Robust Watermarking Scheme for Vector Geographic Data Based on the Ratio Invariance of DWT–CSVD Coefficients

Chengyi Qu \(^1\), Xu Xi \(^{1,*}\), Jinglong Du \(^1\) and Tong Wu \(^2\)

\(^1\) School of Geography Science and Geomatics Engineering, Suzhou University of Science and Technology, Suzhou 215000, China

\(^2\) National Marine Environmental Monitoring Center, Dalian 116085, China

\(^*\) Correspondence: xixu@usts.edu.cn

Abstract: Traditional frequency-domain watermarking algorithms for vector geographic data suffer from disadvantages such as the random watermark embedding position, unpredictable embedding strength, and difficulty in resisting multiple attacks at the same time. To address these problems, we propose a novel watermarking algorithm based on the geometric invariance of the ratios of discrete wavelet transform (DWT) and complex singular value decomposition (CSVD) coefficients, which embeds the watermark information in a new embedding domain. The proposed scheme first extracts feature points from the original vector geographic data using the Douglas–Peucker algorithm, and then constructs a complex sequence based on the feature points set. The two-level DWT is then performed on the complex sequence to obtain the low frequency coefficients (L\(_2\)) and high frequency coefficients (H\(_2\)). On this premise, the CSVD algorithm is utilized to calculate the singular values of L\(_2\) and H\(_2\), and the ratio of the singular values of L\(_2\) and H\(_2\) is acquired as the watermark embedding domain. During the watermark embedding process, a new watermark sequence is created by the fusion of the original watermark index value and bits value to improve the recognition of the watermark information, and the decimal part at different positions of the ratio is altered by the new watermark sequence to control the watermark embedding strength. The experimental results show that the proposed watermarking algorithm is not only robust to common attacks such as geometric, cropping, simplification, and coordinate point editing, but also can extract watermark images with a high probability under random multiple attacks.

Keywords: discrete wavelet transform; complex singular value decomposition; embedding domain; digital watermarking; vector geographical data

1. Introduction

The vector geographic data are widely utilized in urban planning, traffic navigation, disaster prevention and control, military applications, and other fields with a high socio-economic value \([1–4]\). With the rapid development of software, hardware, and internet technology, the sharing and transmission of vector geographical data is becoming simpler and more efficient, which may result in many copyright issues such as illegal storage, copying, and trafficking \([4–8]\). As a cutting-edge technology in the field of information security, digital watermarking is attracting increasing attention and, over time, has become an important technology for the copyright protection of vector geographical data \([9,10]\).

Digital watermarking works by hiding meaningful (e.g., text or images containing producer information) or meaningless (e.g., randomly generated binary sequences) copyright information into the host data by preset information-concealing algorithms. In cases of copyright dispute, the information can only be extracted by re-producing the preset watermarking algorithms, and copyright can then be authenticated through an information comparison \([9]\). This characteristic determines that when digital watermarking is applied to vector geographic data, it is necessary to find a suitable watermark embedding domain.
that can meet the demand of using vector geographic data with a high accuracy, but also ensures that the watermarking algorithm is robust enough to resist complex editing of the vector geographic data. Generally speaking, digital watermarking algorithms for the vector geographic data are classified into spatial or frequency domain algorithms according to the position of the embedded watermark [11]. Cox et al. [12] directly encoded the coordinate values and embedded the watermark information into each vertex, and this was the first spatial domain watermarking approach for the vector data. Despite its lack of robustness, it has been a focus for further research. As the research has progressed, scholars have provided more solutions for improving the performance of the algorithm and have also mined increasing numbers of watermark embedding domains. Wang et al. [13] embedded watermark information by adjusting the value of the spatial topological metric to ensure the performance of the watermarking procedure, using the stability of the topological relations, which also guaranteed that the embedding of the watermark did not result in changes to the topological relations. Lee et al. [14] used relative and absolute position accuracy as constraints to embed watermarks in the perimeter of the polylines and the area of the polygons, ensuring that the vertices embedded in the watermark were within the allowable tolerance. This method could more successfully control the accuracy and topology changes after the watermark was embedded. Wu et al. [15] used polygonal similarity transformation, modulating watermark information in the feature set of spatial relationships of polygonal elements based on quantized index modulation (QIM), effectively maintaining the geometric features. Ren et al. [16] used the angular value between the line segments as the watermark embedding domain and embedded the watermark information using QIM, which had very little impact on the data fidelity. Generally, in the current research of spatial domain watermarking algorithms for vector geographic data, the geometric feature information or statistical information of various data are mined as the embedding domain of the watermark, and a good invisibility can be achieved by controlling the watermark embedding strength and method. However, in this type of algorithm, the watermark information is directly hidden in the coordinates of the vector geographic data, which can be easily removed via precision erase, and as a consequence this type of algorithm tends to be less secure and less practical [17].

Frequency domain watermarking algorithms are generally more robust and covert than spatial domain watermarking algorithms [10,18]. The watermark information is embedded by modifying the frequency domain transform coefficients of the coordinate values in the transformation domain scheme, which brings about a stronger concealment [19]. There are three classical frequency domain algorithms: (1) discrete Fourier transform (DFT), (2) discrete wavelet transform (DWT), and (3) discrete cosine transform (DCT), all of which have been widely studied in vector geographic data watermarking [20]. Xu et al. [21], Neyman et al. [20], and Zhang et al. [2] exploited the geometric invariance of the DFT coefficients to embed the watermark information. These algorithms were robust to geometric attacks but difficult in resisting vertex attacks such as vertex deletion, vertex insertion, or map cropping attacks. Zhang [22] enhanced the previous literatures by embedding watermarking information both in the DFT coefficients and spatial domain. The scheme was resistant to geometric attacks and vertex attacks but unable to resist combination attacks. Kitamura et al. [23] uniformly subdivided the vertexes into rectangular grids, and then taken each rectangular grid as the “pixel” of the grid image. Then, DWT for the images was utilized to embed the watermarks. Yang et al. [24] and Zope-Chaudhari et al. [25,26] achieved a copyright protection of the vector geographic data by embedding the watermarking information in the DWT coefficients. These schemes showed a strong robustness against noise, vertex deletion, and format transformation attacks. However, those methods were susceptible to geometric attacks and had a high impact on the data fidelity. Schemes [27,28] concealed information by fine-tuning the middle frequency of the DCT coefficients. The experimental results showed that their watermarking algorithm had a good resistance to geometric attacks but a poor invisibility. Wang et al. [29] embedded the watermark in the high-frequency coefficients of the DCT domain while also developing an error control
method, which helps to reduce distortion after the watermark embedding. In summary, the frequency domain watermarking of the vector geographic data is gaining interest, but due to the nature of the algorithms, it is challenging to withstand multiple attacks at once, and the position of the watermark information during embedding is highly random, making it challenging to control the invisibility. These drawbacks seriously limit its potential for a practical application. To address these problems, we will dig deeper into the feature invariants in the DWT coefficients, aiming to design a digital watermarking scheme that can withstand multiple attacks while controlling the invisibility.

