



УДК 514.86

АЛГОРИТМ ПРОСТРАНСТВЕННОГО ХЕШИРОВАНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧ ПРИБЛИЗИТЕЛЬНОГО ПОИСКА БЛИЖАЙШИХ СОСЕДЕЙ

Дашкевич А.А., к.т.н.,

Национальный технический университет «Харьковский политехнический институт»

Тел. (057) 707-64-31

Аннотация – предложена реализация алгоритма и структур данных для пространственного хеширования в многомерных пространствах. Разработанный алгоритм позволяет проводить приблизительный поиск ближайших соседей на многомерных сетках, что позволяет проводить оптимизацию таких задач геометрического моделирования, как аппроксимация и выделение поверхностей из точечных множеств, сегментация поверхностей.

Ключевые слова – приблизительный поиск ближайших соседей, пространственное хеширование, аппроксимация, выделение поверхностей, точечное множество, сегментация поверхности.

Постановка проблемы. Во многих задачах вычислительной геометрии, компьютерной графики и геометрического моделирования возникает необходимость нахождения объектов, которые находятся в некоторой окрестности от заданного. Для этого используются различные методики. Например: триангуляция полигональных моделей, обнаружение столкновений объектов, аппроксимация и выделение поверхностей из необработанных пространственных данных, заданных в виде точечных множеств.



На данный момент точное решение для такой задачи дают алгоритмы линейного поиска, которые являются вычислительно затратными по времени выполнения. В связи с этим существует необходимость разработки алгоритмов поиска приближенных ближайших соседей с достаточной точностью и вычислительной сложностью в пространствах большой размерности.

Анализ последних исследований. Для оптимизации поискового перебора 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 методы, основанные на использовании хеширующих функций, которые однозначно указывают на область пространства, которой принадлежит точка. Одной из проблем при разработке хеширующих функций является возможность возникновения коллизий хешей – получение одинаковых значений хеша для разных областей пространства, поэтому необходимы методы вычисления хешей без коллизий.

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

$$h = \frac{x}{C} \cdot 2^k + \frac{y}{C} \cdot 2^m + \frac{z}{C} \cdot 2^n,$$

где C – размер сетки;

k , m и n – некоторые константы, такие что, $k > m > n$
(или $k < m < n$).



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

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

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.

В работе [10] предложен алгоритм приблизительного поиска ближайших соседей на двумерной сетке для применения в решении задач молекулярной динамики. Предложен следующий алгоритм вычисления пространственного хеша:

1) пространство разбивается на сетку размером $C \times C$;

2) для координат всех точек (x, y) вычисляются индексы (j – номер колонки, i – номер ряда в сетке) клетки, в которой находится данная точка:

$$j = \frac{x}{C}, \quad i = \frac{y}{C}$$

3) по заданным индексам вычисляются значения хеша:

$$h = j \cdot 10^n + i \cdot 10^m,$$

где n и m – небольшие целые числа, значения которых зависят от количества разрядов в j и i , соответственно (в общем случае $n = m + 1$, $m = 0$).

Одним из преимуществ данного алгоритма является отсутствие коллизий хешей. Однако, данный алгоритм может быть использован



для пространств с $N = 2$, что приводит к необходимости его расширения на пространства большей размерности.

Формулирование целей статьи. Разработка алгоритма поиска приблизительных ближайших соседей методом упрощённого пространственного хеширования в N -мерных пространствах.

Основная часть. В работе предлагается следующее расширение алгоритма, предложенного в работе [10] для пространственного хеширования в пространствах с $N > 2$:

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

$$i_i = \frac{x_1}{C} \dots i_N = \frac{x_N}{C}$$

где N – размерность пространства;

- 3) по заданным индексам вычисляются значения хеша:

$$h_k = i_i \cdot 10^{P_i} / N$$

где $P_i = 10^{(N-i) \cdot d}$,

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

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

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

$$t = N$$

до тех пор, пока $t > 0$, повторять:

$$i_i = h \bmod 10^d$$

$$h = h / 10^d$$

$$t = t - 1$$



Для проверки данные алгоритмы были реализованы средствами языка Python. В качестве тестовых данных были использованы данные карты глубины, полученной методом стереозрения (рис. 1). На рис. 2 представлена визуализация полученного множества хешей, которые соответствуют пространственному разбиению исходного множества на ячейки размером $C = 8$ для $N = 3$. Также были разработаны дополнительные алгоритмы для проведения дальнейших исследований, в частности алгоритм определения ближайших точек в окрестности данной точки, алгоритм фильтрации исходного пространства по критерию количества точек в пределах одной ячейки, что позволит, например проводить сегментацию точечных множеств на отдельные, независимые области.

