Assessing the Impact of Graphical Quality on Automatic Text Recognition in Digital Maps

Yao-Yi Chiang*,[1], Stefan Leyk, Narges Honarvar Nazari***, Sima Moghaddam***, Tian Xiang Tan*****

* University of Southern California, Spatial Sciences Institute
** University of Colorado at Boulder, Department of Geography
*** University of Southern California, Department of Computer Science and Spatial Sciences Institute

Abstract. Converting geographic features (e.g., place names) in map images into a vector format is the first step for incorporating cartographic information into a geographic information system (GIS). With the advancement in computational power and algorithm design, map processing systems have been considerably improved over the last decade. However, the fundamental map processing techniques such as color image segmentation, (map) layer separation, and object recognition are sensitive to minor variations in graphical properties of the input image (e.g., scanning resolution). As a result, most map processing results would not meet user expectations if the user does not "properly" scan the map of interest, pre-process the map image (e.g., using compression or not), and train the processing system, accordingly. These issues could slow down the further advancement of map processing techniques as such unsuccessful attempts create a discouraged user community, and less sophisticated tools would be perceived as more viable solutions. Thus, it is important to understand what kinds of maps are suitable for automatic map processing and what types of results and process-related errors can be expected. In this paper, we shed light on these questions by using a typical map processing task, text recognition, to discuss a number of map instances that vary in suitability for automatic processing. We also present an extensive experiment on a diverse set of scanned historical maps to provide measures of baseline performance of a standard text recognition tool under varying map conditions (graphical quality) and text representations (that can vary even within the same map sheet). Our experimental results help the user understand what to expect when a fully or semi-automatic map processing system is used to process a scanned map with certain (varying) graphical properties and complexities in map content.

[1] Corresponding author, email: yaoyic@usc.edu

Keywords: Digital Map Processing, Scanned Maps, Geographic Information System, Text Recognition, Optical Character Recognition, Accuracy Assessment

1. Introduction

Digital map processing refers to a set of techniques for converting map images (created through scanning of paper maps or produced as electronic raster maps) into the vector format. This conversion is usually the first step for incorporating geographic information encapsulated in maps (e.g., place names, place types, build-up areas, contour lines) into a spatial-analytic environment, such as a geographic information system (GIS). Since the early 80s, various map processing systems (including both software and hardware tools) were developed to facilitate manual map processing tasks. Today, the efficiency, accuracy, and degrees of automation of map processing systems have been increased considerably (concerning processing speed and the capability to process a variety of maps and map features). The systems that are in place nowadays can be classified by their capabilities into four categories: (1) Basic raster-to-vector conversion tools with a minimum of automation (e.g., Esri ArcScan[2]), which can be applied to a wide variety of map types with different graphical conditions (by leveraging human vision), (2) Semi-automatic systems, which provide some degrees of automation to reduce manual digitization efforts (e.g., AutoCAD RasterDesign[3]), (3) Fully automatic systems for processing a specific map type; this type-dependency often has the disadvantage that the system relies on the user to fine tune the digitization settings (requiring expert knowledge in image processing and graphics recognition, e.g., Map Vectorizer[4]), and (4) Fully or semi-automatic systems that are not limited to a particular map type but designed to extract only specific types of map features (e.g., map labels (Chiang and Knoblock, 2014)). The reader is referred to Henderson (2014) and Chiang, Leyk, and Knoblock (2014) for detailed reviews on map processing techniques and systems.

Despite the exponential growth in computational power and advancement in graphics recognition algorithms in the last decade, most fundamental

---

[2] http://www.esri.com/software/arcgis/extensions/arcscan
[3] http://www.autodesk.com/products/autocad-raster-design/overview
[4] https://github.com/NYPL/map-vectorizer

techniques that support automatic map processing such as color segmentation, (map) layer separation, and object (or symbol) recognition are still limited when processing low quality or complex map images (Cherkassky and Mulier, 1998; Cordella and Vento, 2000; Llados et al., 2002). These techniques are sensitive to minor variations in graphical properties of the input image (e.g., different scanning parameters such as resolution) (Marr, 1982; Cherkassky and Mulier, 1998) and usually require a priori knowledge of the map properties and content (e.g., size of map objects, and cartographic styles). As a result, most map processing systems would fail if the user does not "properly" prepare the map document for processing and train and tune the underlying algorithms. Since the general user rarely has expert knowledge of the underlying map processing techniques, a map processing system is often perceived as a black box that converts a map image into spatial data that are readily accessible in a GIS. One significant implication is that after a few attempts to use a map processing system, the user would give up if the results do not meet user expectations and move to less sophisticated tools for manual raster-to-vector conversion. Not only does this create a discouraged user community, but it also slows down further development of advanced map processing techniques as less sophisticated tools would be seen as more viable solutions.

Therefore, it is critical for a user to understand *what kinds of maps are suitable for automatic (or semi-automatic) map processing* and *what types of results can be expected*. This directly relates to further questions concerning the reliability and objectivity of accuracy assessments. Knowing how sensitive the performance of map processing techniques will be based on variations in graphical quality will inform the user how accuracy could vary across map types and even within one map image in which target features may show differences in graphical properties. In this article, we shed light on such questions. We choose a typical map processing task, text recognition, and discuss how the degree of suitability for text recognition varies across map instances that differ graphically. Furthermore, we carry out an experiment on text recognition in scanned historical maps of various types and origins to demonstrate the impact such variations can have on performance across different levels of graphical quality. This experiment enables accuracy assessment of automatic text recognition results for map labels in a variety of graphical conditions and provides a guideline for estimating the suitability of a given map for automatic text processing.

In the next section, we review various types of maps tested in the literature on text recognition using automatic or semi-automatic map processing systems.

These maps carry different forms and types of text and show varying degrees of complexity due to overlapping map layers and density of cartographic information. Then we discuss in detail the most relevant properties of map images affecting text recognition accuracy. Next, we introduce an automatic text recognition system from our previous work (Chiang and Knoblock, 2014), and describe an experiment on a set of scanned historical maps including Ordnance Survey maps[5] produced in the United Kingdom and several other maps produced in the United States. The experiment demonstrates the baseline performance of this text recognition system on maps with a variety of text representations. We discuss how potential users can evaluate the suitability of a map of interest for text recognition tasks. Finally, we present future outlooks on how text processing in digital maps should further evolve to reach higher degrees of automation and more robust recognition results.

2. Common Map Types Subject to Automatic Text Recognition and Related Accuracy Issues

Text recognition from digital map images is one of the most common map processing tasks, which determines the locations (e.g., bounding boxes or center points) of text objects and generates machine editable strings for individual text labels in the map (Ye and Doermann, 2014). A large number of studies on text recognition in digital maps can be found in the literature (e.g., Nagy et al., 1997; Velázquez and Levachkine, 2004; Gelbukh et al., 2004; Pouderoux et al., 2007; Chiang and Knoblock, 2014; Simon et al., 2014). These studies in which typically text labels are extracted from map images and incorporated into subsequent processing steps of Optical Character Recognition (OCR) have a wide range of applications such as building gazetteers, carrying out historical research on location name changes or studying changes in the landscape and land-use. In addition, extracting and removing map text can improve the recognition of other geographic features such as cadastral boundaries (Cao and Tan, 2002), vegetation features (Leyk et al., 2006), elevation contours (Khotanzad and Zink, 2003) or roads (Li et al., 2000; Chiang and Knobock, 2013).

A variety of map types that have been tested in the literature either for text recognition or for removing map text labels include: cadastral or land register

[5] http://www.ordnancesurvey.co.uk/

maps (e.g., Raveaux et al., 2008), road maps (e.g., Bin and Cheong, 1998; Itonaga et al., 2003; Dhar and Chanda, 2006; Bucha et al., 2007; Chiang et al., 2013; Chiang and Knoblock, 2013), hydrographic maps (e.g., Trier et al., 1997), city maps (e.g., Chen et al., 1999), utility maps (e.g., den Hartog et al., 1996), as well as topographic or other survey maps (e.g., Bessaid et al., 2003; Miyoshi et al., 2004; Chen et al., 2006; Leyk et al., 2006; Leyk and Boesch, 2009; Xin et al., 2006; Henderson et al., 2009). We show several examples of the above map types in the next section to illustrate key characteristics and conditions relevant for text recognition in detail.

Most map processing systems cannot process different types of maps automatically, which is, in particular, true for text recognition. This is because maps have a complex layout in which text labels appear in various forms, colors and size categories, which requires manual identification of processing parameters and system training. Recent studies show an increasing potential to establish text recognition systems that provide reliable solutions across different types of maps, but their accuracy can vary significantly across map types (e.g., Chiang and Knoblock, 2014; Simon et al., 2014). Moreover, variations in text label characteristics (e.g., text color) can also occur within maps of the same types or even a single map page as a result of the scanning and image compression process, differences in map complexity, and inconsistencies of graphical quality in the original map (due to aging or bleaching). Thus, the same recognition method may perform differently in various parts of one map. Understanding such recognition sensitivities to variations in graphical properties can further improve the ability to forecast the potential for automatic text recognition and highlight possible recognition errors automatically. Importantly, this will also lead to realistic and objective accuracy assessments by differentiating graphical quality levels found among text labels in maps.

