

矢量地图水印的数据质量评价研究

赵 晶 林鸿飞 王 凡

(大连理工大学计算机科学与工程系, 大连 116023)

摘 要 含水印数据的质量评价是衡量水印嵌入隐蔽性和数据可用性的重要指标。峰值信噪比(PSNR)等基于能量的度量指标在应用于矢量地图水印系统时具有一定的局限性。从形状的角度考虑了矢量地图水印的数据质量评价问题,借鉴时间序列聚类 and 形状相似性匹配的思想,提出了基于距离度量的水印地图数据评价指标。算法从2维矢量地图中提取1维特征函数,通过度量水印嵌入前后特征函数的形状差异来评价含水印地图的数据质量。实验结果证明,本文提出的度量方法更符合矢量数据的特点,能够得到比现有方法更准确的度量结果。

关键词 矢量地图水印 数据质量评价 距离度量 形状差异

中图法分类号: TP309 文献标志码: A 文章编号: 1006-8961(2010)07-1121-05

Study on Data Quality Evaluation of Vector Map Watermarking

ZHAO Jing, LIN Hongfei, WANG Fan

(Department of Computer Science and Technology, Dalian University of Technology, Dalian 116023)

Abstract Quality of the watermarked data is an important metric for measuring the invisibility of the watermark and the validity of the marked data. Some energy based metrics, e. g. peak signal to noise ratio(PSNR), have some limitations while applied to vector map watermarking systems. This paper considered this issue in the aspect of map shape. The idea of time series clustering and shape similarity matching techniques has been used and the watermarked-map quality measurement based on distance metric is presented. The proposed algorithm extracts 1D feature functions from 2D vector maps and evaluates the quality of the watermarked data by measuring the shape difference between two feature functions before and after watermark embedding. Experiment results indicate that the proposed metric fits the characteristic of vector map data better and, furthermore, it can achieve more reasonable measuring results than previous methods.

Keywords vector map watermarking, data quality evaluation, distance metric, shape difference

0 引 言

在数字水印技术中,对含水印数据的质量评价是衡量水印算法优劣的一项重要指标。通过度量水印嵌入前后载体数据的变化,可以衡量水印的隐蔽性和含水印数据的可用性。栅格图像水印常用的度量指标包括均方差(MSE)、 L_p 距离、峰值信噪比(PSNR)^[1]等。上述指标从本质上都取决于水印嵌入的能量,本文将其归类为基于能量的度量指标。

在矢量地图水印中,地图数据的可用性是由水印嵌入算法来保证的,但即使在含水印地图精度均可用的前提下,不同的嵌入算法对地图的扰动情况也不尽相同,这将直接影响到含水印地图的视觉质量和水印算法的隐蔽性。根据目前的研究进展,对含水印地图的质量评价问题关注较少,只有 Kang 等人^[2]根据栅格图像水印的相关方法定义了一种用于地图坐标的峰值信噪比,这个指标本质上也是基于能量的。实际上,依靠误差能量来评价矢量地图数据质量具有很大的局限性,容易产生不恰当的评

基金项目:国家自然科学基金项目(60673039);国家高技术研究发展计划(863)项目(2006AA01Z151)

收稿日期:2009--; **改回日期:**2009--

第一作者简介:赵 晶(1961—),女,讲师。1998 年于大连理工大学机械系获机械制造专业硕士学位。主要研究领域为模式识别与图像处理、信息检索和信息过滤。E-mail: rfrm@dlut.edu.cn

估结果,这是由矢量地图水印的两个特点决定的:

1) 矢量地图元素通常由 2 维平面上的有序顶点连接而成,水印嵌入会使顶点产生位置偏移。然而,对任一顶点,即使施加同样的位置偏移量,如果偏移的方向不同,对地图元素形状产生的影响也不同。

2) 矢量地图支持无损缩放,当含水印地图被充分放大时,水印的隐蔽性主要取决于水印嵌入对地图细节形状的扰动程度,而不是地图顶点的偏移量。上述两方面因素决定了仅采用基于能量的度量指标来衡量矢量地图的数据质量是不恰当的。

本文借鉴时间序列聚类 and 形状相似性匹配的相关思想,提出了两种基于地图形状的度量指标,通过比较水印嵌入前后地图元素细节形状的差异来衡量含水印地图的质量。实验结果证明,基于形状度量能够有效克服基于能量度量的缺陷,因而更加适用于矢量地图数据。

