

На правах рукописи

**МАЗУРОВ Николай Владимирович**

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ ПРИНЦИПОВ ПОСТРОЕНИЯ ЦЕНТРА  
ВЫСОКОПРОИЗВОДИТЕЛЬНОЙ ОБРАБОТКИ ГЕОПРОСТРАНСТВЕННЫХ  
ДАННЫХ**

Специальность 25.00.35 –  
Геоинформатика

**Автореферат  
диссертации на соискание ученой степени  
кандидата технических наук**

Москва - 2005



Работа выполнена на кафедре информационно-измерительных систем Московского государственного университета геодезии и картографии (МИИГАиК)

Научный руководитель: доктор технических наук, профессор

**Майоров Андрей Александрович**

Официальные оппоненты: доктор технических наук, профессор

**Троицкий Владимир Иванович**

кандидат технических наук

**Негурица Дмитрий Леонидович**

Ведущая организация: Государственный научно-исследовательский институт информационных технологий и телекоммуникаций «Информика»


Защита состоится «29» декабря 2005 г. в 12 час. на заседании диссертационного совета Д212.143.03 при Московском государственном университете геодезии и картографии по адресу: 105064, Москва К-64, Гороховский переулок, д. 4, МИИГАиК, аудитория 321.

С диссертацией можно ознакомиться в библиотеке МИИГАиК

Автореферат разослан «\_\_» ноября 2005 г.

Ученый секретарь

диссертационного совета



Климков Ю.М.

2006-4  
25152

2240430

### Общая характеристика работы

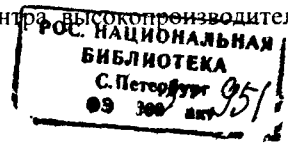
**Актуальность темы исследования.** В настоящее время в науках о Земле информационно-вычислительные системы приобретают решающее значение. В области обработки геопространственных данных требования к мощности и производительности вычислительных систем растут опережающими темпами. Современные потребности в результатах обработки различных геопространственных данных способствуют появлению новых, ресурсоемких методов решения поставленных задач.

К сожалению, современный этап развития информационных технологий не позволяет обеспечить всех заинтересованных в высокопроизводительной обработке геопространственных данных сторон необходимым программным обеспечением и аппаратными средствами для удовлетворения их потребностей. Это связано с дороговизной вычислительных систем, программного обеспечения, отсутствием квалифицированного персонала и многими другими факторами.

Для удовлетворения потребностей различных организаций в результатах обработки геопространственных данных целесообразно создавать специализированные вычислительные центры, предоставляющие коллективный доступ к ресурсам центра. Это позволит решить ряд проблем, возникающих при организации процесса обработки геопространственных данных самими организациями.

В процессе создания центра высокопроизводительной обработки геопространственных данных может возникнуть ряд сложностей, связанных со спецификой задач обработки геопространственных данных. Стандартные подходы к созданию подобного центра не будут удовлетворять выдвигаемым требованиям. Поэтому при проектировании и создании центра необходимо учитывать специфику решаемых задач, требования к скорости и производительности вычислительных систем, объемы обрабатываемой информации и многие другие.

**Целью настоящей работы** является разработка и исследование принципов построения специализированного центра высокопроизводительной обра-



ботки геопространственных данных, позволяющего решать ресурсоемкие задачи обработки геопространственных данных, отвечающего требованиям производительности, надежности, безопасности и ряду другим, описанным в диссертации, являющихся неотъемлемой частью системы обработки геопространственных данных. Кроме того, центр высокопроизводительной обработки геопространственных данных должен обеспечивать доступ к ресурсам удаленным пользователям с высокой степенью доступности и, в то же время, высоким уровнем безопасности.