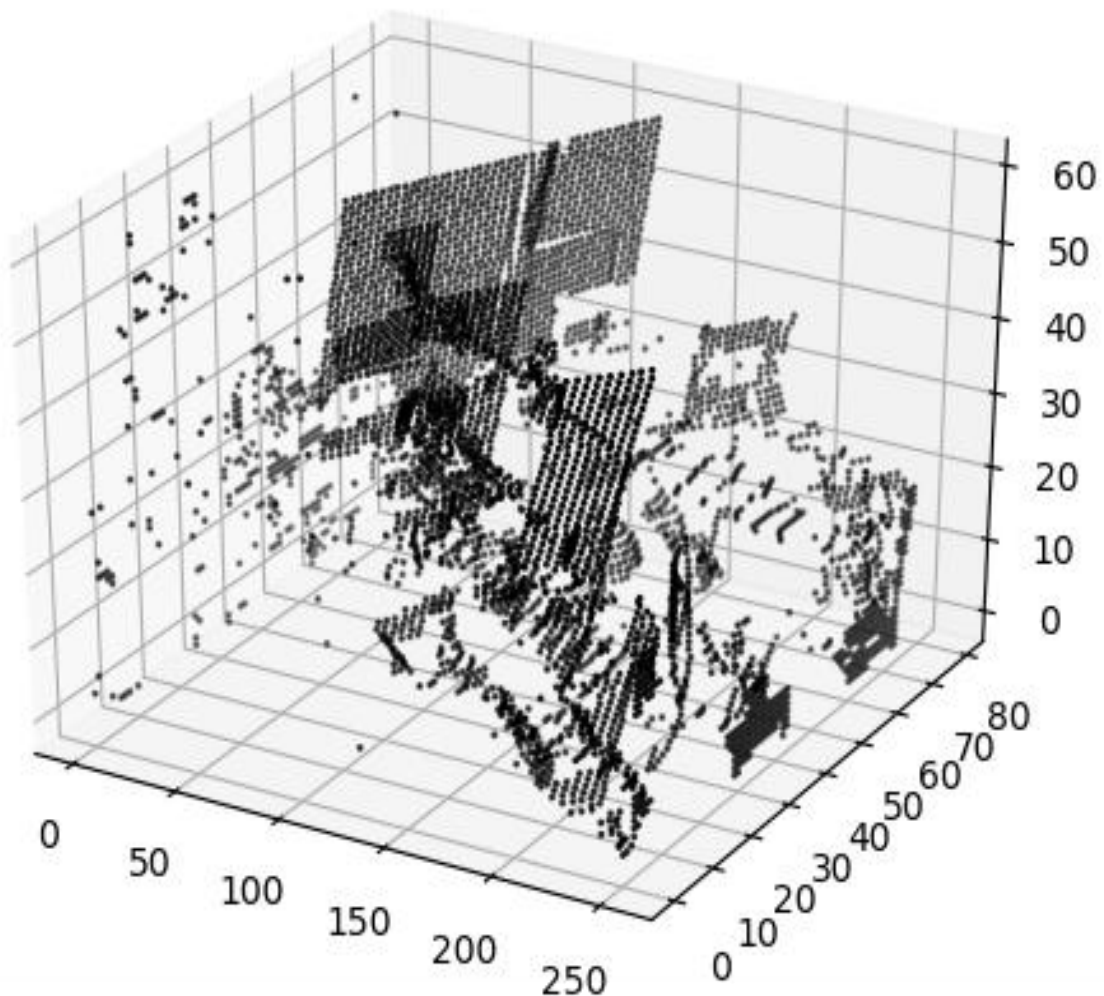


Рис. 1. Исходное множество точек

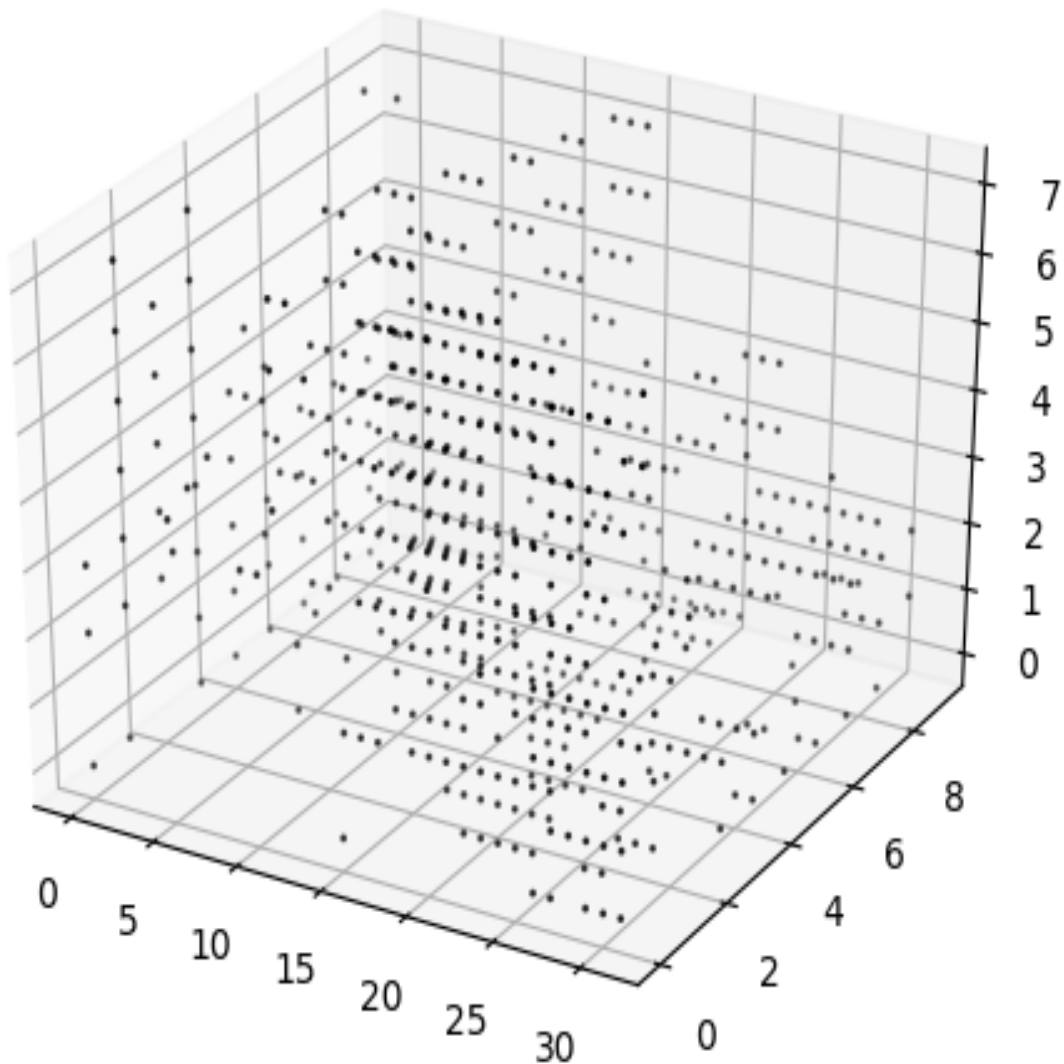


Рис. 2. Визуализация пространственного хеша для множества

Для реализации быстрого поиска точек в пределах одной ячейки разработана структура данных, которая представляет собой хеш-таблицу, ключами в которой выступают значения хешей, а значениями — номера точек исходного множества, которые находятся в пределах заданной ячейки хеша.

Выводы и перспективы дальнейших исследований. Разработанный метод позволяет быстрое вычисление приблизительных ближайших соседей для точек в N -мерном пространстве. Алгоритм требует линейного времени от количества точек для построения хеша. Также одним из преимуществ



разработанного метода является отсутствие коллизий хешей, т.е. каждая ячейка исходного пространства будет иметь уникальный хеш.

Література

1. *Bentley J.L.* Multidimensional Divide and Conquer / J.L Bentley.// Communications of the ACM, 1980. – Vol. 23. – Is. 4. – Pp. 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. – Pp. 209–226.
3. *de Berg M.* Computational Geometry: Algorithms and Applications / de Berg M. // – 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 // Proc. ACM SIGMOD International Conference on Management of Data, 1984. –Pp. 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.* WhenIs “NearestNeighbor” Meaningful /K.Beyer, J.Goldstein, R.Ramakrishnan, U.Shaft // InInternational Conferenceon Database Theory: Springer, 1999. –Pp. 217–235.
