ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.18, No.3, March 2007, pp.739−745
DOI: 10.1360/jos180739

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

数字高程模型数据的信息伪装和信息隐藏技术*

罗 永[1+], 杨岳湘[2], 成礼智[1], 徐志宏[1]

[1](国防科学技术大学 理学院,湖南 长沙 410073)

[2](国防科学技术大学 计算机学院,湖南 长沙 410073)

Information Disguising and Hiding Technique to Protect Digital Elevation Model Data

LUO Yong[1+], YANG Yue-Xiang[2], CHENG Li-Zhi[1], XU Zhi-Hong[1]

[1](School of Science, National University of Defense Technology, Changsha 410073, China)

[2](School of Computer Science, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: Phn: +86-730-4215550, E-mail: yngluo@163.com

Luo Y, Yang YX, Cheng LZ, Xu ZH. Information disguising and hiding technique to protect digital elevation model data. *Journal of Software*, 2007,18(3):739−745. http://www.jos.org.cn/1000-9825/18/739.htm

Abstract: An information disguising and hiding technique is used to protect DEM (digital elevation model) data for store and transformation in this paper. A method is designed to compress the DEM data with a very low bit ratio which would be hidden in the disguised data. The integer wavelet with parameter is used, and a wavelet coefficients set which can embed the information is built. This paper expands the just noticeable distortion (JND) analysis based on the human visual system (HVS) to apply it in the DEM data. Furthermore this paper constructs Hash one-way by the Rabin method, and the algorithm can protect two groups of the DEM data at a time and be public.

Key words: information disguising; information hiding; digital elevation mode; integer wavelet transform with parameter

摘 要: 提出了一种结合信息伪装和隐藏技术来保护数字高程模型(digital elevation model,简称 DEM)数据的方法,保证了 DEM 数据的安全传输和存储.设计了 DEM 数据极低比特率压缩方法,压缩数据隐藏在伪装数据中.应用带参数整数小波变换,提出可嵌入隐藏信息的小波系数集合生成方法.扩展了只针对图像的基于视觉系统(human visual system,简称 HVS)小波域量化噪声的视觉权重(just noticeable distortion,简称 JND)分析方法,使其适用于 DEM 数据,自适应地确定信息隐藏的强度.该方法可同时保护两组 DEM 地形数据.应用 Rabin 方法生成单向 Hash 函数,算法可以完全公开.

关键词: 信息伪装;信息隐藏;数字高程模型;带参数整数小波变换

中图法分类号: TP309 文献标识码: A

数字地形模型(digital terrain model,简称 DTM)[1]是以数字的形式按一定的结构组织在一起,表示实际地形特征的空间分布模型.DTM 主要由栅格(regular square grid,简称 RSG)与不规则三角网(triangulated irregular network,简称 TIN)两种数据格式来表示.它们是地形形状大小和起伏特征的数字描述.栅格数据格式充分表现

* Supported by the National Natural Science Foundation of China under Grant Nos.60573027, 60603014 (国家自然科学基金)
Received 2004-04-04; Accepted 2006-05-11

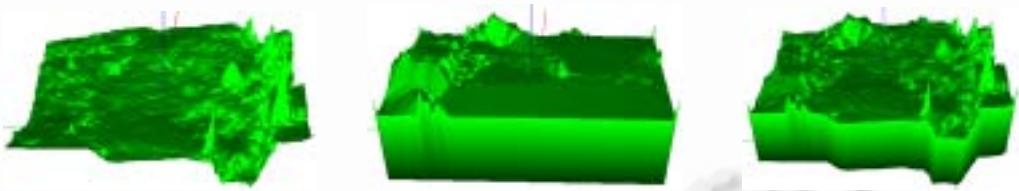
了高程的细节变化,拓扑关系简单,算法实现容易,空间操作及存储方便.

