

A key points-based blind watermarking approach for vector geo-spatial data

Haowen Yan^{a,b,*}, Jonathan Li^b, Hong Wen^c

^aSchool of Mathematics, Physics and Software Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

^bDepartment of Geography & Environmental Management, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

^cNational Key Laboratory of Science and Technology on Communications, UESTC, Chengdu 611731, China

ARTICLE INFO

Article history:

Received 30 April 2010

Received in revised form 25 October 2010

Accepted 26 October 2010

Available online 30 November 2010

Keywords:

Blind watermarking

Vector data

Key points

Similarity degree

ABSTRACT

This paper presents a blind watermarking approach to protecting vector geo-spatial data from illegal use. By taking into account usability, invisibility, robustness, and blindness, the approach firstly determines three feature layers of the geo-spatial data and selects the key points from each layer as watermark embedding positions. Then it shuffles the watermark and embeds it in the least significant bits (LSBs) of the coordinates of the key points. A similar process for selecting the feature layers and the key points in the watermark embedding process is carried out to detect the watermark followed by obtaining the embedded watermark from the LSBs of the coordinates of the key points. Finally, the similarity degrees of three versions of the watermark from three feature layers are calculated to check if the data contains the watermark. Our experiments show that the method is rarely affected by data format change, random noise, similarity transformation of the data, and data editing.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Generally speaking, vector geo-spatial data is of great value because the acquisition of such data is a high cost process (López, 2002), in which high precision instruments and a large amount of physical labour resources are required, and the digitization and vectorization of original data also need hard work. Consequently, vector geo-spatial data normally can not be directly 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 brings a lot of trouble for the data owners to protect the data from piracy.

Digital watermarking, coming from steganography, provides a viable solution for data security. A digital watermark is defined as an imperceptible but identifiable digital signal or mode permanently embedded in other data, namely 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 & Miller, 2002; Zhou, Ren, & Pan, 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 extraction process should be blind if possible, i.e. the watermark can be detected by the data owner without the original data and the original watermark at hand.

Have not only the techniques of digital watermarking received a great deal of attention to ensure copyright protection for multimedia message, such as video data (Hartung & Girod, 1998; Langehaar & Lagendijk, 2001), audio data (Kirovski & Malvar, 2003; Seok, Hong, & Kim, 2002; Wang, Niu, & Qi, 2008) and image data (Aslantas, 2008; Cox, Kilian, Leighton, & Shamoan, 1997; Langelaar & Lagendijk, 2001), and been a focus in network information security (Cox & Miller, 2002), but also it has become a hot issue in the community of Geo-Science for protecting vector geo-spatial data from piracy (Lafaye, Béguet, Gross-Amblard, & Ruas, 2007; López, 2002; Niu, Shao, & Wang, 2006). Generally, there are two categories of watermarking algorithms for two-dimensional (2D) geo-spatial data, i.e. space domain and frequency domain.

Many frequency domain algorithms (e.g., Solachidis & Pitas, 2004; Zhu, Yang, & Wang, 2008) embed the watermarks in the Fourier descriptors of the curves or polygonal lines, causing invisible distortions of the vertices coordinates. Ohbuchi, Ueda, and Endoh (2003) presented a method embedding watermarks in the frequency-domain representation of a 2D vector digital map. The method treats vertices on the map as a point set, and imposes connectivity among the points using Delaunay triangulation, then computes the mesh-spectral coefficients from the mesh created. Modifications of the coefficients according to the message bits, and inverse transforming the coefficients back into the coordinate domain produces the watermarked map. Voigt, Yang, and Busch

* Corresponding author at: School of Mathematics, Physics and Software Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China.

E-mail address: haowen2010@gmail.com (H. Yan).

(2004) addressed a reversible watermarking scheme that exploits the high correlation among points in the same polygon in a map and achieves the reversibility of the whole scheme by an 8-point integer discrete cosine transform, which ensures that the original 2D vector data can be watermarked during the watermark embedding process and perfectly restored during the watermark extraction process, with the watermark accurately extracted at the same time. The watermarks generated by these frequency domain algorithms can generally withstand rotation, translation, scaling, and reflection, and their detection process is blind. However, the spatial precision change (e.g. spatial topological relations among objects) in the watermarked data is difficult to control because the watermark is not directly embed in spatial data. In addition, such algorithms are generally fragile to the attacks from the noise and the revision of the data (e.g. insertion and removal of points).

