

一种支持多版权的地理数据增量水印方法

廖璋 熊胜超 彭智勇 彭煜玮
(武汉大学计算机学院 武汉 430072)

摘要 目前地理数据水印方法有很多,但大都缺少对地理数据同步保护的考虑,同时没有在同步保护的基础上进行多版权水印嵌入的方法。分析了地理数据增量水印和多版权水印的问题,基于量化调制技术,提出了一种能兼顾增量水印和多版权水印的地理数据水印方法。该方法采用基于距离划分的策略,将特征点集划分成两个可重现的不相交子集,通过分别缩放两个子集嵌入双重水印,保证了所有顶点都只在精度范围内修改一次,使保真性得到控制。此外,缩放调制策略使地理数据形状尽可能得到保证。实验表明,该方法能满足增量水印和多版权水印的要求,同时具有很好的鲁棒性。

关键词 地理数据,增量水印,多版权水印,保真性

中图分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.7.006

Geographical Watermarking Method Supporting Incremental and Multiple Watermark Embedding

LIAO Zhang XIONG Sheng-chao PENG Zhi-yong PENG Yu-wei
(School of Computer, Wuhan University, Wuhan 430072, China)

Abstract There are a lot of geographical data watermarking methods, but most of them lack consideration for the protection of fractional geographical data. This paper analyzed incremental and multiple watermarking problems on geographical data. Based on variable step quantization modulation technique, this paper proposed a watermarking method which can be used for both incremental and multiple watermark embedding. By dividing the feature set points into two disjoint subsets and zooming up/out all points in the subsets respectively, the method ensures the watermark embedding without disturbing the multiple watermarks, and guarantees that all the vertices are modified within the scope of fidelity for one time. So fidelity is controlled. In addition, by using the zoom modulation strategy, the shape of geographical data is preserved as well as possible. Experiments show that the method can meet the requirements of the incremental and multiple watermarks, and has good robustness as well.

Keywords Geographical data, Incremental watermark, Multiple watermark, Fidelity

地理数据是地理信息系统的基础,是国家基础设施建设和国防建设不可或缺的战略资源。GIS的发展使得地理数据在很多领域拥有广阔的应用场景,譬如车载导航、城市规划、Web地图等。对于地理数据生产者来说,数据采集是一项非常繁重的工作,需要使用昂贵的专业设备,同时需要耗费大量的人力物力,因此它是生产者的宝贵财富。然而,作为一种数字资源,地理数据非常容易被非法复制和传播,保护地理数据版权非常重要。

地理数据水印技术是保护地理数据版权、追踪盗版行为的重要手段,它通过适当修改原始数据,将水印信息隐藏在数据之中,在经过无意或有意的数据处理后,水印信息仍能被准确鉴别。水印信息可以是版权标识和附加信息,用来标识地理数据产品的所有权归属和进行产品描述;也可以是用户序列号,用来标识买方的身份信息,当数据被非法传播时,可通过检测地理数据中的用户序列号来追踪数据的非法泄露者。

目前,国内外学者提出了很多适用于矢量地理数据的水

印方法。这些方法很好地实现了地理数据版权保护,但在以下两方面存在不足。

1) 缺少对增量水印的考虑

地理数据生产周期一般较长,数据生产是分批、协同进行的,而非一次性完成,过长的生产周期增大了数据泄露的可能。对于用户来说,当前数据尽管不完整,但依旧具有可用性。为了防止不完整数据的版权丢失,每生产一批地理数据都需及时嵌入水印信息;同时为了管理方便,每次嵌入的水印信息应该相同。然而,现有的地理数据水印方法大都基于全局数据(所有顶点)嵌入水印^[1-9],无法达到增量水印嵌入的目的。有些方法不基于全局数据,而是基于地物对的空间关系嵌入水印^[14,15],但由于在数据的生产过程中,地物的生产时序是不定的,若使用该方法进行增量水印嵌入,很难保证地物对在嵌入和检测过程中的同步,因而该类方法同样无法实现增量水印。

到稿日期:2014-06-18 返修日期:2014-09-15 本文受国家自然科学基金项目(61100019)资助。

廖璋(1989-),男,硕士生,主要研究方向为地理数据管理、水印技术,E-mail:zhangliao@whu.edu.cn;熊胜超(1991-),男,硕士生,主要研究方向为轨迹数据管理;彭智勇(1963-),男,博士,教授,博士生导师,主要研究方向为复杂数据管理、地理数据管理;彭煜玮(1980-),男,讲师,CCF会员,主要研究方向为地理数据管理和地理数据水印,E-mail:ywpeng@whu.edu.cn(通信作者)。

