Министерство общего и профессионального образования
Российской Федерации

Томский государственный университет

На правах рукописи

УДК 519.683

Скворцов Алексей Владимирович

ЭФФЕКТИВНЫЕ АЛГОРИТМЫ ВЫЧИСЛИТЕЛЬНОЙ ГЕОМЕТРИИ И ИХ РЕАЛИЗАЦИЯ В ГЕОИНФОРМАЦИОННОЙ СИСТЕМЕ

Специальность 05.13.16 –
«Применение вычислительной техники, математического моделирования и математических методов в научных исследованиях»

Автореферат

диссертации на соискание учёной степени
кандидата технических наук

Томск – 1998

Работа выполнена в Томском государственном университете

Научный руководитель:

кандидат технических наук,

доцент Костюк Ю.Л.

Официальные оппоненты:

доктор физико-математических наук, профессор Ю.М. Полищук,

кандидат технических наук, доцент С.В. Быкова.

Ведущее предприятие:

ПО «Инженерная геодезия» (г. Новосибирск).

Защита состоится 21 января 1999 г. в 14 на заседании Специализированного совета по защите диссертаций Д 063.53.03 при Томском государственном университете (634050, г. Томск, пр. Ленина, 36).

С диссертацией можно ознакомиться в библиотеке Томского государственного университета.

Отзывы на автореферат (2 экз.), заверенные печатью, высылать по адресу: 634050, г. Томск, пр. Ленина, 36, учёному секретарю ТГУ.

Автореферат разослан 11 декабря 1998 г.

Учёный секретарь

Специализированного совета,

к. ф-м. н., доцент Б.Е. Тривоженко

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. Геоинформатика и вычислительная геометрия являются в последнее время одними из наиболее развивающихся областей информатики. Их достижения в явном виде используются в двух основных классах программных систем: в геоинформационных системах (ГИС) и системах автоматизированного проектирования (САПР). До последнего времени эти прикладные направления развивались почти независимо друг от друга, и только в последние годы наметилась тенденция к их сближению и взаимообогащению идеями, принципами функционирования, алгоритмической проработкой. В связи с этим открылся новый пласт практических задач, которые удаётся эффективно решать, используя комбинированные возможности ГИС и САПР. Таким образом, в настоящее время созрела необходимость создания универсальной графической системы, объединяющей потенциальные возможности упомянутых систем.

Проведённое в работе исследование существующих функций ряда отечественных и зарубежных геоинформационных систем и систем автоматизированного проектирования позволило выявить набор базовых функций систем обоих классов, присущие им недостатки и преимущества друг перед другом, выявить ряд нерешённых задач, а также предложить более эффективные способы решения существующих задач, таких как построение цифровой модели рельефа на основе нерегулярного набора отсчётов с использованием триангуляции Делоне, построение оверлеев, буферных зон, взвешенных зон близости и изоконтуров на основе триангуляции с ограничениями, а также разработать усовершенствованные алгоритмы для пространственного индексирования объектов с помощью R-деревьев.

Проектирование и создание универсальной графической системы на базе современных достижений информатики, в частности, геоинформатики, вычислительной геометрии, теории программирования, использование для реализации современных программных объектно-ориентированных технологий позволяет охватить потенциально более широкий круг задач, нежели традиционные системы ГИС и САПР и уменьшить стоимость программно-технических решений задач для конечных пользователей.

Цель работы заключается в разработке методов повышения эффективности графических информационных систем, исследование возможностей расширения их функций, а также реализации предложенных методов в виде базовой универсальной графической информационной системы. В рамках этой общей задачи требовалось разработать и исследовать базовые алгоритмы двумерного пространственного индексирования объектов с помощью R-деревьев, построения нерегулярной модели рельефа на основе триангуляции, решения задач пространственного анализа на плоскости, в частности, алгоритмы построения триангуляции Делоне, триангуляции с ограничениями, построения оверлеев, буферных зон, взвешенных зон близости и изоконтуров.

Научная новизна. На основе обзора ряда современных реализаций ГИС и САПР выявлен набор базовых функций данных систем и выработаны требования к универсальной графической системе, объединяющей функции ГИС и САПР.

Предложена группа новых алгоритмов построения триангуляции Делоне, имеющих трудоёмкость в среднем на равномерном распределении $O(N)$ и $O(N \log N)$, усовершенствованы известные алгоритмы.

Предложен алгоритм построения триангуляции с ограничениями, на основе которого получены новые эффективные алгоритмы решения задач построения оверлеев, буферных зон, построения зон близости и изоконтуров.

Разработана группа новых алгоритмов построения индексной структуры R-дерева, которые позволили повысить скорость графического поиска.

Практическая ценность работы. Разработана универсальная графическая информационная система ГрафИн, в которой реализованы основные возможности, присущие любой полнофункциональной ГИС, базовые функции САПР, а также воплощены предложенные в работе алгоритмы и структуры данных.

На базе системы ГрафИн разработан ряд прикладных систем, использующих в совокупности возможности геоинформационных систем и систем автоматизированного проектирования. Например, созданная информационная система городских коммуникаций позволяет изображать технологические схемы инженерных сетей на полноценной карте города, анализировать технологические аспекты сети, используя информацию по географическому расположению инженерных объектов.