Most of the existing space domain algorithms (e.g. Kang, 2001; Li & Xu, 2004; Ohbuchi, Ueda, & Endoh, 2002) 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 some 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, Kim, Kang, & Han, 2002; Sonnet, Isenberg, Dittmann, & Strothotte, 2003) that insert new points into the original data and take them as the watermark embedding positions. Moreover, an algorithm proposed by Jia, Chen, Ma, and Zhu (2006) inserts the bits of the watermark in 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 frequency domain ones in preserving the precision of watermarked data. This should be emphasized because the precision is of most importance to geo-spatial data. (2) The existing space domain algorithms do not differentiate among point, linear and areal features, and take into little consideration of the spatial characteristics of geo-spatial data. This is important for our new watermarking approach and will be further discussed in Section 2.1. (3) None of the existing space domain algorithm is blind in the watermark detection process.

For these reasons, our research will make improvements at the above three points and aim at proposing a blind watermarking approach in space domain that allows copyright protection of vector geo-spatial data.

The remainder of this paper is organized as follows. Section 2 details an algorithm that allows watermark embedding in vector geo-spatial data. Section 3 describes an algorithm for the watermarking detection. These two sections comprise the main parts of the work. Section 4 describes two examples to test the validity of the proposed watermarking approach. Finally, conclusions are drawn in Section 5.

2. Watermark embedding algorithm

2.1. General approach

To overcome the shortcomings of the existing watermarking algorithms in the space domain, the following strategies have been employed in our new approach:

- (1) *Inserting the watermark multiple times in the feature layers.* It is a common sense in the community of cartography and geographic information science that vector geo-spatial data are divided into multiple feature layers for the purpose of storage and management. If the watermark can be multiply embedded in different feature layers, it must become more difficult to be removed. Therefore, its robustness can be improved. More importantly, multiple embedding means multi-versions of the watermarks can be detected from different feature layers. This provides a potential for judging if the data contains the watermark by comparing the extracted watermarks without the original data and original watermark, i.e. blind detection of the watermark.
- (2) *Utilizing the key points as the watermark embedding positions.* Key points are more important in their geometric and/or attribute aspects than the other ones; so they own less probability to be removed and/or edited by the users. Without doubt to embed the watermark in the key points are favorable to the robustness of the watermarking technique.
- (3) *Using the LSBs of the key point coordinates to embed the watermark.* The number of LSBs used for watermarking in each coordinate can be adjusted according to the data precision requirements from practical applications; hence, the data precision change caused by watermark embedding is controllable. In this sense, the quality of the watermarked data can be maintained.

Based on the above strategies, an algorithm for embedding watermarks in vector geo-spatial data is proposed. Fig. 1 demonstrates its principles and procedures: determination of the feature layers for watermark embedding, selection of the key points, preparation of the watermark, and watermark embedding.

2.2. Determination of embedding feature layers

There are generally multiple feature layers in a vector geo-spatial database. For example, China's topographic map database at scale 1:1000,000 comprises of 14 layers, including control points, hydrography, settlements, roads, topography, boundaries, vegetations, etc. Deciding how many and which feature layers should be used for watermark embedding is the key in this procedure. Our experiences and experiments identify the following four rules for the selection:

Rule 1: The ideal number of the feature layers for watermark embedding should be three for the purpose of the robustness

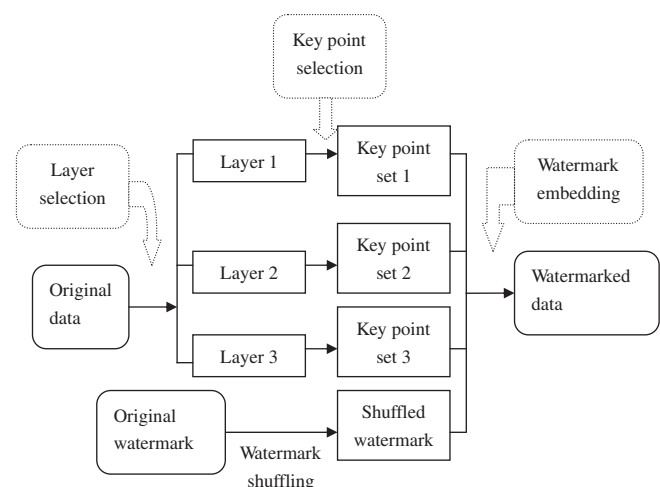


Fig. 1. Procedures of watermark embedding.

and blindness of the watermarking technique. The reason for this is that (suppose that the watermark will be embedded in one feature layer for only one time): one feature layer cannot make the watermark detection blind; and two feature layers make the detection blind, however the detection process is fragile; and four or more feature layers can make the approach blind and robust, nevertheless the approach obviously becomes much more complicated to implement in the meantime. In this sense, the selection of three feature layers makes a balance between the blindness and robustness of the approach.

Rule 2: The number of the points in each selected feature layer should be greater than N (N is the bit number of the watermark, and it equals to the number of points used for embedding the watermark in this paper. A detailed discussion of N will be presented in Section 2.6.1), to ensure the capacity of the space domain is enough for embedding the watermark. Here, the “points” means the individual points of point feature layers and the vertices of the curves in linear and polygonal feature layers.

