

针对二维矢量图形数据的盲水印算法

陈晓光¹, 李岩^{1,2}

(1. 华南师范大学 计算机学院, 广州 510631; 2. 华南师范大学 空间信息技术与应用研究中心, 广州 510631)

(jackytiger1986@163.com)

摘要: 矢量数字水印是图形数字产品和二维矢量地图知识产权保护的重要手段之一。提出了一种针对二维矢量图形和二维矢量地图数据的盲水印算法。首先通过遍历矢量图形数据, 动态获取容差值, 并利用道格拉斯—普克算法提取出二维矢量图形数据中曲线的所有特征点, 然后在容差值范围内对特征点嵌入水印信息。提取水印时, 通过算法的逆变换过程, 可以获得嵌入的水印信息。通过对嵌入水印后的矢量图形进行随机增加、删除坐标点, 以及矢量图形的裁剪、坐标点的压缩攻击实验, 分析提取出的水印图像的相关系数, 结果证明所提的算法具有较强的鲁棒性。

关键词: 二维矢量地图; 数字水印; 精度测试; 鲁棒性; 相似度

中图分类号: TP301.6; TP391.41 **文献标志码:** A

Blind watermarking algorithm for 2D vector map

CHEN Xiao-guang¹, LI Yan^{1,2}

(1. School of Computer, South China Normal University, Guangzhou Guangdong 510631, China;

2. Research Center of Spatial Information Technology and Application, South China Normal University, Guangzhou Guangdong 510631, China)

Abstract: The vector digital watermark is one of the most important means of copyright protection for graphics and vector maps. The authors discussed a blind watermarking method for 2D vector map. First, the entire vector map was traversed to get the tolerance dynamically, then the classical Douglas-Peucker algorithm was used to get the entire feature nodes from vector map, and finally, the watermark was embedded into feature nodes in tolerance range. And with the inversed procedure of the embedding process, the watermark could be extracted. Through the attack method including random adding points, random deleting points, compression and cropping, the correlation coefficient of original watermark bits and the extracted watermark bits from attacked watermarked maps were calculated. The experimental results show that the proposed method has a great robustness.

Key words: 2D vector map; digital watermark; accuracy test; robustness; similarity

0 引言

随着空间信息共享的理念逐渐形成共识, 二维图形或矢量地图数据作为主流数据格式之一, 其版权保护的研究也逐渐成为研究热点。

有关二维矢量地图数据的水印算法, 根据水印嵌入的方法特点及位置差异, 分为空域算法和变换域算法。

空域矢量数据水印算法常采用直接修改矢量数据的坐标顶点值或添加新的冗余坐标点来嵌入水印。变换域算法中水印数据不是通过直接修改顶点坐标嵌入的, 而是先从矢量图中选取坐标点的坐标, 并按一定的规则进行排序; 然后, 再进行域变换, 将水印嵌入到它们的变换系数中。Cox 等人^[1]提出的矢量数据水印算法, 该算法把水印信息直接嵌入到矢量数据各顶点坐标上, 是一种脆弱的空域水印算法, 不能抵抗任何简单的攻击; 文献[2]中提出了一种地理信息系统 (Geographic Information System, GIS) 领域中的水印模型, 但并没有针对此模型作相关的实验; 文献[3-4]中通过先细分矢量图, 然后在细分的矢量图中添加水印; 文献[5]中提出的算法, 并不能抵抗普遍的攻击方式; 文献[6]利用 B-spline 曲线, 在控制点处加入水印, 但是在嵌入水印后有可能会影响图形的精度而降低其可用性。

本文针对二维矢量地图数据精度要求较高且需要在实际应用中以较快捷有效的方式检测出水印信息的特点, 提出了一种针对二维矢量地图数据的盲水印算法。以往算法中对于嵌入水印后引起的图形形变, 都只是提出感官描述, 本文针对此问题提出了精度分析。

1 水印算法

1.1 水印的嵌入过程

在对矢量地图进行水印嵌入之前, 需要选择水印文件生成水印信息, 然后对矢量地图数据进行预处理。

1.1.1 水印信息的生成

水印序列就是一串数字序列, 既可以采用彩色图也可以采用二值位图来生成这串数字序列。但由于彩色图所包含的信息更为丰富, 所生成的水印序列相对会较长。如图 1(a) 所示为大小为 60×40 的二值图像, 图 1(b) 为大小为 60×40 的 256 色图像。

