**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc5272923)

[1 ТЕХНОЛОГИЯ SSCCIP 5](#_Toc5272924)

[1.1 Требования 5](#_Toc5272924)

[1.2 Композиция 6](#_Toc5272924)

1.2.1 Клиент-серверное взаимодействие7

1.2.2 Клиентская компонента8

1.2.3 Серверная компонента8

1.3 Структурная схема программного комплекса9

1.4 Вывод10

[2 АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ ИЗОБРАЖЕНИЯ НА ОСНОВЕ МНОГОМЕРНЫХ ГИСТОГРАММ 11](#_Toc5272925)

2.1 Основные принципы построения алгоритма11

2.2 Особенности классификации по текстурным признакам12

2.3 Принципы выбора признаков13

2.4 Управление программой14

[ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ НА GPU 15](#_Toc5272930)

[ЗАКЛЮЧЕНИЕ 16](#_Toc5272931)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 17](#_Toc5272932)

сПисок определений обозначений и сокращений

ПО – программное обеспечение

ДЗЗ – дистанционное зондирование Земли

ГИС – геоинформационная система

ДДЗЗ – данные дистанционного зондирования Земли

ОС – операционная система

ПК – персональный компьютер

ЭВМ – электронно-вычислительная машина

SFTP – Secure FTP – защищенный протокол передачи файлов на основе SSH

GPU - graphics processing unit – графический процессор

# **ВВЕДЕНИЕ**

Целью преддипломной практики является изучение некоторых основных используемых на данный момент алгоритмов обработки данных дистанционного зондирования Земли.

Для достижения цели в качестве задач предполагается рассмотрение технологии SSCCIP, решающий проблемы взаимодействия ЭВМ и пользователя, обзор алгоритма обработки ДДЗЗ путем автоматической классификации изображений на основе многоспектральной гистограммы, а также с помощью графических процессоров.

.

1. **ТЕХНОЛОГИЯ SSCCIP**

Привлечение высокопроизводительных ЭВМ к решению трудоемких задач обработки данных ДЗЗ неизбежно влечет проблему "враждебности" супер-ЭВМ к пользователю: как правило, многопроцессорная ЭВМ управляется операционной системой семейства Unix и функционирует как вычислительный сервер, взаимодействуя с внешними клиентами через защищенное SSH-соединение; распространенные же программы удаленного SSH-доступа (типа PuTTy или SSH Secure Shell) предоставляют пользователю лишь сервис командной строки [1]. Это приводит к тому, что использование многопроцессорных ЭВМ оказывается доступным только для пользователей с опытом работы в Unix и недоступным рядовому пользователю ОС семейства Windows, которыми оснащено подавляющее большинство ПК в нашей стране. И даже с учетом этого обработка данных на многопроцессорной ЭВМ выполняется в несколько приемов: загрузка данных с помощью SFTP-клиента, запуск команды обработки на удаленной машине (с помощью SSH-терминала), закачка на локальную машину результатов (также с помощью SFTP-клиента).

Таким образом, для массового привлечения супер-ЭВМ к решению трудоемких задач обработки изображений необходимо создавать не только программные средства, реализующие широкий круг методов и алгоритмов обработки изображений на многопроцессорной технике, но также технологии, упрощающие трудоемкий для рядового пользователя процесс использования удаленной многопроцессорной Unix-машины. Попыткой решения этой задачи стал рассматриваемый в настоящей работе экспериментальный программный комплекс SSCCIP.

1.1 Требования

Командой разработчиков комплекса были сформулированы следующие требования к разрабатываемой программной системе SSCCIP:

* Система должна представлять собой автоматизированное рабочее место оператора под управлением ОС Windows и обеспечивать сервис высокопроизводительной обработки данных ДЗЗ на удаленной многопроцессорной ЭВМ.
* Система должна обеспечивать визуализацию исходных данных и результатов обработки.
* Система должна обеспечивать возможность интерактивного формирования задания оператора на обработку.
* Система должна обеспечивать полностью прозрачное для оператора выполнение вычислительного задания на многопроцессорной ЭВМ.
* Система должна обеспечивать возможность своей настройки на конкретную вычислительную среду (многопроцессорную ЭВМ, вычислительную компоненту, систему управлениями пакетными заданиями и пр.).
* Система должна по возможности обеспечивать максимальную легкость расширения новыми алгоритмами.

