kurashige's article published in the Southern California Quarterly won the 2015 Carl I. Wheat Award for the best demonstration of scholarship in that journal from 2012-2014 by a senior historian.

[17] http://www.oldhkphoto.com/coast/
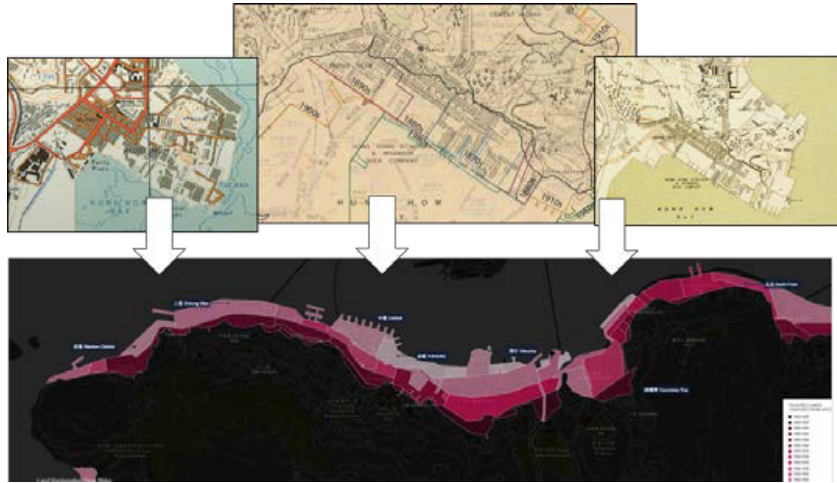
Figure 5. Building an interactive visualization of land reclamation in Hong Kong from historical maps.

The Spatial Sciences Institute at the University of Southern California (USC) is collaborating with an insurance company to automatically read historical Ordnance Survey maps (ca. 1900 – 1970) covering the entire U.K. to identify likely locations of subterranean contamination, such as factories, mines, quarries, and gas works which no longer exist and otherwise would not be known today (Figure 6).[18]
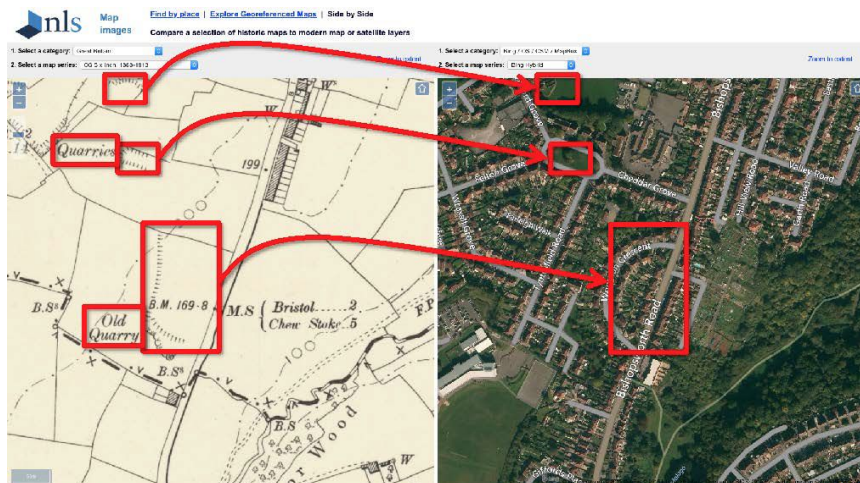

Figure 6. Quarries and infill lands in a historical map and the contemporary satellite imagery.

---

[18] Spatial technology opens a window into history . (2016). [News article]. Retrieved from https://news.usc.edu/91625/spatial-technology-opens-a-window-into-history/

In a joint effort, the USC Shoah Foundation Visual History Archive (VHA) works with the USC Spatial Sciences Institute to link historical maps to places mentioned in genocide survivor testimonies in the VHA archive. The linkages enrich the personal stories of the survivors by using the spatial and temporal context in historical maps to enable the viewers to "go back in time" to recreate the physical world of the historical experience of the survivors (Figure 7).[19]

