

# Joint simplification of various types spatial objects while preserving topological relationships

O. P. Yakimova<sup>1</sup>, D. M. Murin<sup>1</sup>, V. G. Gorshkov<sup>1</sup>

DOI: [10.18255/1818-1015-2023-4-340-353](https://doi.org/10.18255/1818-1015-2023-4-340-353)

<sup>1</sup>P.G. Demidov Yaroslavl State University, 14 Sovetskaya str., Yaroslavl 150003, Russia.

MSC2020: 68W99

Research article

Full text in Russian

Received November 13, 2023

After revision November 21, 2023

Accepted November 22, 2023

Cartographic generalization includes the process of graphically reducing information from reality or larger scaled maps to display only the information that is necessary at a specific scale. After generalization, maps can show the main things and essential characteristics. The scale, use and theme of maps, geographical features of cartographic regions and graphic dimensions of symbols are the main factors affecting cartographic generalization. Geometric simplification is one of the core components of cartographic generalization. The topological relations of spatial features also play an important role in spatial data organization, queries, updates, and quality control. Various map transformations can change the relationships between features, especially since it is common practice to simplify each type of spatial feature independently (first administrative boundaries, then road network, settlements, hydrographic network, etc.). In order to detect the spatial conflicts a refined description of topological relationships is needed. Considering coverings and mesh structures allows us to reduce the more general problem of topological conflict correction to the problem of resolving topological conflicts within a single mesh cell. In this paper, a new simplification algorithm is proposed. Its peculiarity is the joint simplification of a set of spatial objects of different types while preserving their topological relations. The proposed algorithm has a single parameter – the minimum map detail size (usually it is equal to one millimeter in the target map scale). The first step of the algorithm is the construction of a special mesh data structure. On its basis for each spatial object a sequence of cells is formed, to which points of this object belong. If a cell contains points of only one object, its geometric simplification is performed within the bounding cell using the sleeve-fitting algorithm. If a cell contains points of several objects, geometric simplification is performed using a special topology-preserving procedure.

**Keywords:** simplification algorithm; topological relationships; mesh data structure; spatial data; consistent cartographic generalization

## INFORMATION ABOUT THE AUTHORS

Olga P. Yakimova | [orcid.org/0000-0001-8816-2802](https://orcid.org/0000-0001-8816-2802) . E-mail: [olga\\_pavl02@mail.ru](mailto:olga_pavl02@mail.ru)  
corresponding author | Associate professor, Ph.D. in Mathematics.

Dmitriy M. Murin | [orcid.org/0000-0002-8068-0784](https://orcid.org/0000-0002-8068-0784). E-mail: [d.murin@uniyar.ac.ru](mailto:d.murin@uniyar.ac.ru)  
Associate professor, Ph.D. in Mathematics.

Vladislav G. Gorshkov | [orcid.org/0000-0002-2424-3942](https://orcid.org/0000-0002-2424-3942). E-mail: [gorshkov.vladik@mail.ru](mailto:gorshkov.vladik@mail.ru)  
Programmer.

**Funding:** P.G. Demidov Yaroslavl State University, project No. GM-2023-03.

**For citation:** O. P. Yakimova, D. M. Murin, and V. G. Gorshkov, “Joint simplification of various types spatial objects while preserving topological relationships”, *Modeling and analysis of information systems*, vol. 30, no. 4, pp. 340-353, 2023.

## Совместное упрощение пространственных объектов различного типа с сохранением топологических отношений

О. П. Якимова<sup>1</sup>, Д. М. Мурин<sup>1</sup>, В. Г. Горшков<sup>1</sup>

DOI: [10.18255/1818-1015-2023-4-340-353](https://doi.org/10.18255/1818-1015-2023-4-340-353)

<sup>1</sup>Ярославский государственный университет им. П.Г. Демидова, ул. Советская, д. 14, г. Ярославль, 150003 Россия.

УДК 004.67+528.91

Научная статья

Полный текст на русском языке

Получена 13 ноября 2023 г.

После доработки 21 ноября 2023 г.

Принята к публикации 22 ноября 2023 г.

Картографическая генерализация включает выбор отображаемых на карте объектов и явлений и их упрощение (обобщение) с сохранением основных типичных черт и характерных особенностей, а также взаимосвязей в соответствии с критериями, задаваемыми в запросе пользователем, в том числе решаемой задачей и масштабом отображаемой карты. Различные преобразования карт могут изменить отношения между объектами, тем более что общепринятой является практика упрощения каждого типа пространственных объектов независимо (сначала административные границы, потом дорожная сеть, населенные пункты, гидрографическая сеть и т. д.). Разрешение топологических конфликтов — одна из важнейших задач цифровой генерализации карт, решению которой уделяется особое внимание с начала исследований в этой области. Рассмотрение покрытий и сеточных структур позволяет свести более общую проблему коррекции топологических конфликтов к задаче разрешения топологических конфликтов внутри одной ячейки сетки.

В настоящей работе предлагается новый алгоритм геометрического упрощения. Его особенностью является совместное упрощение множества пространственных объектов различного типа с сохранением их топологических отношений. Предлагаемый алгоритм имеет единственный параметр минимальный размер отображаемой на карте детали (обычно он равен одному миллиметру в целевом масштабе карты). Первым шагом алгоритма является построение специальной сеточной структуры данных. На ее основе для каждого пространственного объекта формируется последовательность ячеек, которым принадлежат точки данного объекта. Если в ячейке находятся точки только одного объекта, то его геометрическое упрощение происходит в рамках ограничивающей ячейки по алгоритму *sleeve-fitting*. Если в ячейке содержатся точки нескольких объектов, то геометрическое упрощение осуществляется с помощью специальной, сохраняющей топологию, процедуры.

**Ключевые слова:** алгоритм упрощения; топологические отношения; сеточная структура данных; пространственные данные; согласованная картографическая генерализация

### ИНФОРМАЦИЯ ОБ АВТОРАХ

Ольга Павловна Якимова | [orcid.org/0000-0001-8816-2802](https://orcid.org/0000-0001-8816-2802) . E-mail: [olga\\_pavl02@mail.ru](mailto:olga_pavl02@mail.ru)  
автор для корреспонденции | доцент, канд. физ.-мат. наук.

