

УДК 004.023

ПРОСТОРОВІ ПЕРЕТВОРЕННЯ ТОЧКОВИХ МНОЖИН НА ЦІЛОЧИСЕЛЬНИХ РЕГУЛЯРНИХ СІТКАХ

Дашкевич А.О., канд. техн. наук,

dashkewich.a@gmail.com, ORCID: 0000-0002-9963-0998

Національний технічний університет «Харківський політехнічний інститут» (м. Харків, Україна)

Роботу присвячено розробці узагальненого підходу до розробки методів просторових перетворень на цілочисельних регулярних сітках для розв'язання задач ефективної з точки зору час обробки точкових множин, що представлені в вигляді дискретизованих даних на регулярних сітках. Прикладами таких перетворень є просторова індексація та візуалізація на двовимірну площину. Постійне зростання як обсягів так і розмірності даних, що оброблюються в практичних задачах призводить до зростання вимог до обчислювальної ефективності алгоритмів, які використовуються для їх розв'язання. Існує широке кола задач, в яких використання просторових перетворень може бути використано для підвищення ефективності розрахунків, наприклад, зберігання та використання ресурсів, швидка обробка даних, візуалізація даних, підвищення швидкості сучасних алгоритмів інтелектуального аналізу даних. Іншими прикладами задач є задачі покриття множини, пошук зображень за змістом, обробка геопросторових даних. Методи для оптимізації обчислювальної ефективності в алгоритмах розв'язання подібних задач найчастіше базуються на використанні методів впорядковування точкових множин через використання просторових структур даних і алгоритмів їх обробки. В роботі представлено узагальнений підхід до створення методів просторових перетворень на основі методів просторової індексації цілочисельних регулярних сіток, який дозволяє зводити алгоритми обробки точкових даних до лінійної складності по часу незалежно від розмірності вхідних даних. В роботі запропоновані схеми просторових перетворень на основі лінійних та ступеневих функцій, а також на основі визначення належності точок до деякої області простору. Основою підходу є операції дискретизації та індексації точкових даних. В роботі розглянуто спосіб вирішення проблеми експоненційного зростання кількості комірок на цілочисельній регулярній сітці для зберігання точок. Запропоновано базові операції та алгоритми прямого та зворотного перетворень в метричних просторах. В роботі задано базові характеристики перетворень та наведено підхід до оцінювання ефективності перетворень на їх основі.

Ключові слова: просторові перетворення, цілочисельна регулярна сітка, точкова множина, дискретизовані дані, просторова індексація,

впорядковування точкових множин, візуалізація.

Постановка проблеми. Зростання обсягів та розмірності даних в сучасному світі призводить до підвищених вимог до алгоритмів обробки таких даних для оптимізації використання ресурсів [1, 2] або для їх швидкої обробки [3] та візуалізації [4], підвищення ефективності сучасних алгоритмів інтелектуального аналізу даних [5] та обробки зображень [6]. Під час розв'язання практичних задач у просторах довільної розмірності часто виникає необхідність визначення просторових відносин між точками. До таких задач можна віднести, наприклад, задачу покриття множини точками, що охоплюють таку множину [7], задачу пошуку схожих зображень [8, 9], задачі обробки геопросторових даних [10, 11]. При цьому зростання обсягів даних в таких задачах вимагає пошуку ефективних з точки зору обчислювальної складності по часу методів та алгоритмів визначення просторових характеристик на цілочисельних регулярних сітках.

Аналіз останніх досліджень і публікацій. Поширеними способами підвищення обчислювальної ефективності при визначенні просторових відносин між точками у метричних просторах є впорядковування точкових множин через використання просторових структур даних і алгоритмів їх обробки [12, 13]. Одним із ефективних прикладів використання просторових структур є індексація методами просторового хешування [6, 8], в тому числі й через нейронні мережі [14]. В роботі [15] запропоновано схему хешування на основі лінійних перетворень для просторів низької розмірності, в роботі [16] проведено її узагальнення для просторів довільної розмірності на основі лінійно-ступеневих перетворень. Іншим способом хешування є індексація на основі належності точки до заданої області простору [17]. Існуючі підходи до розв'язання практичних задач із дискретизованими точковими множинами на основі методів просторової індексації потребують подальших досліджень для створення уніфікованих схем застосування при обробці великих та багатовимірних даних.