Применение в системе ГрафИн новых глобальных алгоритмов построения R-деревьев позволило существенно увеличить скорость выполнения основных операций системы, в том числе скорость визуализации карт и чертежей при самом распространённом режиме работы – при просмотре и навигации по изображению.

Предложенные в работе алгоритмы построения триангуляции Делоне с линейной в среднем трудоёмкостью представляют самостоятельную практическую ценность, позволяя принципиально улучшить производительность операций по работе с моделями поверхности, построенной по нерегулярному набору точек, в сравнении с существующими алгоритмами, лучшие из которых имеют трудоёмкость $O(N \log N)$.

Предложенные в работе алгоритм построения триангуляции с ограничениями, а также на его основе эффективные алгоритмы построения оверлеев, буферных зон, взвешенных зон близости и изоконтуров, имеют несомненную практическую ценность, позволяя значительно уменьшить среднюю трудоёмкость основных операций пространственного анализа в геоинформационных системах, имея в тоже время значительно более простую реализацию по сравнению с известными алгоритмами.

Графическая информационная система ГрафИн в настоящее время является коммерческим программным продуктом. Система как инструментальная оболочка и созданные на её основе прикладные комплексы внедрены в ряде организаций Томска и других городов.

Апробация работы. Диссертационная работа и её разделы докладывались и получили положительную оценку специалистов на 14 конференциях и семинарах областного, всероссийского и международного уровней, в том числе на:

1. Всероссийских научно-практических конференциях «ГИС-Форум», г. Москва, 1997, 1998.
2. Всероссийской научно-практической конференции «Непараметрика-97», г. Красноярск, 1997.
3. Международной научно-практической конференции ИНПРИМ-98, г. Новосибирск, 1998.
4. Международной научно-практической конференции SCOR-98, г. Новосибирск, 1998.
5. Международной научно-практической конференции Интеркарто-4, г. Барнаул, 1998.

Объём работы. Диссертация состоит из введения, пяти глав, заключения, двух приложений и списка цитируемой литературы. Общий объём работы составляет 177 страниц, из них 8 страниц – приложения, 15 страниц – список литературы (150 названий). Текст работы иллюстрируется 47 рисунками и 7 таблицами.

СОДЕРЖАНИЕ РАБОТЫ

Во введении кратко описаны актуальность проблемы, цель работы и характеристика полученных в работе результатов.

В первой главе проведён анализ функциональных возможностей отечественных и зарубежных геоинформационных систем и систем автоматизированного проектирования, который позволил выявить сходные черты двух классов программного обеспечения, а также свойственные им недостатки, наиболее ярко проявляющиеся при решении задач, находящихся на стыке ГИС и САПР.

Так, системы САПР, как правило, не могут работать с большими массивами пространственных данных, свойственных ГИС, не умеют выполнять задачи пространственного анализа, типичные для геоинформатики, слабы в атрибутном описании графических объектов.

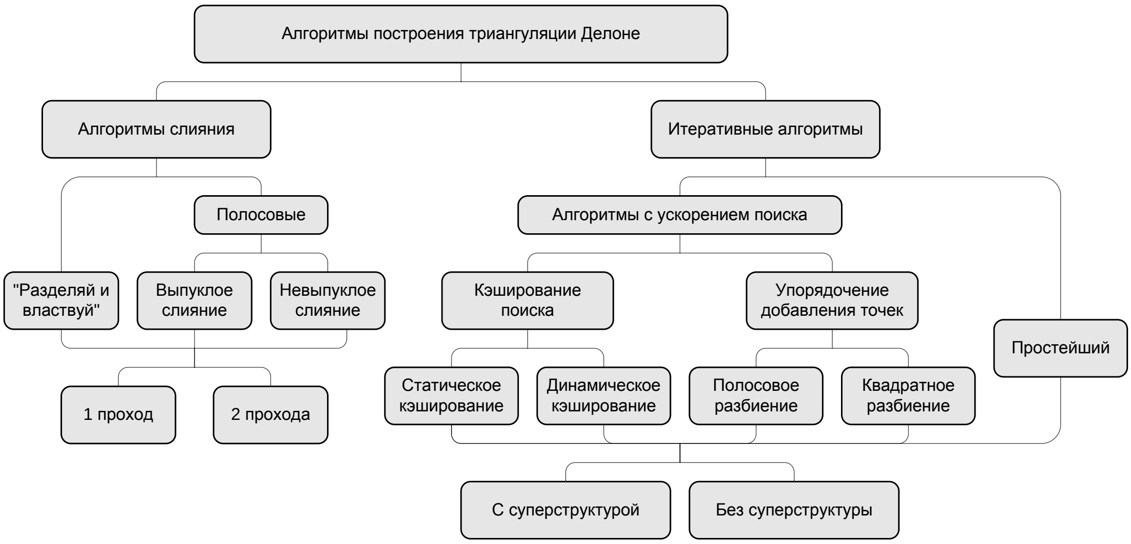
С другой стороны, геоинформационные системы имеют недостаточный набор графических примитивов, и, как следствие, применяют упрощённый подход к моделированию конкретной предметной области и решению в ней задач.

