

ТОПОЛОГІЧНИЙ ПІДХІД ДО ВЕКТОРИЗАЦІЇ РАСТРОВОЇ ГРАФІКИ НА ОСНОВІ СИМПЛІЦІАЛЬНИХ КОМПЛЕКСІВ

Sitko D.O., Hnidenko M.P. A topological approach to raster graphics vectorization based on simplicial complexes. This paper elaborates a topological approach to raster graphics vectorization using methods of Topological Data Analysis (TDA): simplicial complexes and persistent homology. The primary focus is on the development of an algorithm for the construction of simplicial complexes, which is an essential tool for identifying the structural and spatial characteristics of an image. The properties of simplicial complexes in the context of pixel-based data analysis are investigated, and an algorithm for generating simplicial structures from raster images is proposed. The developed approach enables effective analysis of topological relationships between pixels and facilitates the extraction of significant components for subsequent transformation into a vector representation. A simplicial homology search algorithm is also described. The proposed method enhances image segmentation accuracy, which is critical for high-quality vectorization, particularly in cases of complex geometry, structural heterogeneity, and image noise. The results demonstrate the potential of topological methods for improving image preprocessing and segmentation, offering new perspectives in vectorization tasks. The research lays a foundation for further algorithmic development and practical implementation in computer graphics and image processing applications.

Keywords: image vectorization, simplicial complexes, topological data analysis, image segmentation, raster graphics, TDA, topological methods, Vietoris-Rips complex, persistent homology

Сітко Д.О., Гніденко М.П. Топологічний підхід до векторизації растрової графіки на основі симпліціальних комплексів. У статті розглядається топологічний підхід до векторизації растрової графіки з використанням методів симпліціальних комплексів та пошуку сталих гомологій. Основну увагу зосереджено на розробці алгоритму побудови симпліціальних комплексів – ефективного інструменту для виявлення структурних та просторових характеристик зображення. Досліджено властивості симпліціальних комплексів у контексті сегментації піксельних даних та розроблено алгоритм формування топологічних структур на основі пікселів растрового зображення. Запропоновано алгоритм пошуку сталих гомологій. Запропонований підхід дозволяє ефективно аналізувати топологічні зв'язки між пікселями, виокремлювати значущі компоненти та підвищити точність сегментації зображень. Перспективність результатів полягає в покращенні якості векторизації, особливо у випадках складної геометрії, структурної неоднорідності та наявності шуму в даних.

Ключові слова: векторизація зображень, симпліціальні комплекси, топологічний аналіз даних, сегментація зображень, кластеризація, растрова графіка, TDA, топологічні методи, метод сталих гомологій, комплекс Вієторіса-Ріпа

Вступ

Топологічний аналіз даних (ТАД) є потужним інструментом для вивчення структури і форми даних з метою отримання корисної інформації. Використовуючи дані методи можна представити певні аспекти структури даних через просту та зрозумілу топологічну характеристику. Однією з основних переваг ТАД є можливість аналізу зв'язків між елементами даних, що дозволяє розкривати приховані закономірності та особливості. У даній статті увагу буде зосереджено на дослідженні поширених топологічних методах: побудови симпліціальних комплексів та пошуку сталих гомологій.

У даній роботі розглядається застосування топологічних підходів до векторизації растрових зображень за допомогою симпліціальних комплексів. Основною метою є розробка алгоритмів побудови топологічних структур, які дозволяють ефективно аналізувати та сегментувати зображення, що відкриває нові можливості для обробки зображень з високими рівнями складності геометрії та шумів даних.

Аналіз останніх досліджень. У своїх роботах Зомородян та Карлсон досліджують основи персистентної гомології, пропонуючи методи для обчислення та спрощення топологічних характеристик даних. Вони зосереджуються на розробці алгоритмів для обчислення топологічних інваріантів, що дозволяють аналізувати структури даних через їхні топологічні

зв'язки – гомології. Ці роботи також включають методи для побудови комплексів Вісториса-Ріпса, що є основами аналізу великих наборів даних, маючи помірну обчислювальну складність [1-3].

Манч дає розширене введення в топологічний аналіз даних (ТАД) у своєму посібнику, надаючи інформацію про різні топологічні методи, такі як діаграми сталості та граф Маррер. Це дослідження демонструє, як застосування топології може допомогти виявляти структури в даних та полегшити їхню візуалізацію, роблячи ці методи доступними для широкого кола дослідників [4].

Тауз та Карлсон вводять концепцію гомологічної кординатизації, що дозволяє представляти в топологічно адаптованому вигляді для полегшення їх подальшої обробки. Цей метод особливо корисний для покращення результатів, отриманих за допомогою персистентної гомології, і може бути застосований для аналізу складних структур у даних [5].