按平面上等间距规则采样或内插所建立的 DTM,为栅格数据的 DTM,可以写成矩阵$(Z_{ij})_{n\times n}$,其中,Z_{ij} 为格网结点 i, j 上的地形属性数据.当该属性为海拔高程时,该模型就成为数字高程模型(digital elevation model,简称 DEM). DEM 使用十分广泛,由于 DEM 数据是实际的海拔高度,在军事和经济上有重要意义,对它的保护是一个很重要的问题.

目前来说,对 DEM 数据保护的方法主要包括加密和隐藏两种方式,其中:文献[2]设计了一种针对 DEM 数据的数字水印算法,通过隐藏水印标志或者数据的关键信息,达到限制非法用户使用和保护版权的目的;文献[3]提出了通过生成可再现的随机地形和构造模糊关系矩阵,采用伪装的方式保护一组 DEM 数据的方法.本文结合信息伪装技术和信息隐藏技术来保护 DEM 数据,实现了对 DEM 数据的伪装和信息隐藏,并且在不需要任何原始 DEM 信息的条件下,可以从伪装数据中恢复两组真实的 DEM 数据的技术.非法的用户只能获得一个伪装过的且没有实际用处的 DEM 数据.

信息伪装定义为:改变信息的原有特征,从而降低或消除信息的可探测和被攻击的特征,并且可以通过逆变换将原信息还原.该技术对于保护三维地理数据的安全是有重要意义的,通过伪装技术来保护 DEM 数据主要有两个方面的优势:(1) 隐蔽性.伪装以后的数据,从形式上来说是一种有意义的数据,非法用户不能分辨数据的真伪;(2) 安全性.非法用户即使知道该数据是一个伪装数据,仍然不能将真实数据恢复出来.

为了实现这种信息伪装,要解决几个问题:(1) 如何伪装,以便不被察觉.本文将两组原数据归一化到相同的范围内,通过单向 Hash 函数生成一个随机的比例序列加权平均来实现伪装;(2) 如何隐蔽并恢复 DEM 数据.将其中的一组 DEM 数据压缩以后隐藏到伪装数据中.在恢复过程中,先提取隐藏在伪装数据中的一组 DEM 数据,然后再现随机的比例,将另一组正确的高程数据恢复出来;(3) DEM 数据极低比特率的压缩.DEM 数据庞大(512×512 的文本格式的 DEM 数据文件大小大于 3.5MB),而隐藏的信息量是有限的,因此,高效的压缩是必需的.并且,这种压缩要保持地形形状和起伏特征不变.



(a) Original DEM data 1 (a) 原 DEM 数据 1
(b) Original DEM data 2 (b) 原 DEM 数据 2
(c) Disguised DEM data (c) 伪装以后的 DEM 数据

Fig.1 The technology of disguising DEM data

图 1 DEM 数据伪装技术

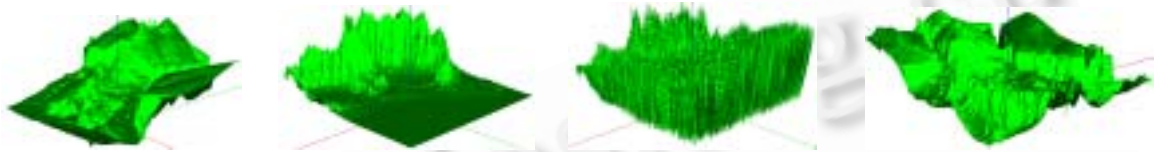
本文采用带参数整数小波变换,为了减少运算便于硬件实现,从构造的带参数小波变换[4]中优选 t=1 整数小波基,只需要移位和加法来实现.压缩质量接近浮点的 CDF9/7 小波[5](图像压缩标准小波变换).

基于视觉系统(human visual system,简称 HVS)小波域量化噪声的视觉权重(just noticeable distortion,简称 JND)分析方法,根据小波系数所在频段提供了一个统一的量化标准以达到视觉效果和量化强度最佳的平衡点.传统的 JND 是针对图像设计的,而 DEM 数据本身的数据范围与图像灰度不同,因此需要扩展.自适应体现在嵌入强度会根据小波系数所在的频段自动确定,而且这个强度是根据 HVS 最优设计的.本文采用扩展了的基于视觉系统小波域量化噪声视觉权重分析方法[6],它适应于高程数据.自适应地确定压缩的高程信息的嵌入强度,提高了隐藏信息的抗攻击能力.