Дмитрий Михайлович Мурин | [orcid.org/0000-0002-8068-0784](https://orcid.org/0000-0002-8068-0784). E-mail: [d.murin@uniyar.ac.ru](mailto:d.murin@uniyar.ac.ru)  
| доцент, канд. физ.-мат. наук.

Владислав Геннадьевич Горшков | [orcid.org/0000-0002-2424-3942](https://orcid.org/0000-0002-2424-3942). E-mail: [gorshkov.vladik@mail.ru](mailto:gorshkov.vladik@mail.ru)  
| программист.

**Финансирование:** Ярославский государственный университет им. П.Г. Демидова, проект № GM-2023-03.

**Для цитирования:** О. П. Yakimova, D. M. Murin, and V. G. Gorshkov, "Joint simplification of various types spatial objects while preserving topological relationships", *Modeling and analysis of information systems*, vol. 30, no. 4, pp. 340-353, 2023.

## Введение

Геоинформационные системы (ГИС) представляют собой совокупность содержащейся в базах данных картографической информации и обеспечивающих ее обработку информационных технологий и технических средств. Такие системы обеспечивают сбор, актуализацию, хранение, анализ и графическое отображение (представление) картографических и иных, связанных с ними, данных. Конечными пользователями геоинформационных систем могут выступать как физические лица, так и иные информационные системы и процессы. Технические средства и информационные технологии, входящие в состав ГИС можно разделить на серверные, реализующие основной функционал ГИС, и пользовательские. Используемые физическими лицами технические средства и информационные технологии в первую очередь рассчитаны на отображение (визуализацию) данных, получаемых от серверной части ГИС.

Число (в том числе прикладных) задач, решаемых с помощью ГИС, непрерывно растет и охватывает различные сферы экономики и жизнедеятельности, например, государственное и муниципальное управление, землеустройство, налогообложение, логистика, транспорт, сельское хозяйство, туризм, метеорология, геологоразведка и т. д. Для эффективного решения задач в каждой области представляют интерес свои данные. Например, для решения задач землеустройства наибольший интерес представляют границы земельных участков, а для управления транспортом — дорожные сети. При этом одни и те же данные могут иметь отличную ценность для различных областей. Например, для определения маршрута достаточно знать карту дорожной сети и возможные варианты путей, а для расчета стоимости ремонта дорожного полотна требуются детальные сведения о протяженности участка, текущем состоянии, типе дорожного покрытия, прогнозируемом транспортном потоке и т. д. Возможность оперативного решения задач с помощью картографических данных может быть обеспечена только в том случае, если выдаваемые серверной частью ГИС данные могут успешно отображаться (обработываться) на пользовательских устройствах. Большой объем информации, который может быть выдан серверной частью ГИС, не всегда с успехом может быть обработан клиентским устройством.

Цифровая (компьютерная) картографическая генерализация (от лат. *generalis* — общий, главный) — это процесс обработки картографических данных, в целях предоставления по запросу пользователя информации, которая может быть воспроизведена (обработана) на пользовательском устройстве, достаточной для обеспечения эффективного решения стоящих перед пользователем задач. Картографическая генерализация включает выбор отображаемых на карте объектов и явлений и их упрощение (обобщение) с сохранением основных типичных черт и характерных особенностей, а также взаимосвязей в соответствии с критериями, задаваемыми в запросе пользователем, в том числе решаемой задачей и масштабом отображаемой карты.

Одни из первых работ [1–3], связанные с цифровой генерализацией карт были опубликованы в открытой печати в конце 1950-х, начале 1960-х годов. Параллельно с этим в начале 1960-х годов начинают появляться первые геоинформационные системы, которые на начальном этапе являлись по сути структурами данных, содержащих информацию о географических объектах и их координатах. В таких системах использовалось растровое (ячейстое) представление информации. В конце 1960-х годов Бюро переписи США был разработан формат Geographic Base File, Dual Independent Map Encoding (GBF-DIME), который позволил учитывать топологические отношения между полилинейными и полигональными картографическими объектами. В этом формате впервые были проиндексированы узловые точки, а также присвоены идентификаторы некоторым объектам.

В 1970-х годах основные усилия исследователей были сосредоточены на разработке алгоритмов генерализации линейных и полилинейных объектов. Один из подходов к упрощению полилиний, до сих пор не потерявший актуальность, был независимо предложен Рамером [4], Дугласом и Пекером [5] в 1972 и 1973 годах соответственно. Идея данного алгоритма заключается в том,

чтобы по заданной полилинии построить другую, содержащую меньшее число вершин, однако сохраняющую некоторые характеристики исходной. Близость исходной и упрощенной полилиний определяется по максимальному расстоянию между ними.

Единого мнения о том существует ли «оптимальная» упрощенная линия для каждой заданной нет. Так в работе [6] приводятся аргументы в пользу того, что некоторые упрощенные вершины более точно передают форму исходной полилинии. Однако в работе [7] утверждается, что ни одно конкретное множество вершин не может быть единственно верным для представления полилинии. С нашей точки зрения, различные множества вершин могут быть более репрезентативны при решении различных задач, что, в том числе, должно стимулировать исследования в сфере адаптирующихся алгоритмов [8, 9].

В 1974 году в работе Р. Финкеля и Дж. Бетли [10] была предложена структура данных, с помощью которой можно эффективно решать задачу хранения и извлечения информации по ключам, подходящая для хранения картографических данных (более точно, данных на плоскости) — квадродерево (или дерево квадрантов). Кроме самой структуры квадродерева в работе также рассматривался ее оптимизированный вариант. Квадродерева нашли широкое применение в пространственных базах данных.

В начале 1980-х годов были исследованы алгоритмы генерализации для полигональных объектов. При этом уже в это время было осознано, что представление картографических данных в различных масштабах является одной из важнейших проблем вычислительной картографии. В 1982 году появляется первая коммерческая ГИС — ARC/INFO компании ERSI.

В 1984 году А. Гуттман в работе [11] предложил структуру данных, представляющую многомерное пространство в виде множества иерархически вложенных и, возможно, пересекающихся, прямоугольников (параллелепипедов), обладающую свойством сбалансированности, —  $R$ -дерево. При хранении картографических данных в такой структуре близко расположенные объекты должны попадать в один лист дерева. Каждый лист  $R$ -дерева при этом должен хранить данные, описывающие картографический объект и ограничивающий прямоугольник объекта (объектов).

