doi:10.3969/j.issn.1672－0504.2019.06.004

抗拼接攻击的矢量瓦片数据水印算法

唐 伟
(浙江省测绘资料档案馆,浙江 杭州 311100)

摘要:网络环境下矢量瓦片数据应用广泛,引发多用户、多线程的非法爬取等安全问题。该文针对网络环境矢量瓦片数据版权保护和用户追溯的需求,提出了一种抗拼接攻击、鲁棒性强的矢量瓦片水印算法。首先,在水印信息生成阶段对两段编码进行设计,在保证区分度的同时,提供尽可能多的用户编码;其次,在水印嵌入和检测阶段,改进了经典的量化索引调制方法,根据矢量瓦片的精度划分更多数量的区间,以提高水印容量。实验结果表明,该文提出的矢量瓦片水印算法能够保证数据的可用性,并且对增加、删除、更新以及多用户的拼接攻击均具有较好的鲁棒性,实现了矢量瓦片的版权保护和用户追溯功能。

关键词:数字水印;矢量瓦片;鲁棒性;拼接攻击
中图分类号:TP309.7 **文献标识码**:A **文章编号**:1672－0504(2019)06－0020－05

0 引言

随着WebGIS和地理信息服务的飞速发展,矢量瓦片数据在GIS行业中的分发、共享和传输愈发多见,在带来便利的同时,矢量瓦片数据共享会引发非法下载、盗版、泄密等安全问题,这些侵权行为不仅打击了版权所有者共享数据的积极性,更有可能危及国家安全,因此,必须采用切实可行的技术措施对矢量瓦片数据进行版权保护[1,2]。

数字水印将版权信息、用户信息等以隐蔽的方式嵌入至载体数据中,实现了分发后数据的水印检测和水印提取[3－8],在地理数据的安全保护中取得了许多研究成果[9－16],但针对矢量瓦片数据的研究相对较少[17,18]。与常见的地理数据相比,矢量瓦片数据承载水印容量小、下载方便,同时具有数据分块、空间拓扑关系等特征[19]。为避免用户对数据的非法大量获取,矢量瓦片数据管理者会限制用户获取的瓦片数量[20],相应地,盗版者会通过多个用户账号爬取瓦片数据,再将其拼接成较大图幅的数据,因此,水印算法能否抵抗这类拼接攻击,是保护矢量瓦片数据版权的关键。

抗拼接攻击的关键是水印信息的生成和水印容量,前者决定了能支持的用户数量及水印提取的稳健性,后者决定了水印算法能否在瓦片数据嵌入完整的水印信息并进行检测。目前,大多数矢量数据水印算法并未顾及这两种特性,也难以直接适用于当前网络环境下的矢量瓦片数据版权保护。鉴于此,本文提出一种支持多用户、大水印容量的矢量瓦片水印算法,提高对拼接以及其他攻击的抵抗能力,以期保护矢量瓦片数据的版权和安全。

1 分段水印信息生成

水印信息的长度决定了支持用户的数量。通常以二进制为基础,二值水印信息长度为L时,支持的用户数量为2^L,当$L\approx30$时,则30位的二值水印信息可支持10亿用户。但在这种情形下,不同水印信息之间的汉明距离最低为1,在水印检测时极易发生误检,影响用户追溯和责任认定的权威性和正确性。因此,水印信息长度应尽可能大,生成的二值水印之间的汉明距离尽可能的高。但考虑到瓦片本身数据量的限制,长度过大的水印信息也难以嵌入至瓦片数据中,因此,需设计适当的水印信息长度参数。

考虑到地理数据的用户有其空间分布特征,不同区域的用户访问数据往往集中在特定的区域。因此,可以将水印信息分成两段:1)第一段代表组编码,可根据用户的地理分布特征、注册IP地址、注册时间等进行分组,此组之间的汉明距离较大,实现组与组之间的明显区分;2)第二段水印信息隶属于该组的用户编码,编码之间的汉明距离较小,从而提供更多的组内用户数。设组编码长度为L_1,用户编码

收稿日期:2019－04－17; 修回日期:2019－07－01
基金项目:国家自然科学基金项目(41971338)
作者简介:唐伟(1984－),女,硕士研究生,高级工程师,研究方向为地理信息分发服务及数据安全管理。E－mail:351212085@qq.com

