Appl Geomat (2012) 4:225–229
DOI 10.1007/s12518-011-0064-y

ORIGINAL PAPER

Blind watermarking technique for topographic map data

Haowen Yan · Jonathan Li

Received: 9 June 2010 / Accepted: 20 September 2011 / Published online: 1 October 2011

Abstract A blind watermarking technique for protecting topographic data from illegal use is proposed, taking into account four rules, i.e., usability, invisibility, robustness, and blindness. The technique firstly determines two feature layers and selects the key points from each layer as watermark embedding positions; then it shuffles the watermark and embeds the watermark in the two layers, respectively. To detect the watermark, a similar process for obtaining the feature layers and the key points in the watermark embedding process is carried out first; then the coordinates of the key points are detected to extract the embedded watermark; finally, the similarity degrees of the two versions of the extracted watermark is calculated, by which the conclusion on whether the data contains the watermark is made. Our experiments show that the technique can resist the attacks from data format change, random noise, similarity transformation, and data editing to some extent.

Keywords Blind watermark · Topographic data · Key point · Similarity degree

Introduction

Topographic map data is of great value because the acquisition of such data is a high-cost process. Consequently, it cannot be freely used without the owner's permission. Nevertheless, the rapid development of computer communication and Internet techniques make it easy to duplicate and distribute digital data via networks, which troubles the data owners for protecting the data from piracy.

H. Yan (✉) · J. Li
Department of Geography & Environmental Management,
University of Waterloo,
Waterloo, ON N2L 3G1, Canada
e-mail: haowen2010@gmail.com

Digital watermarking provides a viable solution for this dilemma. A digital watermark is an imperceptible but identifiable digital signal or mode embedded in the host data, while it does not affect the host data's usability (Ahmed 2004). There are four important rules that should be obeyed in any successful watermarking techniques (Cox and Miller 2002; Zhou et al. 2006). First of all, the embedded watermark should not degrade the quality of the host data. Secondly, the watermark should be perceptually invisible to data users to maintain its protective secrecy. Next, the technique must be robust enough to resist common data processing attacks and not be easily removable by illegal users, but only the data owners ought to be able to extract the watermark. Finally, the watermark should be blind if it is difficult for the data users to obtain the original data and the original watermark. Not only have the techniques of digital watermarking received a great deal of attention to ensure copyright protection for video, audio, and image data, but also it has become a hot issue in the community of geoscience for protecting vector geospatial data from piracy. Generally, there are two categories of watermarking algorithms in this area, i.e., space domain and transform domain. The following presents a critical analysis of the later, for this paper aims at proposing a watermarking approach in space domain.

Most of the existing space domain algorithms (e.g., Samoa et al. 2000; Kang 2001; Ohbuchi et al. 2002; Li and Xu 2004; Schulz and Voigt 2004) are based on the idea of changing the positional relations of the points in vector maps. The principle of these algorithms are as follows: subdivide a vector map into rectangular blocks of adaptive sizes according to the density of vertices and embed each bit of the watermark by displacing the coordinates of a set of the vertices in the block. There are also some algorithms (Park et al. 2002; Sonnet et al. 2003) insert new points into the original data and take them as the watermark embedding positions. Moreover, an algorithm proposed by Jia et al. (2006) inserts the bits of the watermark into the least significant bits (LSBs) of the coordinates to make the watermark capable of resistant to the data revision. The advantage of space domain algorithms is the precision of the watermarked data is controllable, and the watermarks generated by these algorithms are generally resistant against additive random noise, similarity transformation, and vertices revision, to some extent. However, none of such algorithms are blind in detection process.

To critically sum up the above review on the watermarking techniques in space domain: (1) the space domain algorithms prevail over the transform domain ones in preserving the precision of watermarked data. (2) They do not differentiate among point, linear, and areal objects, and little of the spatial characteristics of geospatial data is taken into consideration in the existing algorithms. (3) None of the space domain algorithms are blind in detection processes.