2) 缺少对多版权水印的考虑

一些地理数据水印方法基于单个地物构造水印载体并生成水印嵌入^[10-13], 通过为每个地物构造鲁棒的地物标识, 保证水印嵌入和检测的同步。该类方法满足增量水印的一般要求, 但自身没有对增量水印问题进行深入分析, 而且只能嵌入单重水印。为了追踪数据的非法传播路径, 需在分发或出售的地理数据中同时嵌入买卖双方水印信息。为了使地理数据中蕴含买卖双方水印信息, 文献[10-13]只采用水印拼接的方式形成新水印, 然后将其重新嵌入到地理数据中。但是这样做会覆盖已有的水印信息, 且它实际上是对已修改过的数据作再次修改, 数据的保真性无法控制。

针对现有水印方法的上述不足, 本文提出了一种新的地理数据水印方法。本文贡献如下:

1) 首次提出地理数据增量水印问题和多版权水印问题, 并对此进行了详细的分析;

2) 针对矢量地理数据, 提出一种支持多版权的地理数据增量水印方法, 该方法能对当前生产过程中的部分数据同步嵌入水印, 在数据分发或出售时, 在不破坏已有水印信息的前提下, 再次嵌入买方水印信息, 实现数据的版权跟踪保护;

3) 通过实验验证了该方法的可行性。结果表明该方法具有一定的鲁棒性, 能抵抗平移、旋转、化简等攻击, 同时还能抵抗一定的顶点扰动攻击。

本文第1节分析了增量水印和多版权水印的问题; 第2节详细介绍了本文方法; 第3节给出了实验结果及分析; 最后给出了结论。

1 问题定义

1.1 增量水印

在实际生产过程中, 地理数据的采集是以地物为原子单位进行的, 即地物录入一次完成, 不存在一个地物分多次完成的情况。地理数据可以看成地物的集合, 即 $P = \{P_1, P_2, \dots, P_n\}$, n 表示地物个数, P_i 表示第 i 个地物, $P_i \in \{Polyline, Polygon\}$ 。传统水印方法大都只针对完整数据 P 单次嵌入水印, 若使用传统水印方法同步嵌入水印, 地物将会被重复修改, 地理数据的保真性无法控制, 其过程如式(1)所示。

$$M_i' = E(M_i, w) = E(M_{i-1}' \cup \Delta M_i, w) \quad (1)$$

其中, ΔM_i 表示第 i 批生产的增量数据, M_i 表示生产第 i 批时已拥有的全部地物集合, M_i' 表示 M_i 嵌入水印后返回的结果, E 表示水印嵌入过程, w 表示拟嵌入水印。可以看出, 每次对版本 M_i 嵌入水印时, 需要对 M_{i-1}' 中所有已嵌入水印的地物再次修改以嵌入水印, 若整个地理数据生产分 k 次完成, 则 ΔM_i 中的每个地物 P_L 被修改的期望次数为 $k-i+1$, 数据的保真性受到影响。

采用增量水印嵌入可避免数据因多次修改致使保真性无法控制的问题, 增量水印嵌入过程如式(2)、式(3)所示。

$$\Delta M_i' = E(\Delta M_i, w) \quad (2)$$

$$M_i' = M_{i-1}' \cup \Delta M_i' \quad (3)$$

其中, $\Delta M_i'$ 表示 M_i 嵌入水印后返回的结果。为了支持增量水印, 必须解决以下 3 个问题。

1) $E(\bigcup_{i=1}^k \Delta M_i, w) = \bigcup_{i=1}^k \Delta M_i'$, 即增量水印嵌入完毕后的地理数据与针对完整数据一次性嵌入水印后返回的地理数据一致。

2) 对于任意一个地理数据版本 M_i' , 若在版本 M_i' 中能检

测到某水印比特位, 则该水印比特位一定与拟嵌入水印串中对应位置上的水印比特位一致。也就是说, 检测到的水印串必须是拟嵌入水印串的子集, 即 $\forall i \in G, \text{若 } w \nabla D(M_i') = G \neq \emptyset, \text{必有:}$

$$w, b_i = D(M_i'), b_i$$

其中, D 表示水印提取操作, w 表示拟嵌入的水印串, b_i 表示水印串中下标 i 上的水印比特位。

操作符 ∇ 的定义如下: 对于两等长的水印比特串 w_1, w_2 , $w_1 \nabla w_2 = \{i | 1 \leq i \leq \text{length}(w_1), w_1, b_i \neq '*' \&\& w_2, b_i \neq '*' \}$, 其中 '*' 为填充位, 表示在该位置上未嵌入或检测不到水印比特位。譬如 $w = 110001$, 若检测结果 $D(M_i') = 11 * 0 * 1$, 则说明版本 M_i' 中检测到的水印串中位置 1、2、4、6 上的水印比特位为有效水印位, 由于它与 w 相应位置的水印比特位保持一致, 我们称 $D(M_i')$ 是 w 的子集。

3) $w = D(M_k')$, 即针对最后一个增量集 ΔM_k 嵌入水印后, 得到最终可出售的完整版本 M_k' , 对 M_k' 进行水印检测可以得到拟嵌入的原始水印信息 w 。

