

平均误差控制下的频率域矢量数据压缩方法

黄伟明, 杨建宇*, 岳彦利, 杜 萌, 张 超, 朱德海

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 国土资源部农用地质量与监控重点实验室, 北京 100035)

摘要 矢量数据压缩对于受网络带宽限制的 WebGIS 有着重要意义, 其可减少数据存储空间, 提高网络传输与处理效率。传统的矢量数据压缩方法主要从空间关系的角度出发, 根据原始矢量数据点之间的距离、角度等指标, 判断如何对矢量线要素或面要素进行化简, 略去冗余的端点。而本文则利用已在图像压缩领域被广泛应用的变换编码的频率域压缩技术, 利用能实现能量保持的离散余弦变换和特殊的量化方法, 以及无损熵编码, 对矢量数据实现了能控制压缩后平均误差的有损压缩。该方法无需事先根据不同的误差限值设计量化表, 且能处理指定过大平均误差限值时所出现的问题, 有很强的适应能力。最后, 使用 C# 实现了该方法, 并验证了方法的可靠性, 完成了方法的性能测试。实验结果表明, 本文提出的矢量数据压缩方法能获得较大的压缩比, 且能较好地保持原始矢量数据所具有的地理形态结构特征。

关键词 矢量地图数据; WebGIS; 离散余弦变换; 平均误差; 空间数据压缩

DOI:10.3724/SP.J.1047.2015.00883

1 引言

网络地理信息系统 (WebGIS)^[1] 是传输、处理地理数据的重要载体, 但随着地理数据获取能力的不断增强, 其所需要处理和传输的数据量也不断增大, 导致系统整体效率一定程度的降低。因此, 有必要对地理数据根据具体的应用需求进行压缩。

矢量数据是 GIS 中一种重要的数据结构, 对其进行合理地压缩, 能大大提高 WebGIS 的传输和处理效率。矢量数据压缩按是否能将压缩后的数据完全重建为原数据可分为有损压缩和无损压缩^[2]; 同时, 按照矢量数据压缩方法所采用的原理可将其分为空间域方法和频率域方法^[3]。

空间域的矢量数据压缩方法, 主要是从组成矢量数据的端点之间的空间关系 (如距离、角度等) 方面, 对原始矢量数据点集进行抽稀和化简, 其实现的主要过程为设置具有某种特定含义的阈值, 进而去除在某些特定应用或尺度下冗余的端点。常见的空间域矢量数据压缩方法有 Douglas-Peucker 算法^[4] 及其相对应的一系列改进算法^[5-8], Visvalingam-Whyatt 算法^[9], 黄培之提出的具有预测功能的曲线

矢量数据压缩方法^[10] 等, 这些方法经常被应用于矢量数据渐进传输中^[11-12]。频率域的矢量数据压缩方法出现时间相对较晚, 其原理和方法主要来源于数字信号处理领域部分较为成熟的理论和方法, 其中包括图像压缩中常用的离散余弦变换 (DCT)、小波变换等。

余先川等^[13] 先将矢量数据的点坐标值的差值转换为整形的偏移量, 再对偏移量进行整形小波变换, 最后对变换后的系数进行无损熵编码来实现矢量数据的无损压缩。王玉海等^[14-17] 探索了多种小波基函数的矢量数据压缩模型和算法, 并对它们进行了比较。马伯宁等^[18] 提出了一种具有误差修正的线矢量数据小波变换, 修正了用小波变换进行矢量数据压缩时出现的过大误差。朱海军、吴华意提出利用 DCT 和无损熵编码, 实现矢量数据有损压缩的流程和方法^[3,19-20], 可控制重建后的矢量数据在 x 和 y 方向上的最大位移, 且能实现较大的压缩比。

本文提出一种不依赖于量化表的基于 DCT 的频率域矢量数据压缩方法, 能灵活地控制重建后矢量数据点坐标的平均误差。通过理论与实验分析, 初步验证了该方法的可行性。

收稿日期 2015-01-14; 修回日期: 2015-02-19.

