中国科学: 地球科学 2018 年 第 48 卷 第 3 期: 340 ~ 352
SCIENTIA SINICA Terrae earthcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
论 文

三孔六边形全球离散格网系统代数编码方法

贲进[1,2*], 李亚路[1,3†], 周成虎[2], 王蕊[1], 杜灵瑀[1]

1. 信息工程大学地理空间信息学院, 郑州 450001;
2. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101;
3. 中国矿业大学地球科学与测绘工程学院, 北京 100083

* 通讯作者, E-mail: benj@lreis.ac.cn
† 通讯作者, E-mail: li.yalu@foxmail.com

收稿日期: 2017-01-09; 接受日期: 2017-09-08; 网络版发表日期: 2017-12-15
国家自然科学基金面上项目(批准号: 41671410)、中国博士后基金特别资助项目(编号: 2013T60161)和信息工程大学优秀青年基金项目(编号: 2016610802)资助

摘要 建立三孔六边形全球离散格网系统代数表达模型, 根据其数学性质设计编码方案. 首先, 引入"格元-格点-格边-格心"的概念, 描述格网各种要素. 其次, 以格心作为格元等效研究对象, 通过相邻层次几何递推关系, 证明可采用复数平面上的混合定位计数系统严格描述不同层次的格心, 揭示格元本质上是一种特殊形式的"复进制数". 第三, 通过间隔层次几何递推关系, 证明在特定约束条件下, 仍可采用混合定位计数系统唯一描述不同层次的格心, 据此设计格元编码方案. 第四, 将编码方案扩展到正二十面体的20个三角面上, 完整描述二十面体封闭表面的多分辨率离散格网. 对比实验结果表明, 本文提出的代数编码方案具有理论严密、利于高效编程实现的优点, 单元编码跨面邻近搜索效率是同类方案的242.9倍. 本文采用的复进制数是理想的格网形式化描述工具, 提出的研究思路可用于创建具有普适性的全球离散格网系统理论框架.

关键词 全球离散格网系统, 六边形, 定位计数系统, 代数编码

1 引言

"全球离散格网系统"(Discrete Global Grid System, DGGS)是一种面向全球的新型多分辨率数据融合与地学模拟解决方案(Dutton, 1984; Sahr等, 2003; Peterson, 2006). 它借助特定方法将地球均匀剖分, 形成无缝无叠的多分辨率格网层次结构, 采用单元地址编码代替传统地理坐标参与数据操作(Goodchild和Yang, 1992). 相较于局部区域的传统空间数据组织、应用模式, 全球离散格网系统更适合解决大尺度的问题, 且在结构上支持多分辨率数据的高效处理(周成虎等, 2009).

全球离散格网系统的单元形状主要有三角形、四边形和六边形三种. 三角形是二维平面上的单纯形, 能够较好地拟合曲面, 因此三角形全球离散格网最早用于大区域空间数据建模及可视化(Dutton, 1984; Fekete, 1990; Goodchild和Yang, 1992; White等, 1998). 尽管如此, 三角形格网仍然具有单元非一致相邻、方向不唯一的缺陷, 不利于空间分析(Sahr等, 2003). 四

中文引用格式: 贲进, 李亚路, 周成虎, 王蕊, 杜灵瑀. 2018. 三孔六边形全球离散格网系统代数编码方法. 中国科学: 地球科学, 48: 340–352, doi: 10.1360/N072017-00013
英文引用格式: Ben J, Li Y L, Zhou C H, Wang R, Du L Y. 2018. Algebraic encoding scheme for aperture 3 hexagonal discrete global grid system. Science China Earth Sciences, 61: 215–227, https://doi.org/10.1007/s11430-017-9111-y

© 2017 《中国科学》杂志社

边形是构建平面格网的最佳图形, 几何结构与常见栅格显示设备相似, 单元存储方式与矩阵一致, 因此四边形全球离散格网广泛应用于海量遥感、地形数据的存储管理(吕雪锋等, 2011). 四边形格网也是非一致相邻且无法完全覆盖全球, 高纬地区的单元畸变或退化严重, 极易导致数据存储冗余(赵学胜等, 2012).

与三角形和四边形格网相比, 六边形格网的拓扑关系一致, 有利于实现邻近、连通等空间分析功能(Sahr等, 2003); 拓扑距离与欧氏空间直线距离的模式最匹配, 有助于建立格网度量空间(Birch等, 2007); 空间采样效率最高, 有助于数据可视化表达(Birch等, 2007). 因此, 作者认为六边形全球离散格网系统为海量空间数据的建模与分析提供了一种新颖的技术思路——融合多源信息, 利用六边形格网的优异特性, 在球面空间更方便地处理多分辨率数据, 更准确地表达计算结果.

到目前为止, 基于正多面体的球面递归剖分仍是构建六边形全球离散格网系统的最有效方案. Kimerling等(1999)在环境监测和评估项目(Environmental Monitoring and Assessment Program, EMAP)中提出了著名的"Kimerling理想格网标准", 最终选中去顶二十面体(形如足球)的六边形全球等积离散格网(White等, 1998). Sahr等(2003)提出了全球离散格网系统设计的5大要素, 分析认为基于正二十面体的三孔等积六边形格网(Icosahedral Snyder Equal Area aperture 3 Hexagonal grid, ISEA3H)具有较好的几何属性. 贲进等(2007)设计了正二十面体上的多种六边形格网生成算法, 谢文军(2010)采用半蜂窝梯形构造类似的格网系统.

六边形不具备自相似性, 很难直接建立六边形格网系统层次关系. Gibson和Lucas(1982)提出了一种多维可扩展的数据结构, 其二维特例可通过"广义平衡三进制"(Generalized Balanced Ternary, GBT)运算实现不同层次六边形单元的索引. 然而, 该结构是由高分辨率子单元聚合为低分辨率父单元, 且相邻层次单元方向发生旋转, 无法直接移植到球面(张永生等, 2007). Sahr(2005)结合ISEA3H格网系统特点, 设计了一种特殊的六边形分解/聚合方案, 借助与GBT类似的算术运算实现不同层次单元快速索引. Vince(2006a)证明采用"平衡三进制"(Balanced Ternary, BT)可有效描述正八面体的三孔六边形格网系统层次关系并设计了"重心索引", 后续又提出"进制索引"并给出部分结论(Vince, 2006b). Zheng(2007)、Vince和Zheng(2009)设计了正二十面体的三孔六边形格网系统编码方案和傅里叶变换算法. 张永生等(2007)通过"单元隶属图形"建立四孔六边形格网系统层次关系并设计了索引算法. 白建军(2011)、贲进等(2011)在BT算法的基础上设计了效率极高的四孔六边形格网系统索引算法. Tong等(2013)设计了"六边形四元平衡结构"(Hexagonal Quaternary Balanced Structure, HQBS)描述四孔六边形格网系统单元中心和顶点构建的混合四叉树, 实现了单元编码和空间操作算法. Amiri等(2015a)在正多面体上建立"菱形"逻辑结构, 实现了适用于多种六边形全球离散格网系统的高效可视化算法. Amiri等(2015b)给出了全球离散格网系统索引的分类, 设计了不同索引之间的转换方法.

在工程实践方面, 欧空局(European Space Agency, ESA)在"生命行星计划"(Living Planet Programme)的"土壤湿度和海洋盐分任务"(Soil Moisture and Ocean Salinity Mission, SMOS)中采用ISEA4H格网系统采集、记录和处理数据(Brockmann等, 2013). 加拿大的PYXIS Innovation公司研发了基于ISEA3H格网系统的空间数据整合软件, 其核心技术是PYXIS格网索引专利(Peterson, 2006).