Watermark-embedding algorithm

General thought

To overcome the shortcomings in the existing space domain algorithms for topographic map data, some appropriate strategies are employed in the new approach. These include: embedding of the watermark twice to make its detection blind, using key points of the features as the watermark embedding positions to improve its robustness, and utilizing the LSBs of the key point coordinates to maintain the quality of the watermarked data. Based on these strategies, an algorithm for embedding watermarks in topographic map data is proposed, including three procedures: determination of the feature layers, selection of the key points, and embedding of the watermark.

Embedding feature layer selection

The feature layers selected for the watermark embedding should abide by the following rules.

- At least two feature layers need to be selected for watermark embedding.
- The number of the points in each selected feature layer should be greater than N (N is the bit number of the watermark).
- The more important a layer is, the more probable the layer should be selected.
- Control point layer is not allowed to be selected.

In this approach, two feature layers are selected. They may be linear or/and areal.

Key point selection method for linear feature layers

To detect key points from linear features, both geometric and geographic characteristics of the feature should be taken into consideration (Douglas and Peucker 1973; McMaster 1987; Nickerson 1988; Beard 1991), and different methods are needed for different feature layers. Here, roads are taken as a representative, and a new method for key point detection in the road layer is addressed, comprising the following three steps.

Step 1, calculation of topological relations. This includes the calculation of the connectivity and adjacency relations among the lines and the construction of the road entities according to their topological relations.

Step 2, selection of road terminals. Suppose that there are total N_1 roads. The length values of the roads are sorted in decreasing order. If $N_1 > \frac{N}{2}$, select $\frac{N}{2}$ terminals of the roads that own greater length values as the watermark insertion positions, and end the procedure; or else, select all of the terminals and go to step 3.

Step 3, selection of the intersections. Firstly, select $\frac{N}{2} - N_1$ roads that each road satisfies the following two criterions: (1) it has intersections with the other roads and (2) at least one intersection has not been selected for watermark insertion. Secondly, obtain the key point from each of the roads by: (1) calculating the distance between each unselected intersection and the line segment linking the two terminals of the road and (2) select the intersection with the greatest distance as the watermark insertion position (see Fig. 1).

Key point selection method for areal feature layers

The areal feature layers that consist of connected polygons and that consist of disjoined polygons should be differen-

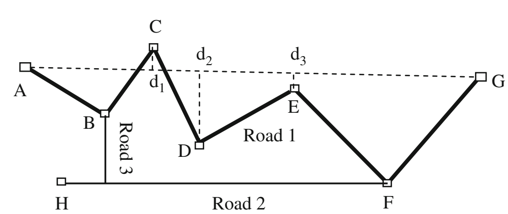

Fig. 1 Demonstration of the intersection selection: for *F* and *B* are the terminals of *Road 2* and *Road 3*, respectively, and have been selected, *D* is selected here, due to its greatest distance to *AG*

tiated in the key point selection methods, due to their different geometric characteristics.

Method for connected areal features

The method for key point selection from connected polygonal objects consists of the following five steps.

Step 1, construct topological polygons of the feature layer.

Step 2, calculate the number of the joint points. A joint point means the point is owned by at least three polygons.

Step 3, calculate the degrees of the joint points and the length values of the common edge owned by two polygons and then sort the joint points in decrease order by their degrees. If two joint points have same degree, they are sorted in decrease order by the sum length values of the edges they joint. Save the sequence number of the sorted joint point in a 1D array *B*.

Step 4, let the total number of the joint points be N_j. If $N_j \geq N$, the joint points whose sequence numbers between B_0 and B_{N-1} are selected as watermarking positions and the procedure is ended; or else, select all of the joint points and go to step 5.

Step 5, sort the length values of the common edges of the polygons in decreasing order, and select $N-N_j$ edges with greater length values and extract one key point from each edge, using a distance-based method shown in Fig. 2.

Method for disjoined areal features

A disjoined areal feature layer consists of topologically separated polygons. Suppose that the polygon number is N_d. To obtain the key points, the following steps are needed.

Step 1, the area of each polygon is calculated.

Step 2, the areas are sorted in decreasing order.

Step 3, if $N_d \geq N$, take *N* polygons with greater areas and select only one key point from the vertices of each polygon; or else, select one or more than one key points from each of the N_d polygons so that the total number of the key points equals to *N*.