The objectives of this study are to: (1) propose a robust watermarking scheme for vector geographical data based on the ratio stability of the DWT and CSVD coefficients, and (2) to evaluate the performance of the proposed scheme through attack experiments and statistical analysis.

2. Methods

2.1. Ratio Invariance of DWT-CSVD Coefficients

The frequency transformation is implemented in the polyline which can be donated as \( L = (E, V) \) with an edge set \( E \) and a vertex set \( V = \{v_1, v_2, \ldots, v_n\} \), where \( v_i = (x_i, y_i) \), \( x_i, y_i \in \mathbb{R} \). A polygon can be expressed as a closed polyline.

The proposed ratio invariance of DWT–CSVD transformation coefficients refers to the property that the ratio of the singular values, which are obtained from the CSVD algorithm performed on the low- and high-frequency coefficients of 2-level Haar wavelet-based DWT, is invariant under rotation, scaling, and translation (RST) operations. Details of the invariance are as follows: firstly, the complex sequence \( C = \{c_1, c_2, \ldots, c_n\} \) is constructed from vertex set \( V \) according to Equation (1):

\[
c_k = x_k + y_k \times i (k = 1, 2, \ldots, n)
\]  

Then, the complex sequence \( C \) is transformed by Haar wavelet-based DWT to obtain the multi-resolution representations: \( L \) and \( H \), where \( L \) is the low-frequency coefficient, also known as the approximate component which includes most energy of the original vertexes, and \( H \) is the high-frequency coefficient, also known as the detail component, which includes little energy. After that, perform Haar wavelet-based DWT on \( H \) to obtain the multi-resolution representations of \( H \): \( L_2 \) and \( H_2 \), where \( L_2 \) is the approximate component of \( H \) and \( H_2 \) is the detail component of \( H \).

Based on \( L_2 \) and \( H_2 \), the complex matrixes \( \text{Mat}_{L_2} \) and \( \text{Mat}_{H_2} \) are constructed, respectively. The CSVD algorithms performed on \( \text{Mat}_{L_2} \) and \( \text{Mat}_{H_2} \) are shown in Equations (2) and (3).

\[
\text{Mat}_{L_2} = \mathbf{U}_{L_2} \times S_{L_2} \times V_{L_2}^H = \mathbf{U}_{L_2} \times sl \times V_{L_2}^H
\]

\[
\text{Mat}_{H_2} = \mathbf{U}_{H_2} \times S_{H_2} \times V_{H_2}^H = \mathbf{U}_{L_2} \times sh \times V_{H_2}^H
\]

where \( \mathbf{U}_{L_2} \) and \( V_{L_2} \) represent the unitary matrix consisting of the left and right singular vectors, respectively. \( V_{L_2}^H \) is the Hermite matrix of \( V_{L_2} \), and \( S_{L_2} \) is the singular value matrix. Because \( \text{Mat}_{L_2} \) is a one-dimensional matrix, after CSVD, the singular value matrix, \( S_{L_2} \), has only one element, i.e., the singular value \( sl \). Equation (3) follows the same principle as Equation (2). Finally, the ratio \( K \) of the DWT–CSVD transformation coefficients can be calculated by Equation (4) after acquiring the singular values of \( L_2 \) and \( H_2 \).

\[
K = \frac{sl}{sh}
\]

The ratio \( K \) will keep invariant, after the original vertexes are attacked by RST. The invariance outlined above will be proved theoretically by the following:
First, the one-dimensional Haar wavelet-based DWT can be expressed by the form of matrix $W_\eta$ ($\eta$ is an even number), which is shown in Equation (5) [30]:

$$
W_{\eta \times \eta} = \begin{bmatrix}
\sqrt{2}/2 & \sqrt{2}/2 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \sqrt{2}/2 & \sqrt{2}/2 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & \cdots & \sqrt{2}/2 & \sqrt{2}/2 \\
\sqrt{2}/2 & -\sqrt{2}/2 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \sqrt{2}/2 & -\sqrt{2}/2 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & \cdots & \sqrt{2}/2 & -\sqrt{2}/2
\end{bmatrix}
$$

(5)

The following is an expression for the transformation of the original signal using Haar wavelet-based DWT:

$$
W_\eta \times \begin{bmatrix}
x_1 + y_1 i \\
x_2 + y_2 i \\
\vdots \\
x_\eta + y_\eta i
\end{bmatrix} = \frac{\sqrt{2}}{2} \times \begin{bmatrix}
(x_1 + x_2) + (y_1 + y_2)i \\
(x_3 + x_4) + (y_3 + y_4)i \\
\vdots \\
(x_{\eta-1} + x_\eta) + (y_{\eta-1} + y_\eta)i
\end{bmatrix}
$$

(6)

$$
H_{\eta/2} = \frac{\sqrt{2}}{2} \times \begin{bmatrix}
(x_1 + x_2) + (y_1 + y_2)i \\
(x_3 + x_4) + (y_3 + y_4)i \\
\vdots \\
(x_{\eta-1} + x_\eta) + (y_{\eta-1} + y_\eta)i
\end{bmatrix}
$$

(7)

$$
G_{\eta/2} = \frac{\sqrt{2}}{2} \times \begin{bmatrix}
(x_1 - x_2) + (y_1 - y_2)i \\
(x_3 - x_4) + (y_3 - y_4)i \\
\vdots \\
(x_{\eta-1} - x_\eta) + (y_{\eta-1} - y_\eta)i
\end{bmatrix}
$$

(8)

where $H_{\eta/2}$ and $G_{\eta/2}$ denote the approximate component and detail component, respectively.