尽管当前六边形全球离散格网系统的研究在格网构建、编码及索引等方面取得了长足进展, 但从诸多成果的研究方法来看, 大多仍以"归纳-猜想-实验"为主. 虽然Vince(2006b)引入"进制"思想描述三孔六边形层次格网, 但是部分结论缺乏严谨的证明. 张永生等(2007)基于单元隶属图形设计的索引方法则完全根据六边形的几何分解规律归纳得出, 尽管通过实验验证了它的正确性, 但在理论上并不严密. 六边形全球离散格网系统高效可视化(Amiri等, 2015a)、不同格网系统转换(Amiri等, 2015b)侧重算法的研究, 缺少理论的提炼和总结. HQBS(Tong等, 2013)同样是归纳得出相邻层次格网单元中心和顶点的继承关系, 据此设计编码索引算法. 实验证实: HQBS在运算过程中频繁出现编码"正则化"失败需要回退重算的情况(图1a), 且回退单元比例随格网层次的增加呈升高趋势(图1b), 严重影响计算效率. 然而, 由于HQBS编码运算规则由归纳得出, 缺乏必要的理论依据和严谨证明, 几乎不可能从原理上分析导致这一缺陷的根源, 对算法的改进和完善更加困难.

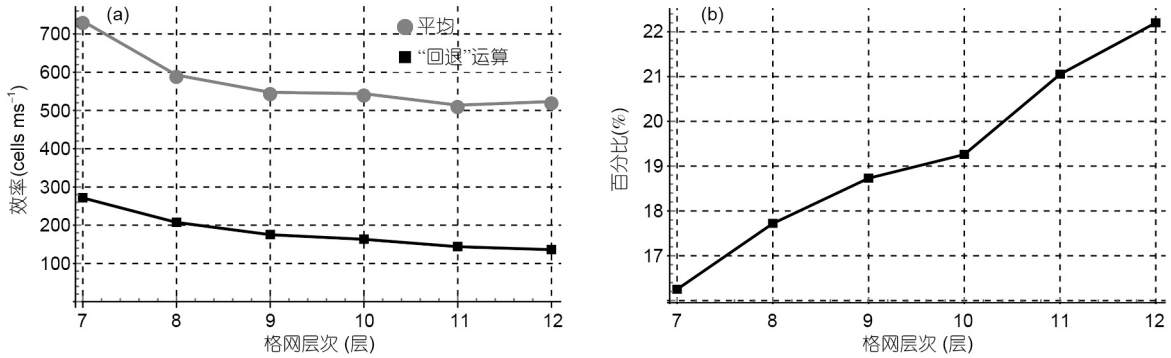


图 1　HQBS编码加法运算实验结果

(a) 平均运算效率与回退运算效率对比; (b) 回退运算单元所占百分比. 硬件配置: Intel Core i7-2720QM@2.2GHz四核心, 16G RAM, SAMSUNG mSATA 128G SSD; 操作系统: Win 7 X64旗舰版 SP1; 开发工具: Microsoft Visual C++ 2012 Update 5, 代码编译成Release版本运行

与计算机系统中广泛应用的二进制编码类似, 全球离散格网系统中空间位置的表示、计算完全依赖单元编码实现, 故编码自身性质、编码相互作用规律是影响格网数据处理效率和可靠性的核心问题. 理想的研究思路应当是建立严密数学模型, 通过严格理论推导, 而不仅仅是归纳, 发现其中的规律. 尤其对于六边形格网系统, 其层次关系描述及编码方案设计仍是学术界有待突破的研究难点. 因此, 本文尝试在周成虎等(2009)提出的"格元-格点-格边-格心"概念模型基础上, 以格心作为描述格元(单元)的等效对象, 借助"进制数"思想, 建立三孔六边形格网系统代数编码模型, 通过严格证明揭示其数学本质. 正如二进制编码的表示和性质奠定了计算机系统的理论基础那样, 本文提出的单元编码方案及相关性质初步构建了六边形全球离散格网系统编码运算的理论体系, 从根本上保证了编码运算的可靠性, 同时也提高了编码运算的效率.

2　格网要素定义

为全面描述格网的各种属性, 定义格元、格点、格边和格心要素, 如图2所示. 格网中的面状单元称为格元, 代表了地表区域的面状特征; 格点是格元的角点, 确定了格元的形状和最大覆盖范围; 格边是格元的边, 通常用于度量格元间的通量关系和相互作用; 格元的中心称为格心, 是格元覆盖区域的定位点, 也是对象、数据与格元建立联系的常用关联点(周成虎等, 2009).

在正八面体或正二十面体的三角面上可进行三孔、四孔六边形层次递归剖分(Sahr等, 2003), 根据单元中心连线与三角面三边的关系, 相同孔径的剖分(格网)又可进一步细分为I类(连线与边平行)、II类(连线与边垂直)和III类(连线与边既不平行也不垂直)(Kenner, 2003). 三孔六边形格网系统格元分辨率的渐变较四孔格网系统平滑, 方向随剖分层次在I类和II类之间交替变化, 如图3所示.

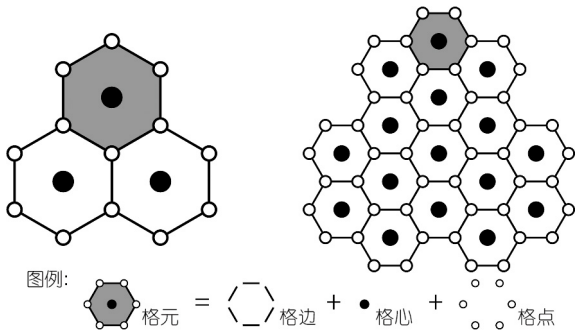


图 2　格网要素示意图

三角面上的三孔六边形格网系统是平面上的多分辨率规则格网, 在同一层次上格元排列紧密且形状、大小完全相同. 对于每个格心x, 格元是平面上到x的距离比到其他格心距离更短的点的集合, 即x的Voronoi图(Vince和Zheng, 2009). 因此在平面规则格网中, 格心与格元一一对应, 下文在不引起歧义的情况下, 采用格心等效替代格元建立格网数学模型.

3　格心层次关系的进制描述

格心位置通常采用笛卡尔坐标描述(Sahr, 2005; 贲进等, 2007; 张永生等, 2007; 白建军, 2011; Tong等,

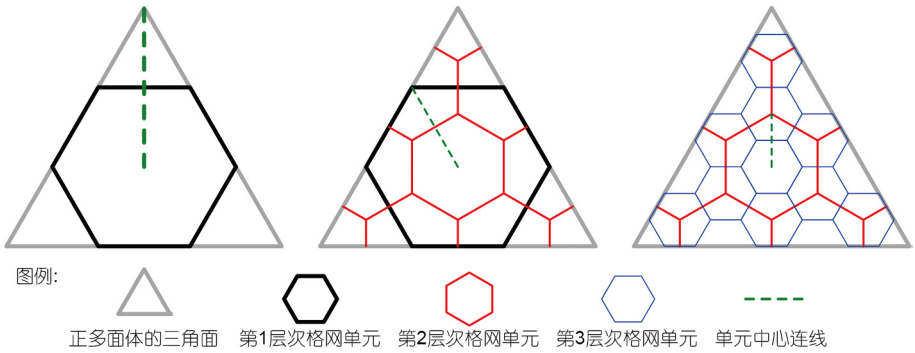


图 3 三孔六边形格网系统示意图

2013), 这种方法虽然直观形象、便于理解, 但由于采用了二维独立变量(如x、y), 表达形式不简洁且不易与编码建立联系. 为了便于推导证明及编码代换, 本文采用复数表示格心位置(Vince, 2006a, 2006b). 根据三孔六边形格网系统生成规律, 第$n(n\geq1)$层格心是由第$n-1$层格心与$n-1$层格点组合而成, 第$n-1$层格边长度是第n层格边长度的$\sqrt{3}$倍, 且格元方向随格网层次交替变化, 将奇数层和偶数层的格元分别命名为"I类型"、"II类型"以示区分, 如图4所示.