问题 1) 要求保证增量水印分批嵌入与全局水印一次嵌入对数据的修改一样。问题 2) 要求每个地物增量集都蕴含能标识版权的水印信息。问题 3) 要求在全局水印检测时, 能提取完整的水印信息。

数据生产过程中, 地物的采集是无序的, 当前地物后面将生产哪些地物也是未知的。因而为了实现增量水印, 必须基于单个地物构造水印载体。若水印嵌入不基于单个地物而依赖于多个地物, 那么地物的逆序录入将破坏水印嵌入时地物之间的依赖关系, 进而导致水印嵌入和检测不同步。

同时针对单个地物来说, 顶点的采集也是无序的, 以 *Polygon* 地物为例, 可以从 *Polygon* 顶点中的任何一个位置开始存储顶点, 所以水印载体的构造不能依赖于地物中的点序。若水印载体依赖于点的生产时序, 那么点的逆序或指定任意起始点录入同样将破坏嵌入的水印。

1.2 多版权水印

为了使地理数据中同时蕴含买卖双方甚至多方水印信息, 传统的地理数据单重水印方法只能采取水印拼接再嵌入的方式。不失一般性, 本文以二次水印为例进行讨论, 其形式化描述如式(4)、式(5)所示。

$$wm = wm_{seller} + wm_{buyer} \quad (4)$$

$$P'' = E(P', wm) \quad (5)$$

其中, wm_{seller} 、 wm_{buyer} 分别表示拟嵌入买卖双方水印信息, P' 为已经蕴含卖方水印 wm_{seller} 后的地物集合。水印嵌入本质上是在地图精度范围内对顶点进行修改。从上述过程可以看出, P'' 为在已修改过的 P' 上再次修改, 若直接采用传统的水印方法, 将无法避免对 P' 中已修改过的顶点再次修改。尽管 $P''(P')$ 是基于版本 $P'(P)$ 在精度范围内对顶点合法改动后的版本, 但是较原始数据 P 而言, 很难控制 P'' 中的顶点修改不超出精度范围。

在考虑增量水印场景的情况下, 为了解决二次水印嵌入时数据因多次修改而导致空间精度无法控制的问题, 可采用如下解决方案:

以单个地物为水印调制对象, 首先将地物顶点集划分成两个互不相交的子集, 即 $P_i = P_{i1} \cup P_{i2}$, 且 $P_{i1} \cap P_{i2} = \emptyset$, 其中 P_i 表示地图中第 i 个地物的顶点集合。然后分别在两个子

集上嵌入独立的水印信息。增量水印嵌入和二次水印嵌入过程可分别描述成式(6)、式(7)。

$$P_i' = \text{incre_embed}(P_i, P_{i1}), P' = \bigcup_{i=1}^n P_i' \quad (6)$$

$$P_i'' = \text{dual_embed}(P_i', P_{i2}), P'' = \bigcup_{i=1}^n P_i'' \quad (7)$$

其中, *incre_embed* 表示对地物进行增量水印调制, *dual_embed* 表示对地物进行二次水印调制, P_i, P_{i1} 和 P_i, P_{i2} 分别表示地物 P_i 顶点集的子集, P_i', P_i'' 分别表示嵌入增量水印、二次水印后的地理数据。

从式(6)、式(7)可以看出, 为了保证增量水印和二次水印的可逆检测, 必须保证增量水印和二次水印嵌入完毕后, 顶点子集划分的前后一致性, 即保证:

$$P_i, P_{i1}(P_i, P_{i2}) = P_i', P_{i1}(P_i', P_{i2})$$

$$P_i'', P_{i1}(P_i'', P_{i2}) = P_i, P_{i1}(P_i, P_{i2})$$

2 一种支持多版权的地理数据增量水印方法

2.1 方法概述

文献[13]基于单个地物通过调制平均特征距离来嵌入水印, 为增量水印提供了思路; 然而该方法只能嵌入单重水印, 具有一定的局限性。本文借鉴可变速长量化调制技术, 采用逻辑分层的思想, 将地物特征点划分成两个可重现的不相交子集, 每个子集蕴含一个水印信息, 最终通过缩放子集实现水印嵌入。

首先抽取地物 P_L 的特征点集合 FS , 然后基于平均特征距离将特征点集合分成两个不相交子集 FS_1, FS_2 ; 接着分别求取两个子集的顶点中心和平均中心距离, 以两子集的顶点中心距离与平均中心距离的比值构造地物标识。根据地物标识将地物 P_L 划分到指定分组中, 一个分组对应一个水印位, 将水印位重复嵌入到分组的每个地物中。同时以两子集的平均特征距离分别作为增量水印和二次水印的载体, 采用量化调制技术对水印载体进行调制, 最终通过顶点缩放达到水印嵌入的目的。

2.2 地物标识和分组

为了使算法具有一定的抵抗化简攻击的能力, 通常选取地物的特征点来构造水印载体并调制。道格拉斯-普克算法是一种被广泛应用于提取特征点的方法, 为了有效地保留曲率变化较大的特征点, 本文采用一种基于偏角限制的道格拉斯-普克算法求取特征点, 详见文献[17]。地物标识生成过程如下:

(1) 设通过特征点抽取得到地物 P_L 的特征点集 $FS_L = \{p_{L1}, p_{L2}, \dots, p_{Lk}, \dots, p_{Lk}\}$ 。求取 FS_L 的顶点中心 O_L , 并计算所有顶点中心距离, 得到距离集 $FD_L = \{fd_{L1}, fd_{L2}, \dots, fd_{Lk}, \dots, fd_{Lk}\}$, 其中 fd_{Lk} 表示点 p_{Lk} 到顶点中心的距离。

(2) 计算 FS_L 平均中心距离 $\overline{fd_L}$, 根据 $\overline{fd_L}$ 将 FS_L 分成两部分 FS_1, FS_2 , 如下所示。

$$FS_1 = \{p_{Lk} | p_{Lk} \in FS_L \& \& fd_{Lk} > \overline{fd_L}\}$$

$$FS_2 = \{p_{Lk} | p_{Lk} \in FS_L \& \& fd_{Lk} < \overline{fd_L}\}$$

(3) 分别求取 FS_1, FS_2 的顶点中心 O_{s1}, O_{s2} , 以及 FS_1, FS_2 的平均中心距离 $\overline{fd_1}, \overline{fd_2}$, 同时计算 O_{s1}, O_{s2} 的距离 $Dist$ 。

顶点中心表示顶点集合中所有点坐标的平均值, 其计算如式(8)所示。

$$x_{O_L} = \frac{\sum_{i=1}^n x_i}{n}, y_{O_L} = \frac{\sum_{i=1}^n y_i}{n} \quad (8)$$

平均中心距离表示集合中所有顶点到顶点中心的距离的平均值, 具体计算如式(9)、式(10)所示, 其中 d_{Lj} 表示地物 P_L 中第 j 个点到地物中心的距离。

$$d_{Lj} = \sqrt{(x_j - x_{O_L})^2 + (y_j - y_{O_L})^2} \quad (9)$$

$$\overline{fd} = \frac{\sum_{j=1}^n d_{Lj}}{n} \quad (10)$$

(5) 计算两子集的顶点中心距离与子集平均中心距离和的比值 DR , 取 DR 的高 h 有效位作为地物标识 Pid_L 。

$$DR = \frac{Dist}{fd_1 + fd_2} \quad (11)$$

$$Pid_L = \text{hsb}(DR, h) \quad (12)$$

其中, h 表示小数点后的位数, 即选取 DR 所有整数位上数字及小数点后 h 位作为地物标识。

由于距离不会因地物平移、旋转而发生改变, 因此 DR 具有很好的几何稳定性。同时因为 Pid_L 基于特征点构建, 从而保证了在经受一定的化简攻击后, Pid_L 仍能保持稳定。此外, 由于顶点扰动对 DR 的影响是从低位向高位传播的, 当选取合适的 h 值时, 在精度范围内对顶点进行扰动将不会影响 DR 高 h 位上的数值, 因此在一定程度的顶点扰动下, 地物标识仍能同步。

为了增强算法对地物攻击的抵抗能力, 本方法采用分组策略, 将水印位重复嵌入到分组的每个地物中。水印检测时重复提取分组中每个地物的水印位, 最终采用投票表决的方式决定分组所蕴含的水印位, 如式(13)所示。

$$Partition_L = Pid_L \% m \quad (13)$$

其中, m 为水印串长度, $Partition_L$ 表示该地物所在分组的组号。在整个数据集中, 所有地物的 DR 是随机分布的, 从而保证了分组尽可能均匀。

特征点子集划分如图1所示。

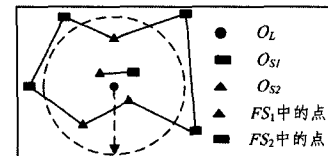


图1 特征点子集划分

2.3 增量水印嵌入

根据地物 P_L 所在分组为 $Partition_L$, 得到拟嵌入该地物 P_L 的水印位, 选择 FS_1 的平均特征距离 $\overline{fd_1}$ 作为水印载体。为了增强算法抗顶点扰动的能力, 本方法使用可变速长量化调制技术[13]对其进行调制, 过程如下:

(1) 根据 Pid_L 生成针对该地物 P_L 的步长 d_L 。生成过程其实是一个函数映射过程, 因为地物标识的鲁棒性, 所以步长也相应具有很好的稳定性。由于顶点修改只能在精度范围内进行, 且在调制时以步长为单位进行顶点移动, 因此步长必须被控制在精度范围内, 如式(14)、式(15)所示。

$$lf = \text{globalFunc}(Pid_L) \quad (14)$$

$$d_L = (gf_{lower} + (gf_{upper} - gf_{lower}) * lf) * \tau \quad (15)$$

其中, $0 \leq lf \leq 1$, τ 表示地理数据空间精度误差, 是一个已知值, 参数 gf_{lower} 和 gf_{upper} 为私密的用于控制步长范围的全局