Развитие ГИС породило вопрос об унификации форматов представления информации для хранения и обмена картографическими данными. В связи с чем в середине-конце 1980-х годов был разработан и принят стандарт стран Варшавского договора «Единая система классификации и кодирования картографической информации» (ЕСКККИ), а в 1991 году Digital Geographic Information Exchange Standard (DIGEST) — стандарт стран НАТО. При этом единого, всеми признаваемого стандарта представления геопространственных данных не существует до сих пор [12].

Различные преобразования карт, в том числе изменение масштаба или генерализация, могут изменить отношения между объектами. Например, могут измениться топологические отношения, отношения порядка или соотношения размеров [13]. Разрешение топологических конфликтов — одна из важнейших задач цифровой генерализации карт, решению которой уделяется особое внимание с начала исследований этой области [14, 15]. С 1990-х годов исследуются подходы к генерализации линейных картографических объектов на основе различных разбиений (покрытий) карты, например, триангуляции Делоне [16], квадратных [17] и гексагональных [18]. Рассмотрение покрытий и сеточных структур позволяет свести более общую проблему коррекции топологических конфликтов к задаче разрешения топологических конфликтов внутри одной ячейки сетки.

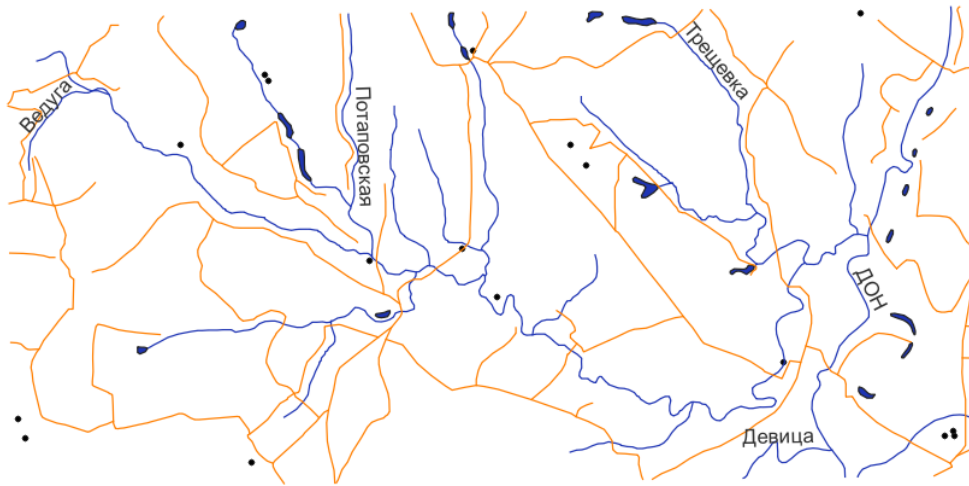
Следует отметить, что алгоритмы генерализации часто рассматривают объекты исключительно одной геометрии: либо полилинии, либо полигоны. Можно сказать, что общепринятой является классификация алгоритмов по этому признаку [19]. У такого подхода есть некоторые основания, поскольку в ГИС информация часто хранится «по слоям», причем в одном слое хранятся данные одной геометрической природы, например, дорожные сети или речные сети, представленные полилиниями, водоемы или здания, представленные полигонами. С одним слоем карты существенно

проще работать в силу относительно небольшого (по сравнению со всей картой) объема данных. Тем не менее, нередко могут возникать ситуации, при которых один географический объект при таком подходе оказывается сегментирован, а его элементы хранятся в разных слоях карты. Так происходит, например, в случае, если в течении реки находится водоем: озеро, болото и т. п. или на отдельном участке русло имеет существенно большую ширину по сравнению с соседними участками. В связи с вышеизложенным представляет интерес алгоритм, способный решать задачи по разрешению топологических конфликтов для объектов различного вида: и точек, и полилиний, и полигонов.

В настоящей работе предлагается новый алгоритм геометрического упрощения. Его особенностью является совместное упрощение множества пространственных объектов различного типа с сохранением их топологических отношений.

## 1. Алгоритм согласованной генерализации

Исходными данными для предлагаемого алгоритма является набор слоев картографических данных. Каждый слой содержит информацию о некотором виде пространственных объектов (границах, реках, озерах, дорогах, лесных массивах и т. д.) и имеет один тип геометрии, то есть содержит только полилинии, только полигоны или только точки. На рисунке 1 показан пример набора исходных данных.



**Fig. 1.** Input data example. Different line layers are shown by different types of polylines

**Рис. 1.** Пример исходных данных. Различные линейные слои показаны разными полилиниями

Предлагаемый алгоритм имеет единственный параметр – минимальный размер отображаемой на карте детали – *Minsize* (обычно он равен одному миллиметру в целевом масштабе карты). Основными шагами алгоритма являются:

1. *Построение сеточной структуры данных.* Все слои данных последовательно просматриваются и объекты каждого слоя заносятся в сеточную структуру данных.
2. *Упрощение.* Для каждого пространственного объекта формируется последовательность ячеек сеточной структуры данных, которым принадлежат точки данного объекта. Если в ячейке находится точки только одного объекта, то его геометрическое упрощение происходит в рамках ограничивающей ячейки по алгоритму «облегающего/обтягивающего рукава», предложенному в работе [20]. Если в ячейке содержатся точки нескольких объектов, то геометрическое упрощение осуществляется с помощью специальной, сохраняющей топологию, процедуры. Далее каждый шаг алгоритма рассматривается более подробно.

### 1.1. Построение сеточной структуры данных

Предлагаемая структура данных на основе квадродерева может содержать следующие типы объектов:

1. Точечные объекты.
2. Полилинейные объекты.
3. Полигональные объекты (тип полилинейных объектов, у которых начало и конец совпадают).

Каждый тип объекта представляет собой структуру данных, включающую:

- *ID* – идентификатор объекта.
- *Geometry* – геометрия объекта (*Point* (точка), *Polyline* (полилиния), *Polygon* (полигональный объект)).
- *Points* – список точек, которые составляют объект.
- *Path* – представляет собой маршрут по ячейкам сеточной структуры данных (проинициализирован \* – корневой ячейкой).

*Вход алгоритма:*

1. *Minsize* – минимальный размер различимой детали на карте.
2. Список всех картографических объектов.

