An Adaptive Watermark Detection Algorithm for Vector Geographic Data

Yingying Wang[1], Chengsong Yang[2*], Na Ren[3,4,5], Changqing Zhu[3,4,5], Ting Rui[2], and Dong Wang[2]

[1]College of Intelligent Science and Control Engineering, Jinling Institute of Technology, Nanjing 211169, China
[e-mail: wyychs@163.com]
[2] Institute of Field Engineering, Army Engineering University of PLA
Nanjing 210007, China
[e-mail: ycsdongshang@163.com]
[3] Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University
Nanjing 210023, China
[4] State Key Laboratory Cultivation Base of Geographical Environment Evolution
Nanjing 210023, Jiangsu Province
[5] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application
Nanjing 210023, China
*Corresponding author: Chengsong Yang

Received March 27, 2018; revised December 6, 2018; accepted August 20, 2019;
published January 31, 2020

Abstract

With the rapid development of computer and communication techniques, copyright protection of vector geographic data has attracted considerable research attention because of the high cost of such data. A novel adaptive watermark detection algorithm is proposed for vector geographic data that can be used to qualitatively analyze the robustness of watermarks against data addition attacks. First, a watermark was embedded into the vertex coordinates based on coordinate mapping and quantization. Second, the adaptive watermark detection model, which is capable of calculating the detection threshold, false positive error (FPE) and false negative error (FNE), was established, and the characteristics of the adaptive watermark detection algorithm were analyzed. Finally, experiments were conducted on several real-world vector maps to show the usability and robustness of the proposed algorithm.

Keywords: Vector geographic data, watermark detection, adaptive, usability, robustness, threshold

The work was supported by the National Natural Science Foundation of China (grant No. 41401518), the Natural Science Foundation of Jiangsu Province (grant No. BK20140066) and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

http://doi.org/10.3837/tiis.2020.01.018 **ISSN : 1976-7277**

1. Introduction

Vector geographic data are fundamental achievements of national information infrastructure and earth science research and have been widely used in cartography, navigation, spatial analysis and many other areas. With the rapid development of computer and communication technology, vector geographic data can be conveniently replicated and distributed. The acquisition of vector geographic data is a costly process; therefore, the copyright protection of vector geographic data has become a hot issue in the geospatial data security domain.

As an effective algorithm for copyright protection, watermarking for vector geographic data has been studied for more than ten years [1]. A number of watermarking algorithms for vector geographic data have been proposed. Generally, digital watermark techniques for vector geographic data contain three parts: watermark generation, embedding, and detection. The first step is selecting a watermark, which could be an image, character, binary sequence, or voice. Preprocessing is performed when necessary, and examples include watermark encryption and watermark compression. The second step is to hide the watermark in the cover data used to embed the watermark. A watermark may be contained in the cover data or not. According to the different embedded domains, digital watermarking is categorized as spatial and transform domain watermarking. Spatial watermarking hides the watermark in the spatial domain component, and transform domain watermarking hides the watermark in the discrete cosine transform (DCT) [2]-[3], discrete Fourier transform (DFT) [4]-[6], or discrete wavelet transform (DWT) domain [7]-[8]. To date, many studies on digital watermark embedding for vector graphic data have been undertaken [9]-[16]. The third step of the digital watermark technique includes extracting and detecting the watermark. Extraction is mostly the inverse operation of the embedding process, and it is used to extract the embedded watermark. The detection step employs various methods to judge the specific content of the watermark. If the original data are not available for detection, the process is called “blind detection”, whereas if the data are available, then the process is called “informed detection.” Compared with research on embedding, the research on detection algorithms is limited [17]-[23]. For most existing watermarking algorithms, robustness has been qualitatively analyzed rather than quantitatively analyzed. However, the detection algorithm is an important part of digital watermark techniques and affects the robustness, reliability, and practicability of the watermark system. A better detection algorithm is useful for improving the robustness and capability of the watermark system.

Data deletions and data additions are the most common types of data processing performed in geographic information systems (GIS), and quantitatively analyzing the robustness of the watermark against data deletion and data addition attacks is a critical process. Therefore, we focus on the detection algorithm, and the design of a watermark detection algorithm based on statistical characteristics is presented. This paper proposes an adaptive watermark detection algorithm that a) can calculate the detection threshold, false positive error (FPE) and false negative error (FNE), b) is robust to data deletions and data additions and c) can qualitatively analyze the robustness of watermarks against data adding attacks.

The remainder of the paper is organized as follows. Section 2 briefly describes the watermark generation and embedding algorithms. Section 3 presents the fixed watermark detection process and establishes the adaptive watermark detection. Section 4 analyzes the characteristics of the proposed watermark detection method. Section 5 describes the experiments and results, and Section 6 presents the conclusions of the paper.

2. Watermark Generation and Embedding

2.1 Watermark Generation

Research manuscripts reporting on large datasets deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, a note should be included stating that the numbers will be provided during review. The numbers must be provided prior to publication.

Watermark information can be classified into two main categories: significant watermarks and insignificant watermarks. A significant watermark is usually represented as a binary logo, and this logo is extracted to detect the presence of the watermark by visual inspection by a third entity, whereas an insignificant watermark is usually represented as a pseudo-random sequence. The presence of the watermark is detected using statistical correlations so that the detection of an insignificant watermark is more objective [11]. Moreover, the length of an insignificant watermark is usually much shorter than that of a significant watermark; thus, an insignificant watermark is more suitable for small datasets, which are common in vector geographic data. To establish the watermark detection model and quantitatively analyze the robustness of the proposed watermarking algorithm, we used a pseudo-random sequence as a watermark for this paper.

Let the watermark information be $W = \{w[i], 0 \le i < N\}$, where $w[i]$ is the watermark bit, i is the watermark bit index, N is the length of the watermark, and $w[i] \in \{-1,1\}$. The statistical characteristics of $w[i]$ are $P(w[i] = -1) = 1/2$ and $P(w[i] = 1) = 1/2$, which means that the probabilities that $w[i]$ are equal to -1 or 1 are the same. We set N to 200 for this paper.

The steps used to generate watermark information are as follows: first, a random integer is generated and used as a watermark seed; second, according to the watermark seed, the watermark is generated by the pseudo-random sequence generator; and third, the watermark seed is saved for the potential watermark detection. A one-to-one relationship is observed between the watermark seed and the corresponding watermark; therefore, we only need to save the watermark seed for the potential watermark detection.

In the subsequent section, the terms watermark seed, watermark, watermark bit, and watermark bit index are frequently referenced. The relationships among the watermark, watermark bit, and watermark bit index are shown in **Fig. 1**, in which N is the length of the watermark.