1 信息伪装

为了实现高程信息的伪装,本文采用伪随机序列控制两组 DEM 数据的合成.设两组 DEM 数据分别为 $A=\{a_{i,j}|i\in[0,N_1-1],j\in[0,M_1-1]\}$(尺寸 $N_1\times M_1$)与 $B=\{b_{i,j}|i\in[0,N_2-1],j\in[0,M_2-1]\}$(尺寸 $N_2\times M_2$).为了达到伪装的效

果,又能保证伪装的安全性,采取随机加权平均的方式.设λ为权重($0<\lambda<1$),目标 DEM 数据 C 的高程值 $c(i_1,j_1)=a(i_2,j_2)\times\lambda+b(i_3,j_3)\times(1-\lambda)$.为了防止非法用户去掉伪装(分离出真实的 DEM 数据),本文采用多个权重随机选取.由一个单向 Hash 函数作为伪随机序列发生器(本文第 3 节给出了该函数的形式和产生随机序列的方法)来控制,从包含 D 种权值的集合$\{\lambda_1,\lambda_2,\ldots,\lambda_D\}$中随机选取.每个点的加权平均值是不一样的.也就是说,即使非法用户知道这是一个伪装数据,也不能将正确的、有意义的数据恢复出来.另外,如果相邻的点之间没有一定的相关性,伪装出来的数据就很容易被察觉.如图 2 所示,图 2(c)不符合一个正常的 DEM 地形数据的条件,因此这样的伪装是不成功的.因此在伪装过程中,如果当前伪装点选取的权重与临近点的权重没有一定的相关性(二者的差$|\lambda_{k_1}-\lambda_{k_2}|<\alpha$称为相关的,$\alpha$为选定的阈值),就跳过,重新由 Hash 函数生成一个,直到满足相关性条件为止.在同样的原则下,这个序列仍然是可以再现的,从而保证了两组 DEM 数据可以准确地分离.



(a) DEM data 1 (a) DEM 数据 1　(b) DEM data 2 (b) DEM 数据 2　(c) Disguising without relationship (c) 不带相关性条件伪装　(d) Disguising with relationship (d) 带相关性条件伪装

Fig.2 The importance of relationship in disguising

图 2 相关性条件在伪装中的重要性

2 DEM 数据压缩

良好的压缩性能是保证信息能够隐藏的关键.如果数据量过大,信息隐藏是无法做到的.由于 DEM 数据量庞大,在保证地形形状和地表起伏特征不变的前提下,必须要对 DEM 数据进行极低比特率的压缩.任意选取一组 DEM 数据,压缩之前将 DEM 数据放大 10 000 倍以保证 DEM 数据的精度,然后对高程数据进行取整(整数小波变换要求数据为整数).

图 3 表示数字高程模型数据归一化到[0~255]整数以后,将其转化为灰度模拟图像.从图 3(b)中可以发现,其图像的灰度变化非常缓慢,表明 DEM 数据间相关度大,有大量的冗余信息.所以,可以利用小波图像压缩的原理,取极低比特率来压缩 DEM 数据.实验证明,可以达到非常好的效果.

(a) 3-D show of DEM data (a) DEM 数据三维显示图　(b) 2-D show of digital image transformed from DEM data (b) DEM 数据转化为灰度模拟图像

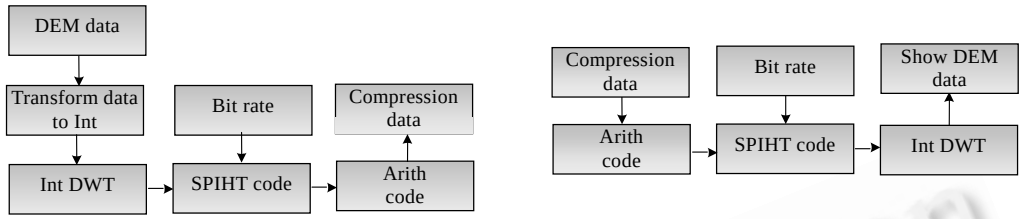
Fig.3 The show of 3-D DEM data and 2-D digital image