*Стадия преобработки:*

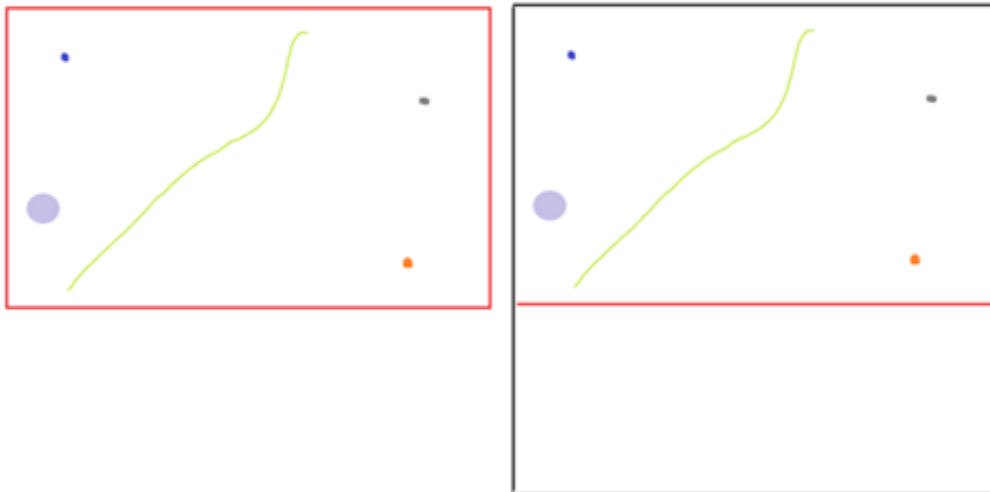
Находятся минимальные и максимальные значения по осям *X* и *Y* от охвата карты.

$$y_{max}, y_{min}, x_{max}, x_{min}.$$

Создается корневая ограничивающая квадратная ячейка, куда помещаются все картографические объекты. Сторона этой ячейки равна:

$$\max(y_{max} - y_{min}, x_{max} - x_{min}).$$

Левый нижний угол ячейки помещается в точку  $(x_{min}, y_{min})$ .



**Fig. 2.** Map coverage (left — before preprocessing, right — after)

**Рис. 2.** Охват карты (слева до преобработки, справа — после)

Таким образом, получаем охват карты, который представляет собой ограничивающий квадрат, в котором находятся все рассматриваемые объекты, с координатами левого нижнего, левого верхнего, правого верхнего, правого нижнего углов.

Тело алгоритма:

1. Создаем очередь  $Q$  на обработку ячеек.
2. Помещаем в  $Q$  корневую ячейку.
3. Пока очередь не пуста:
  - (a) Извлекаем из очереди  $Q$  ячейку на обработку, помещаем указатель на нее в переменную  $currentNode$ .
  - (b) Если в ячейке больше одного уникального объекта (то есть ячейке принадлежат два или более объекта с различными  $ID$ ) и сторона ячейки —  $side > Minsize$ , то:
    - i. Делим ячейку  $currentNode$  на четыре одинаковых квадрата с длиной стороны  $side/2$  (полученные ячейки обозначим  $node_1, node_2, node_3, node_4$ ).
    - ii. Для каждого объекта из  $currentNode$  определяем, в каких дочерних ячейках присутствуют его элементы, и корректируем его поле  $Path$ , заменяя  $currentNode$  на последовательность дочерних ячеек, по которым проходит данный объект.
    - iii. Помещаем ячейки  $node_1, node_2, node_3, node_4$  в  $Q$ .
  - (c) Иначе продолжаем алгоритм.

Каждый точечный объект лежит в какой-то определенной единственной ячейке сетки. Полилинейный объект сетка разбивает на сегменты, каждый из которых принадлежит определенной ячейке. Сегменты полигонального объекта также будут расположены в нескольких ячейках, в силу малого размера которых, сегмент полигонального объекта внутри ячейки можно рассматривать как полилинейный (считаем ячейку замкнутым множеством). Пример построенной сеточной структуры данных изображен на рисунке 3.

Каждая ячейка сетки может быть пустой, т. е. не содержать ни одного пространственного объекта, содержать один или несколько пространственных объектов.

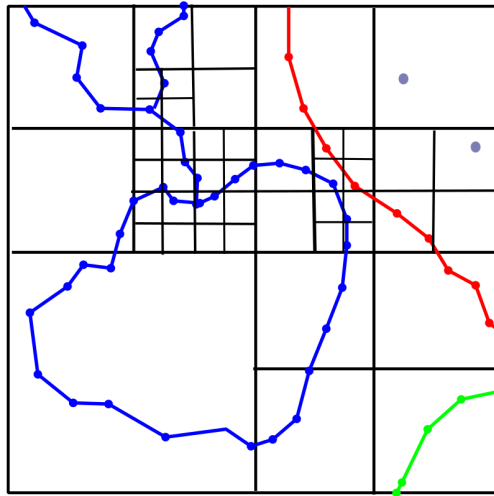


Fig. 3. Example of building a grid data structure

Рис. 3. Пример построения сеточной структуры данных

Рассмотрим подробнее шаг ii алгоритма. Дочерние ячейки имеют номера:

- 0 — левая нижняя ячейка,
- 1 — левая верхняя ячейка,
- 2 — правая верхняя ячейка,
- 3 — правая нижняя ячейка.

Будем рассматривать точки линейного или полигонального объекта в порядке их следования. Допустим, что первая точка  $S$  объекта, принадлежащая данной ячейке, находится в левом нижнем

углу, то есть в дочерней ячейке с номером 0. Если такая точка находится на границе сетки, то рассматривается следующая за ней, до ситуации, при которой дочерняя ячейка может быть однозначно определена. Обозначим через  $F$  следующую точку объекта.

На рисунке 4, показаны возможные варианты расположения вершин  $S$  и  $F$  при условии что  $S$  лежит в нулевой дочерней ячейке. Очевидно, что для дочерних ячеек с номерами 1 – 3 рассуждения будут аналогичными.