#### **Задачи работы:**

1. Определить специфику геопространственных данных, влияющих на методику построения центра высокопроизводительной обработки геопространственных данных.
2. Описать модели процессов обработки, передачи и хранения геопространственных данных.
3. Проанализировать существующие системы высокопроизводительной обработки данных, в том числе и геопространственных.
4. Разработать методику и принципы построения центра высокопроизводительной обработки геопространственных данных.
5. Разработать структуру центра высокопроизводительной обработки геопространственных данных.
6. Исследовать и выбрать технологии, программное и аппаратное обеспечение, позволяющее наиболее рационально решить поставленные задачи.
7. Экспериментально проверить положения, предложенные в диссертационной работе.

**Предметом исследования** является исследование принципов построения центра высокопроизводительной обработки геопространственных данных для задач обработки данных дистанционного зондирования Земли (ДЗЗ), создания и

обновления картографических продуктов, обработки данных наземных полевых исследований, хранения и предоставления доступа к геопространственной информации, прикладных задач, решаемых с использованием геоинформационных систем и многих других.

**Объектом исследования** выбраны существующие методы обработки геопространственной информации, алгоритмы, модели и методики повышения производительности вычислительных систем в целом и систем обработки геопространственной информации в частности.

Результаты проведенных исследований были апробированы на базе Центра высокопроизводительных вычислений МИИГАиК.

**Теоретической и методической основой исследования** послужили отечественные и зарубежные разработки по геоинформатике, высокопроизводительной обработке данных, вычислительным системам и комплексам, сетевым технологиям.

В качестве базы разрабатываемого центра послужил Центр высокопроизводительной обработки геопространственных данных МИИГАиК. При экспериментальной апробации результатов исследований были задействованы программные и аппаратные ресурсы центра.

**Научная новизна диссертационного исследования** заключается в разработке принципов создания специализированного центра для высокопроизводительной обработки геопространственных данных.

Научную новизну представляют следующие результаты:

- определение специфики, влияющей на обработку геопространственных данных;
- принципы построения специализированного вычислительного центра для обработки геопространственных данных;
- методика повышения эффективности работы компонентов центра высокопроизводительной обработки геопространственных данных;

- методика эффективной передачи спутниковых снимков по глобальным сетям передачи данных;
- оценка эффективности применения методов высокопроизводительной обработки для геопространственных данных.

**Основные положения выносимые на защиту:**

- обоснование целесообразности построения специализированного центра высокопроизводительной обработки геопространственных данных с возможностью предоставления удаленного доступа заинтересованным организациям;
- методы организации эффективной передачи растровых и векторных изображений через сети передачи данных;
- методы организации эффективного хранения больших объемов геопространственной информации;
- выбор технологий, стандартов, программного и аппаратного обеспечения для построения специализированного центра высокопроизводительной обработки геопространственных данных;
- структура центра высокопроизводительной обработки геопространственных данных.

**Практическая значимость диссертационного исследования.** Рассматриваемый центр высокопроизводительной обработки геопространственных данных имеет большую практическую значимость. Он будет полезен в различных научно-исследовательских и прикладных проектах и программах, связанных с обработкой геопространственной информации, такой как результаты дистанционного зондирования Земли, наземной геодезической и топографической съемки, при создании и обновлении картографических материалов, решении прикладных задач как, например, поиск полезных ископаемых или ведение кадастрового учета и многих других приложений. Использование центра позволит существенно сократить время получения результатов обработки геопр-

странственных данных. Также центр может быть использован в демонстрационных и учебных целях.

**Публикации результатов исследования.** Основные положения и результаты диссертационной работы отражены в 5 публикациях. Некоторые положения, выдвинутые в диссертационной работе, были реализованы в Центре высокопроизводительной обработки геопространственных данных МИИГАиК.

