

СТРУКТУРЫ ПРОСТРАНСТВЕННЫХ ДАННЫХ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ

Шурмилов П. К.

Кафедра систем управления, Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: pkeycalculator@gmail.com

В данной статье автором рассматриваются основные структуры пространственных данных, характеризующих пространственные объекты в геоинформационных системах.

ВВЕДЕНИЕ

Объектом информационного моделирования в геоинформационных системах является пространственный объект. Некоторое множество цифровых данных о пространственных объектах образует пространственные данные. В практике геоинформатики определен набор базовых моделей пространственных данных: растровая, регулярно-ячеистая, квадродерево, векторная [1].

I. ХРАНЕНИЕ РАСТРОВЫХ ДАННЫХ

Растры применяются для хранения и обработки данных дистанционного зондирования, для представления цифровых моделей рельефа, при визуализации геоданлогоритмы. Существует множество вариантов кодирования растровых структур. Некоторые из них более экономно расходуют память, другие позволяют получать более быстрые алгоритмы. Растровая модель соответствует двумерному ячеистому изображению, которое хранится в памяти компьютера в виде одномерной последовательности значений. Растровые изображения обычно разлагаются по строке сверху – слева. Далее будут описаны другие способы эффективного представления растров [2]. В некоторых форматах графических файлов используется сжатие изображения, основанное на замене длительных последовательностей повторяющихся значений парой <значение, количество повторов> (см. рис. 1-а). Географические данные обычно автокоррелированы. В растровой модели это означает, что соседние ячейки имеют большую вероятность быть одинаковыми, чем разобценные. При обычном порядке сканирования в конце каждой строки происходит скачок на начало следующей строки. Предложим простое изменение порядка сканирования. Нечетные строки будем кодировать слева направо, а четные – в обратном направлении (см. рис. 1-б). Направление сканирования напоминает движение быка, вспахивающего поле. Отсюда название этого способа сканирования – бустрофедон (греч. – бык, вспахивающий поле). Теперь при переходе к новой строке первая ячейка является смежной последней ячейке старой строки. Так в линейном разложении растра

сохраняется автокорреляция и повышается эффективность кодирования [3].

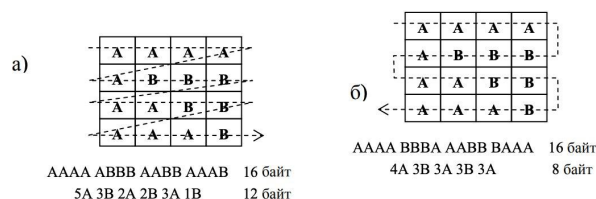


Рис. 1 – Порядки сканирования растров, их линейное разложение и сжатие: а) обычный порядок; б) бустрофедон

Порядок сканирования Мортон основан на иерархическом разбиении карты. В предыдущих способах сканирования учитывалась автокорреляция значений ячеек только по одному направлению (по строке). Географические объекты образуют на растровом изображении пятна. В порядке Мортон предпринимается попытка сканирования ячеек таким образом, чтобы охватить линией обхода эти двумерные пятна.

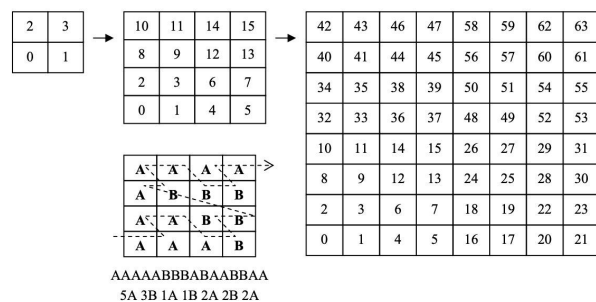


Рис. 2 – Порядок сканирования растра по Мортону

II. ИЕРАРХИЧЕСКИЕ СТРУКТУРЫ ДАННЫХ

Рассмотренные выше порядки сканирования растровых изображений дают незначительные различия в компрессии данных. Основное преимущество Мортон-сканирования и других иерархических структур данных заключается в более быстром доступе к данным. Информация распределена по карте неравномерно. Увеличение разрешения растрового изображения приводит к увеличению размеров файлов, а уменьшение – к потере информации. Далее пойдет речь

об адаптивных методах представления растровых данных с разной плотностью информации. На рис. 4 изображена растровая матрица размера 16 x 16, в которой содержатся 255 значений "А" и одно "В". Индексируем растр следующим способом. Разделим матрицу на четыре подматрицы размера 8 x 8 и нумеруем их 0, 1, 2, 3 в порядке Мортон. Назовем подматрицу гомогенной, если в ней содержатся одинаковые значения. Будем рекурсивно разбивать негомогенные подматрицы до тех пор, пока не достигнем гомогенности всех подматриц. Таким способом получим адаптивное разрешение растрового изображения, где участки с меньшей плотностью информации представлены крупными блоками ячеек, а с большей плотностью – мелкими блоками ячеек. Идея выделения гомогенных блоков растра тождественна кодированию растра по Мортону. Гомогенный блок растра размера $m \times m$ при сканировании по Мортону соответствует коду $\langle m_2, A \rangle$

