

文章编号:1007-5321(2017)04-0048-06

DOI:10.13190/j.jbupt.2017.04.008

# 抗数据合并攻击的矢量地理数据数字水印

王莹莹<sup>1,2,3</sup>, 杨成松<sup>4</sup>, 朱长青<sup>1,2,3</sup>, 任娜<sup>1,2,3</sup>, 方虎生<sup>4</sup>

(1. 南京师范大学 虚拟地理环境教育部重点实验室, 南京 210023; 2. 江苏省地理环境演化国家重点实验室培育建设点, 南京 210023;  
3. 江苏省地理信息资源开发与利用协同创新中心, 南京 210023; 4. 解放军理工大学 野战工程学院, 南京 210007)

**摘要:** 提出了一种抗地理要素合并攻击的矢量地理数据水印算法. 在分析了基于地理要素数据合并攻击对水印检测影响的基础上,按照单个地理要素检测—合并—检测的思想,设计并实现了一种抗数据合并攻击的矢量地理数据数字水印算法. 实验分析结果表明,该算法能有效抵抗常见的数据合并攻击,且可检测出合并后数据中的多个不同水印信息.

**关键词:** 矢量地理数据; 数字水印; 数据合并攻击; 鲁棒性

**中图分类号:** TN319.41

**文献标志码:** A

## Digital Watermarking Against Data Merging Attack for Vector Geographic Data

WANG Ying-ying<sup>1,2,3</sup>, YANG Cheng-song<sup>4</sup>, ZHU Chang-qing<sup>1,2,3</sup>, REN Na<sup>1,2,3</sup>,  
FANG Hu-sheng<sup>4</sup>

(1. Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China;

2. Institute of Field Engineering, PLA University of Science and Technology, Nanjing 210007, China;

3. State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China;

4. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China)

**Abstract:** A watermarking algorithm against geographic-feature-based data merging attack was presented. And a watermarking algorithm against data merging attack was presented based on analyzing the influence of geographic-feature-based data merging attack according to the detection-merging-detection thinking for single geographic feature. Experiments show that the proposed algorithm has good performance on the robustness against data merging attack. Additionally, the different watermarks in the merged vector geographic data can be detected by using the algorithm.

**Key words:** vector geographic data; digital watermarking; data merging attack; robustness

数字水印是信息安全领域中发展起来的前沿技术,为保护矢量地理数据版权并且追踪侵权行为提供了有效保障<sup>[1-2]</sup>. 许多学者对如何运用数字水印技术来保护矢量地理数据的安全进行了研究,并取得了一系列的研究成果<sup>[3-11]</sup>.

在矢量地理数据鲁棒数字水印算法的研究过程中,主要考虑的是数据点的增减、噪声(坐标抖动)和几何变换等常见的矢量地理数据水印攻击. 其中,常将数据合并攻击作为数据点的增加攻击,没有考虑数据合并攻击自身的特点. 目前,专门针对数

收稿日期: 2016-06-23

基金项目: 国家自然科学基金项目(41401518); 江苏省自然科学基金项目(BK20140066); 江苏高校优势学科建设工程资助项目

作者简介: 王莹莹(1985—),女,博士生; 杨成松(1982—),男,讲师, E-mail: 36946046@qq.com.

据合并攻击的矢量地理数据水印算法研究相对较少<sup>[12-13]</sup>, 算法的针对性也不强, 且鲁棒性也难以满足具体应用的需求。

笔者基于矢量地理数据的特征, 分析基于地理要素数据合并攻击的特点, 研究抗数据合并攻击的矢量地理数据水印算法, 并进行了实验分析。

## 1 矢量地理数据合并攻击特点分析

在矢量地理数据的生产处理过程中, 数据合并处理较为常见, 根据矢量地理数据水印算法的特点, 可以把水印算法设计中需要考虑的数据合并攻击分为2种基本类型: 基于区域的数据合并和基于地理要素的数据合并。

### 1.1 基于区域的数据合并

基于区域的数据合并是指参与合并的数据区域互不重叠或重叠部分较小, 如图1所示, 其中来自不同数据源的数据分别用虚线、实线表示。对于基于区域的数据合并攻击, 可以首先实现一种抗数据删减(裁剪)的水印算法, 进而通过采取合适的分块搜索检测策略来检测待测数据中是否含有水印信息<sup>[12]</sup>。