同样大小的图像, 所包含的色彩丰富程度不同, 同样, 所生成的水印序列长度也不同。它们对应生成的水印序列长度如表 1。

从表 1 中可以看出, 同样大小的 256 色位图所产生的水印序列长度, 比二值位图产生的水印序列长度要长很多。

收稿日期: 2011-02-25; 修回日期: 2011-04-21。

作者简介: 陈晓光 (1986-), 男, 湖北咸宁人, 硕士研究生, 主要研究方向: 空间信息技术与应用; 李岩 (1955-), 女, 天津人, 教授, 主要研究方向: 空间信息技术与应用、图形图像处理。

为了保证对于一些数据量较小的图形或矢量地图也能成功完整地嵌入水印,本文采用二值图像生成研究中的水印序列。首先,通过对二值图像进行解析,获取图像的矩阵值;然后,将每一个像素点的值转换为 $\{0,1\}$ 序列 α_i 。

根据式(1)得到水印信息 $w = (w_1, w_2, \dots, w_n)$,其长度记为 M :

$$w_i = \begin{cases} 1, & \alpha_i = 1 \\ -1, & \alpha_i = 0 \end{cases} \quad (1)$$



图1 60×40的二值图像和256色图像

表1 不同的图片格式生成的水印序列长度对比

文件格式	水印序列长度
二值图像	2400
256色图像	19200

1.1.2 特征点的提取

前人所采用对点进行偏移或添加点的方法均不适用于这些高精度矢量地图数据的要求,因此为了保证矢量图形或矢量地图的精度,本文采用道格拉斯—普克法^[7]。它的基本思想是:将曲线的起点和终点连接,找出曲线上离直线距离最近的点,若该最大距离大于限差,则将该点保留,并以该点为界,将曲线分成两部分,对这两部分重复使用该方法。最后,就会找到曲线中的所有特征点,按照点的先后顺序,排列成一个特征点序列。

假设在矢量图中,其中一条曲线中坐标点的存放顺序为从左至右,分别为 $P(x_1, y_1), P(x_2, y_2), \dots, P(x_n, y_n)$,其中起始点 $A = P(x_1, y_1)$,终点 $B = P(x_n, y_n)$,则在曲线中寻找特征点的过程如图2所示。

1) 首先连接首尾顶点组成线段 AB ,找到离线段 AB 最近的点 C ,如图2(a);

2) 判定点 C 离线段 AB 的垂直距离是否大于阈值。若否,则需将终点设定为 $B = C$,然后连接点 A, B ,寻找离新线段 AB 垂直距离最近的点 C ,回到步骤1),如图2(b);

3) 判断 B 是否等于 $P(x_n, y_n)$,若否,则将点 A 放入特征点序列中,同时,设定 $A = C, B = P(x_n, y_n)$,连接线段 AB ,并回到步骤1),如图2(c)所示;

4) 将 A, B 提取为特征点,结束。

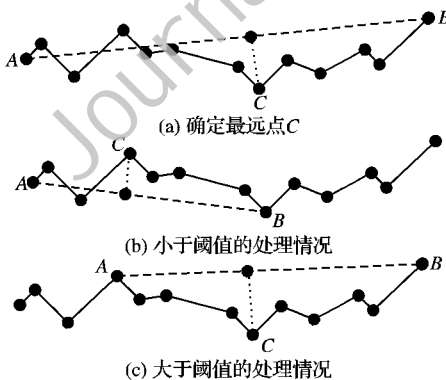


图2 寻找特征点示意图

这里,阈值的选择和矢量地图的比例尺有关,在该阈值内找到的特征点,不仅对于图形形变的控制,具有非常重要的作用,而且通过这些特征点上添加水印的方式,还能有效抵抗压缩攻击。通过这种方式,将矢量地图中每条曲线中的特征点都提取出来,组成特征点序列。

1.1.3 动态容差的选定

本文将通过遍历的方法来动态地获取矢量数据的容差值。通过对二维矢量地图数据的分析,发现这些数据中既有描述坐标系中具体位置的绝对坐标,也有描述相对前一个坐标点的相对坐标。为了准确地分析这些数据,本文算法在进行遍历的过程中,将这些数据变换为绝对坐标,找到所有坐标点中的绝对坐标的最小值 T_{\min} ,然后通过式(2)来确定容差值,记为 g :

$$g = \frac{1}{1000} \left(T_{\min} - \frac{T_{\min}}{1000} \right) \quad (2)$$

1.1.4 水印信息的嵌入

首先,对二维矢量地图每条曲线中提取出的曲线特征点进行排列,若每条曲线中提取出的特征点个数为 n ,用 $v = \{v_1, v_2, \dots, v_n\}$ 来表示这些特征点的坐标值。