Последнее требование представляется особенно важным, так как легкость наращиваемости подобного рода систем и их адаптации к потребностям конкретных пользователей являются необходимыми условиями возможности их дальнейшего развития [2].

1.2 Композиция

Вычислительная система SSCCIP состоит из трех основных компонент:

* Клиентской, функционирующей на клиентской ЭВМ и управляемой пользователем.
* Серверной, функционирующей на удаленной многопроцессорной ЭВМ и выполняющей непосредственно вычисления.
* Коммуникационной, связывающей клиентскую и серверную компоненты защищенным каналом.

1.2.1 Клиент-серверное взаимодействие

Для подавляющего большинства многопроцессорных вычислительных серверов единственным разрешенным способом клиентского сетевого подключения является защищенный SSH-канал. Такой канал в системе SSCCIP обеспечивается:

Со стороны сервера: системным SSH-демоном, поддерживающим прикладные SSH-протоколы SFTP приема-передачи файлов и rexec удаленного выполнения программ/команд на сервере.

Со стороны клиента: библиотекой CryptLib, созданной Питером Гутманном в университете Окленда (Новая Зеландия) [3]. На базе данной библиотеки авторами реализованы прикладные SSH-протоколы SFTP и rexec.

На верхнем уровне клиентская компонента взаимодействует с вычислительным сервером по следующей схеме:

* Клиентская компонента закачивает на вычислительный сервер файлы с исходными данными и файл с описанием задания (далее «файл-задание»), включающий указание имен исходных и результирующих файлов, алгоритмов обработки и их параметров.
* Клиентская компонента запускает (ставит в очередь задач) на сервере серверную компоненту и ожидает завершения вычислений (с некоторым интервалом опрашивает очередь задач многопроцессорной ЭВМ).
* Серверная компонента разбирает файл-задание и выполняет указанную в нем операцию над указанными данными, сохраняя результат в указанных файлах.
* После завершения вычислений клиентская компонента скачивает результаты обработки с вычислительного сервера на клиентскую ЭВМ и выполняет  их визуализацию.

1.2.2 Клиентская компонента

Клиентская компонента представляет собой Windows-приложение SSCCIP\_Client, реализованное на языке C++ . Ее код организован в виде технологического каркаса (framework), обеспечивающего легкость введения новых операций обработки: код, выполняющий общие для всех операций обработки действия (загрузку файлов данных и задания на сервер, запуск и отслеживание задания, загрузка результатов с сервера), вынесен в обобщенную подпрограмму обработки executeParallelTask(); для кодирования конкретной операции обработки необходимо реализовать только действия, специфичные для данной операции: интерактивное конфигурирование операции (выбор обрабатываемых файлов и файлов для сохранения результата, параметризация алгоритма обработки) и визуализацию результата обработки. Для этого в каркасе объявлен интерфейс IParallelTask с двумя методами: Configure() и Visialize(), которые должен определить реализующий данный интерфейс конкретный класс обработки. Метод Configure(), возвращает содержание файла-задания, список файлов для загрузки на сервер перед вычислениями и список файлов для загрузки с сервера после вычислений. С помощью метода Visialize(), вызываемого каркасом после загрузки с сервера результатов вычислений, конкретная операция обработки выполняет визуализацию результатов. Для выполнения конкретной операции обработки необходимо параметризовать обобщенную операцию executeParallel Task() классом конкретной операции.

Каркас также содержит средства визуализации изображений различных графических форматов.

1.2.3 Серверная компонента

Серверная компонента представляет собой Unix-приложение SSCCIP\_Server, разработанное на языке C++ с использованием созданной авторами библиотеки высокопроизводительной обработки изображений на многопроцессорной ЭВМ SSCC\_PIPL [4]. SSCCIP\_Server – это MPI-приложение, запускаемое на каждом вычислительном узле многопроцессорной ЭВМ загрузчиком MPI mpirun. Код SSCCIP\_Server также реализован с использованием идеологии технологического каркаса. В частности, для практически важных типов обрабатывающих процедур «пиксел-пиксел» и «окрестность-пиксел» созданы обобщенные подпрограммы, выполняющие общие для всех операций обработки действия (разбор файла-задания, загрузка исходного изображения с поддержкой множества графических форматов, определение параметров распараллеливания и собственно распараллеливание вычислений, сохранение результата обработки в графическом файле). Конкретная операция обработки реализуется классом алгоритма, интерфейс которого согласован с библиотекой SSCC\_PIPL.