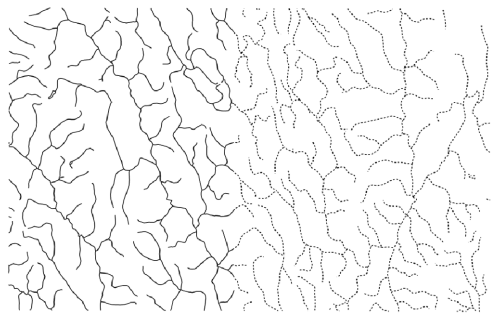


图1 基于区域的数据合并

### 1.2 基于地理要素的数据合并

基于地理要素的数据合并是指合并前数据中的地理要素在合并后大部分依然以单个地理要素的形式存在, 如图2所示。对于基于地理要素的数据合并攻击, 合并前的数据在合并后的数据中不以数据块的形式存在, 同一区域中可能存在来自多个不同数据源的地理要素, 因此, 难以通过分块检测策略检测出载体数据中含有的水印信息, 需要设计新的水印检测算法以应对基于地理要素的数据合并攻击。

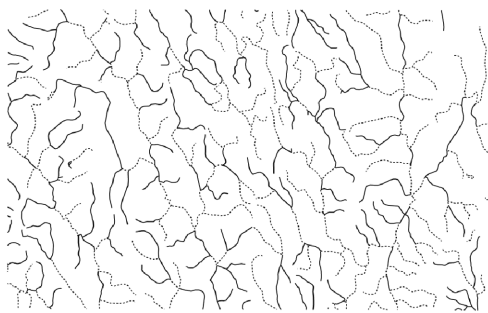


图2 基于地理要素的数据合并

## 2 抗数据合并的矢量地理数据数字水印算法

### 2.1 水印信息生成

根据水印信息有无明确的意义可以把水印信息分为有意义和无意义水印信息两类, 有意义水印信息指具有一定意义的文本、声音、图像或者视频信号, 无意义水印指水印信息没有明确的意义, 如伪随机二值序列等。相对于有意义水印, 无意义水印信息具有长度较小和统计特性好的特点, 考虑到矢量地理数据水印容量较小的特点, 采用统计特性较好的伪随机二值序列为待嵌水印信息<sup>[14]</sup>。

设水印信息为  $W$ , 并且  $W = \{w[i], 0 \leq i < M\}$ 。其中,  $w[i]$  表示水印信息位(比特位),  $w[i] \in \{-1, 1\}$ , 且  $w[i]$  满足  $P(w[i] = -1) = 1/2, P(w[i] = 1) = 1/2$ ;  $i$  表示水印信息位索引;  $M$  表示水印信息长度, 为了使算法适合小数据量的矢量地理数据, 水印信息的长度不宜过长, 取为 100。

### 2.2 水印信息嵌入

为了使水印算法能抵抗数据合并攻击的同时还能有效抵抗矢量地理数据处理中常见的数据压缩、删点、增点、裁剪、要素删除以及乱序(载体数据中坐标顺序的变化)等攻击, 并具有盲水印的特点, 采用基于集合映射与量化思想的水印嵌入算法<sup>[14]</sup>。

集合映射算法的基本思想: 把载体数据(矢量地理数据)视为一系列坐标点的集合, 对于载体数据中的任意坐标点  $(x, y)$ , 建立映射函数  $f(x, y)$ , 使得  $0 \leq f(x, y) < M$ , 即建立坐标点  $(x, y)$  到水印信息位索引  $i$  的映射关系。映射函数与数据点坐标  $(x, y)$  有关, 与数据点在载体中的前后顺序无关, 载体数据的增删并不会引起嵌入了水印的坐标  $(x, y)$  的变化, 因此,  $f(x, y)$  的值相对稳定, 这使得水印算法能有效地抵抗数据增删攻击和坐标的乱序攻击。

## 水印嵌入模型为

$$D \oplus W = \{x_i \oplus w[f(x_i, y_i)]\} \quad (1)$$

其中:  $D$  为载体数据,  $W$  为水印信息,  $x_i$  和  $y_i$  分别为第  $i$  个坐标点的  $x$  坐标和  $y$  坐标,  $\oplus$  表示水印嵌入规则. 采用量化思想来嵌入水印信息, 并将水印信息嵌入到  $x$  坐标中.