1 距离度量

将含水印地图与原始地图之间的形状差异定义为一种距离。假设需要度量的两幅地图 U 和 V 所在的空间为 S , $d(U, V): S \times S \rightarrow \mathbf{R}$ 为定义在该空间上的一个映射, $d(U, V)$ 称为距离函数或者距离度量。根据实际应用环境,当用 $d(U, V)$ 表示地图形状差异程度时,希望它满足以下性质:

1) 非负性 $d(U, V) \geq 0$, 即两幅地图的形状差异是一个非负实数。

2) 一致性 如果 $U = V$, 则 $d(U, V) = 0$, 即两幅完全相同的地图间的距离为 0。

3) 对称性 $d(U, V) = d(V, U)$, 即与人的感官特征相吻合, U 与 V 的形状差异应该等于 V 与 U 的形状差异。

2 基于形状的度量指标

提出的度量指标通过两个步骤得到。首先,从两幅待比较的 2 维地图中分别提取一个 1 维特征函数,该函数反映了地图的细节形状;然后,针对两个特征函数定义一个距离度量,该度量反映两幅地图的形状差异。

2.1 2 维形状的 1 维表示

以地图中单个多边形曲线元素为例说明特征函

数的提取过程(多个元素的情况与之相似)。假设 A 和 B 分别是两幅待比较地图中的某个对应元素。 A 和 B 都由 N 个顶点组成,分别记为 $\{v_1, \dots, v_i, \dots, v_N\}$ 和 $\{u_1, \dots, u_i, \dots, u_N\}$ 。首先采用特征点提取算法提取 A 的特征点集(本文采用 Douglas-Peucker 法^[3]),得到 A 的简化曲线 A_s, A_s 由 A 的特征点连线构成。然后计算 A 中每个顶点 v_i 到 A_s 中对应简化线段的距离,得到长度为 N 的距离序列 $L_A = \{l_{A_1}, \dots, l_{A_i}, \dots, l_{A_N}\}$ 。同时,令 Δs_{A_i} 为 A 中两个相邻顶点的距离,即

$$\Delta s_{A_i} = \begin{cases} 0 & i = 1 \\ d_E(v_{i-1}, v_i) & i \in \{2, \dots, N\} \end{cases} \quad (1)$$

式中, d_E 为两个顶点的欧氏距离。由 Δs_{A_i} 可以进一步得到曲线 A 的弧长序列 $S_A = \{s_{A_i} \mid i = 1, 2, \dots, N\}$, 其中:

$$s_{A_i} = \sum_{k=1}^i \Delta s_{A_k} \quad i \in \{1, \dots, N\} \quad (2)$$

S_A 为单调递增序列。对于 $i = 1, 2, \dots, N$, 距离序列 L_A 与弧长序列 S_A 的元素一一对应。于是可以将距离 L_A 为纵坐标, 弧长 S_A 为横坐标构造一个分段线性函数 $L_A(s)$, 其中 s 为弧长, 其取值范围为 $s \in [0, s_{A_N}]$ 。

$$L_A(s) = \begin{cases} l_{A_i} & s = s_{A_i} \\ l_{A_{(i-1)}} + (s - s_{A_{(i-1)}})k & s \in (s_{A_{(i-1)}}, s_{A_i}) \end{cases} \quad (3)$$

式中, $k = (l_{A_i} - l_{A_{(i-1)}}) / \Delta s_{A_i}$ 为各段的斜率。将函数 $L_A(s)$ 作为曲线 A 的特征函数, 它充分体现了 A 的形状细节。

对于曲线 B , 仍以简化对象 A_s 为参照, 采用同样的方法构造关于 B 的特征函数 $L_B(s)$, s 的取值范围为 $s \in [0, s_{B_N}]$ 。注意: 距离序列 L_B 的求取仍然以 A 的简化曲线 A_s 为参照, 而不是以 B 的简化曲线为参照。这是为了避免由于 A 和 B 的特征点检测偏差造成两个线性函数 $L_A(s)$ 和 $L_B(s)$ 之间差异的增大, 从而确保形状评价的准确性。此时称曲线 A 为参照曲线。图 1 用包含 7 个顶点的简单曲线说明了以 A 为参照曲线时特征函数的提取过程。

2.2 方向距离度量及其对称化

特征函数 $L_A(s)$ 和 $L_B(s)$ 是地图元素 2 维形状的 1 维表示形式, 通过度量 $L_A(s)$ 和 $L_B(s)$ 的距离来衡量曲线 A 和 B 的形状差异。首先定义方向距离度量: 曲线 A 和 B 的方向距离度量记为 $d(A, B)$, 该