设某个I类、II类格元的格心均位于复数平面的原点处, 单位圆上与实数轴R夹角为$\frac{\pi}{3}$的向量对应的复数记为$\omega=\frac{1}{2}+\frac{\sqrt{3}}{2}i$. 易知I类格点可表示为$\frac{1}{\sqrt{3}}\{\omega+\omega^2,\omega^2+\omega^3,\omega^3+\omega^4,\omega^4+\omega^5,\omega^5+\omega^6,\omega^6+\omega\}$, II类格点表示为$\{\omega,\omega^2,\omega^3,\omega^4,\omega^5,\omega^6\}$, 如图5所示. 令$D_o=\{0,\omega,\omega^2,\omega^3,\omega^4,\omega^5,\omega^6\}$、$D_e=\{0,\omega+\omega^2,\omega^2+\omega^3,\omega^3+\omega^4,\omega^4+\omega^5,\omega^5+\omega^6,\omega^6+\omega\}$, 根据网络版附录(http://earthcn.scichina.com)定理一的证明, 三孔六边形格网系统的格心集合:

$$\mathbb{L}_n=\sum_{i=1}^{\left\lfloor\frac{n+1}{2}\right\rfloor}\frac{1}{3^i}D_o+\sum_{j=2}^{\left\lceil\frac{n+1}{2}\right\rceil}\frac{1}{3^j}D_e\quad(n\geq1),$$

其中, "Σ"、"+"是集合之间的累加、加法运算符; "⌊⌋"、"⌈⌉"是向下取整、向上取整运算符.

令$D_o^*=\{\omega,\omega^2,\omega^3,\omega^4,\omega^5,\omega^6\}$, $D_e^*=\{\omega+\omega^2,\omega^2+\omega^3,\omega^3+\omega^4,\omega^4+\omega^5,\omega^5+\omega^6,\omega^6+\omega\}$, 在定理一的基础上, 推导出格心$\alpha\in\mathbb{L}_n$所在格元的格点($v$)、格边($\epsilon$)数学表达式:

$$v=\begin{cases}\alpha+\frac{1}{3}\left(\frac{1}{\sqrt{3}}\right)^n D_o^* & n=2k-1\\ \alpha+\frac{1}{3}\left(\frac{1}{\sqrt{3}}\right)^{n+1} D_e^* & n=2k\end{cases}\quad(k=1,2,3,\cdots),$$

$$\epsilon=\left\{\overline{v_iv_j}\;\middle|\;i=1,2,\cdots,6, j=i\%6+1\right\}.$$

定理一表明, $\mathbb{L}_n$是复数平面上的有限集合, 经计算可得共有格心$6(3^n-2^n)+1$个, 且不同层次格心集合之间满足单向包含关系, 即$\mathbb{L}_1\subset\mathbb{L}_2\subset\cdots\subset\mathbb{L}_n$.

不仅如此, 定理一还揭示了一个重要的事实—格心与实数域上的二进制、八进制、十进制等数字的本质完全相同, 是特定形式的"复进制数"或"定位计数系统"(Knuth, 2002). 对于该定位计数系统, 从两个角度加以解释.

第一, 认为该定位计数系统的进制是3, 奇数层数字位取自集合D_o, 偶数层数字位取自集合D_e, 数字位集合$D=D_o\cup D_e$. 除第一层格心是一个数字之外, 后续都是两个数字表示一个进制位. 对于$\alpha\in\mathbb{L}_n(n\geq1)$,

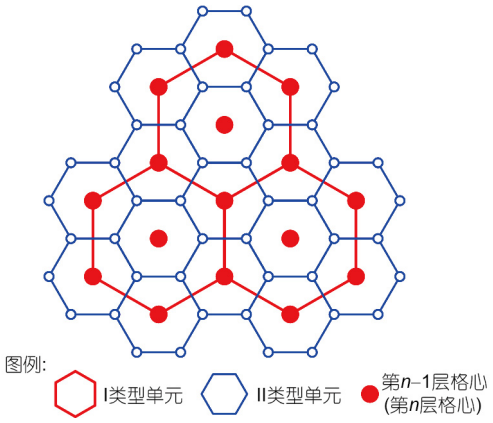


图 4 格心继承关系示意图

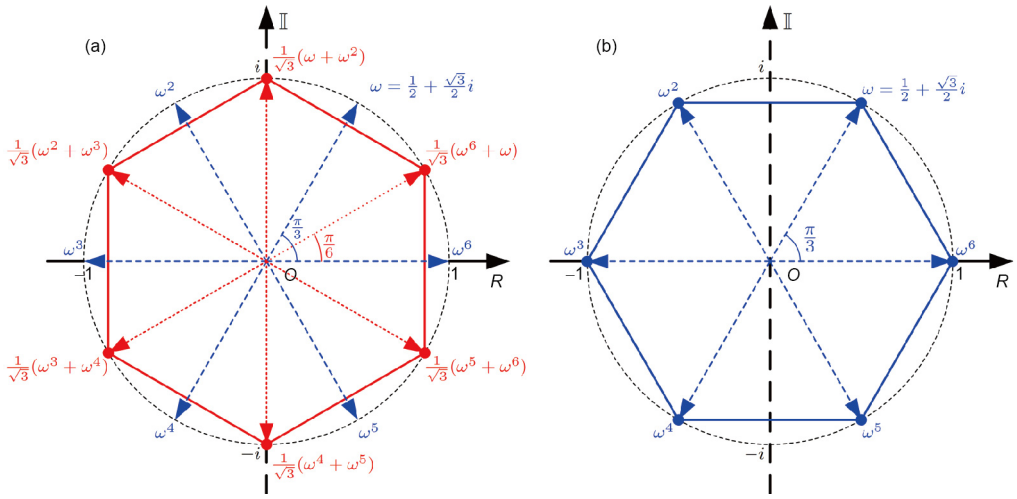


图 5　三孔六边形格元在复数平面上的表示

(a) I类; (b) II类

其具体形式为

$$\alpha = \begin{cases} \dfrac{1}{3}d_1 + \displaystyle\sum_{i=2}^{\frac{n+1}{2}} \frac{1}{3^i}\left(d_{2i-2} + d_{2i-1}\right) & n = 1,3,5,\cdots \\ \dfrac{1}{3}d_1 + \displaystyle\sum_{i=2}^{\frac{n}{2}} \frac{1}{3^i}\left(d_{2i-2} + d_{2i-1}\right) + \frac{1}{3^{\frac{n+2}{2}}}d_n & n = 2,4,6,\cdots \end{cases} \quad \left(d_{2k-1} \in D_o, d_{2k} \in D_e, k = 1,2,3,\cdots,\left\lceil \frac{n}{2} \right\rceil\right), \tag{1}$$

参照十进制数的表达, 省略(1)式中幂指数的底, 记为$\alpha \equiv .d_1(d_2d_3)(d_4d_5)\cdots(d_{n-1}d_n)$(奇数层)或$\alpha \equiv .d_1(d_2d_3)(d_4d_5)\cdots(d_{n-2}d_{n-1})d_n$(偶数层), 计数点可进一步省略.

第二, 认为该定位计数系统由奇数层、偶数层两个进制为3的子系统组合而成, 奇数层数字位取自集合D_o, 偶数层数字位取自集合D_e. 各个子系统的所有层次均是一个数字表示一个进制位. 对于$\alpha \in \mathbb{L}_n (n \geq 1)$, 其具体形式为

$$\alpha = \begin{cases} \displaystyle\sum_{i=1}^{\frac{n+1}{2}} \frac{1}{3^i}d_{2i-1} + \sum_{j=2}^{\frac{n+1}{2}} \frac{1}{3^j}d_{2j-2} & n = 1,3,5,\cdots \\ \displaystyle\sum_{i=1}^{\frac{n}{2}} \frac{1}{3^i}d_{2i-1} + \sum_{j=2}^{\frac{n+2}{2}} \frac{1}{3^j}d_{2j-2} & n = 2,4,6,\cdots \end{cases} \quad \left(d_{2k-1} \in D_o, d_{2k} \in D_e, k = 1,2,3,\cdots,\left\lceil \frac{n}{2} \right\rceil\right), \tag{2}$$

省略计数点后, 格心可记为$\alpha \equiv (d_1d_3d_5\cdots d_n, d_2d_4d_6\cdots d_{n-1})$ 或$\alpha \equiv (d_1d_3d_5\cdots d_{n-1}, d_2d_4d_6\cdots d_n)$.