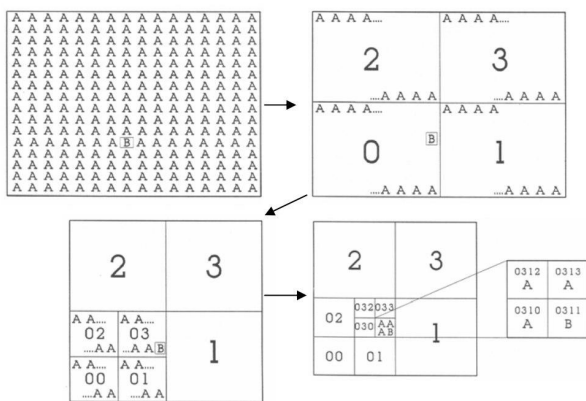


Рис. 3 – Разбиение растра на гомогенные блоки

Соответствие двумерных растровых координат ячейки и адреса ячейки в последовательном файле похоже на аналогичное преобразование при кодировании растра по Мортону. Единственное отличие в том, что используется система счисления с основанием четыре. В примере на рис. 4 ячейка "В" имеет код 0311. В двоичной системе счисления $N = 0311 = (00110101)_2$. Разделим биты между растровыми координатами и выясним, что ячейка лежит в четвертой строке и седьмом столбце. Представленные таким способом растровые данные соответствуют квадродереву, вершина которого – исходное изображение, а листья – гомогенные блоки ячеек. При кодировании квадродеревьев ячейки на каждом уровне могут содержать либо значение гомогенного блока, либо указатель на следующий уровень. Дерево, показанное на рис. 5 может быть представлено в виде линейной последовательности следующим образом.

Позиция:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Содержание:	2	6	А	А	А	А	А	А	10	А	14	А	А	А	В	А	А
Уровень:	0	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4

Рис. 4 – Кодирование квадродерева

Как уже отмечалось выше, основное преимущество иерархической организации данных в ГИС заключается в пространственном упорядочении информации и более быстром ее поиске. Поэтому рассматриваются две задачи ГИС, связанные с индексированием квадродеревьями: первая – поиск всех частей карты с заданным значением и вторая – определение содержимого некоторой ячейки [4].

Обозначим n – число уровней квадродерева (тогда размер растра $2n \times 2n$) и через m – число листьев в дереве. Чтобы найти части карты с некоторым значением B , необходимо проверить каждый лист дерева, что потребует m шагов. Определение значения ячейки происходит путем спуска по квадродереву до тех пор, пока не будет получен гомогенный блок. В худшем случае, когда ячейка находится на самой вершине дерева (как, например, ячейка B на рис. 4), поиск займет n шагов. Сравним теперь трудоемкости обеих задач на квадродереве с трудоемкостями этих задач при различных вариантах сканирования растра.

Таблица 1 – Трудоемкость алгоритмов при различной организации растров

Структура данных	Поиск частей с заданным значением	Определение значения ячейки
Квадродерево	m	n
Обычный порядок	$4n^{2*}$	1^{**}
Бустрофедон	m^{***}	m^{***}
Мортон	m^{***}	m^{***}

Прим. * – проверяется каждая ячейка матрицы; ** – непосредственное вычисление позиции ячейки; *** – число цепочек примерно соответствует числу листьев; **** – проверяется каждая цепочка.

III. ЗАКЛЮЧЕНИЕ

Были рассмотрены и проанализированы некоторые способы кодирования пространственных данных в геоинформационных системах, а именно: обычный порядок, бустрофедон, кодирование Мортон и квадродерево. Оценена трудоемкость задачи поиска всех частей карты с заданным значением и задачи определения содержимого некоторой ячейки при использовании каждого из вышеуказанных способов.

1. Варфоломеев И.В., Савельев А.С. Представление и обработка пространственных данных в ГИС: Методические указания / Варфоломеев И.В., Савельев А. – Красноярск: КГТУ, 2001.
2. Коновалова Н.В., Капралов Е.Г. Введение в ГИС / Коновалова Н.В., Капралов Е. – Петрозаводск: Издво Петрозаводского Госуниверситета, 1995.
3. Кошкарёв А.В., Тикунов В.С. Геоинформатика / Под ред. Д.В. Лисицкого. М. – Картгеоцентр – Геоиздат, 1993.
4. Якубайлик О.Э. Методы и приемы пространственного анализа в геоинформационных системах: Учебное пособие / Якубайлик О. – Красноярск: Издательство КрасГУ, 2001.