

С.В. ЕРЕМЕЕВ, Е.А. СЕЛЬЦОВА

**Алгоритм топологического анализа  
пространственных структур в  
геоинформационных системах**

Муромский институт  
(филиал) ФГБОУ ВО  
«Владимирский  
государственный  
университет имени  
А.Г. и Н.Г. Столетовых»,  
г. Муром

*В статье рассматривается анализ структур географических объектов с помощью метода персистентной гомологии. Представлен алгоритм поиска двумерных симплексов и дыр для построения баркода.*

*Исследование выполнено при финансовой поддержке РФФИ и администрации Владимирской области в рамках научного проекта № 17-47-330387.*

### **Введение**

Анализ топологии городских объектов – сложная и ресурсоёмкая задача, требующая автоматизации. Географические объекты, даже те, что не связаны явно, как правило, составляют собой абстрактные структуры [1,2,3]. Одним из способов обнаружения особых свойств географических объектов и их структур является персистентная гомология [4].

Выявление структур, образованных объектами, является одной из основных задач геоинформационных систем [5]. Главная трудность заключается в отделении топологических особенностей от ошибок, искажающих представление о структурах, составляемых объектами. Успешным решением данной проблемы является применение метода персистентной гомологии [6].

Персистентная гомология используется для качественного анализа топологических особенностей в инженерных и научных областях. Она является относительно новым методом вычисления топологических особенностей пространства в различных

измерениях. Самые стойкие особенности обнаруживаются на больших диапазонах расстояний и представляют собой особые черты объектов на карте. Особенности, которые проявляются на малых диапазонах расстояний, являются так называемым «шумом» и не подлежат дальнейшему рассмотрению [7,8].

В статье рассматривается актуальная задача выявления абстрактных пространственных структур на городской карте на основе методов персистентной гомологии.

### **Основные этапы определения особенностей расположения объектов**

В работе входными параметрами для анализа топологических структур является векторная карта, состоящая из пространственных объектов. Для отображения особенностей расположения объектов используется баркод [9]. Баркод отображает структуру объектов на карте как совокупность сегментов на плоскости, где горизонтальная ось соответствует значению расстояния поиска соседних объектов, а вертикальная ось представляет информацию о количестве обнаруженных структур.

Построение баркода можно разделить на четыре этапа:

1. Считывание данных с карты и построение топологических связей между объектами при увеличении расстояния поиска соседних объектов для формирования связи.

2. Составление треугольников – двумерных симплексов на основе полученных связей и занесение результатов в массив (двумерный симплекс – это простейший двумерный выпуклый многогранник с количеством вершин равным трём).

3. Составление дыр, образованных симплексами.

4. Построение баркода.

Для построения топологических связей на первом этапе необходимо рассчитать расстояния между объектами. Для каждой точки на карте рассчитывается евклидово расстояние до остальных точек в двумерном пространстве.

Если точка находится в пределах допустимого расстояния  $d$ , то между данными двумя объектами строится топологическая связь. На рисунке 1 объект 1 будет связан с объектом 2, но не с объектом 3, так как тот находится на слишком большом расстоянии. Затем

заполняется массив объектов и их связей, где первый номер в строке-массиве – объект на карте, а последующие номера – объекты, с которыми он соединён.

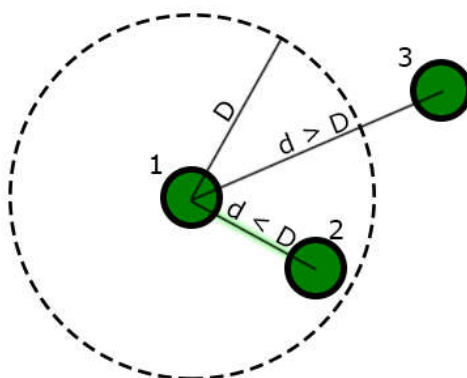


Рис. 1. Поиск соседних объектов для объекта 1

На втором этапе осуществляется работа с полученным на предыдущем шаге списком объектов и их связей. Анализируются связи между точками и, если какие-либо три точки соединены между собой, то составляются симплексы. На рисунке 2 симплекс составляют три точки: 1, 2 и 3. Производятся два первых шага с наименьшего расстояния поиска по максимальное с указанным шагом.

