

基于二维矢量地图属性特征的数字水印算法^①孙建国^② 门朝光

(哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)

摘 要 针对空域和频域数字水印对矢量地图造成不同程度的精度损失以及抗压缩性能差的特点,提出了一种新的矢量地图数字水印算法。该算法利用矢量对象的属性特征构造水印信息,采用网格密度空间聚类方法选取特征结点,并通过对特征结点的二维坐标进行微调,将数字水印嵌入到地图属性文件中。实验和分析表明,该算法达到了矢量地图精度零损伤的目标,同时,提高了数字水印的鲁棒性和实用性。

关键词 水印, 矢量, 地图, 特征

0 引 言

矢量地图是属性信息描述下,通过拓扑关系约束,以图形图像为表现形式的数字产品。典型矢量地图数字水印算法的研究主要集中于空域和频域(变换域)研究^[1]。空域水印对矢量地图的数据精度损失较大,研究人员虽进行了很多改进^[2,3],但其鲁棒性和抗数据压缩能力仍然较弱;频域方法如离散傅立叶变换(DFT)^[4],离散余弦变换(DCT)^[5],离散小波变换(DWT)^[6]等在数字水印的各应用领域均是研究热点,但频域水印对矢量地图精度仍存在一定损伤,且无法抵抗复杂的矢量数据压缩攻击^[7]。对此,文献[8]提出了通过坐标漂移方式向属性文件中嵌入水印的方法。同比以上算法,该方法大大降低了精度损失,但抵抗数据压缩攻击的性能没有获得明显提升。

鉴于现有矢量地图的水印不足,本文提出了一种面向地图属性特征的数字水印算法。与上述水印算法相比,该算法具有以下特点:不对坐标以及地图图像图形进行任何变换操作,使地图精度始终保持零损失;运用网格密度空间聚类的分析方法提取特征结点,结点的规模决定水印容量;在属性信息中嵌入水印,使算法具有突出的稳健性(鲁棒性)和安全性。

1 水印嵌入目标群

水印嵌入目标群是指由算法确定可嵌入水印标

识的点集。众所周知,理想的水印嵌入点可以增强算法的稳健性,对于数据压缩等去除型攻击具有较好的抵抗能力。聚类分析方法能够辅助选取并生成嵌入目标群。

1.1 改进的网格密度空间聚类分析方法

本文设计了一种改进的基于网格密度的空间聚类分析方法。其输入为网格步长和密度阈值。算法保持了网格方法在处理大数据量时的高效率以及密度方法可获取任意形状簇的优势^[9]。为了更好地锁定目标,对该算法主要进行了两方面改进:

改进 1——简化聚类空间。在矢量地图中,其表现形式约包含 14 类图层^[10],去掉各种图形图像形式的渲染图层,将其简化为单一的拓扑关系(如图 1(a)所示),可以获得更可靠的聚类结果。

改进 2——优化属性密度阈值。传统的密度阈值仅以网格内结点分布情况来考查簇的密集性,属性密度阈值结合结点属性特征(属性熵值)来表示结点的密度权重。一般来说,矢量地图由孤立结点、线结点和区域结点复合而成^[11]。孤立结点的属性熵值最大,线路中端点熵值大于其它结点,区域内所有结点的熵值均等。

基于网格密度的空间聚类分析的基本思路是:

- (1) 移除渲染图层,矢量空间简化为拓扑空间;
- (2) 根据聚类空间规模,设定网格,输入网格步长,用式

$$D_{avg} = (n + l + c) / 3 \times (N/M) \quad (1)$$

计算聚类空间的平均属性密度阈值。 D_{avg} 为平均属

① 国家自然科学基金(60873138)和国家博士后基金(20060390461)资助项目。

② 男,1981 年生,博士生,讲师;研究方向:信息安全,数字水印;联系人,E-mail:sunjianguo@hrbeu.edu.cn
(收稿日期:2008-09-23)

性密度, n, l, c 分别代表孤立结点、线结点和区域结点的属性熵值, N 为聚类空间规模, M 为网格数。

(3) 结合拓扑空间聚类结果(如图 1(b)所示), 生成候选嵌入目标集。表 1 所示为算法与文献[12]中 K-means 算法的聚类结果比较。当两种算法对某个样本的集簇分配不同时, 称该样本为差异样本。结果显示, 二者差异较小。

