

A Blind Watermarking Approach to Protecting Geospatial Data from Piracy

Haowen Yan and Jonathan Li

Abstract—Vector geospatial data is of great value, due to the cost processes of acquiring such data. Thus, how to protect vector geospatial data from piracy has become a hot issue in the community of geographic information science, and among which watermarking has been proven a feasible tool. This paper proposes a blind watermarking technique for protecting vector geospatial data from illegal use, mainly taking into consideration four rules in watermarking, i.e. usability, invisibility, robustness, and blindness. The technique consists of two processes: a watermark embedding process and a watermarking extracting process. In the watermarking process, 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. At the beginning of the watermark extracting process, a step similar to that in the watermarking embedding process is carried out to obtain the two feature layers and the key points that have been used for embedding the watermark; then the coordinates of the key points are checked to extract the embedded watermark from the two feature layers, respectively; finally the similarity degree of the two watermarks extracted from two feature layers is calculated, by which the conclusion on whether the data contains the watermark can be made. Our experiments show that the technique can resist the attacks from data format change, random noise, similarity transformation, and data editing.

Index Terms—Blind Watermarking Technique, Key Point, Similarity Degree, Vector Geospatial Data

I. INTRODUCTION

Without doubt, the rise of geographic information industry has greatly improved the value of geospatial data in recent years. Nevertheless, the rapid development of computer communication and Internet techniques make it easy to duplicate and distribute such digital data via networks, which troubles the data owners for protecting the data from free use without their permission (i.e. piracy). For this reason, copyright protection of geospatial data has become a hot issue in the community of geographic information science.

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 [1]. There are four important rules that should be obeyed in any successful watermarking techniques [2], [3].

- Usability: the embedded watermark should not degrade the quality of the host data, i.e. the watermarked data still can be used.
- Invisibility: the watermark should be perceptually invisible to data users to maintain its protective secrecy.
- Robustness: 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.
- Blindness: the watermark should be blind if it is difficult for the data users to obtain the original data and the original watermark.

Have not only 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 been introduced into the area of cartography and geographic information science for protecting vector geospatial data from piracy. Two categories of watermarking algorithms have been mainly studied in this area, i.e. space domain and transform domain. The following presents a critical analysis of the former, for this paper aims at proposing a watermarking approach in space domain.

Most of the existing space domain algorithms [4]-[8] 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 [9], [10] insert new points into the original data and take them as the watermark embedding positions. Moreover, an algorithm proposed by Jia et al. [11] 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, we have the following conclusion (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

Manuscript received May 18, 2011.

Haowen Yan, School of Mathematics, Physics and Software Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China, (e-mail: h24yan@uwaterloo.ca).

Jonathan Li, Department of Geography & Environment Management, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada, (e-mail: Junli@uwaterloo.ca).

characteristics of geo-spatial data is taken into consideration in the existing algorithms. (3) None of the space domain algorithms are blind in watermark detection processes.

To overcome the above shortcoming in the existing algorithms, this paper focuses on proposing a blind watermarking algorithm in space domain for vector geospatial data. The characteristic of the new algorithm is: it selects and uses key points from different feature layers of geospatial data as the watermark embedding position, which makes it feasible to extract watermarks from feature layers of geospatial data without depending on original data and to recognize the extracted watermarks without depending on the original one.

The remainder of the paper is organized as follows: after this introduction, the new watermarking algorithm is presented in detail (section II), including key points selection methods and watermark processing and embedding methods; then the approach to watermarking detection and extraction is addressed (section III). These two parts comprise the main body of the paper. To demonstrate the soundness of the new technique, two experiments are shown in section IV. Finally, some conclusions are drawn.

II. APPROACH TO WATERMARK EMBEDDING

A. Line of thoughts

To make the new watermarking algorithm for geospatial data a successful one, a series of strategies are employed in the new approach.

- Embed the watermark twice to make the watermark detection a blind process (blindness); in the meanwhile, multiple embedding improves the robustness of the watermark.

- Use feature layers with greater importance and the key points of the features as the watermark embedding positions to ensure the robustness of the watermark.