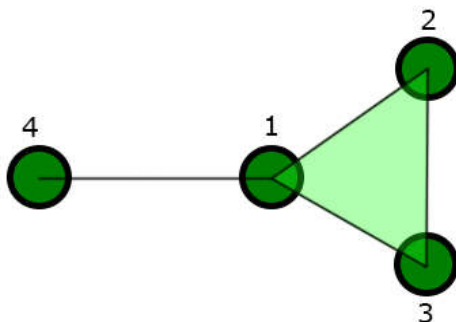


Рис. 2. Симплекс, образованный вершинами 1, 2 и 3

Третий этап будет производиться с «наименьшего» треугольника, так как в противном случае «наибольший» треугольник может «поглощать» меньшие, что приведёт к тому, что в большинстве случаев баркод будет отображать наличие только одной единственной дыры.

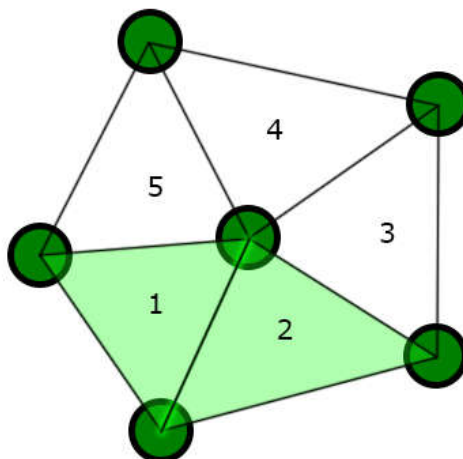


Рис. 3. Дыра из двух симплексов

Третий этап начинается с того, что первый симплекс из массива становится дырой. Затем берётся следующий симплекс. Ищется их общее ребро. В случае, когда это ребро найдено, симплекс приписывается к дыре, а количество дыр не изменяется (рисунок 3). Если общее ребро не было найдено, симплекс записывается второй дырой в массиве дыр, а количество дыр соответственно увеличивается на единицу. Существует случай, когда симплекс может иметь два или три общих ребра с другими дырами. Тогда симплекс заносится в первую подходящую дыру, остальные подходящие дыры приписываются к первой и удаляются из массива. Соответственно на максимальном расстоянии  $d$  между объектами карты в конце останется только одна дыра, в которую будут входить все симплексы, найденные на карте.

### Алгоритм поиска двумерных симплексов

Поиск двумерных симплексов осуществляется по объектам, находящимся на карте, и их связям – эта информация заносится в двумерный массив  $T$  топологических связей объекта. Первым элементом строки-массива является рассматриваемый объект, далее перечисляются связанные с ним объекты. Пример формирования строки-массива  $T$  для рисунка 2 следующий:

$$T[0][0] = 1,$$

$$T[0][1] = 2,$$

$$T[0][2] = 3,$$

$$T[0][3] = 4.$$

Алгоритм поиска двумерных симплексов выполняется для каждой строки-массива  $T$ .

Сначала происходит подсчёт количества объектов в рассматриваемой строке-массиве. Если в ней находятся три объекта, то все они заносятся в контрольный массив  $C$ . Он предназначен для хранения идентификаторов объектов, которые могут составлять двумерный симплекс, и их связей. Массив  $C$  заполняется с помощью следующей формулы:

$$C[0] = T[i]$$

$$C[1] = T[j],$$

$$C[2] = T[k]$$

где  $i, j, k = 0, 1 \dots n$ .

В строке-массиве  $T$  также может находиться и больше трёх объектов. В этом случае берётся первый элемент строки и по два следующих. Для элементов «3;1;2;4;» рассматриваются последовательности:

3, 1, 2;

3, 2, 4;

3, 1, 4;

1, 2, 4.

Таким образом, в контрольный массив заносятся элементы исследуемой строки-массива, имеющие связь с первым объектом той же строки.