**Апробация результатов исследования.** Основные положения диссертации докладывались и обсуждались:

- на Международной конференции Geomatics 2004 (Куба, Гавана, 8-12 мая 2004 г.);

- на Международной научно-технической конференции, посвященной 225-летию МИИГАиК (Россия, Москва, 24-27 мая 2004 г., Московский государственный университет геодезии и картографии);

- на 57-й научно-технической конференции студентов, аспирантов и молодых ученых МИИГАиК (Россия, Москва, 8-9 апреля 2002 г., Московский государственный университет геодезии и картографии);

- на 58-й научно-технической конференции студентов, аспирантов и молодых ученых МИИГАиК (Россия, Москва, 4-5 апреля 2003 г., Московский государственный университет геодезии и картографии);

- на 59-й научно-технической конференции студентов, аспирантов и молодых ученых МИИГАиК (Россия, Москва, 9-10 апреля 2004 г., Московский государственный университет геодезии и картографии);

- на 60-й научно-технической конференции студентов, аспирантов и молодых ученых МИИГАиК (Россия, Москва, 11-12 апреля 2005 г., Московский государственный университет геодезии и картографии);

**Объем и структура диссертации.** Работа состоит из введения, четырех глав, заключения, библиографического списка литературы (90 наименований) и

приложений. Диссертация содержит 119 стр. основного текста вместе с 38 рис. и 3 табл.

### Краткое содержание работы

#### Введение

Во введении обоснована актуальность темы диссертационной работы, определены цели, поставлены задачи исследования, намечены основные этапы выполнения диссертационных исследований.

**Глава 1. Исследование существующих методов и технологий создания высокопроизводительных центров для решения задач обработки геопро-  
странственных данных.**

Дано определение геопро-  
странственных данных и геоинформационных систем. Описана концептуальная математическая модель систем обработки геопро-  
странственных данных на базе теории информации.

Показано, что в общем виде систему обработки геопро-  
странственных данных можно представить в виде группы подсистем. Каждое преобразование под-  
систем можно представить оператором:

$$\bar{x} = P \bar{a}, \quad (1)$$

$$\bar{\beta} = \Phi \bar{x}, \quad (2)$$

$$\bar{\beta}^* = \Omega \bar{\beta}, \quad (3)$$

$$\bar{r}^* = \Psi \bar{\beta}^*, \quad (4)$$

Где  $P$  – оператор первичного преобразования,  $\Phi$  – оператор представления,  $\Omega$  – оператор обработки,  $\Psi$  – оператор хранения и передачи.

В общем случае  $\bar{r}$  отличается от  $\bar{x}$ , и поэтому преобразования, выполняемые всей системой могут, быть представлены в виде:

$$\bar{r} = F \bar{x}.$$



Оператор F характеризует основную задачу обработки геопространственных данных и является ее математической моделью при обработке данных без учета искажений.

Системой обработки геопространственной информацией можно управлять. Управляющие воздействия изменяют внутреннее состояние системы и представляются:

$$\gamma_{t_1} = \varepsilon(\alpha, \gamma, t),$$

где  $\gamma$  - состояние системы в момент времени  $t$  и  $t_1$ ,  $\varepsilon$  - оператор изменения состояния системы обработки геопространственных данных,  $t$  - время.

Общие выражения математической модели системы обработки геопространственных данных можно представить в виде:

$$\bar{r} = G(\alpha, \gamma),$$

$$\gamma_{t_1} = \varepsilon(\alpha, \gamma, t),$$

Основной задачей систем обработки геопространственных данных является представление пользователю оценок геопространственных данных  $\bar{r}^*$ , позволяющих извлечь интересующую информацию.

Любое геопространственное сообщение  $\bar{x}$ ,  $\bar{\beta}$ ,  $\bar{\beta}^*$  и  $\bar{r}^*$  можно представить в виде многомерного вектора  $\bar{V}$  с координатами различной длины  $\bar{v}_k$ ,  $k = 1, 2, \dots, n$ , кодируемые цифрами и знаками. Если слово содержит  $m_{qk}$  цифровых знаков, то объем сообщения

$$J = \sum_{k=1}^n m_{qk} \quad \text{цифровых знаков}, \quad (5)$$

$q$  – основание цифровой системы.