Карлсон в своїй роботі [6] здійснює загальний огляд топології та даних, зокрема звертаючи увагу на важливість топологічних підходів для обробки структурованих даних. Він демонструє, як топологічні інваріанти можуть бути застосовані для вирішення задач з великими наборами даних у різних наукових та інженерних галузях.

В роботі Янга та інших пропонується метод Dense RepPoints, що використовує щільні набори точок для представлення візуальних об'єктів. Це дозволяє отримати точні результати при виявленні форми та структури зображень, що є корисним для поєднання з іншими методами виявлення гомологій [7].

У статті Русселя та інших авторів розглядається топологічно сегментація для розпізнавання дорожніх мереж із даних ALS. Це дослідження демонструє доцільність та ефективність топологічних методів для сегментації просторових даних та покращення їхньої обробки в геопросторових науках [8].

Книга Едельсбруннера та Харера надає ґрунтовне введення до комп'ютерної топології, що включає аналіз основних методів і алгоритмів, які застосовуються в топологічному аналізі даних. Це джерело є базовим для тих, хто хоче освоїти основи цієї галузі та зрозуміти зв'язок між математичною теорією і її реальними застосуваннями [9].

Робота Карлсона та Вейдемо-Йоханссона надає огляд застосувань топологічного аналізу даних для різних наукових задач. Вони детально описують практичне використання персистентної гомології для вирішення задач у науці та інженерії, пропонуючи нові підходи до обробки та аналізу даних [10].

У статті Єсілли та Хасавнеха пропонується автоматичний метод для аналізу поверхневих текстур, використовуючи персистентну гомологію. Це відкриває нові можливості для текстурного аналізу в комп'ютерному зору та машинному навчанні, де точне виявлення текстур є критичним для подальших застосувань [11].

Наукова стаття Коркорана та Джоунса аналізує методи пошуку виокремлених топологічних структур для аналізу географічних даних. Це дослідження демонструє сучасні підходи до обробки географічної інформації та просторового аналізу, що є важливим для наукових досліджень у галузі географії [12].

Постановка завдання. У сучасних задачах аналізу зображень все більшої актуальності набувають підходи, що дозволяють не лише виявляти локальні особливості зображення, а й досліджувати його глобальну структуру. Одним із перспективних напрямів є топологічний аналіз даних, який забезпечує інструментарій для вивчення топологічних та геометричних властивостей даних за допомогою таких структур, як симпліціальні комплекси. У контексті векторизації растрової графіки це відкриває нові можливості та підходи для формалізації виявлення контурів, збереження просторових зв'язків та виділення значущих компонент.

Однак, постає низка дослідницьких викликів: як адаптувати методи побудови симпліціальних комплексів до дискретної природи растрових зображень? Як інтерпретувати гомологічні характеристики у контексті графічних об'єктів? Яким чином використати ці топологічні репрезентації для ефективної сегментації та векторного представлення складних структур?

У зв'язку з цим необхідним є:

- дослідити наявні підходи побудови симпліціальних комплексів на основі піксельних даних;
- адаптувати алгоритми топологічного аналізу до потреб векторизації;
- розробити алгоритм, що забезпечує формування симпліціальної структури з растрового зображення;
- проаналізувати ефективність побудованого алгоритму в умовах складної геометрії, наявності шуму та фрагментарності зображення.

Результати дослідження можуть бути використані для підвищення якості автоматизованої векторизації, а також у суміжних галузях, таких як комп'ютерний зір та цифрова обробка зображення, картографія, археологія та аналіз медичних зображень.

Метою роботи є розробка методики застосування інструментів топологічного аналізу даних (TDA) та побудови симпліціальних комплексів в задачах дослідження форми графічного об'єкту та векторизації растрової графіки. Основна увага приділяється виявленню топологічних структур на зображенні, що дозволяє точніше зберігати просторові й логічні взаємозв'язки об'єктів при переході від піксельного до векторного подання. У рамках дослідження передбачається побудова моделі, яка формалізує процес векторизації на основі гомологічних характеристик, таких як зв'язність множин точок, що формують кільцеві структури та порожнини.

Очікується, що впровадження запропонованого підходу підвищить точність векторизації, зменшить втрату важливої просторової інформації та забезпечить кращу адаптивність до складних графічних об'єктів і нечітких зображень.

Виклад основного матеріалу дослідження

Топологічний простір. Топологічний простір – це множина точок, та пов'язаних із сусідами. Топологія на множині X – це підмножина $T \subseteq 2^X$ така, що:

1. Якщо $S_1, S_2 \in T$, то $S_1 \cap S_2 \in T$.
2. Якщо $\{S_j \mid j \in J\} \subseteq T$, то $\bigcup_{j \in J} S_j \in T$.
3. $\emptyset, X \in T$.

