

АВТОМАТИЗИРОВАННАЯ ГЕНЕРАЛИЗАЦИЯ

АВТОМАТИЗИРОВАННАЯ ГЕНЕРАЛИЗАЦИЯ

Генерализация — неотъемлемое свойство всех картографических изображений, даже самых крупномасштабных.

Она проявляется в обобщении качественных и количественных характеристик объектов, замене индивидуальных понятий собирательными, отвлечении от частных и деталей для показа главных черт.

При этом генерализация ведет не только к исключению части информации, имеющейся на исходном изображении, но и к появлению качественно новой информации на генерализованной карте. По мере генерализации все отчетливее проступают наиболее важные черты объекта, ведущие закономерности, главные взаимосвязи, выделяются геосистемы все более крупного ранга.

При обычном картографировании на бумажном носителе качество генерализации прежде всего зависит от понимания картографом содержательной (географической, геологической и т. п.) сущности изображаемых объектов и явлений, умения отразить главные, типичные их особенности.

Стремление к автоматизации процессов картографической генерализации отмечалось уже на первых этапах применения компьютеров в картографии.

С этой точки зрения *генерализация* — это группа методов, позволяющая сохранить объем информации даже при уменьшении объема данных. Например, при сокращении числа точек на линии, остающиеся должны быть выбраны так, чтобы внешний вид линии не изменился.

Геоинформатика. Толковый словарь основных терминов (1999) дает следующее определение: ***автоматическая, или алгоритмическая генерализация*** – формализованный отбор, сглаживание или фильтрация изображения в соответствии с заданными алгоритмами и формальными критериями.

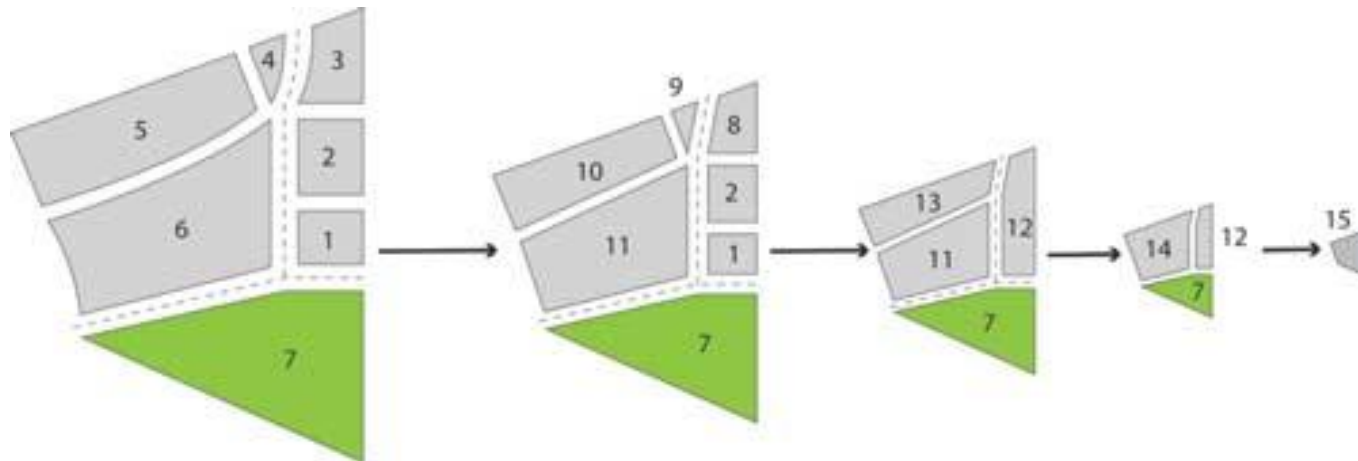
Генерализация пространственных данных — обобщение позиционных и атрибутивных данных о *пространственных объектах* в ГИС в автоматическом или интерактивном режиме с использованием **операторов генерализации**, или **генерализационных операторов** (generalization operators), их наборов или последовательностей, часть которых имеет соответствие в приемах и методах картографической *генерализации*.

Основные операторы генерализации:

- **упрощение** (simplification),
- **сглаживание** (smoothing),
- **утонышение линий** (line thinning),
- **разрядка линий**, т. е. устранение избыточных промежуточных точек в цифровой записи линий (line weeding),
- **отбор** (reselection),
- **переклассификация** (reclassification),
- **агрегирование** (aggregation), в частности, объединение смежных полигонов с уничтожением границ между ними (polygon dissolving/merging),
- **слияние** (amalgamation),
- **маскирование** (masking),
- **прерывание линий** (omissing),
- **утрирование размера или формы** (exaggeration),
- **уменьшение мерности объектов, или свертка, коллапс** (collapse).

Операторы генерализации пространственных данных (ГПД) могут применяться глобально (к слою в целом) или локально (к фрагменту слоя, сегменту линии и т.п.), обслуживать чисто графические (позиционные) или структурные преобразования данных.