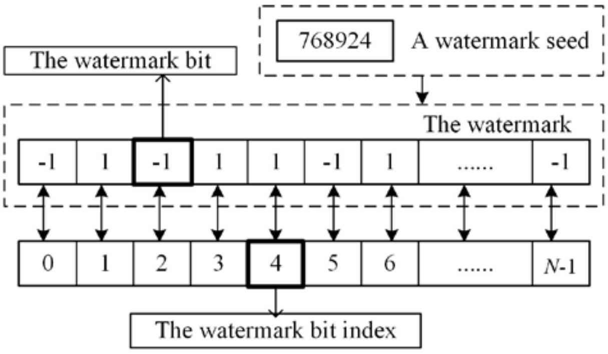


Fig. 1. Relationships among the watermark seed, the watermark, the watermark bits, and the watermark bit index

2.2 Watermark Embedding

The vector geographic data consist of the vertex coordinates. Vertex coordinates are the fundament units of points, polylines, and polygons that describe geographical objects, such as wells, rivers, and residential areas. To obtain a watermarking algorithm that is robust against the most common watermarking attacks such as data cropping, data reordering, data simplification, vertex adding and vertex deleting, we establish the mapping relationships between the vertex coordinates and the watermark bit index [23], and the watermark bits are then embedded into the corresponding vertex coordinates according to the mapping relationships.

In this paper, let the vertex coordinates of the vector geographic data be the set VC , $VC = \{vc_i \,|(x_i, y_i), 0 \le i < M\}$, where vc_i is the ith vertex, (x_i, y_i) is the coordinate of the ith vertex, and M is the number of vertex coordinates. The details of the steps for establishing mapping relationships are as follows.

1) The vertex coordinates VC are mapped to region R , which has a size of $R\_x \times R\_y$ as shown as equation (1):

$$\begin{cases} x_i' = \lfloor x_i * s \rfloor \% R\_x \\ y_i' = \lfloor y_i * s \rfloor \% R\_y \end{cases}, \tag{1}$$

where (x_i', y_i') is the mapped coordinate and s is a scaling factor used to control the data distortions in the watermark embedding.

2) The region R is divided into $N\_c \times N\_r$ grid cells and by establishing the mapping relationships between coordinate (x_i', y_i') and the grid cell index (ic, ir) , which are shown in equation (2), where N_r , N_c , LD_x , and LD_y are all positive integers, R_x can be divided by N_c , and R_y can be divided by N_r . Additionally, $N\_r = R\_\_y / LD\_y$, $N\_c = R\_x / LD\_x$, and $N\_c \times N\_r = N$.

$$\begin{cases} ir = \left\lfloor y_i' / LD\_y \right\rfloor \\ ic = \left\lfloor x_i' / LD\_x \right\rfloor \end{cases} \tag{2}$$

3) The 2-dimensional index of grid cell (ic, ir) is mapped to the 1-dimensional index of the watermark bit per equation (3).

$$i = ir \cdot N\_c + ic \tag{3}$$

According to the preceding steps, “many-to-one” mapping relationships are established between the vertex coordinates and the index of the watermark bit, the watermark bits are embedded into the corresponding vertex coordinates, and the quantization algorithm is adopted for embedding the watermark bit to ensure the blind extraction of the watermark bit [24].

3. Watermark Detection

The extraction of the watermark is the inverse process of watermark embedding. The basic steps for extracting watermark bits are as follows: for arbitrary vertex coordinates (x_i, y_i) , the watermark bit is extracted from (x_i, y_i) according to the quantization, and the index for the extracted watermark bit is then calculated based on the mapping function we established. In most cases, the number of vertex coordinates is far larger than the watermark length, which leads to the “many-to-one” mapping relationships between the vertex coordinates and the index of the watermark bit such that one watermark bit index will correspond to many extracted watermark bits.

Let the extracted watermark bits be the set W'; $W' = \{w'[i][j], 0 \le i < N, 0 \le j < L_i\}$, and $w'[i][j] \in \{-1,1\}$, where $w'[i][j]$ is the jth extracted watermark bit corresponding to the ith watermark bit index, N is the length of watermark, L_i represents the number of extracted watermark bits corresponding to the ith watermark bit index, and M represents the number of vertex coordinates in the detected vector geographic data. Additionally, $L_i \ge 0$ and $\sum_{i=0}^{N-1} L_i = M$.

3.1 Fixed Watermark Detection Algorithm

Currently, the most common algorithm used for watermark detection is fixed watermark detection, and the basic principle underlying this algorithm is as follows: first, establish a watermark W'' with a fixed watermark length of N based on the detected watermark bits W', and then calculate the correlation coefficient between W'' and the original watermark W to judge whether a watermark is contained in the detected data. Let the established watermark be W'' such that $W'' = \{w''[i], 0 \le i < N\}$ and $w''[i] \in \{-1,1\}$. In addition, $t_i = \sum_{j=0}^{L_i-1} w'[i][j]$, and watermark W'' can be established according to expression (4).

$$\begin{cases} w''[i] = 1 & if \;\; t_i > 0 \\ w''[i] = -1 & if \;\; t_i < 0 \end{cases} \tag{4}$$

Additionally, when $t_i = 0$, we set $w''[i]$ to 1 or -1 randomly.

When the extracted watermark W'' is acquired, the normalized correlation coefficient between W'' and the original watermark W is calculated. Let the normalized correlation coefficient be c_1, and we can then judge whether a watermark is contained in the detected data from the value of c_1. The formula for calculating the correlation coefficient is shown in equation (5).

$$c_1 = \frac{\sum_{i=0}^{N-1} w[i] * w''[i]}{N} \tag{5}$$

When a watermark is not contained in the detected vector geographic data, c_1 can be approximated by the following normal distribution with a mean of 0 and standard deviation of $1/N$ and shown in expression (6).

$$c_1 \sim N(0, \frac{1}{N}) \tag{6}$$

When a watermark is contained in the detected vector geographic data and watermarking attacks on the watermarked data are not observed, then correlation coefficient c_1 is equal to

1.0. Therefore, the detection threshold can be set to $4*\sqrt{1/N}$ to ensure that the FPE is less than 0.0001.

Based on the derivation of the formula of the above detection algorithm, we can obtain the detection threshold by controlling the FPE, although calculating the FNE is difficult because the influence of the watermarking attack on the detection result is not clear; therefore, quantitatively analyzing the robustness of the watermarking algorithm is also difficult. To solve this problem, we design an adaptive watermark detection algorithm for vector geographic data.

3.2 Adaptive Watermark Detection Algorithm

To solve the above-mentioned problem, another normalized correlation coefficient, c_2, is directly established based on the extracted watermark bit set W' and the number of detected vertex coordinates M as shown in equation (7).

$$c_2 = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{L_i-1} w[i]*w'[i][j]}{M} \tag{7}$$

We will now discuss the statistical characteristics of the normalized correlation coefficient c_2 under the following two conditions.

3.2.1 Non-watermarked Data

When a watermark is not contained in the detected data, for watermark bit index i, the extracted watermark bit $w'[i][j]$ is randomly equal to 1 and -1; therefore, we can suppose that the probabilities that $w'[i][j]=1$ and $w'[i][j]=-1$ are equal, i.e., $P(w'[i][j]=-1)=1/2$ and $P(w'[i][j]=1)=1/2$.