Пара (X, T) – це топологічний простір. Множина $S \in T$ називається відкритою множиною, а її доповнення в X – замкненою. Підмножина $A \subseteq T$ з індукованою топологією $T_A\{S \cap A \mid S \in T\}$ є підпростором A від X . Відомим прикладом топологічного простору є d -вимірний Евклідовий простір R^d . Таким чином можливо перетворити будь-яку підмножину Евклідового простору у топологічний простір використовуючи індуковану топологію [1, 2].

Функція $f: X \rightarrow Y$ є неперервною якщо для кожної відкритої множини A в Y , $f^{-1}(A)$ є відкритою в X . Гомеоморфізм $f: X \rightarrow Y$ є бієкцією такою, що f та f^{-1} обидві є неперервними. Враховуючи гомеоморфізм $f: X \rightarrow Y$, можемо казати, що X є гомеоморфний до Y . Оскільки гомеоморфізм є співвідношенням еквівалентності на топологічних просторах, ми також скажемо, що X та Y мають однаковий топологічний тип. Топологічний тип є найкращим рівнем класифікації, доступним у топології.

Гомотопія – це сімейство відображень $f_t: X \rightarrow Y, t \in [0, 1]$, таких, що пов'язане відображення $F: X \times [0, 1] \rightarrow Y$, задана як $F(x, t) = f_t(x)$ є неперервною. Тут $X \times [0, 1]$ є топологічним простором, відкриті множини кого є продукцією відкритих множин з X та відкритих множин з простору $[0, 1]$, який є підпростором R . Тоді $f_0, f_1: X \rightarrow Y$, є гомотопними через гомотопію f_t . Відображення $f: X \rightarrow Y$ називається гомотопною еквівалентністю, якщо існує відображення $g: X \rightarrow Y$, таке, що $f \circ g \cong 1_Y$ та $g \circ f \cong 1_X$, де $1_Y, 1_X$ – одиничні відображення на відповідних просторах. Маючи гомотопну еквівалентність $f: X \rightarrow Y$, ми кажемо, що X та Y гомотопно еквівалентні та мають однаковий гомотопний тип. Гомотопний тип є більш жорсткою класифікацією, ніж топологічний тип.

Симпліціальний комплекс. Симпліціальні комплекси (СК) є одним із основних методів топологічного аналізу даних через їх структурну простоту. СК нагадує гіперграф, в якому зв'язки між $(n+1)$ вузлами за представлені за допомогою n -вимірного симплексу. Так

СК – множина K , яка складається із скінченного набору множин, що знаходяться в підмножинних відносинах: якщо $\sigma \in K$ та $\tau \subseteq \sigma$, то $\tau \in K$. Так σ – це симплекс, а τ – його грань. (-1) -симплекс $\emptyset \in K$ є гранню будь-якого симплекса. Симплекс є максимальним, якщо він не є гранню жодного симплекса вищого порядку в K . Якщо $\sigma \in K$ має потужність $|\sigma| = n + 1$, то σ – n -симплекс розмірності n ($\dim(\sigma) = n$). Узагальнюючи, якщо максимальна розмірність симплекса в K рівна d , то K називають d -вимірним комплексом $\dim(K) = d$.

Уявлення розмірності симплексів обумовлене нашою здатністю уявляти n -вимірний симплекс геометрично, як n -вимірний підпростір R^d , $d \geq n$, наприклад, випукла оболонка з $(n+1)$ афінно-незалежних точок. З такої точки зору n -симплекси називають точкою, відрізком, трикутником та тетраедром для $0 \leq n \leq 3$ відповідно. Ключовою властивістю реалізованого симплексу є його стягуваність. Симплиціальний комплекс K може бути вбудований в Евклідовий простір як об'єднання геометричних реалізацій його симплексів, так, щоб вони перетинались лише через спільні грані. Таке об'єднання є основним простором $|K| = \bigcup_{\sigma \in K} \sigma$ топологічного простору K . Топологічні інваріанти, такі як гомотопічний тип, не залежать від певної геометричної реалізації комплексу.

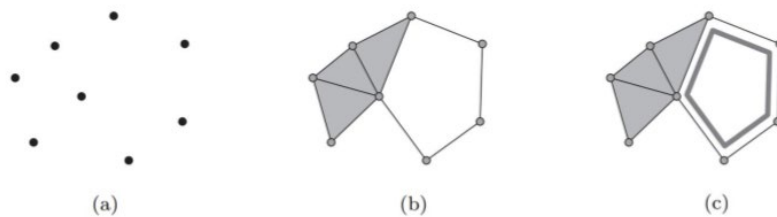


Рис. 1.1.(а) Набір точок S . (б) Відображення K . (в) Інваріанта.

Рис.1.1.(б) зображає геометричну реалізацію симплиціального комплексу з 8 вершинами, 11 гранями та 3 трикутниками. Трикутники та 4 грані, що визначають дірку є максимальними.