When the original vertexes are translated along the x-direction by $\alpha$ and along the y-direction by $\beta$, there is

$$
W_\eta \times \begin{bmatrix}
(x_1 + \alpha) + (y_1 + \beta)i \\
(x_2 + \alpha) + (y_2 + \beta)i \\
\vdots \\
(x_\eta + \alpha) + (y_\eta + \beta)i
\end{bmatrix} = \frac{\sqrt{2}}{2} \times \begin{bmatrix}
(x_1 + x_2 + 2\alpha) + (y_1 + y_2 + 2\beta)i \\
(x_3 + x_4 + 2\alpha) + (y_3 + y_4 + 2\beta)i \\
\vdots \\
(x_{\eta-1} + x_\eta + 2\alpha) + (y_{\eta-1} + y_\eta + 2\beta)i
\end{bmatrix}
$$

(9)

From the above equations, as can be seen, the detail component is unaffected when the original vertexes are translated. On this basis, it is straightforward to prove that the approximate and detail components of the 2-level DWT are invariant when the original signal is translated. Clearly, the ratio of the singular values will not change.
When the original vertexes are scaling by \( \gamma (\gamma > 0) \), there is:

\[
W_\eta \times \begin{pmatrix}
(x_1 \times \gamma) + (y_1 \times \gamma)i \\
(x_2 \times \gamma) + (y_2 \times \gamma)i \\
\vdots \\
(x_\eta \times \gamma) + (y_\eta \times \gamma)i
\end{pmatrix} = \frac{\sqrt{\eta}}{2} \times \gamma \times \begin{pmatrix}
(x_1 + x_2) + (y_1 + y_2)i \\
(x_3 + x_4) + (y_3 + y_4)i \\
\vdots \\
(x_{\eta-1} + x_\eta) + (y_{\eta-1} + y_\eta)i
\end{pmatrix}
\]

(10)

It shows that both the approximate component and detail component are \( \gamma \) times the original value. The following analyze the characteristics of the singular values. The singular value for a complex matrix \( A_{\eta \times 1} \) can be used to calculate Equations (11) and (12).

\[
det \left( \lambda I - A^H \times A \right) = 0
\]

(11)

\[
s_A = \sqrt{\lambda}
\]

(12)

where \( det \) is an operation of obtaining the determinant, \( \lambda_A \) is the eigenvalue of matrix \( A^H \times A \), \( I \) is a unit matrix, and \( s_A \) is the singular value of \( A \). In case of the matrix \( B = \gamma \times A \), there is:

\[
det \left( \lambda I - \gamma^2 A^H \times A \right) = 0
\]

(13)

\[
s_B = \sqrt{\gamma^2 \lambda} = \gamma s_A
\]

(14)

where \( s_B \) is the singular value of \( B \). After scaling, the singular value is also scaled by the scaling ratio. The ratio of the singular values of the approximate component and detail component is identical to the original ratio since both the approximate component and detail component are scaled by the same ratio when the original vertexes are scaled.

When the original signal is rotated by \( \theta \), there is:

\[
W_\eta \times \begin{pmatrix}
(x_1 \cos \theta - y_1 \sin \theta) + (x_1 \sin \theta + y_1 \cos \theta)i \\
(x_2 \cos \theta - y_2 \sin \theta) + (x_2 \sin \theta + y_2 \cos \theta)i \\
\vdots \\
(x_\eta \cos \theta - y_\eta \sin \theta) + (x_\eta \sin \theta + y_\eta \cos \theta)i
\end{pmatrix} =
\]

\[
\begin{pmatrix}
(x_1 + x_2) \cos \theta - (y_1 + y_2) \sin \theta + \{(x_1 + x_2) \sin \theta + (y_1 + y_2) \cos \theta\}i \\
(x_3 + x_4) \cos \theta - (y_3 + y_4) \sin \theta + \{(x_3 + x_4) \sin \theta + (y_3 + y_4) \cos \theta\}i \\
\vdots \\
(x_{\eta-1} + x_\eta) \cos \theta - (y_{\eta-1} + y_\eta) \sin \theta + \{(x_{\eta-1} + x_\eta) \sin \theta + (y_{\eta-1} + y_\eta) \cos \theta\}i
\end{pmatrix}
\]

(15)

Based on Equations (11) and (12), the singular values of the aforesaid approximate component and detail component are as follows:

\[
s_{Ap} = \frac{\sqrt{2}}{2} \times \sqrt{\sum_{k=1}^{\eta} (x_{2k-1} + x_{2k})^2 + (y_{2k-1} + y_{2k})^2}
\]

(16)

\[
s_{Dc} = \frac{\sqrt{2}}{2} \times \sqrt{\sum_{k=1}^{\eta} (x_{2k-1} - x_{2k})^2 + (y_{2k-1} - y_{2k})^2}
\]

(17)

where \( s_{Ap} \) and \( s_{Dc} \) are the singular values of the above approximate component and detail component, respectively. When performing the same option on the original approximate
component and detail component, the singular values acquired are equal to the singular values above. Additionally, the ratios of the singular values are constant.

Since the ratio of the singular values of the approximate component and detail component is not altered by RST attacks, it is robust and provides a theoretical basis for digital watermarking algorithms for vector geographic data.