Вмешательство пользователя в процесс автоматической ГПД обычно преследует цель индикации и устранения графических конфликтов в отображениях однотипных и разнотипных объектов путем их **смещения**, или **перемещения**, уменьшения или устранения геометрических и топологических погрешностей, контроля целостности данных и ненарушенности связи позиционной и атрибутивной части данных.



Процесс автоматизированной генерализации.

На первых трех этапах происходит упрощение и объединение (слияние) объектов. На последнем этапе качественные характеристики обобщаются, вводится новая категория объектов.

СЕМАНТИЧЕСКАЯ И ГЕОМЕТРИЧЕСКАЯ ГЕНЕРАЛИЗАЦИЯ

В картографической генерализации выделяют две стороны: семантическую и геометрическую.

Семантическая сторона генерализации проявляется в обобщении содержания карты, качественных и количественных характеристик объектов.

Генерализация качественных характеристик происходит путем обобщения классификационных признаков объектов, агрегирования позиций в легенде с учетом существующей иерархической структуры географических данных, и выполняется с применением методов классификации одно- и многопараметрически определенных данных.

Обобщение количественных характеристик проявляется в укрупнении количественных градаций отображаемого явления, укрупнении шкал, в переходе от непрерывных шкал к ступенчатым, от равномерных — к неравномерным. Наиболее просто реализуется в автоматическом режиме цензовый отбор картографируемых объектов и явлений.

Геометрическая (пространственная) сторона генерализации связана в первую очередь с правилами отображения формы, размера и положения географических объектов в плоскости карты.

Она проявляется в обобщении геометрических очертаний объектов, спрямлении границ, отказе от мелких деталей, группировке контуров, связанной с обобщением их качественных и количественных характеристик.

Основные требования картографической генерализации накладывают жесткие ограничения на автоматизацию этой стороны процесса. Здесь не годятся методы формального (механического) сглаживания очертаний. Автоматическое распознавание иерархических структур в геометрических данных составляет основную задачу компьютерного распознавания образов, далекую от решения в общем случае.

Методы автоматизации обобщения данных зависят от способа представления — *растрового* или *векторного*, так как подходы к генерализации для них разные.

Генерализация растровых данных не модифицирует топологическую структуру карты, так как она зафиксирована на сетке пикселей, значения пересчитываются с учетом соседних пикселей. Алгоритмы оперируют с количественными и качественными характеристиками статистических или атрибутивных данных. Генерализация при космической съемке происходит автоматически.

Наиболее сложен процесс пространственной генерализации с использованием векторных данных. В векторном формате она выполняется с помощью процедур агрегирования (объединения) смежных объектов, которые меняют топологию объектов, требуют удаления смежных границ и изменения атрибутивных значений.

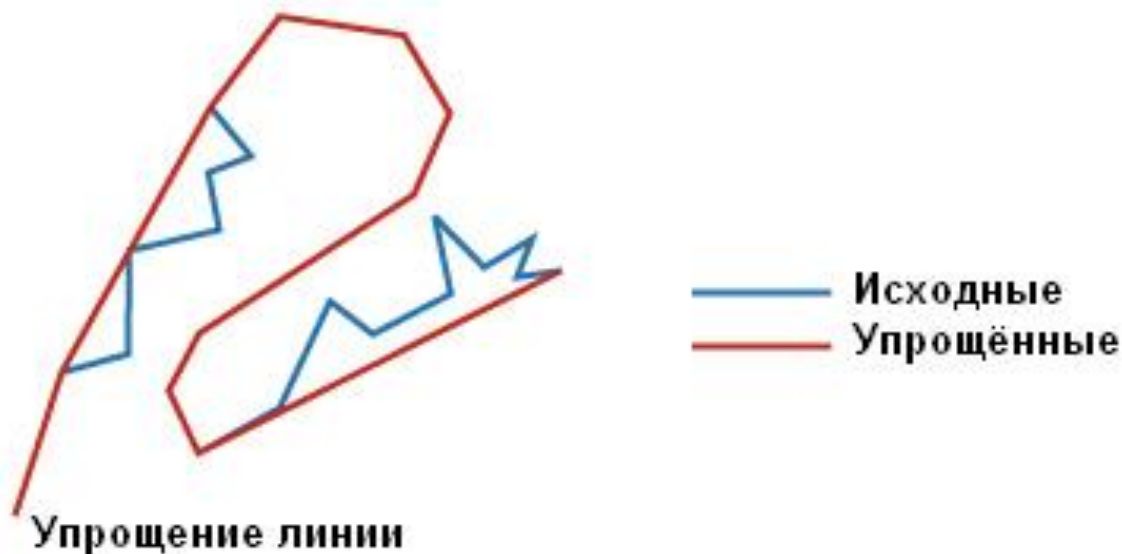
Элементы генерализации линий

Рассмотрим выполнение процесса автоматизированной генерализации на примере упрощения линий.