Rule 3: The more important a layer is, the more probable the layer should be selected. This is in favor of the robustness of the watermarking technique. The importance degrees of the layers generally are determined by the data owner.

Rule 4: Control point layer (ground-based points whose positions and elevations have been precisely and accurately determined) is not allowed to be selected because permission usually is not given to change the coordinate values in this layer.

After this procedure, three feature layers are selected, and the information of each layer are recorded for the watermark detection. Because each feature layer may be point, linear, or areal, so it is pertinent to present the method for selecting key points from each type of layers, respectively.

2.3. Key point selection method for point feature layers

Key points of a point feature layer are those that are more important than the other ones in preserving the distribution and structure characteristics of the point clusters. Therefore, each of them owns less probability to be edited than the other ones. Here, a method based on Voronoi diagrams is presented for detecting the key points in a specific point set, which consists of the following steps:

Step 1: construction of Voronoi diagrams. There have been a variety of algorithms proposed for generating Voronoi diagrams in the computational geometry community (Yan & Weibel, 2008). Here, a commonly accepted one presented by Ahuja (1982) is employed for constructing Voronoi diagrams of the point set.

Step 2: ordering and selection of the boundary points. A point with a divergent Voronoi polygon (if a Voronoi polygon has an open boundary, it is a divergent Voronoi polygon) is called a boundary point because it is at the distributional boundary of the point set. Boundary points are generally regarded as more important than the inner ones on the map and each of them owns less probability to be edited, so it is appropriate to select them as the watermarking positions.

To make blind detection of the watermark possible, the boundary points should be put in a specific order. They are ordered as follows, supposing that there are K boundary points and N points are needed for embedding the watermark:

- Find the boundary point (say P_1) that has the least coordinate x and the least coordinate y and the boundary point (say P_2) that has the greatest coordinate x and the least coordinate y .

- Calculate the distance between each boundary point and P_1 , and each angle that rotates ray P_1P_2 in counter clockwise to the ray linking P_1 and the boundary point.
- Sort the boundary points in the increasing order of the calculated angle values. If two points have same angle values, the one with less distance value is arranged before the other one (an example is shown in Fig. 2). If $K > N$, select all the boundary points as the key points; or else, select N boundary points starting from the beginning of the ordered boundary point array and end this step.

Step 3: selection of the internal points. In this step, $N - K$ internal points should be selected. The selection probabilities of each point in the inner point set (i.e. the difference of the original point set and the boundary point set) can be calculated by:

$$p_i = \frac{I_i A_i}{\sum_{k=1}^N I_k A_k} \quad (1)$$

where P_i is the selection probability of the i th point, I_i is the importance value of the i th point (if the points are thematic features, e.g. schools, their importance values are determined by their thematic attributes, e.g. the number of students; otherwise, their default values are all 1), A_i is the area of the Voronoi polygon of the i th point, and N is the total number of the points in the point feature layer.

A point may be “free”, “fixed”, or “deleted”. Firstly, let the initial status of every point be “free”. Secondly, sort the selection probabilities of all inner points in increasing order. Thirdly, choose a “free” point that has the greatest selection probability and none of its Voronoi neighbor is currently “deleted”, and let it be “selected” and all of its Voronoi-neighboring points be “fixed”, and record the sequential number of this point in an array. And then, iteratively select “free” points and mark them as “selected” until the total selected number of points is $N - K$. Fig. 3 shows an example of this process.

2.4. Key point selection method for linear feature layers

A linear feature layer on the map generally consists of open lines (e.g. hydrography), closed lines (e.g. contours and boundaries), or a mixture of both (e.g. roads). The key points of a line feature are those that may represent the caricature of the line feature (Douglas & Peucker, 1973); hence they have less probability to be modified than the other ones. To detect key points from such layers, both geometric and geographic characteristics of the feature should be taken into consideration (Beard, 1991), and different methods are needed for different feature layers.

To facilitate the discussion, roads are taken as the representative, and a method for detecting key points from road features is proposed here, comprising the following three steps, The idea of the method is based on the fact that the end-points and the

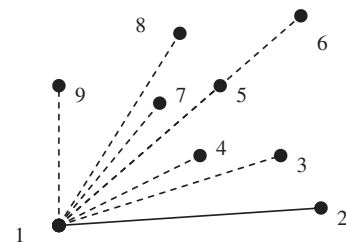


Fig. 2. Principle of boundary point ordering: after calculation of the distance and angle values of each point, points 1 and 2 is arranged first; then the others are sorted in turn. Point 5 is arranged before point 6, for its distance value is less though its angle value is equal to that of the later.