3. Key Characteristics Indicating the Potential for Automated Text Recognition in Maps

Much of the potential for a certain map to be processed with a high degree of automation is directly related to the number of studies that focus on this type of map (e.g., more studies exist on maps with Latin scripts compared to other languages). In this section, we present example maps of different types and discuss a variety of characteristics that can be used to estimate the suitability of these maps for automatic text recognition and those that would indicate the need for user intervention and manual digitization efforts.

The discussion is structured by the major characteristics of text labels and map content: language (script), font, curvature and spacing, print and image quality, text color as well as map complexity. In general, the aim in most studies on text recognition in maps is to detect, extract, and transfer text labels to an OCR component, which then performs the final recognition process (Nagy et al., 1997; Cao and Tan, 2000; Li et al., 2000; Velázquez and Levachkine, 2004; Gelbukh et al., 2004; Pouderoux et al., 2007, Chiang and Knoblock, 2014). How well map labels can be identified and recognized heavily depends on the characteristics described below.

3.1. Map Language

Current OCR software packages, such as the open source Tesseract-OCR[6] or commercial ABBYY FineReader,[7] support a wide range of language scripts, including Latin, Chinese, Korean, Japanese, Hebrew, Arabic, and Indian scripts. However, most of the text recognition work for processing raster maps is limited to Latin scripts, including Spanish (e.g., Gelbukh et al., 2004), French (e.g., Pouderoux et al., 2007), and English (e.g., Chiang and Knoblock, 2014). The main reason is that the document analysis techniques used for detecting locations of text labels in maps are well developed for Latin scripts but less so for other scripts. However, just as OCR progresses over the years from handling only Latin scripts (Rice et al, 1995; Smith, 2007) to more complex scripts, such as degraded Indian scripts (Shukla and Banka, 2014), we expect further progress in developing automatic recognition methods that can handle a variety of scripts in maps. Of course, the performance of text recognition methods in maps with Latin script also depends on other graphical conditions and map characteristics. Lower levels of general image quality will always impact the extraction (e.g., coarse resolution images carry a limited potential for automatic text recognition for any script).

3.2. Map Fonts

Maps with common typewritten fonts usually show the best results in automatic text recognition (Figures 1 and 2) compared to maps with less common fonts (e.g., Fraktur, Antiqua) or stenciled and handwritten text. Text with uncom-

[6] https://code.google.com/p/tesseract-ocr/

[7] http://finereader.abbyy.com/

mon typewritten fonts requires additional training on specific character sets and yields lower OCR accuracy (Helinski et al., 2012). Figure 3 shows an example map with stenciled text. Historical maps are traditionally prepared with manually written or stenciled text, which adds to the challenges in text recognition in older cartographic documents that can suffer from inferior graphical quality and archiving effects (e.g., Gelbukh et al., 2004; Raveaux et al. 2007, 2008; Simon et al., 2014).

Figure 1. An example of typewritten fonts in a scanned map for which OCR performs well (Panama, USGS National Imagery and Mapping Agency (NIMA) ref. no. E762X38382).

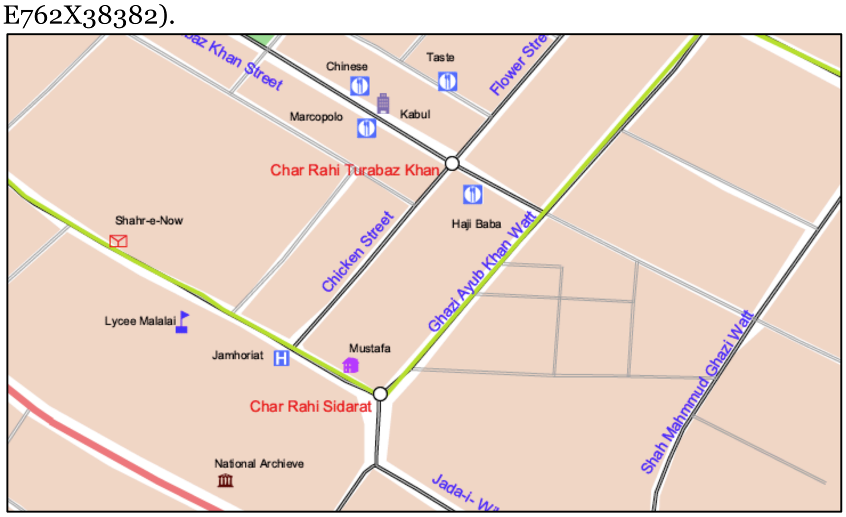


Figure 2. An example of typewritten fonts in a computer generated map for which OCR can perform well (Kabul city center, Afghanistan Information Management Service).

3.3. Character Spacing, Label Curvature and Orientation

OCR software works most robustly if the input text labels are geometrically straight (vertically positioned characters) with regular character spacing and horizontal orientation. Such text labels also have a higher chance to be detected automatically compared to labels with non-horizontal orientation (Figure 2), curved labels (Figure 4) or labels with wide or irregular character spacing (Figures 3 and 5). Automatic systems often break curved labels and labels with wide character spacing into separate string segments, which then require manual post-processing to regroup these string segments (e.g., Velázquez and Levachkine, 2004; Chiang and Knoblock, 2014).

Figure 3. Stenciled text in a historical map of Denmark.

Figure 4. Examples of curved labels in an Afghanistan map.
Source: United Nations

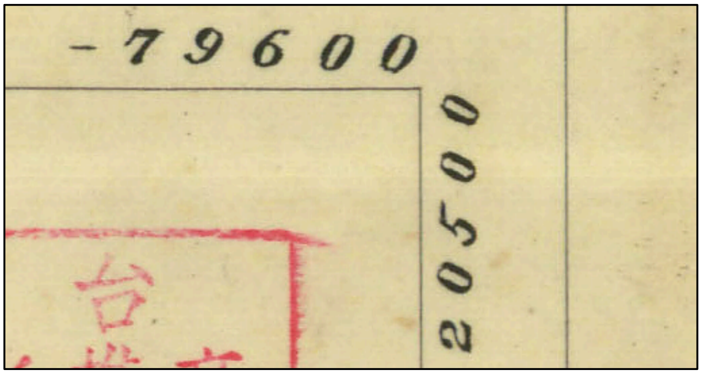


Figure 5. Text labels with wide character spacing in a historical map of Taiwan.

3.4. Print Quality

In general, automatic map processing systems rely on superior print quality of the original paper maps with a minimum of blurring and false coloring to produce accurate results (Henderson, 2014; Chiang, Leyk, and Knoblock, 2014). However, old printing technology was limited in quality and the final printout often suffered from such problems. Print quality is often related to and can be further decreased through bleaching of the map as a direct consequence of aging paper material and the archiving practice. How sensitive the paper material can be to the archiving conditions becomes obvious in historical maps of more than 100 years of age (Leyk et al., 2006). Figure 6 shows an example of blurring and false coloring. The quality of a printed map also depends on the engraving techniques (e.g., stone and copper engraving) used to produce older maps. The transition to modern production techniques varies among countries. Unfortunately, the original plates used for engraving have been disposed in many cases making the paper maps the only sources left. In summary, the degree of blurring, false coloring, and mixed colors provides a strong indication of the potential of automated recognition on a given map. Text in maps often overlaps with other map layers (e.g., Figures 4 and 6), which makes text recognition particularly sensitive to such general printing quality issues.

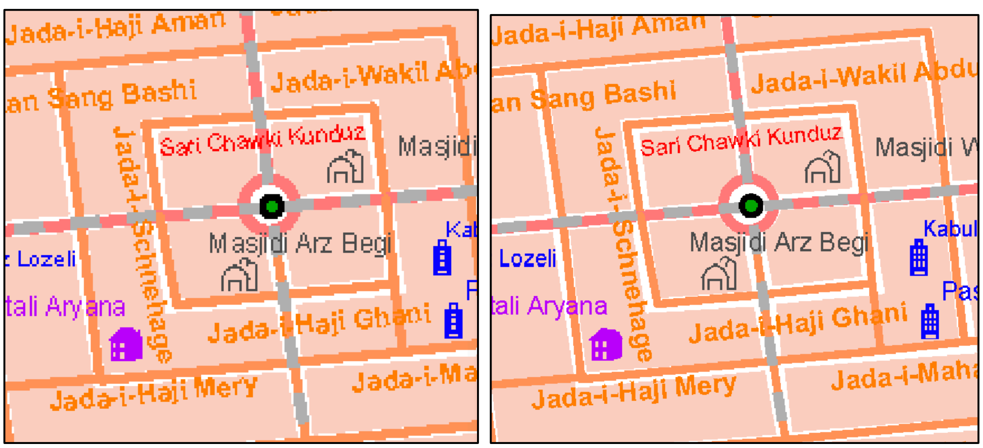
3.5. Image Quality

State-of-the-art OCR software (e.g., Tesseract-OCR and ABBYY FineReader) requires an image resolution of the scanned input image of at least 300 dots-per-inch (DPI) to achieve the best results in "well-conditioned" documents (e.g., see Yin and Huang, 2001; Liu, 2002; Pouderoux et al., 2007). This number increases for maps of high density and complexity such as topographic maps (see Section 3.6). Figures 7 and 8 show a comparison of the text appearance in a map scanned with 150 DPI and 300 DPI, respectively. There are several instances in which images in digital map archives would be stored with a resolution too coarse to differentiate the smallest elements shown in a map. One of the main reasons is hardware limitations as scanners capable of scanning large format documents are expensive and scanning with high resolution is a time-consuming process. Since priority is generally given to a timely completion of a scanning project, such key parameters are often underestimated. As a guideline, the resolution of a scanned map image subject to automated information extraction should facilitate the graphical and visual distinction of the smallest entities in that map. This guideline relates to the concepts of resolution vs. detection in remote sensing imagery, i.e., to detect an object of a certain size the resolution has to be fine enough to be able to spatially and spectrally identify and characterize this object and reduce mixed pixel effects. Text in maps often has varying dimensions (i.e., line thickness) and thus represents a highly sensitive map element regarding resolution. Characters or character chains may become disconnected because thin object parts cannot be represented graphically with the pixel size given. In contrast, creating extremely high-resolution images may result in inefficient map processing. Also, a map image should not be processed by lossy image compression algorithms (e.g., JPEG[8]) as important structural elements become compromised and cannot be reproduced. Figure 9 illustrates how lossy compression of a map image results in pixelated map objects and increased color confusion.