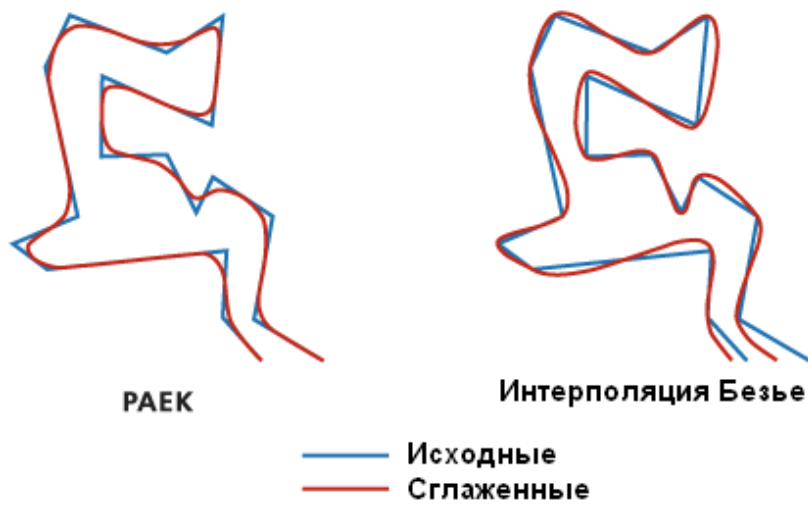
Генерализация линий — малая, но чрезвычайно важная часть проблемы генерализации в картографии, поскольку линиями на карте характеризуется широкий набор типов географических объектов. Более общая задача включает, например, генерализацию ареалов до уровня точки.

При генерализации линий происходит геометрическое манипулирование с цепочками координатных пар (x, y) , которое можно представить в виде ряда элементарных компьютерных процедур. Среди них выделим пять основных, по содержанию схожих с традиционными методами картографической генерализации.

Упрощение — алгоритмы упрощения позволяют исключить лишние или ненужные координатные пары, исходя из определенного геометрического критерия, в качестве которого используют, например, величину расстояния между точками или смещения от *центральной линии*.

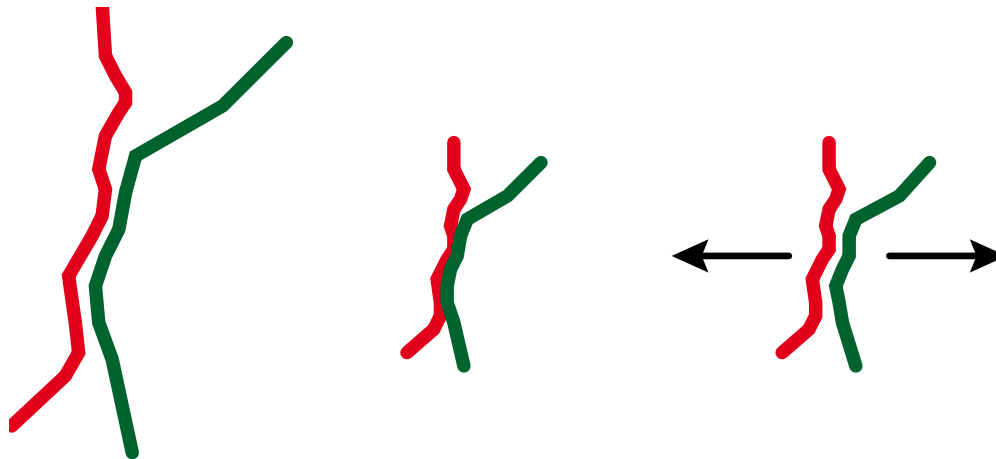


Сглаживание — программы сглаживания позволяют переместить или сдвинуть координатные пары для устранения мелких нарушений и выделить только наиболее значимые тенденции изменения линии. Как правило, сглаживание оцифрованной линии используется для уменьшения влияния процесса кодирования; считается, что сглаживание качественно улучшает такую линию. При сглаживании количество необходимых координат возрастает, поэтому обычно оно применяется только для вывода информации.