The mean and standard deviation of $w[i]*w'[i][j]$ can be expressed as $E(w[i]*w'[i][j])=0$ and $D(w[i]*w'[i][j])=1$, respectively. According to the central limit theorem, correlation coefficient c_2 can be approximated by the normal distribution with a mean of 0 and a standard deviation of $1/M$ as shown in expression (8).

$$c_2 \sim N(0,\frac{1}{M}) \tag{8}$$

3.2.2 Watermarked Data

Suppose that the detected data consist of two different parts, the watermarked data and the non-watermarked data. The extracted watermark bit set W' can then be expressed as two parts, $W'=W_1'\cup W_2'$, where W_1' is the watermark bit set extracted from the watermarked data and W_2' is the watermark bit set extracted from the non-watermarked data. Let W_1' and W_2' be $W_1'=\{w_1'[i][j], 0\le i<N, 0\le j<L_{i\_1}\}$ and $W_2'=\{w_2'[i][j], 0\le i<N, 0\le j<L_{i\_2}\}$, respectively, where $W'=W_1'\cup W_2'$ and $L_i=L_{i\_1}+L_{i\_2}$. Additionally, let $M_1=\sum_{i=0}^{N-1}L_{i\_1}$ and $M_2=\sum_{i=0}^{N-1}L_{i\_2}$, where $M=M_1+M_2$.

The normalized correlation coefficient c_2 can therefore be expressed as equation (9).

$$c_2 = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{L_{i\_1}-1} w[i]*w_1'[i][j] + \sum_{i=0}^{N-1}\sum_{j=0}^{L_{i\_2}-1} w[i]*w_2'[i][j]}{M} \tag{9}$$

For any watermark bit $w_1'[i][j]$, the equation $w[i]=w_1'[i][j]$ is correct, and we can therefore obtain expression (10) from expression (9).

$$c_2 = \frac{M_1 + \sum_{i=0}^{N-1}\sum_{j=0}^{L_{i\_2}-1} w[i]*w_2'[i][j]}{M} \tag{10}$$

According to the deduction of expression (8), correlation coefficient c_2 can be approximated by the normal distribution with a mean of $\frac{M_1}{M}$ and standard deviation of $\frac{M-M_1}{M*M}$ as shown in expression （11）.

$$c_2 \sim N(\frac{M_1}{M}, \frac{M-M_1}{M*M}) \tag{11}$$

Based on expressions (8) and (11), correlation coefficient c_2 can be approximated by a normal distribution. For example, the probability distributions of c_2 when $M_1 = 100$ and $M = 200$ are shown in **Fig. 2** when the detected data partly contain or do not contain a watermark.

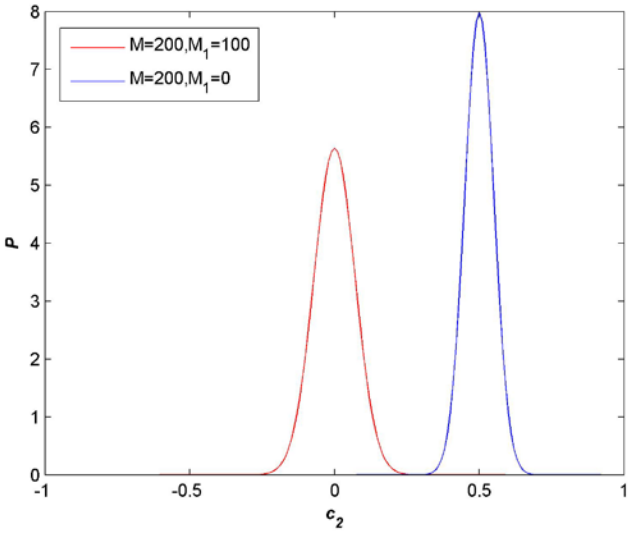


Fig. 2. Probability distributions for c_2 when the detected data partly contain or do not contain a watermark

Expression (8), equation (11) and **Fig. 2** show that the means and standard deviations are different when the detected data partly contain or do not contain a watermark, thus representing the mechanism underlying the watermark detection model.

1) Calculating thresholds using discriminant analysis when M_1 is known.

Based on expressions (8) and (11), the Mahalanobis distance discrimination analysis can be used to calculate the detection threshold. Let the distributed population for expression（8） be π_1 and the distributed population for expression 11） be π_2. The Mahalanobis distance of correlation coefficient c_2 to π_1 and π_2 can be depicted as shown in expressions （12）and （13）.

$$d(\pi_1, c_2) = \frac{|c_2 - 0|}{\frac{1}{\sqrt{M}}} \tag{12}$$

$$d(\pi_2, c_2) = \frac{\left|c_2 - \frac{M_1}{M}\right|}{\frac{\sqrt{M - M_1}}{M}} \tag{13}$$

When $0 < c_2 < \frac{M_1}{M}$, according to the Mahalanobis distance discrimination analysis, the discriminant function used to detect watermark can be illustrated as shown in equation (14).

$$\begin{aligned} T(c_2) &= d(\pi_1, c_2) - d(\pi_2, c_2) = c_2 * \sqrt{M} - \frac{M_1 - c_2 * M}{\sqrt{M - M_1}} \\ &= \frac{\sqrt{M} * \sqrt{M - M_1} + M}{\sqrt{M - M_1}} * (c_2 - \frac{M_1}{\sqrt{M} * \sqrt{M - M_1} + M}) \end{aligned} \tag{14}$$

The detection threshold is $\mu^* = \frac{M_1}{\sqrt{M} * \sqrt{M - M_1} + M}$, and the watermark detection rule is as shown in expression (15).

$$\begin{cases} \text{non-watermarked data} & \mathrm{c}_2 \le \mu^* \\ \text{watermarked data} & \mathrm{c}_2 > \mu^* \end{cases} \tag{15}$$

The FPE e_1 and FNE e_2 can be calculated via expression (16).

$$e_1 = e_2 = \Phi(-\frac{M_1}{\sqrt{M - M_1} + \sqrt{M}}) \tag{16}$$

2) Calculating the threshold by controlling the FNE when M_1 is unknown.

For the most part, the value of parameter M_1 is unknown in the process of watermark detection, and the discriminant analysis mentioned above is not suitable for calculating the detection threshold. In this situation, we can calculate the detection threshold by controlling the FNE. On the condition that FNE (e_1) and the number of detected vertex coordinates of carrier data (M) are fixed, detection threshold, μ^*, can be determined according to equation (17). Therefore, in watermark detection processing the detection probability of virtual detection is equal to e_1.

$$\Phi(\frac{\mu^*}{\sqrt{\frac{1}{M}}}) = 1 - e_1 \tag{17}$$

In practical applications, the value of e_1 is generally relatively small to ensure the effectiveness of watermark detection results. Generally, "3σ" principle or "4σ" principle can be used to determine the detection threshold. Using the "3σ" principle, the detection threshold is $\mu^* = 3*\sqrt{\frac{1}{M}}$. And the virtual detection probability is $e_1 = 0.0013$. Using the principle of "4σ", the detection threshold is $\mu^* = 4*\sqrt{\frac{1}{M}}$, the FPE, e_1, and FNE, e_2, are as shown in equation (18). In the experimental part of this paper, the "4σ" principle is used to determine the detection threshold.

$$\begin{cases} e_1 = 1 - \Phi(4) = 0.000031671 \\ e_2 = \Phi(\frac{4*\frac{1}{\sqrt{M}} - \frac{M_1}{M}}{\frac{\sqrt{M - M_1}}{M}}) = \Phi(\frac{4*\sqrt{M} - M_1}{\sqrt{M - M_1}}) \end{cases} \tag{18}$$

Compared with the fixed watermark detection algorithm mentioned above, equation (18) can be used to calculate the FNE, which is useful for quantitatively analyzing the robustness of the watermarking algorithm against data adding attacks, which will be verified in experimental sections.

