

УДК 514.86

АНАЛИЗ ГЕОМЕТРИЧЕСКИХ ХАРАКТЕРИСТИК ТОЧЕЧНЫХ МНОЖЕСТВ НА ОСНОВЕ АЛГОРИТМА ПРОСТРАНСТВЕННОГО ХЕШИРОВАНИЯ

Дашкевич А.А., к.т.н.,

Шоман О.В., д.т.н.

Национальный технический университет

«Харьковский политехнический институт» (Украина)

В работе предложен подход к решению задач определения и анализа геометрических характеристик точечных множеств. Разработан алгоритм пространственного хеширования. Получена возможность анализировать геометрическую структуру множества на основе пространственных индексов ближайшего соседства на многомерных сетках. Результаты направлены на оптимизацию задачи компьютерного зрения по сегментации точечных множеств на отдельные элементы.

Ключевые слова: точечное множество, пространственное хеширование в многомерных пространствах, геометрическая структура множества, индекс ближайшего соседства, сегментация множества.

Постановка проблемы. В задачах компьютерного зрения данные с сенсоров камеры часто представлены в виде точечных множеств: матрицы яркости пикселей при обработке изображений, облака точек в задачах стереозрения. При этом возникает необходимость сегментации таких точечных множеств, т.е. их разбиения на независимые компоненты. Задача сегментации предполагает решение двух связанных подзадач:

- нахождение точек, которые расположены в некоторой окрестности заданной точки;
- нахождение геометрических характеристик точечного множества в локальных окрестностях.

На данный момент точное решение для задачи поиска ближайших соседей дают переборные алгоритмы поиска, которые отличаются квадратичной вычислительной сложностью. Выделение геометрических признаков в локальной окрестности основано на анализе пространственного распределения точек и статистических показателях ближайшего соседства. Таким образом, качество выделения геометрических признаков тесно связано с эффективностью определения ближайших соседей точек. Отсюда

вытекает необходимость разработать алгоритмы определения геометрических характеристик на основе пространственного анализа с достаточной точностью и вычислительной сложностью в пространствах произвольной размерности.

Анализ последних исследований и публикаций. Для оптимизации поиска ближайших соседей используются специальные структуры данных, например kd-деревья [1, 2], BSP-деревья [3, 4], R-деревья [5] и др., а также алгоритмы по обработке таких структур [6]. Однако эффективность подобных структур данных и алгоритмов падает при росте числа измерений N и становится сравнимой с линейным поиском [7] при некотором критическом значении мерности пространства поиска $N > N_{кр}$. В настоящее время развиваются алгоритмы приближенного поиска ближайших соседей в многомерных пространствах, среди которых можно выделить те, что основаны на использовании хеширующих функций, однозначно указывающих на область пространства, которой принадлежит точка.

В работе [8] представлен метод пространственного хеширования, в котором предлагается заменить координаты x , y и z точек множества на одну координату – хеш и вычислять ее по следующей схеме:

$$h = x/c \cdot 2^k + y/c \cdot 2^m + z/c \cdot 2^n,$$

где c – размер сетки, k , m и n – некоторые константы ($k > m > n$ или $k < m < n$).

В работе [9] предложен следующий подход для вычисления хеша для точки:

$$h = ((x \cdot p_1) \text{ XOR } (y \cdot p_2) \text{ XOR } (z \cdot p_3)) \text{ MOD } n,$$

где $p_1 = 73856093$, $p_2 = 19349663$, $p_3 = 83492791$ – большие простые числа, XOR – побитовая операция «Исключающее ИЛИ», MOD n – операция нахождения остатка от деления на n , n – общее количество точек.

Формулирование целей статьи. Проведение пространственного анализа N -мерных точечных множеств на основе алгоритма упрощенного пространственного хеширования.

Основная часть. В работе [10] предложен алгоритм приближенного поиска ближайших соседей на двумерной сетке для применения в решении задач молекулярной динамики. Одним из преимуществ данного алгоритма является отсутствие коллизий хешей.