式(1)直接根据相邻层次格元几何关系建立定位计数系统及编码方案, 其特点是简单直观、便于理解. 式(2)则通过隔层格元几何关系建立定位计数系统, 其特点是将奇、偶层次分而治之, 兼顾各层的联系, 这一思想在下文中将有重要应用.

以上定位计数系统说明, 格心随格网层次向内部生长, 越来越密集. 用不同大小和颜色的点表示$\mathbb{L}_1$~$\mathbb{L}_6$中的元素, 它们在复数平面上的分布如图6所示.

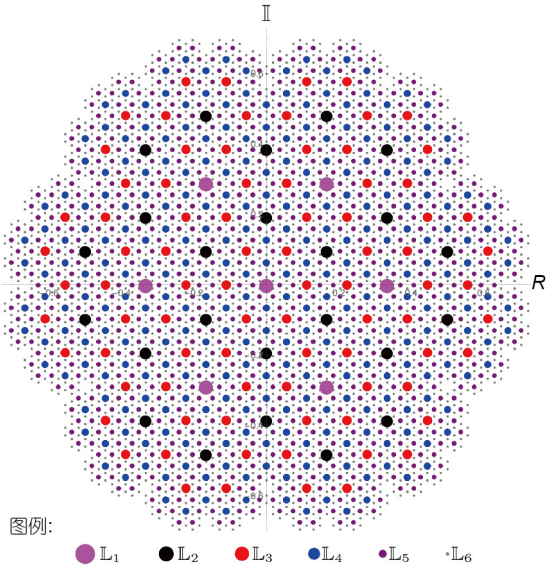


图 6　$\mathbb{L}_1$~$\mathbb{L}_6$在复数平面上的分布

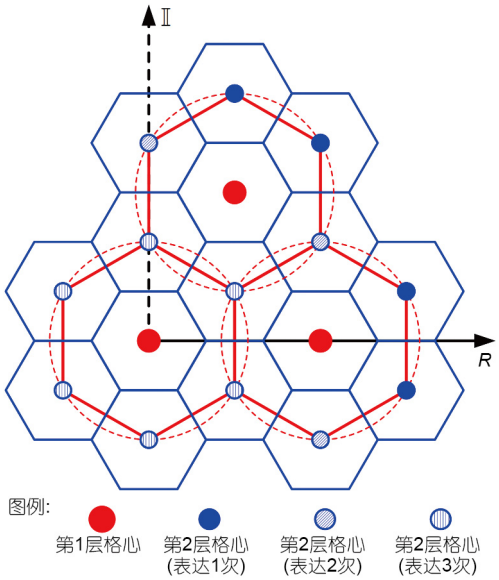


图 7　$\mathbb{L}_2$中格心的多个表达形式

4　格心层次关系的唯一编码

虽然复进制数能够准确描述格心集合, 但α的多个表达式可能对应同一个格心, 或同一个格点分别在式(1)和(2)中有多个表达形式. 例如, 图7中位于$\mathbb{L}_1$格元外接圆交点、属于$\mathbb{L}_2$的格心就有多个表达形式, 后续各层都存在类似现象.

导致这一问题的根本原因是六边形不具备自相似性, 不可避免地会出现一个子格元隶属于多个父格元的情况. 这种“一对多”的映射不利于格网的标识, 必须在定理一的基础上给出限定条件, 使其满足唯一性要求.

令$\mathbb{P}_0=\{0\},\mathbb{P}_1=\frac{1}{3}D_o^*\cup\{0\}$. 根据网络版附录定理二的证明, $\mathbb{L}_n$中部分格点可通过递推关系式

$$\mathbb{P}_n=\begin{cases}\mathbb{P}_{n-1}\cup\left(\mathbb{P}_{n-2}+\frac{1}{3^{\frac{n+1}{2}}}D_o^*\right) & n=2k-1\\ \mathbb{P}_{n-1}\cup\left(\mathbb{P}_{n-2}+\frac{1}{3^{\frac{n+2}{2}}}D_e^*\right) & n=2k\end{cases}\quad(k=1,2,3,\cdots)$$

表示, 且存在唯一确定的

$$\mathrm{e}_i\in\begin{cases}\frac{1}{3^{\frac{i+1}{2}}}D_o^*\cup\{0\} & i=2k-1\\ \frac{1}{3^{\frac{i+2}{2}}}D_e^*\cup\{0\} & i=2k\end{cases}\quad(k=1,2,3,\cdots),$$

使得$\beta\in\mathbb{P}_n$可表示为$\sum_{i=1}^{n}\mathrm{e}_i$, 其中$\mathrm{e}_i=0$或$\mathrm{e}_{i+1}=0$ $(1\le i\le n-1)$.

定理二进一步揭示了三孔六边形格网系统的层次结构: $\beta_{n-2}\in\mathbb{P}_{n-2}$共有7个子格元, 可视为在$\mathbb{P}_{n-1}$上只有1个中心子格元的格心$\beta_{n-1,0}$(位置与$\beta_{n-2}$相同), 在$\mathbb{P}_n$上有6个顶点子格元的格心, 如图8所示.

不仅如此, 定理二还为格网编码方案的设计提供了直接理论支撑. 在奇数层用$k(1\le k\le 6)$替换$\omega^k\in D_o^*$, 在偶数层用$k(1\le k\le 6)$替换$\omega^k+\omega^{k+1}\in D_e^*$, 用编码0替换数字0, e_i中的系数由层次隐含表示. 则$\beta\in\mathbb{P}_n$可用编码唯一表示为$\beta(a_1a_2\cdots a_n)(a_i\in\{0,1,2,\cdots,6\},i=1,2,\cdots,n)$, 且编码中不存在两个连续的非零位. 显然, β在$\mathbb{P}_{n+1}$上的1个中心子格心的编码为$\gamma(a_1a_2\cdots a_n0)$, β在$\mathbb{P}_{n+2}$上的6个顶点子格心的编码为$\eta_i(a_1a_2\cdots a_n0i)(i=1,2,\cdots,6)$.

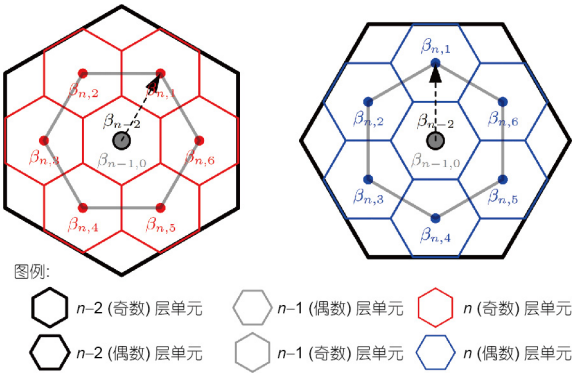


图 8　隔层格元递推关系

用不同大小和颜色的点表示$\mathbb{P}_0$~$\mathbb{P}_5$中的元素, 它们在复数平面上的分布如图9所示.