A method based on deviation angles (the definition is shown in Fig. 3) and the polygon's edge length values is used for selecting key points (Fig. 4) from each polygon. Firstly, sort the deviation angle values in increasing order and save the sequence numbers of the corresponding vertices in a 1D array, say *V*; and then delete the sequence number in *V* whose corresponding vertex owns a joint edge shorter than the mean length of the edges. Finally, select the required vertices according to the sequence number of the vertices recorded in *V*.

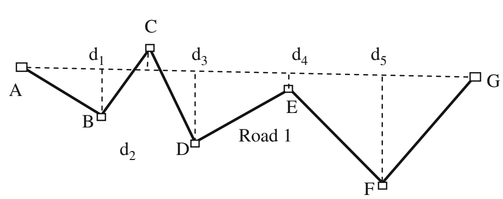

Fig. 2 Principle of the distance-based method: the point *F* is selected, for it has the greatest distance to the line segment linking the two terminals of the edge

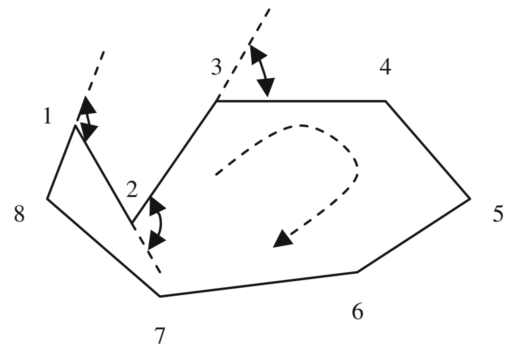

Fig. 3 Deviation angle means an angle between the extension line and the other edge of the same joint vertex (the extension line is in anticlockwise of the polygon). The deviation angles of vertices *1*, *2*, and *3* are shown

Watermark embedding

The watermark used in this technique is a string. To increase the difficulty of removing the watermark, the bits of the American Standard Code for Information Interchange codes of the characters in the string are shuffled first; and then the bits are embedded in the LSBs (Jia et al. 2006) of the coordinate *x* (or *y*) of the selected key points.

Watermark detection approach

The purpose of watermark detection is to find if the data set contains the specific watermark. Four steps are needed in this process: determination of the embedding positions, extraction of the bit chain, reconstruction of the watermark, and comparison of different versions of the extracted watermarks and the decision making.

Step 1, determination of the embedding positions. Firstly, the information about the watermarked layers is obtained from the previous recording. Then the key points are selected from each layer using the same methods as the ones used in watermark embedding. The obtained sequence numbers of the key points from the feature layers are recorded.

Step 2, extraction of the embedded bits. Read each LSB of the coordinate *x* of each key point in turn and form a bit chain using the bits extracted from each layer.

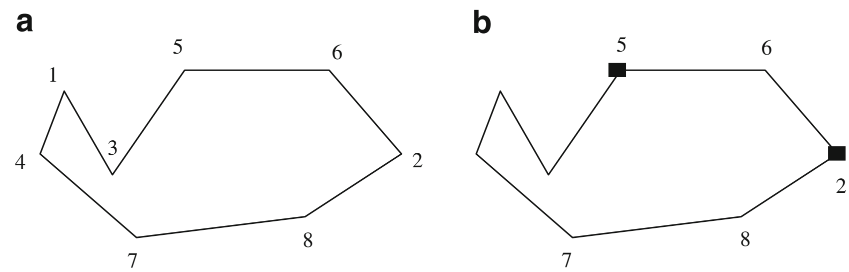

Fig. 4 Demonstration of key point selection from polygons: **a** sort the deviation angles in decreasing order and **b** delete the vertices with shorter joint edges, and select vertices *2* and *5* as the key points

Step 3, reconstruction of the watermark. This is an inverse operation of the watermark shuffling in the preparation of the watermark.

Step 4, comparison and decision making. The equation for the calculation of the similarity degree of any two bit chains with same number of bits is as follows.

$$D = \frac{N_s}{N} \quad (1)$$

where, D is the similarity degree of the two bit chains; N_s is the number of the bits with equal value in the two bit chains, and N is the total number of bits in one bit chain.