基金项目 国家自然科学基金项目 (41171309)。

作者简介 黄伟明 (1990-), 男, 山东济南人, 硕士生, 研究方向为地理信息系统。E-mail: huangweim@hotmail.com

*通讯作者: 杨建宇 (1974-), 男, 博士, 教授, 研究方向为 3S 技术及其土地资源方面的应用。E-mail: ycjyyang@cau.edu.cn

2 平均误差控制下的频率域矢量数据压缩方法

朱海军等在文献[3]中提出了基于DCT变换的矢量数据压缩流程,但其在对DCT系数进行量化之前,需根据不同的误差限值计算得到特定的量化表,然而对于取值范围非常大的矢量数据来说,很难找到像图像压缩中所采用的效果较好的推荐量化表,因此量化表的可用性不佳。另外,其所控制的误差是 x 和 y 2个方向位移的最大值而不是压缩后点在二维空间内的实际位移,且没有探讨当出现很大的误差限值时所需进行的处理。本文针对上述问题进行探讨,在继续沿用文献[3]中压缩流程的同时,提出了新的量化方法与出现过误差限值时的处理方法。

2.1 矢量数据分块

虽然无需使用量化表对DCT系数进行不同步长的均匀量化,但数据分块对提高算法的时间效率与压缩后文件格式的设计仍有重要意义。

压缩比与分块大小 N 直接相关。一般来说, N 的增大可使得原始数据的能量集中于更少的系数,从而实现更大的压缩比,但同时也会使得计算量增大,算法时间效率下降^[21]。因此,如何权衡压缩比与算法效率是使用频率域压缩方法的一个重要问题。对于矢量数据压缩,不仅要注意上述的权衡,同时更要注意矢量数据本身的地理特征,即在分块处理时,同一分块内不能包含不同要素的坐标数据。另外,由于本方法无需使用量化表,因此,无需在每个要素的最后一个分块区域中补充数据,使该分块内数据个数补足为分块大小 N 。

2.2 DCT变换

作为一种常用的正交变换,DCT具有可实现能量保持、能量重新分配与集中的特性,且算法快速,一直被广泛应用于图像压缩领域。

本文将矢量数据分块后,对每个分块中的 x 和 y 坐标分别进行DCT变换,其表达式如式(1)所示。

$$X(k) = c_k \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k\pi}{2N} \quad (1)$$

式中, N 代表该分块内 x 或 y 坐标个数。当 $k=0$ 时, $c_k = \sqrt{1/N}$;当 $1 \leq k \leq N-1$ 时, $c_k = \sqrt{2/N}$ 。

需要说明的是,因为矢量数据的坐标值取值范围很大(约为 2^{18}),且不同的数据集坐标的取值范围

差别很大,因此,本文未采用JPEG标准的图像压缩流程^[22]中消除电平偏移的方法。

2.3 DCT系数的量化

经DCT变换后,对于 x 和 y 坐标每一个分块的DCT变换系数来说,都有1个DC系数,其余的则是AC系数。因DC系数中含有数据分块内的均值信息,相关性较强,所以对 x 和 y 坐标各自的第1个DC系数取整后单独编码,其他的DC系数则在取整后编码为与前一个已取整量化的DC系数的差值。对于AC系数,则使用不同的特殊量化方式。

本文中矢量数据压缩的失真准则选取表征平均误差的均方根误差(RMSE),与误差的算术平均值相比,RMSE可以更好地表征误差的离散程度,抑制过大误差的出现。又因为DCT变换为正交变换,符合帕塞瓦尔定理,因此可以推导得出:

$$\begin{aligned} \varepsilon &= \sqrt{\frac{\sum_{i=0}^{n-1} d_i^2}{n}} = \sqrt{\frac{\sum_{i=0}^{n-1} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]}{n}} \\ &= \sqrt{\frac{\sum_{i=0}^{n-1} [(X_i - \hat{X}_i)^2 + (Y_i - \hat{Y}_i)^2]}{n}} \end{aligned} \quad (2)$$

