J. Vis. Commun. Image R. 25 (2014) 1262–1274

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Accurate segmentation of land regions in historical cadastral maps

Nam Wook Kim [a], Jeongjin Lee [b,*], Hyungmin Lee [c], Jinwook Seo [c]

[a] Mobile Communication Research Center, LG Electronics, 219-24 Gasan-dong, Geumcheon-gu, Seoul 153-801, Republic of Korea
[b] School of Computer Science & Engineering, Soongsil University, 369 Sangdo-Ro, Dongjak-Gu, Seoul 156-743, Republic of Korea
[c] School of Computer Science and Engineering, Seoul National University, 599 Kwanak-ro, Kwanak-gu, Seoul 151-742, Republic of Korea

ARTICLE INFO

Article history:
Received 17 June 2013
Accepted 31 December 2013
Available online 11 January 2014

Keywords:
Cadastral map
Historical geographical information system
Line reconstruction
Line extraction
Polygonal approximation
Land segmentation
Character recognition
Grid removal

ABSTRACT

Historical cadastral maps are valuable sources for historians to study social and economic background of changes in land uses or ownerships. In order to conduct large-scale historical research, it is essential to digitize the cadastral maps. As being established in antiquity, however, they suffer from significant noise artifacts attributed to hand-drawn cartography. In this paper, we propose a novel method of extracting land regions automatically in historical cadastral maps. First, we remove grid reference lines based on the density of the black pixel with the help of the jittering. Then, we remove land owner labels by considering morphological and geometrical characteristics of thinned image. We subsequently reconstruct land boundaries. Finally, the land regions of a user's interest are modeled by their polygonal approximations. Our segmentation results were compared with manually segmented results and showed that the proposed method extracted the land regions accurately for assisting cadastral mapping in historical research.

1. Introduction

From the past to the present, land tenure and parcels have been crucial information for land administration such as land valuation and taxation [1]. To maintain the property records of a country, many nations adopted a cadastre survey, which involves the documentation of land registration such as location, area and ownership [2,3]. Two famous historical examples are the Domesday book from early England and the Napoleonian cadastre from 19th century France. Both of them laid the historical foundation for modern cadastral systems [4]. Historians use those records to study the evolving histories of land parcellation in conjunction with social and economic aspects of changes in land uses or ownerships [5].

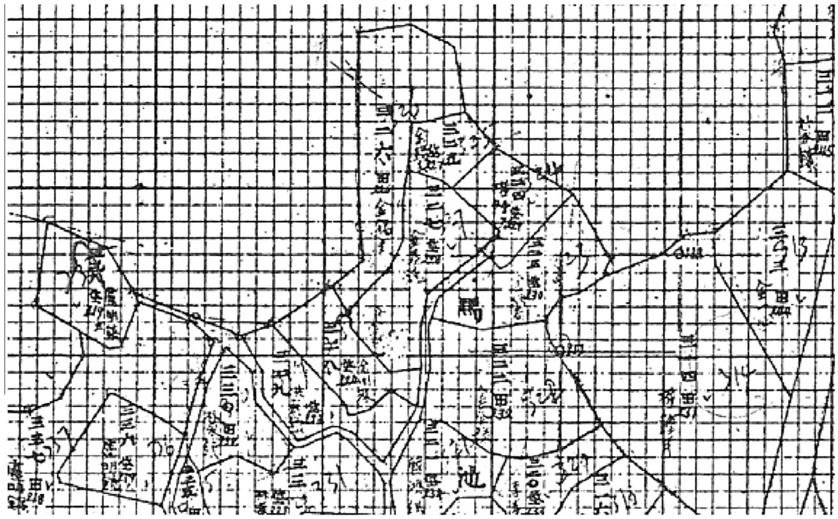
In Korea, the Kyujanggak Institute for Korean Studies (KIKS) preserves a significant number of cadastres from the 17th to 19th century, which cover major cities and suburban areas in the Joseon dynasty of Korean history [6]. A cadastral map has geographical boundaries of land ownership, while a textual cadastre is a tabular data including survey direction, neighborhood, area and owner (Fig. 1). Unfortunately, since they were recorded independently, the integrated cadastral research has been difficult. The researchers in the KIKS are currently working on constructing a mapping between the textual cadastre and cadastral map (i.e., cadastral mapping), which were recorded for same areas but at different times. Through this mapping, they expect to understand the temporal and spatial changes of the land ownership, development status, and residential areas in an integrative manner.