### 2.3 水印信息检测

从图2中基于地理要素的数据合并方式可以看出, 在大多数情况下, 合并前的地理要素在合并后的地理数据中依然以单个地理要素存在. 根据这一特点, 可以设计针对单个地理要素的水印检测算法, 检测出可能含水印的地理要素, 然后把所有可能含水印的要素合并成一个数据集, 并对该数据集进行水印检测, 进而判断出载体数据中是否含有水印信息. 水印信息检测基本流程如图3所示. 其中, 单个地理要素检测起到了筛选的作用, 剔除了不可能含有水印信息的矢量地理数据要素对最终的水印检测结果的影响.

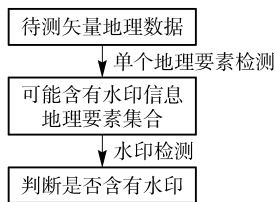


图3 水印信息检测流程

#### 2.3.1 单个地理要素水印检测

单个地理要素水印检测是指对单个的地理要素进行水印检测, 可以用于初步判断该地理要素是否含有水印信息. 为了实现单个地理要素水印检测, 需要对地理要素包含的坐标进行水印提取和相关统计工作. 设地理要素含有  $N$  个坐标点, 则水印提取统计流程如下:

- 1) 初始化统计量  $X=0, k=0$ ;
- 2) 对地理要素中的第  $k$  个数据点  $(x_k, y_k)$  进行水印信息提取, 设提取到的水印信息位为  $b(b \in \{-1, 1\})$ ,  $k=k+1$ ;
- 3) 根据映射关系  $f(x, y)$  获得坐标点  $(x_k, y_k)$  对应的水印信息位的索引  $i$ ;
- 4) 比较提取到的水印信息位  $b$  和原始水印信息位  $w[i]$ , 如果两者相等, 则  $X=X+1$ ;
- 5) 如果  $k < N$ , 则转到步骤2), 否则水印信息提取统计工作完成.

其中  $X$  用于统计某一地理要素中单个坐标点

水印提取结果正确的个数. 当水印提取并获得统计量  $X$  之后, 可以通过判断  $X$  和  $N$  之间的关系来判断矢量地理要素是否含有水印信息. 当单个地理要素中不含水印信息时, 提取到的单个水印信息位与原始水印信息位相等的概率是 0.5, 即统计流程中  $P(b=w[i])=0.5$ . 由此可知, 统计量  $X$  的大小服从二项分布:

$$P(X=n) = B(n, N, 0.5) = \binom{N}{n} (0.5)^N \quad (2)$$

其中:  $P(X=n)$  为地理要素不含水印信息条件下  $X$  取不同值时的概率分布,  $N$  为地理要素中数据点总数. 基于概率密度分布函数式(2), 可以通过假设检验的方法来初步判断地理要素中是否含有水印信息, 但是这种方法的计算量较大. 显然, 当  $X$  越接近  $N$  时表示待测地理要素含有水印信息的概率越大, 为了计算方便, 可以计算  $X$  与  $N$  的比值, 当  $X/N$  大于某一阈值时, 认为待测地理要素中可能含有水印信息. 考虑到统计特性的需要, 一般要求  $N$  的值要大于或者等于 10, 当  $N$  大于 10 小于 20 时, 检测阈值取 0.9; 当  $N$  大于等于 20 时, 检测阈值取 0.8.

#### 2.3.2 地理要素合并检测

通过对单个地理要素进行水印信息检测, 能初步判断出某一地理要素中是否含有水印信息. 但是, 由于单个地理要素所含数据点较少, 准确判断出单个地理要素中含有何种水印信息较难. 如图4所示, 当某一载体数据中可能嵌入了水印信息1或者水印信息2, 且某一可能含有水印信息的地理要素上的坐标点全部映射到水印信息位的第3位(索引为2)时, 单个地理要素检测结果对应的水印信息位都是1, 此时就很难判断出嵌入的是水印信息1还是