参数,且满足 $0 \leq gf_{lower} \leq gf_{upper} \leq 1$, 显然 $d_L \in (gf_{lower} * \tau, gf_{upper} * \tau)$, 保证了步长被控制在精度范围内。

(2) 根据步长 d_L 生成 0(1) 量子子集合, 计算 $\overline{fd_1}$ 所在量化区间, 根据拟嵌入水印比特位, 对 $\overline{fd_1}$ 进行适当调制。为了提高算法抗顶点扰动的能力, 在调制时将 $\overline{fd_1}$ 调制到量化区间的中点。

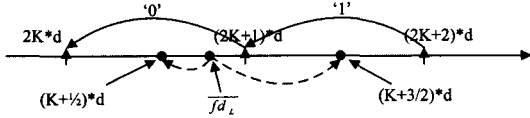


图 2 量化调制策略

如图 2 所示, 每个 0(1) 量子子代表一个区间, 设 $q_0 (q_1)$ 表示 0(1) 量子子, 对于该量子子区间中的每一个值 v , 满足如下定义。

定义 1 $v \in q_0$ 当且仅当存在 $k \in N$, 使得 $2k * d < v \leq (2k+1) * d$ 。

定义 2 $v \in q_1$ 当且仅当存在 $k \in N$, 使得 $(2k+1) * d < v \leq (2k+2) * d$ 。

(3) 以 O_{s1} 为中心, 通过适当缩放 FS_1 中的所有点嵌入水印。设调制后的平均特征距离为 $\overline{fd_1}'$, 则可得到 FS_1 的缩放比例 θ , 如式(16)所示。

$$\theta = \frac{\overline{fd_1}'}{\overline{fd_1}} \quad (16)$$

综上, 增量水印嵌入算法如图 3 所示。

输入: 地物增量集 ΔP

拟嵌入水印串 $W_S = b_1 b_2 b_3 \dots b_m$

步长控制参数 gf_{lower}, gf_{upper}

空间精度误差: τ

输出: 嵌入水印 W_S 后的地物集 $\Delta P'$

for each P_L in ΔP

$FS_L = \text{get_featurePoint_Set}(P_L, \tau)$

$O_L = \text{compute_FS_center}(FS_L)$

$mfd_L = \text{compute_fd}(FS_L, O_L)$

$\{FS_1, FS_2\} = \text{sub_partition}(FS_L, mfd_L)$

$\{OS_1, OS_2\} = \text{compute_center}(\{FS_1, FS_2\})$

$Mfd_{s1} = \text{compute_fd}(FS_1, OS_1)$

$Mfd_{s2} = \text{compute_fd}(FS_2, OS_2)$

$Dist = \text{compute_dist}(OS_1, OS_2)$

$Pid_L = \text{compute_pid}(Mfd_{s1}, Mfd_{s2}, Dist)$

$Partition_L = Pid_L \text{ mod } m$

$b_L = W_S.at(Partition_L)$

$lf = \text{globalFunc}(Pid_L)$

$d_L = (gf_{lowe} + (gf_{upper} - gf_{lower}) * lf) * \tau$

$Mfd'_{s1} = \text{quantize}(Mfd_{s1}, b_L, d_L)$

$\theta = Mfd'_{s1} / Mfd_{s1}$

$FS_1' = FS_1.scale(OS_1, \theta)$

$P_L' = (P_L - FS_1) \cup FS_1'$

$P_L'.store()$

endfor

return $\Delta P'$

图 3 增量水印嵌入算法

2.4 增量水印检测

尽管增量水印嵌入对象为地物增量集 ΔM_i , 水印检测的

对象为当前地理数据的某个版本 M_i (即 $\cup_i \Delta M_i'$), 两者表面上存在差别, 但本质上由于嵌入和检测都是基于单个地物, 因此水印检测可看成是水印嵌入的逆过程。

针对单个地物, 首先计算 FS_1 的平均特征距离 $\overline{fd_1}$, 同时计算得到该地物的调制步长 d_L , 得到该地物相关的量子子区间, 通过判断 $\overline{fd_1}$ 所在的量化区间 loc , 决定该地物蕴含的水印位 b_L , 如式(17)所示, 具体通过原则 I 来判断。

原则 I: 如果存在 $k \in N$, 使得 $2 * k < loc \leq 2 * k + 1$, 则 $b_L = '0'$; 否则 $b_L = '1'$ 。

$$loc = \lfloor \frac{\overline{fd_1}}{d_L} \rfloor \quad (17)$$

综上, 增量水印检测算法如图 4 所示。

输入: 地物集 $M_i' = \{P_{L1}, P_{L1}, \dots\}$

步长控制参数 gf_{lower}, gf_{upper}

空间精度误差 τ

投票阈值 Ω

已嵌入水印串的大小 m

输出: 长度为 m 的水印串 $W' = b_1 b_2 \dots b_m$

$vote[m] = \{0, \dots, 0\}; W' = 0 \dots 0; G_N[m] = \{0, \dots, 0\};$

for each P_L in M_i'