为了增强水印算法添加删除点的鲁棒性,根据生成的水印序列的位置信息,对特征点进行分组,采用 c (本文中 c 值约定为4)种水印嵌入策略。

在矢量地图中任意一条曲线中的第 k 个特征点 (vx_k, vy_k) 嵌入水印时的公式如式(3)所示:

$$\begin{cases} vx'_k = vx_k - vx_k \% g + policy_x(k, g) \\ vy'_k = vy_k - vy_k \% g + policy_y(k, g) \end{cases} \quad (3)$$

其中: (vx_k, vy_k) 是嵌入水印之前的特征点的坐标值, (vx'_k, vy'_k) 是嵌入水印后的特征点坐标值,具体的嵌入策略 $policy$ 由特征点的位置信息以及容差值确定。在本文方法中,水印序列信息位的位置信息和特征点的位置信息一一对应。表2中, $policy_x$ 表示针对坐标点 x 轴上的嵌入策略, $policy_y$ 表示针对坐标点 y 轴上的嵌入策略。

表2 具体嵌入策略 $policy_x, policy_y$ 确定方式

水印序列信息位 w_i	$policy_x$	$policy_y$
-1	$0.5g \cdot \{1 + \sin[(i\%c)\pi/10]\}$	$0.5g \cdot \{1 + \cos[(i\%c)\pi/10]\}$
1	$0.25g \cdot \sin[(i\%c)\pi/10]$	$0.25g \cdot \cos[(i\%c)\pi/10]$

例如,在一条曲线中提取出的特征点序列长度为500,其中,第50个特征点的坐标值为(5320.231, 2543.764),且此时水印序列信息位为 $w_i = -1$,容差 $g = 1.0$ 时,则水印嵌入后,该点的坐标值为(5320.655, 2543.976)。同样,若此时水印序列信息位为 $w_i = 1$,容差 $g = 1.0$ 时,水印信息嵌入后,该点的坐标值为(5320.077, 2543.238)。利用这种方法,对曲线中的所有特征点依次嵌入水印信息。

同时,必须考虑到第 i 条曲线上特征点个数 n 与水印序列长度 M 之间的关系,若 $M > n$,则需将此条曲线先搁置,直到遍历到下一条符合此条件的曲线时,一并嵌入水印。

1.2 水印的提取过程

由于嵌入水印过程中,对矢量图形中的坐标点进行偏移时,本文算法首先通过在容差范围内的偏移,因此,矢量图形中的特征点并没有发生改变。当提取水印信息时先用Douglas-Peucker法对矢量图形中数据进行特征点提取,将提取出的特征点组成一个序列;然后利用嵌入水印过程中的逆过程从这些特征点中提取出水印信息。

提取水印时,确定提取出来的水印序列,如式(4)如下:

$$\begin{cases} 0.5g < (vx'_k - vx_k/g) < g \\ 0.5g < (vy'_k - vy_k/g) < g \end{cases} \Rightarrow w_i = -1 \quad (4)$$

$$\begin{cases} 0 < (vx'_k - vx_k/g) < 0.25g \\ 0 < (vy'_k - vy_k/g) < 0.25g \end{cases} \Rightarrow w_i = 1 \quad (5)$$

嵌入水印后的矢量地图中的特征点的坐标值满足式(4)中的不等式时,则该点提取出来的水印值设定为-1,满足式(5)中的不等式时,则该点提取出来的水印值设定为1。其中: (vx'_k, vy'_k) 为特征点序列中第 k 个特征点的坐标值, g 为容差值。

然后将提取出来的水印数据重新分组,对应的转换为图像数据中的矩阵,然后根据矩阵数组生成相应的图像文件。将水印序列转换为图像矩阵的转换公式如下:

$$\alpha_i = \begin{cases} 1, & w_i = 1 \\ 0, & w_i = -1 \end{cases} \quad (6)$$

其中 α_i 表示要生成的图像矩阵中的对应每个像素的像素值。

1.3 水印信息的检测

水印信息的检测是在水印信息提取后进行的,在水印的嵌入过程中,针对矢量地图中的每一条曲线都进行了水印的嵌入,因此,在进行水印检测时,会有针对每一条曲线中提取出的水印信息的检测。通过取得最优相关系数,获得最有利的检测结果。相关系数 $cor(m, \bar{m})$ 的计算公式如式(7)所示:

$$cor(m, \bar{m}) = \frac{\sum_{i=0}^{n-1} (m_i \cdot \bar{m}_i)}{\sqrt{\sum_{i=0}^{n-1} (m_i^2)} \cdot \sqrt{\sum_{i=0}^{n-1} (\bar{m}_i^2)}} \quad (7)$$