В главе исследуются возможности построения универсальной графической информационной среды, интегрирующей функции ГИС и САПР, а также формулируется ряд обоснованных требований к такой системе.

Для существующих реализаций систем ГИС и САПР выявлен ряд подзадач, определяющих, в конечном счёте, эффективность работы системы в целом, и для которых необходима разработка новых, более совершенных алгоритмов. Основная сложность таких задач связана со слишком большой трудоёмкостью, не позволяющей решать задачи с большим объёмом входных данных. Это задача регионального поиска объектов, задача построения планарной триангуляции Делоне, а также задачи пространственного анализа на плоскости, решения которых рассматриваются в следующих главах.

Во второй главе исследуется задача построения планарной триангуляции, являющейся базовой в подсистеме моделирования поверхностей любой геоинформационной системы. Наиболее часто для такого моделирования используется триангуляция Делоне, обладающая рядом оптимальных свойств среди всех возможных триангуляций. Лучшие из известных автору алгоритмов построения триангуляции Делоне имеют трудоёмкость в худшем и среднем $O(N \log N)$, но при этом имеют, как правило, сложную алгоритмическую реализацию и относительно большое время работы на больших наборах точек.

В работе рассматривается ряд существующих и вновь предлагаемых алгоритмов построения триангуляции Делоне, структурно изображённых на рисунке. Все эти алгоритмы были реализованы на одной программно-аппаратной платформе и протестированы на различных по размеру и закону распределения наборах точек.



Классический алгоритм рекурсивного слияния «Разделяй и властвуй» до настоящего времени в литературе считается наиболее эффективным из всех алгоритмов построения триангуляции Делоне, хотя и сложным по реализации. Этот алгоритм имеет трудоёмкость в среднем и худшем случаях $O(N \log N)$ (здесь и далее по тексту оценки трудоёмкости в среднем для алгоритмов триангуляции приводятся для равномерного распределения исходных точек в единичном квадрате). В главе исследуются другие алгоритмы, использующие в своей работе принцип слияния.

Предлагаемые *алгоритмы полосового слияния* основаны на разбиении множества входных точек на полосы по признаку одинакового количества точек в полосах (с помощью алгоритма вычисления квантилей множества чисел) или по одинаковой ширине полос (цифровой сортировкой). После этого на множествах точек внутри полос специальным линейным по сложности алгоритмом строятся полосовые триангуляции. На следующем этапе алгоритма полученные частичные триангуляции склеиваются в единую сплошную триангуляцию. В более простом *алгоритме выпуклого слияния* для полосовых триангуляций достраивается выпуклая оболочка и затем применяется алгоритм слияния двух триангуляций, применяющийся в классическом алгоритме «Разделяй и властвуй». В другом более сложном варианте *невыпуклого слияния* частичные триангуляции не достраиваются выпуклой оболочкой, а непосредственно сливаются, используя специальный механизм, предотвращающий возникновение пересекающихся треугольников.

Полученные алгоритмы полосового слияния имеют в худшем случае трудоёмкость $O(N^2)$, а в среднем на равномерном распределении – $O(N \log N)$ или даже $O(N)$ в зависимости от используемого алгоритма сортировки точек внутри полос. Даже при использовании сортировки с трудоёмкостью $O(N \log N)$ преимущество этих алгоритмов перед классическим «Разделяй и властвуй» ярко проявляется на равномерном распределении, т.к. полученные алгоритмы имеют меньший коэффициент пропорциональности в формуле зависимости количества операций алгоритма относительно числа входных точек. Так, на различных по количеству наборах точек алгоритм выпуклого слияния даёт выигрыш около 5% времени, а невыпуклого ~ 15%.

Для полосовых алгоритмов слияния получена формула выбора оптимального числа полос для равномерного распределения точек в прямоугольнике размером $a \times b$ при условии минимизации манхеттенновской длины рёбер получаемой триангуляции:

$$m^* = \sqrt{s_{сл.пол.} \cdot \frac{a}{b} \cdot N}\ , \tag{1}$$

где $s_{сл.пол.}$ – коэффициент разбиения на полосы алгоритма полосового слияния, значение которого $\approx 0{,}13$ установлено экспериментально для различных распределений исходных точек.

Алгоритмы двухпроходного построения триангуляции Делоне опираются на известный в литературе факт, что любая триангуляция может быть преобразована в триангуляцию Делоне последовательными локальными преобразованиями пар треугольников, не удовлетворяющих критерию Делоне. На первом этапе таких алгоритмов строится триангуляция, для треугольников которой условие Делоне может не выполняться. После этого анализируются все пары соседних треугольников в триангуляции и при необходимости перестраиваются.