Формулювання цілей статті. Розробка узагальненого підходу та методів перетворень точкових множин на цілочисельних регулярних сітках на його основі до розв'язання задач просторової обробки точкових множин у просторах довільної розмірності.

Основна частина. Перетворення простору одної розмірності d_1 в простір розмірності d_2 для дійсних чисел визначається як:

$$f: R^{d_1} \rightarrow R^{d_2}, \quad (1)$$

при чому d_1 може дорівнювати d_2 . У випадку перетворення дійснозначного простору у цілочисельний на основі (1) отримуємо:

$$f: R^{d_1} \rightarrow Z^{d_2}. \quad (2)$$

Випадок перетворення між цілочисельними просторами визначатиметься як:

$$f: Z^{d_1} \rightarrow Z^{d_2}. \quad (3)$$

Визначимо основні види перетворень на основі виразів (2)-(3), які

розглядаються в роботі:

- *просторова індексація (просторове хешування)* – відображення на одновимірний простір:

$$h: M^d \rightarrow Z. \quad (4)$$

- *візуалізація* – відображення на двовимірний простір:

$$f_v: M^d \rightarrow Z^2. \quad (5)$$

де $M \in \{R, Z\}$ – вхідний метричний простір.

Для розв'язання деяких практичних задач може виникнути необхідність визначення значення дискретизованих координат вхідного простору за відомим значенням результату перетворення. Визначимо *обернене* перетворення:

$$\bar{f}: G_M \rightarrow G_O, \quad (6)$$

де $G_M \in Z^{\{1,2\}}$, $G_O \in Z^d$ – перетворена та вхідна сіткові геометричні моделі (СГМ), відповідно.

Функція перетворення, для якої існує зворотне перетворення (6) є *двосторонньою*, функція, для якої такого перетворення не існує є *односторонньою*.

Для хеш-функцій обернене перетворення на основі (6) буде:

$$\bar{h}: H \rightarrow G, \quad (7)$$

або:

$$\bar{h}(h(g_j)) = g_j, \quad (8)$$

де g_j – комірка СГМ.

Аналогічно, хеш-функція, для якої існує зворотне перетворення є двосторонньою, а хеш-функція, для якої такого перетворення не існує є односторонньою.

На основі підходів, представлених в роботах [15, 16] узагальнено просторову індексацію можна визначити як лінійну комбінацію зважених координат точки $p = (x_1, \dots, x_d)$ приведену до цілого значення. Можна визначити сім'ю *лінійних* хеш-функцій для довільного простору:

$$h(p) = \left\lfloor \sum_{i=1}^{d-1} t_i x_i + x_d \right\rfloor, \quad (9)$$

де t_i – параметр дискретизації координатної осі (кількість комірок вздовж осі). Наприклад, індексація точок на двовимірній СГМ на основі виразу (9) може бути представлена як:

$$h(p_D) = y_d t_x + x_D, \quad (10)$$

де $p_D = (x_D, y_D)$ – дискретизована точка на площині, t_x, t_y – параметри дискретизації координатних осей Ox та Oy , відповідно.

Обернення індексу, отриманого за виразом (10):

$$y_D = h \bmod t_y, \\ x_D = \left\lfloor \frac{h - y_D}{t_y} \right\rfloor.$$

Лінійні хеш-функції на основі виразу (9) є двосторонніми тільки для випадку $d = 2$, на основі представленого в роботі [16] методу хешування