2.2. Watermark Information Generation

The watermark used in this paper is a binary image. Figure 1a shows a sample of a binary image with a size of 4 × 4 bits, and Figure 1b,c are the watermark index and watermark bits at the corresponding positions of the watermark image.

![Watermark Image Sample](image)

**Figure 1.** Sample of watermark image. (a) Watermark image; (b) watermark index; (c) watermark bit.

To enhance the security of the watermark, it is necessary to scramble the watermark using the Arnold transformation before embedding the watermark [31]. Arnold transformation, also called cat mapping which is proposed by Russian mathematician V.I. Arnold [32], works by mapping the original pixel position \((p, q)\) to a new position \((p', q')\). Arnold transformation is now generalized by introducing two parameters \(a\) and \(b\), as shown in Equation (18) [33].

\[
\begin{pmatrix} p' \\ q' \end{pmatrix} = \begin{pmatrix} 1 & a \\ b & 1 + ab \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} (\text{mod } N) \tag{18}
\]

where \(N\) is the size of the watermark image and \text{mod} is the modulo operation, \(a, b \in \mathbb{N}^+\). The parameters \(a\) and \(b\) can be used as a secret key for watermark detection.

Inverse Arnold scrambling transformation can be expressed as:

\[
\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} ab + 1 & -b \\ -a & 1 \end{pmatrix} \begin{pmatrix} p' \\ q' \end{pmatrix} (\text{mod } N) \tag{19}
\]

To improve the recognition of the watermark information, the original watermark bits are mapped to new values according to Equation (20):

\[
\text{bit}_m = \begin{cases} 3, & \text{bit}_o = 0 \\ 7, & \text{bit}_o = 1 \end{cases} \tag{20}
\]

where \(\text{bit}_o\) is the original watermark bit and \(\text{bit}_m\) is the mapped bit. Obviously, the mapping process is reversible.

The final watermark information is created by combining watermark indexes with a specified fixed “length” and mapped watermark bits in order to ensure watermark synchronization. The “length” of the watermark index is the total number of digits, including the “0” value at the beginning. For example, suppose the fixed “length” is set to 3, the watermark index 16 will be expressed as “016” and suppose the fixed “length” is set to 4, the watermark index 16 will be expressed as “0016”. Take the watermark image in Figure 1 as the original watermark, the final watermark information is shown in Figure 2 when the fixed “length” is set to 3.
watermarked ratio where the $\varepsilon$ will be embedded into the coefficient ratio. The specific embedding rule is shown in Equation (21).

$$K' = F(K, 1/\varepsilon, w)$$

where $\varepsilon > 0$, $\varepsilon \in \mathbb{N}$, $1/\varepsilon$ can be regarded as the watermark embedding strength, by function $F$, the watermark information $w$ is embedded to the $\varepsilon$-th decimal place of the original DWT–CSVD transformation coefficients ratio, i.e., $K$, and $K'$ refers to the new ratio containing the watermark information. Figure 3 illustrates the embedding processing where the $\varepsilon$-th to the $(\varepsilon + p - 1)$-th decimal places of $K$ are replaced by $w$. For example, if $K = 3.141592653589793$, $\varepsilon = 4$, and $w = '0157'$, then $K' = 3.141015753589793$.

\[\begin{array}{cccc} 01 & 05 & 09 & 13 \\ 02 & 06 & 10 & 14 \\ 03 & 07 & 11 & 15 \\ 04 & 08 & 12 & 16 \end{array}\]

(a)  \[\begin{array}{cccc} 7 & 3 & 7 & 3 \\ 7 & 7 & 3 & 7 \\ 3 & 7 & 7 & 3 \\ 7 & 3 & 7 & 7 \end{array}\]

(b)  \[\begin{array}{cccc} 017 & 053 & 097 & 133 \\ 027 & 067 & 103 & 147 \\ 033 & 077 & 117 & 153 \\ 047 & 083 & 127 & 167 \end{array}\]

(c)  

Figure 2. Watermark with fused watermark indexes and bits. (a) Fixed length index; (b) mapped bits; (c) final watermark values.

2.3. Watermark Embedding Rules

The ratio of the DWT–CSVD coefficients remains constant when rotating, scaling, and translating the vector geographic data, which is the basis of the robustness of the proposed watermarking algorithms. Based on this theory, the watermark information will be embedded into the coefficient ratio. The specific embedding rule is shown in Equation (21).

\[K' = F(K, 1/\varepsilon, w)\]

2.4. Process of Watermark Embedding and Detection

2.4.1. Process of Watermark Embedding

The proposed watermark embedding process is provided in Figure 4, and the specific algorithm steps are as follows:

Step 1: the Douglas–Peucker algorithm [34] is executed on the vertexes to obtain the feature points.

Step 2: the complex sequence of the feature points is constructed using Equation (1), then the 2-level DWT coefficients with a low frequency ($L_2$) and high frequency ($H_2$) are calculated from the complex sequence.

Step 3: the CSVD algorithm is applied for $L_2$ and $H_2$, and the corresponding singular values are denoted as $SL$ and $SH$, respectively. Then, their ratios $K$ are calculated.

Step 4: generate the watermark information according to Section 2.2, and the watermark information is embedded into the ratio $K$ by Equation (21), so as to obtain the watermarked ratio $K'$.

Step 5: the watermarked singular value of the approximate component, denoted as $SL'$, is calculated according to Equation (22).

$$SL' = K' \times SL$$ (22)
Step 5: the watermarked singular value of the approximate component, denoted as $S_L'$, is calculated according to Equation (22).

$$S_L' = K \times S_L$$

Then, the watermarked approximate component, referred to as $L_2'$, is obtained using the inverse complex singular value decomposition (ICSVD) algorithm. The ICSVD algorithm is described as (23).

$$L_2' = U_{L_2} \times S_L' \times V_{H_L}^H$$

where $U_{L_2}$ and $V_{H_L}^H$ denote the unitary matrixes consisting of the left and right singular vectors of $L_2$, respectively.