A typical task involved in the cadastral mapping is to use sticky notes and highlighters to mark the mapping from a land owner's name in the textual cadastre to the corresponding land region in the cadastral map [7]. Since this task is done on a physical copy of each cadastral map, it is difficult to undo if the mapping result turns out to be wrong. Also, such manual work is cumbersome for editing and searching the existing mappings. And the large quantity of cadastres makes maintenance tasks laborious.

To facilitate the cadastral mapping task, it is desirable to digitize the historical cadastral maps by allowing land regions to be searchable and manageable. Accordingly, there have been many works in building historical geographical information system (GIS) to assist cadastral mapping. A notable example is the Great Britain historical GIS [8]. It holds the changing boundaries of administrative units with historical statistics recorded from 1840 to 1970. Its GIS database is structured in a way that the land boundaries, if exist, are provided for specified dates. It was constructed by combining geographical maps with textual sources that provided specific dates for boundary changes. Another example is the China historical GIS which covers 2000 years of dynastic

* Corresponding author. Fax: +82 2 886 7589.
E-mail addresses: namw.kim@samsung.com (N.W. Kim), leejeongjin@ssu.ac.kr (J. Lee), hmlee@hcil.snu.ac.kr (H. Lee), jwseo@hcil.snu.ac.kr (J. Seo).

http://dx.doi.org/10.1016/j.jvcir.2014.01.001
1047-3203/



Global Index	Survey Direction	4-Neighbor Information				Area	Owner
		East	West	North	South		
212	West	JRW	LBK	Road	Mountain	783	KMB
213	West	KMB	LSW	Road	LKS	1445	LBK
214	West	LBK	Creek	Road	LKS	4080	LSW
215	North	Road	KSS	LSW	LYS	1924	LKS
216	West	LKS	Creek	LSW	LYS	2940	KSS

Fig. 1. A pre-modern cadastre in KIKS – cadastral map (top) and textual cadastre (bottom).

history in China [9]. Because of the poor accuracy in the records of many administrative regions, it did not attempt to reconstruct the exact boundaries. Instead, it used the locations of administrative units and their relative positions to approximate their boundaries. By integrating statistical data and geospatial data collected at different times (e.g., textual cadastre and cadastral map) into a single computer system, this system expedites historical research such as geo-referenced demographic study [10]. However, building those systems have never been easy particularly with historical cadastral data. It involves manual vectorization of spatial data into points, lines and polygons, which is a highly time-consuming and costly process [5]. To alleviate this problem, we need a more efficient way to perform the vectorization of geographical information.

This paper proposes a segmentation method that extracts the land regions accurately in historical cadastral maps. We use the cadastral maps from the KIKS (Fig. 1), which are currently being used by the Korean historians. As being established in pre-modern era, the maps suffer from significant noise artifacts attributed to hand-drawn cartography. They have not only compact grid lines and label characters, which are considered noise, but also eroded region boundaries. We designed a staged segmentation pipeline by devising a series of image processing techniques. We first eliminate noises in a scanned map image by removing grid reference lines based on the density of the black pixel with the help of the jittering. Then, we remove land owner labels by considering morphological and geometrical characteristics of thinned image. Then, we subsequently reconstruct broken land boundaries which are originated from both the eroded map and noise reduction phase. Finally, we extract the land regions of user's interest by generating their polygonal approximations.

The remainder of this paper is organized as follows. The next section describes the proposed method of automatically extracting the land region. Section 3 presents the experimental results. Section 4 describes one of applications for assisting historical research using the proposed method, followed by conclusion and future work in Section 5.

2. Related work

In the traditional practice of cartography, maps such as topographic or cadastral maps were hand-drawn on papers [11]. Nowadays, due to the scalability and efficiency, geographical information for urban planning or resource management is processed through computer systems [12]. With the recent proliferation of GIS applications, there has been an increasing need for converting existing analog maps to vector forms [13]. The vector data has many advantages over the raster data by encoding the topological structure of a map only in the form of points and lines. In addition, it takes less storage. It is not limited by spatial resolution, and is easier to manage and update [14].

The digitization is often performed manually either through a digitizing tablet with a paper map on its surface or an on-screen digitization using a scanned map [15]. Since the manual digitization is time consuming with intrinsic human errors, more advanced semi-automated or fully-automated procedures of the extraction of cartographic information from maps have been proposed [16]. Unfortunately, a fully-automatic vectorization is still a challenging task as types of maps vary considerably and human verification is almost always necessary [17]. In particular, it is considerably difficult to accurately vectorize historical maps because of its poor graphical quality caused by scanning or image compression processes, as well as the aging of the archived paper material, which often causes false coloring, blurring or bleaching problems [16,17].

A typical automatic vectorization procedure involves (a) digitization of paper documents using a scanner, (b) filtering, (c) thres-