Під-комплексом називається підмножина $L \subseteq K$, що також є симплиціальним комплексом. Важливим під-комплексом є n -скелет, що складається з симплексів K , розмірності менше ніж, або рівній n . 1-скелет симплиціального комплексу – це граф.

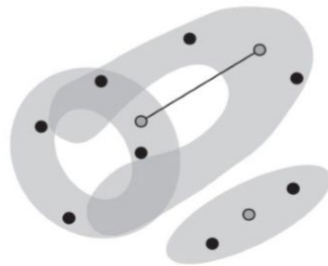


Рис. 1.2. Множина 8 чорних точок, відкрите покриття з 3 множин, та нерв цього покриття: симплиціальний комплекс із 3 вершин та однієї грані.

Припустимо, що ми маємо скінченну множину точок S , отриману з невідомого простору X , та вбудовану в деякий топологічний простір Y . Ключова ідея топологічного аналізу полягає в локальній апроксимації X , використовуючи фрагменти побудованого простору Y . Відкрите покриття S – це $U = \{U_i\}_{i \in I}$, $U \subseteq Y$, де I – індексуєча множина, $S \subseteq \bigcup_i U_i$, та U_i – відкриті. Нерв N покриття U – це:

1. $\emptyset \in N$.
2. Якщо $\bigcap_{j \in J} U_j \neq \emptyset$ для $J \subseteq N$, то $J \in N$.

Таким чином що нерв – це симплиціальний комплекс. Об'єднання множин в відкритому покритті є нашою апроксимацією невідомого X . Його нерв слугує скінченим комбінаторним представленням, яке може бути використане при обрахунках. Якщо множини покриття не приховують цікаву топологію, або в собі, або в їх схемах перетину, вся топологія зображується в нерві. Формально, покриття U є хорошим, якщо усі U_i є стягуваними та такими ж є усі їх непорожні скінченні перетини. Згідно з леми нерва, нерв хорошого покриття є гомотопно

еквівалентним до самого покриття, тобто до об'єднання множин в покритті [1, 2]. Ця лема слугує базою для декількох методів представлення множини точок.

Комбінаторне представлення симпліціальних комплексів. На вході маємо скінченну множину точок $S \subseteq Y$. Ми будемо комбінаторне представлення K що апроксимує простір X з якого була отримана множина S . Будемо вважати, що Y є метричним простором з метрикою $d: Y \times Y \rightarrow R$.

Комплекс Чеха. Нехай $B_\varepsilon(x)$ - відкрита куля з радіусом ε та центром в точці x . Тобто, для $\varepsilon \in R, x \in Y$,

$$B_\varepsilon(x) = \{y \in Y | d(x, y) < \varepsilon\}$$

Маючи $\varepsilon \in R$ та $S \subseteq Y$, ми будемо ε -кулю на кожній з точок щоб отримати поверхню:

$$U_\varepsilon = \{B_\varepsilon(x) | x \in S\}.$$

Комплекс Чеха C_ε є нервом цього покриття. Оскільки кулі є опуклими, а опуклі множини є стягуваними, то покриття є хорошим, а його нерв фіксує топологію покриття.

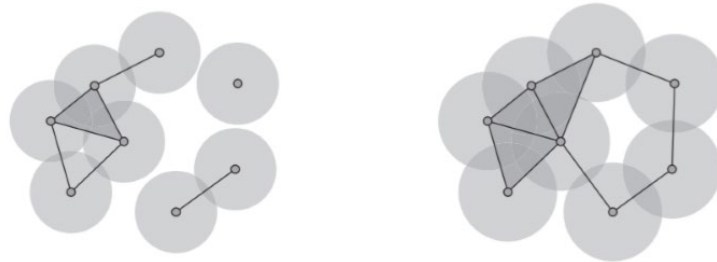


Рис. 1.3. Комплекс Чеха C_ε є нервом покриття з ε -куль. На рисунку зображено два комплекси при двох різних значеннях ε .

Рисунок 1.3. демонструє покриття U_{ε_1} та U_{ε_2} з двома різними значеннями $0 < \varepsilon_1 < \varepsilon_2$. Нерв кожного покриття намальований поверх. Зауважимо, що кожний нерв є гомотопно еквівалентним до свого покриття: покриття та нерв на першому малюнку обидва мають 3 компоненти та одну дірку, а покриття та нерв на другому малюнку 1 компоненту та 1 дірку.

Ми можемо побудувати комплекс Чеха для будь-яких значень ε . Очевидно, що $C_0 = \emptyset$ та C_∞ - це $(|S|-1)$ -симплекс. Тобто комплекс Чеха може мати набагато вищу розмірність ніж вбудований простір Y . Оскільки n -симплекс має 2^{n+1} граней, комплекс може стати масивним при великих значеннях ε .