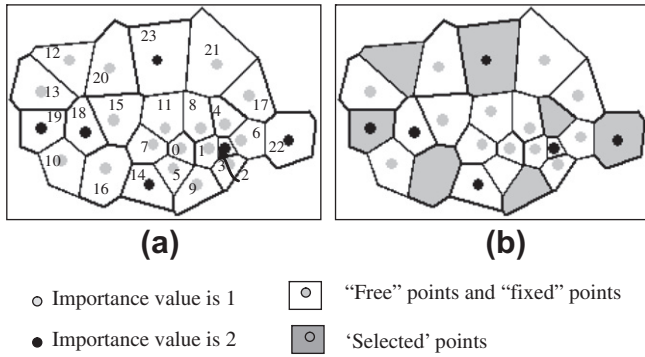


Fig. 3. Principle of inner point selection: (a) sort the selection probabilities of the points in increasing order (the increasing order of selection probability is numbered); and (b) select points in turn by the selection probabilities.

intersections of roads are generally more important than the other points of the roads, and they have less probability to be modified than the other ones, so they may be selected as the key points of road features:

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 end-points. Suppose that there are totally N_1 roads. The length values of the roads are sorted in decreasing order. If $N_1 \geq N/2$ (N is the number of the points used for embedding the watermark), select N end-points (i.e. start and ending points) of the roads that own greater length values, and end the procedure; or else, select all of the end-points and go to Step 3.

Step 3: selection of the intersections. Firstly, select $N - N_1$ roads. Each selected road should satisfy 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 end-points of the road, and (2) selecting the intersection with the greatest distance as the watermark insertion position (see Fig. 4).

After this step, totally N key points are selected.

2.5. Key point selection method for areal feature layers

The areal feature layers that consist of polygon coverages and that consist of disjointed polygonal objects should be differentiated

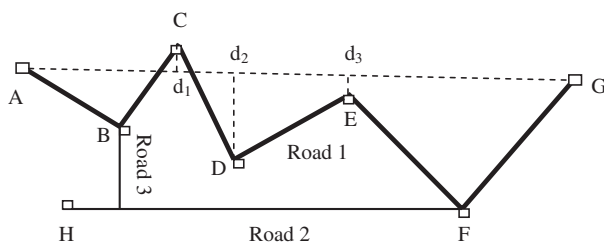


Fig. 4. Demonstration of road intersection selection: There are five intersections B–F on Road 1. Since F and B are also the end-points of Road 2 and Road 3, respectively, and have been selected as watermarking positions; therefore, D whose distance d_2 to AG (the line linking the end-points of Road 1) is greater than that of C and E is selected here.

in the key point selection methods due to their different geometric characteristics. Here, the key points of polygon coverages are the joint points of polygons, while the key points of disjointed polygonal objects are those that are similar to the key points of linear features (Douglas & Peucker, 1973). The common characteristic of the two types of key points is both of them have more probability to be retained on the map after map generalization or/and map editing.

2.5.1. Method for polygon coverages

The method for key point selection from polygon coverages comprises 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 two 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 numbers of the sorted joint points in a one-dimensional (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 the key points 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 edges of the polygons in increasing order, and select $N - N_j$ edges with greater length values and extract one point from each edge as the key point, using a distance-based method (Fig. 5) similar to the one proposed by Douglas and Peucker (1973). An example of this method is demonstrated in Fig. 6.

2.5.2. Method for disjointed areal features

A disjointed areal feature layer consists of topologically separated polygons, such as settlements on large scale maps. Suppose that the number of the polygons 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 sum number of the key points equals to N .

A method based on deviation angles and the polygon's edge length values is used for selecting key points from each polygon. Suppose that the vertices of the polygon are saved in clockwise. The deviation angle of a vertex can be defined as an angle rotating in clockwise from the extension line of the edge linking the previous vertex and the vertex itself to the next neighboring edge (see Fig. 7).

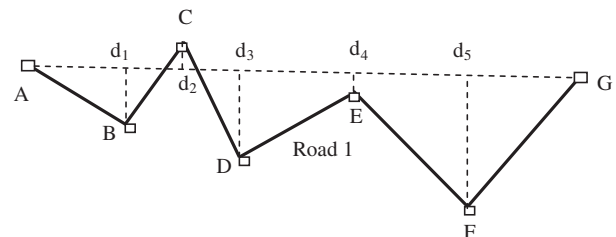


Fig. 5. Principle of the distance-based method: point F is selected, for it has the greatest distance to the line segment linking the two ending points of the edge.