长度为 L_2,分组编码机制如图 1 所示。

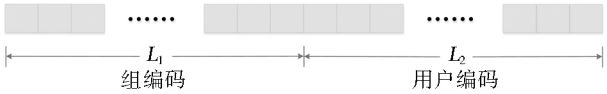


图 1　分组编码机制示意

Fig. 1　Diagram of group-based coding mechanism

由于组编码中组与组之间的汉明距离较大,编码系统存在一定的冗余,因此,在进行水印检测时,可以对检测到的水印信息进行一定程度上的验证。若组编码正确,则可判断检测的水印信息正确;若组编码不正确,则可能有两种情形:1)检测的水印信息不足,即瓦片数据量过低或者水印容量较低;2)瓦片数据遭受到攻击,导致部分坐标点提取的水印信息不正确。在这两种情形下,均可丢弃不正确的错误编码,进一步增加水印检测的正确性。

2　顾及坐标精度的多段量化调制方法

量化索引调制(Quantization Index Modulation, QIM)作为经典的数字水印嵌入方法,具有鲁棒性强、盲检测、运算复杂度低等特性,在矢量地理数据水印中应用较多。设原始数值为 c,量化步长为 l,经典 QIM 嵌入的水印信息 w 分为 0 和 1 两种,嵌入后的数值为 c',则嵌入方法为:

$$\begin{cases} c'=c & c\%l<l/2, w=0 \text{ 或 } c\%l\geqslant l/2, w=1 \\ c'=c-l/2 & c\%l\geqslant l/2, w=0 \text{ 或 } c\%l<l/2, w=1 \end{cases} \tag{1}$$

式中:%为取余运算。可以看出,对于每个数值 c,QIM 嵌入的水印容量为 1 bit,由于每个坐标点有 x 和 y 两个坐标,因此,理论上经典 QIM 的水印容量为每节点 2 bit;而单个矢量瓦片含有的矢量节点较少,往往只有几十个,因此,经典 QIM 难以满足矢量瓦片的水印容量嵌入要求。

为提高水印容量,对经典 QIM 的量化调制区间重新划分。原有的量化区间为两种,对应 $w=0$ 和 $w=1$ 的两种水印信息,现予以划分为 k 段,则每个量化区间对应的水印信息为 $w\in\{0,1,2,\cdots,k-1\}$,此时量化水印嵌入可表示为:

$$c'=c+\frac{l}{k}\times\left(w-\left\lfloor\frac{kc-\left\lfloor\frac{c}{l}\right\rfloor\times kl}{l}\right\rfloor\right) \tag{2}$$

式中:$\lfloor\ \rfloor$为向下取整操作。设提取的水印信息为 w',水印检测方法为:

$$w'=\left\lfloor\frac{kc'-\left\lfloor\frac{c'}{l}\right\rfloor\times kl}{l}\right\rfloor \tag{3}$$

由式(3)可以计算出,当量化区间划分为 k 段时,改进的 QIM 水印容量为每节点 $2\lfloor\log_2 k\rfloor$ bit,例如,当 $k=8$ 时,每节点可嵌入的水印容量为 6 bit。改进的 QIM 水印容量与区间划分数量 k 成正比,因此,增加 k 可以提高 QIM 的水印容量。但因为瓦片数据中节点坐标有其精度特性,故水印容量不能无限制提高。设坐标值的表达形式为:

$$x=a_0.a_1a_2a_3\cdots a_n \tag{4}$$

式中:a_0 表示整数部分,小数部分从 a_1 开始,每位用不同的下标表示。设 a_i 为坐标可以更改的精度位,a_j 为坐标存储的最低有效精度位,则理论上的水印容量极限为 $2(j-i+1)$ bit。考虑鲁棒性和盲检测特性,QIM 每个区间的长度应满足:

$$\frac{l}{k}\gg 10^{-j} \tag{5}$$

考虑到浮点数的一般表达方式,可取:

$$\frac{l}{k}\geqslant 10^{-j+3} \tag{6}$$

为实现水印容量和鲁棒性的平衡,有 $l\leqslant 10^{-i}$,即对节点坐标的量化调制不能超过允许更改的精度位。此时区间数 k 可表示为:

$$k=10^{j-i-3} \tag{7}$$

为便于水印容量计算,k 应为 2 的整数次方,即:

$$k=2^{\lfloor\log_2 10^{j-i-3}\rfloor} \tag{8}$$

对于参数 k 的确定,举例说明如下:设节点坐标可更改精度位 $i=3$,坐标最低有效精度位为 8,由式(8)可知 $k=64$,即 QIM 中量化区间数为 64 个,此时水印容量提高为每节点 8 bit。