If the similarity degree is greater than the given threshold value, it can be concluded the data contains the watermark.

Experimental studies

The proposed approach has been implemented in Visual C++ (version 6.0). To verify its correctness and soundness, a set of experiments have been done using various data sets. The following presents one of the experiments in detail.

The data set used in this experiment is a topographic map at scale 1:1,000,000 (Fig. 5a), in SHAPE file format, freely downloaded from the National Fundamental Geographic Information System of China. There are totally seven feature layers in the dataset including two point layers (cities and lakes), three linear layers (high roads, railroads, and rivers) and two polygonal layers (hydrobodies and boundaries). The watermark used in the experiment is the string "National Fundamental GIS" (the number of the bits $N=24\times7=168$).

In light of the criterion given in the watermark-embedding layer selection, the provincial and national boundaries (polygonal layer) and the railways (linear layer) were selected (Fig. 5b). One hundred sixty-eight key points were chosen from each of the two layers using above proposed methods, and the bits of the string were embedded in the coordinates x of the key points.

The approach is evaluated from the following four aspects.

1. Usability: the usability of the watermarked data can be evaluated at scientific level by means of analyzing the relative errors of the data. According to the calculation and statistic of the positional changes of all coordinates x used for watermark embedding, none of the relative

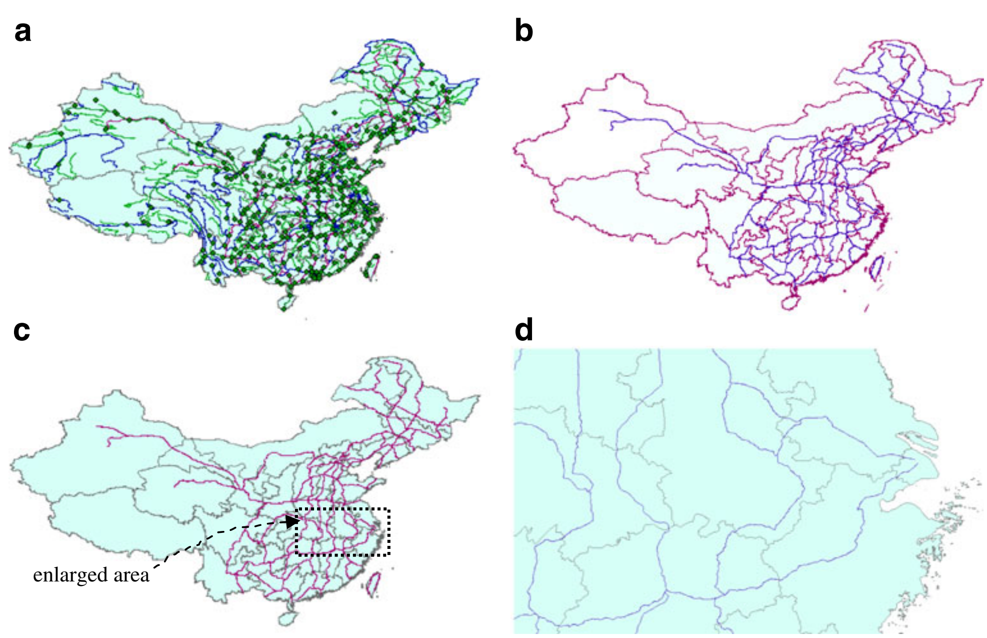

Fig. 5 Test of the watermarking technique using topographic map data from the National Fundamental Geospatial Database at 1:1,000,000. **a** Original map data with seven feature layers, **b** selected two feature layers for watermark embedding, **c** overlap of the original data and the watermarked data, and **d** an enlarged area for showing the quality of the watermarked data

error is greater than two times of the mean square error (the tolerance value of most standards for spatial data) of the coordinates *x*; so the data with the watermark can still be used.