Предлагается расширение алгоритма, приведенного в работе [10], для пространственного хеширования в пространствах с $N > 2$:

- 1) пространство разбивается на сетку размером c ;
- 2) для каждой точки с координатами $(x_1, \dots, x_i, \dots, x_N)$, заданной в диапазоне $[0, G_i]$, вычисляются индексы клетки, в которой находится данная точка:

$$i_1 = \frac{x_1}{c}, \dots, i_N = \frac{x_N}{c},$$

где N – размерность пространства;

3) по заданным индексам вычисляются значения хеша:

$$h_k = \frac{i_i \cdot 10^{(N-i)d}}{N},$$

где d – максимальное количество разрядов хеша, которое приходится на одну пространственную координату x_i ,

$$i_i = \max\left(\frac{G_i}{c}\right).$$

Был разработан алгоритм разбиения полученного хеша на отдельные индексы по каждой из координатных осей:

$t = N$
 до тех пор, пока $t > 0$, повторять:
 $i_t = h \bmod 10^d$
 $h = h / 10^d$
 $t = t - 1$

Данный алгоритм разбивает пространство на многомерную регулярную сетку и позволяет определять количество элементов множества в ячейках сетки. Это дает возможность проводить вычисление пространственных геометрических характеристик множества.

В работе проведено исследование геометрических характеристик точечного множества на основе предложенного алгоритма. На рис. 1 представлено исходное множество и соответствующее хеш-разбиение на ячейки размером $c = 8$ для $N = 3$.

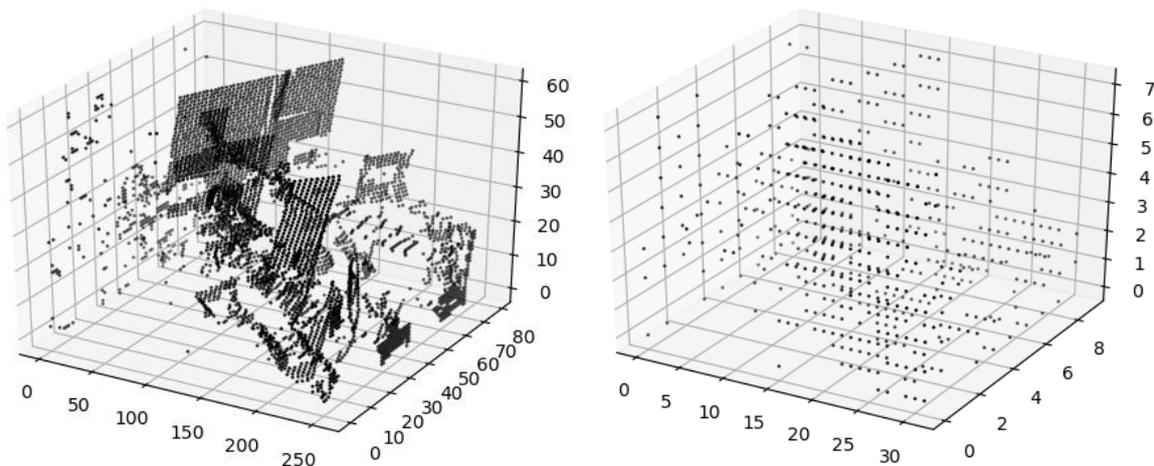


Рис. 1. Исходное множество точек и его хеш

В качестве меры пространственного распределения предлагается использовать плотность точек в ячейке. На рис. 2 представлена визуализация пространственной плотности в ячейках, что позволяет проводить пространственный анализ множества.

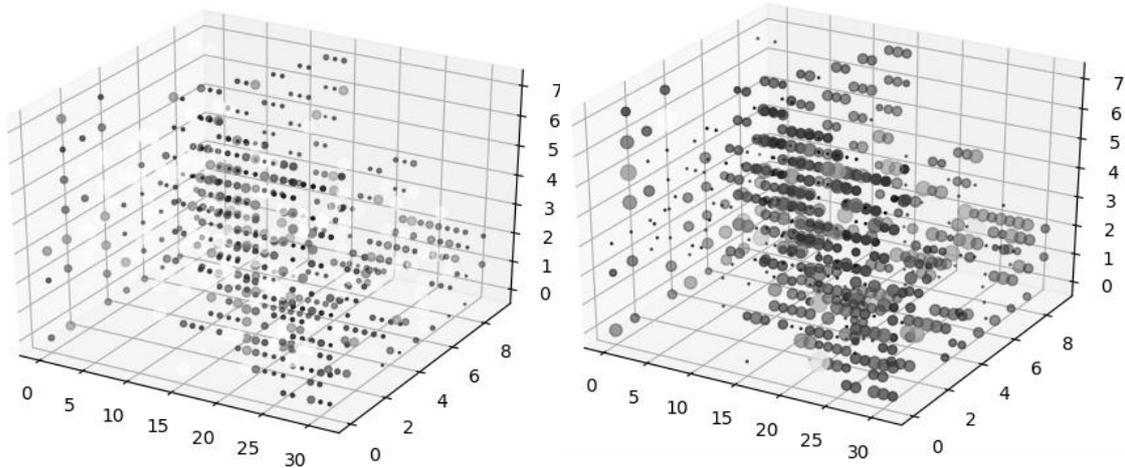


Рис. 2. Визуализация пространственной структуры ячеек на основе распределения плотностей множества