其中: \bar{m}_i 是从加水印后的图形中提取出来的第 i 位水印信息, m_i 表示原始水印序列中第 i 位水印信息。通常,相关系数 cor 取阈值0.75。

2 实验仿真及精度分析

2.1 鲁棒性测试

为了验证本文中提出的水印算法的性能,本文选取了广东省1:25万比例尺的矢量地图数据进行实验,该图中总共有坐标点66761个。同时,为了对水印的鲁棒性进行测试,本文对矢量地图数据做了增加数据点、删除数据点、裁剪以及压缩等攻击方式的测试。

图3(a)是未嵌入水印之前的广东省1:25万矢量地图数据;图3(b)是用来生成水印序列的水印文件。



图3 广东省1:25万矢量地图数据及水印文件

利用本文中提出的水印算法流程对广东省1:25万矢量地图数据进行水印嵌入之后的矢量地图数据如图4所示。

图5(a)是随机增加了5000个数据点后的矢量地图数据,图5(b)是从随机增加5000个坐标点后的矢量地图数据中提取出来的水印图像。

图5(c)是随机删除5000个坐标点后的矢量地图数据,图5(d)是从删除了5000个坐标点后的矢量地图数据中提取出来的水印图像。

图5(e)是对水印图裁剪50%后的矢量地图数据,图5(f)是从裁剪后的矢量地图数据中提取出来的水印图像。

图5(g)是进行压缩处理后的矢量地图数据,图5(h)从

压缩后的矢量地图数据中提取出来的水印图像。



图5 各种攻击方式后的矢量地图数据和提取的水印图像

由于水印嵌入过程中,是针对矢量地图中的每一条符合条件的曲线都嵌入水印,因此,在获取相关系数的过程中,可以选择计算结果中的最优值作为水印提取的结论。通过计算检测提取出来的水印图像和原始水印图像之间的矩阵相关系数,如表3所示。

可以看出,算法有较强的抗攻击能力。通过随机增加5000个坐标点,生成水印的原始图像和增加5000个坐标点之后提取出的结果图像之间的相关系数达到0.855,而随机

删除5000个坐标点时,该相关系数的值能达到0.919。这主要是由于采用随机的方式添加或删除水印点,且水印算法在提取过程中,单个特征点的水印提取与该点在特征点序列中的位置关系有着非常重要的联系,因此,通过随机的方式来增删坐标点,会产生一些差异,然而,由于水印的嵌入,是在所有符合条件的曲线中重复嵌入的,这种随机性也就会随着重复的嵌入而消失,也就是说,对于数据量越大的矢量图形,水印在提取过程中,就越能取得较好的实验效果。同样,对于裁剪和压缩攻击,算法也具有较好的实验效果。在不同的攻击方式下,基本上都能正常地提取出水印信息。通过相关系数的计算结果发现,各种攻击方式下,该系数的值均能高于经验阈值0.75。

表3 生成水印的原始图像和提取的结果图像相关系数

攻击方式	相关系数
随机增加5000个坐标点	0.855 120 732 722 731 0
随机删除5000个坐标点	0.918 817 651 956 702 7
裁剪50%	0.875 104 079 933 388 9
压缩	0.955 870 108 243 130 7

2.2 不可感知性分析与精度分析

图形形变包含了两部分内容,一方面需要保证图形在嵌入水印之后,原始图形和嵌入水印后的图形在主观的视觉差异和客观的数值计算上具有较小的差异,另一方面需要保证坐标点的改变是在该图形所对应比例尺的精度范围内。

1) 不可感知性分析。

主观的视觉差异分析通常会受到测试者的身份、经验、年龄等严重影响,因此,本文通过客观的数值计算的方式来验证。本文将引入数字图像水印领域中的保真度(image Fidelity, IF)计算公式如式(8)所示:

$$IF = 1 - \frac{\sum_{i=0}^{N-1} [(x_i - x'_i)^2 + (y_i - y'_i)^2]}{\sum_{i=0}^{N-1} (x_i^2 + y_i^2)} \quad (8)$$

其中:\$(x_i, y_i)\$ 为原始坐标点, \$(x'_i, y'_i)\$ 为水印嵌入后图形中的坐标点, \$N\$ 为坐标点的个数。经计算后求的嵌入水印后的图形保真度达到0.9999999999997772。经过主观的视觉差异测试和数值计算,表明该算法的不可感知性良好。