3　水印嵌入与检测算法流程

基于上述分段水印信息生成和水印容量提高的 QIM 机制,能够将水印信息较为完整地嵌入至瓦片数据中。矢量地理数据水印算法可分为空域水印算法和变换域水印算法,考虑到瓦片数据集合数量大、单个瓦片数据量小的特点,本文采用空域水印算法对瓦片数据进行水印嵌入。具体流程为:1)基于分段的水印信息生成策略,生成组编码长度为 L_1、用户编码长度为 L_2、总编码长度为 L_1+L_2 的水印信息;2)读取瓦片数据中的坐标点 $V(x,y)$;3)根据当前瓦片数据的层级和应用环境,设定坐标可更改的精度位 i 和最低有效精度位 j 等参数,根据式(8)计算 QIM 划分区间数 k;4)根据式(2)分别对 x 和 y 嵌入水印信息;5)循环嵌入瓦片数据中的坐标点,至该瓦片嵌入完毕时,判断水印信息是否已完整地嵌入该瓦片,未完整嵌入则略去对该瓦片的水印嵌入,降低检测水印时的误检率;6)重复上述过程,直至完成所有的瓦片数据水印嵌入。

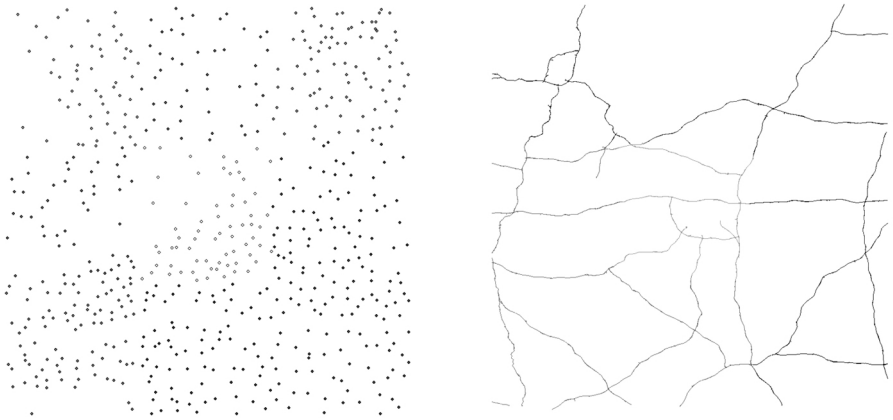
水印检测是水印嵌入的逆过程,根据水印的嵌

入算法，同样在空域中提取水印信息，具体流程为：1）读取待检测瓦片数据中的坐标点 $V(x,y)$；2）与水印嵌入的步骤 3）相同，确定 QIM 的相关参数；3）根据式(3)提取出水印信息；4）检测完当前瓦片后，根据多数原则生成最终的水印信息，判断组编码是否正确，若正确则保留该水印信息，否则跳过该瓦片；5）重复上述过程直至所有瓦片检测完毕，列举出各个瓦片对应的组信息和用户信息，实现版权鉴定和用户追溯。

4　实验结果与分析

4.1　实验概述

为验证本文算法对瓦片数据的有效性，通过实验考察水印算法的检测结果、不可感知性以及鲁棒性。实验采用某省瓦片数据集，为 3×3 瓦片拼接而成的点数据 A 和线数据 B(图 2)，精度要求均为 0.1 m。



(a) 数据A　　(b) 数据B

图 2　实验数据
Fig. 2　Experimental data

水印嵌入时，参数设定为：$L_1=L_2=100$，即水印信息总长为 200 位，此时可提供较多的用户组数目和各组内用户数，能更好地抵抗拼接攻击。根据数据精度，结合式(8)可推得水印嵌入和提取参数 $i=2$、$j=7$，QIM 区间数 $k=64$，采用自相关系数(NC)对提取的水印进行定量评价。设原始水印信息为 w，检测出的水印信息为 w'，NC 的计算公式为：