In addition to image resolution, the color encoding (if the map contains color layers) used for scanning and processing as well as the bit-depth of the image data are also important factors with regard to image quality. Color encoding is most relevant in preprocessing steps such as color image segmentation (Leyk, 2010; Leyk and Boesch, 2010) for generating clear character representations input to OCR (Chiang and Knoblock, 2014). Choices of color spaces include RGB (red, green, and blue), HSL (hue, saturation, and luminance), or CIE 1976 $L*u*v$. The bit-depth of the image indicates the maximum number of unique colors that can be represented in an image, which is important in recognition tasks in which objects to be distinguished are very similar in color. In most text recognition tasks, the use of 24-bit data during the scanning process is sufficient to produce clear text appearance (e.g., crisp character edges) for OCR.

[8] http://www.jpeg.org/

Figure 6. An example of poor print quality in a NIMA evasion chart (EVC NH-36A, NIMA ref. no. EVCXXNH36A).



(a) 150 DPI (b) 300 DPI

Figure 7. Comparison of text appearance under different image resolutions (Kunduz city map, Afghanistan Information Management Service).

(a) 150 DPI (b) 300 DPI

Figure 8. Comparison of text appearance under different resolutions chosen for the scanning process; NIMA tactical pilotage chart (Australia, TPC Q-15A, NIMA ref. no. TPCXXQ15A).

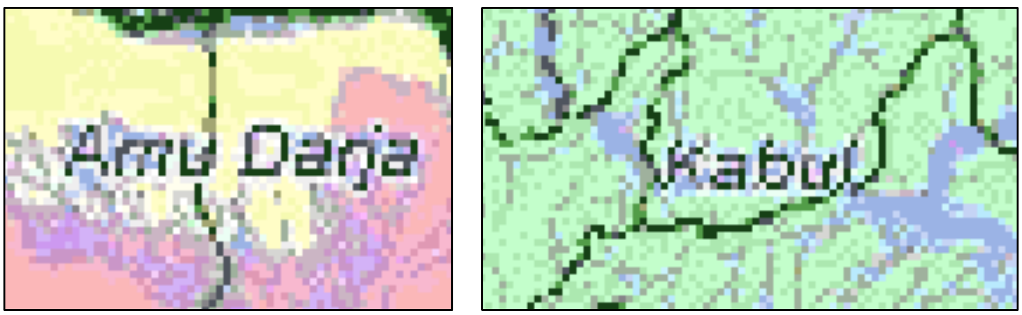


Figure 9. Low image resolution and lossy image compression compromise the appearance of text and map features (United Nations Environment Programme and United Nations Institute for Training and Research Operational Satellite Applications Programme map).

3.6. Map Complexity

Maps can contain dense and overlapping map features (of the same or different color layers) and text (e.g., Figure 10), which makes map images a challenging document type for recognition tasks (Cordella and Vento, 2000; Llados et al., 2002). As a consequence, frequent instances of mixed colors and merged map objects may occur impeding the identification or separation of features or symbols. For highly complex maps, such as topographic maps, an image resolution of at least 500 DPI has been demonstrated suitable in recent research (e.g., Li et al., 2000; Liu, 2002; Leyk and Boesch, 2009; Chiang et al., 2014) in order to ensure that map processing techniques (including text recognition) produce robust results. Issues of image and print quality (as described above) in combination with map complexity can be found in historical maps, which therefore represent particularly challenging documents for recognition tasks including text recognition (Simon et al., 2014).

Figure 10. A sample map with complex and dense content, text with small fonts and in different colors (Muqdisho, Somalia, NIMA ref. no. EVTXXMUQDISHO).

3.7. Color of Map Features

Ideally, map features of the same type should have a distinct color avoiding merging and color mixing effects as mentioned above under print and image quality. However, Figure 11 shows one of many examples where the text labels and the road edges are both drawn in black. In this case, the recognition task would likely require manual post-processing for recovering the text labels that overlap with road edges. Even if text color would be different from other map layers, there may still be significant problems regarding color variations and mixed colors, i.e., colors may not be clearly differentiated everywhere as an issue of print quality. Image quality issues (e.g., bleaching, blurring, resolution, and color space used for scanning) may add to these points. In general, if text appears in the same color as other map layers, the success of text recognition will depend on the degree of complexity of the map and the frequency of overlaps between these layers.

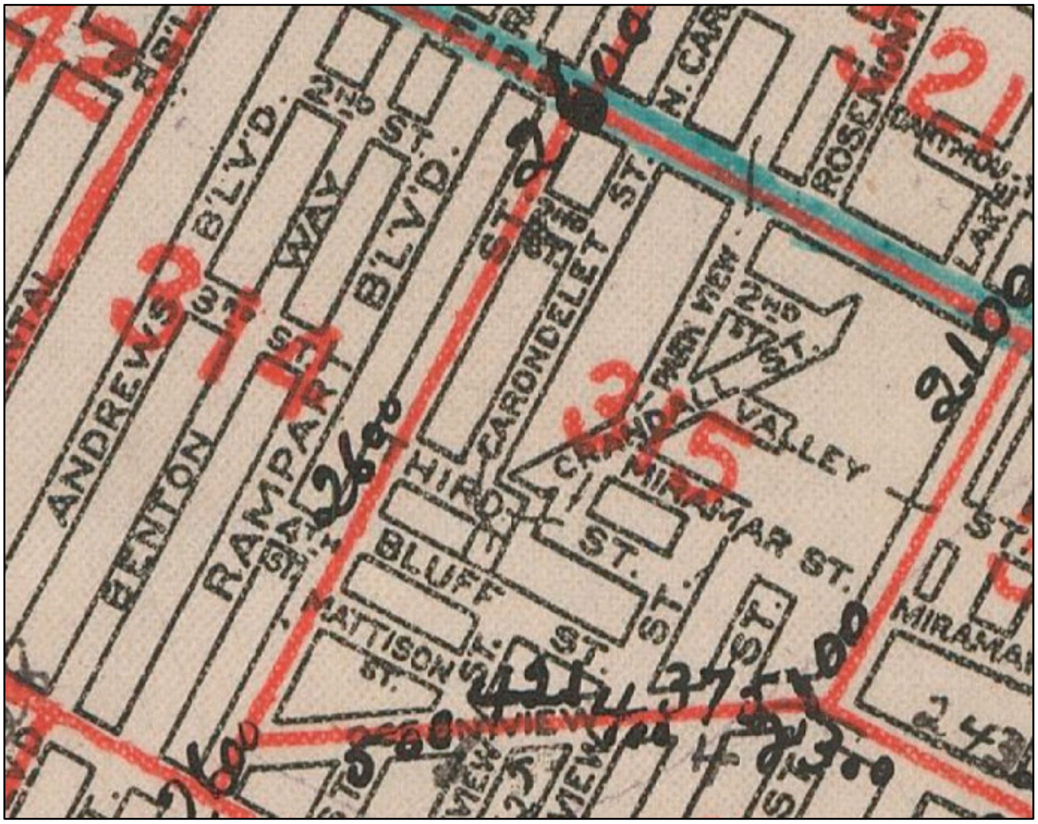


Figure 11. Both text and roads are drawn in black color; red precinct boundaries and black text labels overlap resulting in mixed colors (1920 Los Angeles precinct map, Los Angeles City Archive).

4. Data and Experimental Setting

This section describes the tested map products, the characteristics of the map content (including map labels), and the test system.

4.1. Tested Map Products and Their Characterization

To demonstrate the differences in text recognition outcomes under varying graphical conditions and text properties as discussed in Section 3, we tested the performance of a text recognition tool for six different map products (Table 1), including the 1920 6-inch Ordnance Survey topographic maps from the

National Library of Scotland,[9] and United States historical railway, auto road and mileage maps from the David Rumsey Map Collection.[10]

Table 1. The metadata of the six tested map products.