Координата  $\bar{v}_k$  может принимать  $M_k$  значений и кодируется простой последовательностью цифр. Если на одно  $k$ -е сообщение тратится  $m_{qk}$  цифровых символов, то их число будет равно  $q^{m_{qk}}$ . Отсюда максимальная численность различных кодов для  $\bar{v}_k$  будет соответствовать  $M_k = q^{m_{qk}}$ . Тогда  $m_{qk} = \log_q M_k$  и по формуле (5) имеем:

$$J = \sum_{k=1}^n \log_q M_k \text{ цифровых знаков.} \quad (6)$$

Формулу (6) можно выразить через двоичные знаки

$$J = \sum_{k=1}^n m_{qk} \log_2 q \text{ двоичных знаков.} \quad (7)$$

Был сделан вывод, что объем геопространственных сообщений (6) и (7), не зависящий от смыслового содержания, позволит вычислить пропускную способность системы обработки геопространственных данных как максимально возможный поток первичных сообщений  $\bar{x}$ , которые данная система может обработать, обеспечивая заданную оценку  $\bar{r}^*$ .

Здесь же дан обзор задач обработки геопространственных данных, требующих значительных вычислительных ресурсов, таких как предобработка спутниковых снимков (атмосферная коррекция, сегментация и др.). Показаны вычислительные затраты на решения этих задач. Исходя из этого автор считает необходимым применение высокопроизводительных систем, таких как суперкомпьютеры и вычислительные кластеры для решения подобных задач.

Исходя из рассмотренных задач обработки геопространственных данных, была поставлена задача создания центра обработки данных, отвечающего требованиям по скорости работы, объемам обрабатываемой и хранимой информации, требованиям к надежности и защите данных, а также экономической целесообразности и выгоде.

## **Глава 2. Разработка модели центра высокопроизводительной обработки геопространственных данных.**

Была предложена структура центра высокопроизводительной обработки геопространственных данных (Рис. 1). Центр состоит из нескольких модулей: модуль загрузки данных, модуль обработки данных, модуль хранения данных и модуль доступа к обработанной информации. Доступ к ресурсам центра может быть осуществлен как через Интернет, так и из локальной сети университета.

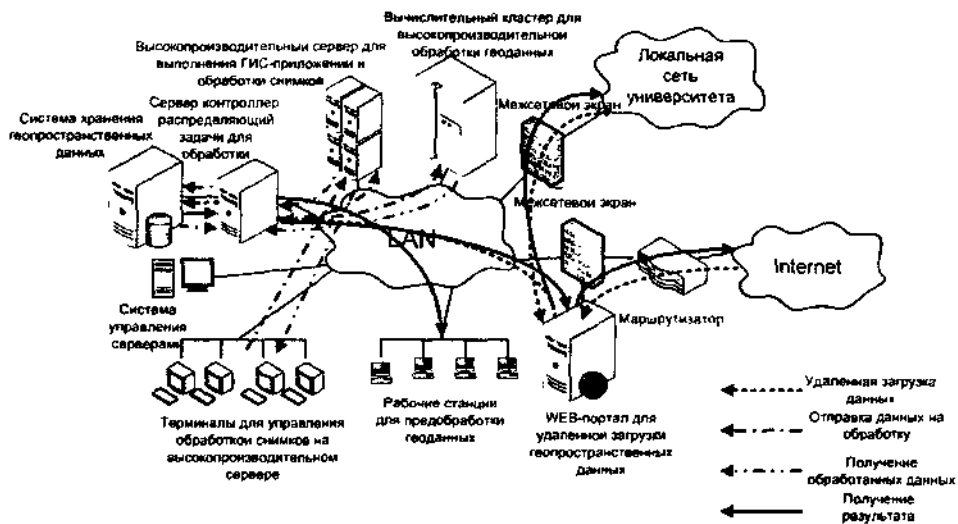


Рис. 1. Общая структура центра высокопроизводительной обработки геопроизводственных данных

Была предложена концепция мультимасштабных изображений (Рис. 2), позволяющих оптимизировать процесс загрузки растровых и векторных изображений через коммуникационные сети с низкой пропускной способностью.

