Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 29 (2012) 1331 – 1337

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

Image Watermark Similarity Calculation of GIS Vector Data

LI Sha-Sha, ZHOU Wei, LI An-Bo *

Key Laboratory of Virtual Geographical Environment, Ministry of Education Nanjing Normal University Nanjing 210046, China

Abstract

In recent years, digital watermarking is playing an increasingly important role in the geographical information copyright protection. However, usually there is a high undetected rate in image watermarking authentication of GIS (Geographic Information System) vector data, because in watermark similarity calculation there are some problems, for example, the extracted watermark displacement and length changing, etc. This paper based on these focused research: the extracted watermark's position correction methods on the basis of the original watermark; the non-equal length strings similarity calculation methods, formed an image watermark similarity calculation of GIS vector data. This method effectively solved the problem of the high undetected rate in watermark certification, and to some extent improved the vector data's copyright authentication theoretical basis and technology.

Keywords: GIS; Image watermark; Similarity calculation; Vector data;Watermarking Authentication

1. Introduction

In recent years, digital watermarking, one of represented copyright mark technology, is playing an increasingly important role in geographical information copyright protection. At present, both at home and abroad, the research on how to embed a watermark in vector data is relatively mature. GIS (Geographic Information System) vector data, compared to the raster data which attributes are clear and position is implied, has its own characteristics: (1) Position clear, attributes implied; (2) various express, complex storage structure; (3) no fixed storage order; (4) complicated structure, it contains not only the geometric information, attribute information, as well as topology information; (5) with high precision, less

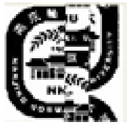
* Corresponding author. Tel.:+86-25-13645161205 ; fax:+86-25-85898270
E-mail address: mrlab@126.com.

doi:10.1016/j.proeng.2012.01.136

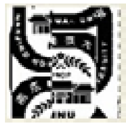
redundancy; (6) with a hierarchical organization of features. In GIS vector data's copyright protection, these above characteristics lead to two difficulties in watermark authentication.

On the one hand, according to the GIS vector data's characteristic of un-fixed storage order, delete points and add points could easily undermine the order of points, thus have a great influence on the location of the embedded watermark information. When extracted watermark, because of the blind test method is usually used, can not correctly determine the location of the watermark information, lead to the phenomenon of the extracted watermark produced displacement and its length has changed etc (Fig.1). So, it is bound difficult to correctly and reasonably calculate the similarity, and result a high undetected rate.

(a)



(b)



(c)

Fig.1. Comparison graph of extracted watermark before and after attack (a) Original watermark ;(b) Extracted watermark after 1% points be deleted; (c) Extracted watermark after 1% points be increased

In order to solve the problem of displacement of the extracted image watermark, to make the extracted watermark image pixel to the right place and enhance the reliability of watermark authentication results. Zhu Chang-Qing (2006), Wang Zhong-Jun(2008) presented recorded the location of the point when embedded the watermark information in the coordinate points, then by comparing the difference between the watermarked data and the original data to extract the watermark information[1-2]. However, this algorithm is a non-blind watermarking algorithm, and it has enhanced data redundancy and not conducive to the stability of copyright protection. There is no any stable and simple algorithm to correct the extracted image watermark's position. However, image watermark's position correction has a direct impact on similarity calculation and digital watermarking authentication.

On the other hand, the watermark similarity refers to the similarity of the original watermark and the extracted watermark. It is one of the key steps in watermarking authentication. In recent yeas, scholars both at home and abroad proposed several methods about the measure of watermark similarity, the most commonly used metrics are Normalized correlation (NC) [3-4], Normalized Hamming similarity (NHS) [5-6], Bit Error Rate (BER) [7-8] , Bit Correct Ratio (BCR) [9], Normalized Correlation Index (NCI) [10], these metrics are all required to the extracted watermark and original watermark with equal length, some of these indicators have no uniform detection threshold. During the watermark authentication process for GIS vector data, if direct use these indicators to measure the similarity of the watermark, will inevitably lead to a higher undetected rate. So this method of not stable enough watermark authentication, which applies to images, audio and other traditional areas seriously affecting the application of watermark authentication technology.

Therefore, it is necessary to propose a more correct and reasonable method of watermark similarity calculation. This paper based on these focused research: the extracted watermark's position correction methods on the basis of the original watermark; the non-equal length strings similarity calculation methods, formed a more correct and reasonable method of image watermark similarity calculation on GIS vector data, to reduce the undetected rate in watermarking authentication, and to promote GIS vector data watermarking research and application.