Step 6: the complex sequence $C'$ containing the watermark is computed using the inverse discrete wavelet transform (IDWT), as shown in Equations (24) and (25) where $W_{Ha1}$ and $W_{Ha2}$ are the inverse matrixes of the Haar wavelet-based DWT matrix. Then, the real and imaginary parts of $C'$ are separated to obtain the watermarked feature points coordinates. Finally, the obtained feature points coordinates are used to replace the original feature points coordinates to gain the watermarked vector geographical data.

$$H' = \begin{bmatrix} L_2' \\ H_2 \end{bmatrix} \times W_{Ha1}$$

$$C' = \begin{bmatrix} L' \\ H' \end{bmatrix} \times W_{Ha2}$$

2.4.2. Watermark Extraction and Detection

The distortion of the watermarked data is commonly used to assess the quality of the digital watermarking algorithms. When the watermark is implanted in the host data, there are inaccuracies between the original vector geographic data and the watermarked data. To evaluate the accuracy of the watermarked data, the maximum error ($\text{Max.Error}$) and mean square error ($\text{MSE}$) are employed as the error metrics to gauge the variability. The $\text{Max.Error}$ and $\text{MSE}$ equations are provided below.

$$\text{Max.Error} = \max \left( \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \right) i = 1, 2, \ldots, n$$

Figure 4. Flow chart of watermark embedding algorithm.
\[ \text{MSE} = \frac{\sum_{i=1}^{n} (x_i - x'_i)^2 + (y_i - y'_i)^2}{n} \]  

(27)

where \((x_i, y_i)\) represents the original coordinates, \((x'_i, y'_i)\) represents the watermarked coordinates, and \(n\) denotes the number of vector geographic data vertexes.

Multiple attacks will be implemented on the watermarked data and then the robustness of the proposed scheme was verified by the extracted watermark images. The similarity between the extracted watermark image and the original watermark image is determined using normalized correlation (NC). The NC value ranged from 0 to 1. The higher the NC value, the closer the similarity between the two. Specifically, when the NC is equal to 1, it means that the extracted watermark is same as the original watermark. The NC equation is as follows [35]:

\[ \text{NC} = \frac{\sum_{ij} \text{XNOR}(A_{ij}, \tilde{A}_{ij})}{pq} \]  

(28)

where \(A_{ij}\) and \(\tilde{A}_{ij}\) represent the bits of the original and extracted watermark which are the size of \(p \times q\), respectively, \(\text{XNOR}(\bullet)\) denotes the operator of the exclusive NOR.

The watermark extraction process is the inverse operation of the watermark embedding, and the detail process of the watermark extraction of the proposed algorithm is shown in Figure 5. The procedure for computing the singular value ratio in the watermark extraction algorithm is identical to that in the watermark embedding algorithm, hence it is not discussed here. The steps for extracting the watermark from the singular value ratio are as follows:

1. The watermark index and mapped bit (\(b_{it_o}\)) are obtained from the ratio \(K\).
2. The real watermark bit (\(b_{it_r}\)) is calculated by Equation (29).

\[ b_{it_r} = \begin{cases} 0, & b_{it_o} < 5 \\ 1, & b_{it_o} \geq 5 \end{cases} \]  

(29)

Figure 5. Flow chart of watermark extraction algorithm.
Step 2: repeat Step 1 until all the watermark indexes and bits are extracted.

Step 3: through Steps 1 and 2, one watermark index may correspond to multiple watermark bits. A voting mechanism, expressed as Equation (30), is utilized to determine the final bit.

\[
W'_i = \begin{cases} 
0, & a < b \\
1, & a \geq b 
\end{cases}
\]  

(30)

where \(i\) denotes the watermark index, \(a\) denotes the number of bits with value 1 and index \(i\), \(b\) denotes the number of bits with value 0 and index \(i\). \(W'_i\) denotes the extracted watermark value, which is scrambled.

Step 4: the final extracted watermark image, \(W''\), is obtained by the inverse Arnold transformation of \(W'\).

Step 5: the NC value between the extracted image \(W''\) and the original watermark image \(W\) is chosen to measure the watermark extraction effect.

3. Experiments and Results

3.1. Experimental Data

To verify the reliability of the proposed watermarking scheme, three data sets, rivers, roads, and village administrative division data were selected as the experimental data (Figure 6). These vector geographic data, including polylines and polygons, were in the ESRI shapefile format, which is commonly used in geographical information systems. For convenience, the experimental data were named as "rivers", "roads", and "boundary", respectively. Table 1 provides the details of the experimental data.

![Figure 6. The original vector geographic data: (a) rivers; (b) roads; (c) boundary.](image)

Table 1. Detail of the experimental data.

| Datasets    | Types  | Compression Ratios | Feature Point Counts | Object Counts | Vertexes Counts |
|-------------|--------|--------------------|----------------------|---------------|-----------------|
| Roads       | Polyline | 70.44%             | 84,002               | 58,294        | 429,463         |
| Rivers      | Polyline | 70.12%             | 7301                 | 2703          | 36,724          |
| Boundary    | Polygon | 72.14%             | 24,144               | 20,799        | 135,185         |

The copyright graph used in the experiments is a binary image with a size of 16 × 16 bits. Figure 7 depicts the watermark image and scrambled watermark image.
3.2. Invisibility Evaluation and Parameters Settings

The embedding of the watermark information causes a certain amount of interference to the host data, influencing the accuracy of the original vector data and presenting the risk that the presence of the watermark information may be perceived by the data receivers. Invisibility as a technical feature of the vector data digital watermarking is a factor that must be considered when evaluating the performance of digital watermarking algorithms.