式中, ε 代表压缩后的均方根误差; d_i 代表第 i 个点与压缩前对应点的距离; n 代表某一要素中点的个数; x_i, y_i 表示压缩前点的横纵坐标; \hat{x}_i, \hat{y}_i 代表压缩后点的横纵坐标; X_i, Y_i 代表量化前的DCT系数值; \hat{X}_i, \hat{Y}_i 代表量化后的DCT系数值。

由式(2)可得出,频率域系数改变所产生的平均误差,直接对应于压缩后点的平均位移,因此,在一定的平均误差限值 h 下,可只对一些较大的DCT系数进行采样并取整量化,对于较小的DCT系数则直接置为“0”。其实现的具体步骤如下:

(1)对于要素类中的某一个要素,将其所对应的所有分块中 x 和 y 坐标的所有AC系数按照由小到大进行排序得到系数序列 $C\{C_0, C_1, \dots, C_{2n-m-1}\}$ (n 代表该要素中点的个数; m 代表该要素坐标变换得到DC系数的个数),同时存储各个系数对应于排序前的分块和顺序信息。另外,也需计算取整所有DC系数所产生的误差:对于这一要素的 x 和 y 坐标所有DC系数 $\{D_0, D_1, \dots, D_{m-1}\}$ 来说,取整量化所产生的整体误差 ε_{DC} 计算公式如式(3)所示:

$$\varepsilon_{DC} = \sqrt{\sum_{k=0}^{m-1} (|D_k| - \lfloor |D_k| \rfloor)^2} \quad (3)$$

式中, $\lfloor \cdot \rfloor$ 代表向下取整运算。

(2)从头开始遍历第一步所得到系数序列 C ,直到找到一个 C_i 满足下列2个条件(式(4)、(5))。

$$\sum_{k=0}^{i-1} C_k^2 + \sum_{k=i}^{n-1} (|C_k| - \lfloor |C_k| \rfloor)^2 + \varepsilon_{DC}^2 \leq nh^2 \quad (4)$$

$$\sum_{k=0}^i C_k^2 + \sum_{k=i+1}^{n-1} (|C_k| - \lfloor |C_k| \rfloor)^2 + \varepsilon_{DC}^2 > nh^2 \quad (5)$$

式中, n 代表该要素内点的个数; h 代表平均误差(均方根误差)限值。

(3)找到 C_i 后,将 C 的子序列 $\{C_0, C_1, \dots, C_{i-1}\}$ 所有的值直接量化为“0”,对另一个子序列 $\{C_i, C_{i+1}, \dots, C_{2n-m-1}\}$ 中的每一个数则简单取整。

(4)将上述步骤处理后的序列 C ,参照存储的各个系数对应于排序前的分块和顺序信息按顺序排列,以便于压缩后输出文件结构的设计。

最后,遍历要素集中的其他要素,按照同样方法处理直至所有要素的DCT变换系数量化结束。

2.4 过大误差限值的处理

按照上述方法,可灵活地根据平均误差限值,决定使哪些系数直接量化为“0”,其与无损熵编码结合可大幅度降低矢量数据所需的存储空间。

但依上述步骤进行处理的过程中,若出现过大的平均误差限值,则可能会使得某一分块或多个分块中的所有AC系数被置为“0”,只剩下DC系数。该情况下,会出现重建的矢量数据中,在这一数据分块内所有的 x 或 y 坐标为同一个值(压缩后点数不发生变化,但是点的坐标一致),甚至极端情况下会出现一个线要素或面要素被压缩成一个点,或一个面要素被压缩为一条线的情况。因此,有必要对过大误差限值的出现进行判断和处理。

(1)若被压缩的矢量数据为线要素类,则对每一分块区域的 x 和 y 坐标的所有DCT变换系数中最大的AC系数只进行简单的“取整处理”,不能将其置为“0”;