Nanetti et al. (2015) manually transcribed and georeferenced the textual content in the *Mappa Mundi by* Fra Mauro (ca. 1450). They use the transcribed data as a knowledge aggregator to represent the world as seen from Venice in the fifteenth centuries. They also plan to use the map data for automatic provenance and validation assessment of large and heterogeneous collections of other historical sources.
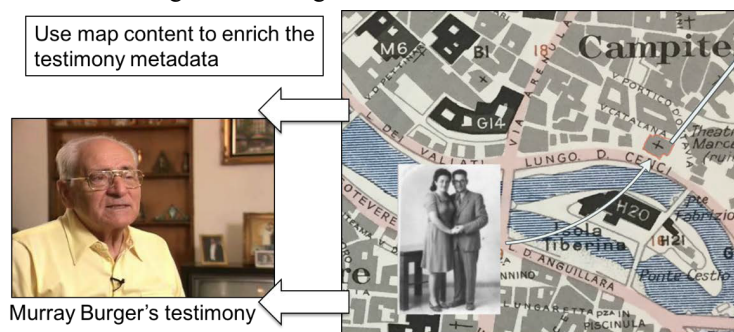
Figure 7. Using historical maps to identify the wedding location (historical synagogue) described by Holocaust survivor Murray Burger.

The example studies in this section demonstrate the power of historical maps in social science research. They also show that not only extracting and recognizing map content is important, but providing semantic annotations to the map content and linking the map content to other data sources will enable researchers to investigate complex social science problems at a scale that cannot be done today.

4. Current Trends in Text Recognition from Historical Maps

Text recognition from maps is a difficult task, especially for historical maps. This is because map labels often overlap with other map features, such as road lines, do not follow a fixed orientation within a map, and can be stenciled and handwritten text (Chiang et al., 2014, 2016; Nagy et al., 1997). Also, many historical scanned maps suffer from poor graphical quality due to bleaching of the original paper maps and

archiving practices. This section presents a number of trends in text recognition from historical maps.

Classic Approaches Traditional, the approaches on text recognition from historical maps follow the classic document recognition strategy that first analyze the map to identify the potential text areas. Then the approaches use optical character recognition (OCR) algorithms or tools (e.g., TesseractOCR)[20] to convert the detected areas into machine-readable text (Honarvar Nazari et al., 2016; Chiang & Knoblock, 2014; Li et al., 2000; Pezeshk & Tutwiler, 2010, 2011; Raveaux et al., 2007, 2008; Li et al., 2000). This line of work has demonstrated promising results on single map sheet or map series but still does not handle large numbers and various types of historical maps because of the significant heterogeneity of historical maps.

Crowdsourcing Approaches To handle the vast variety of historical map types and sources, recent work has adopted the crowdsourcing strategy. Crowdsourcing is not a new idea (but can be very difficult to implement and popularize) in document recognition and map processing. The David Rumsey Map Collection held crowdsourcing events to georeference their map collections by crowdsourcing. The New York Public Library provides semi-automatic tools for the public to extract parcel polygons from their historical insurance maps.[21] They also noted that even with the crowdsourcing approach, semi-automatic tools are required to process their map collections in a reasonable time (Arteaga, 2013). Specifically for converting textual content in historical maps using crowdsourcing, the Pelagios Commons[22] is a notable community that provides tools and online infrastructures to facilitate annotating historical locality references in digital materials. Their tools allow semi-automatic extraction, recognition, annotation, and linking of place names in historical maps (Simon et al., 2010, 2014, 2015). These tools and the online infrastructure allow them to provide full-text searchable place data ranges from ancient times to 1500 AD and from Europe to East Asia.

To make the crowdsourcing strategy more effective and efficient, it will be necessary to build adaptive semi-automatic techniques that improve the level of automation as more maps are processed. Also, as the crowdsourcing strategy is used, approaches for cross-validating between user generated content and the "gold data" as well as recording the provenance information are required (e.g., Garijo, Gil, & Harth, 2015).

Multi-Model Approaches Another line of work in the recent development of text recognition from historical maps uses additional data sources as the dictionary to help correct the recognition errors. While this dictionary strategy is common in OCR,

---

[20] Tesseract-OCR. (2016). [Computer software]. Retrieved from https://github.com/tesseract-ocr
[21] NYPL map-vectorizer. (2016). [Computer software]. https://github.com/NYPL/map-vectorizer
[22] Plageois Commons. (2016). [Website]. Retrieved from http://commons.pelagios.org/

compiling and effectively using a dictionary for recognizing historical map text is difficult. This is because a dictionary built using contemporary data sources does not contain place names that no longer exist. Also, without knowing the map coverage beforehand, multiple dictionary entries can match to a partially recognized label. For example, a partially recognized label "Glas wo" near London could be matched to "Glasgow" when the label is "Glassworks". Even the map coordinates are known, the map text might not be at the exact location of the geographic features depending on the cartographic labeling practice of the map. Weinman (2013) presents an approach that overcomes this challenge. His approach recognizes text labels in maps to then match the recognized text to a gazetteer by their position patterns using a RANSAC variant called MLESAC (Torr & Zisserman, 2000). He showed that this approach could automatically georeference historical maps and improve the recognition accuracy even when the gazetteer only contains 70% of the text in the test maps.