(4) 对每个集簇, 选取密度差异值最小的结点为密度中心, 密度差异值的定义为

$$Mavg = | A_i \times D_i - \sum_{j=1}^k (A_i \times D_j - Davg) / K \times S^2 |$$

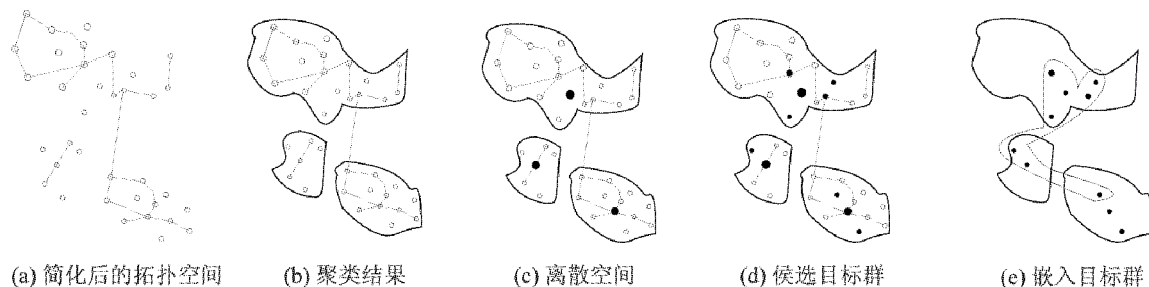


图 1 水印嵌入目标群生成示意图

表 1 本文算法与 K-means 聚类结果对比

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Total
样本总数	910	870	620	1011	768	834	697	881	813	948	8352
差异样本	30	42	36	44	28	42	26	34	40	62	384
差异度	3.3%	4.82%	5.8%	4.36%	3.65%	5.04%	3.74%	3.86%	4.92%	6.54%	4.6%

1.2 水印嵌入目标群

水印容量决定嵌入目标群的规模。一个目标结点至少能嵌入一比特的水印编码, 所以对于容量为 k 的水印编码, 至多需要 k 个目标结点, 而候选目标群的规模应为 $n, n > k$ 。算法的主要思路是:

(1) 在由拓扑关系诱导的 Hilbert 空间^[13]内, 按照距离度量方式, 计算集簇内每个结点到密度中心的距离, 描述如下:

$$Distance(x, C) = A_x \times \sqrt{x^2 + C^2} \quad (3)$$

式中 x 为考察结点, C 为密度中心, A_x 为考察结点的属性熵值。Distance(x, C) 为评价标准, 距离密度中心越小表明该结点被选定概率越大(如图 1(d)所示)。

(2) 确定目标群的初步规模。在各集簇内, 平均选取一定数量距离值较小的结点组成候选目标群。

(3) 根据 Distance(x, C) 值, 不断剔除候选目标群中的结点并调整群的密度中心, 最终确定水印嵌入目标群, 其中, 目标群的元素个数等于水印编码的

(2) 式中 $Mavg$ 为密度差异, D_i 为某集簇的第 i 个结点, A_i 为结点的属性熵值, $Davg$ 为平均属性密度, K 为该集簇的结点数, S 表示网格步长。

图 1(c)所示的黑色圆点即为各集簇的密度中心, 每个集簇只有一个密度中心。根据各结点到密度中心的相对距离, 从每个集簇中选择一定规模的优化(距离密度中心较近)结点做为嵌入目标集的元素。

结点选取的规模直接决定水印容量, 反之, 可根据水印容量扩充嵌入目标集的元素数目。

位数(如图 1(e)所示)。

2 基于属性特征的数字水印算法

2.1 水印标识嵌入

由于利用矢量对象的属性信息来承载水印标识, 水印标识必然写入地图的属性文件中。算法对结点坐标值采取“0 比特复合”的微调操作来完成水印嵌入, 描述如下:

(1) 对于结点 (x_i, y_i) , 为保证矢量对象属性信息及地图属性文件内容的完整性, 约定 (x_i, y_i) 至多复合 4 比特编码。

(2) x_i 坐标表示水印编码类型, 即嵌入的是水印编码为[0]或[1]; y_i 表示相应类型编码的嵌入数量。 (x_i, y_i) 只嵌入连续的同一类型的编码。