$FS_L = \text{get_featurePoint_Set}(P_L, \tau)$

$O_L = \text{compute_FS_center}(FS_L)$

$Mfd_L = \text{compute_fd}(FS_L, O_L)$

$\{FS_1, FS_2\} = \text{sub_partition}(FS_L, Mfd_L)$

$\{OS_1, OS_2\} = \text{compute_center}(\{FS_1, FS_2\})$

$Mfd_{s1} = \text{compute_fd}(FS_1, OS_1)$

$Mfd_{s2} = \text{compute_fd}(FS_2, OS_2)$

$Dist = \text{compute_dist}(OS_1, OS_2)$

$Pid_L = \text{compute_pid}(Mfd_{s1}, Mfd_{s2}, Dist)$

$Partition_L = Pid_L \text{ mod } m$

$G_N[Partition_L]++$

$lf = \text{globalFunc}(Pid_L)$

$d_L = (gf_{lower} + (gf_{upper} - gf_{lower}) * lf) * \tau$

$loc = \lfloor Mfd_{s1} / d_L \rfloor$

if $(1 = loc \% 2)$

$vote[Partition_L]++$

else

$vote[Partition_L]--$

endfor

initialize $i=0$

while $(i++ < m)$

if $(vote[i] > (2\Omega - 1) * G_N[i])$

$W'.at(i).set('1')$

end while

return W'

图 4 增量水印检测算法

其中, Ω 一般选择 50%, 表示当分组中超过一半的地物投票为 '1' 时, 将该分组所携带的水印位判定为 '1', 否则为 '0'。

2.5 二次水印嵌入和检测

二次水印嵌入时, 以 FS_2 的平均特征距离 $\overline{fd_2}$ 为水印载体。在增量水印嵌入时, 只基于 FS_1 的顶点中心缩放 FS_1 中的所有点, 由于缩放是在精度范围内进行的, 对特征中心和平均中心距离 $\overline{fd_L}$ 的影响很小, 因此水印嵌入后, 以 $\overline{fd_L}$ 为依据划分 FS_1, FS_2 是可重现的。对于增量水印嵌入后导致 FS_1

和 FS_2 不可重现的极端情况,由于采用分组投票策略,单个地物水印位破坏不影响分组水印位的检测结果,使得对这种极端情况的容忍度增强。

除水印载体和调制策略不同外,二次水印嵌入选择地物标识、分组等步骤与增量水印嵌入过程一致;此外,二次水印嵌入时可以选择独立调制参数。尽管二次水印方法基于完整地理数据嵌入水印,但实际上可以将其看成是增量水印的一种特殊情况。

如图3所示,将 ΔP 换成 P' ,以 $Mfd_{i,2}$ 为水印载体,以 OS_2 为中心缩放 FS_2 ,即为二次水印嵌入算法。如图4所示,将 M_i 换成最终地理数据版本 M_k ,以 $Mfd_{i,2}$ 为水印检测载体,且基于原则 I 判断水印位,即为二次水印检测算法。

3 实验结果及分析

本节通过实验来验证所提出方法的可行性,并评估该方法抵抗几何攻击、顶点扰动、顶点化简的性能。

3.1 实验数据集

实验数据集为上海地图,包含 10692 个 *Polygon* 类型地物,共 583598 个点,原始地图如图5所示。

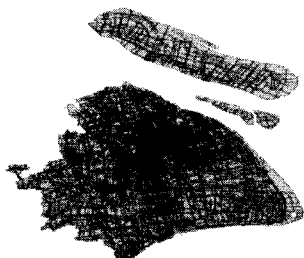


图5 上海市地图

3.2 增量水印实验

本文通过以下方法来模拟地理数据增量的生产过程:首先将 10692 个地物随机分配到 K 个子集中,每个子集不为空;然后依次对每个子集中的地物嵌入相同的水印信息,同时在每次对第 i 个子集中的所有地物嵌入水印 w 完毕后,针对当前所有已嵌入水印的地物(即地理数据版本 M_i'),提取其中的水印信息。

拟嵌入的水印串 w_1 大小为 128 位, $K=10$,步长调制参数 gf_{upper} 和 gf_{lower} 分别为 0.8、0.6,特征点化简阈值为地图精度 τ 。实验结果如表1所列,从中可以看出该方法能达到增量水印的要求。

表1 增量水印检测

版本号	总地物数	水印检测率
1	599	100%
2	1300	100%
3	3501	100%
4	4022	100%
...
10	10692	100%

3.3 二次水印实验

利用增量水印方法向地理数据中嵌入水印信息 w_1 ,然后再向其中嵌入水印信息 w_2 。其中水印串 w_1 、 w_2 大小都为 256,增量水印嵌入中参数 gf_{upper} 、 gf_{lower} 分别为 0.8 和 0.6,二次水印嵌入参数 gf_{upper} 、 gf_{lower} 分别为 0.7 和 0.5,特征点化简阈值为空间误差精度 τ 。