4. Characteristic Analysis for the Adaptive Watermark Detection Algorithm

In Section 3.2, the adaptive watermark detection algorithm for vector geographic data was established based on the different probability distributions of correlation coefficients in different situations. In this section, we will quantitatively analyze the anti-noise capability of the proposed watermark detection algorithm.

From equation (18), we can determine that a change in M_1 has an influence on the FNE e_2. **Fig. 3** shows the influence of M_1 on the FNE e_2 when M is 1000, 3000, 5000 and 7000 based on equation (18). Vector geographic data are the fundamental achievements of national information infrastructure and earth science research and have been widely used in cartography, navigation, spatial analysis and many other areas. With the rapid development of computer and communication techniques, vector geographic data can be conveniently replicated and distributed. In addition, the acquisition of vector geographic data is a costly process, and as a consequence, the copyright protection of vector geographic data has become a hot issue in the geospatial data security domain.

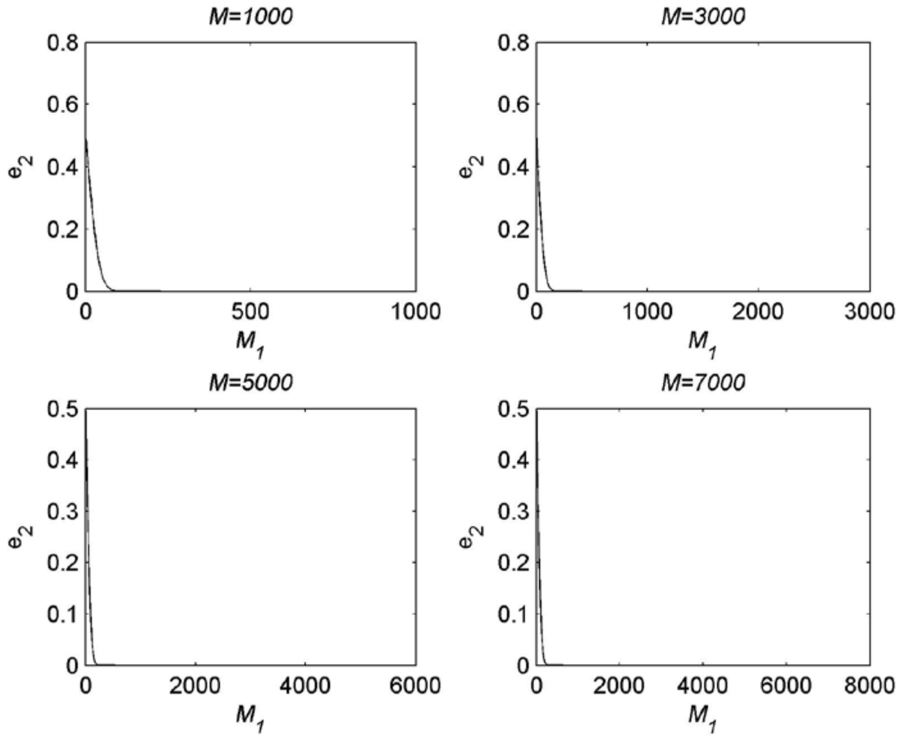


Fig. 3. Influence of M_1 on the FNE e_2 for M equal to 1000, 3000, 5000 and 7000

Fig. 3 shows that the FNE e_2 is equal to approximately 0.5 when M_1 is close to 0; in addition, the FNE e_2 decreases with increasing M_1 and eventually approaches 0.0, which means that watermark detection will not be missed.

Although correlation coefficient c_2 will increase as M_1 increases, the most important problem is determining a threshold value for M_1 to ensure that a watermark can be identified in the detected data. We will quantitatively discuss this issue under the FNE constraint.

Let the FNE constraint condition be e_2c. The minimum values of M_1 that can ensure a FNE greater than the e_2c for different values of M were calculated using expression (18), and the results are shown in **Table 1** for a FNE constraint of 0.001, where $\min M_1$ is the minimum value of M_1 and FNE_C is the corresponding FNE with parameters M and M_1.

Table 1. Minimum values of M_1 under the FNE constraint ($e_2c = 10^{-3}$)

ID	M	min M_1	FNE_C (10^{-4})
1	200	40	8.4498
2	400	58	8.8762
3	600	72	8.8513
4	800	83	9.8602
5	1000	94	9.0814
6	2000	134	9.6767
7	4000	191	9.8825
8	6000	235	9.8614
9	8000	272	9.8872
10	10000	305	9.7680

Table 1 shows that the $\min M_1$ value increases with increasing M under the FNE constraint. To further analyze the anti-noise capability, the $\min M_1$ values for different M values under FNE constraints of 0.0001, 0.001 and 0.01 were calculated. **Fig. 4** shows the relationships between $\min M_1$ and M under the different FNE constraints, and **Fig. 5** shows the relationship between $\min M_1/M$ and M .

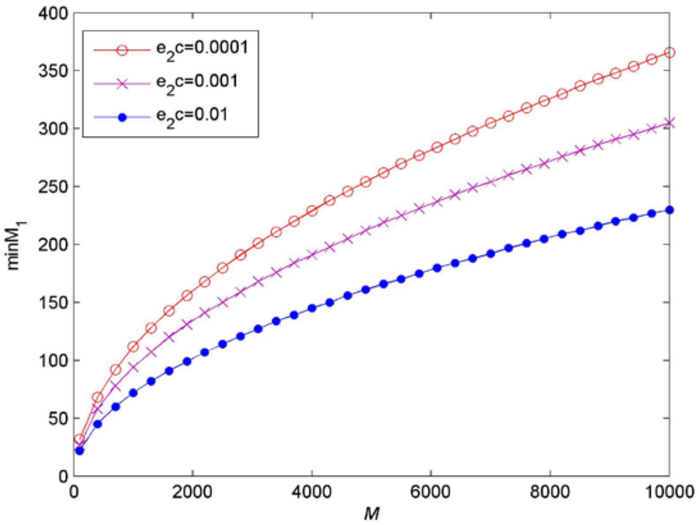


Fig. 4. Relationships between $\min M_1$ and M under different FNE constraints

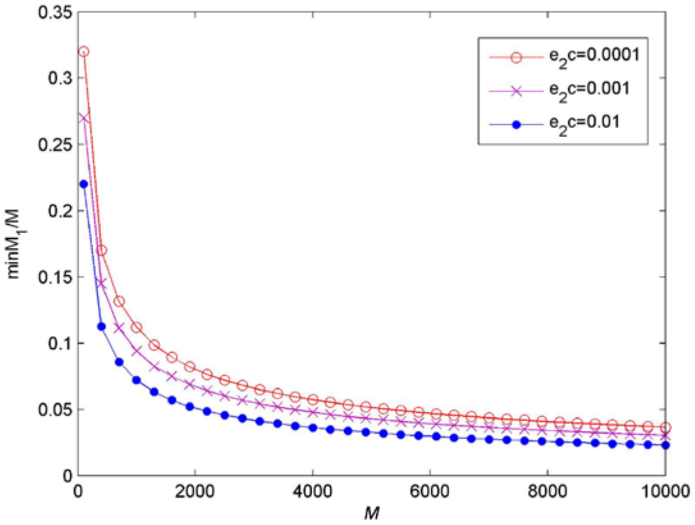


Fig. 5. Relationships between $\min M_1/M$ and M under different FNE constraints

Fig. 4 and 5 show that $\min M_1$ increases with increasing M and decreases with increasing e_2c and $\min M_1/M$ decreases with increasing M , indicating that smaller ratios of watermarked data to detected data are needed to ensure watermark detection from the detected data for increasing M .