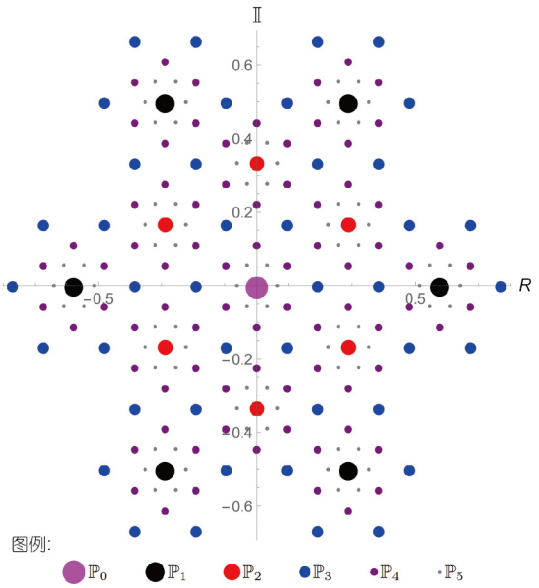
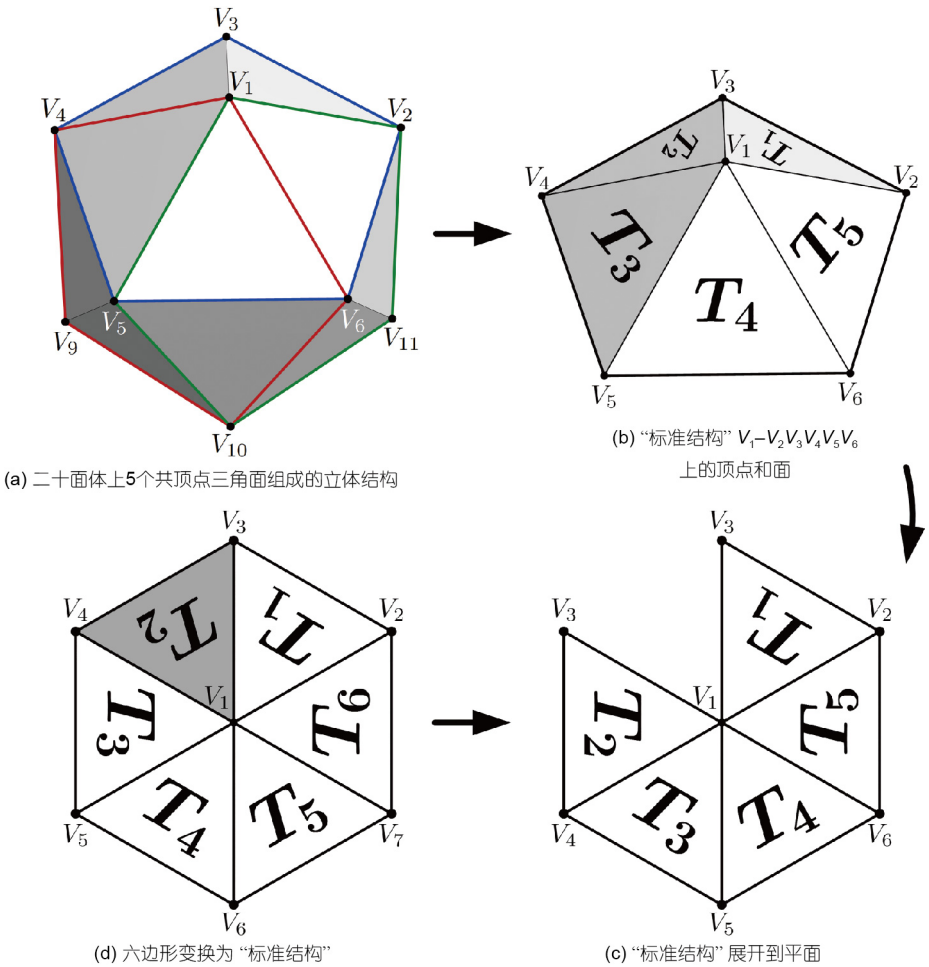


图 9　$\mathbb{P}_0$~$\mathbb{P}_5$在复数平面上的分布

5　格心在正二十面体上的扩展

上文仅讨论了平面多分辨率格心的层次关系, 而在全球离散格网系统的研究中, 通常在正二十面体三角面上构建多分辨率格网, 因此上述格心的描述还必须扩展到正二十面体上. 考察正二十面体的几何结构, 任一顶点周围均匀分布了5个相邻顶点, 形成了由5个共顶点三角面组成的立体结构, 如图10a中V_1–$V_2V_3V_4V_5V_6$、 V_5–$V_1V_4V_9V_{10}V_6$、 V_6–$V_1V_5V_{10}V_{11}V_2$等. 这一结构是描述二十面体顶点、边和面相互关系的关键, 为了便于论述和计算, 定义V_1–$V_2V_3V_4V_5V_6$为“标准结构”, 如图10b所示. 其中仅有V_1与其他5个顶点不共面, 沿V_1V_i(i=2,3,4,5,6)中的任意一条边(假设是V_1V_3)



(a) 二十面体上5个共顶点三角面组成的立体结构

(b) “标准结构” V_1–$V_2V_3V_4V_5V_6$ 上的顶点和面

(c) “标准结构” 展开到平面

(d) 六边形变换为 “标准结构”

图 10　“标准结构”到平面的变换

将标准结构“剪开”, 可将其展开到平面上, 如图10c所示. 这是围绕公共顶点V_1的5个平面三角形, 也可等效看作从V_1周围连续排列、完整覆盖平面的6个三角形T_1~T_6中去除T_2得到, 如图10d所示.

令$H^* = \bigcup_{i=1}^{6} T_i$表示围绕$V_1$的$T_i$拼合而成的六边形, H_i'表示从H^*中去除T_i得到的图形, 记为$H_i' = H^* \setminus T_i$. 定义算子$\mathcal{R}(H'_i)$是对H_i'中三角形、顶点按逆时针顺序编号操作, 则标准结构在平面上的展开图形$H_i = \mathcal{R}(H'_i)$. 根据二十面体的几何特点, 按照表1轮换标准结构中的各个顶点, 封闭表面均可采用图11所示标准结构描述.

借助标准结构, 建立二十面体上的格心数学模型. 令$\mathbb{P}_n^k (k = 1,2,\cdots,6$, 下同)为$\mathbb{P}_n$中第一个非零位编码为$k$的格心组成的集合,$\mathbb{P}'_{n,k}$表示从$\mathbb{P}_n$中去除$\mathbb{P}_n^k$后得到格心组成的集合, 即$\mathbb{P}'_{n,k} = \mathbb{P}_n \setminus \mathbb{P}_n^k$. 去除$\mathbb{P}'_{n,k}$中格心在原点处格元的一边, 原点处的六边形变为五边形. 因此,$\mathbb{P}_{n,k+1}$与$\mathbb{P}_{n,k-1}$(k+1、k−1在{1,2,…,6}中循环取值, 下同)相邻, 相互拼合后, 形成标准结构上的格心集合$\mathbb{P}_{n,k}$, 如图12所示.

为了便于计算机程序在$\mathbb{P}_n$中计算$\mathbb{P}_{n,k}$的接边情况, 用β^+表示对β中每个非零位加1并对6取模运算得到的结果, 即格心围绕原点逆时针旋转60°. 设$\beta \in \mathbb{P}_n^{k-1}$、$\gamma \in \mathbb{P}_n^k$对应的格元有公共边, 则在$\mathbb{P}_n$中有$\beta^+ \in \mathbb{P}_n^k$、$\gamma^+ \in \mathbb{P}_n^{k+1}$对应的格元也有公共边, 且$\gamma^+$为$\beta$在$\mathbb{P}_{n,k}$上的相邻格元的格心.

表 1　标准结构在正二十面体上的分布

二十面体顶点	标准结构顶点(按顺序逆时针排列)	标识
V_1	V_1−$V_2V_3V_4V_5V_6$	H_2
V_2	V_2−$V_3V_1V_6V_{11}V_7$	H_2
V_3	V_3−$V_4V_1V_2V_7V_8$	H_2
V_4	V_4−$V_5V_1V_3V_8V_9$	H_2
V_5	V_5−$V_6V_1V_4V_9V_{10}$	H_2
V_6	V_6−$V_2V_1V_5V_{10}V_{11}$	H_2
V_7	V_7−$V_8V_3V_2V_{11}V_{12}$	H_5
V_8	V_8−$V_9V_4V_3V_7V_{12}$	H_5
V_9	V_9−$V_{10}V_5V_4V_8V_{12}$	H_5
V_{10}	V_{10}−$V_{11}V_6V_5V_9V_{12}$	H_5
V_{11}	V_{11}−$V_7V_2V_6V_{10}V_{12}$	H_5
V_{12}	V_{12}−$V_{11}V_{10}V_9V_8V_7$	H_5