Зазвичай комплекс Чеха рідко використовується на практиці через складність його обчислення. Однаковий радіус для усіх куль відповідає припущенню про рівномірність відбору вхідних даних, що не є дійсним для реальних наборів даних. Ми можемо використовувати неоднорідні радіуси для формування покриття. Ця ідея була досліджена іншими методами, такими як альфа-комплекс.

Альфа комплекс. Щоб зменшити розмір комплексу, ми обмежуємо розмірність, використовуючи геометрію вбудованого простору. Маючи $S \subseteq Y$, регіон Вороного $R(x)$ точки $x \in S$ - це множина точок з Y , що є найближчими до x :

$$R(x) = \{y \in Y | d(x, y) \leq d(x', y), \forall x' \in S, x' \neq x\}$$

Діаграма Вороного – це множина усіх регіонів Вороного для точок з S . Ця діаграма може бути розглянута як замкнута поверхня для Y . Комплекс Делоне є нервом діаграми Вороного. Покриття Вороного та його нерв є фундаментальними геометричними об'єктами і були широко вивчені в рамках обчислювальної геометрії.

Рисунок 1.4.(а) зображає діаграму Вороного для нашої множини точок, і покриває її нервом – комплексом Делоне.

Тепер ми можемо використати діаграму Вороного, щоб обмежити взаємодію покриття ε -куль з попереднього розділу. Для кожної точки $x \in S$, ми об'єднуємо ε -кулю та регіон Вороного, щоб отримати обмежений регіон Вороного. Множина усіх обмежених регіонів формує покриття:

$$U_\varepsilon = \{B_\varepsilon(x) \cap R(x) | x \in S\}.$$

Альфа-комплекс A_ε є нервом цього покриття. Згідно з означення, $A_\varepsilon = \emptyset$, A_∞ - це комплекс Делоне, а A_ε – під-комплекс комплексу Делоне, для будь-якого ε . Більш того, альфа-комплекс та комплекс Чеха є гомотопно еквівалентними. Проте, на відміну від комплексу Чеха, максимальна розмірність альфа-комплексу обмежена вбудованою розмірністю, якщо S знаходиться в загальному положенні.

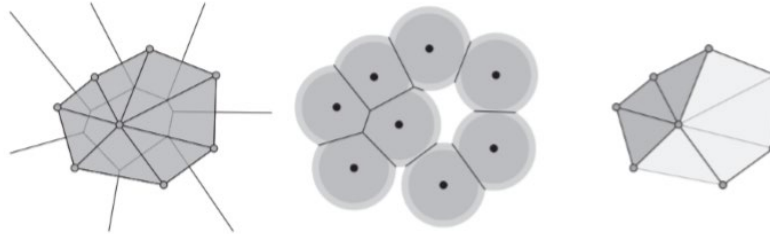


Рис. 1.4. (а) Набір даних, діаграма Вороного та її нерв, комплекс Делоне. (б) Обмежені регіони Вороного, що формують покриття, нервом якого є альфа-комплекс (в), зображений як під-комплекс світлішого комплексу Делоне.

Рис. 1.4.(б) накладає обмежені регіони Вороного для нашої множини точок для двох різних значень ε , що були використані в Рис. 1.3.

Рис. 1.4.(в) зображає альфа-комплекс A_{ε_2} як підкомплекс комплексу Делоне. При такому значенні ε , комплекс є таким самим як і комплекс Чеха $2C_{\varepsilon_2}$ на рис. 1.3.(б).

Ми починаємо побудову альфа-комплексу з побудови комплексу Делоне. Далі для кожного симплекса комплексу Делоне ми обчислюємо мінімальне значення ε , при якому симплекс увійде до альфа-комплексу. Після цього ми сортуємо симплекси по мініальному ε , щоб отримати часткову упорядкованість симплексів. Тепер ми можемо побудувати альфа-комплекс для будь-якого значення ε використовуючи цю впорядкованість. Оскільки комплекс Делоне є скінченим, альфа-комплекс змінюється лише у скінченній кількості критичних значень ε при зміні від 0 до нескінченності.

Використання рівномірних радіусів передбачає неявне припущення про рівномірність збору даних. Це припущення може бути зняте шляхом узагальнення альфа-комплексу. Ми можемо припустити, наприклад, що множина точок є зваженою, а вага точки прив'язана до локального розміру об'єкта. В результаті ми можемо використати показник потужності для визначення матриці потужності покриття та його нерву, регулярної триангуляції.

Сьогодні існують ефективні алгоритми та програмне забезпечення для обрахунку комплексу Делоне, а отже і альфа-комплексу в просторах розмірності 2 та 3, тож комплекс добре підходить для топологічного аналізу у низьких розмірностях. Проте його побудова у просторах високої розмірності є важкою.