Размер контрольного массива  $C$  всегда равен трём, так как любой треугольник (двумерный симплекс) имеет только три вершины. В дальнейшем при подтверждении составления найденными объектами симплекса элементы массива будут занесены в отдельную строку-массив симплексов. Заполнение массива производится для каждой строки-массива  $T$  с каждым набором вершин из неё, где первый элемент набора вершин всегда является первым объектом в исследуемой строке-массиве  $T$ .

После заполнения массива  $C$  проверяется связь последних двух объектов, имеющих в нём. Поскольку массив  $C$  формируется только по связям первого объекта, необходимо проверить, связаны ли оставшиеся два объекта между собой. Если представлять этот процесс графически, вначале только первая точка соединялась с двумя другими. Для составления треугольника необходимо также

соединить вторую и третью вершины, сформировав ребро. Поэтому необходимо проверить, содержит ли третья строка-массив  $C$  первый элемент второй строки-массива. Если  $C[1][0] = C[2][j]$ , то:

$$S[i][0] = C[0][0]$$

$$S[i][1] = C[1][0],$$

$$S[i][2] = C[2][0]$$

где  $i = 0, 1 \dots n$ ,  $j = 0, 1 \dots n$ .

Если проверка пройдена, происходит попытка записи элементов массива  $C$  в строку-массив  $S$  – это массив всех симплексов. Записать строку не удаётся, если в массиве  $S$  уже есть такой симплекс, причем последовательность объектов в нём роли не играет. То есть при наличии строки-массива «1;2;3;» строка-массив «2;3;1;» записана не будет.

### **Алгоритм формирования информации для построения баркода**

Для построения баркода необходимо собрать информацию о дырах и расстояниях, на которых они образуются.

Вначале происходит считывание информации с карты, все объекты и их координаты записываются в соответствующие массивы. Формирование массива симплексов  $S$  происходит в цикле от самого малого радиуса поиска по максимальный.

Если расстояние  $d$  между объектами на карте меньше или равно радиусу поиска, то эти объекты связываются, и информация о них заносится в массив объектов и связей  $T$ . Затем осуществляется поиск двумерных симплексов на текущем расстоянии.

Процесс поиска симплексов на увеличивающемся расстоянии поиска  $d$  отображён на рисунке 4.

