Watermarking Algorithm Based on Invisible Unicode Characters for GIS Attribute Database

ZHU Chang-qing, TONG De-yu, REN Na

Institution:Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, Jiangsu Province, Nanjing, 210023

Email: zcq88@263.net

1. Introduction

Geographic data consists of two different data types which are spatial data and attribute data. The attribute data is an essential feature of geographic objects, containing variables and significances in GIS (Tang GA, 2007). The attribute data is characterized by massive, non-numerical, diverse forms and it is often stored in relational database. The technology of watermarking algorithm is potential of being applied in database to protect its copyright, and many scholars are doing researching work on it. After the watermark on numeric database was first proposed (Rakesh 2002) and improved later (Zhu 2006, Gauray 2009), some new algorithms for non-numeric database have been proposed. Such new methods include replacing similar words (Radu 2004, Agusti 2006), improved by statistics (Dong 2008), and inserting spaces (Ali 2008). But those methods will affect the attribute data themselves, since they have been modified the data integrity, which may lead to uselessness of attribute data. From the point view of encoding, Unicode characters are used by certain researches in information hiding (Huang 2011) and in watermarking (Li 2012).

2. The Generation of Watermark Information

2.1 The advantages of the invisible Unicode characters

The typical data form of the non-numerical database is the character string, which can be implemented with the Unicode encoding method. There exist some invisible characters for the Unicode encoding method, and those characters have the features of zero width and invisibility. By using such invisible Unicode characters for watermark embedding, the following advantages can be offered: (1) The performance of the database is not affected. (2) The generated watermark information is invisible. (3) More watermark information can be embedded.

2.2 Generating watermark information based on mapping index

The meaningless watermark information has the better statistical features, randomness and anti-attacking ability. This paper proposes a novel meaningless watermark information generating mechanism, and its working principle is given as follows (Ren 2011):

(1) A seed number is produced by using the random number generator at first, and this number works as the initial value for generating the pseudo-random binary sequence in the next step.
(2) With the initial value provided by the seed and the pseudo-random binary sequence generator, an N-length pseudo-random binary sequence *W* is generated as:

$$W = G(Seed) = \{w_i, i = 0,1,2,\ldots,N-1\}, w_i \in \{0,1\}$$

(3) The relationship between the pseudo-random binary sequence *W* and the copyright information is built.

It should be mentioned the length of watermarking information is very important for watermark embedding and detecting. The special features of the operation for database, such as deleting columns, deleting rows and the view generation should be considered during setting the length for watermark information. If the length of watermark information can be divisible by quite a few integrals, the regularity will be induced in the watermark embedding, so that the anti-attacking ability of watermark will be affected. Therefore, the prime number is proposed to be the length of the watermark information in this paper. It can be concluded that no matter how many columns exist in the attribute table of database, the watermark information can be distributed randomly.

3. Watermarking algorithm based on the invisible characters

There are two key issues in the watermarking algorithm: the watermarking value and the embedding position of watermark information. To address these two issues, a novel watermark embedding procedure is proposed: Encode the invisible Unicode characters firstly. Each character corresponds to a certain value between 0 and 14. Three of those characters are used to present the value of watermark and the embedding position of watermark. The first two characters present the embedding position of the watermark, and the last character corresponds to the value of watermark.

3.1 Embedding Watermark

The procedure of embedding watermark is proposed as follows:
(1) The binary sequence W is obtained by using the mapping index. The information of watermark value and embedding position in the *W* sequence are analyzed. According to the aforementioned encoding principle for invisible characters, an array, namely *WM*[0~198] is generated, where 199 elements are contained.
(2) By connecting to the database, its attributes such as the table name, the field type, the prime key and the external key are obtained. It is noted that the prime key and the external key do not embed watermark so as to maintain the integrity.

(3) By repeatedly reading the data in the GIS attribute database, the array *WM*[0~198] with watermark information can be embedded effectively.