1.3 Структурная схема виртуального кластера SSCCIP

Виртуальный кластер программного комплекса SSCCIP используется для обработки данных физических экспериментов, осуществляемых в Институте Ядерной Физики Сибирского Отделения Российской Академии Наук. Задачи характеризуются использованием однопоточных программ и параллелизацией на уровне данных. Для создания виртуальных ресурсов на локальном кластере используются штатные возможности PBS Pro. PBS Pro запускает требуемую виртуальную машину, описанную в скрипте, и останавливает ее по завершению задания, после чего вычислительный узел может использоваться в обычном режиме другими заданиями. В качестве среды виртуализации используется KVM. Обмен данными между Институтом и программным комплексом осуществляется через суперкомпьютерную сеть Новосибирского Научного Центра (10 Гбит/с).

На рисунке 1 изображена схема виртуального кластера SSCCIP.

C:\Users\User\Downloads\Untitled Diagram (1).png

Рисунок 1 – Структурная схема виртуального кластера программного комплекса

1.4 Вывод

Общим результатом работ по разработке программного комплекса SSCCIP стало создание технологического каркаса программного комплекса высокопроизводительной обработки данных ДЗЗ на удаленной многопроцессорной ЭВМ. В рамках данного комплекса могут решаться различные задачи обработки и анализа данных ДЗЗ, причем от разработчика технологии требуется лишь кодирование деталей, специфичных для погружаемого алгоритма.

1. **АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ ИЗОБРАЖЕНИЯ НА ОСНОВЕ МНОГОМЕРНОЙ ГИСТОГРАММЫ**

Изучаемая программа предназначена для автоматической классификации многоспектральных и текстурных изображений. Данная программа разработана Сидоровой Валерией Сергеевной в лаборатории Обработки Изображений Института Вычислительной Математики и Математической геофизики Сибирского отделения Российской Академии Наук.

2.1 Основные принципы построения алгоритма

Векторное пространство признаков  классифицируется предложенным новым алгоритмом кластерного анализа с применением оценок качества классификации. Под качеством понимается изолированность полученных унимодальных кластеров. В основе нового алгоритма лежит быстрый непараметрический метод Нарендры, разделяющий векторное пространство на унимодальные кластеры, соответствующие локальным максимумам гистограммы. Он не использует никаких предположений о функциях распределения различных классов, не требует никаких априорных данных (число кластеров, количество итераций и т.д.). Поиск максимумов ведется методом наискорейшего спуска с учетом дискретности данных. Быстрота алгоритма достигается за счет представления векторов в форме линейного списка, упорядоченного по их возрастанию. Алгоритм Нарендры предполагал предварительное квантование векторного пространства, но число уровней квантования задавалось произвольно [5].

Новый алгоритм автоматизирует выбор числа уровней квантования, основываясь на оценке изолированности кластеров по предложенной мере. Строится ряд классификаций методом Нарендры для различного числа уровней квантования, и из этого ряда  автоматически выбираются лучшие распределения, соответствующие минимумам меры изолированности кластеров. Такой подход позволяет получить лучшие распределения по унимодальным кластерам, сократив при этом их число.

Предложенная индикаторная мера оценки качества распределений основана на вычислении среднего отношения значения многомерной гистограммы на границе унимодального кластера и в области модального вектора.  Показано, что она позволяет корректно оценить изолированность кластеров  данных высокой плотности, характерной для дистанционного зондирования. Мера соответствует требованиям  достоверности неконтролируемой классификации, в отличие от традиционно применяемых мер, основанных на соотношении дисперсий и расстояний между кластерами, которые могут быть корректно использованы лишь в задачах контролируемой классификации для оценки разделимости классов.

2.2 Особенности классификации по текстурным признакам

Статистические текстурные признаки являются пространственными характеристиками, а не точечными в отличие от спектральных, поэтому есть некоторые особенности классификации. Текстурные признаки вычисляются в окрестности (окне) каждой точки изображения. Предложен автоматический выбор окна одного размера для всех точек изображения. От размера окна существенно зависят результаты классификации. Его выбор осуществляется путем перебора различных значений размера, начиная с некоторого небольшого, и построения распределения,  лучшего по изолированности кластеров (выше описан алгоритм)  для каждого из размеров. Предполагается, что для некоторого размера окна значения признаков стабилизируются для всех текстур изображения, и число кластеров перестает меняться. Этот размер и выбирается.