2. Image Watermarking Authentication of GIS Vector Data

Digital watermarking authentication refers to determine whether the carrier data contains watermark information by a certain algorithm and key. The first step of watermarking authentication usually is watermark extraction, and then the watermark decision. Watermarking decision is usually by comparing the threshold with the similarity that between the extracted watermark and the original watermark, if the similarity above a specified threshold can be determined to contain designated watermark in the test data, otherwise, determined the watermark-free [11].

However, in allusion to the extracted watermark's disordered and length change problems, which exist in the vector data's digital watermark, the traditional watermarking authentication process in image and audio etc. is difficult to directly apply. So a watermark position correction part must be increased on the basis of traditional watermarking authentication process, and improve the method of non-equal length string's similarity calculation, to form a complete GIS vector data watermark authentication process (Fig.2). Specific process is as follows: first, using the key through the watermark extraction algorithm to extract the image from the watermarked works, in the process may also be involved the original vector data; Second, correct the position of the extracted image watermark reference to the original watermark; third, calculate the similarity between the original watermark and the corrected watermark; Finally, compare the similarity with the specified threshold, if the similarity above the threshold, can be determined to contain designated watermark in the test data, otherwise, determined the watermark-free.

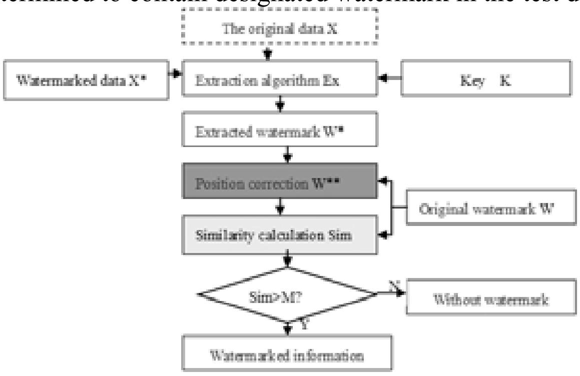


Fig.2. Watermarking authentication process of GIS vector data

3. Watermark Similarity Algorithm of GIS Vector Data

3.1. The extracted watermark position correction on the basis of referencing original watermark

Since the extracted watermark may emerge confusion, how to restore each pixel is the key problem. Therefore, with reference to the original watermark, we take the following optimization algorithm to rapidly recover each pixel location of the extracted watermark. Specific implementations as follows: First, initializes the sequence W'' as the corrected watermark and calculate the two watermarks' length difference d; Second, take one character from the shorter watermark to match the longer-watermark's *d+1* characters. Third, if the match is successful, save it to W'', else mark the position in W'' and begin to loop match in watermarks' next position. The algorithm process (Fig.3) as follows:

Step 1: the original watermark W_O by embedding algorithm embed the watermark W_O into the carrier data D, and then get D' by attacking D, then extracted watermark W_E from D', denoted by $Loc = 0$;

Step 2: compare the length of W_O and W_E, let the longer watermark be W, and length of L_W, let the shorter watermark be W', and length of;

Step 3:the length difference d of W and W' is gotten from the formula (1), initialization sequence W'' ,that is used to store the corrected watermark, length of L_W , $W''(i) = '*'$,where '*' represents the modified position, i=0,1,2,…. L_W -1;

$$d = |L_W - L_{W'}| \quad (1)$$

Step 4: i) read a character $W'(i)$ from W' , i=0, 1, 2… $L_{W'}$ -1;

ii) take $d+1$ characters from the Loc one in W , while $L_W < Loc + d + 1$, take L_W - Loc characters, denoted by $W(dis)$;

iii) match $W'(i)$ and $W(dis)$, if match is successful, then record the starting position of match, denoted by Pos , go to step iv), else, $i = i + 1$, $Loc = Loc + 1$, go to step v);

iv) to assign a value to W'' , that is $W''(Loc + Pos - 1) = W'(i)$, $i = i + 1$, $Loc = Loc + Pos$;

v) cycle through Steps i) through iv) until the original watermark W or the extracted watermark W' is finished reading, then save the corrected watermark W'' .