для просторів довільних розмірностей для випадків більших розмірностей простору було розроблено метод *ступеневого лінійного хешування*:

$$h(p) = \sum_{i=1}^d \frac{x_{D_i} t_i^{(d-1)n_d}}{d}, \quad (11)$$

де x_{D_i} – i -та дискретизована координата точки $p = (x_1, \dots, x_d)$, $x_i \in [0, \max(x_i)]$:

$$x_{D_i} = \left\lfloor \frac{x_i}{c_i} \right\rfloor.$$

де c_i – розмір комірки вздовж i -ї координатної осі, n_d – максимальна кількість розрядів хешу на одну просторову координату x_{D_i} :

$$n_d = \max \left(\frac{\max(x_i) - \min(x_i)}{c_i} \right).$$

Лінійні хеш-функції на основі виразу (11) є двосторонніми для будь-яких значень d . Обернення ступеневого лінійного хешу здійснюється наступним чином [16]:

```

D = d
while D > 0
     $x_{D_i} = h \bmod t_i^{n_d}$ 
     $h = \left\lfloor \frac{h}{t_i^{n_d}} \right\rfloor$ 
    D = D - 1
end
```

Недоліком лінійних функцій індексації простору є експоненційне зростання кількості необхідних комірок для зберігання об'єктів, що призводить до зростання відповідних вимог до об'єму оперативної пам'яті для зберігання, цей факт обмежує використання сімей лінійних хеш-функцій просторами розмірності $d \leq 10 \div 12$ в залежності від наявної оперативної пам'яті. Таким чином, необхідно знаходити ефективні способи індексації для просторів більш високих розмірностей. Слід відзначити, що внаслідок "прокляття розмірності" у просторах великих розмірностей, навіть при малих значеннях t , необхідний об'єм пам'яті для зберігання комірок СГМ буде суттєво вищий за існуючі на даний момент можливості. Таким чином, можливе лише розбиття на комірки, які включатимуть доволі великі об'єми вхідного простору. Для реалізації такого підходу існують нелінійні схеми просторової індексації, на основі визначення належності до деякої частини простору, до таких методів можна віднести:

- сім'ю методів Locality-sensitive hashing (LSH) [6,8];
- метод сферичного хешування [17].

Розглянемо метод сферичного хешування. *Визначення 1.* Для заданої точкової множини $P = \{p_1, \dots, p_N\}$, $p_i = (x_1, \dots, x_d)$, $i = 1, \dots, N$ та деякої множини сфер $S = \{s_1, \dots, s_K\}$, $s_j = (c_j, r_j)$, c_j, r_j – координати центру та радіус j -ї сфери, відповідно, $j = 1, \dots, K$, $c_j = (x_{c_1}, \dots, x_{c_d})$ визначимо хеш як суму двійкових розрядів:

$$h(p_i) = \sum_{j=1}^K h_j 2^{j-1}, \quad (12)$$

де кожен двійковий розряд хешу h_j визначає належність точки $p_i \in P$ до сфери $s_j \in S$:

$$h_i = \begin{cases} 0 & \text{if } \rho(p_i, c_j) > r_j \\ 1 & \text{if } \rho(p_i, c_j) \leq r_j \end{cases}$$

де $\rho(\cdot)$ – задана метрика.

Радіуси та центри сфер задаються або випадковим чином, або можливе використання спеціальних методів, наприклад, алгоритму кластеризації k -середніх для пошуку сфер, що охоплюють всю точкову множину. Слід відзначити, що метод сферичного хешування може бути легко розширеним на використання довільних геометричних об'єктів замість сфер, наприклад еліпсоїдів або паралелепіпедів тощо, що дозволяє керувати балансом точності охоплення множини точок та швидкістю роботи алгоритму.

Якщо зменшити в методі сферичного хешування радіуси сфер до нуля, залишивши лише точки-центри сфер і знаходити відстані від вхідних точок множини до цих центрів, то можливе поєднання схеми лінійного (або лінійного ступеневого) хешування та схеми сферичного хешування в вигляді методу метричного хешування:

$$h(p) = \sum_{i=1}^K \rho_D(x, c_i), \quad (13)$$

де $\rho_D(\cdot)$ – дискретизоване значення метрики, значення метрик при цьому можна використовувати як координати точок в перетвореному просторі і за рахунок цього проводити зменшення розмірності вхідної точкової множини.