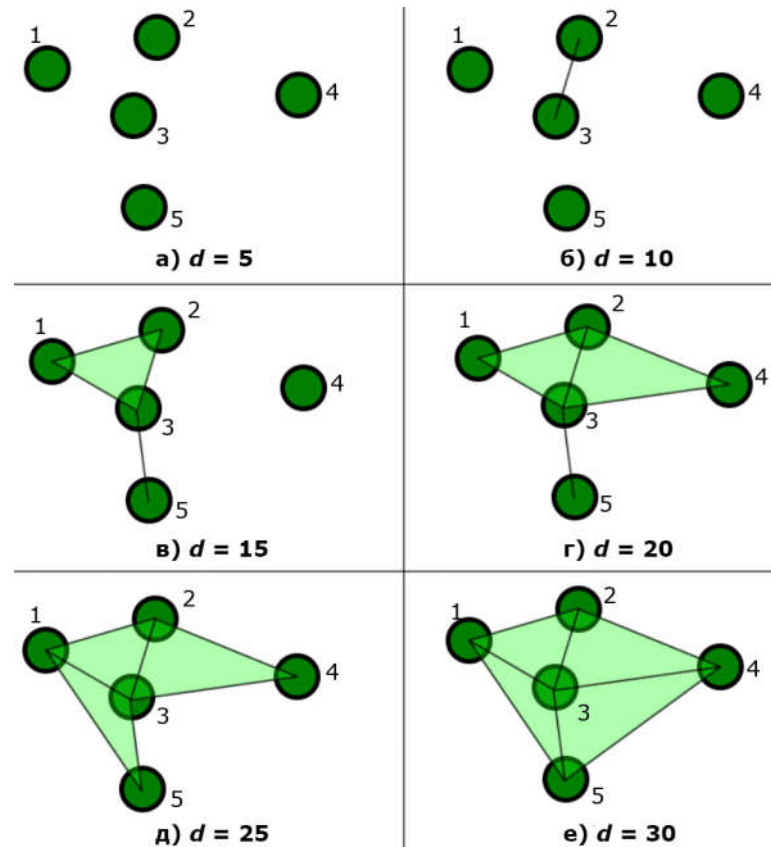


Рис. 4. Формирование связей и симплексов при увеличении расстояния  $d$

Если количество симплексов на текущем шаге стало больше, чем на предыдущем, то это расстояние запоминается для дальнейшего составления баркода. Эти расстояния интересны тем, что появление новых симплексов говорит об изменении количества или состава дыр.

Для каждой строки-массива  $S$  производится поиск дыр. Информация о них содержится в массиве  $H$ , элементами которого являются симплексы. Первый симплекс из массива  $S$  автоматически становится первой дырой. Затем рассматривается второй симплекс. Если у него и уже имеющейся дыры есть хотя бы одно общее ребро, то рассматриваемый симплекс «приписывается» к этой дыре:

$$\begin{cases} H[i] = H[i] + H[k] + S[j], & \text{если } H[k][l] \text{ и } H[i][l] \text{ содержат общее ребро с } S[j] \\ H[i] = H[i] + S[j], & \text{если } H[i][l] \text{ и } S[l] \text{ содержат общее ребро,} \\ H[i] = S[j], & \text{в противном случае} \end{cases}$$

где  $i = 0, 1 \dots n$ ,  $j = 0, 1 \dots n$ ,  $k = 0, 1 \dots i$ ,  $l = 0, 1 \dots n$ .

Если рассматриваемый симплекс не имеет ни одного общего ребра со всеми имеющимися дырами, он заносится в конец массива

$H$  как новая дыра. Также возможен случай, когда симплекс имеет общие рёбра с несколькими дырами сразу. Тогда он приписывается к первой дыре, как и оставшиеся связанные дыры, которые затем удаляются из массива  $H$ .

На основе проведённого анализа дыр запоминается информация об их количестве и составе на отмеченных ранее расстояниях. В результате собирается вся необходимая информация для построения баркода и дальнейшего анализа.

### Экспериментальная часть

Для осуществления поиска дыр необходимо знать идентификаторы объектов, расположенных на карте, и их координаты. В ГИС ИнГЕО у каждого объекта имеется свой уникальный идентификационный номер, также можно определить его координаты.

В процессе работы программы составляется список симплексов  $S$ , в каждой строке находятся номера трёх объектов, входящих в один и тот же симплекс, или вершин треугольника. Все симплексы уникальны. Это означает, что симплексы с другим порядком тех же вершин не добавляются в список.

Также составляется список дыр  $H$ , где каждая запись представляет собой набор симплексов, которые вошли в дыру.

Для построения баркода имеется список расстояний, на которых были найдены новые симплексы, и список количества дыр, отображающий количество дыр, найденных на определённом расстоянии из списка расстояний.

Результатом работы программного модуля является баркод, представляющий собой график, на котором показана информация о количестве дыр в зависимости от расстояния  $d$  поиска соседних объектов. Соответственно баркод отображает структуру расположения объектов на карте.

На вход в программный модуль подаётся карта ГИС ИнГЕО, на которой расположены какие-либо постройки в виде полигонов – замкнутых геометрических фигур, состоящих из отрезков прямых линий. При выполнении алгоритма здания представляются в виде точечных объектов (рисунок 5). При переходе от полигонов к точкам



должно соблюдаться условие расположения точек в центре полигона – постройке.