以符号“ $\circ$ ”表示编码的连接复合操作, 若 $x_i = x_i \circ 0$, 则 $y_i = y_i \circ "0_1 \cdots 0_k"$; 若 $x_i = x_i \circ 00$, 则 $y_i = y_i \circ "1_1 \cdots 1_k"$ 。其中, k 为 y_i 嵌入 x_i 所约定类型的编码数量($k \leq 4$)。

图2为一个6比特水印利用3个结点进行嵌入的过程示意图。显而易见,对于相同的连续的编码可以使用一个结点来嵌入,且至多嵌入4位编码,超过4位则需嵌入下一个结点 (x_{i+1}, y_{i+1}) 。

	110100	$(x_p, y_p) = (x_p, y_p) = (x_k, y_k) = (X.xxxxxxxxxx, Y.xxxxxxxxxx)$
a.	110100	$(x_p, y_p) = (X.xxxxxxxxxx00, Y.xxxxxxxxxx00)$
b.	110100	$(x_p, y_p) = (X.xxxxxxxxxx0, Y.xxxxxxxxxx0)$
c.	110100	$(x_p, y_p) = (X.xxxxxxxxxx00, Y.xxxxxxxxxx0)$
d.	110100	$(x_p, y_p) = (X.xxxxxxxxxx0, Y.xxxxxxxxxx00)$

图2 矢量结点的水印标识嵌入操作

(3) 将结点序列存储为向量 $N_0 = [n_1, n_2, \dots, n_i]^T$, 并将 N_0 转换为二进制应用 Arnold 置乱:

$$N' = \text{Arnold}(N_0, k) \quad (4)$$

式中 N' 为置乱后输出编码, Arnold 置乱为

$$x' = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} x \pmod{N} \quad (5)$$

x' 和 x 分别是置乱前后的编码序列, 序列长度为 N 。
 k 为迭代次数, 作为水印检测密钥保存。

水印嵌入过程如图3所示。

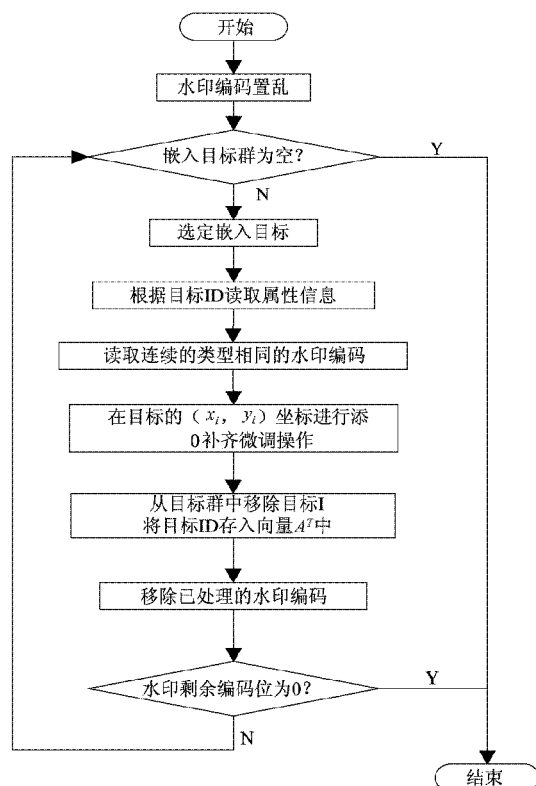


图3 水印嵌入流程

很明显,从实现方式可以认定:“0 比特复合”的嵌入方式对地图精度是零损失的。

2.2 水印信息检测

水印载体是矢量地图的属性文件,所以水印信息检测是对属性文件中对应结点属性的一次遍历。由密钥 k 及置乱后的二进制字符串 N' , 可重新生成向量 $N_0 = \text{Arnold}(N', T - k)$, 根据 N_0 所保存的结点 ID 序列, 遍历属性空间可检测到水印编码 W_t 。当携带水印编码的地图遭受攻击后, 一些结点可能被去除, 水印编码会出现不一致情况, 为此, 提出水印相似度计算方法

$$\text{Sim}(W, W_t) = \sum_{i=1}^n \langle W_i, W_{t_i} \rangle / N \quad (6)$$

进行版权判定。如图4所示, W 为原始水印, W_t 为检测出的水印。 N 为水印容量, 当对应位置数值相同时, $\langle W_i, W_{t_i} \rangle = 1$; 若 W_{t_i} 为空, 或数值不同, 则 $\langle W_i, W_{t_i} \rangle = 0$ 。