In previous work, we developed a semi-automatic approach that extracts and recognizes text labels in map images in a system called Strabo (Apache Version 2 License) (Chiang & Knoblock, 2014). While Strabo could achieve over 90% precision and recall in recognizing text labels in scanned contemporary maps, it could only produce 47.6% precision and 83.5% recall on well-conditioned text from historical Ordnance Survey six-inch maps (Chiang et al., 2016). The result is that very often only partial labels could be recognized from a historical map (Figures 9(a) and 9(b)) and manual post-processing is required to correct the recognition results.

In an effort to test higher levels of automation in text recognition from historical maps (Yu et al., 2016), we exploit the fact that geographic names for the same area found in different data sources is not independent and use geographic names in OpenStreetMap and other maps covering the same area as the "dependent" knowledge source. Given a historical map, the task at hand is to recognize all map labels in the map accurately without user intervention. First, the system queries a map repository to find all map editions covering the same area and then extracts and recognizes labels in the identified maps. Second, the system compares and uses a fuzzy matching algorithm to match the recognized (imperfect) labels using their locations and string similarity. Finally, the system uses two million geographical names extracted from OpenStreetMap to generate an improved recognition result. For example, by matching "Cltureh" from the 1935 map to "urch" in the 1900 map, the system finds the word "Church" in the geographic names extracted from OpenStreetMap to replace "Cltureh" and "urch" in the 1935 and 1900 recognition results, respectively.
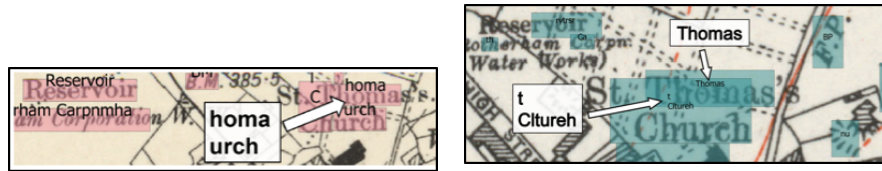
(a) Strabo incorrectly recognizes "St. Thomas Church" as "homa urch" in a 1900 map.

(b) Strabo incorrectly recognizes "St. Thomas Church" as "Thomas t Cltureh" in a 1935 map.

Figure 9. Matching imperfect OCR results from two map editions to improve recognition accuracy

For the multi-model approaches, the current challenges include how to exploit string similarity measures between the extracted map text (which contain recognition errors) and other sources (e.g., gazetteer entries) to 1) prune the search space for finding the matching pattern efficiently and 2) using matches between the OCR text and dictionary entries to learn potential OCR errors specifically for each map type. For example, the character sequences "ni" and "in" is commonly recognized as one character "m" during OCR. With enough training data (matches between OCR text and dictionary entries), the algorithm should be able to learn that the OCR results "Baldwm Hills" is highly likely to be "Baldwin Hills" for a specific map type or condition.

5. Future Outlook

This paper presents studies in natural and social sciences demonstrating the opportunities for the image processing & pattern recognition community to transform conventional research practices in using historical maps. For example, a new technology that automatically generates machine-readable or -understandable (e.g., LinkedData (Bizer, Heath, & Berners-Lee, 2009)) place name databases from historical maps and to do so at scale will enable biology scientists to minimize the time and effort for geo-locating their data records and to efficiently query and analyze historical records by location and time. These opportunities also present unique possibilities for researchers in image processing & pattern recognition to identify collaborators in other scientific domains. This type of interdisciplinary collaboration allows the researchers in image processing & pattern recognition to create algorithms and applications to solving "wicked" research problems and addressing real-world challenges facing our society. Further, the paper discusses a number of trends and their challenges in text recognition from historical maps. These trends have already shown promising results in the automatic unlocking of textual content in

heterogeneous historical maps. Solving the challenges in these trends will make it possible to use a large number of heterogeneous historical maps efficiently and study historical spatiotemporal datasets on a large scale.