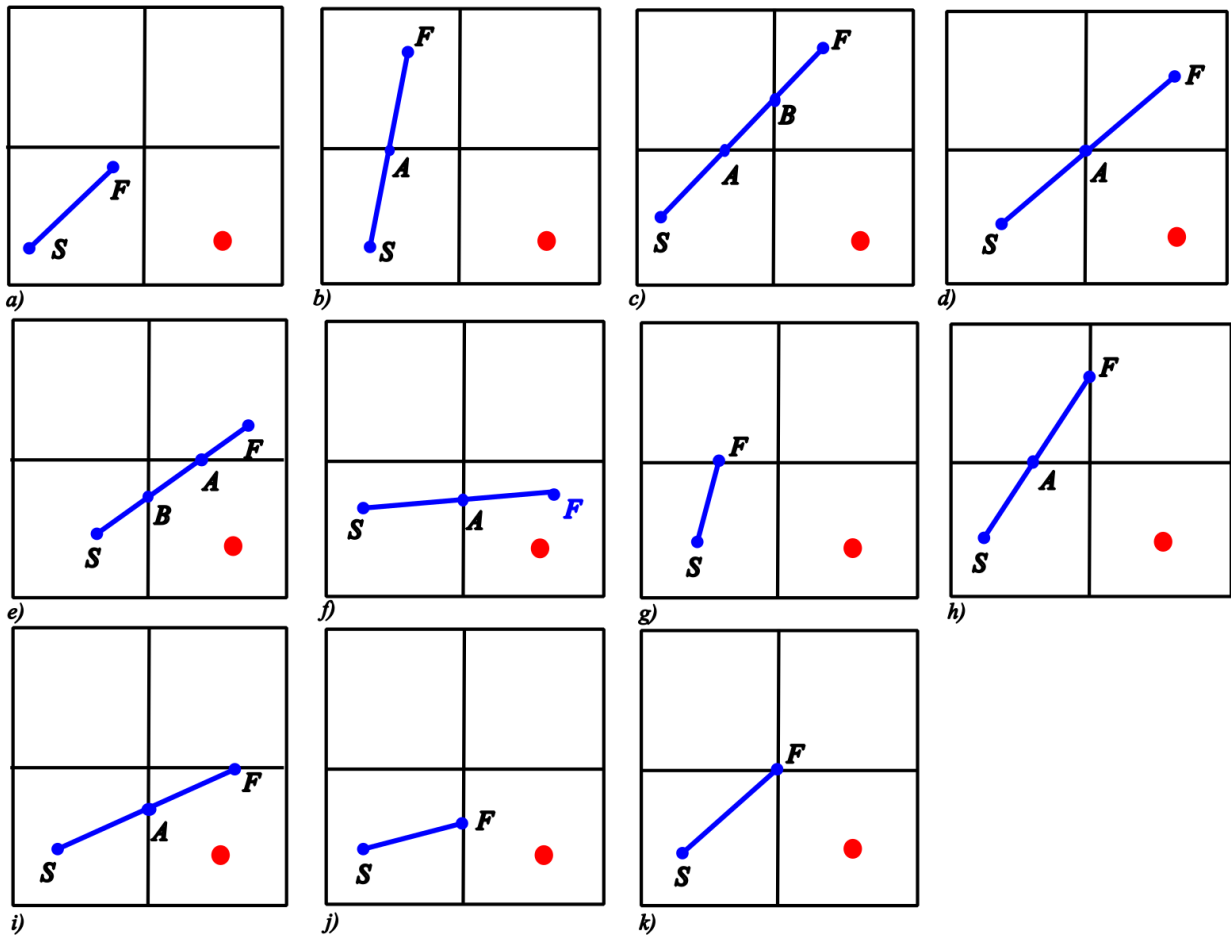


Fig. 4. Options for the location of polyline vertices

Рис. 4. Варианты расположения вершин полилинии

Допустим, что два объекта находятся в одной ячейке сетки (на рисунке 4 — линейный и точечный), и тогда необходимо разбить текущую ячейку на четыре дочерних.

Для случаев a), g), j), к) нам не нужно искать точки пересечения с ячейками сетки, так как вершина  $F$  (как и вершина  $S$ ) лежит в левой нижней ячейке.

Для остальных ситуаций нужно искать точки пересечения с сеткой для того, чтобы в дальнейшем учитывать их топологические отношения, если потребуется делить дочернюю ячейку на еще более мелкие. Обозначим через  $A$  и  $B$  точки пересечения с дочерними ячейками сетки.

Для случаев b) и h) точка пересечения  $A$  находится пересечением отрезка  $SF$  с горизонтальным «терминатором», а для случаев f), i) — с вертикальным «терминатором». Каждая из этих ситуаций однозначно идентифицируется положением точки  $F$ .



Ситуации с), d) и e) несколько сложнее. Для них требуется определить каким образом отрезок  $SF$  проходит по ячейкам сетки. Во всех этих случаях точка  $F$  лежит в правом верхнем углу, следовательно, у нас есть следующие варианты прохождения отрезка  $SF$ :

- Отрезок  $SF$  проходит через левую верхнюю ячейку – с).
- Отрезок  $SF$  проходит через точку пересечения терминаторов – d).
- Отрезок  $SF$  проходит через правую нижнюю ячейку – e).

Для идентификации одной из этих ситуаций (с, d, e) используем следующий, разработанный нами, алгоритм:

1. Находим точку пересечения отрезка  $SF$  с горизонтальным терминатором. Обозначим эту точку через  $A$ .
2. Если точка  $A$  лежит левее, чем точка пересечения терминаторов, то это ситуация с). Находим точку пересечения отрезка  $SF$  с вертикальным терминатором. Обозначим эту точку через  $B$ . В соответствующие ячейки помещаем точки пересечения полилинии с терминаторами (точку  $A$  помещаем в левую нижнюю ячейку,  $A$  и  $B$  – в левую верхнюю ячейку,  $B$  – в правую верхнюю ячейку).
3. Иначе, если точка  $A$  совпадает с точкой пересечения терминаторов, то это ситуация d). В соответствующие ячейки (левую нижнюю, правую верхнюю) помещаем точку  $A$  (пересечения терминаторов).
4. Иначе (точка  $A$  находится правее, чем точка пересечения терминаторов) – это ситуация e). Находим точку пересечения отрезка  $SF$  с вертикальным терминатором. Обозначим эту точку через  $B$ . В соответствующие ячейки помещаем точки пересечения полилинии с терминаторами (точку  $A$  помещаем в правую верхнюю,  $A$  и  $B$  – в правую нижнюю,  $B$  – в левую нижнюю ячейку).

*Результат работы алгоритма:*

1. Построенное квадродерево.
2. Для каждого объекта создан свой маршрут *Path* по ячейкам сетки.