Комплекс Вісторіса-Ріпа. Комплекс Вісторіса-Ріпа є популярним в топологічному аналізі, через легкість його побудови навіть у просторах великої розмірності. На відміну від більшості інших комплексів, він базується на графі, а не на покритті. Дано $S \subseteq Y$, $\varepsilon \in R$, тоді $G_\varepsilon = (S, E_\varepsilon)$ - це граф ε -сусідів на S , де:

$$E_\varepsilon = \{\{u, v\} | d(u, v) \leq \varepsilon, u \neq v \in S\}.$$

Кліка в графі – це підмножина вершин, що створює повний під-граф. Кліка є максимальною, якщо її не можна зробити більшою. Кліковий комплекс, який також називають флаговим комплексом, має максимальні кліки графа як його максимальні симплекси. Комплекс Вісторіса-Ріпа V_ε є кліковим комплексом графа ε -сусідів. Надалі будемо його називати ВР комплексом.

Рис.1.5. зображає побудову ВР комплексу для нашої множини точок. Ми починаємо з побудови графа $2\varepsilon_1$ -сусідів (рис. 1.3.(а)). Граф має 5 максимальних кліків, позначених сірими овалами. Кожний з максимальних кліків стає максимальним симплексом в ВР комплексі $V_{2\varepsilon_2}$ (рис. 1.3.(б)).



Рис.1.5.(а) Граф та його кліки. (б) VR комплекс.

Щодо схожості з іншими комплексами, VR комплекс є схожим до комплексу Чеха, проте не є з ним гомотопно еквівалентним, тому ми можемо розглядати його як апроксимацію. Як і комплекс Чеха, VR комплекс може мати максимальну розмірність симплекса - $(|S| - 1)$. Такий граничний випадок виникає тоді, коли граф є повним, тобто будь-яка пара вершин є суміжною, а отже існує ребро, що сполучає ці вершини. На практиці нам зазвичай потрібно побудувати лише n -скелет для $n \leq |S|$. Ми також обраховуємо VR комплекс V_ε для максимального значення $\varepsilon \in R$. Для кожного симплекса $\sigma \in V_\varepsilon$ ми також обраховуємо мінімальне значення ε при якому симплекс входить в VR комплекс, при цьому 0-симплекси (точки) входять до нього при $\varepsilon = 0$, а 1-симплекси (відрізки) при ε рівному їх довжині.

Алгоритм побудови комплексу Вісторіса-Ріпа. Класичний топологічний аналіз даних виконується у два основні етапи, як зображено на рисунку 1.1. Вхідними даними є множина точок S (рис. 1.1.(а)).

Перший крок передбачає апроксимацію невідомого простору X за допомогою комбінаторної структури K (рис. 1.1.(б)). Одним із найпоширеніших представлень є комплекс Вісторіса-Ріпа.

Далі розраховуються топологічні інваріанти структури K , такі як цикл c (рис. 1.1.(в)). Топологічні інваріанти забезпечують апроксимацію властивостей простору X у скінченному представленні S .

Хоча алгоритм є ефективним, він характеризується високою обчислювальною складністю при великій кількості точок у просторі з багатьма вимірами.

Вхідними даними є множина точок S та значення максимальної відстані P . Алгоритм побудови комплексу складається з двох кроків:

- 1) Поєднання точок, які знаходяться на відстані не більше ніж P . Це визначає 1-скелет комплексу, що еквівалентний графу сусідів.
- 2) Розширення комплексу на основі графа сусідів до симплексів вищої розмірності.

***GetClosestNeighbours* (G, u)**

1 return $\{v \in G.V \mid u > v, \{u, v\} \in G.E\}$

Граф сусідів представлений у вигляді ієрархічної структури, де ключами є точки з множини S , впорядковані за віддаленістю від початку координат. Значення – множини точок-сусідів, впорядковані за зростанням відстані.

Інформація про існуючі ребра (1-симплекси) зберігається у впорядкованій множині, де елементи впорядковані за довжиною. Для цього використовується механізм порівняння об'єктів за довжиною ребра.

Розширення комплексу. На даному етапі ми маємо граф сусідів (1-скелет). Задача - побудова симплиціального комплексу більшої розмірності. Використовується індуктивний підхід:

- 1) Ініціалізація масиву симплексів G . Починаємо з графа сусідів, що містить усі симплекси розмірності 1. Встановлюємо $i = 1$.
- 2) Для кожного симплекса τ розмірності i знаходимо множину N – точки, що є сусідами всіх точок із τ та йдуть перед ними у S .

3) Для кожної вершини $v \in N$ стверджується, що $\delta = \tau \cup \{v\}$ – це симплекс розмірності $(i + 1)$. Додаємо δ до G .