- Utilize the LSBs of the key point coordinates as the watermark insertion positions to ensure the quality of the watermarked data (usability) and the imperceptible of the watermark to data users (invisibility).

The new algorithm for watermark embedding consists of three steps (Fig.1), including: (1) determination of the embedding feature layers, (2) selection of the key points, and (3) embedding of the watermark.

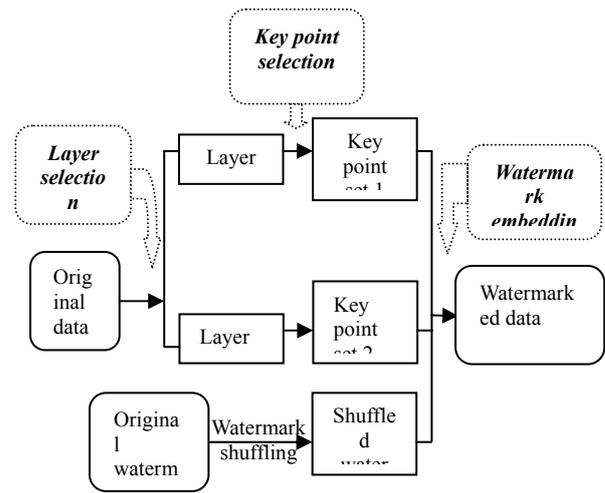


Fig.1. Flowchart of the watermark embedding algorithm

B. Embedding feature layer selection

A geospatial database generally consists of multiple feature layers (e.g. road, contour, building, etc.). To maintain the usability of the watermarked geospatial data and improve the robustness of the watermark, the feature layers used for watermark embedding should abide by the following rules.

- Two feature layers need to be selected for watermark embedding for the purpose of embedding the watermark twice.

- The number of the points (i.e. the vertices of lines or polygons) 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 used for watermarking.

After this step, two feature layers are selected. They may be linear and polygonal feature layers, so the following will present the key point selection methods for linear and polygonal (areal) features.

C. Key point selection method for linear feature layers

Both geometric and geographic characteristics of the feature should be taken into consideration in order to detect key points from linear features, [12]-[15], and different methods are needed for different linear features. Here, roads are taken as a representative, and a new method for key point selection (Fig.2) is proposed, comprising the following three steps.

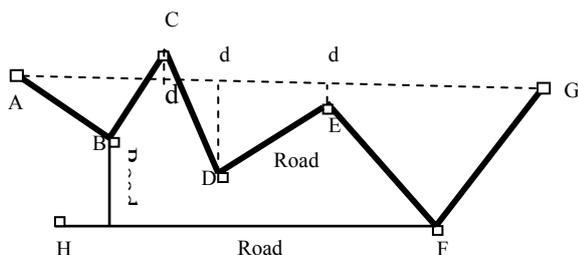


Fig.2. Demonstration of the intersection selection: for F and B are the terminals of Road 2 and Road 3, respectively, and have been selected, D is selected here, due to its greatest distance to AG.

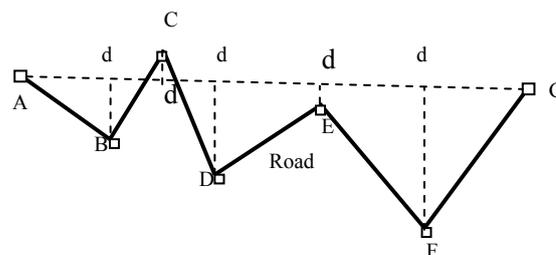


Fig. 3. Principle of the distance-based method: the point F is selected, for it has the greatest distance to the line segment linking the two terminals of the edge.

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 totally 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 criteria: (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.

D. Key point selection method for areal feature layers

There are two kinds of areal feature layers: connected polygons and disjointed polygons. They should be differentiated 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.

Step1, 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.3.

• Method for disjointed areal features

A disjointed 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 .

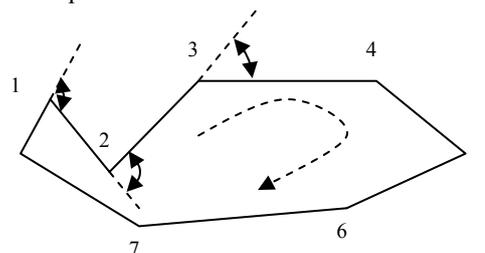


Fig.4. Deviation angle means an angle between the extension line and the other edge of the same joint vertex (the extension line is in anti-clockwise of the polygon). The deviation angles of vertices 1, 2, and 3 are shown.