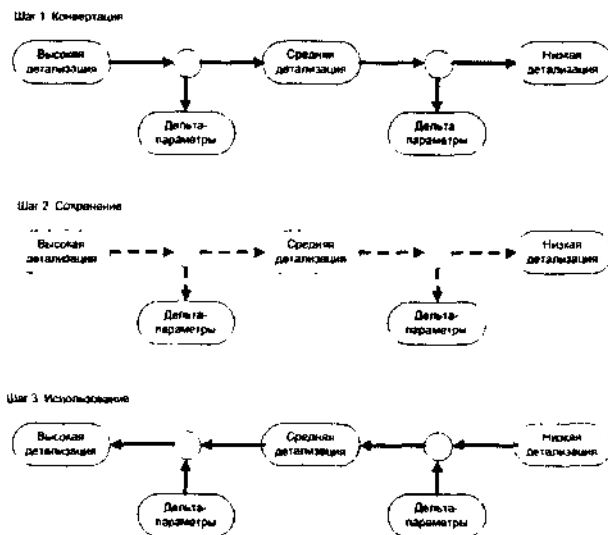


Рис. 2. Процедура использования мультимасштабных снимков

Для растровых изображений схема имеет следующий вид. На первом этапе изображение подвергается сжатию при помощи алгоритма сжатия. При этом получается изображение с меньшей разрешающей способностью, но меньшее по объему. “Потерянная” информация сохраняется в виде дельта-параметров, которые в последствии помогут восстановить снимок до исходного разрешения. Эта процедура повторяется до тех пор, пока не получается изображение с минимальным разрешением, которое целесообразно использовать.

На втором этапе полученный окончательный снимок, имеющий небольшой объем, но и слабую разрешающую способность, сохраняется в базе данных, при этом сохраняются соответствующие этому изображению дельта-параметры.

На третьем этапе инициатор запроса указывает разрешающую способность, необходимую для решения конкретной задачи. После этого из базы данных извлекается необходимое изображение и восстанавливается до нужной разрешающей способности. При восстановлении изображения используются дельта-параметры.

При создании мультимасштабных форматов векторных изображений основной задачей является минимизация количества объектов на изображении, и тем самым уменьшение объема, занимаемого файлом. При этом выделяется базовое изображение, минимального объема, но недостаточно информативное. В процессе уточнения изображения к нему будут подгружаться остальные объекты, улучшающие качество изображения (Рис. 3).

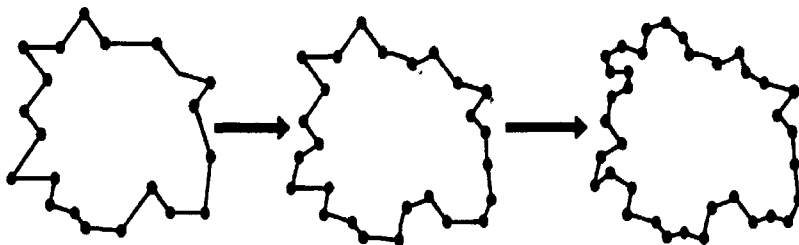


Рис. 3. Динамическая загрузка векторного изображения

В работе предложен способ определения базисных вершин, составляющих каркас объекта. Предложен метод ранжирования вершин по их наименьшему влиянию на очертание объекта. Метод основан на принятии за наименее важную вершину, образующую треугольник с минимальной площадью.

Таким образом загрузка изображения будет иметь вид (Рис. 4):

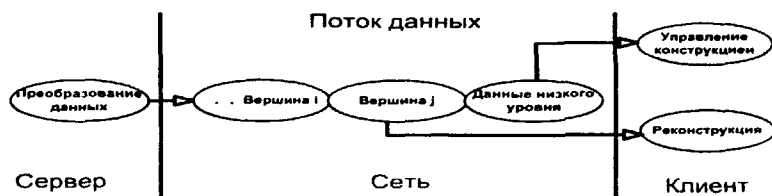


Рис. 4. Загрузка изображения по сети

Алгоритм организации хранения растровых и векторных изображений представлен на рис. 5.



Рис. 5.

Приведена модель оценки отказоустойчивости центра высокопроизводительной обработки геопространственных данных как информационной системы.