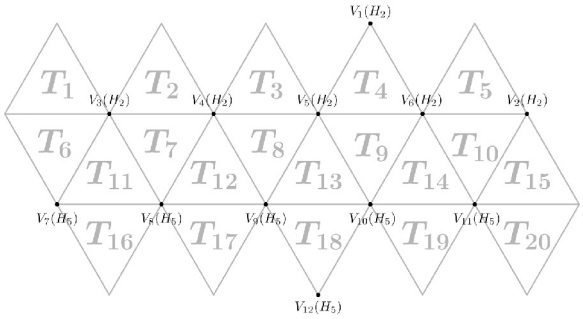


图 11　“标准结构”在二十面体上的分布

记正二十面体封闭表面的三孔六边形离散格网系统第n层对应的格心集合为$\mathbb{H}_n (n \geq 1)$, 根据上文分析,$\mathbb{H}_n$由两部分组合而成: 第一部分是原点位于20个三角面几何中心的$\mathbb{P}_{n-2}$, 记为$\mathbb{P}_{n-2}^{(i)} (i = 1,2,\cdots 20)$; 第二部分是原点位于12个顶点处的$\mathbb{P}_{n,k} (k = 1,2,\cdots,12)$. 用公式表示:

$$\mathbb{H}_n = \bigcup_{i=1}^{20} \mathbb{P}_{n-2}^{(i)} \cup \bigcup_{k=1}^{12} \mathbb{P}_{n,k}, \tag{3}$$

式(3)$\mathbb{P}_{n-2}^{(i)}$之间除原点不同之外完全等价, 可视为$\mathbb{P}_{n-2}$在不同三角面上的复制, 需要在格点编码前添加面索引标识加以区别, 如$\beta(F_i, a_1a_2\cdots a_{n-2})$. 而$\mathbb{P}_{n,k}$则与顶点有关, 编码结构也略有差异, 同样需要添加顶点标识加以区别, 如$\gamma(V_k, a_1a_2\cdots a_n)$. $\mathbb{P}_{n-2}^{(i)}$与$\mathbb{P}_{n,k}$相差的2位编码, 需根据F_i、V_k的实际位置关系在β编码的最前面添加.

6　实验与分析

作者采用Visual C++ 2012 Update 5开发了正二十面体的格网编码运算原型系统GridView 1.2, 核心二十面体类的定义如表2所示, DgStdUnit类和DgTriFacet类分别对应式(3)中的$\mathbb{P}_{n,k}$和$\mathbb{P}_{n-2}^{(i)}$. 该系统可根据用户输入层次$n(n \geq 2)$生成格网, 同时实现了单元编码之间的相加、相乘运算. 格网生成的基本原理是: 首先根据定理二生成编码集合$\mathbb{P}_n$、$\mathbb{P}_{n-2}$, 然后根据图12的原理处理$\mathbb{P}_n$得到12个顶点上的$\mathbb{P}_{n,k}$, 接着复制$\mathbb{P}_{n-2}$到20个三角面上得到$\mathbb{P}_{n-2}^{(i)}$, 最后显示完整的$\mathbb{H}_n$的三维图形. 算法流程如图13所示, 运行结果如图14、15所示, 采用Snyder等积多面体投影(Snyder, 1992; 赁进等, 2006)生成球面格网, $\mathbb{P}_{n-2}^{(i)}$用红色填充以示区别.

为了验证本文提出的标准结构在单元编码跨面

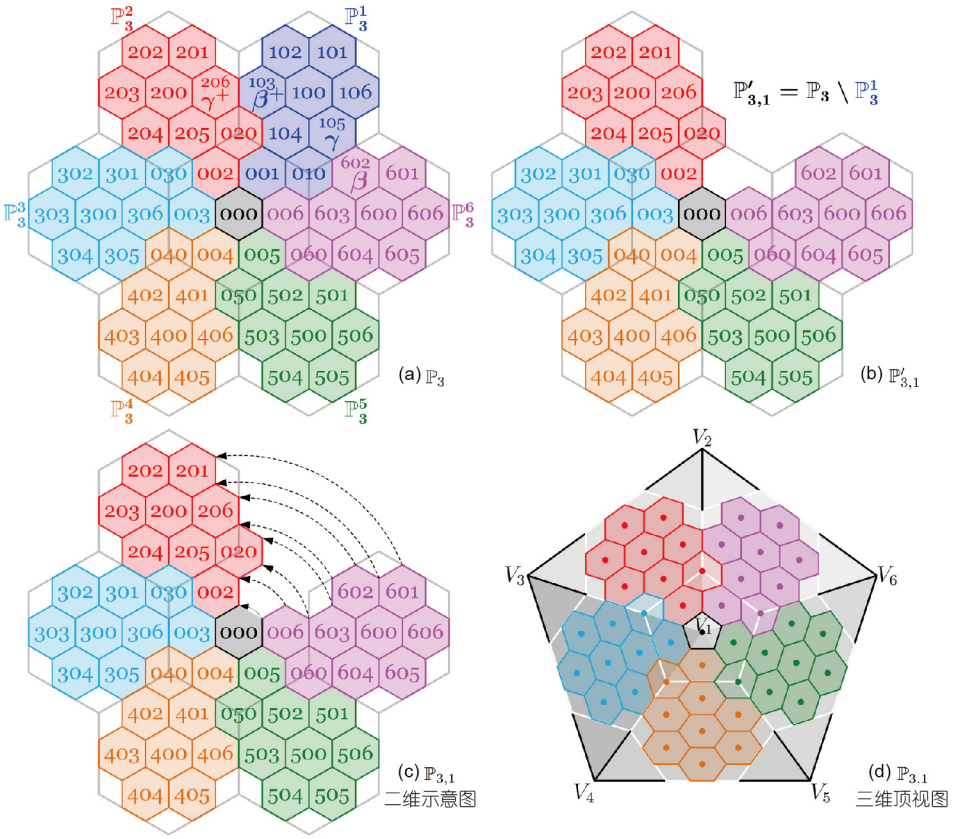


图 12　$\mathbb{P}_3$、$\mathbb{P}'_{3,1}$、$\mathbb{P}_{3,1}$示意图

表 2　二十面体核心类定义

// 正二十面体类	// 标准结构类

class DgIcosahedron
{
public:
// 构造函数
DgIcosahedron(double L, GeoCoord pt, double azh);
DgIcosahedron(GeoCoord pt, double azh);
.....
protected:
double L_;// 边长
DgDVec3D vtx_[12];// 12个顶点
DgStdUnit su_[12];// 12个标准结构
DgTriFacet tri_[20];// 20个三角面
.....
};
	class DgStdUnit
	{
	public:
	// 构造函数
	DgStdUnit(unsigned short cid, unsigned short* vid, DgType t);
	DgStdUnit( );
	.....
	protected:
	// 中心顶点的编号(索引号)
	unsigned short cid_;
	// 四周顶点编号(索引号)
	unsigned short vid_[5];
	// 标准结构的类型
	unsigned short type_;
	.....
	};