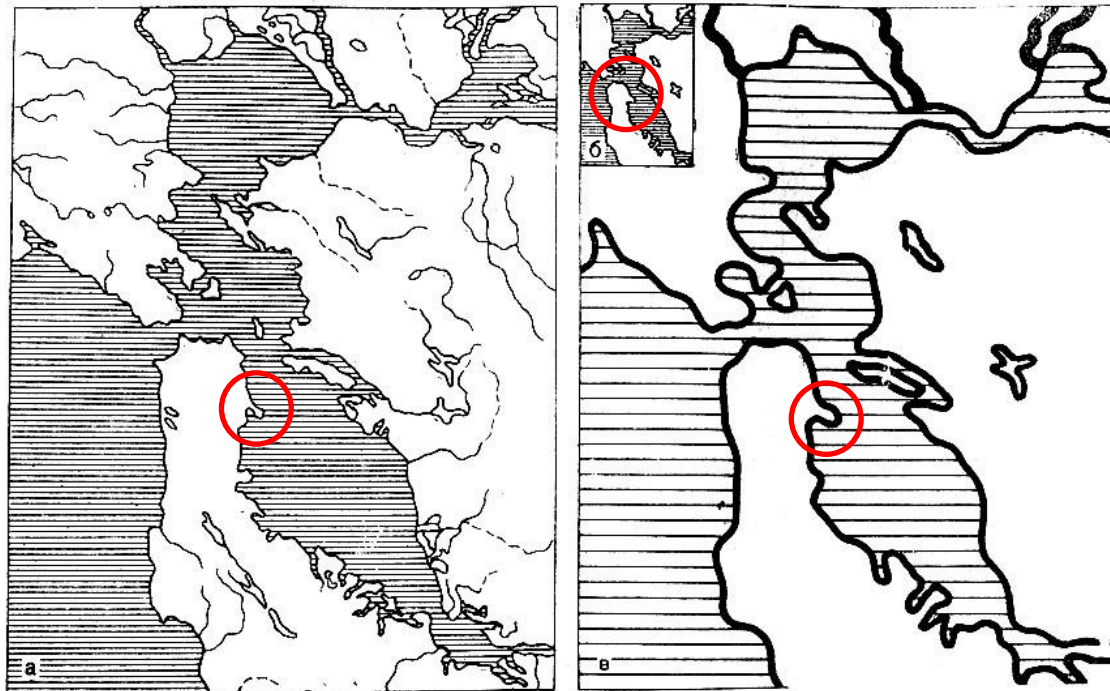


Улучшает визуальное и картографическое качество линий посредством сглаживания их углов.

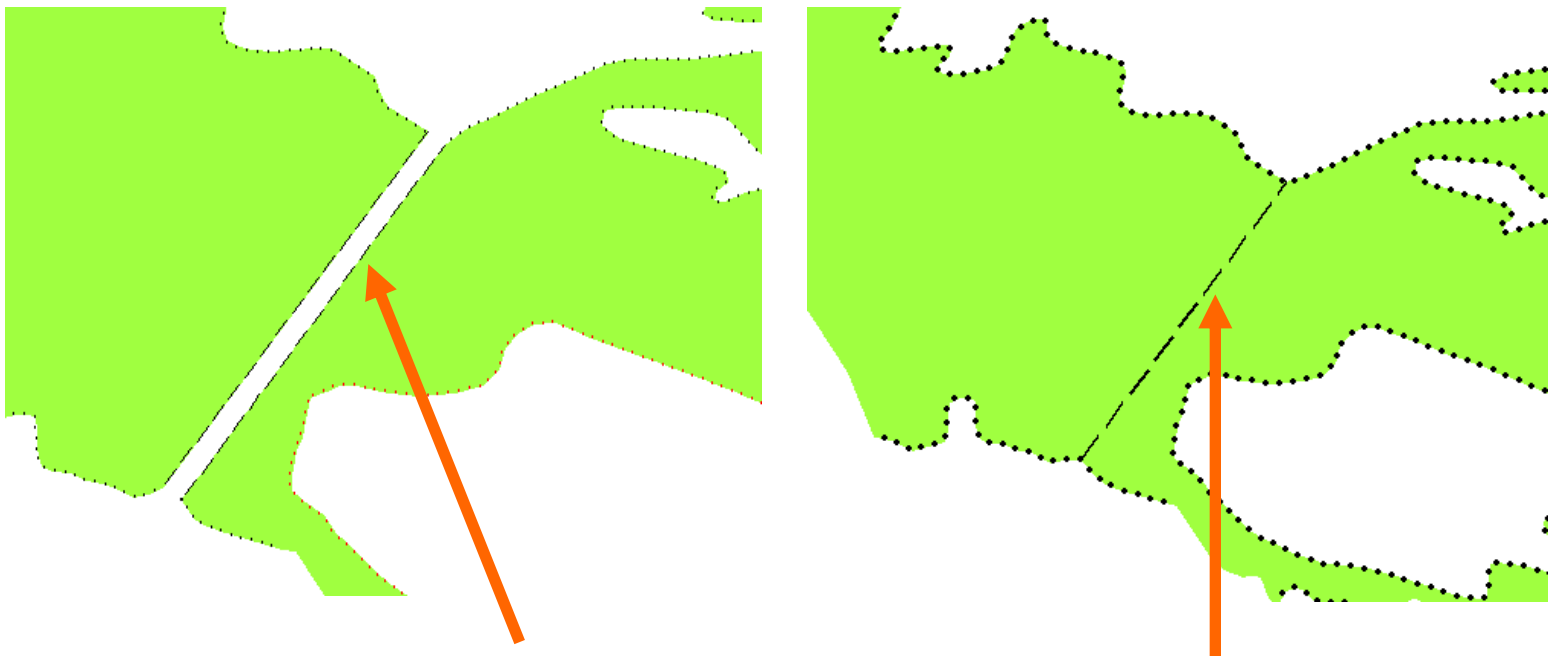
Перемещение объектов — предполагает сдвиг двух объектов, чтобы при уменьшении масштаба избежать их слияния или наложения; большинство алгоритмов перемещения объектов в векторном формате ориентировано на интерактивный режим, когда векторы начального перемещения задаются специалистом-картографом; в других случаях для регулирования процесса перемещения используется уменьшенная копия объекта.