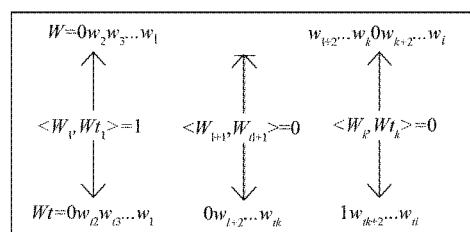


图4 水印相似度计算流程

3 实验及分析

基于 MAPX 插件^[14]的 VC.NET 应用程序测试表明, 使用携带水印的属性文件对地图精度及应用无任何影响, 对地图精度无损。

3.1 鲁棒性分析

对于水印 W_0 , 嵌入率界定于 40% 到 60% 之间, 本文的嵌入率计算采取水印编码与载体编码的比值形式。

在同等嵌入率的情况下, 对应用不同水印算法的矢量地图进行鲁棒性实验测试。实验内容主要包括: 数据压缩攻击^[15], 压缩比例为 40%; 剪切攻击 (如图5所示), 扭曲变形, 噪声攻击 ($\alpha = 50$)。

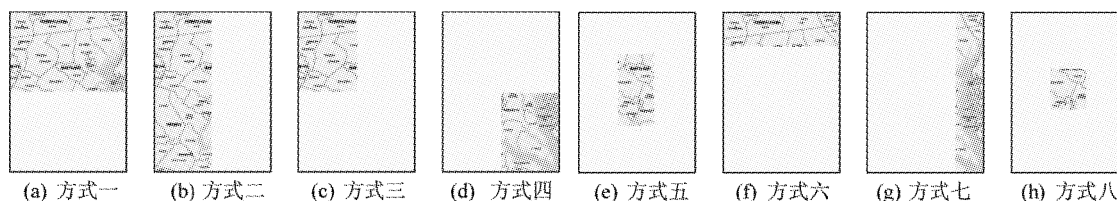


图 5 八种剪切模式

评价指标为误码率,被定义为在水印的实现过程中,接收的信息发生错误的概率。

实验结果如表 2 所示。本文算法的各项指标都

优于文献[2]空域和文献[5]DFT 频域水印算法。特别是对于数据压缩攻击,本文算法的误码率约为空域算法的 1/10,频域算法的 1/8。

表 2 各种攻击下本文算法与同类算法的比较

攻击类型	空域算法误码率 [文献 2](%)	DFT 算法误码率 [文献 5](%)	本文算法误码率 (%)
数据压缩 40%	69.2	48.1	8.6
方式 1	12.5	2.2	1.1
方式 2	14.2	5.4	1.7
方式 3	22.8	7.2	4.5
方式 4	25.1	8.2	6.1
方式 5	42.6	12.4	10.6
方式 6	44.6	18.8	12.5
方式 7	79.3	29.8	18.3
方式 8	81.2	34.3	17.1
变形扭曲	47.8	34.6	0
噪声 ($\alpha = 50$)	28.3	14.6	0

3.2 安全性分析

保证水印信息不被非法提取和检测,关系到水印的实际应用效果。应将所有嵌入水印信息的矢量结点 ID 序列添加到向量 V 中,并将向量 V 转化为二进制字符串,采用 Arnold 置乱的方式进行数据加密。

对于编码长度为 l 的水印标识,攻击者在截获到密钥信息 A^T 的情况下,其付出的攻击代价为 $2^N \times k! \times l!$ 。而在正常情况下,我们只需迭代 $2(T - k)$ 次就可以得到原始的水印信息,可以 $2^N \times k! \times l! \gg 2(T - k)$,在这样高昂的攻击代价下,水印的安全性得到了保证。