为了便于对比,将 w_1 、 w_2 分别表示为像素为 16×16 的位图,实验结果如图6、图7所示。



图6 增量水印 w_1 嵌入前后 图7 二次水印 w_2 嵌入前后

检测得到增量水印 w_1 匹配正确率为 98.8%,二次水印 w_2 匹配正确率为 99.2%。从图6、图7可以看出,嵌入二次水印 w_2 不影响已嵌入增量水印 w_1 。实验表明,当步长控制参数 gf_{upper} 、 gf_{lower} 选择区间 $[0,1]$ 的不同值时,水印正确率不受影响。

3.4 鲁棒性评估

3.4.1 几何攻击

平移、旋转不影响地物中顶点之间的距离,因而不影响变换前后特征点集合的重现;此外,平均特征距离也不会因为平移、旋转而受到影响,该方法能抵抗这两种几何攻击。由于缩放破坏了特征点集合的可重现性,缩放后水印载体和地物标识将无法同步,即该方法无法抵抗缩放攻击。然而地理数据生产过程中,顶点坐标都遵循一定的比例尺,因而将地图缩放到已知的原始比例尺上,即可检测到嵌入地理数据中的水印信息。实验表明,地图缩放后,将其还原到原始比例尺上仍能检测出完整水印。

3.4.2 顶点扰动

一般来说,考虑到地理数据的可用性,顶点扰动只能在地图精度 τ 内进行。在实验中,为实施扰动,以顶点为中心,以扰动值 ρ 为半径画圆,将顶点随机调整到该圆区域的某个位置上。然后从扰动后的数据中分别提取增量水印和二次水印,并计算水印的匹配正确率。实验中使用的参数与3.3节保持一致。

如图8所示,横轴单位为地图精度 τ ,纵轴表示水印匹配正确率。从中可以看出,顶点随机扰动使得匹配正确率呈现一定的波动性,但总的来说,顶点扰动对水印的正常检测没有太大影响。对于增量水印来说,匹配正确率基本在 95.7% 左右,对于二次水印来说,匹配正确率在 95.3% 左右。本方法具有较强的抵抗顶点扰动的能力。

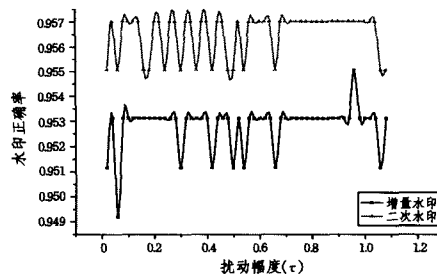


图8 顶点扰动

3.4.3 顶点化简

本文采用道格拉斯-普克算法对已嵌入双重水印的地理数据进行化简,然后从化简后的地理数据中检测水印并计算匹配正确率。实验参数与3.3节一致。图9显示了不同化简幅度下水印检测的正确率,其中横轴单位为地图精度 τ 。

从图9可以看出,当化简幅度小于 τ 时,化简攻击对水印检测没有影响;当化简幅度大于 τ 时,正确率呈直线下降趋势;当化简幅度在 τ 和 2τ 范围内时,水印匹配率大于 80%,可以通过检测出的增量水印和二次水印声明版权。

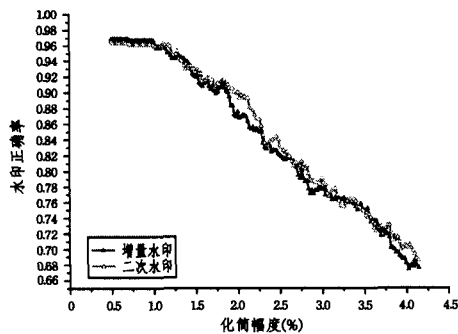


图9 顶点化简

结束语 本文首次提出了地理数据增量水印和多版权水印的问题。针对该问题,以量化调制技术为基础,设计了一种支持多版权的地理数据增量水印方法。实验表明,该方法能够实现增量水印和二次水印的嵌入及检测,同时具有较强的鲁棒性。

参考文献

[1] Samoa M, Matsuura Y, Takashima Y. A Scheme of digital watermarking for geographical map data[C]//Proc. of the Symposium on Cryptography and Information Security, Okinawa, Japan, 2000;26-28

[2] Schulz G, Voigt M. A high capacity watermarking system for digital maps[C]//Proc. ACM Int. Workshop on Multimedia and Security, Magdeburg, Germany, Sep. 2004;180-186

[3] Kang H. A vector watermarking using the generalized square mask[C]//Proc. of the International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA, 2001;234-236