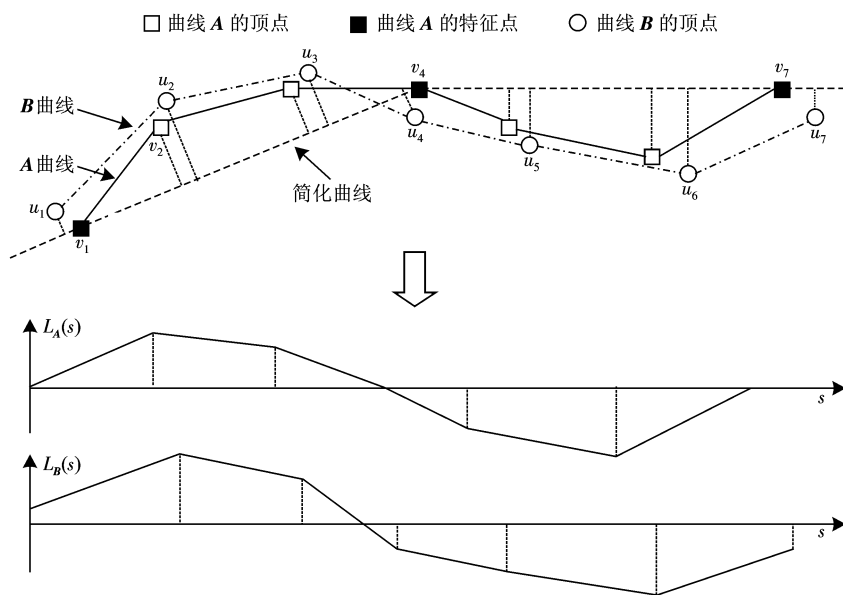


图 1 特征函数的提取过程

Fig. 1 Procedure of feature function extraction

距离表示为 $L_A(s)$ 和 $L_B(s)$ 的函数:

$$d(A, B) = f(L_A(s), L_B(s)) \quad (4)$$

在提取 $L_A(s)$ 和 $L_B(s)$ 时,默认以 $d(\cdot)$ 中第 1 个参数代表的曲线为参照曲线,因此上述定义具有方向性,即一般来说, $d(A, B) \neq d(B, A)$ 。 $f(\cdot)$ 是分段线性函数的某种距离定义,它应该对函数 $L_A(s)$ 和 $L_B(s)$ 的形状差异敏感。下面基于不同的 $f(\cdot)$ 给出两种 $d(A, B)$ 定义:

1) 基于短序列距离(记为 d_{STS})的度量

借鉴时间序列聚类的相关技术^[4],通过度量 $L_A(s)$ 和 $L_B(s)$ 的平均短序列距离来衡量 A 和 B 的形状差异,即

$$d(A, B) = d_{\text{STS}}(L_A(s), L_B(s)) =$$

$$\sqrt{\frac{\sum_{j=2}^N \left(\frac{l_{A_j} - l_{A_{(j-1)}}}{s_{A_j} - s_{A_{(j-1)}}} - \frac{l_{B_j} - l_{B_{(j-1)}}}{s_{B_j} - s_{B_{(j-1)}}} \right)^2}{N-1}} \quad (5)$$

由式(5)可见,分段线性函数 $L_A(s)$ 和 $L_B(s)$ 的短序列距离实质上是二者斜率序列的平均欧氏距离。

2) 基于转动函数面积差距离(记为 d_{TFA})的度量

除短序列距离以外,还可以借鉴形状匹配技术中基于转动函数^[5]的度量方法。首先求取 $L_A(s)$ 和 $L_B(s)$ 的斜率函数(可以看做是转动函数的一种形式),记为 $\Theta_A(s)$ 和 $\Theta_B(s)$ 。由于 $L_A(s)$ 和 $L_B(s)$

是分段线性的, $\Theta_A(s)$ 和 $\Theta_B(s)$ 必然是分段恒定的,它们可以表示成两个 2 维序列:

$$\begin{cases} \Theta_A(s) = \{ (k_A(i), \Delta s_A(i)) \mid i = 1, \dots, N-1 \} \\ \Theta_B(s) = \{ (k_B(i), \Delta s_B(i)) \mid i = 1, \dots, N-1 \} \end{cases} \quad (6)$$