В работе установлено, что применение двухпроходного подхода к построению триангуляции Делоне позволяет в ряде случаев значительно улучшить производительность, но, что самое главное, существенно упростить алгоритмы построения триангуляции, т.к. на первом этапе работы алгоритма нет необходимости контролировать выполнение условия Делоне, а также поддерживать дополнительные структуры, без которых трудоёмкость алгоритма слияния становится хуже оптимальной. Применение двухпроходного подхода к построению триангуляции Делоне позволяет улучшить скорость работы упомянутых выше алгоритмов слияния, что объясняется тем, что эти алгоритмы в отличие от рассматриваемых ниже итеративных алгоритмов даже без проверки условия Делоне строят треугольники, которые с большой вероятностью в дальнейшем не будут перестраиваться.

Вторая большая группа – *итеративные алгоритмы построения триангуляции*. Эти алгоритмы по сравнению с алгоритмами слияния обычно гораздо проще в реализации. Трудоёмкость классического простого итеративного алгоритма равна в среднем $O(N^{3/2})$, а в худшем случае $O(N^2)$. Основное время в этих алгоритмах тратится на поиск треугольника, в который попадает очередная добавляемая точка. Если бы треугольники находились мгновенно, то трудоёмкость итеративного алгоритма на равномерном распределении была бы в среднем линейной. Поэтому в работе исследуются различные варианты ускорения поиска треугольников.

В литературе предлагается некоторым образом упорядочить порядок добавления точек в триангуляцию так, чтобы очередная добавляемая точка была близка к предыдущей и, следовательно, треугольник, в который попадёт точка, будет близок к предыдущему найденному треугольнику. В предлагаемом автором *итеративном полосовом алгоритме* все точки разбиваются на полосы вдоль одной координаты, а затем сортируются внутри полос по другой. В *итеративном квадратном алгоритме* всё множество точек разбивается на квадраты (количество $\sim N$) с помощью цифровой сортировки. После этого точки добавляются в триангуляцию группами, соответствующими квадратам, последовательно перебирая смежные квадраты разбиения. Полученные алгоритмы имеют трудоёмкость на равномерном распределении в среднем $O(N \log N)$ или $O(N)$ в зависимости от используемого алгоритма сортировки, в худшем – $O(N^2)$.

Для квадратного алгоритма получена формула выбора оптимального числа квадратов $m \times m$ для равномерного распределения точек в прямоугольнике размером $a \times a$ при условии минимизации расстояний между последовательно добавляемыми в триангуляцию точками:

$$m^* = \sqrt{2N - 3}\,. \qquad (2)$$

Другой вариант ускорения поиска треугольников использует идею перехода от сложного планарного треугольного разбиения плоскости к другому более простому (например, к регулярной прямоугольной сети), покрывающему первое, но в котором возможно очень быстро проводить локализацию точки. Для этого, необходимо сопоставить каждому элементу второго разбиения (например, прямоугольнику) некоторый элемент первого – треугольник, который бы находился в некотором смысле как можно ближе к исходному (прямоугольнику). После этого для локализации треугольника по заданной точке плоскости необходимо найти элемент вторичного разбиения (прямоугольник), от него по ссылке перейти к треугольнику, который с большой вероятностью находится близко к цели. После того как искомый треугольник найден, таблица ссылок между разбиениями (кэш) может быть скорректирована на основании результата последнего поиска. Построенные на такой идее алгоритмы построения триангуляции Делоне названы автором *алгоритмами кэширования*.

В работе предлагается два варианта кэширования – *алгоритм статического кэширования*, в котором перед началом работы алгоритма выделяется фиксированного размера кэш, и *алгоритм динамического кэширования*, в котором кэш растёт в размере по мере увеличения количества точек в частично построенной триангуляции. Трудоёмкость этих алгоритмов в худшем случае составляет $O(N^2)$, в среднем же на равномерном распределении для статического кэширования $O(N^{\frac{9}{8}})$, а для динамического – $O(N)$.

Для алгоритма статического кэширования получена формула выбора оптимального размера кэша для равномерного распределения точек в прямоугольнике размером $a \times a$ при условии минимизации количества переходов при поиске треугольников:

$$m^* = \frac{1}{3} N^{\frac{3}{8}}. \tag{3}$$

В аналогичных условиях для алгоритма динамического кэширования экспериментально получено значение коэффициента роста $r_{кэш} \approx 5$ в формуле условия увеличения размера кэша $n = r_{кэш} \cdot m^2$, где n – номер добавляемой точки, а m – текущий размер кэша.

В главе анализируются *суперструктуры* (множества дополнительно вводимых в триангуляцию точек) как средство упрощения итерационных алгоритмов и повышения в ряде случаев скорости их работы.

В конце главы приводятся результаты экспериментального моделирования всех представленных в работе алгоритмов построения триангуляции Делоне, подтверждающие их высокую эффективность на различных распределениях точек.

В третьей главе структура триангуляции исследуется для решения ряда задач пространственного анализа на плоскости, имеющих широкое применение в графических системах.

Основная идея предлагаемых алгоритмов заключается в разбиении всей задачи на элементарные задачи, решаемые на одном треугольнике триангуляционного разбиения плоскости с последующим объединением полученных результатов.

Для решения поставленных задач кроме алгоритма построения триангуляции дополнительно необходимы следующие общие алгоритмы.