Корректировка (утрирование) — с помощью корректировки в уже упрощенный набор данных можно снова ввести некоторые детали; например, сглаженная линия может потерять сходство с побережьем, тогда для улучшения ее вида будет проведено утрирование в случайных точках; один из методов предполагает расчленение линии путем введения дополнительных точек и придания большего сходства с оригиналом.



Слияние — ведет к объединению двух параллельных объектов при уменьшении масштаба; например, берега реки или обочины дороги в мелком масштабе сливаются, остров превращается в точку; алгоритмы слияния соединяют два параллельных линейных объекта.



Наиболее «узким местом» автоматизированных систем является время построения карты; с уменьшением числа координатных пар в процессе упрощения увеличивается скорость черчения. Упрощение позволяет сократить набор данных на 70 %, не изменяя при этом основных параметров линии; в результате значительно экономятся ресурсы памяти.

При уменьшении масштаба карты координатные пары сближаются, и при значительном уменьшении разрешение данных может даже превысить графическое разрешение выводного устройства. При упрощении такие координатные пары будут отброшены еще до уменьшения масштаба.

Время выполнения ряда векторных операций (сдвиг, поворот, масштабирование, картометрический анализ) при упрощении набора данных намного сокращается; ускоряются и многие процессы создания условных знаков, штриховок.

Алгоритмы упрощения линий

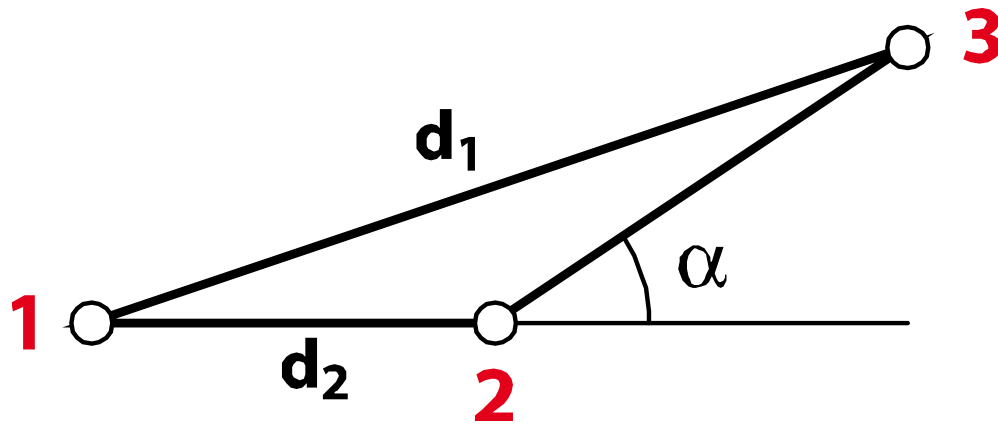
Алгоритмы упрощения линий отличаются критериями выбора точек.

Алгоритм независимых точек очень прост и ни в коей мере не учитывает топологические связи с другими координатными парами, При его выполнении отбрасывается либо каждая n -я пара (например, третья, десятая). Либо случайно отбирается $1/n$ часть набора координатных пар.

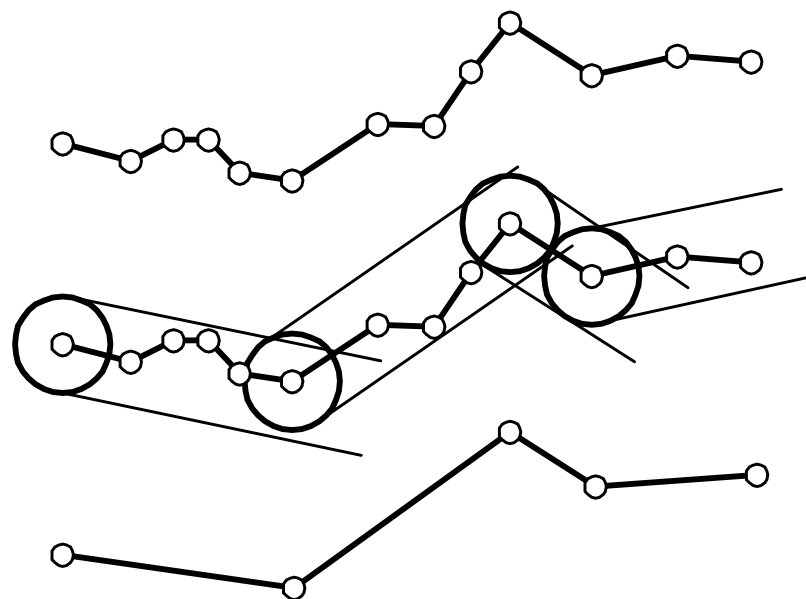
Алгоритмы локальной обработки для принятия решения об отбрасывании координатных пар используют данные о положении соседних точек:

- а) евклидово расстояние между точками (длина отрезка оцифрованной линии);
- б) изменение углов между соседними отрезками линии;
- в) комбинацию первых двух критериев.