Для оцінювання результату перетворень точкових множин на СГМ з точки зору дослідження їх заповненості та обчислювальної ефективності необхідно визначити такі базові статистичні характеристики точкової множини на СГМ:

- щільність заповнення комірок:

$$d_G = \frac{\sum n_j}{N_G},$$

де $n_j = 1 \mid w_j > 0$, $n_j = 0 \mid w_j = 0$ – двійковий стан заповненості комірки, $N_G = |G|$ – загальна кількість комірок в СГМ;

- середня кількість точок в комірці:

$$\tilde{w} = \frac{\sum w_j}{N_G};$$

- квадрат відхилення від середнього в комірці:

$$w_{vj} = (w_j - \tilde{w})^2;$$

- рівномірність розподілу точок -- середньоквадратичне відхилення в комірці:

$$w_{var} = \frac{1}{N_G} \sum_{j=1}^{N_G} w_{vj};$$

- щільність точок у комірці:

$$D_j = \frac{w_j}{V_C},$$

де V_C – об'єм комірки (однаковий для усіх комірок СГМ).

На основі визначених просторових характеристик введемо такі спеціальні види просторової індексації:

- *щільне* хешування – всі комірки СГМ є заповненими, $d_G = 1$;
- *рівномірне щільне* хешування – в усіх комірках СГМ міститься однакова кількість точок, $\tilde{w} \geq 1$, $w_{var} = 0$;
- *впорядковування* – рівномірне щільне хешування, при якому в кожній комірці СГМ міститься тільки одна точка, $\tilde{w} = 1$, $w_{var} = 0$.

Висновки. В ході виконання роботи було запропоновано узагальнений підхід до розв'язання задач просторової обробки точкових множин у просторах довільної розмірності та методи перетворень точкових множин на цілочисельних регулярних сітках на його основі. Запропонований підхід дозволяє зменшувати обчислювальну складність до константної для операції обробки однієї точки множини, що призводить до зменшення складності до рівня лінійної у випадку обробки усієї множини. Здійснено програмну реалізацію розроблених методів із застосуванням мови програмування Python.

Література

1. Simić S.D., Tanković N., Etinger D. Big data bpmn workflow resource optimization in the cloud. *Parallel Computing*. 2023. Vol. 117. P. 103025.
2. Martens J., Blankenbach J. A decomposition scheme for continuous level of detail, streaming and lossy compression of unordered point clouds. *Graphical Models*. 2023. Vol. 130. P. 101208.
3. Shi P., Li J., Zhang Y. LiDAR localization at 100 fps: a map-aided and template descriptor-based global method. *International Journal of Applied Earth Observation and Geoinformation*. 2023. Vol. 120. P. 103336.
4. Liu D., Peng J., Wang Y., et al. Implementation of interactive three-dimensional visualization of air pollutants using webgl. *Environmental Modelling & Software*. 2019. Vol. 114. P. 188–194.
5. Jing W., Zhao C., Jiang C. An improvement method of DBSCAN algorithm on cloud computing. *Procedia Computer Science*. 2019. Vol. 147. P. 596–604.
6. Winter C., Steinebach M., Yannikos Y. Fast indexing strategies for robust image hashes. *Digital Investigation*. 2014. Vol. 11. P. S27–S35.
7. Liu Y., Zhou C., Cheng Y. S2U: an efficient algorithm for optimal integrated points placement in hybrid optical-wireless access networks. *Computer Communications*. 2011. Vol. 34, No. 11. P. 1375–1388.