Map Title / Coverage	DPI (approx.)	Map Scale	Publisher	Date
Ordnance Survey Six-inch Map, London, U.K.	406	1: 10,560	Ordnance Survey	1920
Cram's Railroad and Township Map, Florida[11]	336	1: 1,330,560	Cram Atlas Company	1875
Map of the Northern Pacific Railroad and connections[12]	302	1: 7,500,000	Rand McNally	1879
Map Of Missouri, Showing Line and Land Grant of the St. Louis & San Francisco Railway[13]	304	1: 1,966,700	Woodward, Tiernan & Hale	1879
Auto Road Map, Colorado[14]	402	1: 1,700,000	Rand McNally	1927
Black and White Mileage Map, South Dakota[15]	379	N/A	Rand McNally	1924

Within several map pages from the Ordnance Survey six-inch map series of the U.K., we tested ten map subsections near London each covering 1,000 x 1,000 square meters in the TQ grid (the British National Grid), equal to 1,512 x 1,512 pixels. For each of the historical U.S. maps, we selected one map subsection ranging from 753 x 665 to 1,176 x 1,121 pixels for testing. Figures 12-18 show examples of the test maps, which represent a wide range of variations in map conditions and labeling styles. Based on the criteria relevant for text recognition (see Section 3), text labels in these maps can be characterized as follows:

---

9 http://maps.nls.uk
10 http://www.davidrumsey.com
11 http://www.davidrumsey.com/luna/servlet/s/81sbj5
12 http://www.davidrumsey.com/luna/servlet/s/gev3rb
13 http://www.davidrumsey.com/luna/servlet/s/ql012o
14 http://www.davidrumsey.com/luna/servlet/s/1oscg7
15 http://www.davidrumsey.com/luna/servlet/s/8g46i4

Map Language and Fonts:

The Ordnance Survey maps have Latin scripts (English) and use common fonts with the exceptions of some special locations (Figures 12 and 13). The other historical maps have Latin scripts (English) and use uncommon fonts (likely stenciled text) varying within the same map (Figures 14 – 18) (see Section 3.2).

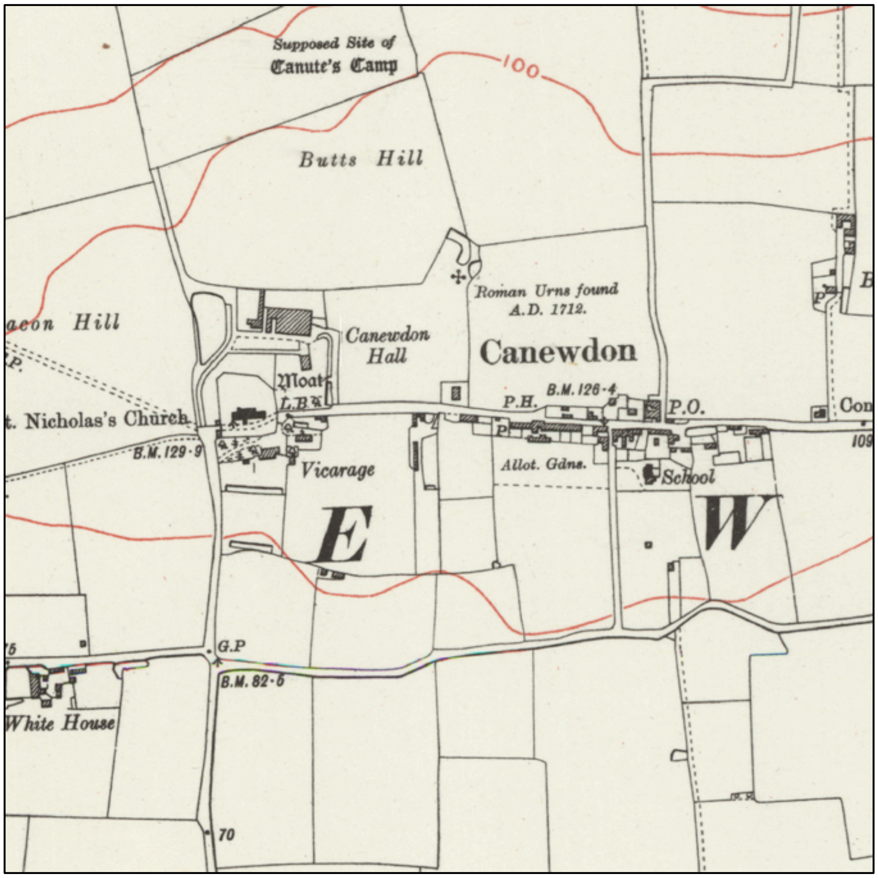


Figure 12. An example area of the tested Ordnance Survey map (TQ) (see Table 1).

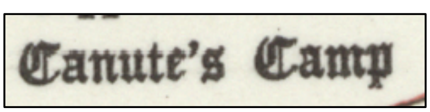


Figure 13. An example of an uncommon font in the Ordnance Survey maps.

Print and Image Quality:
The test map subsections are relatively free from print quality issues (discussed in Section 3.4) with the exceptions of the Map of Missouri that shows a visible fold line (Figure 15), and three other U.S. maps that were scanned out of books and show bleed-through from the back side (Figures 15 – 17). The image format of the test maps is TIFF without lossy compression. The exact scan resolutions of the original maps were not available. We estimated the image resolutions using the dimensions of the scanned images in pixels and the available sizes of the map documents in inches. The estimated resolution for every test map was higher than 300 DPI (Table 1). To test the impact of decreasing image quality for text recognition, we manually scaled the image dimensions of each map to 165%, 132%, 66% (medium), 50% (low), 33%, and 17%, respectively, using the bicubic interpolation. This interpolation method was carried out to simulate different image resolutions and possible compression defects combined. Note that when the map image was scaled up using the bicubic interpolation (165% and 132%), the DPI of the image did not increase. Our goal was to use these enlarged images to simulate the map content scanned at a higher DPI (e.g., larger font sizes and wider character spacing). We tested the performance of text recognition in all 15 map sections for each image quality level.

Label Curvature and Character Spacing, Map Complexity, and Color of Map Features:
The map layers of most maps tested are primarily represented in black (often blurred) color except for the contour lines, hydrography, and railroads. Other characteristics (label curvature, spacing, and map complexity) showed great variation among the test maps and were therefore (together with above characteristics) used to divide the map labels into three groups of general map properties relevant to recognition accuracy. These groups are described in the next subsection.

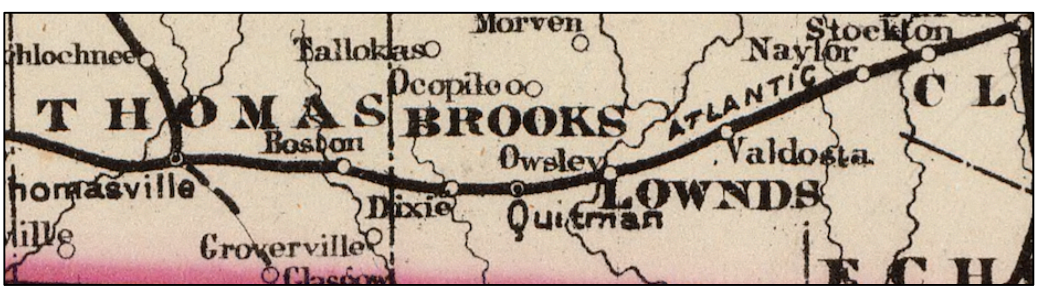


Figure 14. An example area of the tested Cram's Railroad and Township Map, Florida (see Table 1).

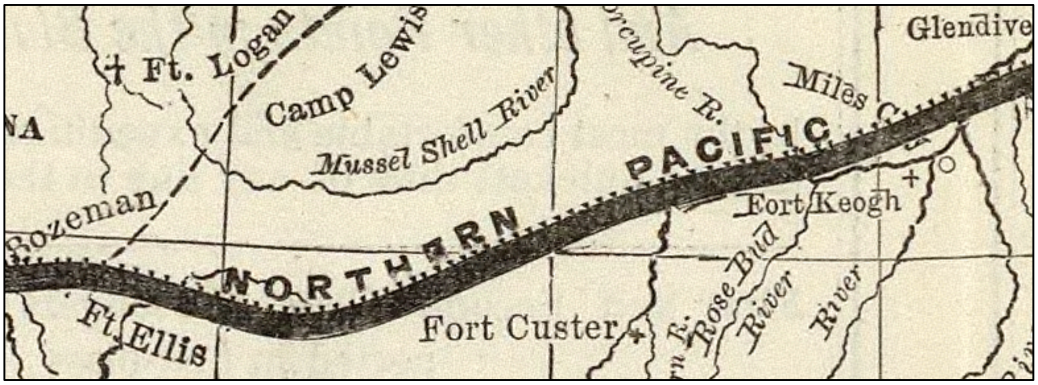


Figure 15. An example area of the tested Map of the Northern Pacific Railroad and connections (see Table 1).