1. Алгоритм построения триангуляции с ограничениями.
2. Алгоритм классификации треугольников по попаданию в полиполигоны.

3. Алгоритм выделения полиполигонов из триангуляции.

Задача построения триангуляции, в том числе с ограничениями, является неоднозначной. Среди всех возможных классов триангуляций на практике чаще всего используется триангуляция Делоне. Однако в случае дополнительного задания множества полилиний, которые должны пройти по рёбрам готовой триангуляции, построение триангуляции Делоне не всегда возможно. Поэтому в работе вводится определение *триангуляции Делоне с ограничениями*, для которой условие Делоне выполняется для всех рёбер, кроме некоторого набора фиксированных.

Для построения триангуляции Делоне с ограничениями по заданным множествам точек и полилиний предлагается использовать какой-либо итеративный алгоритм построения обычной триангуляции Делоне, например, алгоритм динамического кэширования с суперструктурой, предложенный во второй главе настоящей работы.

В предлагаемом алгоритме построения триангуляции Делоне с ограничениями вначале в цикле по всем полилиниям производится последовательная вставка точек полилинии и составляющих её отрезков. При этом для каждого добавляемого отрезка делается проверка на пересечение с уже существующими рёбрами триангуляции, выполняя при необходимости перестроения треугольников и разбиения фиксированных рёбер на части.

Пусть G – количество отрезков полилиний, N – число точек в триангуляции, а P – число точек взаимного пересечения отрезков полилиний. Тогда трудоёмкость предложенного алгоритма составляет в худшем случае $O((G+P+N)^2)$, а в среднем – $O(G+P+N)$, если среднее число треугольников, пересекающихся с вставляемыми фиксированными рёбрами, на каждом шаге ограничено константой.

Для решения задачи классификации всех треугольников триангуляции по признаку их попадания в заданное множество полиполигонов автором предлагается алгоритм, использующий идею алгоритма растеризации произвольных полиполигонов в комбинации с алгоритмом заливки растровой области методом затравки. Трудоёмкость работы данного алгоритма составляет $O(N)$.

В задаче выделения полиполигонов из триангуляции необходимо объединить все треугольники с одинаковыми кодами в полиполигоны. Используя идею алгоритма заливки растровой области методом затравки, автору удалось построить алгоритм, имеющий трудоёмкость $O(N)$.

Задача построения буферных зон требует определения всех точек плоскости, удалённых от множества объектов $\{a_i\}$ (обычно это точки, линии и полигоны) не более чем на расстояние $s_i = s(a_i)$. На основе вышеприведённых алгоритмов в работе предлагается следующий алгоритм. Для всех точечных объектов и вершин ломаных и полигонов строятся полигоны, аппроксимирующие вокруг них круговые буферные зоны. Для всех отрезков ломаных и полигонов вычисляются прямоугольники, которые в объединении с ранее вычисленными круговыми зонами полностью определяют буферные зоны отрезков и ломаных. Все полученные круговые полигоны, прямоугольники и исходные полигоны передаются на вход алгоритма построения триангуляции с ограничениями. Затем все полученные треугольники классифицируются по признаку попадания в эти полигоны. В заключение все положительно проклассифицированные треугольники объединяются в один полиполигон.

Если N – общее число исходных точек и вершин входных ломаных и полиполигонов, а G – максимальное число сегментов аппроксимации круговых буферных зон, то общая трудоёмкость алгоритма составит $O((G \cdot N)^2)$ в худшем случае, и $O(G \cdot N)$ – в среднем.

Задача построения оверлеев определяются на множествах полиполигонов A и B как задача нахождения их объединения, пересечения, разности и симметрической разности. Результат должен быть представлен в виде одного полиполигона. На основе вышеприведённых алгоритмов в работе предлагается следующий алгоритм. Всё множество полиполигонов передаётся в алгоритм построения триангуляции с ограничениями в качестве замкнутых полилиний. Затем каждый треугольник полученной триангуляции классифицируется в зависимости от выполняемой операции. И в конце вызывается алгоритм выделения полиполигонов. Трудоёмкость полученного алгоритма составляет $O(N^2)$ в худшем случае, и $O(N)$ – в среднем.

Задача построения взвешенных зон близости требует определения всех точек плоскости, для которых расстояние s до объектов $\{a_i\}$, помноженное на веса $\{w_i \geq 0\}$, является минимальным.

Если все веса объектов $\{w_i\}$ равны, то задача определяется как задача построения диаграмм Вороного, являющаяся дополнительной по отношению к триангуляции Делоне, и поэтому имеет решение, равное по трудоёмкости задаче построения триангуляции Делоне.

Для решения задачи в общем случае автором в работе предлагается следующий алгоритм. В цикле по всем парам объектов вся плоскость разделяется на две части, достижимые из одной из точек либо из другой (по прямой или окружности). Полученные линии разделения передаются в качестве полилиний на вход алгоритма построения триангуляции с ограничениями. Затем все полученные треугольники классифицируются по признаку близости к исходным точкам. После чего выполняется алгоритм выделения полиполигонов.