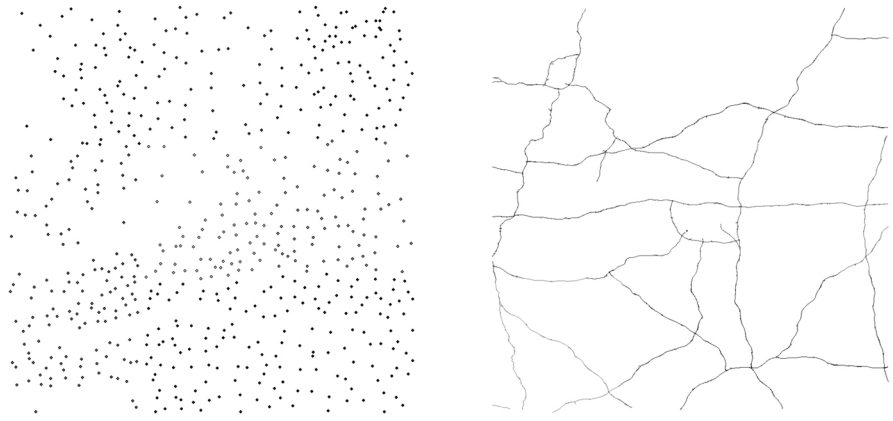
$$NC=\frac{1}{L}\sum_{n=1}^{L}XNOR(w'[n],w[n]) \tag{9}$$

式中：$XNOR$ 为异或操作；$L=L_1+L_2=200$；[]和 n 分别为遍历的水印信息下标索引和索引值。

NC 的经验阈值可设为 0.8，当 $NC\geqslant 0.8$ 时，认为检测的水印信息与原始水印信息相匹配。首先，对数据 A 和数据 B 的瓦片分别运用本文的算法嵌入单一的水印信息，检测结果 NC 均为 1，表明水印信息能够被正确提取，可有效对矢量瓦片数据进行版权保护。

4.2　不可感知性

水印算法的不可感知性是指嵌入的水印主观上难以被察觉，并且不会影响数据的可用性。对于矢量瓦片数据，嵌入的水印不能改变其精度。因此，本文从主观和客观两方面考察水印算法的不可感知性，嵌入水印后的数据A′和数据B′如图3所示。对



(a) 数据A′　　(b) 数据B′

图 3　含水印数据
Fig. 3　Watermarked data

比图 3 和图 2 可知,嵌入水印的数据与原始数据相比并无视觉上的差异。为进一步客观评价水印算法的不可感知性,采用统计的方式度量嵌入水印的数据误差(表 1),可以看出,误差的最大值不超过数据的精度要求,且平均值和标准差均保持在较低水平,表明水印的嵌入不会影响矢量瓦片的可用性。这是因为水印嵌入时考虑了坐标的精度位,只对允许精度变化的位数进行了修改,从而保证了矢量瓦片数据的精度不受水印嵌入的影响。

表 1　误差统计
Table 1　Error statistics　　m

嵌入水印数据	数据精度	误差最大值	误差平均值	误差标准差
数据 A	0.1	0.0632	0.0177	0.0137
数据 B	0.1	0.0546	0.0131	0.0196

4.3　鲁棒性

鲁棒性是水印抵抗各类攻击的能力,也是评价水印算法稳健与否的重要指标。对于矢量瓦片,增加、删除、更新和拼接攻击是常见的攻击方式,因此,鲁棒性实验主要考察水印算法对这些攻击的抵抗能力。以数据 B′为例,对数据进行增加、删除和更新攻击,为客观、公正地评价水印算法的鲁棒性,将本文算法结果与文献[13]算法的结果进行对比(表 2)。从表 2 可知,对于只包含 9 个瓦片的数据 B′,在增加 100%、删除 70%或更新 40%的攻击情形下,水印检测出的 NC 值仍高于 0.8 的阈值,表明水印能够被正确提取,并且在相同攻击程度下,本文算法的鲁棒性均优于对比算法,从而验证了水印算法对增加、删除和更新攻击具有很好的鲁棒性。

表 2　增删更新鲁棒性实验结果
Table 2　Robustness experimental results against the attack modes of addition, deletion and update