4) Якщо не додано жодного симплекса розмірності $(i + 1)$, ітерацію завершено. Інакше: $i = i + 1$ і повертаємося до кроку 2.

У підсумку отримуємо n -скелет комплексу Вісторіса–Ріпа, де n – максимально можлива розмірність симплекса для заданої відстані P .

BuildInductiveVR(G, k)

```

1   $V \leftarrow G.V \cup G.E$ 
2  for  $i \leftarrow 1$  to  $k$  do
3    foreach  $i$  as simplex  $r \in V$  do
4       $N \leftarrow \bigcap_{u \in r} \text{Closest Neighbours}(G, u)$ 
5      foreach  $v \in N$  do
6         $V \leftarrow V \cup \{r \cup \{v\}\}$ 
7  return  $V$ 
```

Алгоритм пошуку гомологій. Гомології являють собою зв'язані групи точок у просторі. Пошук гомологій є останнім етапом у процесі кластеризації даних, і кількість виявлених гомологій визначає кількість кластерів.

Існує кілька підходів до пошуку гомологій. Загальний алгоритм пошуку симплиціальних гомологій реалізується наступним чином:

1. Побудова матриці 1-симплексів, де 0 позначає відсутність 1-симплекса (ребра), а 1 — його наявність.

2. Виконання операцій над матрицями: додавання за модулем 2 (XOR) та перестановка рядків і стовпців для приведення матриці до нормальної форми Смита.

3. Перевірка можливості подальшої оптимізації.

4. Повторення кроку 2, поки умови для виконання кроку 3 не буде виконано.

Після виконання цих перетворень отримуємо нормалізовану матрицю. Одиниці, що залишаються на головній діагоналі, є орієнтирами для визначення приналежності точок до конкретних кластерів. Якщо існує шлях через ребра Вісторіса–Ріпа до одиниці, яка є певною точкою в матриці, то точка належить даному кластеру.

Висновки

У даній роботі було розглянуто можливість використання методів топологічного аналізу даних та симплиціальних комплексів для задач векторизації растрових зображень. Було проаналізовано основні концепції ТАД, такі як симплиціальні комплекси, персистентна гомологія. Розроблено та продемонстровано алгоритм побудови симплиціального комплексу, а саме комплексу Вісторіса–Ріпа, та показано, що вони здатні виявляти важливі топологічні особливості зображень, зокрема компоненти зв'язності, кільця та порожнини, які мають критичне значення для збереження структури об'єктів при переході до векторної форми. Досліджено та викладено алгоритм пошуку симплиціальних гомологій.

Запропонований підхід дозволяє не лише підвищити точність векторизації, а й зменшити втрату просторової інформації, що особливо важливо при обробці зашумлених або історичних картографічних матеріалів, неоднорідних та складних графічних матеріалів. Використання симплиціальних структур забезпечує високу адаптивність до різноманітних складних форм зображень при помірній складності реалізації.

Простота та ефективність методів топологічного аналізу демонструють перспективність застосування в комп'ютерному зорі та цифровій графіці, зокрема в тих випадках, де класичні алгоритми векторизації не здатні зберегти важливі структурні характеристики зображення. У майбутньому доцільним є подальший розвиток цього підходу, зокрема шляхом інтеграції з глибоким навчанням для автоматизації виявлення топологічних ознак та підвищення продуктивності систем обробки графічної інформації.

Список використаної літератури:

1. Zomorodian, A., & Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational Geometry*, 33(2), 249–274. URL: <https://doi.org/10.1007/s00454-004-1146-yOUCI+3ResearchGate+3OUCI+3>.
2. Edelsbrunner, H., Letscher, D., & Zomorodian, A. (2002). Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4), 511–533. URL: <https://doi.org/10.1007/s00454-002-2885-2>.
3. Zomorodian, A. (2010). Fast construction of the Vietoris-Rips complex. *Computers & Graphics*, 34(3), 263–271. URL: <https://doi.org/10.1016/j.cag.2010.03.007>.
4. Munch, E. (2017). A user's guide to topological data analysis. *Journal of Learning Analytics*, 4(2), 47–61. URL: <https://doi.org/10.18608/jla.2017.42.6>.
5. Tausz, A., & Carlsson, G. (2011). *Homological coordinatization*. arXiv. URL: <https://arxiv.org/abs/1107.0511arXiv>.
6. Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2), 255–308. URL: <https://doi.org/10.1090/S0273-0979-09-01249-X>.
7. Yang, Z., Sun, Y., Liu, S., Shen, C., Jia, J. (2020). Dense RepPoints: Representing visual objects with dense point sets. In A. Vedaldi, H. Bischof, T. Brox, & J. M. Frahm (Eds.), *Computer Vision – ECCV 2020* (Vol. 12366, pp. 226–242). Springer. URL: https://doi.org/10.1007/978-3-030-58589-1_14.
8. Roussel, J.-R., Bourdon, J.-F., Morley, I. D., Coops, N. C., Achim, A. (2023). Vectorial and topologically valid segmentation of forestry road networks from ALS data. *International Journal of Applied Earth Observation and Geoinformation*, 118, 103267. URL: <https://doi.org/10.1016/j.jag.2023.103267>.
9. Edelsbrunner, H., & Harer, J. (2010). *Computational topology: An introduction*. American Mathematical Society.
10. Carlsson, G., & Vejdemo-Johansson, M. (2021). *Topological data analysis with applications* (1st ed.). Cambridge University Press. URL: <https://doi.org/10.1017/9781108975704>.
11. Yesilli, M. C., & Khasawneh, F. A. (2021). Data-driven and automatic surface texture analysis using persistent homology. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1350–1356). IEEE. URL: <https://doi.org/10.1109/ICMLA52953.2021.00219>.
12. Corcoran, P., & Jones, C. B. (2023). Topological data analysis for geographical information science using persistent homology. *International Journal of Geographical Information Science*, 37(3), 712–745. URL: <https://doi.org/10.1080/13658816.2022.2155654>.
13. Snášel, V., Nowaková, J., Xhafa, F., & Barolli, L. (2017). Geometrical and topological approaches to big data. *Future Generation Computer Systems*, 67, 286–296. URL: [https://doi.org/10.1016/j.future.2016.06.005\(upcommons.upc.edu\)](https://doi.org/10.1016/j.future.2016.06.005(upcommons.upc.edu)).
14. Юрчук, І. А. (2014). Метод сталих гомологій топологічного аналізу даних. *Наукоємні технології*, (3)23, 289. URL: <https://jrnl.nau.edu.ua/index.php/SBT/article/view/7397/8431>.
15. De Silva, V., Morozov, D., & Vejdemo-Johansson, M. (2011). Dualities in persistent (co)homology. *Inverse Problems*, 27(12), 124003. URL: <https://doi.org/10.1088/0266-5611/27/12/124003>.
16. Edelsbrunner, H., & Harer, J. (2010). *Computational topology: An introduction*. American Mathematical Society.
17. Carlsson, G., & Vejdemo-Johansson, M. (2021). *Topological data analysis with applications* (1st ed.). Cambridge University Press. URL: <https://doi.org/10.1017/9781108975704>.
18. Snášel, V., Nowaková, J., Xhafa, F., & Barolli, L. (2017). Geometrical and topological approaches to big data. *Future Generation Computer Systems*, 67, 286–296. URL: [https://doi.org/10.1016/j.future.2016.06.005\(upcommons.upc.edu\)](https://doi.org/10.1016/j.future.2016.06.005(upcommons.upc.edu)).
19. Huber, S. (2021). Persistent homology in data science. In P. Haber, T. Lampoltshammer, M. Mayr, & K. Plankensteiner (Eds.), *Data Science – Analytics and Applications* (pp. 81–88). Springer Fachmedien Wiesbaden. URL: [https://doi.org/10.1007/978-3-658-32182-6_13\(OUCI\)](https://doi.org/10.1007/978-3-658-32182-6_13(OUCI)).

20. Wong, C.-C., & Vong, C.-M. (2021). Persistent homology based graph convolution network for fine-grained 3D shape segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE. URL: <https://doi.org/10.1109/ICCV48922.2021.00701>(ResearchGate).
21. Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., & Carlsson, G. (2013). Extracting insights from the shape of complex data using topology. *Scientific Reports*, 3, p. 1236. URL: <https://doi.org/10.1038/srep01236>.
22. van Veen, H. J., Saul, N., Eargle, D., & Mangham, S. W. (2019). Kepler Mapper: A flexible Python implementation of the Mapper algorithm. *Journal of Open Source Software*, 4(42), 1315. URL: <https://doi.org/10.21105/joss.01315>(joss.theoj.org).
23. Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science*, 6, 17. URL: <https://doi.org/10.1140/epjds/s13688-017-0109-5>(SpringerOpen).

Автори статті

Сітко Денис – аспірант, Державний університет інформаційно-комунікаційних технологій, Київ, Україна.

ORCID: 0009-0000-8000-2089

Гніденко Микола – кандидат технічних наук, доцент, Державний університет інформаційно-комунікаційних технологій, Київ, Україна.

ORCID: 0009-0002-5261-3581

Authors of the article

Sitko Denys – postgraduate, State University of Information and Communication Technology, Kyiv, Ukraine.

ORCID: 0009-0000-8000-2089

Hnidenko Mykola – Candidate of Sciences (technical), Associate Professor, State University of Information and Communication Technology, Kyiv, Ukraine.

ORCID: 0009-0002-5261-3581