Acknowledgements

REFERENCES

Adams, O. G. (1928). *Place Names In The North Central Counties Of Missouri* (Ph.D.). University of Missouri-Columbia.

Alex, B., Byrne, K., Grover, C., & Tobin, R. (2015). Adapting the Edinburgh Geoparser for Historical Georeferencing. *International Journal of Humanities and Arts Computing*, *9*(1), 15–35.

Arteaga, M. G. (2013). Historical map polygon and feature extractor. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on MapInteraction* (pp. 66–71). ACM.

Chiang, Y.-Y., & Knoblock, C. A. (2014). Recognizing Text in Raster Maps. *GeoInformatica*, *19*(1), 1–27.

Chiang, Y.-Y., Leyk, S., & Knoblock, C. A. (2014). A Survey of Digital Map Processing Techniques. *ACM Computing Surveys (CSUR)*, *47*(1), 1.

Chiang, Y.-Y., Leyk, S., Nazari, N. H., Moghaddam, S., & Tan, T. X. (2016). Assessing the impact of graphical quality on automatic text recognition in digital maps. *Computers & Geosciences*, *93*, 21–35.

Davis, C. C., Willis, C. G., Connolly, B., Kelly, C., & Ellison, A. M. (2015). Herbarium records are reliable sources of phenological change driven by climate and provide novel insights into species' phenological cueing mechanisms. *American Journal of Botany*, *102*(10), 1599–1609.

D'Ignazio, C., Bhargava, R., & Zuckerman, E. (n.d.). Cliff-clavin: Determining geographic focus for news. In *NewsKDD: Data Science for News Publishing*.

Garijo, D., Gil, Y., & Harth, A. (2015). Challenges for Provenance Analytics Over Geospatial Data. In B. Ludäscher & B. Plale (Eds.), *Provenance and Annotation of Data and Processes* (Vol. 8628, pp. 261–263). Springer International Publishing.

Godfrey, B., & Eveleth, H. (2015). An Adaptable Approach for Generating Vector Features from Scanned Historical Thematic Maps Using Image Enhancement and Remote Sensing Techniques in a in a Geographic Information System. *Journal of Map & Geography Libraries*, 18–36.

Gregory, I., Donaldson, C., Murrieta-Flores, P., & Rayson, P. (2015). Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research. *International Journal of Humanities and Arts Computing*, *9*(1), 1–14.

Gregory, I. N., & Ell, P. S. (2007). *Historical GIS: Technologies, Methodologies, and Scholarship* (Vol. 39). Cambridge University Press.

Guralnick, R. P., Wieczorek, J., Beaman, R., Hijmans, R. J., Group, B. W., & Others. (2006). BioGeomancer: automated georeferencing to map the world's biodiversity data. *PLoS Biology*, *4*(11), e381.

Hill, A. W., Guralnick, R., Flemons, P., Beaman, R., Wieczorek, J., Ranipeta, A., … Remsen, D. (2009). Location, location, location: utilizing pipelines and services to more effectively georeference the world's biodiversity data. *BMC Bioinformatics*, *10*(Suppl 14), S3.

Honarvar Nazari, N., Tan, T. X., & Chiang, Y.-Y. (2016). Integrating Text Recognition for Overlapping Text Detection in Maps. In *Electronic Imaging, Document Recognition and Retrieval XXIII* (pp. 1–8).

Khotanzad, A., & Zink, E. (2003). Contour Line and Geographic Feature Extraction From USGS Color Topographical Paper Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(1), 18–31.

Kurashige, L. (2013). Rethinking Anti-Immigrant Racism: Lessons from the Los Angeles Vote on the 1920 Alien Land Law. *Southern California Quarterly*, *95*(3), 265–283.

Lavoie, C. (2013). Biological collections in an ever changing world: Herbaria as tools for biogeographical and environmental studies. *Perspectives in Plant Ecology, Evolution and Systematics*, *15*(1), 68–76.

Leidner, J. L., & Lieberman, M. D. (2011). Detecting geographical references in the form of place names and associated spatial natural language. *SIGSPATIAL Special*, *3*(2), 5–11.

Leyk, S., & Boesch, R. (2009). Colors of the Past: Color Image Segmentation in Historical Topographic Maps Based on Homogeneity. *GeoInformatica*, *14*(1), 1–21.