图 3 DEM 数据三维显示与灰度模拟图像显示

本文采用 SPIHT(多级树集合分裂算法编码)[7]来实现小波系数的编码,再用算术编码对编码数据作熵编码,最后得到压缩数据.图 4(a)是对 DEM 数据进行小波压缩,图 4(b)是解压缩的流程图.

图 5 是 DEM 数据压缩实验效果图(9/7-1 小波与 CDF9/7 小波作了对照实验).实验中,取极低比特率(bit ratio=0.015625),一个 512×512 的 DEM 数据文件压缩成为一个 500 多个字节的文件,并且能够保持解压后的三维地形形状和起伏特征基本不变.如图 5(a)所示为原始的 DEM 数据,图 5(b)和图 5(c)分别为 CDF9-7 和 9/7-1 小波的压缩结果.通过实验可以发现:DEM 数据有良好的压缩效果,两种小波基的压缩效果接近(图 5(b)的 PSNR=

38.3913db,图 5(c)的 PSNR=37.887db).但是,9/7-1 小波较 CDF9/7 小波的运算复杂度大为降低,且仅采用移位和加法实现,极大地提高了运算速度,降低了硬件实现成本.

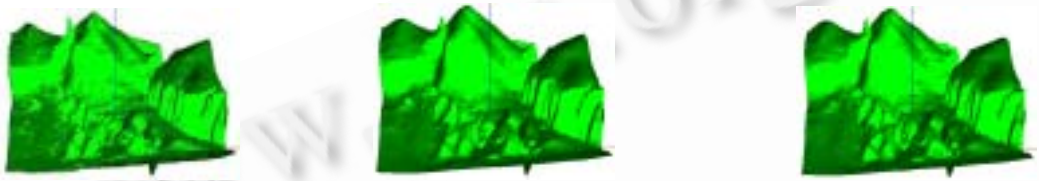


(a) Procedure of compressing DEM data
(a) DEM 数据压缩流程

(b) Procedure of decompressing DEM data
(b) DEM 数据解压流程

Fig.4 Illustration of DEM data compression and decompression

图 4 DEM 数据压缩和解压缩的流程图



(a) The original DEM data
(a) DEM 原始数据

(b) Restored data by compressing with CDF9-7
(b) CDF9-7 压缩数据恢复

(c) Restored data by compressing with 9/7-1
(c) 9/7-1 压缩数据恢复

Fig.5 Experiment comparison between CDF9-7 and 9/7-1 wavelet

图 5 CDF9-7 小波与 9/7-1 小波压缩对照实验

3 信息隐藏

3.1 信息隐藏流程

如图 6 所示,首先将 DEM 高程数据经过前面的方法进行压缩,取极低比特率,使数据压缩到 500 多个字节(512×512);然后,对伪装 DEM 数据进行小波变换,将压缩以后的 DEM 数据隐藏到伪装高程数据的整数小波变换系数中;最后,重构伪装数据,就得到了隐藏信息的伪装高程数据.

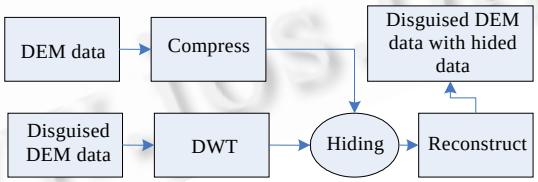


Fig.6 Procedure of hiding information

图 6 信息隐藏流程图

3.2 构造单向Hash函数

根据 Hash 函数的定义,应用 Rabin 方法[8]构造单向 Hash 函数.利用大整数开方取模运算,当大整数位数达到 1 000 比特位时,即使是万亿次(每秒)的计算机也无能为力.