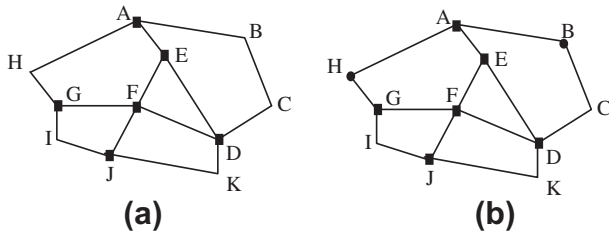


Fig. 6. Demonstration of the key point selection from polygonal coverages (suppose that eight key points are needed): (a) six joint points A, D, E, F, G and J are selected first; and (b) then select two longer edges of the polygons, and the points B and H belonging to the two edges are selected.

Firstly, calculate and 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 all edges. Finally, select the required vertices according to the sequence number of the vertices recorded in V (Fig. 7).

2.6. Watermark embedding

2.6.1. Preparation of the watermark

Most of the watermarks used in the past applications and past research work are images (Cox & Miller, 2002; Zhu et al., 2008). However, that using strings as watermarks appeared in literatures in recent years (Jia et al., 2006; Wang, Wang, Wang, & Qin, 2009). Both images and strings have their advantages, disadvantages, and dominant application fields (Wang et al., 2009). Here, a string is used as the watermark due to its less number of bits compared with that of an image. To increase the difficulty of detecting the embedded watermark from the host data, the following operations are carried out:

- Form a 1D bits chain, say C , using the ASCII codes of the characters in the string;
- Shuffle the bits of C by

$$C_i = \begin{cases} C_i, & \text{if } i \text{ is odd} \\ C_{N-i-1}, & \text{if } i \text{ is even \& } i \leq \frac{N}{2}. \end{cases} \quad (2)$$

where C_i is the i th bit value of C , N is the number of the bits of C . Here, shuffle means exchanging positions of the bits.

- Exert bit operation NOT on the shuffled bit chain.

Bit operation NOT is only for enhancing the robustness of the watermark; so other bit operations, such as XOR, can also be used here (Zhu et al., 2008).

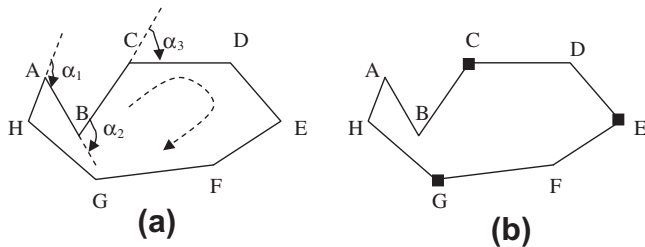


Fig. 7. Demonstration of the key point selection from a disjointed polygon (suppose that three key points are needed): (a) sort the deviation angles in increasing order (α_1, α_2 and α_3 are the deviation angles of corresponding vertices); and (b) delete A, B, D, F, and H from sorted vertex array, for each of them has at least one joint edge shorter than the mean length of all edges, and select C, E and G as the key points.

2.6.2. Method for watermark embedding

With the watermark prepared and the key points selected, the watermark embedding becomes a considerably easy process. To ensure the precision of the data, an embedding method based on LSBs proposed by Jia et al. (2006) is employed, and the bit values in C are in turn written in the LSBs of the coordinate x (or y) of the key points.

3. Watermark detection algorithm

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

Step 1: determination of the embedding positions. Firstly, the information about the three layers used for embedding the watermark 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 three feature layers are recorded in three 1D arrays $S1, S2$ and $S3$.

Step 2: extraction of the bit chains. By reading each LSB of the coordinate x of each key point in light of the three sequence number $S1, S2$ and $S3$, three 1D bit chains $C1, C2$ and $C3$ are formed. The bit numbers of $C1, C2$ and $C3$ should be all equal to the total number of bits of the original watermark.

Step 3: reconstruction of the watermark. This is an inverse operation of the watermark shuffling in the preparation of the watermark. The bits in $C1, C2$ and $C3$ are re-arranged using the reverse operations used in watermark shuffling, and three versions of the original watermark C^*1, C^*2 and C^*3 (i.e. three strings) are remolded.

Step 4: comparison and decision making. The similarity degree of any two bit chains with same number of bits can be calculated by:

$$D = \frac{N_s}{N} \quad (3)$$

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 each bit chain.

Let the similarity degrees of C^*1 and C^*2, C^*1 and C^*3 , and C^*2 and C^*3 be D_{12}, D_{13} , and D_{23} , respectively. If one of the three similarity degrees is greater than a given threshold value, it can be concluded the data contains the watermark.

4. Experiments

The proposed approach was implemented in Visual C++ (Version 8.0). To verify its correctness and soundness, a set of experiments were carried out using various datasets. Two of the experiments are presented here. To demonstrate the adaptability of the approach, two different types of data have been chosen, representative of different scales and formats of geo-spatial data.