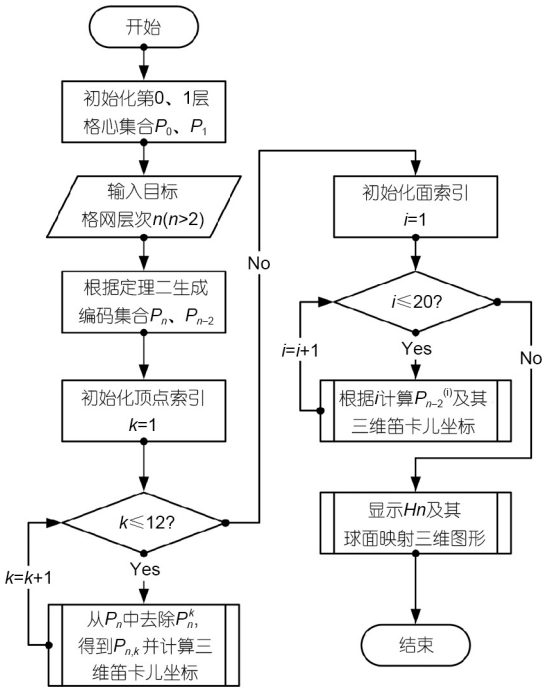


图 13　格网构造算法流程图

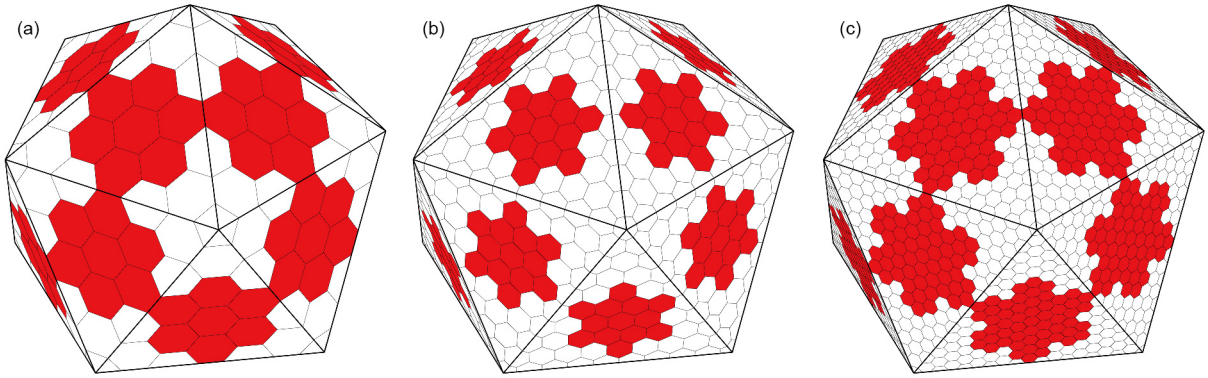


图 14　二十面体上不同层次的六边形格网

(a) n=3; (b) n=4; (c) n=5

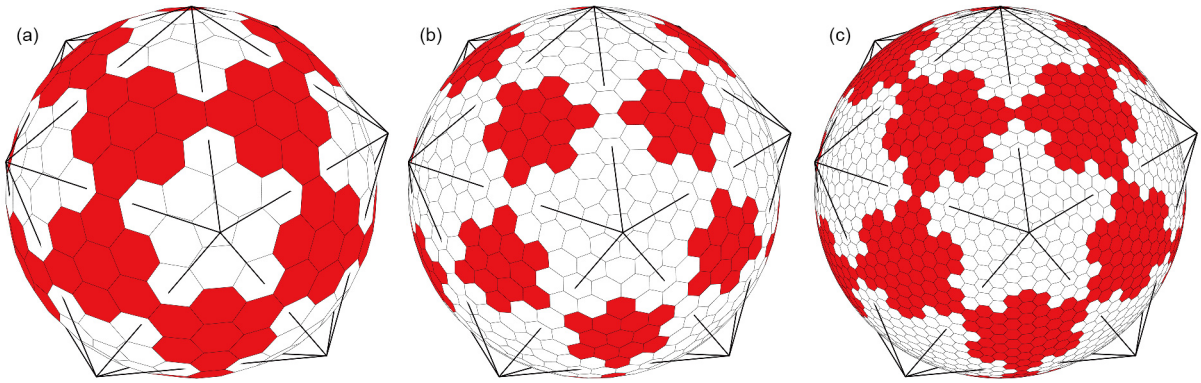


图 15　球面上不同层次的六边形格网

(a) n=3; (b) n=4; (c) n=5

运算中的优势, 作者参照Vince和Zheng(2009)、Tong等(2013)的研究思路, 设计了正二十面体三角面边界处单元的邻近搜索算法, 即通过编码加法运算查找图16中诸如h_0、h_1、h_2、h_3等边界单元的6(5)个虚线标示的跨面相邻单元. 统计该算法在单位时间内查找邻近单元的平均个数(称为"标准化效率"), 并与采用HQBS(Tong等, 2013)方案完成相同操作的标准化效率对比, 消除不同孔径的影响, 保证结果客观、公平. 全部程序均编译为Release版本, 并在一台ThinkPad W520笔记本(硬件配置: Intel Core i7-2720QM CPU@2.20GHz四核心, 16G RAM, SAMSUNG mSATA 128G SSD; 操作系统: Windows 7 x64旗舰版 SP1; 开发工具: Microsoft Visual C++ 2012 Update 5)上运行, 实验结果如表3所示, 效率对比曲线如图17所示.

上述实验结果表明, 本文提出的代数编码方法不仅具有严密的理论基础, 而且有利于高效编程实现. 在测试环境中执行较复杂的单元编码跨面邻近搜索算法, 10层以上的效率稳定在2万单元/ms左右(图17a), 平均可达HQBS的242.9倍(图17b). 这主要得益于算法采用了标准结构, 原本通过浮点数运算(效率低)才能完成的单元跨面旋转、平移等几何操作完全借助编码运算(效率高)实现. HQBS虽然在二十面体的单个三角面上采用编码运算实现几何操作, 但是在跨面运算时仍涉及少量坐标旋转、平移的浮点运算, 且编码运算中会出现"正则化"失败回退重算的情况, 因而效率较低.

7　总结与展望

本文在"格元-格点-格边-格心"概念模型的基础上, 采用格心作为描述格元的等效对象, 揭示了平面三孔六边形多分辨率格网的数学本质, 严格证明了格元表达的唯一性条件并以此为依据讨论了格元编码

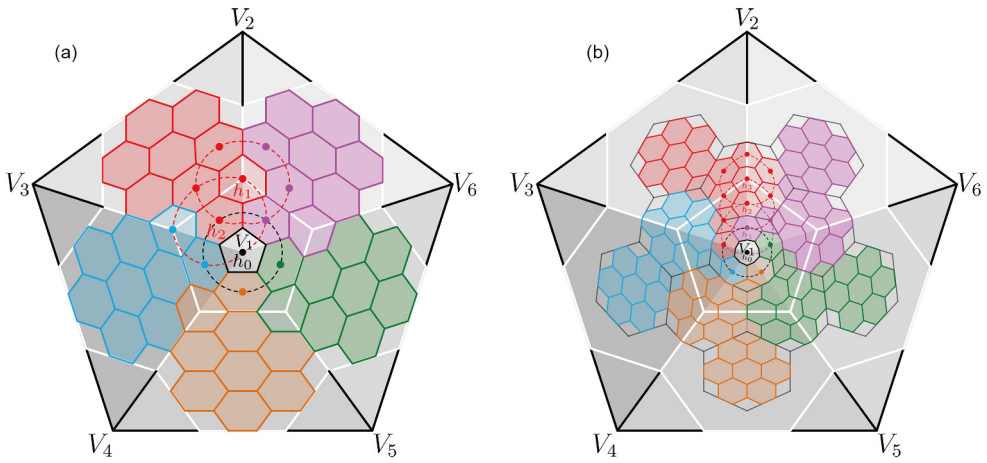


图 16　编码跨面邻近搜索算法示意图

(a) $\mathbb{P}_{3,1}$; (b) $\mathbb{P}_{4,1}$

表 3　编码跨面邻近搜索对比实验结果

层次	HQBS		本文采用方案		效率比
	单元个数	运算效率(单位/ms)	单元个数	运算效率(单位/ms)	
5	42	80.2	27	74713.5	931.5
6	84	162.9	81	59384.6	364.6
7	170	292.6	81	48524.4	165.8
8	340	242.1	243	47109.3	194.6
9	682	238.8	243	42704.1	178.8
10	1364	213.4	729	33704.0	158.0
11	2730	219.3	729	27689.0	126.2
12	5460	209.0	2187	22955.5	109.9
13	10922	200.8	2187	18606.1	92.7
14	21844	192.4	6561	20642.3	107.3