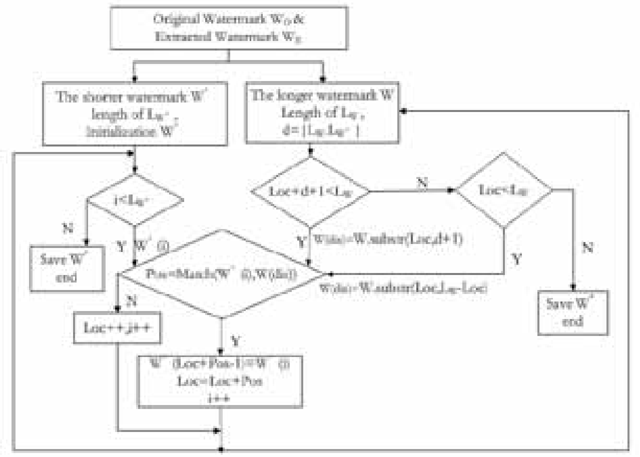


Fig.3 Image Watermark position correction of GIS vector data

3.2. Watermark similarity calculation

The watermark similarity often expresses with the same number of original watermark and extracted watermark[12], its essence is calculate the degree of similarity of two strings, so its algorithm is close relation to the strings' approximate search. Since in the watermark similarity calculation, taking into account the length of the extracted watermark may change, this paper based on the research of non-equal length string approximate matching, decided to use E. Ukkonen (1985) proposed time-based optimization of the dynamic programming algorithm[13], as the image watermark similarity calculation method. The main implementation process as: first, according to the original watermark and the corrected watermark calculate the edit distance matrix M (Equation 2), and then cycle M to get the minimum edit distance, to calculate the watermark similarity *Sim*(Equation 3).

$$M_{i,0} \leftarrow i, M_{j,0} \leftarrow 0$$
$$M_{i,j} \leftarrow \begin{cases} M_{i-1,j-1} & if(W''(i) = W_O(j)) \\ 1 + \min(M_{i-1,j-1}, M_{i-1,j}, M_{i,j-1}) & if(W''(i) \neq W_O(j)) \end{cases} \quad (2)$$

$$Sim = (1 - K/L_{W_O}) * 100\% \quad (3)$$

Where W_o is the original watermark, W'' is the watermark after correction, M is the edit distance matrix; K represents the minimum edit distance, L_{WO} is the length of the original watermark; $K/\ L_{WO}$ indicates the error level, it can measure W'' 's ability of tolerating the error.

Step 1: make the original watermark W_o for the match string, make the corrected watermark W'' for the pattern string, according to the formula (2) to generate M, where $i = 1, 2, ...L_{Wo}$, $j = 1, 2, ...L_{W_o}$;

Step 2: set $K = M_{LWO,0}$, K represents the minimum edit number while W'' convert to W_o , when K=0, it means exactly the same as W'' and W_o ;

Step 3: read $M_{LWO,j}$, determine the size of K and $M_{LWO,j}$, when $K > M_{LWO,j}$, $K = M_{LWO,j}$, j=j+1;

Step 4: if $j \le L_{W_o}$, cycle step 3, otherwise go to step 5;

Step 5: Calculate the watermark similarity *Sim*.

4. Experiments Analysis

Experiment data is ESRI shapefile of JangSu (Fig.4), original watermark is the logo of Nanjing Normal University (Fig.1 (a)), embedding algorithm is the least significant bits and embedding domain is DWT frequency domain.

(a) (b) (c) (d)

Fig.4. Comparison graph of vector data before and after embedding watermark (a) Original data; (b) Original data after sectional enlarging; (c) Watermarked data; (d) Watermarked data after sectional enlarging.

4.1. Image watermark correction

Table 1 Data editing attacks experimental

Number of points	Attacks		
	Insert points	Delete points	Move points
1%			
5%			
10%			

In this paper we used following attacks to attack the watermarked vector data: inserting, deleting, moving points. In this experiment we used program to carry out simulation, random insert, delete or move the point in the watermarked vector data, the experiment layers contained 41737 points in total, the results shown in Table 1. As can be seen from the table, insert and delete point had greater impact on the extracted watermark. Because of inserting and deleting points can change the storage location of the vector data elements, the extracted watermark information maybe displacement. The watermark correction algorithm proposed in this paper can correct the pixel of disordered watermark into the correct position, which will largely improve the watermark similarity, and also improve the watermarking authentication efficiency.

4.2. Watermark similarity calculation

Currently, for the watermark similarity measure has not absolutely conclusive, from the theoretical analysis, according to the changed level of the watermarked data, you can get the appropriate degree of watermark similarity. For example, randomly delete 5% information of the watermarked data, the watermark similarity should be about 95%. However, inserting information into the watermarked data will not change the watermark similarity, because the watermark information that embedded in the data has not been modified.

Based on the above analysis, this paper used Levenshtein distance [14] and non-equal length strings similarity algorithm to compare analysis: attacking the embedded watermark data by inserting, deleting or moving points, and then calculating the similarity between the original watermark and the extracted watermark. The idea of Levenshtein Distance algorithm is to start comparison from the left side of two strings, then record the similarity of the substring that have been compared, accordingly to further get the similarity of the next character position. This algorithm's matched result only from the first character, and is suitable for approximate matching of the equal length strings.