To evaluate the invisibility of the proposed scheme, Max.Error and MSE between the watermarked and original data were used with embedding strengths of 1/4, 1/5, 1/6, and 1/7. For comparison, two other schemes were also tested: (1) the scheme in ref. [25] which proposed a DWT-based algorithm with an embedding strength of 0.5, and (2) the scheme in ref. [36] which proposed a singular value decomposition-based algorithm with an embedding strength of $1 \times 10^{-6}$. Table 2 shows the results, indicating that the smaller the watermark strength value, the stronger the imperceptibility of the watermark. Conversely, the larger the watermark strength value, the less imperceptible the watermark is. Furthermore, the experimental results show that the proposed scheme, at a watermark embedding strength of 1/5, has better invisibility compared with the other two schemes. Figure 8 shows the overlay effect and detail enlargement between the river data with embedded watermark information and the original river data, and it can be seen that the proposed scheme causes little visual difference between the watermarked data and the original data.

![Watermark images](https://via.placeholder.com/150)

**Figure 7.** Watermark images. (a) Original watermark image; (b) scrambled watermark image.

| Datasets   | Indicators | Proposed Scheme (m) | Scheme in Ref. [25] (m) | Scheme in Ref. [36] (m) |
|------------|------------|---------------------|------------------------|------------------------|
|            |            | 1/4 | 1/5 | 1/6 | 1/7 | 0.03 | $1 \times 10^{-6}$ |
| Roads      | Max.Error  | 0.60 | 0.05 | $5.7 \times 10^{-3}$ | $3.7 \times 10^{-4}$ | 0.17 | 0.3 |
|            | MSE        | $6.1 \times 10^{-5}$ | $6.3 \times 10^{-7}$ | $6.6 \times 10^{-9}$ | $5.8 \times 10^{-11}$ | 0.01 | 0.09 |
| Rivers     | Max.Error  | 1.18 | 0.08 | $9.1 \times 10^{-3}$ | $4.8 \times 10^{-4}$ | 0.18 | 2.81 |
|            | MSE        | $1.5 \times 10^{-3}$ | $1.5 \times 10^{-5}$ | $9.9 \times 10^{-8}$ | $5.3 \times 10^{-10}$ | 0.01 | 0.94 |
| Boundary   | Max.Error  | 0.69 | 0.06 | $2.9 \times 10^{-3}$ | $2.8 \times 10^{-4}$ | 0.178 | 1.99 |
|            | MSE        | $1.3 \times 10^{-4}$ | $1.4 \times 10^{-6}$ | $1.2 \times 10^{-8}$ | $1.2 \times 10^{-10}$ | 0.01 | 0.60 |

3.3. Robustness Evaluation

To evaluate the robustness of the proposed scheme, various types of attacks were carried out at different levels on the watermarked data prior to the watermark extraction. The NC value was utilized to assess the difference between the original and extracted watermarks, and the copyright attribution was established by comparing the magnitude of the NC value and the predetermined NC threshold. In this paper, the NC threshold was set to 0.85, in other words, if the NC value of the original watermark and the extracted watermark was greater than 0.85, the watermark was considered to be successfully extracted...
from the vector geographic data, otherwise it failed to do so. For comparison, the robustness of the schemes in ref. [25,36] were also evaluated.

3.3.1. RST Attacks

RST refers to rotation, scaling, and translation, which are common geometric transformations in vector geographic data operations. To verify the robustness against the RST attacks, the vector geographic data to be inspected were rotated, translated, and scaled at different degrees and the corresponding extracted watermark images with NC values were recorded. The results are shown in Table 3.

Table 3. The results of RST attacks.

| Datasets | Rotation (°) | Translation (m) | Scaling |
|----------|-------------|-----------------|---------|
|          | 10          | 60              | 150     | 10    | 1000 | 5000 | 0.5 | 0.95 | 1.05 | 2 |
| Rivers   | NC = 1      | NC = 1          | NC = 1  | NC = 1 | NC = 1 | NC = 1 | NC = 0.49 | NC = 0.91 | NC = 0.90 | NC = 0.56 |
| Roads    | NC = 1      | NC = 1          | NC = 1  | NC = 1 | NC = 1 | NC = 1 | NC = 0.44 | NC = 0.91 | NC = 0.92 | NC = 0.49 |
| Boundary | NC = 1      | NC = 1          | NC = 1  | NC = 1 | NC = 1 | NC = 1 | NC = 0.51 | NC = 0.89 | NC = 0.92 | NC = 0.48 |

As can be observed, the NC values between the extracted watermark and the original watermark are invariably one for both translational and rotational attacks at different scales, which implies that the proposed scheme is fully resistant to both translation and rotation. When scaling the watermarked data by 0.95 and 1.05, the NC values of all the experimental data are higher than the threshold, 0.85, and the copyright can be visually authenticated by the extracted watermark images. All the NC values are lower than the threshold when the scaling factors are 0.5 and 2. Because the proposed scheme, a blind watermarking scheme, cannot predict the scale degree of the experimental data to be evaluated, many of the feature points acquired by the Douglas–Peucker algorithm could not correlate to the feature points utilized for embedding the watermark as the scale degree increased.
Therefore, the proposed scheme could identify a watermark under a minor scaling attack but cannot extract watermark information as the degree of scaling increased.

3.3.2. Simplification Attacks

Simplification is a common operation in geographical analysis. After simplification, the remaining points in the vector geographic data are called feature points, and the deleted points are called non-feature points. A good algorithm should resist attacks from simplification. To test the resistance of the proposed scheme to simplification attacks, the feature points were extracted and the non-feature points were deleted, with simplification rates ranging from 0% to 90%, by the Douglas–Peucker algorithm. Figure 9 depicts the experimental outcomes. The NC values of all the experimental data are invariably one when the simplification rate ranges from 0% to 70%. The outcomes result from that when the simplification rate is lower than the rate at which the original vector geographical data is simplified to extract the feature points, and the proposed scheme could accurately obtain the watermarked points so as to extract the watermark information without errors. When the simplification rate is higher than 70%, where only a few feature points are preserved, the NC values decrease below 0.75. Actually, if a high-intensity simplification attack is performed, the contour of the vector geographic data tends to be straight, and its application value is lost. Therefore, the proposed scheme is particularly resistant to simplification attacks.