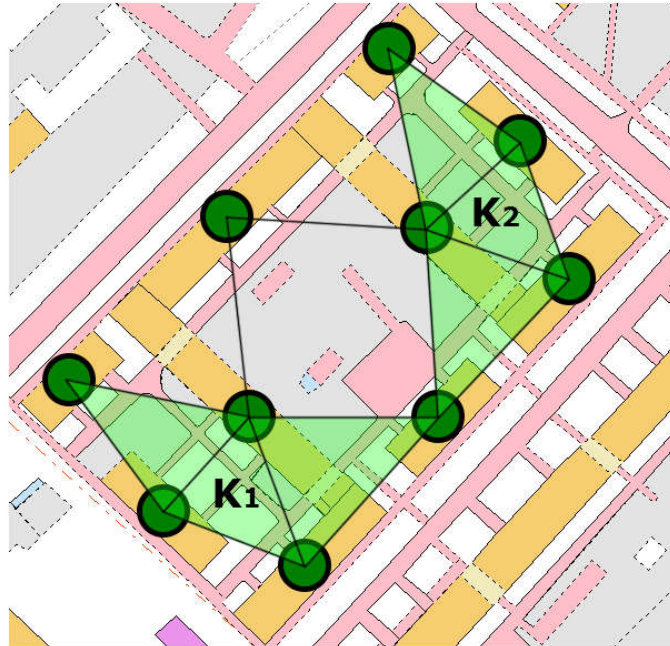


Рис. 5. Расположение точечных объектов на карте и отображение абстрактных структур

Модуль считывает объекты с карты и определяет их координаты, затем в цикле строит связи для меняющегося расстояния  $d$ , на котором осуществляется поиск соседних объектов. На основе этих связей происходит составление двумерных симплексов, которые впоследствии объединяются в дыры.

Как только программа рассчитает данные о симплексах и дырах, можно приступить к построению баркода (рисунок 6).

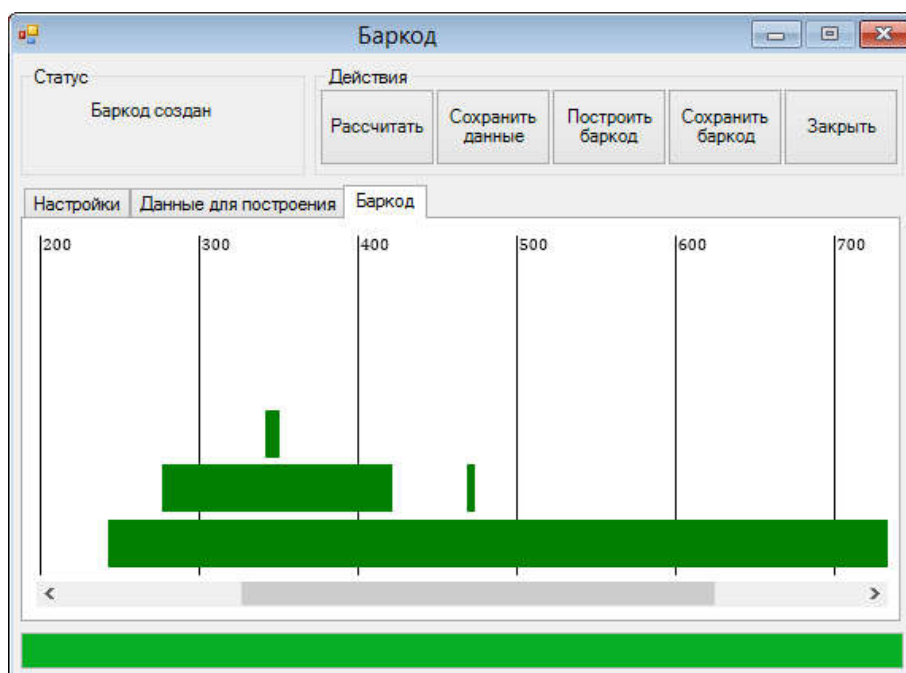


Рис. 6. Баркод, отображающий структуры расположения точечных объектов на карте

По оси  $Y$  отображается количество дыр, по оси  $X$  – расстояния, на которых осуществлялся поиск дыр.