8. *Nguyen H.* (ed.). GPU Gems 3 / Nguyen H. // Addison Wesley, 2007. – P. 1008.
9. *Ize T. Wald I.* 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.



АЛГОРИТМ ПРОСТОРОВОГО ХЕШУВАННЯ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ ПРИБЛИЗНОГО ПОШУКУ НАЙБЛИЖЧИХ СУСІДІВ

А.О. Дашкевич

Анотація

Запропоновано реалізацію алгоритму і структур даних для просторового хешування в багатовимірних просторах. Розроблений алгоритм дозволяє проводити приблизний пошук найближчих сусідів на багатовимірних сітках, що дозволяє проводити оптимізацію таких задач геометричного моделювання, як апроксимація і виділення поверхонь з точкових множин, сегментація поверхонь.

SPATIAL HASHING ALGORITHM FOR APPROXIMATE NEAREST NEIGHBORS SEARCH

A. Dashkevich

Summary

The need of nearest neighbors search arises in many problems of computational geometry, geometric modeling and computer graphics, for example, polygonal mesh triangulation, collision detection, approximation and surface reconstruction from raw point sets. Developing of approximate nearest neighbors search algorithms, that decrease computational complexity, is the actual problem.

Algorithm and data structures for spatial hashing in multidimensional spaces is described in the paper. The algorithm allows to find approximate nearest neighbors on multidimensional grids for optimizing such problems in geometrical modeling, as surface reconstruction and approximation from point-sets, surface segmentation.