式中, k 和 Δs 分别是每个分段的斜率值和段长度,于是可以定义基于 $\Theta_A(s)$ 和 $\Theta_B(s)$ 分段面积差的距离度量指标:

$$d(A, B) = d_{\text{TFA}}(L_A(s), L_B(s)) =$$

$$\sqrt{\frac{\sum_{i=2}^N |k_A(i) \cdot \Delta s_A(i) - k_B(i) \cdot \Delta s_B(i)|}{N-1}} \quad (7)$$

上述两种距离定义均具有方向性。为满足距离度量的对称性要求,采用下式对上述度量进行对称化,得到最终的距离度量 $d(A, B)$:

$$d(A, B) = \frac{d(A, B) + d(B, A)}{2} \quad (8)$$

可以证明 $d(A, B)$ 满足本文第 1 节中对距离度量要求的 3 个性质,可以用于度量两幅地图的形状差异。

3 实验结果

采用如下方法检验上述两种度量指标的有效性。首先,选取一幅大比例的等高线图作为原始地图,图 2 为该图一块截取放大部分。

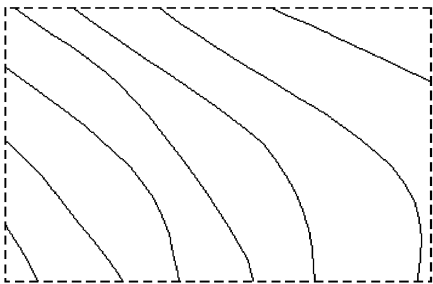


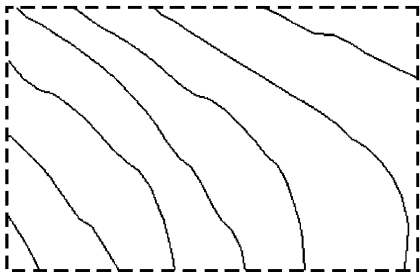
图 2 原始地图数据(部分)
Fig.2 Part of original map

接着,分别采用 4 种不同的算法向地图中嵌入水印信息,这些算法包括:

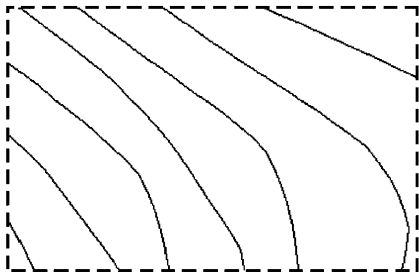
算法 1 统计特征水印算法^[6];

- 算法 2 基于特征点检测的水印算法^[7];
算法 3 基于坐标差值扩大的水印算法^[8];
算法 4 基于线段长度差值扩大的水印算法^[9]。

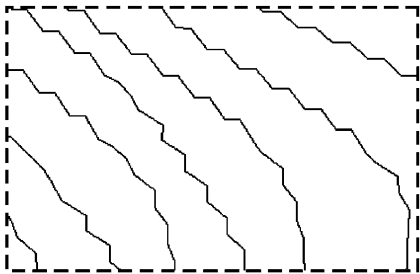
图 3 为各算法得到的含水印地图的放大部分。对于算法 1 和算法 2,通过控制水印嵌入的幅度因子,使二者具有相近的水印嵌入能量,但由图 3 可见,二者对地图形状的扰动程度明显不同;算法 3 和算法 4 则分别具有最大和最小的水印嵌入能量。通过视觉判断可以得出结论,图 3 中,4 幅子图的数据质量由好到差依次为(d) > (b) > (a) > (c), (b)和 (d)具有平滑的外形特征,其视觉质量明显优于(a)和(c)。



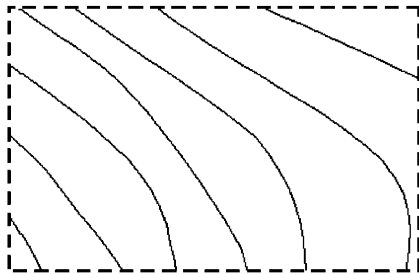
(a) 算法 1



(b) 算法 2



(c) 算法 3



(d) 算法 4

图 3 含水印地图
Fig.3 Watermarked map

一个合理的度量指标,其度量结果应与人的视觉观察结果基本吻合。以原始地图为基准,计算每幅含水印地图的质量评价指标,这些指标包括噪声能量(记为 E)、峰值信噪比、短时序列距离(基于 d_{STS})、以及转动函数面积差距离(基于 d_{TFA})。表 1 为各种指标的度量结果。

注意,上述 4 种度量指标中,PSNR 的数值越大,表示图形的质量越好,而其余 3 种指标则正好相反。由表 1 的数据可见,4 种度量指标都认为图 3(c)的质量最差,而图 3(d)的质量最好。因此,