Выводы. Предложенный метод дает возможность выявления структурных особенностей точечных множеств. Предложенный алгоритм позволяет выявлять такие характеристики множества, как пространственное распределение его элементов на сетке, что позволяет в дальнейшем проводить сегментацию множества на составляющие его элементы.

Литература

1. Bentley J.L. Multidimensional Divide and Conquer / J.L. Bentley // Communications of the ACM. – 1980. – Vol. 23. – Is. 4. – P. 214–229.
2. Friedman J.H. An algorithm for finding best matches in logarithmic expected time / J.H. Friedman, J.L. Bentley, R.A. Finkel // ACM Transactions on Mathematical Software. – 1977. – Vol. 3. – Is. 3. – P. 209–226.
3. de Berg M. Computational Geometry: Algorithms and Applications / M. de Berg // Springer Science & Business Media, 2008. – P. 259.
4. Castelli V. Image Databases: Search and Retrieval of Digital Imagery / V. Castelli, D. Lawrence // John Wiley & Sons, 2004. – P. 422.
5. Guttman A. R.–Trees: A Dynamic Index Structure for Spatial Searching / A. Guttman // Proc. ACM SIGMOD International Conference on Management of Data. – 1984. – P. 47–57.
6. Liu T. An Investigation Of Practical Approximate Nearest Neighbor Algorithms / T. Liu, A. Moore, A. Gray, K. Yang // Neural Information Processing Systems. – 2004.
7. Beyer K. When Is «NearestNeighbor» Meaningful? / K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft // InInternational Conference on Database Theory. –1999. – P. 217–235.
8. Nguyen H. (ed.). GPU Gems 3 / H. Nguyen // Addison Wesley, 2007. – P. 1008.

9. Ize T. Asynchronous BVH Construction for Ray Tracing Dynamic Scenes on Parallel Multi-Core Architectures / T. Ize, I. Wald, S. G. Parker // Proceedings of the 2007 Eurographics Symposium on Parallel Graphics and Visualization. – 2007.
10. Ніцин О.Ю. Спрощений алгоритм просторового хешування для задач молекулярної динаміки / О.Ю. Ніцин, А.О. Дашкевич, О.В. Охотська, О.Є. Мацулевич // Науковий вісник Таврійського державного агротехнологічного університету. – Мелітополь: ТДАТУ, 2016. – Вип. 6, Т. 1. – С. 287–291.

АНАЛІЗ ГЕОМЕТРИЧНИХ ХАРАКТЕРИСТИК ТОЧКОВИХ МНОЖИН НА ОСНОВІ АЛГОРИТМУ ПРОСТОРОВОГО ХЕШУВАННЯ

Дашкевич А.О., Шоман О.В.

В роботі запропоновано підхід до розв'язання задач визначення і аналізу геометричних характеристик точкових множин. Розроблено алгоритм просторового хешування. Одержано можливість аналізувати геометричну структуру множини на основі просторових індексів найближчого сусідства на багатовимірних сітках. Результати спрямовано на оптимізацію задачі комп'ютерного зору з сегментації точкових множин на окремі елементи.

Ключові слова: точкова множина, просторове хешування, геометрична структура множини, індекс найближчого сусідства, сегментація множини.

ANALYSIS OF GEOMETRIC CHARACTERISTICS OF POINT SETS BASED ON SPATIAL HASHING ALGORITHM

Dashkevich A., Shoman O.

An approach to solving problems of determining and analyzing the geometric characteristics of point sets is proposed. An algorithm for spatial hashing has been developed. It is possible to analyze the geometric structure of a set on the basis of spatial proximity indices on multidimensional grids. The results are aimed at optimizing the task of computer vision for segmenting point sets into individual elements.

Key words: point set, spatial hashing, set geometric structure, proximity indices, set segmentation.