5. Experiments and Results

To evaluate the performance of the proposed watermarking detection algorithm, several vector geographic maps at a scale of 1: 1,000,000 were adopted. We performed several experiments to verify the usability and robustness of the proposed watermarking detection algorithm.

5.1 Applicability Analysis

We established the two different probability distributions for c_2 in Section 3.2, and they represent the basic principle for calculating the watermarking detection threshold. However, the premise for this detection model is the Lindeberg conditions, which are difficult to demonstrate by means of theoretical derivation. Therefore, the applicability of expression (8) is verified by making experiments by using 151 different 2D vector maps.

The watermark detection results are mainly influenced by the quantization step and the original watermark when the mapping relationship is fixed. Therefore, the experiments were performed using the following steps.

1) Set the quantization step s for the watermark bit extraction.
2) Extract all watermark bits from the cover data.
3) Generate 10 watermarks randomly, calculate correlation coefficient c_2 between the extracted watermark bits and the watermarks generated according to equation (7), and normalize correlation coefficient c_2 to c_2' using the formula $c_2' = c_2 * sqrt(M)$.
4) Repeat experimental steps 1 – 3 using the 151 different 2D vector maps.

The above experiments are repeated for 4, 6, 8 and 10 quantization steps s, and we obtain 4 groups of experimental results, each of which contains 1510 normalized correlation coefficients c_2'. The flow chart for every quantization step present in **Fig. 6**. i, j, and k are the positive integers. And i is the number of quantization steps, j is the number of 2D vector maps, k is the number of randomly generating watermarks. $Data\,i$ is Data as shown in **Fig. 7**.

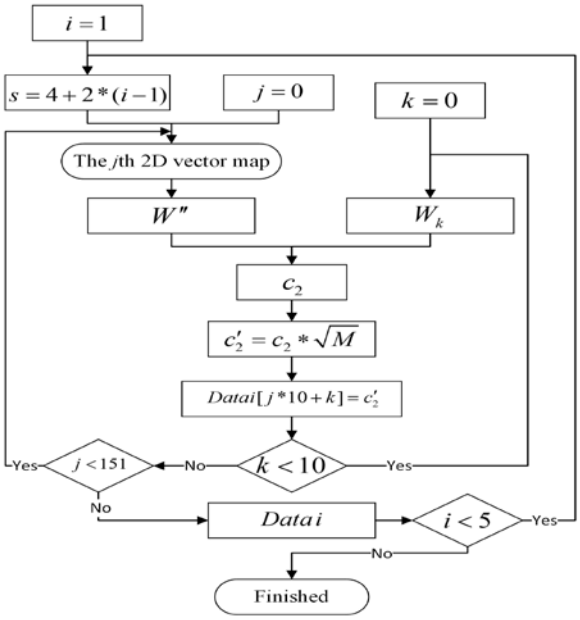
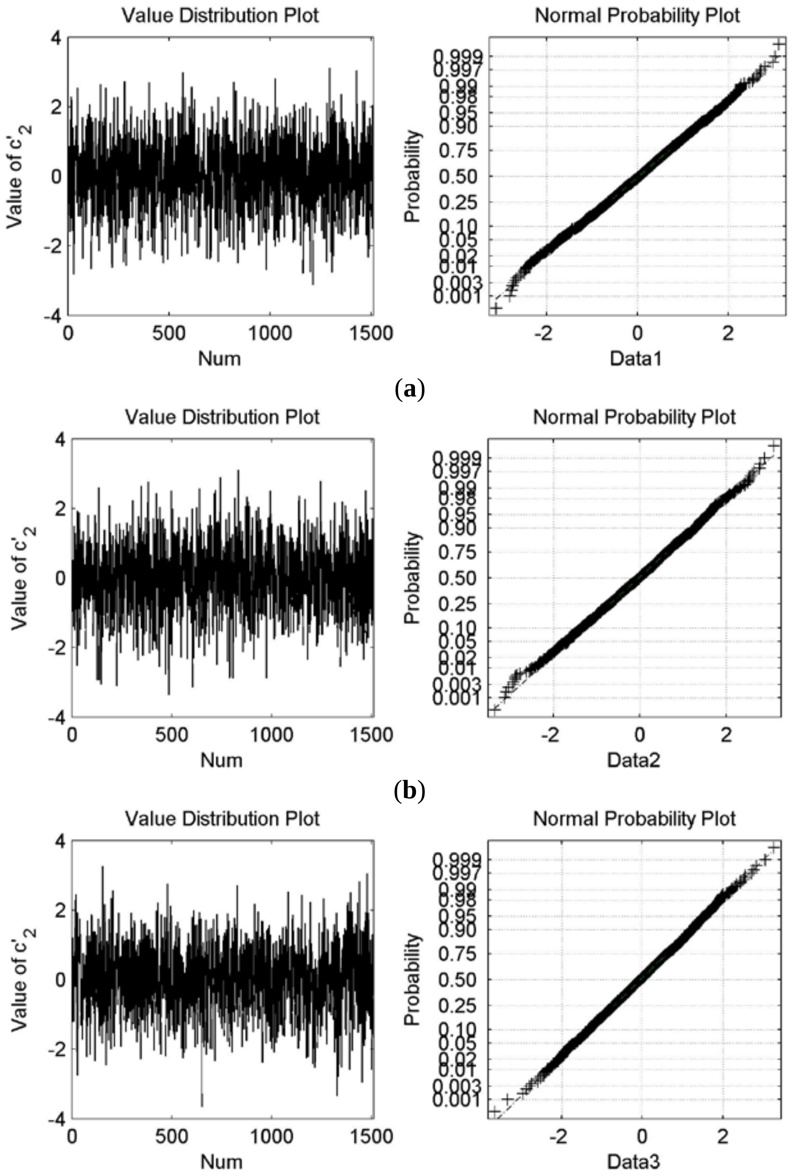