Другая особенность в том, что на границах объектов с разной текстурой появляются ложные кластеры. Для индикации этих узких кластеров применяется пороговый метод, доля граничных точек по отношению к площади превышает порог для ложных кластеров. Предложен алгоритм объединения ложных кластеров  в цепочки по принципу плохой изолированности в пространстве признаков,  с учетом  близости их сегментов в плоскости изображения. Цепочка ложных кластеров выстраивается до тех пор, пока не попадется истинный кластер, и тогда всей цепочке присваивается его номер.

2.3 Принципы выбора признаков

Автоматизация выбора различных моделей или статистик текстуры основана на сравнении значений предложенной ранее меры качества неконтролируемой классификации. Реализовано вычисление текстурных признаков:

* Статистик Харалика.
* Основанных на известной авторегрессионной модели случайного поля SAR (Simultaneus Autoregressive), определенной на двумерном тороидальном массиве изображения с заданной формой соседства пикселей. Параметры модели рассчитываются по известному численному алгоритму в условиях максимального правдоподобия; приближенные значения параметров, характеризующие зависимость пикселя от ближайших соседей и уровня шума, вычисляются в каждой точке изображения и используются как признаки текстуры, формируя многомерный вектор.

В зарубежной литературе рассмотрено применение текстурных признаков, основанных на модели SAR, для контролируемой классификации эталонных текстур альбома Бродатца. Использование этой модели для неконтролируемой классификации – новая разработка, повышающая уровень автоматического распознавания по сравнению с мировыми аналогами.

В качестве примера автором программы приводится анализ и сравнение результатов классификации изображения лесных ландшафтов на аэроснимке для двух систем признаков. Для обеих систем был выполнен автоматический отбор признаков по предложенной мере. Сравнение систем признаков показало, что качество классификаций (по мере) выше для второй системы для любого числа уровней квантования. Сопоставление кластерных карт и карт, полученных наземными измерениями, подтвердило этот вывод. Исследовались различные сочетания признаков Харалика и среднего тона, однако их мощности оказалось не достаточно, чтобы различить некоторые возрастные фазы двух типов леса (сосняков и кедровников). Признаки по модели SAR показали большую различительную мощностью, чем по системе Харалика. Они позволили распознать с точностью наземной таксации возрастные фазы насаждений леса различных типов и возрастов, близких по визуальным свойствам.

2.5 Управление программой

Алгоритм реализован в программной среде системы объектно-ориентированного программирования Visual C++ версии 5.0 фирмы Microsoft c библиотекой классов MFC, разработанная для ОС Windows.   
При разработке программы  использовался механизм многодокументного интерфейса MDI, который позволяет работать одновременно с несколькими изображениями (документами), каждое из которых связано со своим окном.

Параметры выбора режима работы программы подробно описаны в меню Help пользовательского окна программы.  Может быть задан интервал и шаг по числу уровней квантования для поиска лучших распределений с выводом на экран их кластерных карт и характеристик. В другом режиме можно просмотреть подряд все классификации для выбранных уровней квантования.

Статистические текстурные признаки вычисляются в одном из выбранных каналов как дополнение к спектральным, в частных случаях могут быть заданы лишь спектральные, или только текстурные.

Общие характеристики найденных распределений (значение меры изолированности, число уровней квантования, число кластеров) и свойства соответствующих кластеров: значения наиболее вероятных векторов и их плотности вероятности, площади, граничные точки по всем спектральным каналам и др., заносятся в таблицы и могут быть показаны на экране и записаны в файл. Карта кластеров получается в виде BMP-файла (256-цветного) [6].