![Figure 9. The results of simplification attacks.](image)

3.3.3. Vertex Addition and Deletion Attacks

Vertex addition attacks interfere with the watermark information by introducing extra points into the vector geographic data. To verify the robustness of the scheme to attacks from the vertex addition, 10% to 200% of points were added into the watermarked data. The results are recorded in Figure 10. As can be seen, the NC values remain constant at one, although the adding ratio addition ratio reaches 200%. The experimental findings can be explained by the new points being excluded as non-feature points, which has no impact on the watermark extraction. Therefore, the proposed scheme can totally resist attacks from the vertex addition.
Vertex deletion attacks work by randomly deleting a certain percentage of vertexes, including the feature points, which are not removed by a simplification of the vector geographic data. Figure 11 illustrates the results of the deletion attacks with a deletion ratio ranging from 0% to 90%. As can be seen from the findings, the NC values are greater than 0.85 when the deletion rate is less than or equal to 30%. The NC values of the boundary and roads are higher than 0.85 when 30% to 40% of vertexes are eliminated, yet the watermark in the rivers cannot be detected. When the deletion rate reaches 50%, the proposed scheme cannot correctly extract the watermark of any of the experimental data. As the deletion ratio climbs, more feature points carrying watermark information are removed, resulting in a watermark extraction failure. However, an excessively high deletion ratio would have a negative impact on the quality of the vector geographic data, and therefore this approach is robust within a fair range of deletion point attacks.
3.3.4. Cropping Attacks

To verify the robustness of the proposed scheme against map cropping, the watermarked data were cropped by 1/8, 1/4, 1/2, and 3/4, and the proposed scheme were compared with the schemes in refs. [25,36]. Table 4 shows the results of the NC values for the three schemes with different levels of cropping attacks. For rivers data, the schemes in refs. [26,36] fail to extract the watermark information even when the cropping area is 1/8, whereas the proposed scheme could resist the attack with a cutting area of 3/4. For the roads and boundary, the NC values of all the three schemes are one, when cropping off 1/8 and 1/4 of the area. The difference is that as the area of the cropped area increases, the proposed scheme is still able to extract the watermark information accurately, even when 3/4 of the area are cropped off, with NC values higher than 0.9, but the schemes in refs. [25,36] could not. Irregular cropping attacks have a high probability of changing the vertexes ordering, on which a watermark extraction of the schemes in refs. [25,36] rely, resulting in the failure of the watermark extraction. The proposed scheme successfully extracts the watermark bits and indexes by fusing the watermark index and bit values to create a new watermark sequence, guaranteeing a high accuracy for the detection of attacks from cropping.

Table 4. The results of cropping attacks.

| Clip Ratios | Datasets | Scheme in Ref. [25] | Scheme in Ref. [36] | Proposed Scheme |
|-------------|----------|--------------------|--------------------|-----------------|
| 1/8         | rivers   | NC = 0.52          | NC = 0.54          | NC = 1          |
|             | roads    | NC = 1             | NC = 1             | NC = 1          |
|             | boundary | NC = 1             | NC = 1             | NC = 1          |
| 1/4         | rivers   | NC = 0.54          | NC = 0.50          | NC = 1          |
|             | roads    | NC = 1             | NC = 1             | NC = 1          |
|             | boundary | NC = 1             | NC = 1             | NC = 1          |
| 1/2         | rivers   | NC = 0.48          | NC = 0.46          | NC = 1          |
|             | roads    | NC = 0.44          | NC = 0.48          | NC = 1          |
|             | boundary | NC = 0.49          | NC = 0.5            | NC = 1         |
| 3/4         | rivers   | NC = 0.52          | NC = 0.44          | NC = 0.91       |
|             | roads    | NC = 0.48          | NC = 0.51          | NC = 1          |
|             | boundary | NC = 0.55          | NC = 0.52          | NC = 1          |
3.3.5. Combination Attacks

A combination attack is a comprehensive attack that employs multiple attacks on the experimental data at the same time, making it more difficult to defend because of the unknown methods of the attacks and the fact that they are simultaneous. To examine the robustness of the proposed scheme against a combination attack, a 30\% simplification attack \( (d_1) \), 1/2 clipping attack \( (d_2) \), and 60° rotation attack \( (d_3) \) were chosen as the combination attack methods in this experiment. Table 5 displays the experimental results, where × means extracting watermark images unsuccessfully and √ denotes extracting watermark images correctly. From Table 5, it can be seen that the proposed scheme could extract the watermark information accurately under a random combination of \( d_1, d_2, \) and \( d_3 \) attacks, but neither the scheme in ref. [25] nor the scheme in ref. [36] could extract the watermark under the same combination attacks. It can be concluded that the proposed scheme performs well in terms of combination attacks.

| Datasets | Scheme in Ref. [25] | Scheme in Ref. [36] | Proposed |
|----------|---------------------|---------------------|----------|
|          | \( d_1, d_2 \) | \( d_2, d_3 \) | \( d_1, d_3 \) | \( d_1, d_2, d_3 \) | \( d_1, d_2 \) | \( d_2, d_3 \) | \( d_1, d_3 \) | \( d_1, d_2, d_3 \) |
| Rivers   | 0.422              | 0.496              | 0.457    | 0.410    | 0.539    | 0.555    | 0.457    | 0.492    | 1        | 1        | 1    | 1 |
|          | ×                  | ×                  | ×        | ×        | ×        | ×        | ×        | ×        | √        | √        | √    | √ |
| Roads    | 0.547              | 0.457              | 0.449    | 0.508    | 0.632    | 0.590    | 0.423    | 0.445    | 1        | 1        | 1    | 1 |
|          | ×                  | ×                  | ×        | ×        | ×        | ×        | ×        | ×        | √        | √        | √    | √ |
| Boundary | 0.484              | 0.496              | 0.449    | 0.500    | 0.671    | 0.613    | 0.574    | 0.508    | 1        | 1        | 1    | 1 |
|          | ×                  | ×                  | ×        | ×        | ×        | ×        | ×        | ×        | √        | √        | √    | √ |