Leyk, S., Boesch, R., & Weibel, R. (2006). Saliency and Semantic Processing: Extracting Forest Cover From Historical Topographic Maps. *Pattern Recognition*, *39*(5), 953–968.

Li, L., Nagy, G., Samal, A., Seth, S., & Xu, Y. (2000). Integrated text and line-art extraction from a topographic map. *International Journal on Document Analysis and Recognition*, *2*(4), 177–185.

Murphey, P. C., Guralnick, R. P., Glaubitz, R., Neufeld, D., & Ryan, J. A. (2004). Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the Mountain and Plains Spatio-Temporal Database-Informatics Initiative (Mapstedi). *Phyloinformatics*, *1*(3), 1–29.

Nagy, G., Samal, A., Seth, S., & Fisher, T. (1997). Reading street names from maps-Technical challenges. *Procs. GIS/LIS*.

Nanetti, A., Cattaneo, A., Cheong, S. A., & Lin, C.-Y. (2015). Maps as Knowledge Aggregators: from Renaissance Italy Fra Mauro to Web Search Engines. *The Cartographic Journal*, *52*(2), 159–167.

Newbold, T. (2010). Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, *34*(1), 3–22.

Ngo, V., Swift, J., & Chiang, Y.-Y. (2015). Visualizing Land Reclamation in Hong Kong: A Web Application. In *International Cartographic Conference*.

Pezeshk, A., & Tutwiler, R. L. (2010). Improved Multi Angled Parallelism for separation of text from intersecting linear features in scanned topographic maps. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1078–1081). IEEE.

Pezeshk, A., & Tutwiler, R. L. (2011). Automatic Feature Extraction and Text Recognition From Scanned Topographic Maps. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *49*(12), 5047–5063.

Pyke, G. H., & Ehrlich, P. R. (2010). Biological collections and ecological/environmental research: a review, some observations and a look to the future. *Biological Reviews of the*

Cambridge Philosophical Society, *85*(2), 247–266.

Raveaux, R., Burie, J. C., & Ogier, J. M. (2007). A Colour Document Interpretation: Application to Ancient Cadastral Maps. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)* (Vol. 2, pp. 1128–1132). IEEE.

Raveaux, R., Burie, J. C., & Ogier, J. M. (2008). Object Extraction from Colour Cadastral Maps. In *Document Analysis Systems, 2008. DAS '08. The Eighth IAPR International Workshop on* (pp. 506–514). IEEE.

Rios, N. E., & Bart, H. L. (2010). GEOLocate (Version 3.22) [Computer software].

Samy, G., Chavan, V., Ariño, A. H., Otegui, J., Hobern, D., Sood, R., & Robles, E. (2013). Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. *Biodiversity Informatics*, *8*(2). http://doi.org/10.17161/bi.v8i2.4124

Simon, R., Barker, E., & Isaksen, L. (2015). Linking early geospatial documents, one place at a time: annotation of geographic documents with Recogito. *E - Perimetron*.

Simon, R., Pilgerstorfer, P., Isaksen, L., & Barker, E. (2014). Towards Semi-Automatic Annotation of Toponyms on Old Maps. *E - Perimetron*, *9*(3), 105–128.

Simon, R., Sadilek, C., Korb, J., Baldauf, M., & Haslhofer, B. (2010). Tag Clouds and Old Maps: Annotations as Linked Spatiotemporal Data in the Cultural Heritage Domain. In *Workshop On Linked Spatiotemporal Data*. Zurich, Switzerland.

Torr, P. H. S., & Zisserman, A. (2000). MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding: CVIU*, *78*(1), 138–156.

Vellend, M., Brown, C. D., Kharouba, H. M., McCune, J. L., & Myers-Smith, I. H. (2013). Historical ecology: using unconventional data sources to test for effects of global environmental change. *American Journal of Botany*, *100*(7), 1294–1305.

Weinman, J. (2013). Toponym Recognition in Historical Maps by Gazetteer Alignment. In *Proceedings of the 12th International Conference on Document Analysis and Recognition* (pp. 1044–1048).

Yoshida, K., Burbano, H. A., Krause, J., Thines, M., Weigel, D., & Kamoun, S. (2014). Mining herbaria for plant pathogen genomes: back to the future. *PLoS Pathogens*, *10*(4), e1004028.

Yu, R., Luo, Z., & Chiang, Y.-Y. (2016). Recognizing Text On Historical Maps Using Maps From Multiple Time Periods. In *Proceedings of the 23rd International Conference on Pattern Recognition*.