The first dataset is China’s fundamental map at scale 1:1000,000 (Fig. 8), comprising 14 feature layers, publicly provided by the National Geomatics Center of China (NGCC). The dataset is in the Shapefile format (Vector data in DXF format was also used in our experiments, but not shown here). The settlements (points), hydrography (linear), and roads (linear) are selected for watermark embedding. The watermark is the string “National Geographic Data”.

The second dataset is a topographic map at scale 1:50,000 (Fig. 9), comprising 18 feature layers, provided by the Surveying and Mapping Bureau of Gansu Province, China. The dataset is in the DXF file format. The hydrography (linear), roads (linear), and

settlements (areal) is selected for watermark embedding. The watermark is the string “Gansu Surveying and Mapping”.

The approach is evaluated from the four aspects, such as usability, invisibility, robustness and blindness.

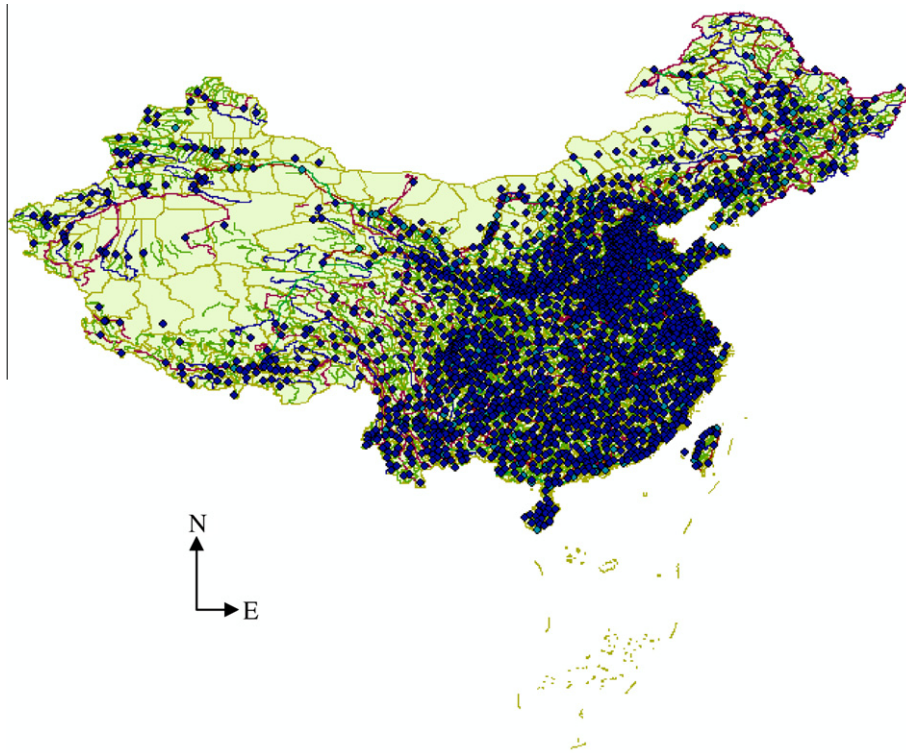


Fig. 8. Data used in the Experiment 1: the map scale is 1:1000,000, but it is not shown to scale here (free data from <http://nfgis.nsd.gov.cn/nfgis/chinese/db/dlg025.htm>).