8. Majhi M., Mallick A. K. Random projection and hashing based privacy preserving for image retrieval paradigm using invariant and clustered feature. *Journal of King Saud University. Computer and Information Sciences*. 2022. Vol. 34, No. 9. P. 6829–6846.
9. Aghav-Palwe S., Mishra D. Feature vector creation using hierarchical data structure for spatial domain image retrieval. *Procedia Computer Science*. 2020. Vol. 167. P. 2458–2464.
10. Grasso I., Ritter M., Cosenza B., et al. Point distribution tensor computation on heterogeneous systems. *Procedia Computer Science*. 2015. Vol. 51. P. 160–169.
11. Iserte S., Macías A., Martínez-Cuenca R., et al. Accelerating urban scale simulations leveraging local spatial 3d structure. *Journal of Computational Science*. 2022. Vol. 62. P. 101741.
12. Bentley J. L. Multidimensional divide-and-conquer. *Commun. ACM*. 1980. Vol. 23, No. 4. P. 214–229.
13. Guttman A. R-trees: a dynamic index structure for spatial searching / A. Guttman. New York, NY, USA : *Association for Computing Machinery*, 1984. 47–57 p. ISBN 0-89791-128-8.
14. Cui Q., Chen Z.-M., Yoshie O. Delving into the representation learning of deep hashing. *Neurocomputing*. 2022. Vol. 494. P. 67–78.
15. Ніцин О.Ю., Дашкевич А.О., Охотська О.В., Мацулевич О.Є. Спрощений алгоритм просторового хешування для задач молекулярної динаміки. *Науковий вісник таврійського державного агротехнологічного університету*. Мелітополь: ТДАТУ, 2016. Вип. 6, Т. 1. С. 287-291.
16. Дашкевич А.А. Алгоритм просторового хешування для рішення задач приблизительного пошуку найближчих сусідів. *Науковий вісник ТДАТУ*. Мелітополь, 2018. Вип. 8. Т. 1. С. 79-86.
17. Дашкевич А.О., Шоман О.В. Метрична сегментація цифрових зображень на основі методу сферичного хешування. *Обуховські читання: XV Міжнародна науково-практична конференція, м. Київ, 10 березня 2020 року: тези конференції*. Київ: Національний університет біоресурсів і природокористування України, 2020. С. 60-61.

SPATIAL TRANSFORMATIONS OF POINT SETS ON INTEGER REGULAR GRIDS

Andrii Dashkevych

The work is devoted to the development of a generalized approach to the development of methods of spatial transformations on integer regular grids for solving the problems of time-efficient processing of point sets presented in the form of discretized data on regular grids. Examples of such transformations are spatial indexing and visualization on a two-dimensional plane. The constant growth of both the volume and dimension of the data processed in practical problems leads to an increase in the requirements for the computational effi-

ciency of the algorithms used to solve them. There is a wide range of tasks in which the use of spatial transformations can be used to increase the efficiency of calculations, for example, storage and use of resources, fast data processing, data visualization, increasing the speed of modern algorithms for intelligent data analysis. Other examples of tasks are set coverage tasks, image search by content, geospatial data processing. Methods for optimizing computational efficiency in algorithms for solving similar problems are most often based on the use of methods for ordering point sets through the use of spatial data structures and algorithms for their processing. The paper presents a generalized approach to creating methods of spatial transformations based on methods of spatial indexing of integer regular grids, which makes it possible to reduce point data processing algorithms to linear time complexity regardless of the dimension of the input data. The paper proposes schemes of spatial transformations based on linear and power functions, as well as on the basis of determining whether points belong to a certain region of space. The basis of the approach is the operations of discretization and indexing of point data. The paper considers a way to solve the problem of exponential growth of the number of cells on an integer regular grid for storing points. Basic operations and algorithms for direct and inverse transformations in metric spaces are proposed. In the paper, the basic characteristics of transformations are given and an approach to evaluating the effectiveness of transformations based on them is given.

Keywords: spatial transformations, integer regular grid, point set, discretized data, spatial indexing, ordering of point sets, visualization.