С помощью разработанной структуры данных можно ускорить работу ГИС при выполнении запросов вида: «найти ближайший объект по отношению к данному» (наивный алгоритм – перебор всех картографических объектов, в то время как достаточно перебрать все объекты в ячейке, в которой находится данный объект и все ближайшие ячейки по отношению к той, в которой находится данный объект), «получить информацию о топологических отношениях объектов» и др.

## 1.2. Упрощение

Как результат работы первого этапа алгоритма для каждого полилинейного или полигонального объекта получаем последовательность ячеек сеточной структуры данных, по которым он проходит. Согласно построению описанной выше структуры только в ячейках минимального размера может находиться более одного объекта. Внутри любых других ячеек производится геометрическое упрощение полилинейного объекта или границы полигонального объекта с помощью алгоритма «обтягивающего рукава» (sleeve-fitting). Идея его работы заключается в методе постепенного «обтягивания» участков упрощаемой полилинии «рукавами» заданной ширины  $d$ . Алгоритм итеративно проходит по точкам и выполняет построение рукава. Если какая-то точка не попадает в заданные рамки, то в ней рукав изламывается. Часть полилинии, попавшей внутрь рукава, будет упрощена до отрезка, соединяющего первую и последнюю точки, попавшие в рукав (см. рис. 5).

Параметр алгоритма sleeve-fitting  $d$  задается равным половине минимального размера различимой детали *Minsize*, который установлен при старте работы предлагаемого алгоритма. Такое соотношение позволяет равномерно сокращать число точек в больших и маленьких ячейках.

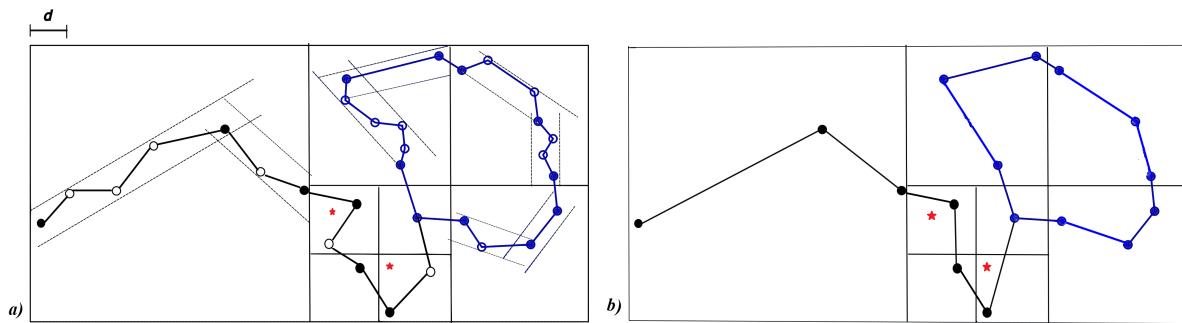


Fig. 5. Simplification process

Рис. 5. Процедура упрощения

В ячейках минимального размера может находиться несколько картографических объектов, причем разной геометрии. Следует отметить, что полигональный объект целиком в маленькой ячейке находиться не может, так как первым этапом картографической генерализации является отбор, где объекты слишком мелкие для целевого масштаба удаляются. Отсюда можно считать, что в ячейке минимального размера могут быть либо точечные, либо полилинейные объекты, представляющие собой как границы полигонов, так и собственно полилинии.

Если в ячейке минимального размера находится один объект, то для его представления выбирается одна точка, что оправданно малостью ячейки в целевом масштабе карты. В случае наличия нескольких объектов для каждой пары строится матрица девяти пересечений [21, 22], позволяющая решить вопрос об их взаимном расположении.

Если полилинейные объекты касаются или пересекаются, то в упрощенном виде сохраняется их точка касания (пересечения). Если совпадают, то для обоих объектов сохраняется линия, соединяющая точки на границах ячейки, остальные точки удаляются. Для непересекающихся точечных и полилинейных объектов отслеживается их взаимное расположение, чтобы точка оставалась в той же полуплоскости относительно полилинии.

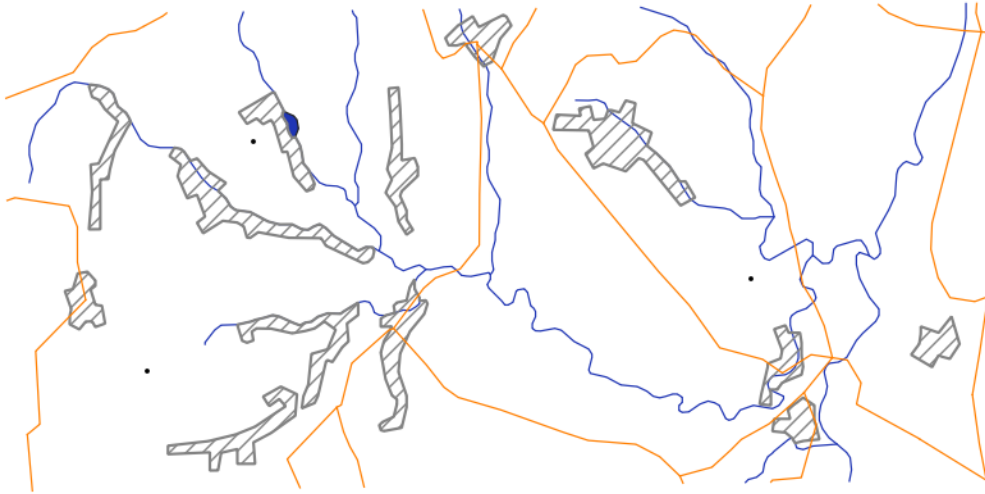
Пример упрощения показан на рисунке 5.

## 2. Эксперименты и результаты

Тестирование разработанного алгоритма производилось на фрагментах карты Российской Федерации масштабов 1:500000 и 1:1000000. Все фрагменты карты проходили предобработку. К сожалению, качество цифровых пространственных данных, доступных в открытых источниках, оставляет желать лучшего.