攻击方式	攻击强度	本文算法 NC	对比算法 NC
增加	25%	1	1
	50%	0.9650	0.9450
	75%	0.9150	0.8700
	100%	0.8400	0.7400
	150%	0.7750	0.6900
删除	25%	1	0.9800
	50%	0.9900	0.8950
	60%	0.9100	0.8100
	70%	0.8250	0.7450
	80%	0.7450	0.6700
更新	10%	1	1
	20%	1	0.9550
	30%	0.9750	0.8650
	40%	0.8800	0.8200
	50%	0.6350	0.5250

多用户的拼接攻击作为网络环境下矢量瓦片常见的攻击方式,也是考察水印算法是否适用于矢量瓦片的重要方面。对本文的水印算法进行拼接攻击的鲁棒性实验,实验数据 A、数据 B 依照 3×3 规格切出的瓦片编号分别为 A1－A9、B1－B9,由于有 9 个切片,因此最多有 9 个用户参与拼接攻击,用户编号为 U1－U9;对不同的瓦片嵌入相同或不同的用户信息,将含水印的瓦片拼接为 3×3 的数据进行水印检测,数据 A 和数据 B 的拼接攻击实验结果如表 3 和表 4 所示。

表 3　数据 A 拼接攻击鲁棒性实验结果
Table 3　Robustness experimental results for data A against mosaic attack

瓦片编号	A1	A2	A3	A4	A5	A6	A7	A8	A9	检测结果
嵌入的用户信息	U1	U1	U1	U1	U2	U2	U2	U2	U2	U1,U2
	U3	U3	U3	U4	U4	U4	U5	U5	U5	U3－U5
	U6	U6	U7	U7	U8	U8	U9	U9	U9	U6－U9
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U1－U9

表 4　数据 B 拼接攻击鲁棒性实验结果
Table 4　Robustness experimental results for data B against mosaic attack

瓦片编号	B1	B2	B3	B4	B5	B6	B7	B8	B9	检测结果
嵌入的用户信息	U1	U1	U1	U1	U2	U2	U2	U2	U2	U1,U2
	U9	U9	U9	U8	U7	U7	U6	U6	U6	U6－U9
	U1	U3	U1	U5	U2	U4	U2	U3	U5	U1－U5
	U3	U8	U7	U2	U1	U4	U6	U9	U5	U1－U9

由表 3 和表 4 可知,对于实验数据 A 和数据 B,无论在各个瓦片中嵌入怎样的用户信息,都能在水印检测时正确提取出嵌入的所有用户信息,实现了对拼接攻击稳定的鲁棒性。这是由于水印信息生成机制中,足够大的汉明距离能够保证组与组之间的分离,分段编码也保证了用户信息之间的区分;而运用水印容量提高的 QIM 方法,能够在每个瓦片中嵌入足够多的水印信息,实现对单个瓦片的水印嵌入和提取,从而保证在拼接攻击时,水印算法能够成功地检测出每个瓦片包含的用户信息。

综上所述,无论是水印算法的检测结果、不可感知性还是鲁棒性,都能满足矢量瓦片数据版权保护和用户追踪的实用性需求,并且解决了拼接攻击下水印算法的鲁棒性问题。

5　结论

本文面向网络环境下的矢量瓦片版权保护需求,提出了一种抗拼接攻击的矢量瓦片水印算法:分段水印信息生成为用户编码提供了良好的分组和区分度设计;改进的 QIM 水印嵌入和检测方法在保证矢量数据精度的前提下,能够在单个瓦片中嵌入足够多的用户信息,大幅提高了水印容量。实验结果表明,本文的瓦片水印方法能够保证数据的可用性,对增加、删除、更新等攻击具有较好的鲁棒性,特别是在多用户参与的拼接攻击下,能检测出瓦片中包含的全部用户信息,有效地实现了矢量瓦片数据的用户追踪,为矢量瓦片数据的安全保护提供了可行

的解决方案。

参考文献：

[1] 朱长青. 地理数据数字水印和加密控制技术研究进展[J]. 测绘学报，2017(10)：1609－1619.

[2] 朱长青，杨成松，任娜. 论数字水印技术在地理空间数据安全中的应用[J]. 测绘通报，2010(10)：1－3.

[3] PARAH S A，SHEIKH J A，LOAN N A，et al. Robust and blind watermarking technique in DCT domain using inter-block coefficient differencing[J]. Digital Signal Processing，2016，53：11－24.