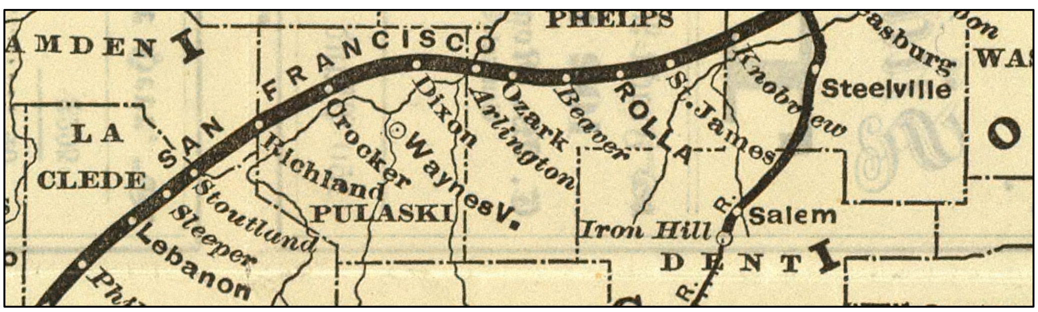


Figure 16. An example area of the tested Map Of Missouri, Showing Line and Land Grant of the St. Louis & San Francisco Railway (see Table 1).

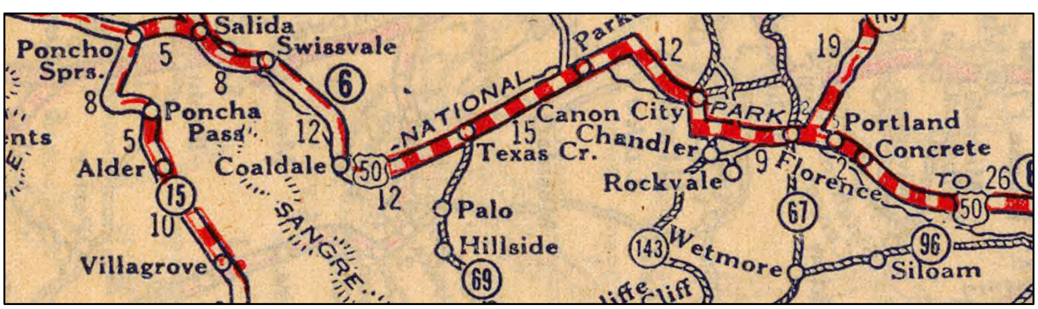


Figure 17. An example area of the tested Auto Road Map, Colorado (see Table 1)..

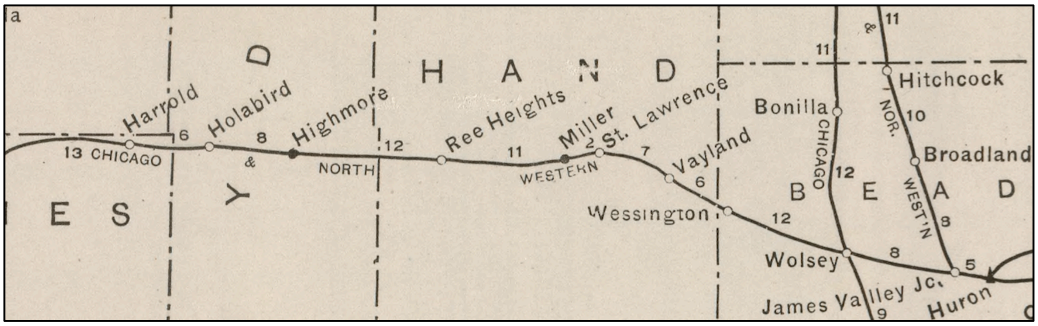


Figure 18. An example area of the tested Black and White Mileage Map, South Dakota (see Table 1).

4.2. Groups of Text Representations Based on Map Characteristics

Here, we define three groups of text representations of varying quality based on general map characteristics relevant to recognition. Each group contains characters in different sizes. Characters with a larger font size do not guarantee to have better recognition results than characters with a smaller font size in a map despite the common expectation that large font size would provide advantages for recognition similar to higher resolution. This is because map text that contains characters with larger font size typically shows wider character spacing, which makes processing this text label very difficult independently on resolution (Section 3.3). The recognition results of each group in Section 5 will demonstrate the impact of the map properties discussed in this article on the recognition accuracy.

Group 1 "suitable" (with high suitability for text recognition):
These are mostly clear and clean (unblurred and saturated) text labels with characters that are in either common, uncommon, or stenciled fonts, do not overlap with other map features, are not surrounded by or close to groups of non-text features, are only slightly curved or multi-oriented, or have regular or slightly wider (than usual) character spacing (Figure 19).

Group 2 "processable" (with moderate suitability for text recognition):

These are text labels that are slightly distorted, moderately curved, or may be surrounded by or close to (but not overlapping with) one or more non-text objects similar in size compared to a character (e.g., tree symbols) (Figure 20).

Group 3 "unsuitable" (with low suitability for text recognition): These are text labels with characters that overlap with non-text objects (Figure 21), are significantly curved[16] (Figure 22), or have wide character spacing (Figure 23).

Figure 19. Example labels that are highly suitable for text recognition (Group 1).

[16] A word that deviates more than 30% from a straight label (Chiang and Knoblock, 2014)

Figure 20. Example labels that are in noisy areas where nearby non-text symbols (e.g., trees, terrain features, circular symbols) could mislead the text detection and recognition algorithms (top), are slightly distorted or moderately curved (bottom) (Group 2).

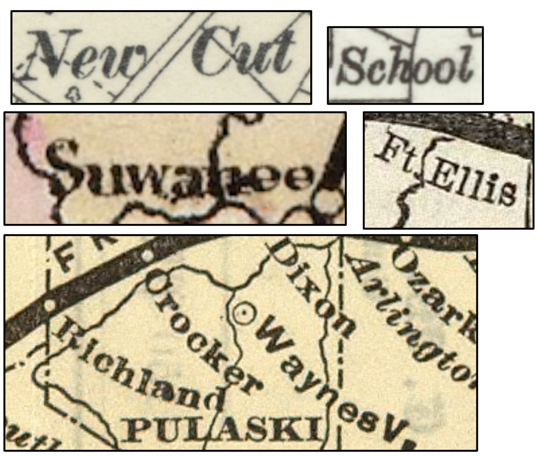


Figure 21. Example labels that overlap with other feature layers (Group 3).

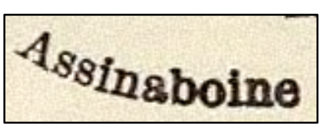


Figure 22. An example text label that deviates more than 30% from a straight label (Group 3).

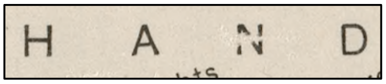


Figure 23. An example label that has a wide character spacing (Group 3).

4.3. A Brief Description of the Text Recognition Method Used

In order to conduct the experiment we used an open source text recognition tool, Strabo, developed in our previous work (Chiang and Knoblock, 2014)[17] that has been tested with a variety of map types (Chiang et al., 2014; Fernandes and Chiang, 2015; Honarvar Nazari et al., 2015). Strabo is a semi-automatic tool that can be trained by a user for processing a map of a certain type for text recognition. Strabo has two main components: (1) A text detector that exploits cartographic labeling principles to identify text pixels, groups the identified text pixels into characters, and then merges characters into text strings, and (2) A text recognizer that automatically determines the orientation of each detected string using a skew detection algorithm, rotates the string to the horizontal direction, and then uses Tesseract-OCR to convert the horizontal labels to machine-readable datasets. A detailed technical description of Strabo can be found in our previous publication (Chiang and Knoblock, 2014).

Recent efforts on integrating the text recognition capabilities in Strabo with a GIS (Chiang et al., 2014; Fernandes and Chiang, 2015) attempt to establish an end-to-end map digitization process from text label detection to OCR to result curation within a single software platform. This direct transition eliminates the need for manual data export/import procedures between GIS and OCR software and facilitates a broader use of such technologies in applied research (e.g., extracting historical location names from maps to better understand landscape conversions).

To train Strabo, the user delineates an example area that contains a map label. Then Strabo detects text pixels in the example area and learns the colors that represent text in the map.[18] In this experiment, since the text layers are primarily in black, we did not need to train Strabo. We used manually identified color

[17] https://github.com/spatial-computing/strabo-command-line-pub

[18] Details of Strabo training steps and demonstration videos can be access from http://spatial-computing.github.io/#projects

thresholds to extract the black layer from the Ordnance Survey maps. We used an automatic color binarization method (Bradley and Roth, 2007) to extract the black layers from the other test maps to save manual effort. Both the manual and automatic color binarization methods generated clear text layers.

In this comparative study we used parameter settings for running processes in Strabo as suggested in Chiang and Knoblock (2014) without parameter tuning for each test map, as follows:

- Two text pixels can only be connected to one another if they are in direct adjacency.

- A character can only be connected to another character (for constituting a text label) if the ratio of sizes between the two characters (larger character divided by the smaller character) is less than two. The size of a character refers to the character width or height whichever is larger.

- In a text label, the space between two connected characters is less than 1/5 of the size of the larger character.

- A text label that is curved and deviating more than 30% from a straight label (i.e., 234 degrees) will be broken into shorter labels for recognition.