A method based on deviation angles (the definition is shown in Fig. 4) and the polygon's edge length values is used for selecting key points (Fig.5) 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 .

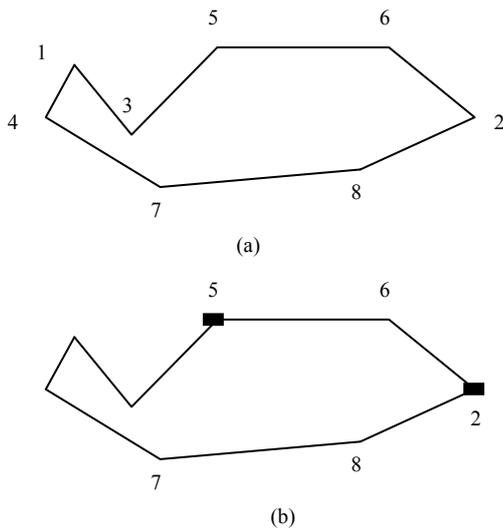


Fig.5. Demonstration of key point selection from polygons: (a) sort the deviation angles in decreasing order; and (b) delete the vertices with shorter joint edges, and select vertices 2 and 5 as the key points.

E. Watermark embedding

The watermark used in this technique is a string. To increase the difficulty of recognizing and removing the watermark, the bits of the ASCII 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.

III. APPROACH TO WATERMARK EXTRACTION

The purpose of watermark detection is to find if the data set contains the specific watermark. Four steps are needed in this process.

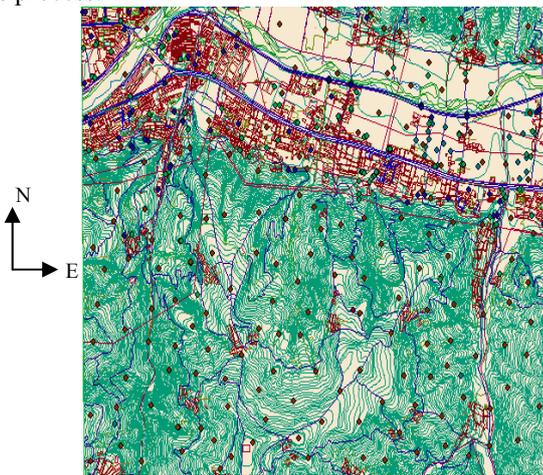


Fig.6. Geospatial data set used in the experiment

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.

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 formula 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.

IV. EXPERIMENTS

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. One of the experiments is presented here. The data set is a topographic map at scale 1: 50,000 (Fig.6), comprising 14 feature layers, provided by the Surveying and Mapping Bureau of Gansu Province, China. The data is in SHAPE file format. The roads (linear) and settlements (areal) are selected for watermark embedding. The watermark is the string "Gansu Surveying and Mapping".

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 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. 6 is the overlapped visualization of the original data set and the corresponding watermarked feature layers. It is not easy to visually tell the difference between them.

(3) Robustness: Five operations are exerted on the watermarked feature layers of the data set. The operations and the corresponding similarity degrees between each pair of extracted watermarks are listed Table 1. Our experiments have proved that if the similarity degree is greater than 0.70, the two layers usually contains the same watermark. Table 1 tells 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.

V. CONCLUSIONS

A blind watermarking approach to the copyright protection of vector geospatial data has been proposed in this paper. The approach embeds the watermark 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.

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 discs 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.

ACKNOWLEDGMENT

The work described in this paper is partially supported by OGS (Ontario Graduate Scholarship), partially supported by the Program for Changjiang Scholars and Innovative Research Team in Universities (IRT0966), and partially supported by an 863 project (2009AA121404).