Примером реализации последнего критерия служит алгоритм Джэнкса. Он требует задания трех исходных параметров: d_1 и d_2 — наименьшие допустимые расстояния от первой точки до двух последующих, и α — наибольший допустимый угол между двумя векторами, соединяющими три точки. Если одно из вычисленных расстояний меньше соответствующего минимума или угол меньше допустимого предела, то промежуточная точка отбрасывается.



В некоторых алгоритмах для отбора точек требуется осуществить некоторые геометрические построения. Например, для обозначения района поиска строится «коридор», определяемый двумя параллельными линиями. Параметрами коридора могут служить его наклон, ширина, задание области поиска координат или расстояний.



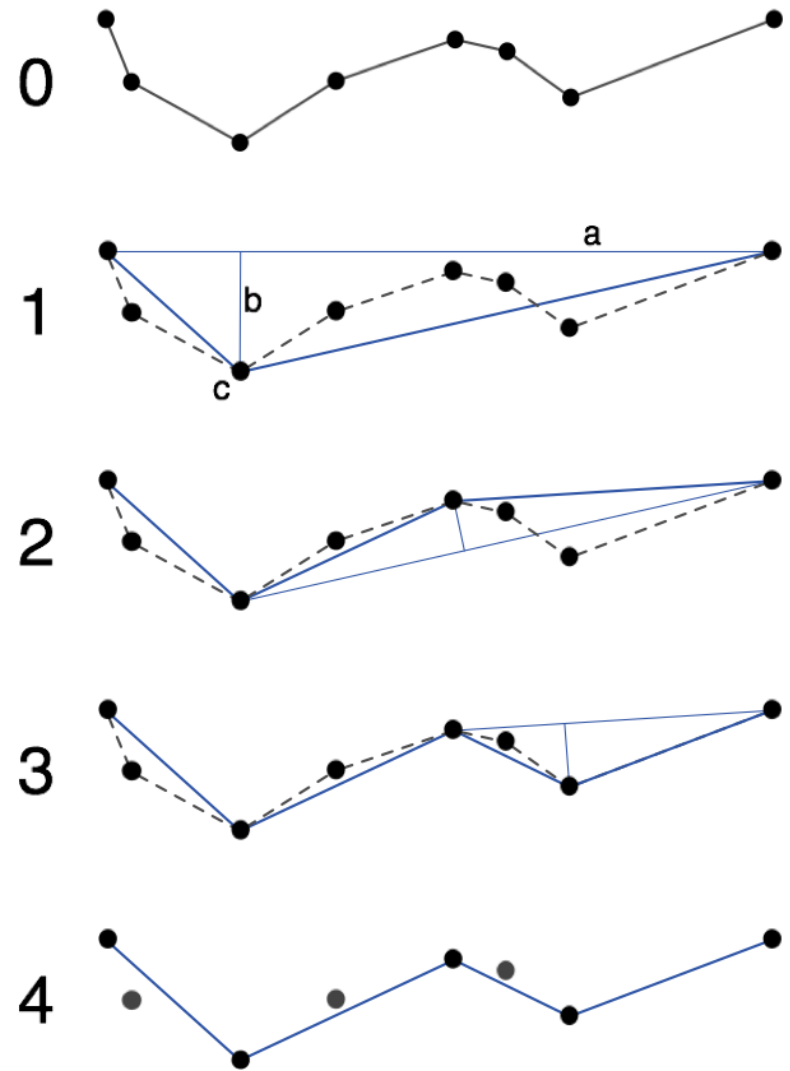
Исходная линия

Схема построения коридора

Генерализованная линия

В алгоритмах глобальной обработки линия рассматривается как целое. Среди них наиболее распространен алгоритм Дугласа–Пейкера. Принцип его работы основан на задании полосы (коридора) допустимых отклонений, многократном просмотре с помощью этой полосы всех точек линии, начиная с первой и последней, и фиксации каждый раз наиболее удаленной от ее центральной линии точки. Подобным образом обрабатывают всю исходную линию, при необходимости возвращаясь назад до тех пор, пока все точки не окажутся внутри своего коридора. Упрощенная линия будет образована центральными линиями всех коридоров, построенных в этом процессе.

Практически доказано, что алгоритм Дугласа–Пейкера дает наилучшие результаты (лучше других сохраняет важные геометрические характеристики данных).