As mentioned, the above steps in Strabo did not require training. For character recognition, we used the Tesseract-OCR engine with its default training data for English script without any additional training on the map font. To demonstrate the impact of pure map characteristics on text recognition, we did not use a dictionary to post-correct the results.

5. Experimental Results and Discussion

We manually transcribed text labels in the test maps and identified their suitability for text recognition (i.e., groups) to create the ground truth for validating the experiments.[19] The 15 test areas from the six map products of various types contain a total of 5,700 characters. The overall character-level precision, recall, and F-Score (the harmonic mean of precision and recall) for the original resolution were 37.32%, 61.79%, and 46.53%, respectively. All three measures dropped when the image resolution was reduced (Figure 24). Precision, recall,

[19] Test maps and ground truth are available at: https://github.com/spatial-computing/map-ocr-ground-truth

and F-Score dropped with decreasing resolution (e.g., the F-Score decreased by 11.98% from the original to the medium resolution and by 5.08% from the medium to the lowest resolution). The F-Score dropped to a mere 0.28% when the image was resized to 17% of the original dimensions. Recall dropped sharply from 61.79% to 42.47% from the original to the medium resolution. The main reason for this observation is that after the first bicubic resampling, the resolution of every test map was lower than 300 DPI, which represents a critical benchmark for OCR (See Section 3.5) in general. Furthermore, resampling introduces noise that reduces graphical quality such as character clarity. This type of noise is similar to the type of error that can be introduced during the original sampling stage (scanning). Also, if the resampling process incorporates a lossy compression algorithm, the medium- and low-resolution images would show even nosier character representations and would have a lower recognition rate.

Figure 25 shows two example results. In these instances, Strabo detected the text locations correctly at all resolution levels, but Tesseract-OCR could not recognize some of the characters in the medium- and low-resolution images. Comparing the two cases, although “Wolsey” has a wider character spacing it has a cleaner representation (fewer smudges and bleedings) than “MADISON” in the original image. Therefore, when the image resolution was reduced to less than 300 DPI, the OCR tool showed a better recognition result for the downsampled text label “Wolsey” than for “MADISON”.

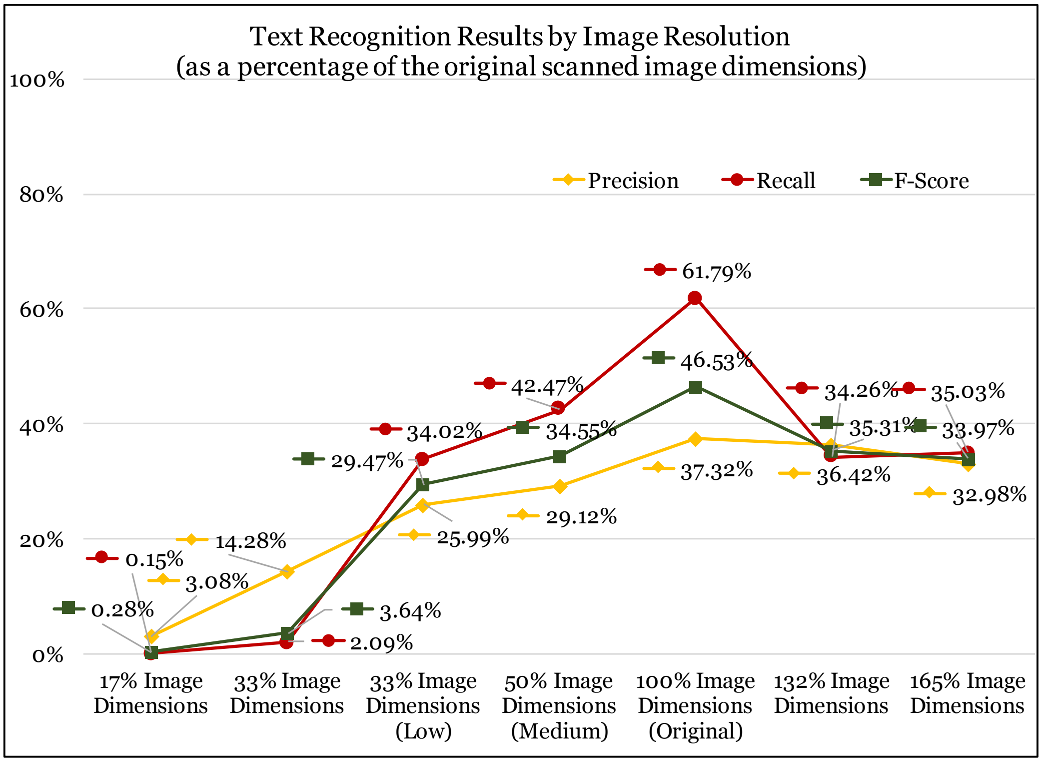
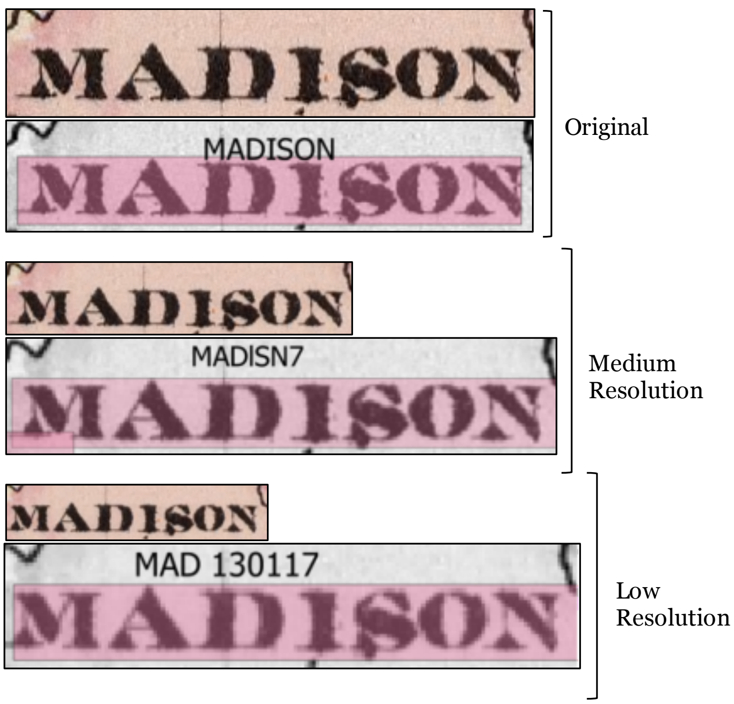


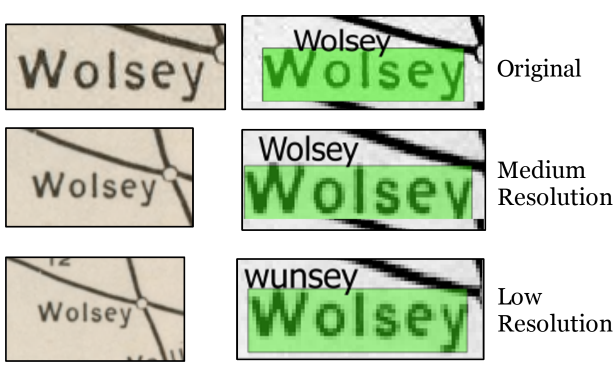
Figure 24. Experimental results by image resolution level

Table 2 shows the character-level precision, recall, and F-Score for each character group (groups 1-3; see Section 4.2) at each of the tested image dimensions, including the original, medium, and low resolutions. Group 1 contains 2,024 characters (35.51% of the total number of characters). Group 2 contains 896 characters (15.72% of the total number of characters). A closer look at the results for Group 2 reveals that non-text objects near existing words could be incorrectly detected as characters and hence a text label could be incorrectly broken into several parts (Figure 26). Also, it should be noted that the F-Score of Group 2 in the original resolution was close to the F-Score of Group 1 in the medium resolution. This illustrates that an improperly prepared map scan could largely reduce the prospect of using an automatic/semi-automatic map processing tool even if the map labels were clean, clear, and noise-free. In addition, when the resolutions were lower than 300 DPI, non-text objects were more likely to be grouped with nearby characters, so the precision of Group 2 was even lower than Group 3 in both the medium and low resolutions.

The third group contains 2,780 characters (48.77% of the total number of characters). In the experiment, this group included mostly text labels that overlap with (or touch) other map features (e.g., lines) or appear significantly curved. Strabo employed a recent method for detecting text labels overlapping with other features (Honarvar Nazari et al., 2016), but such overlaps still pose a major difficulty for OCR. As expected, Group 3 had the lowest values for recall and F-Score across the three image resolutions.



(a) The label “MADISON” in the test map of Florida



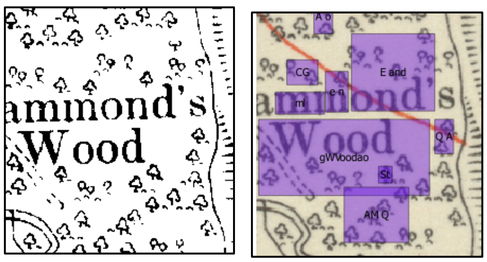
(b) The label "Wolsey" in the test map of South Dakota

Figure 25. Comparison of text recognition results for the same text label at three different image resolutions for two cases. The color images (top in (a) and left in (b)) show the map labels. The purple (a) and green (b) areas in the result images (bottom in (a) and right in (b)) are the Strabo-identified text locations. The black characters on top of the identified locations are the recognition results.