3.4. Comparative Analysis

In this paper, we propose a new robust digital watermarking scheme based on the ratio invariance of DWT–CSVD coefficients. In this section, the algorithm performance was evaluated by being compared with prior schemes. In Table 6, √ indicates that the scheme is robust against an attack, and × indicates that the scheme is not robust against that attack. Table 6 shows that the scheme in ref. [25] is robust to addition and deletion attacks, mainly due to the facts that the watermark is embedded in the DWT coefficients with a low frequency, but the scheme is not robust to common geometric attacks. The scheme in ref. [36] hides the watermark information in the eigenvectors of the SVD, which can better extract the watermark from the geometric attacks; however, the scheme cannot resist cropping, addition, and deletion attacks. Neither the scheme in ref. [25] nor the scheme in ref. [36] can effectively extract the watermark information from the combined attack. The proposed scheme can extract the watermark information more completely in both the above attacks and their combined attacks. In comparison, it can be concluded that the proposed scheme has a stronger robustness.

| Attack Types | Datasets | Proposed | Scheme in Ref. [25] | Scheme in Ref. [36] |
|--------------|----------|----------|---------------------|---------------------|
| Rotation     | Rivers   | √        | ×                   | √                   |
|              | Roads    | √        | ×                   | √                   |
|              | Boundary | √        | ×                   | √                   |
| Translation  | Rivers   | √        | ×                   | √                   |
|              | Roads    | √        | ×                   | √                   |
|              | Boundary | √        | ×                   | √                   |
| Scaling      | Rivers   | ×        | ×                   | √                   |
|              | Roads    | ×        | ×                   | √                   |
|              | Boundary | ×        | ×                   | √                   |
Table 6. Cont.

| Attack Types | Datasets | Proposed | Scheme in Ref. [25] | Scheme in Ref. [36] |
|--------------|----------|----------|---------------------|---------------------|
| Simplification | Rivers | ✓ | × | ✓ |
| | Roads | ✓ | × | ✓ |
| | Boundary | ✓ | × | ✓ |
| Clipping | Rivers | ✓ | × | × |
| | Roads | ✓ | × | × |
| | Boundary | ✓ | × | × |
| Deleting | Rivers | ✓ | ✓ | × |
| | Roads | ✓ | ✓ | × |
| | Boundary | ✓ | ✓ | × |
| Adding | Rivers | ✓ | ✓ | × |
| | Roads | ✓ | ✓ | × |
| | Boundary | ✓ | ✓ | × |
| Combination | Rivers | ✓ | × | × |
| | Roads | ✓ | × | × |
| | Boundary | ✓ | × | × |

4. Conclusions

In this study, a novel watermarking scheme is designed based on the combination with DWT and CSVD for vector geographic data. The proposed scheme puts forward the ratio of the DWT–CSVD coefficients as the watermark embedding domain, which is invariant under geometric attacks. In the watermark embedding stage, a new watermark sequence formed by fusing the watermark index and bit values of the copyright image is embedded to the fractional bit of the ratio value to hide the watermark information. The experimental results show that the proposed scheme has a good invisibility and can extract reliable watermark images under different degrees of translation, rotation, compression, cropping, vertex deletion, and a vertex addition attack. Subjected to a random combination of the above attacks, the proposed scheme can also extract the watermark information with a high probability, demonstrating an excellent robustness. The scheme shows to be credible for the copyright protection of the vector geographic data as tested.

Author Contributions: Chengyi Qu and Xu Xi conceived and designed the experiments; Chengyi Qu wrote the original draft; Jinglong Du and Xu Xi reviewed and edited the paper; Tong Wu contributed framing, ideas, context, and wordsmithing. All authors have read and agreed to the published version of the manuscript.

Funding: This work is jointly funded by the National Natural Science Foundation of China (grant no. 42101420) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant no. SJCX22_1560).

Data Availability Statement: The data used in this study can be accessible by request to the corresponding author.

Acknowledgments: The authors gratefully acknowledge the anonymous referees for their insightful criticism, which definitely enhanced the work.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Qiu, Y.; Gu, H.; Sun, J. High-payload reversible watermarking scheme of vector maps. Multimed. Tools Appl. 2018, 77, 6385–6403. [CrossRef]
2. Zhang, L.; Yan, H.; Qi, J.; Zhang, Y. A Blind Watermarking Algorithm for Copyright Protection of Vector Geospatial Data Under Controllable Errors Based on DFT. Geomat. Inf. Sci. Wuhan Univ. 2015, 40, 990–994. [CrossRef]
3. Peng, F.; Lin, Z.-X.; Zhang, X.; Long, M. Reversible data hiding in encrypted 2D vector graphics based on reversible mapping model for real numbers. IEEE Trans. Inf. Forensics Secur. 2019, 14, 2400–2411. [CrossRef]
32. Arnold, V.I.; Avez, A. Ergodic Problems of Classical Mechanics; Benjamin: New York, NY, USA, 1968.

33. Shang, Z.; Ren, H.; Zhang, J. A block location scrambling algorithm of digital image based on Arnold transformation. In Proceedings of the 2008 the 9th International Conference for Young Computer Scientists, Zhangjiajie, China, 18–21 November 2008; pp. 2942–2947.

34. Douglas, D.H.; Peucker, T.K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartogr. Int. J. Geogr. Inf. Geovisualiz. 1973, 10, 112–122. [CrossRef]

35. Kandi, H.; Mishra, D.; Gorthi, S.R.S. Exploring the learning capabilities of convolutional neural networks for robust image watermarking. Comput. Secur. 2017, 65, 247–268. [CrossRef]

36. Gaata, M.T. Robust watermarking scheme for GIS vector maps. Ibn AL-Haitham J. Pure Appl. Sci. 2018, 31, 277–284. [CrossRef]