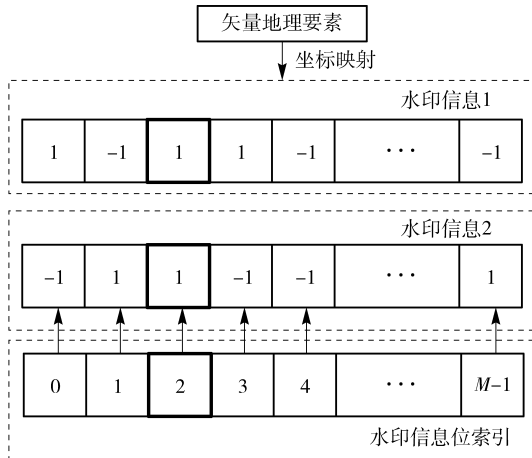


图4 地理要素与水印信息位映射关系



是水印信息 2. 即通过分析提取水印信息的统计特性,单个地理要素水印检测能够在一定程度上判断出该地理要素中是否含有水印信息,但大多数情况下难以确定该地理要素中含有何种水印信息.

为了解决这一问题,运用先筛选后合并检测的方法来实现抗地理要素层次的数据合并水印攻击.其基本思想:首先采用单个地理要素水印检测方法,对待测数据中的所有矢量地理要素进行水印检测;然后合并可能含有水印信息的地理要素,组成一个新的待测数据集合;最后提取新的待测数据集中的水印信息并进行相关检测,确定待测数据中含有何种水印信息.

设提取到的水印信息为  $W' = \{w'[i], 0 \leq i < M\}$ , 计算  $W$  与  $W'$  之间的相关系数  $c$ , 如式(3)所示. 一般地,当相关系数  $c$  大于 0.5 时,可以判定待测数据中含有水印信息  $W^{[14]}$ .

$$c = \frac{\left( \sum_{i=0}^{M-1} w[i] w'[i] \right)}{M}$$

(3)

3 实验与分析

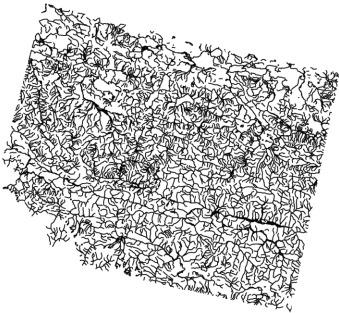
为验证提出算法的鲁棒性,特别是抗数据合并攻击能力,选取 3 幅比例尺为 1:100 万,坐标单位为  $m$  的线状矢量地理数据作为实验数据,分别含有 7 126、7 553 和 7 492 条线状地理要素,如图 5 所示.

实验基本方法:分别采用 Peng 等<sup>[8]</sup>提出的水印嵌入算法和笔者提出的水印嵌入算法,在实验数据中随机选取部分地理要素嵌入水印信息,模拟基于地理要素的数据合并攻击,然后分别运用 Peng 等<sup>[8]</sup>提出的水印检测算法和笔者提出的水印检测算法(算法 2)对嵌入水印后的实验数据进行水印检测,不同数据源和不同攻击条件下水印检测结果如表 1 所示,其中笔者在文献[8]水印提取的基础上增加了水印相关检测环节最终得到了算法 1. 表 1 中,合并攻击强度(%)是指嵌入水印信息的地理要素个数与载体数据实际含有的地理要素个数比值的百分数表示,√表示检测到水印信息,×表示检测不到水印信息.

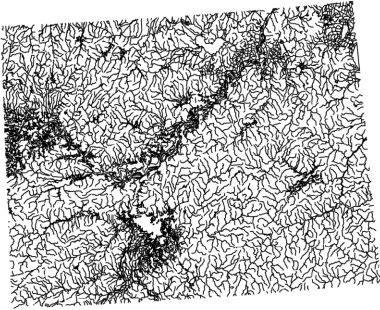
由表 1 检测结果的对比可以看出,在基于地理要素的合并攻击下,当合并后数据中含水印信息的地理要素较多的情况下(如超过 10%),算法 1 和算法 2 均能从合并后的数据中检测出水印信息. 但是,当合并后的数据中含水印信息的地理要素较少



(a) 实验数据1



(b) 实验数据2



(c) 实验数据3

图 5 实验数据

时(如低于 8%),算法 1 很难从合并后的数据中检测出水印信息,而算法 2 依然能有效地检测到水印信息,算法鲁棒性较好. 究其原因,算法 1 在水印检测过程中将不含有水印信息的矢量地理要素作为噪声进行处理,利用相关检测的抗噪能力来抵抗一定强度的数据拼接攻击;而算法 2 通过单个地理要素的水印检测,能将含有水印信息的线状地理要素筛选出来,同时剔除了不含水印信息的地理要素对水印检测结果的影响,提高了算法抗基于地理要素合并攻击的能力.