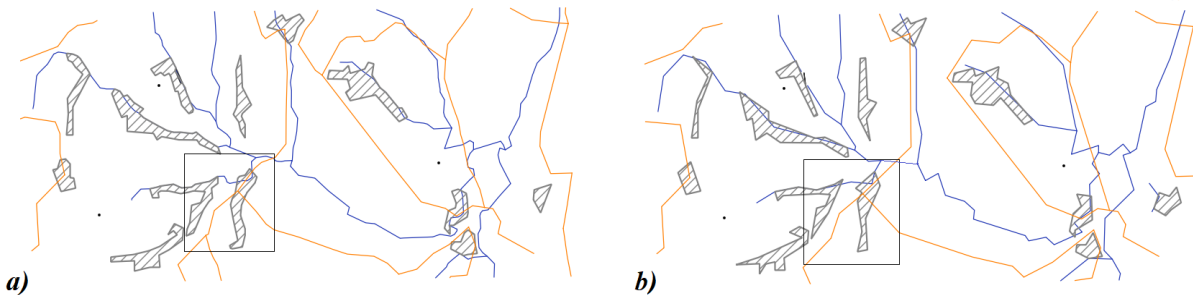
Например, при оцифровке реки полилиния разбивается на отрезки от впадения одного притока до впадения другого. Каждый отрезок имеет свой числовой идентификатор, причем его значения для последовательных участков реки существенно отличаются. Объединить реку в единый объект (с одним и тем же идентификатором) можно по ее имени, а если имя не указано, то только вручную, средствами ГИС, непосредственно выделяя попарно те части объекта, которые надо соединить с одним и тем же идентификатором. Для построения структуры данных, на основе которой работает предлагаемый алгоритм согласованного упрощения, необходимо, чтобы один объект имел один и тот же идентификатор. Для оценки точности пространственного позиционирования будет использоваться модифицированное расстояние Хаусдорфа (МНД), введенное и исследованное в работе [23]. Оно определяется следующим образом. Пусть  $A = \{a_1, a_2, \dots, a_n\}$  и  $B = \{b_1, b_2, \dots, b_m\}$  — два множества точек, а  $d(x, y)$  — евклидово расстояние между двумя точками. Среднее расстояние между  $A$  и  $B$  может быть вычислено по формуле:

$$\bar{d}(A, B) = \frac{1}{n} \sum_{a_i \in A} d(a_i, B),$$



**Fig. 6.** A fragment of a map of the Voronezh region at a scale of 1:1000000

**Рис. 6.** Фрагмент карты Воронежской области масштаба 1:1000000



**Fig. 7.** The result of simplification by the proposed algorithm a) and the sleeve-fitting algorithm b)

**Рис. 7.** Результат упрощения предлагаемым алгоритмом a) и алгоритмом sleeve-fitting b)

где  $d(a_i, B) = \min_{b_j \in B} d(a_i, b_j)$ . С помощью симметричной формулы вычисляется среднее расстояние между  $B$  и  $A$ :

$$\bar{d}(B, A) = \frac{1}{m} \sum_{b_j \in B} d(b_j, A),$$

где  $d(b_j, A) = \min_{a_i \in A} d(b_j, a_i)$ . Тогда модифицированное расстояние Хаусдорфа будет вычисляться так:

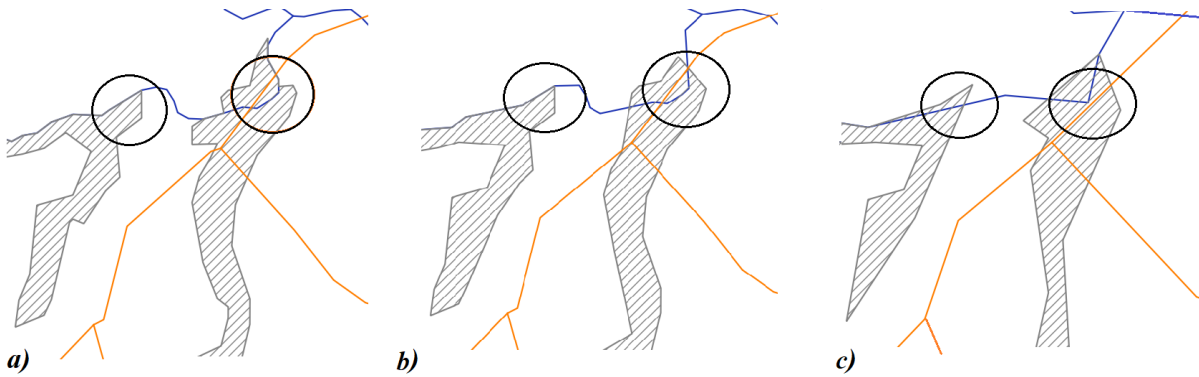
$$MHD(A, B) = \max(\bar{d}(A, B), \bar{d}(B, A)).$$

Чем меньше MHD, тем ближе исходная полилиния к упрощенной. Использование среднего значения расстояния позволяет снизить влияние выбросов — единичных точек, сильно удаленных от исходного положения.

На рисунке 6 представлен фрагмент карты Воронежской области, включающий в себя реки, озера, дороги, границы населенных пунктов и уникальные растительные объекты, которые отображаются черными точками. Этот фрагмент был упрощен алгоритмом sleeve-fitting [20] (каждый слой отдельно) и предлагаемым алгоритмом согласованной генерализации.

Результаты упрощения приведены на рисунке 7.

Выделенный фрагмент был вынесен на рисунок 8. Очевидно, что при упрощении предлагаемым алгоритмом сохраняются правильные топологические отношения картографических объектов.



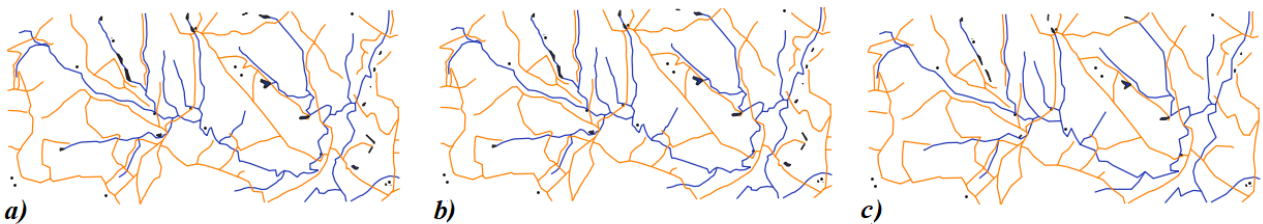
**Fig. 8.** Input data a), after simplification by the proposed algorithm b) and by the sleeve-fitting algorithm c)