Использование теории фракталов при проведении генерализации

Английский исследователь Л. Ричардсон еще в начале 1960-х гг. провел подробное исследование картографических представлений береговых линий и измерений их длин в различных масштабах. Он выявил едва ли не бесконечное число типов извилистости таких линий. Эти исследования привели его к выводам, что важное свойство линий заключается в том, что некоторые их сегменты, видимые при низком разрешении, приблизительно повторяются при последовательно повышающемся разрешении. Согласно *теории фракталов*, такие линии имеют *самоподобие* форм и могут быть промоделированы *фрактальными линиями*.

Термин «фракталы» был введен Бенуа Мандельбротом в 1977 г.

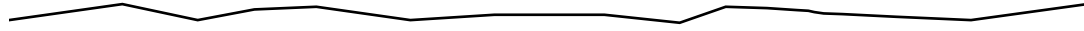
Фракталы — это не столько набор моделей, сколько набор понятий. Они создают основу для понимания того, как изменяются изображения реальных географических объектов при генерализации или изменении масштаба, позволяют перейти к систематическому анализу влияния масштаба и разрешения на процесс генерализации. Многие исследования по фрактальной геометрии природы пока еще только начинаются (особенно в геоморфологии и картографии), но в некоторых областях уже достигнуты поразительные результаты.

В обычной евклидовой геометрии, которую в основном используют для представления географической реальности, действуют с точками, линиями, площадями и объемами. Евклидова размерность (E) выражена положительными целыми числами и отражает количество координат, необходимых для определения точки. Чтобы обозначить какую-либо точку на профиле, нужны две координаты, следовательно, евклидова размерность профиля равна двум. Для обозначения точки на поверхности требуются три координаты, поэтому размерность поверхности — три. С евклидовой размерностью тесно связана так называемая топологическая размерность (D_T) объектов — на плоском листе бумаги (с евклидовой размерностью 2) можно изобразить двумерную фигуру ($D_T = 2$), одномерную линию ($D_T = 1$) и точку с нулевой размерностью ($D_T = 0$).

Англоязычный термин *fractals* происходит от латинского корня *fractus* (дробный). Во фрактальной геометрии тоже действуют с точками, линиями, площадями и объемами, но не ограничиваются целочисленной размерностью — фрактальная размерность (D) может выражаться любым действительным числом. Это число должно быть, по меньшей мере, равно топологической размерности явления, а верхний предел не должен превышать евклидову размерность (т. е. $0 < D_T < D < E$). Для прямой линии топологическая и фрактальная размерности одинаковы и равны 1; слегка изогнутые линии сохраняют топологическую размерность 1, а фрактальная размерность будет несколько больше 1; сильно изогнутые линии при $D_T = 1$ характеризуются большей фрактальной размерностью; линия, полностью «заполняющая» лист бумаги, имеет фрактальную размерность 2.

В природе многие картографические линии характеризуются фрактальной размерностью от 1,15 до 1,30, а поверхности — от 2 (абсолютно плоские) до 3 (целиком заполняющие пространство).