上面的模拟拼接攻击实验是数据源中只有 1 份数据含有水印信息的数据合并攻击的情况. 然而,在很多情况下,合并前的各数据中可能都含有不同的水印信息,结果导致合并后的数据中含有多多个水印信息,如何从合并后的数据中检测出多个水印信

表 1 检测结果对比

序号	合并攻击强度/%	算法 1	算法 2
1	11.23	√	√
	9.82	√	√
	8.42	√	√
	7.02	×	√
	5.61	×	√
	4.21	×	√
	2.81	×	√
	1.40	×	√
2	11.92	√	√
	10.59	√	√
	9.27	√	√
	7.94	×	√
	6.62	×	√
	5.30	×	√
	3.97	×	√
	2.65	×	√
3	1.32	×	√
	10.68	√	√
	9.34	√	√
	8.01	√	√
	6.67	×	√
	5.34	×	√
	4.00	×	√
	2.67	×	√
	1.33	×	√

息,也是抗数据合并攻击需要解决的问题. 为了检测算法是否具有检测多水印信息的能力,对实验数据进行模拟多水印合并攻击,分析算法的鲁棒性.

实验基本方法:从实验数据 1 中随机选取两部分互不相同的矢量地理要素集,利用水印嵌入算法对这两部分矢量地理要素分别嵌入不同的水印信息,模拟多个含有不同水印信息数据源的合并攻击,然后利用算法 1 和算法 2 检测各自嵌入的水印信息,检测结果如表 2 所示. 其中,每 1 次模拟合并攻击均对数据嵌入了 2 个水印信息,分别为水印 1 和水印 2,其后括号中的百分比表示嵌入水印时所使用的地理要素百分比.

例如,水印攻击 1 表示从实验数据 1 中分别选取 28.07% 的互不相同的两部分矢量地理要素,利用各自嵌入算法分别对其嵌入水印 1 和水印 2 两种不同的水印信息,然后利用算法 1 和算法 2 对各自

含有 2 个水印信息的载体数据进行水印信息的检测.

表 2 多水印检测结果对比

序号	水印	攻击强度/%	算法 1	算法 2
1	1	28.07	√	√
	2	28.07	×	√
2	1	11.23	×	√
	2	11.23	√	√
3	1	8.14	×	√
	2	8.14	×	√
4	1	6.31	×	√
	2	6.31	×	√
5	1	4.91	×	√
	2	4.91	×	√
6	1	2.81	×	√
	2	2.81	×	√

从表 2 中可以看出,当合并后的地理数据中含有 2 个水印信息时,在含水印信息的数据量较大的情况下,算法 1 能从合并后的数据中检测出水印信息,但是每次只能检测出 1 个水印信息,在含水印信息的数据量较小的情况下,算法 1 无法检测出任何水印信息;而算法 2 不仅能从合并后的数据中同时检测出嵌入的 2 个水印信息,而且即使含水印信息的数据量较小的情况下,依然能有效地从合并后的数据中完整地检测出这 2 个水印信息,具有较好的鲁棒性. 究其原因,主要是算法 1 在水印检测过程中没有考虑待测数据中存在多个水印信息的情况,在水印检测过程中根据相关检测值最大来确定载体数据中含有的水印信息,当含有水印信息的数据量较大时仅能检测到 1 个水印信息,含水印信息的数据量较小时,由于受到不含水印信息载体数据的影响,检测不到任何水印信息;算法 2 在水印检测过程中,通过筛选剔除了不含水印信息的地理要素以及含有其他水印信息的地理数据对水印检测结果的影响,能有效地从待测数据中检测到 2 个水印信息. 另外,由于在水印攻击过程中随机选择一定比例的数据嵌入水印信息,这种随机性导致在水印检测过程中,当含有水印信息的数据量较大时,算法 1 检测到水印 1 和水印 2 是随机的,如表 2 所示.