Если N – общее число исходных точек, а G – максимальное число сегментов аппроксимации круговых зон близости, то общая трудоёмкость алгоритма составит $O(G^2 \cdot N^4)$ в худшем случае, и $O(G \cdot N^2)$ – в среднем.

В четвёртой главе исследуется другая важная практическая проблема вычислительной геометрии, связанная с задачей эффективного регионального поиска. В графических системах, работающих с большим объёмом пространственно определённой информации, эффективность решения данной задачи определяет, в конечном счете, скорость работы системы в целом.

Так как формы различных объектов в графических системах могут значительно различаться, то для упрощения алгоритмов регионального поиска для всех объектов вычисляется минимальный объемлющий прямоугольник со сторонами, параллельными осям координат. При этом задача регионального поиска в графических системах определяется как задача нахождения всех прямоугольных объектов в заданном прямоугольном регионе.

В настоящее время наиболее эффективной структурой, позволяющей обрабатывать неточечные объекты при эффективном размещении данных во вторичной памяти, считаются *R-деревья* и ряд их модификаций, в частности, наиболее эффективные из них *R*-деревья*, поэтому в дальнейшем в работе в качестве эталона используется именно эта разновидность R-деревьев.

Данная индексная структура предназначена для эффективного выполнения регионального поиска, поэтому основная идея R-дерева заключается в разбиении пространства в каждом узле дерева на минимально пересекающиеся группы, что позволяет при выполнении запросов эффективно отсекать неверные ветви поиска. В настоящее время даже для задачи разбиения на две группы не известен полиномиальный алгоритм, поэтому во всех алгоритмах построения R-деревьев используются приближённые подходы, дающие на реальных данных далеко не оптимальные разбиения.

Высокая степень перекрытия входов в узлах дерева происходит из-за динамичности алгоритмов его построения, когда при вставке очередного элемента его необходимо за минимальное время разместить и при необходимости перестроить дерево. При этом глобальная оптимизация дерева на каждом шаге не производится.

На практике работа с индексными структурами состоит из двух основных этапов: начального построения дерева и последующей оперативной работы. При этом во многих случаях во время второго этапа производится только поиск объектов без вставки новых и удаления старых.

В данной работе автором предлагается учесть специфику начального построения структуры, построив вначале более уравновешенную структуру R-дерева, а на последующих этапах использовать обычные алгоритмы для работы с R-деревьями. В частном случае такого подхода, когда вначале ничего не строится, получается обычное R-дерево.

В работе предлагается *глобальная стратегия построения R-дерева*, близкого к оптимальному. Главной и наиболее сложной частью этой стратегии является задача разбиения множества объектов на минимально пересекающиеся группы, поэтому в дальнейшем эта задача анализируется подробнее.

Во все существующие алгоритмы построения R-деревьев входят алгоритмы разбиения множества объектов на две части с минимальным пересечением. При этом выбираются два базовых объекта, последовательными присоединениями к которым всех остальных строятся необходимые группы. Обобщая данный алгоритм на случай нескольких групп, в работе предлагается выбрать по одному базовому объекту для каждой группы и, используя аналогичные итеративные алгоритмы добавления оставшихся объектов, произвести разбиение. На этом принципе построен предлагаемый в данной работе *базовый алгоритм разбиения*. Глобальный алгоритм построения R-дерева, построенный на базовом алгоритме разбиения, имеет трудоёмкость $O(N \log N)$.

Как показало экспериментальное исследование, проведённое автором, такой алгоритм хорошо ведёт себя на равномерных распределениях непересекающихся объектов. Но на неравномерных распределениях часто происходит переполнение строящихся групп, что приводит к необходимости помещения очередных объектов в неоптимальные группы. В итоге получаются недопустимо большие перекрытия построенных групп. Поэтому в работе предлагается *клеточный алгоритм*, в котором вначале на основе клеточного разбиения плоскости строятся группы объектов, и только затем анализируется их количество и наполнение. В случае необходимости полученные группы делятся пополам или объединяются с другими. Глобальный алгоритм построения R-дерева, основанный на клеточном алгоритме разбиения, имеет трудоёмкость $O(N^2)$.

Другим предлагаемым автором вариантом является *алгоритм, основанный на стратегии «Разделяй и властвуй»*, в котором производится последовательное разбиение всего множества на две части до тех пор, пока всё множество не окажется разбито на требуемое число частей. Глобальный алгоритм построения R-дерева, использующий алгоритм разбиения «Разделяй и властвуй», имеет трудоёмкость $O(N \log^2 N)$.

Для повышения качества полученного разбиения множества объектов на группы в работе автором предлагается *алгоритм уточнения*, позволяющий в ряде случаев значительно улучшить качество получаемого разбиения. Такой подход может использоваться как в обычных динамических алгоритмах манипуляции с R-деревьями (в алгоритме расщепления узла при вставке объектов в дерево), так и в алгоритмах разбиения, применяемых в глобальных алгоритмах. При этом доказывается, что трудоёмкость глобального алгоритма построения R-дерева не увеличивается при использовании уточнения разбиения.

В конце главы приводятся результаты экспериментального моделирования всех представленных в работе алгоритмов построения R-деревьев, подтверждающие их высокую эффективность на различных распределениях объектов и регионов поиска.