2) 精度分析。

通过嵌入水印后的矢量地图和原始矢量地图之间的数值变化来进行精度的评价,通过计算矢量图形中每个点的平均偏移量,如式(9)所示:

$$\bar{E} = \frac{1}{M} \sum_{i=0}^{M-1} \sqrt{(y'_i - y_i)^2 + (x'_i - x_i)^2} \quad (9)$$

其中:\$(x_i, y_i)\$ 为原始坐标点, \$(x'_i, y'_i)\$ 为对应的嵌入水印后的坐标点。该图中总共有点66761个,平均偏移量为0.5597657744167763。由计算结果可知,嵌入水印后的图形,其精度是有保证的。

3 结语

本文提出了一种针对二维矢量地图数据的盲水印算法,通过对矢量地图中的特征点提取,然后在特征点处嵌入水印信息。实验结果表明,算法具有较好的鲁棒性,对裁剪、压缩等攻击都有较好的抗攻击能力,同时,对于数据精度也有一定的保证。然而,本文算法只能证明水印文件中加入了水印,却不能防止未授权者通过伪造原始文件和原始水印来迷惑版权保护,即对IBM攻击存在隐患,这也是下一步研究的重点。

参考文献:

- [1] COX G S, de JAGER G. A survey of point pattern matching techniques and a new approach to point pattern recognition [C]// Proceedings of Symposium on Communication and Signal Processing. Lesotho: [s. n.], 1993: 243 - 248.
- [2] AYBET J, AL-SAEEDY H, FARMER M. Watermarking spatial data in geographic information systems [C]// Proceedings of the 5th International Conference on Global Security, Safety, and Sustainability. Heidelberg: Springer, 2009: 18 - 26.
- [3] OMBUCHI R, UEDA H, ENDOH S. Robust watermarking of vector digital maps [C]// ICME 2002: IEEE International Conference on Multimedia and Expo 2002. Piscataway, NJ: IEEE Press, 2002: 26 - 29.
- [4] 王超,王伟,王泉.一种空间域矢量地图数据盲水印算法[J].武汉大学学报:信息科学版,2009,34(2):163 - 165.
- [5] VOIGT M, BUSCH C. Watermarking 2D-vector data for geographical information systems [C]// Proceedings of the Society of Photographic Instrumentation Engineers, Security and Watermarking of Multimedia Content. Bellingham, WA: SPIE, 2002: 621 - 628.
- [6] GOU HONGMEI, WU MIN. Data hiding in curves with application to fingerprinting maps [J]. IEEE Transactions on Signal Processing, 2005, 53(10): 3988 - 4005.
- [7] DOUGLAS D, PEUCKER T. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature [J]. The Canadian Cartographer, 1973, 10(2): 112 - 122.
- [8] 李媛媛,许录平.用于矢量地图版权保护的数字水印[J].西安电子科技大学学报,2004,31(5):719 - 723.
- [9] 张鸿生,李岩,曹阳.一种采用曲线分割的矢量图水印算法[J].中国图象图形学报,2009,14(8):1516 - 1522.
- [10] 王忠军,王玉海,王豪.一种鲁棒的矢量地图数字水印算法[J].测绘科学,2008,33(4):148 - 150.

(上接第2158页)

- [6] YU L, LIU H. Efficient feature selection via analysis of relevance and redundancy [J]. Journal of Machine Learning Research, 2004, 5(10): 1205 - 1224.
- [7] BAGLIOMI M, FURLETTI B, TURINI F. DrC4.5: Improving C4.5 by means of prior knowledge [C]// Proceedings of the 2005 ACM Symposium on Applied Computing. New York: ACM, 2005: 474 - 481.
- [8] 苟光磊,黄丽丰,邹航.基于信息增益的属性约简算法[J].重庆工学院学报,2007,21(6):90 - 92.
- [9] HALL M A. Correlation-based feature selection for discrete and numeric class machine learning [C]// Proceedings of the 17th International

Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers, 2000: 359 - 366.

- [10] Information and Computer Science University of California. Irving KDD cup 1999 data [EB/OL]. [2010 - 09 - 20]. <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>.
- [11] 罗敏.基于聚类和支撑向量机的网络入侵检测研究[D].武汉:武汉大学,2003.
- [12] 何慧,苏一丹,覃华.基于信息增益的贝叶斯入侵检测模型优化的研究[J].计算机工程与科学,2006,28(6):3288 - 3290.
- [13] 令狐红英,陈梅,王翰虎,等.基于互信息可信度的贝叶斯网络入侵检测研究[J].计算机工程与设计,2009,30(14):38 - 40.