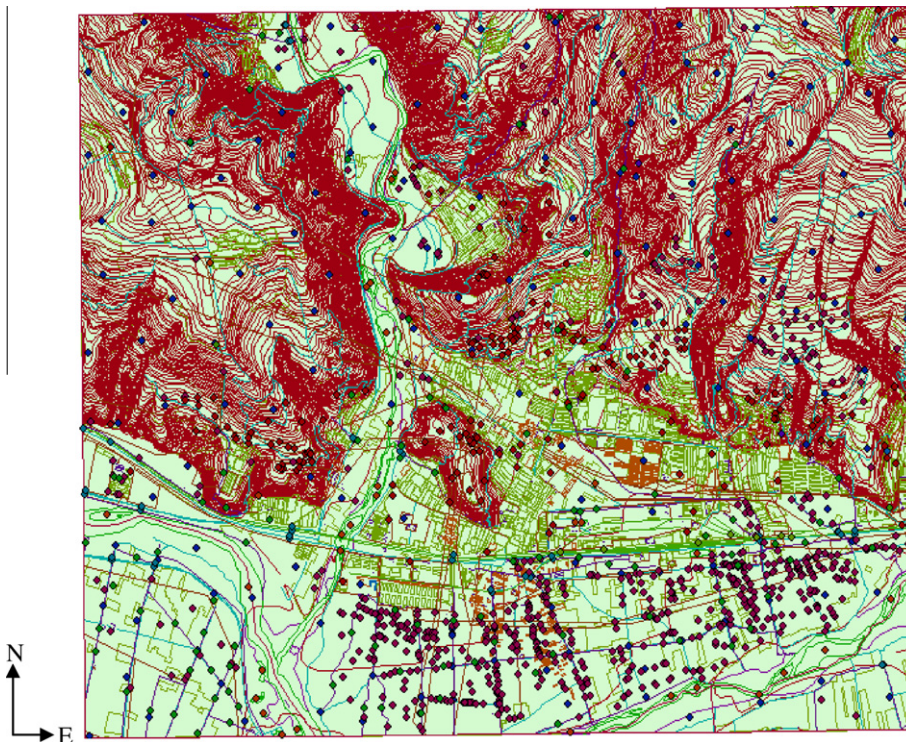


Fig. 9. Data used in the Experiment 2: the map scale is 1:50,000, but it is not shown to scale here (academic use only).

Table 1

Test the robustness using different operations.

	Format change	Similarity transformation	Random noise attack	10% points deletion	10% points insertion
Experiment 1	$D_{s,h} = 1$	$D_{s,h} = 1$	$D_{s,h} = 0.930$	$D_{s,h} = 0.904$	$D_{s,h} = 0.923$
	$D_{s,r} = 1$	$D_{s,r} = 1$	$D_{s,r} = 0.931$	$D_{s,r} = 0.898$	$D_{s,r} = 0.874$
	$D_{h,r} = 1$	$D_{h,r} = 1$	$D_{h,r} = 0.965$	$D_{h,r} = 0.870$	$D_{h,r} = 0.908$
Experiment 2	$D_{s,h} = 1$	$D_{s,h} = 1$	$D_{s,h} = 0.922$	$D_{s,h} = 0.851$	$D_{s,h} = 0.880$
	$D_{s,r} = 1$	$D_{s,r} = 1$	$D_{s,r} = 0.918$	$D_{s,r} = 0.911$	$D_{s,r} = 0.906$
	$D_{h,r} = 1$	$D_{h,r} = 1$	$D_{h,r} = 0.971$	$D_{h,r} = 0.908$	$D_{h,r} = 0.872$

$D_{s,h}$, $D_{s,r}$, and $D_{h,r}$ are the similarity degrees between settlements and hydrography, settlements and roads, and hydrography and roads, respectively.

Similarity transformation includes translation, scaling, and reflection.

Random noise means the deletion and insertion of small number of points, lines, and/or polygons in the data set by mistake.

Usability: The usability of the watermarked data can be evaluated at a scientific level by means of analyzing the relative errors of the data. According to the calculation and statistics of the positional changes of all coordinates x used for watermark embedding, none of the relative error in the two experiments 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 watermarks can still be used.

Invisibility: Each of the maps shown in Figs. 8 and 9 are the overlapped visualization of the original dataset and the corresponding watermarked data set. It is difficult to visually tell the difference between the two data sets. In other words, the inserted watermark is perceptually invisible to the data users.

Robustness: Five operations are exerted on the watermarked feature layers of the two data sets, respectively. Then the watermarks are extracted to test the robustness of the approach. The operations and the corresponding similarity degrees between each pair of extracted watermarks are listed in Table 1.

Our experiments have proved that if the similarity degree is greater than 0.70, the two layers usually contains the same watermark. In light of the similarity degrees in Table 1, it can be concluded the data format changes and similarity transformations have no effects on the watermarked data sets, and the watermarking technique is also robust to resist the attacks from random noise, data deletion, and data insertion. However, this approach cannot resist the data change from map generalization.

Blindness: Neither the original data sets nor the original watermarks is needed in the watermark detection processes of the two experiments, so this is a wholly blind watermarking approach.

5. Concluding remarks

In this paper, we have presented a blind watermarking approach to the copyright protection of vector geo-spatial data. The approach multiply embeds the watermark in three different feature layers of the host data, while it differentiates and takes into account the characteristics of different spatial features. 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.

This approach has been implemented successfully by the authors and the software has been used by the Lanzhou Bureau of Land Resources, Gansu Province, China, for duplicating data using hard discs and CDs, and distributing data via the Internet. However, our approach has several limitations on which future research should be directed. First, its performance is dependent on the data layer. Second, it cannot resist data modification through map generalization. Finally, it performs well on topographic maps only. Future extensions should be applicable to other types of vector geospatial datasets to offer the same protection to other common data types.

Acknowledgements

The work described in this paper was carried out during a research visit by Haowen Yan to the University of Waterloo. The work was partially supported by a grant provided by the National Natural Science Foundation of China (Project No. 40871208), and partially supported by a grant of the “New Century Excellent Talents” Program of the Ministry of Education of China (Project No. NCET-07-0404), and partially supported by Program for Changjiang Scholars and Innovative Research Team in University (IRT0966). The authors are grateful to the anonymous reviewers for their valuable comments and appreciate the editors’ useful advice in improving the wording quality of the paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.compenvurbsys.2010.10.004.