<u>В пятой главе</u> кратко описана созданная автором *универсальная графическая информационная система ГрафИн*, а также приведено её сравнение с рядом известных реализаций САПР и ГИС.

Система ГрафИн в своей работе использует наиболее эффективные современные алгоритмы вычислительной геометрии, что позволяет значительно улучшить скорость реакции всей системы в целом.

Система ГрафИн имеет ряд нововведений, выделяющих её из существующих систем ГИС и САПР. В первую очередь это полноценная возможность создания и редактирования в пределах одной карты или схемы разнообразных графическим изображений, ранее доступных только в каком-нибудь одном из упомянутых классов графических систем.

В таблице 1 приведены сравнение базовых функциональных возможностей реализованной системы ГрафИн с рядом наиболее распространённых в мире систем ГИС и САПР. В таблице 2 приведен ряд базовых операционных характеристик системы, определяющих скорость реакции системы при выполнении основных операций в системе.

Из приведённых таблиц, конечно, не следует полное превосходство системы ГрафИн над другими системами, но при этом можно заметить, что реализованный набор базовых функциональных возможностей находится на уровне ведущих в мире коммерческих систем.

Таблица 1. Сравнение базовых функциональных возможностей системы ГрафИн с лидирующими графическими системами.

Параметры	ГрафИн	ARC/INFO	MapInfo	AutoCAD
Поддерживаемые структуры данных:				
Векторные нетопологические	+	+	+	–
Векторные топологические	+	+	–	–
Растровые	+	+	+	+
Нерегулярные сети данных	+	+	–	–
Регулярные сети данных	+	+	–	–
Сложная графики САПР	+	–	–	+
Векторная серверная графика	+	+	+	–
Косметическое оформление	+	–	–	+
Стандартные возможности:				
Географические координаты	+	+	+	–
Отображение в проекции	–	+	+	–
Поддержка дигитайзера	–	+	–	+
Интерфейс с СУБД	+	+	+	–
Число визуализаторов графики	8	5	4	1
Число форматов данных	30	50	15	10
Функции пространственного анализа:				
Построение буферных зон	+	+	+	–
Построение оверлеев	+	+	–	–
Логические операции со слоями	+	+	+	–
Выделение объектов слоем	–	+	+	–
Построение зон близости	+	+	–	–
Анализ цифровых моделей рельефа:				
Вычисление углов наклона	+	+	–	–
Вычисление экспозиций склонов	+	+	–	–
Генерация изолиний	+	+	–	–
Генерация изоклин	+	–	–	–

Параметры	ГрафИн	ARC/INFO	MapInfo	AutoCAD
Генерация поперечных профилей	+	+	–	–
Подсчёт объёмов земляных работ	+	+	–	–
Определение зон видимости	–	+	–	–
Определение бассейнов стока	–	+	–	–
Инженерные задачи:				
Создание инженерных схем	+	–	–	+
Трассировка схем	+	–	–	+
Атрибутное наполнение схем	+	–	–	–
Совмещение схем с картами	+	–	–	–
Связь объектов карты со схемами	+	–	–	–
Прочие возможности:				
Решение задач на графах	+	+	–	–
Обработка геодезических данных	–	+	–	–
Ввод дистанционных данных	–	+	–	–

Таблица 2. Сравнение базовых операционных характеристик системы ГрафИн с лидирующими графическими системами.

Параметры	ГрафИн	ArcView	MapInfo	AutoCAD
Время полной отрисовки карты с 10 000 полигонами, с.	0,5	3	3	1,5
Время отрисовки фрагмента (1% площади) карты с 10 000 полигонами, с.	0,05	0,8	0,2	0,05
Скорость открытия файла с 10 000 полигонами и первой отрисовки, с.	5	3	6	45
Скорость построения триангуляции по 10 000 случайных точек, с.	2	18	–	–

В заключении кратко рассматривается возможность применения полученных в работе результатов, а также потенциальная область применения системы ГрафИн.

В приложениях приведена структура алгоритма построения триангуляции Делоне методом динамического кэширования, а также приложены акты о внедрении системы ГрафИн в различных организациях.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. На основе описанных в литературе геоинформационных систем и систем автоматизированного проектирования общего назначения выделен и обобщён набор общих функций, который должен быть реализован в универсальной графической системе широкого назначения.
2. На основании комплексного исследования задачи построения триангуляции Делоне предложена группа новых алгоритмов построения триангуляции Делоне, имеющих трудоёмкость в среднем на равномерном распределении $O(N \log N)$ и $O(N)$, и, что особенно важно, на реальных данных показывающих результаты, превосходящие по быстродействию все известные алгоритмы. Это такие алгоритмы, как полосовые алгоритмы слияния, итеративные алгоритмы с изменённым порядком добавления точек, а также итеративные алгоритмы с кэшированием поиска.
3. На основании доказанных в литературе оптимальных свойств триангуляции Делоне предложена двухпроходная стратегия, позволяющая значительно упростить и ускорить алгоритмы построения триангуляции Делоне методом слияния.
4. На основе предложенного алгоритма построения триангуляции с ограничениями предложены алгоритмы решения ряда задач пространственного анализа на плоскости, позволяющие значительно упростить и ускорить их выполнение по сравнению с аналогичными известными. Это алгоритмы решения построения оверлеев, буферных зон, взвешенных зон близости и изоконтуров.
5. На основе исследования реальной работы алгоритмов построения индексной структуры R-дерева предложена глобальная стратегия, позволяющая значительно улучшить скоростные характеристики R-деревьев при выполнении регионального поиска.
6. На основе анализа причин недостаточно высокого качества построения индексной структуры R-дерева предложен алгоритм уточнения разбиения множества объектов на подмножества с минимальным перекрытием объемлющих группы прямоугольников, позволяющий в ряде случаев улучшить качество получаемого разбиения в классических и глобальных алгоритмах построения R-деревьев.
7. Разработана универсальная графическая информационная система ГрафИн, в которой воплощены полученные в работе теоретические результаты.