[4] HAI T，LI C M，ZAINL J M，et al. Robust image watermarking theories and techniques：A review[J]. Journal of Applied Research and Technology，2014，12：122－138.

[5] CHEN S T，HUANG H N. Optimization-based audio watermarking with integrated quantization embedding[J]. Multimedia Tools and Applications，2016，75(8)：4735－4751.

[6] 王潇，任娜，朱长青，等. 基于 QR 码和量化 DCT 的遥感影像数字水印算法[J]. 地理与地理信息科学，2017，33(6)：19－24.

[7] CHEDDAD A，CONDELL J，CURRAN K，et al. Digital image steganography：Survey and analysis of current methods[J]. Signal Processing，2010，90(3)：727－752.

[8] 许惠，任娜，朱长青. 基于防重复嵌入双水印的 DEM 完整性认证算法[J]. 地理与地理信息科学，2016，32(1)：34－38.

[9] 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.

[10] 张黎明，闫浩文，齐建勋，等. 基于归一化的矢量空间数据盲水印算法[J]. 地球信息科学学报，2015，17(7)：816－821.

[11] PENG Z Y，YUE M L，WU X，et al. Blind watermarking scheme for polylines in vector geo-spatial data[J]. Multimedia Tools and Applications，2015，74(24)：11721－11739.

[12] 杨辉，侯翔. 一种抗等角投影变换矢量地图数据水印机制[J]. 测绘科学技术学报，2014，31(5)：524－528.

[13] WANG Y Y，YANG C S. A multiple watermarking algorithm for vector geographic data based on coordinate mapping and domain subdivision[J]. Multimedia Tools and Applications，2017，77(15)：19261－19279.

[14] LEE S H，HUO X J，KWON K R. Vector watermarking method for digital map protection using arc length distribution[J]. IEICE Transactions on Information and Systems，2014，E97. D(1)：34－42.

[15] 佟德宇，朱长青，任娜. 一种不依赖主键的地理数据库水印算法[J]. 地理与地理信息科学，2015，31(5)：86－89.

[16] 林威，翟信德，朱长青，等. 基于 QR 码的遥感影像数字水印算法[J]. 北京邮电大学学报，2015，38(1)：26－30.

[17] WANG B，REN N，ZHU C Q. Watermarking algorithm based on data feature for tile map[A]. International Conference on Geo-Informatics in Resource Management and Sustainable Ecosystem[C]. Berlin：Heidelberg，2014. 186－193.

[18] REN N，ZHU C Q. A digital watermark algorithm for tile map stored by indexing mechanism[A]. Cartography from Pole to Pole[C]. Berlin：Heidelberg，2014. 79－86.

[19] LI L，HU W，ZHU H H，et al. Tiled vector data model for the geographical features of symbolized maps[J]. Plos One，2017，12(5)：e0176387.

[20] LI H T，PENG Q S，LI Y H. Data security analysis of WebGIS based on tile-map technique[A]. Proceedings of the 2009 International Symposium on Web Information Systems and Applications[C]. 2009. 190－193.

Watermarking Algorithm for Vector Tiled Data against Mosaic Attack

TANG Wei

(*Archives of Surveying and Mapping Information of Zhejiang Province，Hangzhou 311100，China*)

Abstract：While vector tiled data have been widely used in the Internet and Web，the security issues such as illegal crawling or copy need to be solved urgently. Aiming at the copyright protection and user tracking for vector tiled data in network environment，a robust watermarking algorithm designed to resist mosaic attack is proposed in this paper. In the proposed algorithm，watermark is generated by two segments，and the watermark generation algorithm can provide as many codes as possible while ensuring the discriminations between groups and users. Moreover，in the watermark embedding and extraction stage，the classical quantization index modulation method is improved by dividing a certain number of intervals according to the data precision. Thereby the watermark capacity is increased largely. The experiments have been conducted to verify the proposed algorithm，and the results show that this algorithm ensures the availability and the precision of the data. Besides，its superior robustness against data addition，deletion，update and multi-user mosaic attack has also been proved by the experiments，meaning the proposed method is applicable for copyright protection and user tracking for vector tiled data.

Key words：digital watermark；vector tiled data；robustness；mosaic attack