(2)若被压缩的矢量数据为面要素类,则对每一分块区域的 x 和 y 坐标分别的DCT变换系数中最大的AC系数只进行简单的“取整处理”,不能将其置为“0”。

若上述只进行“简单取整”处理的DCT系数的绝对值在0-1之间,则采用双精度浮点数将其记录在压缩后的输入文件的特殊位置,同时记录其相应的位置信息。

2.5 无损熵编码

在对矢量数据的DCT变换系数进行前述的量

化处理之后,本文选用无损熵编码方法继续对量化后的变换系数进行处理,减小压缩后矢量数据所需的存储空间。

本文采用与JPEG图像压缩标准^[22]相似的Huffman编码方法,并根据量化后的 x 和 y 坐标的DCT变换系数设计适合于DC系数差值和AC系数的编码表。

3 实验与分析

为验证上述方法的效果和可靠性,作者在Windows 8平台上,使用C#实现了本文所述方法。使用了2个地区的实验数据:(1)北京市通州区矢量等高线图的一部分,要素类中共有261 114个端点,占用4333 KB的存储空间;(2)美国爱达荷州的矢量边界图,共有7127个端点,占用112 KB的存储空间。对上述2个地区的实验数据,以多个不同的平均误差限值,对2个不同的原矢量数据进行压缩。为了兼顾压缩的时间效率和压缩比,本文对矢量数据的分块大小 N 的选取做了量化比较,如表1所示。从实验结果可得出,当分块大小 N 选择为64时,算法效率和压缩比得到了很好的权衡,因此,在本文后续实验中矢量数据的分块大小选用64。以不同平均误差限值压缩后得到的矢量数据的平均误差(RMSE)、压缩后点的最大位移、所占存储空间及压缩比,如表2、3所示。

表1 使用不同的分块大小进行算法的性能测试
(包含算法运行时间和压缩比)

Tab. 1 Compression performance (compressed storage space and time efficiency) in respect to the size of block

| 性能参数 | 分块大小 N | | | | |
|------------|----------|-----|------|------|------|
| | 16 | 32 | 64 | 128 | 256 |
| 压缩所用时间(ms) | 682 | 947 | 1417 | 2089 | 3020 |
| 压缩后大小(KB) | 382 | 378 | 372 | 370 | 369 |

表2 本文方法在不同平均误差限值下的性能参数
(实验数据为部分北京市通州区矢量等高线图)

Tab. 2 Compression performance (compression ratio and compression bias) of our method (from the first test featureclass)

| 性能参数 | 平均误差限值(m) | | | | |
|-----------|-----------|-------|-------|-------|-------|
| | 1 | 2 | 5 | 8 | 10 |
| RMSE(m) | 0.96 | 1.94 | 4.83 | 7.64 | 9.47 |
| 最大误差(m) | 4.85 | 10.93 | 28.28 | 58.49 | 80.50 |
| 压缩后大小(KB) | 594 | 372 | 269 | 212 | 188 |
| 压缩比(%) | 86.3 | 91.4 | 93.8 | 95.1 | 95.7 |

表3 本文方法在不同平均误差限值下的性能参数
(实验数据为美国爱达荷州的矢量边界图)

Tab. 3 Compression performance (compression ratio and compression bias) of our method (from the second test featureclass)

| 性能参数 | 平均误差限值(m) | | | | |
|-----------|-----------|-------|--------|--------|--------|
| | 10 | 20 | 30 | 50 | 80 |
| RMSE(m) | 9.99 | 20.00 | 29.99 | 49.99 | 79.97 |
| 最大误差(m) | 35.73 | 64.50 | 103.36 | 190.22 | 442.31 |
| 压缩后大小(KB) | 19 | 16 | 14 | 11 | 9 |
| 压缩比(%) | 83.0 | 85.7 | 87.5 | 90.2 | 92.0 |