图 7 是单向 Hash 函数生成随机位置的流程图.对每一个伪装数据取一个标识 ID_i,随机选取两个大的素数 p,q,计算参数 $n=p\times q$,这里,p 和 q 是秘密的;n 是公开的.p 和 q 均有 512 比特位,密码设为 K 有 512 比特位.

通过如图 7 所示的循环,可以得到一个序列$\{(t,r)_i\}(0<i<Q)$(Q 表示隐藏信息量,它决定随机序列的长度).为了防止序列中出现重复值,建立一个临时表,记录已经产生的(t,r).每生成一个(t,r),就与表中的序列对照:如果没有与之相同的,就将其写入临时表中,然后计算下一个;如果相同,则不记录到临时表中,然后重新计算.所以,只要

拥有参数和密码,这个序列就可以再现,从而为提取隐藏信息提供了途径.

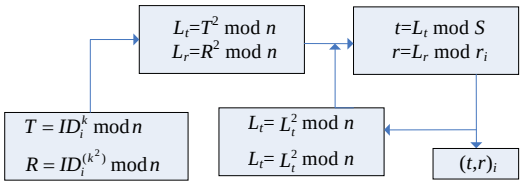


Fig.7 Construct pseudo stochastic sequence based on Rabin method

图 7 应用 Rabin 方法生成随机序列

对于 DEM 数据的伪装,也采用这个 Hash 函数来控制权重的选取.在生成权重的过程中,不需要考虑重复出现值的问题,因为相邻点的合成权重一致是允许的,也不会影响真实高程数据的还原.

3.3 信息隐藏算法

$\{M_i\}(0\le i\le S-1)$为可嵌入隐藏信息的系数集,G 表示集合生成的阈值.生成方法如下:

(1) 搜索绝对值最大的小波系数 $W_{\max}$;

(2) 计算 $T, T=2^{[\log_2 W_{\max}]}$,则 $T<W_{\max}<2T$;

(3) 计算可利用集合数 s,对应于选定的阈值 G,满足$|T/2^s|>G$;

(4) 系数搜索,生成系数集合

$M_0=\left\{w_{k_0^0}, w_{k_1^0}, w_{k_2^0},..., w_{k_{r_0}^0}\right\}$,满足$|w_{k_i^0}|\in[T,2T)$

$M_1=\left\{w_{k_0^1}, w_{k_1^1}, w_{k_2^1},...., w_{k_{r_1}^1}\right\}$,满足$|w_{k_i^1}|\in[T,2T)$

…

$M_{s-1}=\left\{w_{k_0^{s-1}}, w_{k_1^{s-1}}, w_{k_2^{s-1}},...., w_{k_{r_{s-1}}^{s-1}}\right\}$,满足$|w_{k_i^{s-1}}|\in[T/2^{s-1}, T/2^{s-2})$

M_i 中包含的小波系数的数量分别为 $r_i, i=0,1,2,\ldots,S-1$.

嵌入信息过程:

首先,通过单向 Hash 函数产生的伪随机序列$\{(t,r)_i\}(0<i<Q)$(Q 表示压缩数据信息量),根据$(t,r)_i$选取系数集合 M_t;然后,在 M_t 中选取小波系数 $w_{k_r^t}$,通过修改小波系数 $w_{k_r^t}$,将 1bit 信息 b 嵌入([ ]表示取整): $w'_{k_r^t}=[w_{k_r^t}/2T_{l,f}]\times 2T_{l,f}\pm T_{l,f}\times b$;最后,将所有的压缩数据隐藏到伪装高程数据中.

$T_{l,f}$ 为可见阈值(JND)[6]确定的修改幅度,由于 JND 是针对图像(取值范围是 0~255 之间的整数)提出来的,为了使其适用于 DEM 数据,采用对量化因子作缩放调整.

在本算法中,不需要用到任何原始 DEM 数据,合法的用户通过提取隐藏的信息恢复 DEM 数据,从而获得一组真实的高程地理信息.然后通过再现合成的权重序列,将另一组真实的 DEM 数据恢复出来.当然,恢复出来的 DEM 数据与原始高程数据比起来有一定的误差,但是在保证地面起伏特征和地形形状不变的前提下(峰值信噪比在 36db 以上),这种误差是被允许的.