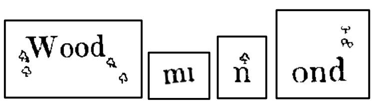
Table 2. Experimental results by character groups and image resolutions.

Image Dimension and Character Groups	Precision	Recall	F-Score
165% of the original image dimensions			
Group 1	41.31%	51.55%	45.87%
Group 2	24.56%	41.02%	30.72%
Group 3	28.82%	21.07%	24.34%
132% of the original image dimensions			
Group 1	44.74%	47.65%	46.15%
Group 2	27.84%	38.41%	32.29%
Group 3	32.71%	23.18%	27.13%
Original image (original resolution)			
Group 1	47.55%	83.50%	60.60%
Group 2	29.57%	71.65%	41.87%
Group 3	32.05%	42.81%	36.65%
66% of the original image dimensions (medium resolution)			
Group 1	37.32%	57.91%	45.39%
Group 2	20.46%	46.43%	28.41%
Group 3	26.51%	29.96%	28.13%
50% of the original image dimensions (low resolution)			
Group 1	31.30%	40.51%	35.31%
Group 2	16.84%	31.92%	22.05%
Group 3	23.07%	20.79%	21.87%
33% of the original image dimensions			
Group 1	19.02%	4.34%	7.07%
Group 2	4.28%	0.26%	0.49%

Group 3	9.09%	1.04%	1.86%
17% of the original image dimension			
Group 1	1.67%	0.06%	0.13%
Group 2	0.00%	0.00%	0.00%
Group 3	3.65%	0.26%	0.48%



(a) The detected text labels (text is part of the black layer) in purple boxes and the recognition results (the text labeled inside the purple boxes in Arial)



(b) Four of the detected text areas

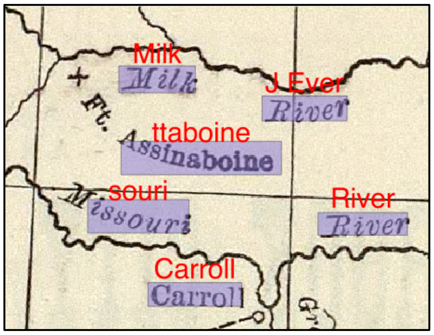
Figure 26. A noisy text area (Group 2) and the text detection and recognition results for these characters and strings are shown.

Further, when the image dimension was increased (132% and 165%), the recognition results showed a decrease in all accuracy measures compared to the results from the original resolution. This shows that after scanning, we could not add more information (i.e., to increase the DPI) to the map image for improving the recognition results (by upscaling the image). Table 2 also shows that when the image resolution dropped to 17% of the original resolution (less than 100 DPI), we could not correctly detect any character in Group 2. This was due to the fact that beyond 100 DPI, most of the characters became too blurry to be detected after the bicubic resampling.

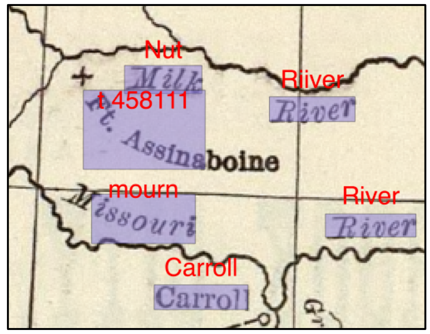
Figure 27 shows some example recognition results for text labels from every group at the three different resolutions. The word “Milk” was only correctly recognized in the original resolution. The uncommon character style of “Milk” resulted in poor OCR results when the resolution decreased. The curved words

“Ft. Assinaboine” and “Missouri” were broken into smaller parts during the text detection steps, so only parts of them were recognized by OCR. Moreover, curved strings were difficult for OCR to process. As an example, all characters except one of the detected label “Ft. Assina” in the medium resolution were recognized incorrectly. As can be seen in Figure 27(c), when the resolution was reduced, Tesseract-OCR was unable to correctly segment individual characters in a detected label because the character spacing was too small. For example, the word “Caroll” was recognized as “cmau” in the low-resolution image because Tesseract-OCR grouped some adjacent characters as single characters. Also, the characters “Ri” in the lower occurrence of the label “River” were incorrectly segmented into the two characters “IN”. The problem of erroneous character segmentation becomes more problematic when a word overlaps with other map features. For example, the characters “Ri” in the top occurrence of the label “River” were incorrectly segmented into the three characters “J E” in the original resolution because of the grid line between “R” and “i”. When the resolution decreased to medium, the characters “Ri” were segmented into “Rii” because the number of pixels between “R”, the gridline, and “i” were smaller (than in the original) and hence the space character was not in the recognition result.

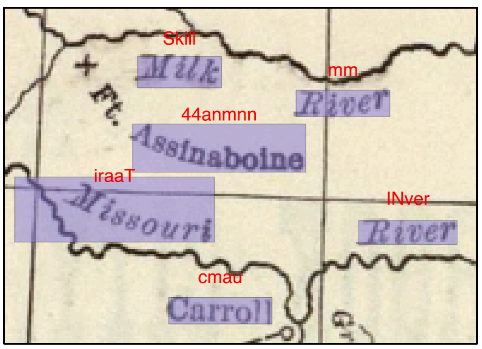
As can be seen in Table 2, even when a map was carefully prepared (scanned) such that high levels of image quality could be warranted, significant challenges remain in recognizing map text in a fully automated setting due to the complexities and variations in map properties. These graphical properties, here of characters and text labels, could even vary considerably across one map sheet, and the performance of map processing techniques directly relates to such properties. Such variations would remain hidden if accuracy would only be assessed over all labels as a whole without distinguishing between levels of graphical quality, feature representations, and map products. If incorporated into accuracy assessments this knowledge provides a more objective basis to estimate the suitability of a considered map for automatic processing (e.g., text recognition). For example, if the vast majority of characters or text labels in the map of interest belong to Group 1 and the resolution satisfies basic benchmarks for robust OCR performance the user could expect a good potential for automated or semi-automated map processing. In contrast, if most characters would be categorized as Group 3 the potential for automation would be expected to be very low without further tuning or training. This potential would be expected to further decrease for lower levels of image resolution.



(a) Original resolution



(b) Medium resolution (66%)



(c) Low resolution (50%)

Figure 27. Example text labels and their recognition results (text labels in red) across the three test image resolutions. The images of the medium and low resolution are enlarged here to better illustrate the results. The labels "Milk" (deformed characters), "River" (top, overlapping with a grid line), "Ft. Assinaboine" (curved over 30%), "Missouri" (overlapping with the grid line and curved over 30%) belong to Group 3. The label "River" (bottom right, uncommon font) belongs to Group 2 and the label "Carroll" is an example of Group 1.

Overall, in the described experiments additional OCR training and incorporating and tuning symbol recognition algorithms to remove non-text objects would likely improve the recognition accuracy in Groups 1 and 2 but still require user intervention to some degree. In Group 3, additional text/graphics separation techniques and dictionaries could be used to recover overlapping text in the OCR step, but this would require great amounts of effort by the user. For example, in a string "House", if the character "s" was removed due to overlapping features and "Hou e" was recognized, a dictionary containing the word

“House” could facilitate the reconstruction of the full word. Finally, crowdsourcing approaches such as CAPTCHA[20] could be used to scale up the result curation task and make it possible for an organization or user to process large volume map series with reasonable degrees of efficiency.

6. Summary and Outlook

This article discussed a variety of criteria to evaluate the suitability of scanned and digitally produced maps for automatic map processing using text recognition as the target application. This discussion fills an important gap in the literature which to-date has not seen an explicit and systematic assessment of the potential impacts of graphical quality issues on automatic or semi-automatic map processing tasks. The usefulness of the map/text criteria was demonstrated in an extensive experiment to test a common text recognition tool for maps, Strabo, for different map products at varying image resolutions. The results for each resolution were assessed, separately, for three groups of text representations defined based on the graphical characteristics of map text. This study is meant to support potential users of map processing tools to better understand (1) whether or not the map images of interest are suitable candidates for higher degrees of automation in map processing, (2) how much user intervention would be required and (3) how much variation in methods performance and thus in intervention needs can be expected. We view this article as a first step to systematically evaluate the potential to successfully process different maps and map series using an automatic or semi-automatic recognition system. Such a state-of-the-art introduction manual, here focused on text recognition, will help users interested in applying digital map processing systems to better understand current possibilities from the perspective of graphical quality and inherent uncertainty. This discussion could be further extended to other processing techniques such as line detection or symbol recognition in scanned maps.

Acknowledgements

This research is based upon work supported in part by the University of Southern California under the Undergraduate Research Associates Program (URAP).

[20] http://www.captcha.net/

We thank Rashmina Ramachandran Menon for creating a part of the experiment ground truth and Ronald Yu for managing the Strabo GitHub repository.

References

Bessaid, A., Bechar, H., and Ogier, J.-M. (2003). Automatic topographic color map analysis system. In *Proceedings of the 3rd International Workshop on Pattern Recognition in Information Systems*. 191–196.