$D = 1,1$



$D = 1,2$



$D = 1,4$



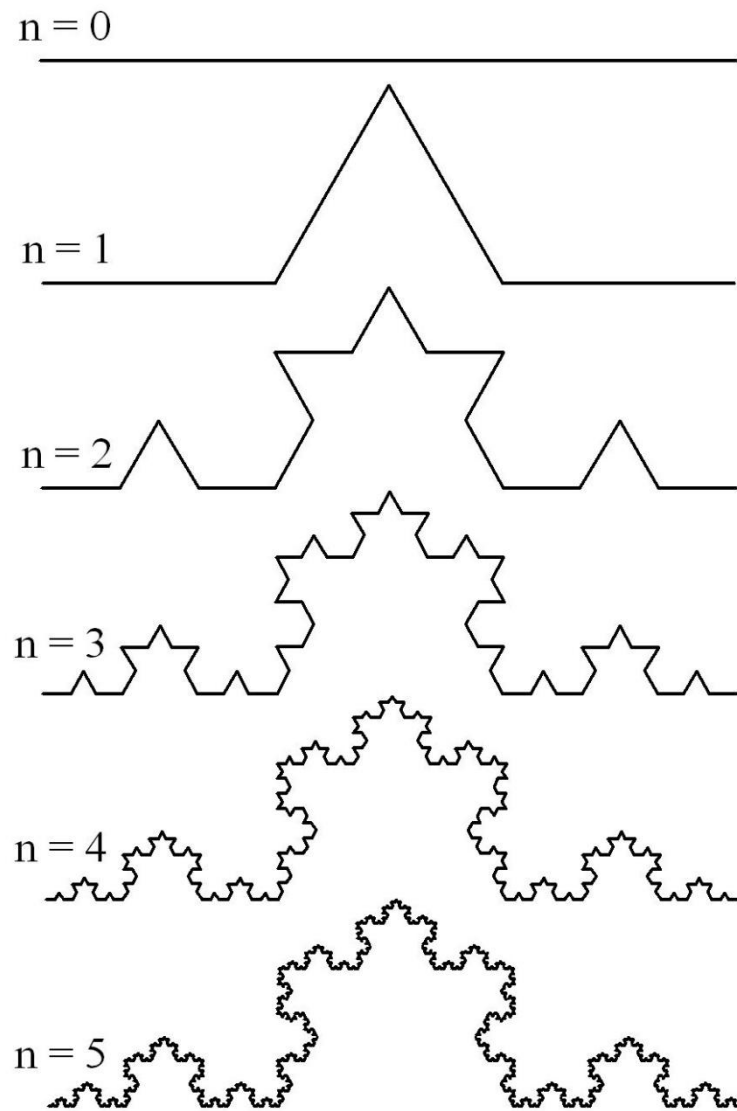
Линии разной фрактальности

Фрактальная размерность является новой и необычной характеристикой метрической информации о линиях и поверхностях.

Самоподобие и масштабирование. С математическим определением фрактала связано и еще одно понятие, отражающее иерархический принцип организации действительности, — *самоподобие* фрактальных объектов. Самоподобие означает, что некоторые черты объекта или явления сохраняются при таких изменениях масштаба, как приближение или удаление.

При геометрическом подобии существует строгое соответствие между крупным и мелким масштабами. Формы, сохраняющиеся при разном изменении масштаба по разным координатным осям, называются *самоподобными*.

Замечено, что такими фрактальными свойствами обладают многие географические объекты — побережья морей и океанов, реки, горные ущелья, государственные границы там, где они проведены по природным контурам



Использование фрактальной геометрии (особенно фрактальной размерности) позволяет проводить сопоставления и оценивать параметры объектов или их отношения в зависимости от масштаба. Эти вопросы особенно важны для геоинформационного картографирования при цифровом представлении картографических характеристик и касаются таких задач, как:

- ▣ определение подходящего интервала кодирования (цифрования);
- ▣ генерализация линий — наилучшим будет такой способ генерализации, при котором сохраняется фрактальная размерность линии;
- ▣ изображение линий в масштабе, крупнее того, в котором проводилась съемка — вводится дополнительная «информация» путем добавления деталей, рассчитанных как функция фрактальной размерности исходной линии;
- ▣ включение фрактальной размерности в число картометрических показателей.

Выводы

Долгое время генерализация считалась субъективным процессом, зависящим от навыков и знаний картографа. Для решения проблемы ее формализации исследователями были привлечены методы вычислительной геометрии, искусственного интеллекта, теории фракталов и др. Однако к настоящему времени уровень автоматизации основных процессов генерализации, предлагаемый в существующих системах, не достаточен для перехода картографов от ручного выполнения операций к машинным (человеко-машинным) процедурам. Не реализованы ГИС-технологии, позволяющие эффективно организовать отбор и обобщение картографических объектов, в полной мере охватив все требования, предъявляемые к генерализации. Наиболее перспективным направлением исследований данной проблемы является применение *интеллектуальных информационных систем*, хорошо зарекомендовавших себя при решении слабоформализованных задач.

С автоматизированной генерализацией напрямую связан вопрос **мультимасштабности карты**.

Мультимасштабные карты, становясь стандартом де-факто в ГИС-среде, требуют соответствующей разработки методик проектирования и наполнения мультимасштабных баз данных.

Алгоритмы автоматической генерализации пока что не удовлетворяют требованиям ни по качеству и географической достоверности результатов, ни по затратам времени на выполнение, поэтому в интерактивной среде неприемлемы. В то же время масштабный ряд карты может включать один-два десятка уровней с диапазоном значений, например, от 1:10 000 до 1:10 000 000. Необходимо каким-то образом обеспечить иерархию масштабов карты данными соответствующей структуры и точности.

Достаточно простое с концептуальной точки зрения решение этой проблемы состоит в создании мультимасштабной базы геоданных (МБГД). БГД — это объектно-ориентированная модель данных, позволяющая описывать поведение объектов. МБГД состоит из наборов данных различной детализации (пространственной и семантической), формирующих так называемые лоды, или уровни детализации (от англ. LoD — Level of Detail). Каждый лод оптимизирован для отображения в определенном диапазоне масштабов, т. е. для него существуют опорный масштаб, к которому приведены (генерализованы) данные, и некая масштабная окрестность, в пределах которой можно использовать данные без ущерба для читаемости карты, скорости ее отображения, эстетичности и т. д.

Мультимасштабное интерактивное картографирование является актуальной задачей. Многие его подзадачи, например интерактивная генерализация данных, пока не имеют удовлетворительного решения и требуют дальнейших исследований. Мультимасштабные карты, становясь стандартом де-факто в ГИС-среде, требуют соответствующей разработки методик проектирования и наполнения *мультимасштабных баз данных*.