1. **ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ НА GPU**

В последние годы появился большой практический интерес к использованию современных графических процессоров в качестве универсального вычислителя. GPU ориентирован на поиски более эффективного решения задач машинной графики, в частности он содержит аппаратные функции, позволяющие эффективно выполнять массовые вычисления (однотипные операции над большим объемом данных). Эти возможности позволяют использовать GPU в задачах, не связанных с визуализацией, но также опирающихся на массовые вычисления, например в задачах обработки и анализа изображений. На ряде практических задач вычисления на GPU обеспечили 70-кратное ускорение по сравнению с вычислениями на центральном процессоре, что соответствует временным результатам, получаемым на суперЭВМ. Концепция универсальных вычислений на GPU получила поддержку производителей графических ускорителей (технологии CUDA от NVIDIA, ATI Stream от AMD), что делает доступным написание программ, использующих GPU, на языке высокого уровня без знаний архитектуры сопроцессора. Это, а также невысокая стоимость современных GPU, делает их вовлечение в процессы обработки и анализа данных ДЗЗ актуальной научной задачей. В частности, возможна разработка распределенной технологии, позволяющей использовать ЭВМ с современным графическим ускорителем как вычислительный сервер, обслуживающий вычислительные запросы клиентов на обработку данных ДЗЗ. При разработке данной технологии неоценимую помощь могут оказать сведения о системе SSCCIP (принципы организации, архитектура и программный код) [7].

# **ЗАКЛЮЧЕНИЕ**

В ходе преддипломной практики были получены знания о некоторых основных используемых на данный момент алгоритмах обработки данных дистанционного зондирования Земли.

В результате практики было проведено рассмотрение многофункциональной системы обработки данных дистанционного зондирования Земли с использованием современных методов и алгоритмов обработки изображений, распределенных параллельных программных технологий и Web-ресурсов, разрабатываемой лабораторией обработки изображений Института Вычислительной Математики и Математической Геофизики СО РАН. В частности, рассмотрен экспериментальный программный комплекс SSCCIP, предназначенный для выполнения задачи высокопроизводительной обработки ДДЗЗ на многофункциональных ЭВМ без снижения работоспособности пользователя. Хотя инновационная разработка, основанная на использовании кластерной архитектуры и параллелизации, заявляет о серьезной конкуренции на международном уровне, также в разработке присутствует ряд недостатков: моральное устаревание коммуникационной среды, недостаточная пропускная способность сети, следовательно, трудности при работе с файлами, размер которых исчисляется сотнями гигабайт.

Также в работе представлен анализ уникальной программы, предназначенной для автоматической классификации многоспектральных и текстурных изображений.

В последнем разделе работы представлен краткий обзор проведения высокопроизводительных вычислений на GPU.

Цель и задачи преддипломной практики можно считать выполненными.

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. PuTTY: Telnet/SSH Клиент [Электронный ресурс], URL: https://putty.org.ru/ (дата обращения: 14.5.19).
2. Энциклопедия Windows [Электронный ресурс], URL: http://windata.ru/windows-xp/faq-xp/ispolzovanie-rexec/ (дата обращения: 14.5.19).
3. CryptLib [Электронный ресурс], URL: http://www.cs.auckland.ac.nz/ ~pgut001/cryptlib (дата обращения: 14.5.19).
4. Лаборатория обработки изображений ИВМиМГ СО РАН // SSCC-PIPL: библиотека обработки изображений на многопроцессорных ЭВМ [Электронный ресурс], URL: http://ipl.sscc.ru/lab/RFFI10/ RU/SSCC-PIPL/main.htm (дата обращения: 14.5.19).
5. Лаборатория обработки изображений института вычислительной математики и математической геофизики // Программа разделения многомерной гистограммы по унимодальным кластерам [Электронный ресурс], URL: http://loi.sscc.ru/lab/Weblab/LeraKlas/DEMRU/ DemonRu (дата обращения: 14.5.19).
6. Лаборатория обработки изображений ИВМиМГ СО РАН // Автоматическая классификация изображения по спектральным и текстурным признакам на основе многомерной гистограммы [Электронный ресурс], URL: http://ipl.sscc.ru/lab/RFFI10/RU/svs10.htm (дата обращения: 14.5.19).
7. Бучнев А.А., Пяткин В.П., Русин Е.В. Программные технологии для обработки данных дистанционного зондирования Земли, 2011 г., 10 с. [Электронный ресурс], URL: https://cyberleninka.ru/article/ v/programmnye-tehnologii-dlya-obrabotki-dannyh-distantsionnogo-zondirovaniya-zemli (дата обращения: 14.5.19).