3.2 Detecting Watermark

The procedure of detecting watermark is proposed as follows:
(1) Similar to the procedure of embedding watermark, by connecting to the database, its attributes such as the table name, the field type, the prime key and the external key are obtained. Then, construct the table for containing the watermarking information.
(2) Repeatedly read the data in the table, which is possible to be embedded watermarks. Judge whether the pre-specified invisible Unicode characters are contained. If contained, the watermark is detected within the corresponding data. Then, the watermark is extracted from the data and recorded according to the encoding principle.
(3) According to the majority principle, the final watermark information is determined. By comparing with the watermarking information obtained with mapping index, the correlation coefficients of statistics are calculated. If the correlation coefficient is detected higher than the presetting value, the detection is successful, and the copyright information is determined.

4. Experiments and Analysis

In this paper, the invisible Unicode characters are adopted to present the watermark information. Thus, the invisibility of the watermark is maintained after its being embedded into the GIS attribute database. The unlawful users will not be aware of the watermark while browsing or using the database. So, the watermarking information can be hidden better. By controlling the embedding position and embedding method properly, the operation of the GIS attribute database after being embedded watermark will not be affected. Those applications of database such as the matching of Chinese character strings, SQL query and processing of external program will keep working properly. The practical use of the GIS attribute database is maintained. The experiments have been carried out to verify that the proposed watermarking algorithm can well resist different external attacks in the database such as the data increment, the column deleting, the row deleting, the view generation, the structure modification, the sorting, etc. It has been proven that the improved robustness is provided due to the better distributing property and redundancy of the proposed watermark algorithm.

5. Conclusion

In this paper, a novel watermarking algorithm is proposed for the GIS attribute database based on the invisible Unicode characters. The propose method implements the embedding and detecting of watermark without affecting the operation of GIS attribute database. It has been verified that the proposed algorithm has the good robustness and invisibility. It can well resist the attacks in the database such as the

data increment, the data deleting, the modification of table, view generation, etc. So, it has the prosperous future in application.

6. Acknowledgements

The work was supported by the National Natural Science Foundation of China (Grant No. 41071245), the University Science Research Project of Jiangsu Province (Grant No. 12KJB420002),and Open Foundation of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing of Wuhan University (Grant No. 12I01)

7. References

Tang GW,Liu XJ, Lv GN, Sheng YH, Wang C and Zhang T, 2007, *The Tutorial of Geographic Information System*. Higher Education Press, Beijing, China.

Rakesh A and Jerry K, 2002, Watermarking Relational Databases, *Proceedings of the 28th VLDB Conference*, Hong Kong, China, 155-168.

Zhu Q, Yang Y, Le JJ and Luo YS, 2006, Watermark Based Copyright Protection of Outsourced Database, *Proceedings of the 10th International Database Engineering and Applications Symposium*, Delhi, India, 301-308.

Gaurav G and Josef P, 2009, Database Relation Watermarking Resilient against Secondary Watermarking Attack, *Proceedings of the 5th International Conference on Information Systems Security*, Kolkata, India, 222-236.

Radu S, 2004, Proving ownership over categorical data, *Proceedings of the 20th International Conference on Data Engineering*, Boston, USA, 584-595.

Agusti S and Josep DF, 2006, Watermarking Non-numerical Databases,*3rd International Conference on Modeling Decisions for Artificial Intelligence*, Tarragona, SPAIN, 239-250.

Dong XM, Tian YP, Li XH and Yu G, 2008, Study of Watermarking Nonnumeric Data in Relational Databases, *Geomatics and Information Science of Wuhan University*, 33(10):1026-1028.

Ali AH and Ashraf O, 2008, Robust and Blind Watermarking of Relational Database Systems, *Journal of Computer Science*, 4(12):1024-1029

Huang GC, Wang YB and Zhang KZ, 2011, Research and Design of Information Hiding Algorithm Based on Encoding of Unicode, *Computer Technology and Development*, 21(10):233-236.

Li QN, Li J and Wu R, 2012, Database Copyright Protection Based on Dual Zero-watermark, *Computer Engineering*, 38(8):107-110

Ren N, 2011, *Research on Digital Watermarking Algorithm for Remote Sensing Image*, Nanjing Normal University, Nanjing, China.