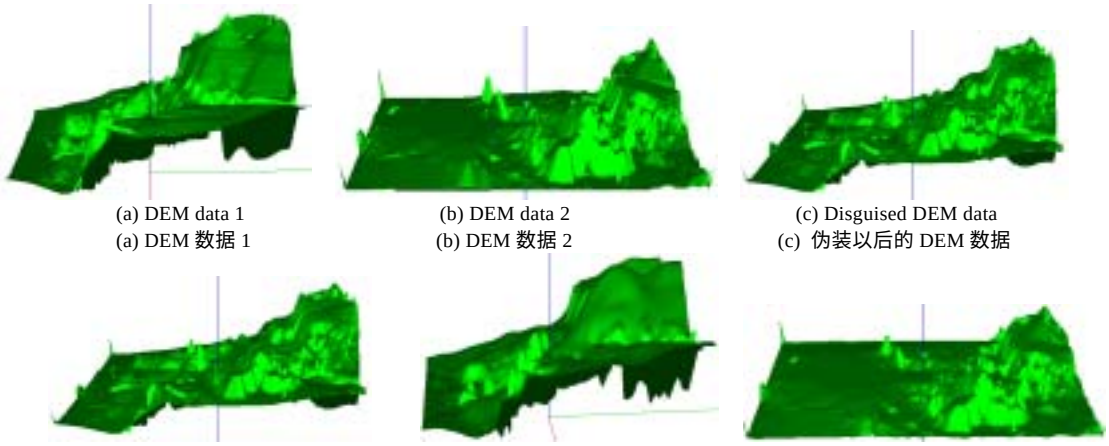
4 实验结果与安全性分析

4.1 实验结果

下面是该算法的演示实验,在实验中,两组 DEM 数据的尺寸都是 512×512.

图 8(a)和图 8(b)表示原始的两组 DEM 数据,图 8(c)是生成的伪装数据,图 8(d)是隐藏了 DEM 数据 1 的伪

装数据.图 8(e)是从图 8(d)中提取并恢复出来的隐藏 DEM 数据.图 8(f)是再现随机比例序列以后,通过隐藏信息的伪装数据(图 8(d))和提取的隐藏数据(图 8(e))恢复出来的 DEM 数据 2.从实验中可以看出,该方法实现了良好的伪装特性,并且两组真实的 DEM 数据都得到了恢复.恢复出来的 DEM 数据保持了地面起伏和地形形状不变.



(a) DEM data 1 (a) DEM 数据 1

(b) DEM data 2 (b) DEM 数据 2

(c) Disguised DEM data (c) 伪装以后的 DEM 数据

(d) Disguised DEM data with hiding information (d) 隐藏信息的伪装 DEM

(e) Restored DEM data 1 (e) 恢复的 DEM 数据 1

(f) Restored DEM data 2 (f) 恢复的 DEM 数据 2

Fig.8 The experiment of disguising DEM data and hiding information

图 8 DEM 数据伪装和信息隐藏实验

4.2 安全性分析

对所有的用户来说,伪装数据是公开的.只有合法的用户或者版权拥有者才可以将真实的两组高程数据恢复出来.由于在隐藏过程中隐藏的是一组 DEM 数据的压缩数据,如果破坏部分压缩数据就不能正常解压,因此,伪装数据不能承受过大的破坏.但是,该方法采用伪装的主要目的是防止非法用户觉察到真实的高程信息的存在(隐蔽性),同时保证真实信息能被非法用户获取.

如果要获得真实的数据就必须知道伪装信息使用的权重序列,而权重序列采用一个单向 Hash 函数控制.没有密码 K,这个序列是不能够准确再现的.当然,如果要这样做还必须知道正确的一组 DEM 数据.如果要获取隐藏的 DEM 数据,就必须知道压缩的 DEM 数据隐藏的位置.如果要获取位置,同样也需要破解单向 Hash 函数.