Таким образом был получен баркод, который отображает структуру расположения зданий на рисунке 5. На рисунке имеются как стойкие особенности, так и особенности, проявившиеся на малых расстояниях, которые представляют собой «шум», который в дальнейшем не рассматривается. Стойкими особенностями являются две структуры объектов  $K1$  и  $K2$ , которые при увеличении расстояния  $d$  объединяются в общую структуру. Структуры  $K1$  и  $K2$  наблюдаются на расстоянии от 200 до 420. При  $d$  больше 420 остаётся одна общая структура объектов.

Кроме построения баркода программный модуль позволяет просматривать и сохранять информацию о полученных симплексах и расстояниях, на которых они были образованы, о составах дыр как на определённых расстояниях, так и полученных в результате анализа. Имеются настройки, в которых можно указать слой, на котором должен осуществляться поиск симплексов и дыр, есть возможность включить демонстрацию процесса составления симплексов, таким образом, есть возможность просмотреть, как и в каком порядке составляются связи и образуются искомые

симплексы. Также с помощью полученного баркода можно просматривать, состав дыр на интересующих расстояниях.

### **Заключение**

В статье рассматривается метод персистентной гомологии для пространственных объектов. Это относительно новый инструмент вычислительной топологии, используемый для выявления особенностей структур, образованных объектами.

Были разработаны алгоритмы поиска двумерных симплексов и поиска дыр.

Начальным этапом поиска двумерных симплексов является вычисление расстояний между объектами с помощью их координат, на основе чего осуществляется установление связей между объектами. После завершения работы со связями структуры объектов анализируются на наличие произвольных треугольников - симплексов. Результат поиска – список симплексов  $S$ .

Алгоритм поиска дыр использует информацию, полученную в результате выполнения первого алгоритма. Рассматриваются структуры, составляемые симплексами, и расстояния, на которых эти структуры были сформированы. Результатом являются полученные данные, используемые для построения баркода: расстояния и обнаруженные структуры.

Заключительный этап – построение баркода, отражающего особенности структур объектов на карте. Далее баркод может быть проанализирован на наличие «шума» и выявление стойких характеристик объектов.

Разработанные алгоритмы позволяют собирать всю информацию для анализа топологических особенностей расположения пространственных объектов на карте. Полученные алгоритмы будут использоваться для сравнения расположения объектов на различных картах для выявления различий и сходств. Кроме того, разработанный метод может также использоваться для кластеризации объектов, основываясь не только на расстоянии между рассматриваемыми объектами, но и на составляемых ими структурах.

## Литература

1. Ereemeev S.V., Andrianov D.E., Komkov V.A. Comparison of Urban Areas Based on Database of Topological Relationships in Geoinformational Systems // Pattern Recognition and Image Analysis. – 2015. – Vol. 25(2). – P. 314–320.
2. Еремеев С.В., Андрианов Д.Е., Титов Д.В. Метод представления информации о топологии карты в структуре идентификаторов пространственных объектов // Известия высших учебных заведений. Приборостроение. – 2015. – №2. – С. 99-103.
3. Еремеев С.В., Андрианов Д.Е., Купцов К.В. Метод идентификации непересекающихся пространственных объектов на основе структурных элементов // Телекоммуникации. – 2016. – №11. – С. 39-44.
4. Макаренко Н.Г., Уртьев Ф.А., Князева И.С. Распознавание текстур на цифровых изображениях методами вычислительной топологии // Современные проблемы дистанционного зондирования Земли из космоса. – 2015. – Т. 12, № 1. – С. 131-144.
5. Еремеев С.В., Андрианов Д.Е., Комков В.А. Алгоритмы формирования графовой модели городской территории в ГИС // Геоинформатика. – 2013. – №4. – С. 19-24.
6. Edelsbrunner H., Harer J. Computational Topology: An Introduction // American Mathematical Society, 2009. – 241 p.
7. Carlsson G. Topology and data // American Mathematical Society. – 2009.– Vol. 46(2). – P. 255-308.
8. Ghrist R. Barcodes: the persistent topology of data // American Mathematical Society. – 2008.– Vol. 45(1). – P. 61-75.
9. Xia K. Persistent Homology for The Quantitative Prediction of Fullerene // Journal of computational chemistry. – 2015.– Vol. 36(6). – P. 408-422.