经实验分析,在基于地理要素的数据合并攻击条件下,当合并后的地理数据中含有 3 个或者 3 个以上的水印信息时,所提算法依然能完整地检测出

其中的水印信息。

## 4 结束语

基于矢量地理数据特征,对基于地理要素合并的矢量地理数据水印算法进行了研究,提出了一种抗基于要素合并的矢量地理数据水印算法,并对算法的鲁棒性进行了实验分析。实验结果和分析表明:① 算法能有效抵抗程度较强的数据合并攻击,在含有水印信息的地理要素较少的情况下依然能有效检测到数据载体中含有的水印信息;② 算法能有效检测出数据合并攻击中可能出现的多个版权(多个水印)信息,具有较好的实用性。

由于算法需要对单个地理要素的水印提取结果进行统计分析,因此只适用于线状、面状地理数据,目前还不适用于抵抗点状地理数据合并攻击,这是下一步需要研究的问题。

## 参考文献:

- [1] 朱长青,周卫,吴卫东,等. 中国地理信息安全的政策和法律研究[M]. 北京: 科学出版社, 2015: 2-45.
- [2] 朱长青,许德合,任娜,等. 地理空间数据数字水印理论与方法[M]. 北京: 科学出版社, 2014: 1-6.
- [3] 闵连权. 一种鲁棒的矢量地图数据的数字水印[J]. 测绘学报, 2008, 37(2): 262-267.  
Min Lianquan. A robust digital watermarking in cartographic data in vector format[J]. Acta Geodaetica et Cartographica Sinica, 2008, 37(2): 262-267.
- [4] 孙建国,门朝光,俞兰芳,等. 矢量地图数字水印研究综述[J]. 计算机科学, 2009, 36(9): 11-16.  
Sun Jianguo, Men Chaoguang, Yu Lanfang, et al. Survey of digital watermarking for the vector maps[J]. Computer Science, 2009, 36(9): 11-16.
- [5] 许德合,朱长青,王奇胜. 利用 DFT 幅度和相位构建矢量空间数据水印模型[J]. 北京邮电大学学报, 2011, 34(5): 25-28.  
Xu Dehe, Zhu Changqing, Wang Qisheng. A construction of digital watermarking model for the vector geospatial data based on magnitude and phase of DFT[J]. Journal of Beijing University of Posts and Telecommunications, 2011, 34(5): 25-28.
- [6] Cao Liujuan, Men Chaoguang, Ji Rongrong. High-capacity reversible watermarking scheme of 2D-vector data[J]. Signal Image and Video Processing, 2015, 9(6): 1387-1394.
- [7] Wang Nana, Zhao Xiangjun. 2D vector map data hiding with directional relations preservation between points[J]. Aeu-International Journal of Electronics and Communications, 2017, 71: 118-124.
- [8] Peng Zhiyong, Yue Mingliang, Wu Xia, et al. Blind watermarking scheme for polylines in vector geo-spatial data [J]. Multimedia Tools and Applications, 2015, 74(24): 11721-11739.
- [9] Ohbuchi R, Ueda H, Endoh S. Robust watermarking of vector digital maps[C]// 2002 IEEE International Conference on Multimedia and Expo (ICME2002). Lausanne: IEEE Press, 2002: 577-580.
- [10] Solachidis V, Nikolaidis N, Pitas I. Watermarking polygonal lines using Fourier descriptors[C]// 2000 IEEE International Conference on Acoustics, Speech and Signal Processing. Istanbul: IEEE Press, 2000: 1955-1958.
- [11] Lee S H, Kwon K R. Vector watermarking scheme for GIS vector map management[J]. Multimedia Tools and Applications, 2013, 63(3): 757-790.
- [12] 杨成松. 矢量地理数据数字水印模型与算法研究[D]. 郑州: 解放军信息工程大学, 2011.
- [13] 杨成松,朱长青. 基于常函数的抗几何变换的矢量地理数据水印算法[J]. 测绘学报, 2011, 40(2): 256-261.  
Yang Chengsong, Zhu Changqing. Research on watermarking algorithm robust to geometrical transform for vector geo-spatial data based on invariant function[J]. Acta Geodaetica et Cartographica Sinica, 2011, 40(2): 256-261.
- [14] 杨成松,朱长青,陶大欣. 基于坐标映射的矢量地理数据全盲水印算法[J]. 中国图象图形学报, 2010, 15(4): 684-688.  
Yang Chengsong, Zhu Changqing, Tao Daxing. A blind watermarking algorithm for vector geo-spatial data based on coordinate mapping[J]. Journal of Image and Graphics, 2010, 15(4): 684-688.