References

1. Simić, S. D., Tanković, N., & Etinger, D. (2023). Big data BPMN workflow resource optimization in the cloud. *Parallel Computing*, 117, 103025. URL: <https://doi.org/10.1016/j.parco.2023.103025> [in English]
2. Martens, J., & Blankenbach, J. (2023). A decomposition scheme for continuous Level of Detail, streaming and lossy compression of unordered point clouds. *Graphical Models*, 130, 101208. URL: <https://doi.org/10.1016/j.gmod.2023.101208> [in English]
3. Shi, P., Li, J., & Zhang, Y. (2023). LiDAR localization at 100 FPS: A map-aided and template descriptor-based global method. *International Journal of Applied Earth Observation and Geoinformation*, 120, 103336. URL: <https://doi.org/10.1016/j.jag.2023.103336> [in English]
4. Liu, D., Peng, J., Wang, Y., Huang, M., He, Q., Yan, Y., Ma, B., Yue, C., & Xie, Y. (2019). Implementation of interactive three-dimensional visualization of air pollutants using WebGL. *Environmental Modelling & Software*, 114, 188–194. URL: <https://doi.org/10.1016/j.envsoft.2019.01.019> [in English]
5. Jing, W., Zhao, C., & Jiang, C. (2019). An improvement method of DBSCAN algorithm on cloud computing. *Procedia Computer Science*, 147, 596–604. <https://doi.org/10.1016/j.procs.2019.01.208> [in English]

6. Winter, C., Steinebach, M., & Yannikos, Y. (2014). Fast indexing strategies for robust image hashes. *Digital Investigation*, 11, pp. 27–S35. URL: <https://doi.org/10.1016/j.diin.2014.03.004> [in English]
7. Liu, Y., Zhou, C., Cheng, Y., 2011. S2U: An efficient algorithm for optimal integrated points placement in hybrid optical-wireless access networks. *Computer Communications* 34, 1375–1388. URL: <https://doi.org/10.1016/j.comcom.2011.02.005>. [in English]
8. Majhi, M., & Mallick, A. K. (2022). Random projection and hashing based privacy preserving for image retrieval paradigm using invariant and clustered feature. *Journal of King Saud University - Computer and Information Sciences*, 34(9), 6829–6846. URL: <https://doi.org/10.1016/j.jksuci.2022.04.018> [in English]
9. Aghav-Palwe, S., & Mishra, D. (2020). Feature Vector Creation Using Hierarchical Data Structure for Spatial Domain Image Retrieval. *Procedia Computer Science*, 167, 2458–2464. URL: <https://doi.org/10.1016/j.procs.2020.03.298> [in English]
10. Grasso, I., Ritter, M., Cosenza, B., Bengler, W., Hofstetter, G., & Fahringer, T. (2015). Point Distribution Tensor Computation on Heterogeneous Systems. *Procedia Computer Science*, 51, 160–169. URL: <https://doi.org/10.1016/j.procs.2015.05.217> [in English]
11. Iserte, S., Macías, A., Martínez-Cuenca, R., Chiva, S., Paredes, R., & Quintana-Ortí, E. S. (2022). Accelerating urban scale simulations leveraging local spatial 3D structure. *Journal of Computational Science*, 62, 101741. URL: <https://doi.org/10.1016/j.jocs.2022.101741> [in English]
12. Bentley, J. L. (1980). Multidimensional divide-and-conquer. *Commun. ACM*, 23(4), 214–229. URL: <https://doi.org/10.1145/358841.358850> {in English}
13. Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data*, 47–57. URL: <https://doi.org/10.1145/602259.602266> [in English]
14. Cui, Q., Chen, Z.-M., & Yoshie, O. (2022). Delving into the representation learning of deep hashing. *Neurocomputing*, 494, 67–78. URL: <https://doi.org/10.1016/j.neucom.2022.04.082> [in English]
15. Nitsyn, A., Dashkevich, A., Ohotskaya, E., & Matsulevich, A. (2016). Simplified spatial hashing algorithm for molecular dynamics problems. *Scientific bulletin of the Tavria agrotechnological state university*, is. 6, vol. 1, 287–291. [in Ukrainian]
16. Dashkevich, A. (2018). Spatial hashing algorithm for approximate nearest neighbors search. *Scientific bulletin of the Tavria agrotechnological state university*, is. 8, vol. 1, 79–86. [in Russian]
17. Dashkevych, A. (2020). Metric segmentation of digital images based on the method of spherical hashing. *Tezy dopovidey XV Mizhnarodnoyi naukovo-praktychnoyi konferentsiyi "Obukhovs'ki chytannya"*, 60–61. [in Ukrainian]