3.3 水印容量分析

本文认为,水印容量本质上是一种弹性容量。由水印编码的长度来调整水印嵌入目标群的规模,以此达到嵌入强度的自适应调整。设每一个矢量结点均至少嵌入一比特的水印信息,则这样的矢量地图必定是强鲁棒的。

4 结 论

本文提出了利用矢量地图的对象属性特征构造水印信息,通过改进的网格空间聚类分析方法确定嵌入水印的矢量对象目标集合的水印算法。实验表明,该水印算法对地图精度无损失,在等量嵌入率情况下,算法能够更好地抵抗矢量数据压缩攻击和几何变换攻击。下一步的研究工作将重点提出一种无密钥参与的盲水印算法,重点提高水印的安全性。

参考文献

- [1] Xu D H, Zhu C Q, Wang Q S. A survey of the research on digital watermark for the vector digital map. *Geomatics World*, 2007, 12(6):42-48
- [2] Ntalianis K S, Doulamis A D, Doulamis N D, et al. An automatic scheme for stereoscopic video object-based watermarking using qualified significant wavelet trees. In: *Proceedings of the IEEE 2002 International Conference on Image Processing*, New York, United States, 2002. 501-504

- [3] Thoen W, Bill H. GIS and steganography part3: vector steganography. <http://www.directionsmag.com>; Directions Magazine, 2007
- [4] Solachidis V, Nikolaidis N, Pitas I. Fourier descriptors watermarking of vector graphics images. In: Proceedings of the International Conference of Image Processing, Vancouver, Canada, 2000. 9-12
- [5] Voigt M, Yang B, Busch C. Reversible watermarking of 2D-vector data. In: Proceedings of the 2004 Multimedia and Security Workshop on Multimedia and Security, Magdeburg, Germany, 2004. 160-165
- [6] 李媛媛, 许录平. 矢量图形中基于小波变换的盲水印算法. 光子学报, 2004, 33(1): 97-100
- [7] Deng S J, Lu L. Research on a digital watermark algorithm suitable to vector map. In: Proceedings of the IEEE International Conference on Automation and Logistics, Jinan, China, 2007. 1236-1240
- [8] 马桃林, 顾种, 张良培. 基于二维矢量数字地图的水印算法研究. 武汉大学学报(信息科学版), 2006, 31(9): 792-294
- [9] 王博. GIS系统中基于网格密度的空间聚类算法的研究与应用: [硕士学位论文]. 大连: 大连理工大学, 2005. 19-25
- [10] Shuh E, Hiroshi M, Ryutarou O. Development of digital watermarking technology for vector digital maps: [technical report]. Tokyo: International Palaeontological Association, 2001. 12-20
- [11] 王涛. 坐标几何中的水印隐藏算法设计与实现: [硕士学位论文]. 北京: 北京邮电大学, 2003. 40-43
- [12] Bagirov A M. Modified global k-means algorithm for minimum sum-of-squares clustering problems. *Pattern Recognition*, 2008, 41(10): 3192-3199
- [13] Tannor D J, Sklarz S E. Quantum computation via local control theory: direct sum vs direct product Hilbert spaces. *Chemical Physics*, 2006, 332(2): 87-97
- [14] Chen L S, Zhu L, ShangGuan W. The research and design of vehicle surveillance system base on MapX. In: Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications, Harbin, China, 2007. 2040-2044
- [15] 胡艳. 矢量地图数据无损压缩算法的研究: [博士学位论文]. 北京: 解放军信息工程大学, 2007. 34-37

A digital watermarking algorithm based on attribute features of 2D vector maps

Sun Jianguo, Men Chaoguang

(College of Computer Science and Technology, Harbin Engineering University, Harbin 150001)

Abstract

Aiming at the problems that spatial domain watermarking and frequency domain watermarking of 2D vector maps cause precision losses and have the poor performance of anti-compression, this paper proposes a new digital watermarking algorithm for 2D vector maps. The algorithm uses the attribute-features of vector objects to construct watermarking information, selects characteristic nodes with the improved grid density clustering method, and embeds the watermark into the attribute files by tuning the coordinates of characteristic nodes. The experiments and analysis show that the algorithm can achieve the zero-loss goal in the precision of vector maps and enhance the robustness and availability of watermarking.

Key words: watermarking, vector, map, feature