Fig. 6. The flow chart of the applicability analysis experiments

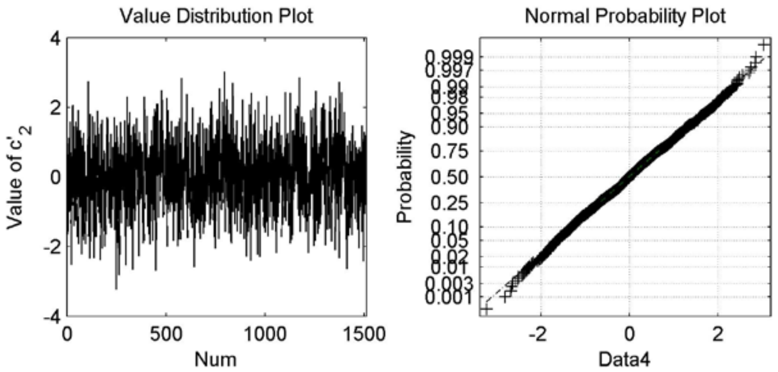
W'' is the extraction watermark. W_k is the randomly generation watermark. According to expression (8), the normalized correlation coefficient c_2' obeys the standard normal distribution, i.e., $c_2' \sim N(0,1)$. The 4 groups of experimental results are denoted Data1, Data2, Data3 and Data4, and the value distribution and normal probability plots for the 4 groups of experimental results are shown in **Fig. 7**.



(**a**)

(**b**)

(**c**)



(**d**)

Fig. 7. Value distribution and normal probability plots for the experimental results. (a) Plots for Data1; the mean value is 0.0216, and the standard deviation is 1.0202. (b) Plots for Data2; the mean value is -0.0394, and the standard deviation is 1.0083. (c) Plots for Data3; the mean value is -0.0485, and the standard deviation is 0.9868. (d) Plots for Data4; the mean value is -0.0031, and the standard deviation is 1.0045

Fig. 7 shows that the experimental results were clustered near 0.0 and all experimental results ranged from -4.0 to 4.0. The mean values and standard deviations of the 4 experimental datasets and the corresponding normal probability plots show that the experimental results can be approximated by a normal distribution with a mean of 0 and standard deviation of 1.0. This result is consistent with the theoretical derivation; therefore, the adaptive watermark detection proposed here is suitable for use with vector geographic data.

5.2 Robustness Analysis

Data deletion and data addition are the most common types of data processing in GIS. We will discuss the robustness of the proposed algorithm against data deletion and data addition attacks in this section.

5.2.1 Data Deletion Attacks

Data deletion processing consists of deleting vertices, compressing data, cropping data, and deleting features. The steps used to analyze robustness against data deletion included embedding the watermark into the experimental data, performing data deletion attacks on the watermarked data, and detecting the watermark from the attacked vector geographic data using the two watermark detection algorithms mentioned in Section 3. We randomly chose 10 different 2D vector maps from the experimental data to perform the experiment. **Fig. 8** shows one of the experimental polyline-based vector maps that consists of 40591 vertex coordinates, and **Table 2** shows the corresponding watermark detection results.

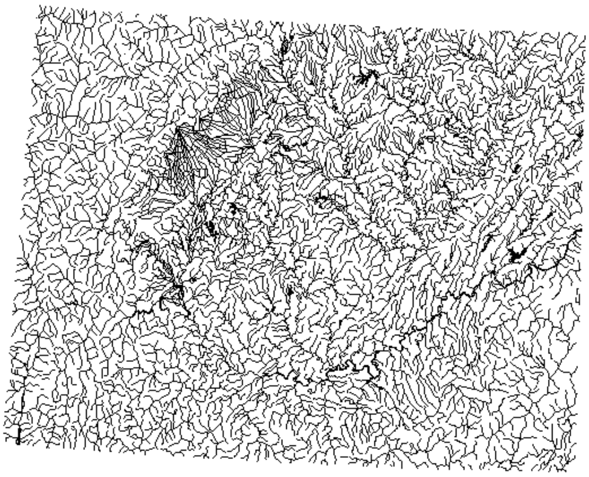
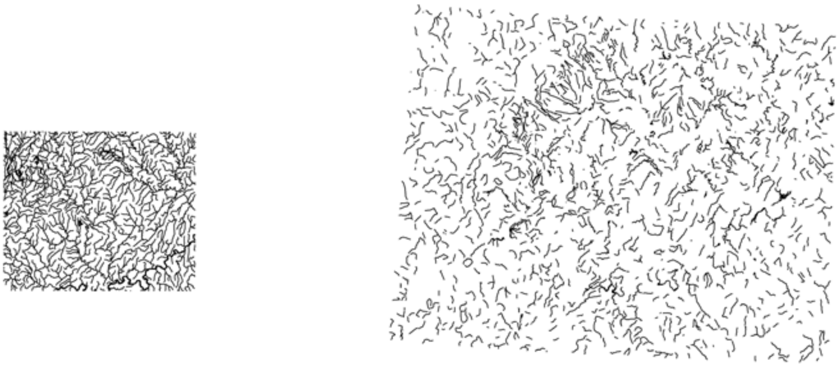


Fig. 8. Polyline-based vector map for the data deletion attack experiments

As expressed in **Table 2**, the deletion attacks are watermarking attacks on the watermarked data, and they consist of deleting vertices, cropping data, compressing data and deleting features. For these watermarking attacks, deleting vertices refers to randomly delete coordinate points from the watermarked data; the data cropping attack crops any regions of watermarked data (as shown in **Fig. 9_(a)**). The data compression attack uses the Douglas-Peucker algorithm to compress the watermarked data [25]. The feature detecting attack randomly deletes features from the watermarked data (as shown in **Fig. 9_(b)**). The vertex number of residual data is the vertex number of detected data after the watermarking attacks, c_1 is the correlation coefficient calculated according to the fixed watermark detection algorithm, c_2 is the correlation coefficient calculated according to the adaptive watermark detection algorithm, and √ indicates that the watermark can be detected from the cover data.



(**a**) (**b**)

Fig. 9. Cropping attack and deleting features attack

Table 2. Watermark detection results for the experimental data

Deletion attacks	**Vertex number of residual data**	$\mathbf{C_1}$	$\mathbf{C_2}$
Deleting vertices	32473	1.0(√)	1.0(√)
	16237	1.0(√)	1.0(√)
	8199	1.0(√)	1.0(√)
	812	1.0(√)	9.7(√)
	406	1.0(√)	0.85(√)
Cropping data	35279	1.0(√)	1.0(√)
	14485	1.0(√)	1.0(√)
	2529	1.0(√)	1.0(√)
	558	1.0(√)	0.93(√)
	201	1.0(√)	0.58(√)
Compressing data	33456	1.0(√)	1.0(√)
	29786	1.0(√)	1.0(√)
	23562	1.0(√)	1.0(√)
	16543	1.0(√)	1.0(√)
	11754	1.0(√)	1.0(√)
Deleting features	15678	1.0(√)	1.0(√)
	4229	1.0(√)	1.0(√)
	1300	1.0(√)	1.0(√)
	728	1.0(√)	0.95(√)
	299	1.0(√)	0.82(√)

Table 2 shows that we can detect the watermark from the attacked data, the correlation coefficients c_1 and c_2 are both equal to 1.0 when the vertex number of residual data is large, correlation coefficient c_2 remains equal to 1.0, and correlation coefficient c_1 decreases as the vertex number decreases when the vertex number of the residual data is small (e.g., less than 1000). Coefficient c_1 is not equal to 1.0 because watermark bits are not extracted for part of the watermark bit indexes when the vertex number of the residual data was small.