References

- Ahmed, A.M. (2004). Digital image watermarking using fuzzy logic and naturalness preserving transform. Ph.D. Thesis, Kansas State University.
- Ahuja, N. (1982). Dot pattern processing using Voronoi neighborhoods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(3), 336–343.
- Aslantas, V. (2008). A singular-value decomposition-based image watermarking using genetic algorithm. *International Journal of Electronics and Communications*, 62(5), 386–394.
- Beard, K. (1991). Theory of the cartographic line revisited. *Cartographica*, 28(4), 32–58.
- Cox, I. J., Kilian, J., Leighton, T., & Shamon, T. G. (1997). Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12), 1673–1687.
- Cox, I. J., & Miller, M. L. (2002). The first 50 years of electronic watermarking. *Journal of Applied Signal Processing*, 56(2), 126–132.
- Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer*, 10(2), 112–122.
- Hartung, F., & Girod, B. (1998). Watermarking of uncompressed and compressed video. *Signal Processing*, 66(3), 283–301.
- Jia, P. H., Chen, Y. Z., Ma, J. S., & Zhu, D. K. (2006). Digital watermark-based security technology for geo-spatial graphics data. *Chinese Geographical Science*, 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).
- Kirovski, D., & Malvar, H. S. (2003). Spread spectrum watermarking of audio signals. *IEEE Transactions on Signal Processing*, 51(4), 1020–1033.
- Lafaye, J., Béguec, J., Gross-Amblard, D., & Ruas, A. (2007). Invisible graffiti on your buildings: Blind and squaring-proof watermarking of geographical databases. In *Proceedings of the 10th international symposium on spatial and temporal databases*, Boston, USA, July 16–18, Lecture notes in computer science (Vol. 4605, pp. 312–329).
- Langelaar, G. C., & Lagendijk, R. L. (2001). Optimal differential energy watermarking of DCT encoded images and video. *IEEE Transactions on Image Processing*, 10(1), 148–158.
- Li, Y. Y., & Xu, L. P. (2004). Vector graphical objects watermarking scheme in wavelet domain. *Acta Photonica Sinica*, 33(1), 97–100.
- López, C. (2002). Watermarking of digital geospatial datasets: A review of technical, legal and copyright issues. *International Journal of Geographical Information Science*, 16(6), 589–607.

- Niu, X. M., Shao, C. Y., & Wang, X. T. (2006). A survey of digital vector map watermarking. *International Journal of Innovative Computing*, 2(6), 1301–1306.
- 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).
- Ohbuchi, R., Ueda, H., & Endoh, S. (2003). Watermarking 2D vector maps in the mesh-spectral domain. In *Proceedings of international conference on shape modeling and application (SMI2003)*, May 2003, Seoul, Korea (pp. 216–228).
- Park, K.T., Kim, K.I., Kang, H., & Han, S.S.(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).
- Seok, J., Hong, J., & Kim, J. (2002). A novel audio watermarking algorithm for copyright protection of digital audio. *ETRI Journal*, 24(3), 181–189.
- Solachidis, V., & Pitas, I. (2004). Watermarking polygonal lines using Fourier descriptors. *IEEE Computer Graphics and Applications*, 24(3), 44–51.
- Sonnet, H., Isenberg, T., Dittmann, J., & Strothotte, T. (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).
- Voigt, M., Yang, B., & Busch, C. (2004). Reversible watermarking of 2D-vector data. In *Proceedings of the ACM international workshop on multimedia and security*, September 2004, Magdeburg, Germany (pp. 160–165).
- Wang, X. Y., Niu, P. P., & Qi, W. (2008). A new adaptive digital audio watermarking based on support vector machine. *Journal of Network and Computer Applications*, 31(4), 735–749.
- Wang, C., Wang, W., Wang, Q., & Qin, Q. (2009). A watermarking algorithm for vector maps in spacial domain. *Geomatics and Information Science of Wuhan University*, 34(2), 163–169 (in Chinese).
- Yan, H. W., & Weibel, R. (2008). An algorithm for point cluster generalization based on the Voronoi diagram. *Computers and Geosciences*, 34(8), 939–954.
- Zhou, X., Ren, Y., & Pan, X. Z. (2006). Watermark embedded in polygonal line for copyright protection of contour map. *International Journal of Computer Science and Network Security*, 6(7B), 202–205.
- Zhu, C.Q., Yang, C.S., & Wang, Q.S. (2008). A watermarking algorithm for vector geospatial data based on integer wavelet transform. In *International archives of the photogrammetry, remote sensing and spatial information sciences, July 2008* (Vol. 37(B4), pp. 15–18). Beijing, China.