REFERENCES

- [1] A.M. Ahmed, "Digital image watermarking using fuzzy logic and naturalness preserving transform". Ph.D. dissertation, Kansas State University, Manhattan, Kan, USA, 2004.
- [2] I. J. Cox, and M. L. Miller, "The first 50 years of electronic watermarking", *Journal of Applied Signal Processing*, Vol. 56, No. 2, pp. 126-123, 2002.
- [3] X. Zhou, Y. Ren, and X. Z. Pan, "Watermark embedded in polygonal line for copyright protection of contour map", *International Journal of Computer Science and Network Security*, Vol. 6, No. 7B, pp.202-205, 2006.
- [4] M. Samoa, Y. Matsuura, and Y. Takashima, "A scheme of digital watermarking for geographical map data". In *Proc. Cryptography and Information Security*, Okinawa, Japan, 2000, pp. 26-28.
- [5] H. Kang, "A vector watermarking using the generalized square mask". In *Proc. International Conf. Information Technology : Coding and Computing*, Las Vegas, USA, 2001, pp. 234-236.
- [6] R. Ohbuchi, H. Ueda, and S. Endoh, "Robust watermarking of vector digital maps". In *Proc. IEEE International Conf. on Multimedia and Expo 2002*, Lausanne, Switzerland, 2002, vol.1, pp. 577-580.
- [7] Y. Y. Li, and L. P. Xu, "Vector graphical objects watermarking scheme in wavelet domain", *Acta Photonica Sinica*, Vol. 33, No. 1, pp. 97-100, 2004.

- [8] G. Schulz, and M. Voigt, "A high capacity watermarking system for digital maps". In *Proc. Multimedia and Security Workshop on Multimedia and Security*, Magdeburg, Germany, 2004, pp. 180-186.
- [9] K. T. Park, K. I. Kim, and H. Kang, et al., "Digital geographical map watermarking using polyline interpolation". In *Proc. IEEE Pacific Rim Conf. on Multimedia, Hsinchou*, Taiwan, 2002, pp. 58-65.
- [10] H. Sonnet, T. Isenberg, and J. Dittmann, et al., "Illustration watermarks for vector graphics". In *Proc. 11th Pacific Conf. Computer Graphics and Applications*, Calgary, Canada. 2003, pp. 73-82.
- [11] P. H. Jia, Y. Z. Chen, and J. S. Ma, et al, "Digital watermark-based security technology for geo-spatial graphics data", *Chinese Geographical Science*, Vol. 16, No. 3, pp. 276-281, 2006.
- [12] D. H. Douglas, and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitised line or its caricature", *Canadian Cartographer*, Vol. 10, No. 2, pp. 112-122, 1973.
- [13] R. B. McMaster, "Automated line generalization", *Cartographica*, Vol. 24, No. 2, pp.74-111, 1987.
- [14] B. G. Nickerson, "Automated cartographic generalization for linear features", *Cartographica*, Vol. 25, No. 3, pp.15-26, 1988.
- [15] K. Beard, "Theory of the cartographic line revisited", *Cartographica*, Vol. 28, No. 4, pp. 32-58, 1991.



Haowen Yan, holds Bachelor's, Master's and Doctor's degrees in cartography and geographic information engineering from Wuhan University, China. His research interests lie in theories of spatial relations, automated map generalization and spatial data security. Yan's is a professor of geographic information science at Lanzhou Jiaotong University and a visiting scholar in the University of Waterloo.

He had taken a research assistant position in Hongkong Polytechnical University for one year and a senior researcher position in the University of Zurich for one year.



Jonathan Li is a professor at the Department of Geography and Environmental Management, the University of Waterloo. He completed his Ph.D degree in geomatics engineering at the University of Cape Town before beginning teaching in 2000 as an Assistant Professor at the Department of Geography, University of Regina followed by being an Assistant Professor and tenured Associate Professor in the Department of Civil Engineering at Ryerson University.

Li's research interests lie mainly in the areas of remote sensing and geographic information science, including high-resolution satellite mapping, airborne and terrestrial mobile lidar mapping, earth observation of global change, remote sensing of inland and coastal waters, remote sensing of renewable energy potential, mapping of climate-induced hazards, Internet GIS and Web Mapping, Terrain Analysis in Hydrogeography, geospatial sensor network, and geospatial information technologies for emergency response and disaster management.