由表2、3可知,在对第1个要素类进行实验的过程中,当平均误差限值设定为2 m时,压缩比达到了91.4%;在对第2个要素类进行实验的过程中,当平均误差限值设定为50 m时,压缩后结果的压缩比达到了90.2%。同时,以上2个矢量数据的形态结构特征,均得到了较好的保持(如图1、2所示,其中,局部对比图中实线表示压缩前的数据,虚线表示压缩后的数据)。由此可见,本文所提出的矢量数据

压缩方法可灵活地控制压缩后的平均误差,且能使得矢量数据所需的存储空间大大减少。

同时,本文方法的运算量主要集中在DCT变换、反变换及上述的量化过程,对于DCT的变换和反变换来说,其快速算法可直接应用于本文方法中。对于量化过程来说,在判断某一点是否满足上述量化方法中所提到的2个条件时,某一端点对应的判断过程可利用其之前端点判断时的计算结果,并在其基础上进行运算量很小的计算,这样可大大加快运算效率。因此,在实际应用中,本文方法获得较好的时间效率。

本文方法与传统空间域的矢量数据压缩方法有较大不同,重建后的矢量数据与原始数据相比,点的个数不会发生改变,但也几乎不会保留原始矢量数据中的点。在某些情况下,其可获得比传统方法更大的压缩比,例如,对于Douglas-Peucker算法压缩后的矢量数据来说,表示一条线要素最少需要4个双精度的浮点数,而本文方法最少只需要3个

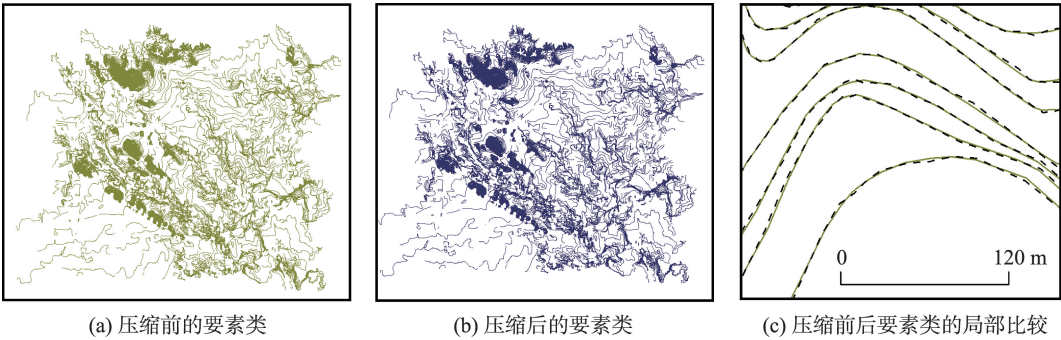


图1 部分北京市通州区矢量等高线图压缩前后全局和局部比较

Fig. 1 The comparison of the original map and compressed map at global and local scales (for the first test featureclass)