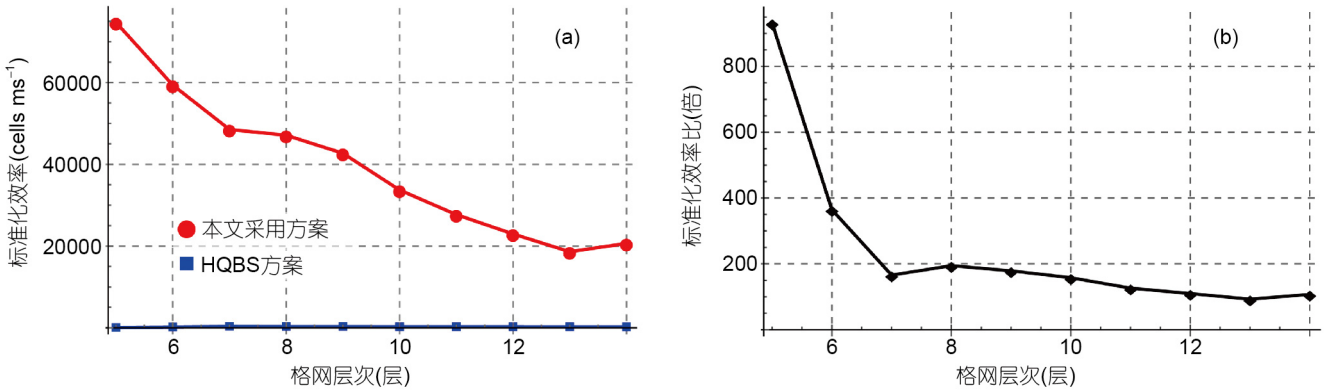


图 17　编码跨面邻近搜索效率对比

(a) 绝对值; (b) 比值

方案. 通过对顶点结构特点的分析, 将平面格网扩展到正二十面体封闭表面, 从而建立了三孔六边形全球六边形离散格网系统的代数编码方案.

本文提出的方法是对Shar(2005)、Vince(2006b)、Vince和Zheng(2009)方案的完善和发展. Vince等学者仅给出了三孔六边形格网系统编码方案及相关结论, 本文则从格网系统的几何属性出发, 通过严格的数学证明揭示其规律, 因此在理论上是完备的. 与Sahr通过实验归纳设计的A3HT(Aperture 3 Hexagon Tree)结构相比, 本文的描述不仅在形式上更简洁, 而且从理论上证明了编码的性质, 为高效编码算法设计提供了根本保证, 这在对比实验中已得到验证.

尽管本文只讨论了一种六边形格网系统, 但相关定理的证明过程已表明多分辨规则格网是有限格心集合的自我复制, 本质上是特定形式的复进制数. 这一结论为格网系统结构特征的抽象提供了强有力的支持: 复进制数是二维复数平面上的点, 准确记录单元几何位置, 有利于几何特征的抽象; 复进制数又是复数域中的数, 天然支持单元之间的运算, 有利于代数特征的抽象; 复进制数还是特殊形式的级数, 其中每一项记录了到达当前单元经过的父单元, 有利于编码特征的抽象. 因此, 本文采用的复进制数是“全息”的格网形式化描述工具, 本文的研究思路亦可方便地拓展到其他规则格网系统上, 创建具有普适性的全球离散格网系统理论框架, 为算法的设计、分析和优化提供理论支撑.

致谢 衷心感谢匿名评审专家对本文提出的真知灼见; 衷心感谢信息工程大学地理空间信息学院童晓冲副教授慷慨提供的HQBS C++源代码及测试程序.

参考文献

白建军. 2011. 基于正八面体的四孔六边形球面格网编码及索引. 遥感学报, 15: 1125–1137

贲进, 童晓冲, 元朝鹏. 2011. 孔径为4的全球六边形格网系统索引方法. 测绘学报, 40: 785–789

贲进, 童晓冲, 张永生, 陈荣国. 2007. 球面等积网格系统生成算法与软件模型研究. 测绘学报, 36: 187–191

贲进, 童晓冲, 张永生, 张衡. 2006. 对施耐德等积多面体投影的研究. 武汉大学学报: 信息科学版, 31: 900–903

吕雪锋, 程承旗, 龚健雅, 关丽. 2011. 海量遥感数据存储管理技术综述. 中国科学: 技术科学, 41: 1561–1573

Knuth D. 2002. 苏运霖, 译. 计算机程序设计艺术, 第2卷: 半数值算法(第3版). 北京: 国防工业出版社. 175–189

谢文军. 2010. 全球多分辨率半蜂窝梯形网格构建及其可视化研究. 硕士学位论文. 合肥: 合肥工业大学. 9–22

张永生, 贲进, 童晓冲. 2007. 地球空间信息球面离散格网——理论、算法及应用. 北京: 科学出版社. 76

赵学胜, 王磊, 王洪彬, 李颖. 2012. 全球离散格网的建模方法及基本问题. 地理与地理信息科学, 28: 29–34

周成虎, 欧阳, 马廷. 2009. 地理格网模型研究进展. 地理科学进展, 28: 657–662

Amiri A, Harrison E, Samavati F. 2015a. Hexagonal connectivity maps for digital Earth. Int J Digital Earth, 8: 750–769

Amiri A, Samavati F, Peterson P. 2015b. Categorization and conversions for indexing methods of discrete global grid systems. ISPRS Int J Geo-Inform, 4: 320–336

Birch C P D, Oom S P, Beecham J A. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecol Model, 206: 347–359

Brockmann C, Ruescas A, Stelzer K. 2013. Using BEAM with SMOS data. Training Presentation. 3rd ESA Advanced Training on Ocean Remote Sensing. Technical Report. National Maritime College of Ireland

Dutton G. 1984. Part 4: Mathematical, algorithmic and data structure issues: Geodesic modelling of planetary relief. Cartogr-Int J Geogr Inf Geovisual, 21: 188–207

Fekete G. 1990. Rendering and managing spherical data with sphere quadtree. In: Proceedings of the 1st Conference on Visualization ’90. California: San Francisco. 176–186

Gibson L, Lucas D. 1982. Spatial data processing using generalized balanced ternary. In: Processings of IEEE Computer Society Conference on Pattern Recognition and Image Processing. Las Vegas. 566–571

Goodchild M F, Yang S R. 1992. A hierarchical spatial data structure for global geographic information systems. CVGIP-Graphical Model Image Processing, 54: 31–44

Kenner H. 2003. Geodesic Math and How to Use It. California: California Press. 32–35

Kimerling J A, Sahr K, White D, Song L. 1999. Comparing geometrical properties of global grids. Cartogr Geogr Inf Sci, 26: 271–288

Peterson P. 2006. Close-packed uniformly adjacent, multi-resolution overlapping spatial data ordering. US Patent, 2006/0265197 A1

Sahr K. 2005. Discrete global grid systems: A new class of geospatial data structures. Doctoral Dissertation. Ashland: University of Oregon. 71–102

Sahr K, White D, Kimerling A J. 2003. Geodesic discrete global grid systems. Cartogr Geogr Inf Sci, 30: 121–134

Snyder J P. 1992. An equal-area map projection for polyhedral globes.

Cartogr-Int J Geogr Inf Geovisual, 29: 10–21

Tong X, Ben J, Wang Y, Zhang Y, Pei T. 2013. Efficient encoding and spatial operation scheme for aperture 4 hexagonal discrete global grid system. Int J Geogr Inf Sci, 27: 898–921

Vince A. 2006a. Indexing the aperture 3 hexagonal discrete global grid. J Visual Commun Image Represent, 17: 1227–1236

Vince A. 2006b. Indexing a discrete global grid. Special Topics in Computing and ICT Research, Advances in Systems Modelling and ICT Applications, Volume II. Kampala: Fountain Publishers. 3–17

Vince A, Zheng X. 2009. Arithmetic and Fourier transform for the PYXIS multi-resolution digital Earth model. Int J Digital Earth, 2: 59–79

White D, Kimerling A J, Sahr K, Song L. 1998. Comparing area and shape distortion on polyhedral-based recursive partitions of the sphere. Int J Geogr Inf Sci, 12: 805–827

Zheng X. 2007. Efficient fourier transforms on hexagonal arrays. Doctoral Dissertation. Florida: University of Florida. 62–99