Bin, D. and Cheong, W. K. (1998). A system for automatic extraction of road network from maps. In *Proceedings of the IEEE International Joint Symposia on Intelligence and Systems*. 359–366.

Bradley, D., & Roth, G. (2007). Adaptive thresholding using the integral image. *Journal of graphics, gpu, and game tools*, *12*(2), 13-21.

Bucha, V., Uchida, S., and Ablameyko, S. (2007). Image pixel force fields and their application for color map vectorisation. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition*. Vol. 2. 1228 –1242.

Cao, R. and Tan, C. L. (2002). Text/graphics separation in maps. In *Proceedings of the Fourth IAPR International Workshop on Graphics Recognition Algorithms and Applications*. 167–177.

Chen, L.-H., Liao, H.-Y., Wang, J.-Y., and Fan, K.-C. (1999). Automatic data capture for geographic information systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,* 29(2), 205 –215.

Chen, Y., Wang, R., & Qian, J. (2006). Extracting contour lines from common-conditioned topographic maps. *IEEE Transactions on Geoscience and Remote Sensing*, 44(4), 1048–1057.

Cherkassky, V. and Mulier, F., (1998), Learning from data: Concepts, theory, and methods. New York, NY, USA: Wiley.

Chiang, Y.-Y. and Knoblock, C. A. (2013). A general approach for extracting road vector data from raster maps. *International Journal on Document Analysis and Recognition*, 16(1), 55–81.

Chiang Y.-Y. and Knoblock C. A., (2014). Recognizing text in raster maps. *GeoInformatica*, 19(1), 1-27.

Chiang, Y.-Y., Leyk, S., Knoblock, C. A. (2013). Efficient and robust graphics recognition from historical maps. In *Graphics Recognition. New Trends and Challenges, Lecture Notes in Computer Science, 7423*, Y.-B. Kwon and J.-M. Ogier, Eds., Springer, Berlin Heidelberg, 25–35.

Chiang Y.-Y., Leyk S. and Knoblock C.A. (2014). A Survey of Digital Map Processing Techniques. *ACM Computing Surveys,* 47(1): 1-44

Chiang, Y.-Y., Moghaddam, S., Gupta, S., Fernandes, R., & Knoblock, C. A. (2014). From Map Images to Geographic Names. In *Proceedings of the 22th International Conference on Advances in Geographic Information Systems.*

Cordella, L. P. and Vento, M. (2000). Symbol and shape recognition. In *Proceedings of the 3rd International Workshop on Graphics Recognition, Recent Advances (GREC'99).* Springer-Verlag, London, UK, 167–182.

den Hartog, J., ten Kate, T., and Gerbrands, J. (1996). Knowledge-based segmentation for automatic map interpretation. In *Graphics Recognition Methods and Applications*, R. Kasturi and K. Tombre, Eds. Lecture Notes in Computer Science, 1072. Springer Berlin, 159–178.

Dhar, D. B. and Chanda, B. (2006). Extraction and recognition of geographical features from paper maps. *International Journal of Document Analysis and Recognition,* 8(4), 232–245.

Fernandes, R., Chiang, Y. Y. (2015). Creating an Intuitive and Effective User Interface for Map Processing in a Geographic Information System. In *Proceedings of the 27th International Cartographic Conference* (to appear).

Gelbukh, A., Levachkine, S., and Han, S.-Y. (2004). Resolving ambiguities in toponym recognition in cartographic maps. In *Proceedings of the 5th IAPR International Workshop on Graphics RECognition*, 104-112.

Helinski, M., M. Kmieciak, and T. Parkola. (2012). Report on the comparison of Tesseract and ABBYY FineReader OCR engines, IMPACT technical report.

Henderson, T. C. (2014) Analysis of Engineering Drawings and Raster Map Images, Springer-Verlag New York, ISBN: 978-1-4419-8166-0.

Henderson, T., C., Linton, T., Potupchik, S., and Ostanin, A. (2009). Automatic segmentation of semantic classes in raster map images. In *Proceedings of the Eighth IAPR International Workshop on Graphics Recognition*. 253–262.

Honarvar Nazari, N., Tan, T. X., Chiang, Y.-Y. (2016) Integrating Text Recognition for Overlapping Text Detection in Maps. In Proceedings of the 23rd Document Recognition and Retrieval Conference (to appear).

Itonaga, W., Matsuda, I., Yoneyama, N., and Ito, S. (2003). Automatic extraction of road networks from map images. *Electronics and Communications in Japan (Part II: Electronics),* 86(4) 62–72.

Kim, N. W., Lee, J., Lee, H., & Seo, J. (2014). Accurate segmentation of land regions in historical cadastral maps. *Journal of Visual Communication and Image Representation*, 25(5), 1262-1274.

Khotanzad, A. and 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.Leyk, S., Boesch, R., and Weibel, R. (2006). Saliency and semantic processing: Extracting forest cover from historical topographic maps. *Pattern Recognition*, 39(5), 953–968.

Leyk, S. and Boesch, R. (2009). Extracting composite cartographic area features in low-quality maps. *Cartography and Geographical Information Scien*ce 36(1), 71–79.

Leyk, S. (2010). Segmentation of colour layers in historical maps based on hierarchical colour sampling. In *Ogier J.-M., Liu W., and Lladós J. (eds.): Graphics Recognition, GREC 2009, Lecture Notes in Computer Science* 6020, pages 231-241.

Leyk, S. and Boesch, R. (2010). Colors of the past: color image segmentation in historical topographic maps based on homogeneity. *GeoInformatica* 14(1):1-21.

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

Liu, Y. (2002). An automation system: Generation of digital map data from pictorial map resources. *Pattern Recognition* 35(9), 1973–1987.

Llados, J., Valveny, E., Sanchez, G. and Marti, E. (2002). Symbol recognition: Current advances and perspectives. In *Graphics Recognition Algorithms and Applications, D. Blostein and Y.-B. Kwon, Eds., Lecture Notes in Computer Science, Vol. 2390*. Springer, Berlin, 104–128.

Shukla, M. K. and Banka, H. (2014) Degraded Script Identification for Indian Language – A Survey. *International Journal of Computer Applications*, 108(6), 11-22.

Marr, D., 1982, Vision. San Francisco, CA, USA: W.H. Freeman and Company.

Miyoshi, T., Li, W., Kaneda, K., Yamashita, H., and Nakamae, E. (2004). Automatic extraction of buildings utilizing geometric features of a scanned topographic map. In *Proceedings of the 17th International Conference on Pattern Recognition*. Vol. 3. 626–629.

Nagy, G., Samal, A., Seth, S., Fisher, T., Guthmann, E., Kalafala, K., Li, L., Sivasubramaniam, S., and Xu, Y. (1997). Reading street names from maps - technical challenges. In *GIS/LIS conference*, 89–97.

Pouderoux, J., Gonzato, J. C., Pereira, A., and Guitton, P. (2007). Toponym recognition in scanned color topographic maps. In *Proceedings of the 9th International Conference on Document Analysis and Recognition*, Vol. 1, 531–535.

Raveaux, R., Barbu, E., Locteau, H., Adam, S., Héroux, P., and Trupin, E. (2007). A graph classification approach using a multi-objective genetic algorithm application to symbol recognition. In *Graph-Based Representations in Pattern Recognition, F. Escolano and M. Vento, Eds. Lecture Notes in Computer Science, vol. 4538*. Springer, 361–370.

Raveaux, R., Burie, J.-C., and Ogier, J.-M. (2008). Object extraction from colour cadastral maps. In *Proceedings of the IAPR International Workshop on Document Analysis Systems*. Vol. 0. 506–514.

Rice, S. V., Jenkins, F. R., Nartker, T. A. (1995) The fourth annual test of OCR accuracy. Technical Report 95-04, Information Science Research Institute, University of Nevada, Las Vegas

Simon, R., Pilgerstorfer, P., Isaksen, L., & Barker, E. (2014). Towards semi-automatic annotation of toponyms on old maps. e-Perimetron, 9(3), 105-112.

Smith, R. (2007) An Overview of the Tesseract OCR Engine. In *Proceedings of the International Conference on Document Analysis and Recognition*, 7(1), 629–633.

Trier, O., Taxt, T., and Jain, A. (1997). Recognition of digits in hydrographic maps: binary versus topographic analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 19(4), 399–404.

Velázquez, A. and Levachkine, S. (2004). Text/graphics separation and recognition in raster-scanned color cartographic maps. In *Graphics Recognition, Lecture Notes in Computer Science, vol. 3088*. J. Llados and Y.-B. Kwon, Eds. Springer, 63–74.

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

Xin, D., Zhou, X., and Zheng, H. (2006). Contour line extraction from paper-based topographic maps. *Journal of Information and Computing Science,* 1(5), 275–283.

Ye, Q., & Doermann, D. (2014) Text Detection and Recognition in Imagery: A Survey. IEEE Transactions on Pattern Analysis & Machine Intelligence. doi: 10.1109/TPAMI.2014.2366765

Yin, P.-Y. & Huang, Y.-B. (2001). Automating data extraction and identification on Chinese road maps. *Optical Engineering* 40(5), 663–673.