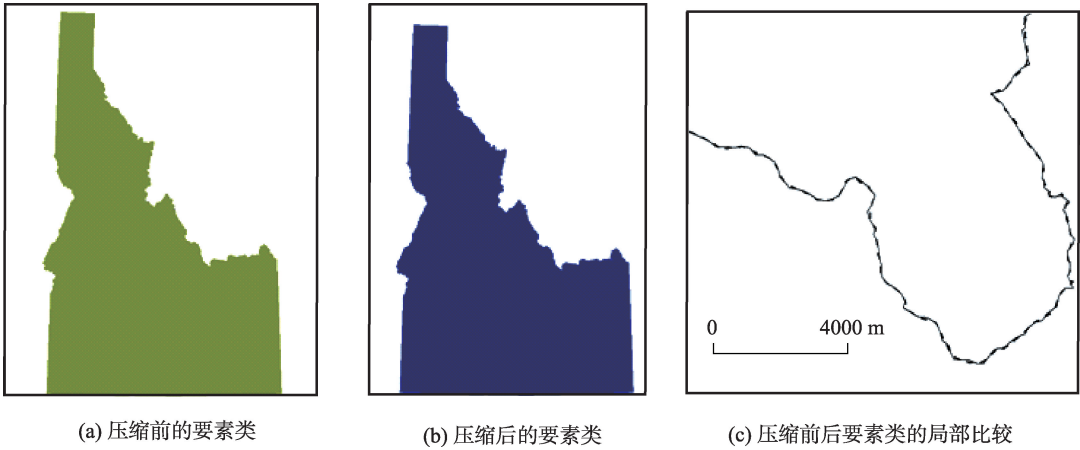


图2 美国爱达荷州的矢量边界图压缩前后全局和局部比较

Fig. 2 The comparison of the original map and compressed map at global and local scales (for the second test featureclass)

16位的整形数据,就可表示一条线要素。另外,传统的空间域的矢量数据压缩方法,主要控制压缩后数据的最大误差,而本文方法可控制平均误差,即可对整体的偏差进行更好地控制。

相对于原有的基于DCT的矢量数据压缩方法来说,本文提出的方法可更加灵活地控制压缩后的平均误差,无需事先根据不同的误差限值设计量化表,且本文提出的方法所控制的是要素级别的平均误差,而非每一分块中的平均误差。

4 结论

本文利用DCT变换和特殊的量化处理方法,以及无损熵编码实现了矢量数据的有损压缩,且无需量化表。实验结果表明,本文方法能实现较高的压缩比,且能将压缩后矢量数据的要素级别的平均误差限制在一定的限值内。相对于空间域的矢量数据压缩方法,本文方法从整体的角度描述压缩后矢量数据的平均失真,为矢量数据压缩的误差控制提供了新的视角。

本文所提出的频率域方法与传统的空间域方法有较大不同,其不会减少矢量数据所存储的点数量。因此,是否能将空间域的方法与频率域的方法相结合,进一步降低矢量数据所需存储空间,提升WebGIS系统的整体效率,有待进一步研究。同时,实验结果表明,使用本文的矢量数据压缩方法时,阈值设置得过小会使得其压缩比不高,而设置得过大则会使失真过于明显或严重扰乱原有拓扑关系,因此,如何选取一个合适的压缩阈值使得压缩结果,可在存储空间和失真方面取得平衡是个值得研究的问题。

参考文献:

- [1] 宋关福,钟耳顺,王尔琪.WebGIS——基于Internet的地理信息系统[J].中国图象图形学报,1998,3(3):251-254.
- [2] 杨建宇,杨崇俊,明冬萍等.WebGIS系统中矢量数据的压缩与化简方法综述[J].计算机工程与应用,2005,40(32):36-38.
- [3] 朱海军,吴华意,李德仁.基于DCT变换的GIS矢量数据压缩技术研究[J].武汉大学学报·信息科学版,2008,32(12):1123-1126.
- [4] 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]. The Canadian Cartographer, 1973, 10(2):112-122.
- [5] 杨得志,闫国年.矢量数据压缩的Douglas-Peucker算法的实现与改进[J].测绘通报,2002(7):18-22.
- [6] 王净,江刚武.无拓扑矢量数据快速压缩算法的研究与实现[J].测绘学报,2003,32(2):173-177.
- [7] Hershberger J, Snoeyink J. An $O(n \log n)$ implementation of the Douglas-Peucker algorithm for line simplification[C]. Proceedings of the tenth annual symposium on Computational geometry, ACM, 1994:383-384.
- [8] Hershberger J E, Snoeyink J. Speeding up the Douglas-Peucker line-simplification algorithm[M]. University of British Columbia, Department of Computer Science, 1992.
- [9] Visvalingam M, Whyatt J D. Line generalisation by repeated elimination of points[J]. The Cartographic Journal, 1993,30(1):46-51.
- [10] 黄培之.具有预测功能的曲线矢量数据压缩方法[J].测绘学报,1995,24(4):16-320.
- [11] 操震洲,李满春,程亮,等.矢量曲线数据的网络渐进传输[J].武汉大学学报·信息科学版,2013,38(4):475-479.
- [12] 操震洲,李满春,程亮,等.矢量数据渐进传输系统的研究与实现[J].计算机应用与软件,2013,30(10):229-231.
- [13] 余先川,张君兰,张立保.基于整数小波变换的空间矢量数据压缩方法[J].地球科学:中国地质大学学报,2011,36(2):381-385.
- [14] 王玉海.基于小波的矢量地图数据压缩模型和方法[D].郑州:中国人民解放军信息工程大学,2001.
- [15] 王玉海,朱长青.基于小波分析的线状要素压缩优化的综合性研究[J].武汉大学学报·信息科学版,2007,32(7):630-632.
- [16] 王玉海,王耀革.基于B样条小波变换的矢量地图数据压缩及边界处理[J].测绘学院学报,2001,18(1):69-71.
- [17] 王玉海,王耀革.基于多进制小波的等高线数据简化[J].测绘学院学报,2001,18(B9):14-16.
- [18] 马伯宁,冷志光,汤晓安,等.具有误差修正的线矢量数据小波变换[J].计算机辅助设计与图形学学报,2011,23(11):1825-1829.
- [19] 吴华意,朱海军.基于DCT的GIS矢量数据压缩技术研究[C].2005年中国地理信息系统协会理论与方法学术会议论文集,2005:68-75.
- [20] 朱海军,吴华意.GIS矢量数据压缩在不同失真准则下频域码率分配算法[C].2005年中国地理信息系统协会理论与方法学术会议论文集,2005:148-153.
- [21] Goyal V K. Theoretical foundations of transform coding[J]. IEEE Signal Processing Magazine, 2001,18(5):9-21.
- [22] ISO/IEC 10918-1. Information technology-digital compression and coding of continuous-tone still images-part 1: Requirements and guidelines[S]. ISO/IEC International Standard 10918-1, ITU-T Rec T81, 1993.

Vector Data Compression of Frequency Domain Based on Tolerance of Average Error

HUANG Weiming, YANG Jianyu*, YUE Yanli, DU Meng, ZHANG Chao and ZHU Dehai

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; 2. Key Laboratory for Agriculture Land Quality, Monitoring and Control of the Ministry of Land and Resources, Beijing 100035, China)

Abstract: With the constant enhancement of data acquisition capability, an increasing amount of spatial data is available to WebGIS. However, the efficiency of spatial data transmission cannot meet the current need for fast access to spatial data. In such a context, the compression of spatial data is important for reducing the space of data storage and improving the efficiency of data transmission and processing for WebGIS. Traditional methods of vector data compression are usually based on spatial relations, such as the distances or angles between different vertices, to determine which vertices are redundant. Whereas in the field of image compression, the method based on the techniques of transform coding (in frequency domain) is more frequently-used. This paper proposes a novel method of spatial vector data compression in frequency domain. This method is composed of several steps. To be specific, we transform the coordinate values of x and y to frequency domain coefficients respectively using Discrete Cosine Transform (DCT), and quantize these coefficients according to the relevant specific thresholds, which could restrict the average distortions (root-mean-square error, RMSE) of the reconstructed vector data. This quantizing method does not need quantization tables and is adaptive to large thresholds that often cause problems such as compressing a polygon or line feature into a point or a polygon feature into a line. In the final stage of the proposed vector data compression method, the Huffman coding is used in a similar way to the corresponding part of JPEG standard. The proposed compression method was implemented by C#.NET and was applied to compress vector data to test its performance from the aspects of compression ratio and geometric shape distortion. The results of the tests show that the proposed method is a feasible solution for vector data compression. It is flexible in terms of restricting distortion, and it can achieve large compression ratios as well as retain the main geographic shape's characteristics which are derived from the original vector data, within the compressed vector data.

Key words: vector data map; WebGIS; DCT; average error; spatial data compression

*Corresponding author: YANG Jianyu, E-mail: ycjyyang@cau.edu.cn