We performed the same experiments on other experimental data, and the experimental results were consistent with those in **Table 2**.

5.2.2 Data Addition Attacks

The data addition process consists of adding vertices, merging data, and smoothing data. The steps used to analyze robustness against data additions included embedding the watermark into the experimental data, performing data addition attacks on the watermarked data, and then detecting the watermark from the attacked vector geographic data using the two watermark detection algorithms mentioned in Section 3. **Fig. 10** shows the experimental point-based vector map that consists of 2613 vertex coordinates. In the experiments, we randomly added vertices into the cover data.

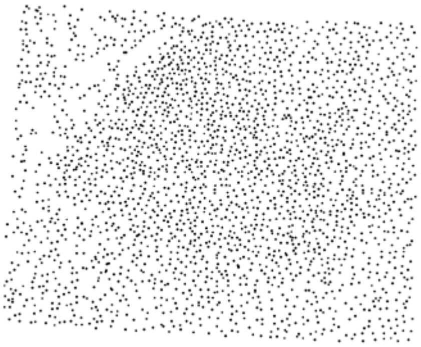
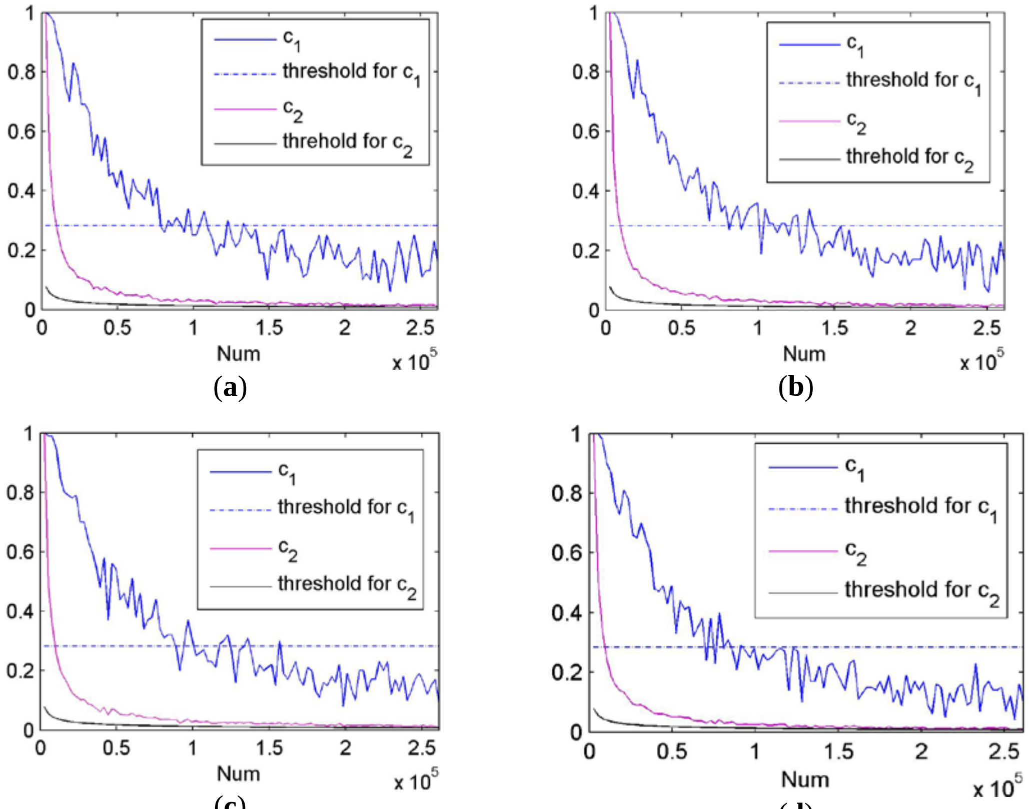


Fig. 10. Point-based vector map for data addition attack experiments

Although the strengths of watermark attacks may be similar, the detection results may differ if the vertices are added randomly. Therefore, we performed 4 experiments using the experimental data. **Fig. 11** shows the 4 experimental results, for which Num is the vertex number of detected data, which increases as the strength of the data addition attack increases.



(a)

(b)

(c)

(d)

Fig. 11. Experimental results for the data addition attacks

Fig. 11 shows that the correlation coefficients c_1 and c_2 decreased as the strength of the data addition attack increased. In **Fig. 11**, the curve for correlation coefficient c_2 is more stable than that of c_1, which means that correlation coefficient c_2 is more predictable.

Correlation coefficient c_2 was distributed around M_1 / M , which is consistent with formula (11). **Fig. 11** shows that the adaptive watermark detection method was more robust than the fixed watermark detection method for data addition attacks.

We randomly chose 10 different 2D vector maps from the 151 experimental maps to perform the same data addition experiments mentioned above, and the experimental results are consistent with the results shown in **Fig. 11**.

6. Conclusion

In this paper, an adaptive watermark detection algorithm for vector geographic data is proposed that can quantitatively analyze the robustness of watermark methods for data addition. The following conclusions were derived from the results.

1) A watermark detection threshold can be calculated based on the proposed detection model, and the FPE and FNE can be determined to evaluate the robustness of the watermarking algorithm.
2) Data deletion attacks (such as cropping data and deleting vertices) have no influence on the watermark bit detection using residual data, and we can detect watermarks from residual watermarked data after deletion attacks when the amount of residual watermarked data is not very small.
3) Data addition attacks (such as merging data and inserting vertices) are equivalent to when a part of the detected data contains a watermark while another part does not contain a watermark, and the robustness against data addition can be quantitatively analyzed using the proposed watermark detection algorithm.
4) Only part of the vector geographic data containing a watermark can ensure that we detect the watermark from the detected data, which provides the ability to only select the modifiable vertex coordinates to embed a watermark when non-modifiable vertex coordinates are contained in the dataset.

In the processing of vector geographic data, rotation, scaling, and translation are also common. These transformations are all attacks for the cover data. Future work will include the development of a watermarking detection algorithm that can be used to quantitatively analyze these watermarking attacks.

References

[1] Cox I. J., Miller M. L. and Bloom J. A., *Digital Watermarking*, Morgan Kaufmann, San Diego, California, 2002; pp. XV. Article (CrossRef Link).

[2] Voigt M., Yang B. and Busch C., “Reversible watermarking of 2D-vector data,” in *Proc. of the 2004 Workshop on Multimedia and Security*, pp.160-165, September 20-21, 2004. Article (CrossRef Link).

[3] Kang H. and Iwamura K., “Information hiding method using best DCT and wavelet coefficients and its watermark competition,” *Entropy*, vol. 17, no. 3, pp. 1218-1235, March, 2015. Article (CrossRef Link).

[4] Solachidis V. N. and Nikolaidis I. P., “Watermarking polygonal lines using Fourier descriptors,” *IEEE Computer Graphics and Applications*, vol. 24, no. 3, pp. 44-51, May, 2004. Article (CrossRef Link).