The results are shown in Table 2. As can be seen from the table, the similarity of non-equal length similarity algorithm, in the same strength of inserting and deleting points, are higher than the Levenshtein Distance algorithm's. It is mainly because of the algorithm presented by this paper, in allusion to the non-equal length watermarks, can dynamic finding the best starting matched position. For this paper's algorithm have the higher watermark similarity after inserting points attack, it is because the extracted watermark's location have been corrected before the watermark similarity calculation. For the similarity after attack by moving points, the results of two algorithms are the same, mainly because of the attack did not change the length of the extracted watermark.

Table 2 Comparison of different watermark similarity algorithms

Attacks / Algorithms	Insert points			Delete points			Move points		
	5%	10%	15%	5%	10%	15%	5%	10%	15%
This paper 's algorithm	0.987	0.969	0.9571	0.9501	0.9001	0.8501	0.9564	0.9159	0.8751
Levenshtein Distance	0.925	0.8562	0.7834	0.931	0.8456	0.7764	0.9564	0.9159	0.8751

5. Conclusion

Watermark similarity calculation is a key element of watermark authentication. However, the watermark similarity calculation of GIS vector data still has many irrational factors, for example, the disordered extracted watermark and the different length of watermarks. The method this paper presented can correct the disordered extracted watermark on the basis of reference the original watermark and calculate the similarity of non-equal length watermarks. Experiments show that: the algorithm can make the image watermark similarity calculation of GIS vector data more reasonable, and reduce the undetected rate of the watermark authentication. In addition, the algorithm be slightly modified can also be used for the text form watermark similarity calculation.

Acknowledgements

Support by the HI-TECH research and development program of china (863 Program) –research on key technique for copyright protection of GIS vector data (NO. 2006AA12Z222), and the project funded by the priority academic program development of Jiangsu higher education institutions.

References

[1] Zhu Chang-qing, Yang Cheng-song, Li Zhong-yuan. An Anti-compression Watermarking Algorithm for Vector Data, *Journal of Zhengzhou Institute of Surveying and Mapping*, China; 2006, p.281-283.

[2] Wang Zhongjun, Wang Yuhai, Wang Hao. A robust water marking algorithm for vector digital maps, *Science of surveying and maping*; China; 2008, p.148-150.

[3] Li Yuanyuan, Xu Luping. Copyright protection of the vector map using the digital watermark. *Journal of XIDIAN university*, China; 2004, P719-723.

[4] Liu Fanghong, Zhang Guicang. Blind Watermarking Algorithm Based on Contourlet Transform, *China science and technology information*, China; 2011. p.135-137.

[5]Chu Shu-Chuan, Roddick John F., Lu Zhe-Ming, Pan Jeng-Shyang. A Digital Image Watermarking Method Based on Labeled Bisecting Clustering Algorithm,*IEICE Transactions*; 2004, p. 282-285

[6]Yan Haowen, Li Jonathan,Wen Hong.A key points-based blind watermarking approach for vector geo-spatial data,Computers, , Environment and Urban Systems; 2010.

[7]Kitamura I.;Kanai S.;Kishinami T.Copyright protection of vector map using digital watermarking method based on discrete Fourier transform;2001,p.1191-1193

[8] B.Gunsel, Y.Ulker, S.Kirbiz. A statistical framework for audio watermark detection and decoding. *MULTIMEDIA CONTENT REPRESENTATION, CLASSIFICATION AND SECURITY*[J]; 2006, p.241-248

[9] Tong Ming,Yan Tao, Ji Hongbing. Strong anti-robust watermarking algorithm, *Journal of XIDIAN university*, China; 2009, P. 22-27.

[10] Zhong Shangping, Gao Qingshi. The Feasibility Analysis of Normalized-correlation-based Vector Maps Watermarking Detection Algorithm and the Improved Watermarking Algorithm, *Journal of Image and Graphics*, Chian; 2006, p.401-409.

[11] Sun Shenghe, Lu Zheming, Niu Xiamu. Digital Watermarking Technology and Application, 2004.

[12] He Bing. An Improved Algorithm of Similarity Formular in Digital Watermarking. *Computer & Digital Engineering*, China; 2010, p.95-98.

[13]Navarro Gonzalo, Raffinot Mathieu.Flexible Pattern Matching in String:Practial On-Line Search Algorithms for Texts and Biological Sequences; 2002.

[14] V. I. Levenshtein. Binary Codes Capable of Correcting Spurious Insertions and Deletions of Ones. Problems of Information Transmission, 1965, p.8-17.