在位置攻击下,假设非法用户已经获得了伪装的 DEM 数据和单向 Hash 函数,但没有密码 K,一般来说,有两种方法可以获取信息:一种就是直接对单向 Hash 函数进行密码分析.如果 $y=f(x)$是用于函数的单向 Hash 函数,非法用户必须找到 f^{-1} 才行.本方法采用 Rabin 方法和取模运算作为单向 Hash 函数, f^{-1} 是无法找到的.从而非法用户不能获取高程信息隐藏的位置.

另一种方法是分解整数 $n=p\times q$,n 有 1 024 位,$2^{1024}>10^{300}$.现在分解的最大整数是 150 位(10 进制).在现有的硬件条件下,分解 300 位的大整数是不可能的.如果穷举密码 K,K 为 512 位,K 有 2^{512} 种组合.如果非法用户用一台 10 000MPS 的电脑计算,计算时间为 $2^{512}/(10000\times10^6\times60\times60\times24\times356)>10^{130}$ 年.

5 总 结

在应用上,地理信息数据是一类经济利益很大的数据类型,对它的保护关系到国家军事和国民经济.本文采用信息伪装和信息隐藏技术对这类数据进行版权保护和使用控制,隐藏的高程数据不易察觉,并且可以防止非法提取和恢复数据.在算法公开的前提下,仍然能够保证安全,有着广阔的应用前景.

References:

[1] Ware JM. A procedure for automatically correcting invalid flat triangles occurring in triangulated contour data. Computers & Geosciences, 1998,24(2):141−151.

[2] Luo Y, Cheng LZ, Chen B, Wu Y. Study on digital elevation mode data watermark via integer wavelets. Journal of Software, 2005, 16(6):1096−1103 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/16/1096.htm

[3] Luo Y, Cheng LZ, Wu Y, Xu ZH. Research on information disguising technique to protect DEM data based on fuzzy relation. Fuzzy Systems and Mathematics, 2004,18(3):116−120 (in Chinese with English abstract).

[4] Luo Y, Cheng LZ, Xu ZH, Wu Y. A visible Watermark based on integer wavelet transform with parameters. Journal of Software, 2004,15(2):238−249 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/15/238.htm

[5] Calderbank AR, Daubechies I, Sweldens W, Yeo BL. Wavelet transforms that map integers to integers. Applied and Computational Harmonic Analysis, 1998,5(3):332−369.

[6] Watson AB, Yang GY. Visibility of wavelet quantization noise. IEEE Trans. on Image Processing, 1997,6(8):1164−1174.

[7] Said A, Pearlman WA, A new, fast, and efficient image code based on set partitioning in hierarchical trees. IEEE Trans. on Circuits and System for Video Technology, 1996,6(3):243−250.

[8] Merkle R. One way Hash functions and DES. In: Brassard G, ed. Advances in Cryptology, Proc. of the CRYPTO'89. LNCS 435, Springer-Verlag, 1989. 428−446.

附中文参考文献:

[2] 罗永,成礼智,陈波,吴翊.数字高程模型数据整数小波水印算法.软件学报,2005,16(6):1096−1103. http://www.jos.org.cn/1000-9825/16/1096.htm

[3] 罗永,成礼智,吴翊,徐志宏.基于模糊关系的 DEM 数据信息伪装技术研究.模糊系统与数学,2004,18(3):116−120.

[4] 罗永,成礼智,徐志宏,吴翊.基于带参数整数小波变换的可见数字水印.软件学报,2004,15(2):238−249. http://www.jos.org.cn/1000-9825/15/238.htm

罗永(1976－),男,博士,讲师,主要研究领域为应用数学,信息安全,信号与图像处理.

成礼智(1962－),男,博士,教授,博士生导师,主要研究领域为信息科学中新型算法与软件,小波变换与图像处理,应用数学.

杨岳湘(1965－),男,博士生,教授,主要研究领域为网络与信息安全,并行算法.

徐志宏(1977－),女,博士,主要研究领域为工程力学,数值模拟.