[5] S. N. Neyman, I. N. P. Pradnyana and B. Sitohang, "A new copyright protection for vector map using FFT-based watermarking," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 12, no. 2, pp. 367-378, June, 2014.

[6] M. Urvoy, D. Goudia and F. Autrusseau, "Perceptual DFT watermarking with improved detection and robustness to geometrical distortions," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1108-1119, July, 2014. Article (CrossRef Link).

[7] Y. Li and L. Xu, "A blind watermarking of vector graphics images," in *Proc. of 5th International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 2003*, pp. 424-429, September 27-30, 2003. Article (CrossRef Link).

[8] A. Benoraira, K. Benmahammed and N. Boucenna, "Blind image watermarking technique based on differential embedding in DWT and DCT domains," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 55-65, July, 2015. Article (CrossRef Link).

[9] R. Ohbuchi, H. Ueda and S. Endoh, "Robust watermarking of vector digital maps," in *Proc. of 2002 IEEE International Conference on Multimedia and Expo*, pp. 577-580, August 26-29, 2002. Article (CrossRef Link).

[10] X. Zhou, Y. Ren and X. 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, July, 2006. Article (CrossRef Link).

[11] Yang, C. S., Zhu, C. Q. and Wang, Y. Y, "Robust watermarking algorithm for geometrical transform for vector geo-spatial data based on invariant function," *Acta Geodaetica Et Cartographica Sinica*, vol. 40, no. 12, pp. 256-261, November, 2011. Article (CrossRef Link).

[12] S. H. Lee, X. J. Huo and K. R. Kwon, "Vector watermarking method for digital map protection using arc length distribution," *IEICE Transactions on Information and Systems*, vol. E97-D, no. 1, pp. 34-42, January, 2014. Article (CrossRef Link).

[13] Z. Peng, M. Yue, X. Wu and Y. Peng, "Blind watermarking scheme for polylines in vector geo-spatial data," *Multimedia Tools and Applications*, vol. 74, no. 24, pp. 11721-11739, December, 2015. Article (CrossRef Link).

[14] N. Wang, "Reversible watermarking for 2D vector maps based on normalized vertices," *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 20955-20956, October, 2017. Article (CrossRef Link).

[15] N. N. Wang and X. J. Zhao, "2D vector map data hiding with directional relations preservation between points," *Aeu-International Journal of Electronics and Communications*, vol. 71, pp. 118-124, January, 2017. Article (CrossRef Link).

[16] Wang, Y. Y., Yang, C. S. and Zhu, C. Q, "Digital watermarking against data merging attack for vector geographic data," *Journal of Beijing University of Posts and Telecommunications*, vol. 40, no. 4, pp.48-53, July, 2017**.** Article (CrossRef Link).

[17] A. Adelsbach, S. Katzenbeisser and A. R. Sadeghi, "Watermark detection with zero-knowledge disclosure," *Multimedia Systems*, vol. 9, no. 3, pp. 266-278, September, 2003. Article (CrossRef Link).

[18] A. Adelsbach, M. Rohe and A.-R. Sadeghi, "Non-interactive Watermark Detection for a Correlation-Based Watermarking Scheme," *Communications and Multimedia Security*, pp. 129-139, September 19-21, 2005. Article (CrossRef Link).

[19] M. Malkin and T. Kalker, "A Cryptographic Method for Secure Watermark Detection," in *Proc. of Information Hiding*, pp. 26-41, July 10-12, 2006. Article (CrossRef Link).

[20] Shao, C. Y., Wang, H. L., Niu, X. M. and Wang, X. T., "A shape-preserving method for watermarking 2D vector maps based on statistic detection," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 3, pp.1290-1293, March, 2006. Article (CrossRef Link).

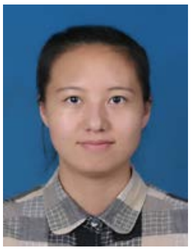
[21] Zhu X. W., "Research of blind watermark detection algorithm based on generalized Gaussian distribution," *Journal of Software*, vol. 5, pp. 413-420, April, 2010. Article (CrossRef Link).

[22] C. Zuo, A. Li and C. Meng, "GIS vector data automatic watermark detection based on mobile agent technology," in *Proc. of 18th International Conference on Geoinformatics*, pp. 1-4, June 18-20, 2010. Article (CrossRef Link).

[23] Yang, C. S., Zhu, C. Q. and Wang, Y. Y., “Self-detection watermarking algorithm and its application to vector geo-spatial data,” *Geomatics and information science of Wuhan University*, vol. 36, no. 12, pp. 1402-1405, December, 2011. Article (CrossRef Link).

[24] B. Chen and G. W. Wornell, “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,” *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423-1442, May, 2001. Article (CrossRef Link).

[25] D. H. Douglas and T. K. Peucker, “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature,” *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, no. 2, pp. 112-122, December, 1973. Article (CrossRef Link).

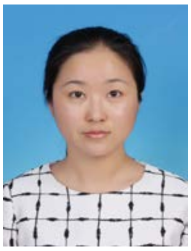


Yingying Wang: She received the Ph.D. degrees in cartography and geographic information system from the Nanjing Normal University, Nanjing, China, in 2018.

Currently, she is Lecturer in Jinling Institute of Technology, Nanjing, China. Her research interest includes the applications of geographic information system and geographic information security.

Chengsong Yang: He received the Ph.D. degrees in geographic information system from the Information Engineering University, Zhengzhou, China, in 2011.

Currently, he is Lecturer in the Institute of Field Engineering, Army Engineering University of PLA, Nanjing, China. His research interests are in the broad areas of geographic information system (GIS) and digital image processing, with a focus on geographic information security.

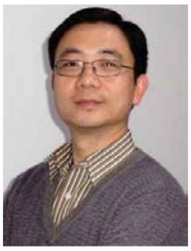


Na Ren: She received the Ph.D. degrees in cartography and geographic information system from the Nanjing Normal University, Nanjing, China, in 2011.

Currently, she is an Associate Professor in Nanjing Normal University, Nanjing, China. Her research interests are in theories and methods of geographic information system and digital image processing, with a focus on spatial information security.

Changqing Zhu: He received the Ph.D. degrees in mathematics from the Information Engineering University, Zhengzhou, China, in 1997.

Currently, he is a Professor in Nanjing Normal University, Nanjing, China. His research interests are in theories and methods of geographic information system and digital image processing, with a focus on spatial information security.



Ting Rui: He received the M.S. degree and Ph.D. from PLA University of Science and Technology, Nanjing, China, in1998 and 2001respectively. Ting RUI is Professor of the Institute of Field Engineering, Army Engineering University of PLA, Nanjing, China. He mainly applies computer vision, machine learning, multimedia, and video surveillance. He has authored and co-authored more than 80 scientific articles.

Dong Wang: He received PhD degree in Vehicle Engineering from PLA University of Science and Technology, Nanjing City, China, in 2013.

Now he works as a lecture at Army Engineering University of PLA, China. And he is a postdoctoral researcher in Second Institute of Engineering Research and Design, Southern Theatre Command, China. His current research interests include Unmanned technology, fault diagnosis and signal processing.