### **Глава 3. Разработка методики создания центра высокопроизводительной обработки геопространственных данных.**

В работе приведена система критериев для выбора программных и аппаратных средств для построения центра высокопроизводительной обработки геопространственных данных. Показано, что выбор той или иной аппаратной платформы и программных средств определяется рядом критериев, к которым относятся:

- отношение стоимость/производительность;
- надежность и отказоустойчивость;
- масштабируемость;
- совместимость.

Удаленный доступ к ресурсам центра со стороны заинтересованных организаций целесообразно осуществлять в двух режимах. Первый – пакетный режим, подразумевает посылку формализованных данных по электронной почте, т.е. данные посылаются электронным письмом определенного формата. В письме устанавливаются флаги, несущие служебную информацию о самой задаче, методе решения, необходимой точности и другая необходимая информация. Второй режим подразумевает загрузку данных через WEB-интерфейс. При этом режиме сотрудник организации, которой необходимо обработать данные, заходит на специализированный сайт, имеющий определенную форму, выбирает метод, параметры обработки данных и загружает их на сервер центра.

Была создана система загрузки данных через интернет и визуализации пространственных данных на базе web-приложений с использованием технологий MySQL и PHP (Рис. 6).

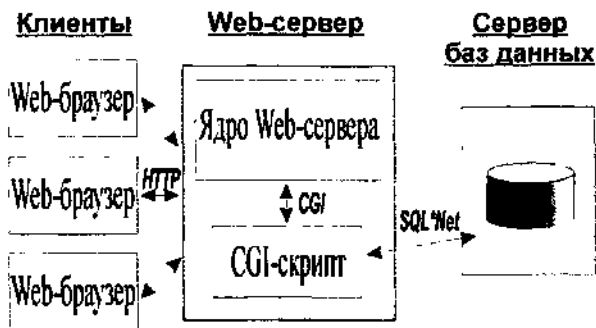


Рис. 6. Клиент-серверная архитектура

Созданная система позволяет загружать данные через web-интерфейс (Рис. 7), используя обычный браузер. Эти данные сохраняются в базе данных MySQL. После этого они доступны для обработки любыми приложениями через стандартные интерфейсы и SQL запросы.

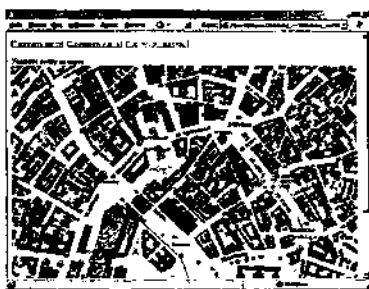


Рис. 7. Веб-интерфейс системы загрузки данных

Обработанные данные также могут быть визуализированы посредством web-интерфейса.

Были рассмотрены вопросы обеспечения безопасности данных и повышения уровня отказоустойчивости компонентов центра. Сформулированы основные требования к обеспечению информационной безопасности.

В качестве метода защиты авторских прав на цифровые карты и защиты их от незаконного копирования были предложены цифровые водяные знаки, которые не заметны на карте в обычном состоянии, но при применении специализированных средств из карты извлекается дополнительная информация, в том числе и об авторах продукта.

При внедрении в карту цифровых водяных знаков можно пользоваться особенностями формата карты или использовать избыточность форматов. Такая встроенная информация будет иметь небольшой объем, но этого достаточно, так как все, что требуется, это встроить информацию о создателях карты, лицах, уполномоченных распространять данный продукт и другую, не очень объемную информацию. Целесообразно всю информацию встраивать не в явном виде, а применяя криптографические методы.

#### **Глава 4. Экспериментальные исследования принципов создания центра высокопроизводительной обработки геопространственных данных.**

Для подтверждения принципов изложенных в диссертационной работе, был проведен эксперимент по обоснованию эффективности предложенного способа доступа к геоданным векторных форматов через Интернет. Целью эксперимента являлось выяснение влияния предложенного метода генерализации на очертание объектов после упрощения векторного изображения.