[4] Ohbuchi. Robust watermarking of vector digital maps [C]//Proc. of the IEEE International Conference on Multimedia and Expo., Lausanne, Switzerland, vol. 1, 2002;577-580

[5] Voigt M, Busch C. Feature-based watermarking of 2D-vector data [C]//Proc. of the SPIE, Security and Watermarking of Multimedia Content, Santa Clara, USA, vol. 5020, 2003;359-366

[6] Zhang L, et al. New robust watermarking algorithm for vector data[J]. Wuhan University Journal of Natural Sciences, 2010, 15(5):403-407

[7] Bazin C, Le Bars J M, Madelaine J. A blind, fast and robust

method for geographical data watermarking[C]//ACM Symposium on Information, Computer and Communications Security, 2007;265-272

[8] Lee S H, Kwon K R. Vector watermarking scheme for GIS vector map management[J]. Multimedia tools and applications, 2013, 63(3):757-790

[9] Yan Hao-wen, Li J, Wen Hong. A key points-based blind watermarking approach for vector geo-spatial data[J]. Computers, Environment and Urban Systems, 2011, 35(6):485-492

[10] 邵承永,汪海龙,牛夏牧,等. 基于统计特征的二维矢量地图鲁棒水印算法[J]. 电子学报, 2005, 33:2312-2316

Shao Cheng-yong, Wang Hai-long, Niu Xia-mu, et al. A Robust Watermarking Algorithm for 2 D Vector Maps Based on Statistic Detection[J]. Acta Electronica Sinica, 2005, 33:2312-2316

[11] Lafaye J B, Gross-Amblard D. Invisible graffiti on your buildings; blind and squaring-proof watermarking of geographical databases[C]//Lecture notes in computer science, advances in spatial and temporal databases, LNCS 4605, 2007;312-329

[12] Jungyeop K. Vector map digital watermarking using angles [C]//2010 Sixth International Conference on Networked Computing and Advanced Information Management (NCM). 2010

[13] 汪传建,葛贺飞,丁卯,等. 一种基于可变步长量化调制的地理数据库水印方法[J]. 计算机研究与发展, 2011, 48(10):1960-1971

Wang Chuan-jian, Ge He-fei, Ding Mao, et al. A Geographic Databases Watermarking Method Based on Quantization Modulation with Variable Steps[J]. Journal of Computer Research and Development, 2011, 48(10):1960-1971

[14] Wang Chuan-jian, et al. Watermarking geographical data on spatial topological relations [J]. Multimedia Tools and Applications, 2012, 57(1):67-89

[15] Wu Bai-yan, Wang Wei-a, Miao Dan-dan. 2D vector map watermarking based on spatial relations[C]//International Conference on Earth Observation Data Processing and Analysis (ICEODPA). 2008

[16] 刘晓红,李树军,黄文睿. 制图综合中偏角限制道格拉斯算法研究[J]. 测绘与空间地理信息, 2006, 29(1):59-60

Liu Xiao-hong, Li Shu-jun, Huang Wen-qian. Study of Douglas-Peucker Algorithm Controlling by the Goniometry in Generalization[J]. Geomatics & Spatial Information Technology, 2006, 29(1):59-60

(上接第14页)

本文通过多个 NP 完全问题到 MSP 问题的归结,揭示了 MSP 问题所反映的 NP 完全问题的共性。MSP 问题暴露了隐含的解空间的搜索过程,具有良好的表达能力。归结的技巧在于:通过顶点边集的合理取值,排除以不符合解局部特征的顶点为端点的边。此外,本文对 MSP 问题的相变进行了理论分析,为相关算法的测试提供了难例产生方法。

参考文献

[1] Jiang Xin-wen. A Polynomial Time Algorithm for the Hamilton Circuit Problem[J]. ArXiv preprint cs/1305.5976, 2013

[2] Jiang Xin-wen, Peng Li-hong, Wang Qi. MSP Problem: Its NP-completeness and Its Algorithm[C]//Proceedings of CUTE. 2010;1-5

[3] Jiang Xin-wen, Liu Wan-wei, Wu Tian-jun, et al. Reductions from MSP to SAT and from SUBSET SUM to MSP[J]. Journal of Computational Information Systems, 2014, 10(3):1287-1295

[4] Fan Shuo, Jiang Xin-wen. Proving NP-completeness of Polynomial Reduction from the SAT Problem to the MSP Problem (English Abstract)[J]. Computer Science, 2012, 39(11):179-182

[5] Wang Shao-wen. The Study of the Number of Colors on the Graph (English Abstract)[J]. Journal of Beijing Institute of Machinery, 1997, 12(1):15-20

[6] Dong An-guo, Gao Lin, Zhao Jian-bang. Algorithms for Subgraph Isomorphism in Graph Pattern Mining (English Abstract)[J]. Mathematics in Practice and Theory, 2011, 41(13):105-112

[7] Gebauer H, Szabo T, Tardos G. The Local Lemma Is Tight for SAT[C]//Proceedings of SODA. 2011;664-674