表 1 4 种算法嵌入水印后地图质量度量结果
Tab.1 The measuring results of the map watermarked by four different algorithms

	图 3			
	(a)	(b)	(c)	(d)
E	0.047 293	0.047 251	0.140 33	0.000 138 2
PSNR	74.566 8	57.423 3	51.644 6	82.159 7
d_{STS}	0.089 571	0.031 642	0.326 75	0.003 1
d_{TFA}	0.349 47	0.223 73	0.936 62	0.046 541

对于水印嵌入能量差别较大的情况,4种度量指标都能够合理区分。但是,当水印嵌入能量十分相近时(图3(a)和图3(b)),基于能量的度量指标性能明显下降。其中, E 的度量结果极为接近,很难通过该指标区分两图的质量高低;PSNR则认为图3(a)的质量优于图3(b),这与人的视觉观察结论相违背。而本文提出的两种度量指标都得出了图3(b)优于图3(a)的度量结果,这与人的视觉观察特征相吻合。上述实验结果表明,基于能量的度量指标在应用于矢量地图时具有一定的局限性,有可能造成不恰当的评估结果,而基于形状的度量指标则更符合矢量地图的数据特征。

4 结 论

由于矢量地图的特殊数据结构和无损缩放的特性,地图的形状扰动成为衡量含水印地图数据质量需要考虑的重要因素。PSNR等基于能量的度量指标本质上是坐标误差能量的反映,对地图形状的变化不敏感,因此用于衡量矢量图形水印性能时具有一定的局限性。本文直接从形状的角度考虑地图数据质量度量问题,通过度量地图特征函数的形状距离来衡量地图的视觉质量差异,避免了PSNR可能出现的不恰当结论,因而更加适合于矢量地图水印的数据质量评价。由于矢量数据的特殊性,地图的质量往往是由很多因素共同决定的。除地图元素的形状以外,重要的因素还包括地图数据的最大误差值、地图对象间的拓扑关系、误差的总能量等。因此,仅采用单一的度量方式很难全面地反映上述特征,地图数据质量的度量需要从多方面来考虑,采用多种度量方式进行。

参考文献 (References)

- [1] Kutter M, Petitcolas F A P. A Fair benchmark for image watermarking systems[C]//Proceedings of SPIE: Security and Watermarking of Multimedia Content. Bellingham WA: ETATS-UNIS SPIE, 1999, 3657: 219-239.
- [2] Kang H. A vector watermarking using the generalized square mask [C]// Proceedings of International Symposium on Information Technology: Coding and Computing. Washington, DC, USA: IEEE Computer Society, 2001: 234-236.
- [3] Douglas D H, Peucker T K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature [J]. Cartographica: the International Journal for

Geographic Information and Geovisualization, 1973, 10: 112-122.

- [4] Möller-Levet C S, Klawonn F, Cho K H, et al. Fuzzy clustering of short time series and unevenly distributed sampling points [C]//Proceedings of the 5th International Symposium on Intelligent Data Analysis. Berlin: Springer, 2003: 28-30.
- [5] Arkin E, Chew P, Huttenlocher D, et al. Anefficiently computable metric for comparing polygonal shapes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(3): 209-215.
- [6] Voigt M, Busch C. Feature-based watermarking of 2d-vector data [J]. Proceedings of SPIE, 2003, 5020: 359-366.
- [7] Shao Chengyong, Wang Xiaotong, Xu Xiaogang. A robust watermarking algorithm for 2D vector maps based on feature point detection [C]//Proceedings of the 12th Conference on Image and Graphics. Beijing: Tsinghua University Press, 2005: 145-148. [邵承永, 王孝通, 徐晓刚. 基于特征点提取的矢量地图鲁棒水印算法[C]//第十二届全国图像图形学术会议论文集. 北京: 清华大学出版社, 2005: 145-148.]
- [8] Shao Chengyong, Wang Xiaotong, Xu Xiaogang. Studyon lossless data hiding algrithm for digital vector maps[J]. Journal of Image and Graphics, 2007, 12(2): 206-211. [邵承永, 王孝通, 徐晓刚. 矢量地图的无损数据隐藏算法研究[J]. 中国图象图形学报, 2007, 12(2): 206-211.]
- [9] Wang Xiaotong, Shao Chengyong, Xu Xiaogang, et al. Reversible data hiding scheme for 2D vector maps based on difference expansion [J]. IEEE Transactions on Information Forensics and Security, 2007, 2(3): 311-320.