2. Invisibility: Fig. 5c is the overlap of the original feature layers and the corresponding watermarked feature layers; Fig. 5d is an enlarged area of Fig. 5c. It is clear that no visual difference can be found between the original feature layers and the watermarked ones. In other words, the embedded watermark is imperceptible to data users.
3. Robustness: five operations, including data format change, similarity transformation, random noise attack, point deletion (10% of the total points are deleted), and point insertion (10% of the total points are inserted), are exerted on the watermarked feature layers of the dataset, and the corresponding similarity degrees between each pair of extracted watermarks are 1, 1, 0.881, 0.905, and 0.874. Our experiments have proved that if the similarity degree is greater than 0.70, the two layers usually contains the same watermark. Therefore, it is obvious that the data format change and similarity transformation have no effects on the watermarked data and the watermarking technique is also robust to resist the attacks from random noise, data deletion, and data insertion.
4. Blindness: neither the original data nor the original watermark is needed in the watermark detection process, so this is a blind watermarking approach.

Conclusions

A blind watermarking approach to the copyright protection of vector topographic data has been proposed in this paper, which fills the gap in this issue. The approach embeds the watermark in twice in the host data. The watermark embedded by this approach does not change the topological relations among spatial objects, is perceptually invisible to data users, and is resistant to data format change, similarity transformation, and data editing, to some extents.

The approach has been implemented by the authors and the software has been used by the Lanzhou Bureau of Land Resource, Gansu Province, China, for duplicating data using hard disks and CDs, and distributing data via internet. Therefore, its soundness and validity has been proved. The data used in our experiments are all topographic maps, so whether this technique is adaptive to other types of vector geospatial data needs further investigation.

Acknowledgments We are grateful to the anonymous reviewers' helpful comments. This work is supported by the Ontario Graduate Scholarship, 863 programs (2009AA121404) and the Program for Changjiang Scholars and Innovative Research Team in University (IRT0966).

References

Ahmed AM (2004) Digital image watermarking using fuzzy logic and naturalness preserving transform. Ph.D. thesis, Kansas State University

Beard K (1991) Theory of the cartographic line revisited. Cartographica 28(4):32–58

Cox IJ, Miller ML (2002) The first 50 years of electronic watermarking. J Appl Signal Process 56(2):123–126

Douglas DH, Peucker TK (1973) Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. Can Cartogr 10(2):112–122

Jia PH, Chen YZ, Ma JS et al (2006) Digital watermark-based security technology for geo-spatial graphics data. Chin Geogr Sci 16 (3):276–281

Kang H (2001) A vector watermarking using the generalized square mask. In: Proceedings of the International Conference on Information Technology:Coding and Computing, April 2001, Las Vegas, USA. pp. 234–236

Li YY, Xu LP (2004) Vector graphical objects watermarking scheme in wavelet domain. Acta Photonica Sinica 33(1):97–100

McMaster RB (1987) Automated line generalization. Cartographica 24(2):74–111

Nickerson BG (1988) Automated cartographic generalization for linear features. Cartographica 25(3):15–26

Ohbuchi R, Ueda H, Endoh S, (2002) Robust watermarking of vector digital maps. In: Proceedings of IEEE International Conference on Multimedia and Expo 2002, September 2002. Lausanne, Switzerland. vol. 1, pp. 577–580

Park KT, Kim KI, Kang H et al (2002) Digital geographical map watermarking using polyline interpolation. In: Proceedings of the IEEE Pacific Rim Conference on Multimedia, December 2002, Hsinchou, Taiwan. pp. 58–65

Samoa M, Matsuura Y, Takashima Y (2000) A scheme of digital watermarking for geographical map data. In: Proceedings of the Symposium on Cryptography and Information Security, January 2000, Okinama, Japan. pp. 26–28

Schulz G, Voigt M (2004) A high capacity watermarking system for digital maps. In: Proceedings of the Multimedia and Security Workshop on Multimedia and Security, September 2004, Magdeburg, Germany. pp. 180–186

Sonnet H, Isenberg T, Dittmann J et al (2003) Illustration watermarks for vector graphics. In: Proceedings of the 11th Pacific Conference on Computer Graphics and Applications, October 2003, Calgary, Canada. pp: 73–82

Zhou X, Ren Y, Pan XZ (2006) Watermark embedded in polygonal line for copyright protection of contour map. Int J Comput Sci Netw Secur 6(7B):202–205