Основным пунктом, который требует обоснования, является то, что при удалении большого числа вершин для сокращения объема загружаемой первоначально информации объекты на изображении теряют свою форму не существенно. Безусловно, при удалении каких-то вершин изображение теряет свою информативность. Но эти потери не существенны по сравнению с выигрышем в объеме передаваемого файла. Для определения количественных показателей различия преобразованного и исходного изображения необходимо выбрать правильный показатель для сравнения. В качестве меры для сравнения изображений на различных стадиях генерализации применялась площадь полигонов. Этот параметр был принят как наиболее полно описывающий изменения объектов при трансформации.

Кроме того, в результате эксперимента предполагалось наглядно показать отсутствие существенных различий в трансформированных изображениях, влияющих на восприятие объектов.



Для проведения эксперимента было сформировано векторное изображение, содержащее 3108 полигонов. Оно было получено путем объединения 133644 вершин. Координаты вершин импортировались из файла ArcInfo. Исходный файл в три этапа преобразовывался по алгоритму, описанному в работе. Причем на первой стадии было удалено 50% вершин от начального количества. На втором этапе было удалено 80% вершин от начального количества. Оставшиеся вершины образовали новое изображение. Результаты этих преобразований представлены на рис. 8 – 10.



Рис. 8. Исходное изображение



Рис. 9. Изображение после первой трансформации (удалено 50% вершин)



Рис. 10. Изображение после второй трансформации (удалено 80% вершин)

При сравнении рис. 8 и 10 видно, что одинаковые полигоны имеют схожие очертания, позволяющие их идентифицировать. Но визуального сравнения не достаточно для того, чтобы утверждать эффективность предлагаемого метода. Нужно формализовать результаты эксперимента.

Для формализации эксперимента был применен метод сравнения количества полигонов, имеющих одинаковые площади. Для этого на всех трех изображениях были посчитаны площади полигонов. Все полигоны были разбиты в соответствии с площадями на категории 0 – 0.01, 0.01 – 0.05, 0.05 – 0.1, 0.1 – 0.3, 0.3 – 0.6, 0.6 – 0.8, 0.8 – 1.0, 1.0 – 2.0, 2.0 – 5.0, 5.0 – 7.0, 7.0 – 10.0, 10.0 – 50.0 квадратных километров. Также был произведен анализ 24-х случайно выбранных полигонов в процессе преобразований.

Анализ поведения полигонов в процессе преобразований показан на рис. 11.

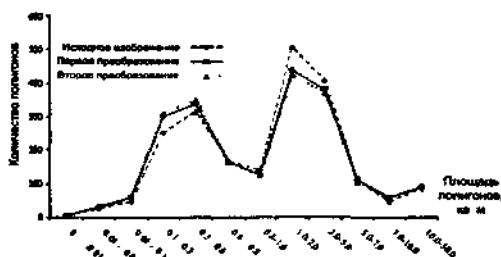


Рис. 11. Изменение количества полигонов с разными площадями

Анализ показал, что наибольшая разница площадей из выбранных диапазонов наблюдается для площадей полигонов 1.0 – 2.0 кв. км. Значения для этого диапазона получились следующие: 497 (исходное изображение), 443 (изображение после первого преобразования) и 431 (изображение после второго преобразования). В процентном отношении максимальная разница по выделенным диапазонам площадей между исходным изображением и изображением, полученным после первого преобразования, составляет 10,86%, а полученным после второго преобразования 13,28%. В целом для всех полигонов разница между количеством полигонов с одинаковыми площадями составила 5,7% для первого преобразования и 6,5% для второго.

Также проводился анализ поведения 24-х полигонов, относящихся к разным диапазонам площадей. Результаты анализа представлены на рис. 12.