ОПУБЛИКОВАННЫЕ РАБОТЫ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Жихарев С.А., Скворцов А.В. Моделирование рельефа в системе ГрафИн. // Геоинформатика: Теория и практика. Выпуск 1. – Томск: Изд-во Томск. ун-та, 1998, с. 194-205.

2. Жихарев С.А., Скворцов А.В. Построение графовых структур в ГИС и САПР. // Геоинформатика: Теория и практика. Выпуск 1. – Томск: Изд-во Томск. ун-та, 1998, с. 139-152.

3. Скворцов А.В. Глобальные алгоритмы построения R-деревьев. // Геоинформатика: Теория и практика. Выпуск 1. – Томск: Изд-во Томск. ун-та, 1998, с. 67-83.

4. Скворцов А.В. Графическая интегрированная система ГрафИн. // Томский государственный университет. Научно-технические разработки. Каталог. – Томск: Изд-во Томского ЦНТИ, 1998, с. 90.

5. Скворцов А.В. Новые технологии от ESRI на платформе Windows. // ArcReview, 1997, № 3, с. 11.

6. Скворцов А.В. Система ГрафИн. // Геоинформатика: Теория и практика. Выпуск 1. – Томск, Изд-во Томск. ун-та, 1998, с. 182-193.

7. Скворцов А.В. Универсальное программное средство для информационных систем электрических сетей. // Современные техника и технологии (труды областной конференции). – Томск: Изд-во ТПУ, 1998, с. 24-25.

8. Скворцов А.В. Ускоритель набора программ. // Информатика и образование, 1997, № 2, с. 24.

9. Скворцов А.В., Бабанов А.М., Слюсаренко С.Г. Кадастровое обеспечение задач управления городом. // Интеркарто-4 (материалы международной конференции). – Барнаул, 1998, с. 416-419.

10. Скворцов А.В., Гриценко Ю.Б. Вопросы построения универсальной графической системы, работающий с территориально определённой информацией. // Геоинформатика: Теория и практика. Выпуск 1. – Томск, Изд-во Томск. ун-та, 1998, с. 170-181.

11. Скворцов А.В., Жихарев С.А., Фукс А.Л. Применение цифровых моделей рельефа для задач планирования территории. // ИНПРИМ-98 (материалы международной конференции), часть V. – Новосибирск, 1998, с. 65.

12. Скворцов А.В., Костюк Ю.Л. Применение триангуляции для решения задач вычислительной геометрии. // Геоинформатика: Теория и практика. Выпуск 1. – Томск: Изд-во Томск. ун-та, 1998, с. 127-138.

13. Скворцов А.В., Костюк Ю.Л. Система ГрафИн как пример современной интегрированной ГИС. // Интеркарто-4 (материалы международной конференции). – Барнаул, 1998, с. 152-157.

14. Скворцов А.В., Костюк Ю.Л. Сравнительный анализ алгоритмов построения триангуляции Делоне. // SCOR-98 (материалы международной конференции). – Новосибирск, 1998, с. 138.

15. Скворцов А.В., Костюк Ю.Л. Универсальная графическая информационная система ГрафИн. // ИНПРИМ-98 (материалы международной конференции), часть V. – Новосибирск, 1998, с. 65-66.

16. Скворцов А.В., Костюк Ю.Л. Эффективные алгоритмы построения триангуляции Делоне. // Геоинформатика: Теория и практика. Выпуск 1. – Томск: Изд-во Томск. ун-та, 1998, с. 22-47.

17. Скворцов А.В., Слюсаренко С.Г. Информационная система городских коммуникаций. // ИНПРИМ-98 (материалы международной конференции), часть III. – Новосибирск, 1998, с. 71.

18. Скворцов А.В., Слюсаренко С.Г., Субботин С.А., Кобрин А.Ю. Информационное обеспечение инженерных сетей. // Геоинформатика: Теория и практика. Выпуск 1.– Томск: Изд-во Томск. ун-та, 1998, с. 206-225.

19. Скворцов А.В., Субботин С.А. Универсальная технология отображения условных знаков. // ИНПРИМ-98 (материалы международной конференции), часть V. – Новосибирск, 1998, с. 66.