**Рис. 8.** Исходные данные a), после упрощения предлагаемым алгоритмом b) и алгоритмом sleeve-fitting c)

**Table 1.** Modified Hausdorff distance between original and simplified objects

**Таблица 1.** Модифицированное расстояние Хаусдорфа между исходными и упрощенными объектами

Объекты	Исходное количество точек	Алгоритм согласованной генерализации		Алгоритм sleeve-fitting	
		Количество точек	MHD	Количество точек	MHD
<b>Фрагмент карты масштаба 1: 1000000</b>					
Реки	397	219	5	98	17
Водоемы	14	6	37	3	34
Границы населенных пунктов	335	231	14	129	16
Дороги	165	150	7	96	29
<b>Фрагмент карты масштаба 1:500000</b>					
Реки	1758	1018	4	382	8
Водоемы	614	324	15	176	21
Трубопроводы	120	112	1	55	5
Железные дороги	392	277	3	77	8



**Fig. 9.** Generalization results with the algorithm parameter a) 1000 m b) 1500 m c) 2000 m

**Рис. 9.** Результаты генерализации с параметром алгоритма a) 1000 м b) 1500 м c) 2000 м

В таблице (см. таблицу 1) приведены значения модифицированного расстояния Хаусдорфа между исходными и упрощенными объектами. Объекты, генерализованные с помощью предлагаемого алгоритма, ближе к исходным данным, но при этом разница количества точек между исходным

данными и результатом работы нашего алгоритма существенно меньше, чем аналогичная разница у алгоритма *sleeve-fitting*.

На рисунке 9 показано последовательное упрощение исходных данных формата 1:500000, представленных на рисунке 1, с параметром алгоритма равным 1 000 м, 1 500 м и 2 000 м.

## Выводы

В работе предложен оригинальный подход к генерализации всей совокупности картографических объектов, сохраняющий их топологические отношения. В качестве перспектив дальнейших исследований следует назвать применение разработанной структуры данных для операций отбора и обобщения при мультимасштабном картографировании, а также всестороннее тестирование и, возможно, модификация предлагаемого алгоритма для упрощения карт крупных масштабов.

## References

- [1] W. Tobler, *Numerical map generalization*. Department of Geography, University of Michigan Ann Arbour, MI, USA, 1966, 78 pp.
- [2] F. Töpfer and W. Pillewizer, “The principles of selection”, *Cartographic Journal*, vol. 3, no. 1, pp. 10–16, 1966.
- [3] J. D. Perkal, “Proba obiektywnej generalizacji”, *Geodezia I Kartografia*, vol. 7, no. 2, pp. 130–142, 1958, in Polish.
- [4] U. Ramer, “An iterative procedure for the polygonal approximation of plane curves”, *Computer Graphics and Image Processing*, vol. 1, no. 3, pp. 244–256, 1972.
- [5] D. H. Douglas and T. K. Peucker, “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature”, *Cartographica: the international journal for geographic information and geovisualization*, vol. 10, no. 2, pp. 112–122, 1973.
- [6] J. S. Marino, “Identification of characteristic points along naturally occurring lines. an empirical study”, *The Canadian cartographer Toronto*, vol. 16, no. 1, 1979.
- [7] G. Dutton, “Scale, sinuosity, and point selection in digital line generalization”, *Cartography and Geographic Information Science*, vol. 26, no. 1, pp. 33–54, 1999.
- [8] A. Chehreghan and R. Ali Abbaspour, “Estimation of empirical parameters in matching of linear vector datasets: An optimization approach”, *Model. Earth Syst. Environ*, pp. 1029–1043, 3 2017.
- [9] A. Chehreghan and R. Ali Abbaspour, “A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm”, *Cartography and Geographic Information Science*, vol. 45, pp. 255–269, 3 2018.
- [10] R. A. Finkel and J. L. Bentley, “Quad trees a data structure for retrieval on composite keys”, *Acta informatica*, vol. 4, pp. 1–9, 1974.
- [11] A. Guttman, “R-trees: A dynamic index structure for spatial searching”, in *Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, 1984, pp. 47–57.
- [12] U. A. Kravchenko, *Information geomodeling: The problem of data and knowledge representation*, Abstract of the dissertation for the degree of Doctor of Technical Sciences. In Russian, 2013.
- [13] G. Dettori and E. Puppo, “How generalization interacts with the topological and metric structure of maps”, in *Proceedings of the 7th International Symposium on Spatial Data Handling*, 1996, pp. 559–570.
- [14] D. Rhind, “Generalization and realism with automated cartographic system”, *Canadian Cartographer*, vol. 10, no. 1, pp. 51–62, 1973.

- [15] M. Monmonier, "Displacement in vector-and raster-mode graphics", *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 24, no. 4, pp. 25–36, 1987.
- [16] P. M. Van Der Poorten and C. B. Jones, "Characterisation and generalisation of cartographic lines using delaunay triangulation", *International Journal of Geographical Information Science*, vol. 16, no. 8, pp. 773–794, 2002.
- [17] Z. Li and S. Openshaw, "Algorithms for automated line generalization<sup>1</sup> based on a natural principle of objective generalization", *International journal of geographical information systems*, vol. 6, no. 5, pp. 373–389, 1992.
- [18] P. Raposo, "Scale-specific automated line simplification by vertex clustering on a hexagonal tessellation", *Cartography and Geographic Information Science*, vol. 40, no. 5, pp. 427–443, 2013.
- [19] L. Zhilin, *Algorithmic Foundation of Multi-Scale Spatial Representation*. Taylor & Francis Group, LLC, 2007, 282 pp.
- [20] Z. Zhao and A. Saalfeld, "Linear-time sleeve-fitting polyline simplification algorithms", in *Proceedings of AutoCarto*, vol. 13, 1997, pp. 214–223.
- [21] M. Egenhofer and J. Herring, "Categorizing binary topological relations between regions, lines and points in geographic databases, the 9-intersection: Formalism and its use for natural language spatial predicates", *Santa Barbara CA National Center for Geographic Information and Analysis Technical Report*, vol. 94, pp. 1–28, 1990.
- [22] V. G. Gorshkov, D. M. Murin, and O. P. Yakimova, "Research of models of topological relations of spatial objects", *Modelirovanie i Analiz Informatsionnykh Sistem*, vol. 29, no. 3, pp. 154–165, 2022, in Russian.
- [23] M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching", in *Proceedings of 12th international conference on pattern recognition*, vol. 1, 1994, pp. 566–568.