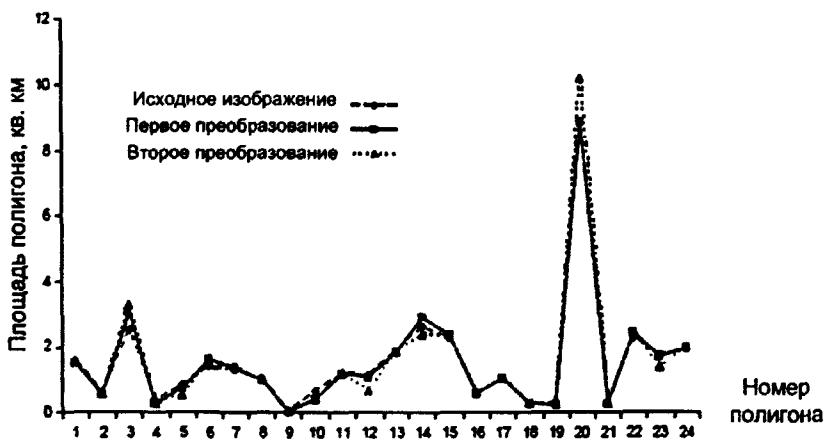


Рис. 12. Анализ произвольных полигонов

Анализ также показывает незначительную разницу между площадями полигонов при применении процедуры генерализации.

Эксперимент показал, что предложенный метод эффективного доступа к изображениям векторных графических форматов показывает хорошие результаты. При изменении количества вершин, образующих полигоны, в 5 раз, а сле-

довательно и объема векторного изображения тоже примерно в 5 раз, характеристики полигонов изменяются не существенно (6,5% по площади полигона).

### **Заключение**

В результате проведенных исследований в соответствии с поставленными целью и задачами автором сформулированы общие итоги диссертационной работы.

1. Рассмотрена специфика геопространственных данных, влияющих на методику построения центра высокопроизводительной обработки геопространственных данных.

2. Описаны модели процессов обработки, передачи и хранения геопространственных данных.

3. Разработаны принципы построения и методика создания центра высокопроизводительной обработки геопространственных данных.

4. Разработана структура центра высокопроизводительной обработки геопространственных данных.

5. Выбраны технологии, программное и аппаратное обеспечение, позволяющее наиболее рационально решить поставленные задачи.

7. Экспериментально проверены и подтверждены основные теоретические результаты, полученные в диссертационной работе.

### **Список опубликованных работ автора по теме диссертации**

1. Diseño de un centro, de alto rendimiento, para el procesamiento de datos geograficos (Geodatos), A. Mayorov, N. Mazurov, J. Vila Ortega, Resumen general de 10 Convencion Internacional y Feria Informatica 2004, 4 Congreso internacional geomatica 2004, Havana, Cuba, 2004. – с. 87 - 89.

2. Защита геопространственной информации в локальных вычислительных сетях с использованием firewall-сервера на базе маршрутизатора CISCO PIX-515, Мазуров Н.В., Известия ВУЗов, Геодезия и аэрофотосъемка, специальный выпуск, Москва, 2002, с. 133 – 141.

3. Принципы организации web-ориентированных баз геопространственных данных, Мазуров Н.В., Известия ВУЗов, Геодезия и аэрофотосъемка, Москва, 2004, с. 114 – 118.

4. Проектирование центра высокопроизводительной обработки геопространственных данных Московского государственного университета геодезии и картографии – Савиных В.П, Майоров А.А., Малинников В.А., Кондауров И.Н., Мазуров Н.В. – Сборник трудов международной научно-технической конференции, посвященной 225-летию МИИГАиК, Москва, 2004. – с. 83 – 88.

5. Возможность использования цифровых водяных знаков для защиты векторных и растровых графических форматов от незаконного копирования, Мазуров Н.В., - Известия вузов, Геодезия и аэрофотосъемка, Москва, 2005. - с. 231 – 237.



1

2

№23754

РНБ Русский фонд

2006-4  
25152

Подписано в печать 24.11.2005. Гарнитура Таймс  
Формат 60х90/16. Бумага офсетная. Печать офсетная.  
Печ. л. 1,5. Уч.-изд. л. 1,5; Тираж 80 экз. Заказ №223 Цена договорная  
УПП «Репрография» МИИГАиК, 105064, Москва, Гороховский пер., 4