

На правах рукописи

Попов Сергей Борисович

**МОДЕЛИРОВАНИЕ И ФОРМИРОВАНИЕ СТРУКТУРЫ
РАСПРЕДЕЛЕННЫХ СИСТЕМ ОБРАБОТКИ
КРУПНОФОРМАТНЫХ ИЗОБРАЖЕНИЙ
НА ОСНОВЕ ДИНАМИЧЕСКОЙ
ОРГАНИЗАЦИИ ДАННЫХ**

Специальность 05.13.18 – Математическое моделирование,
численные методы и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
доктора технических наук

Самара 2010

Работа выполнена в государственном образовательном учреждении высшего профессионального образования «Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет)» (СГАУ) на кафедре технической кибернетики и в Учреждении Российской академии наук Институте систем обработки изображений РАН

Научный консультант: член-корреспондент РАН,
доктор технических наук, профессор
Сойфер Виктор Александрович

Официальные оппоненты: доктор технических наук, доцент
Востокин Сергей Владимирович

доктор технических наук, профессор
Кораблин Михаил Александрович

доктор технических наук, профессор
Султанов Альберт Ханович

Ведущая организация: Учреждение Российской академии наук Институт проблем управления сложными системами РАН,
г. Самара

Защита состоится «29» октября 2010 г. в 10 часов на заседании диссертационного совета Д 212.215.05 при государственном образовательном учреждении высшего профессионального образования «Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет)» по адресу:
443086, г. Самара, Московское шоссе, 34, ауд. 209.

С диссертацией можно ознакомиться в библиотеке СГАУ.

Автореферат разослан «__» _____ 2010 г.

Ученый секретарь
диссертационного совета
д.т.н., профессор

А.А. Калентьев

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Диссертация посвящена моделированию распределенных систем обработки изображений, формированию структуры таких систем, теоретическому и практическому решению проблемы отображения задач обработки изображений на архитектуру распределенных вычислительных систем.

Актуальность темы

Одной из важнейших проблем использования вычислительной техники является "отображение задач вычислительной математики на архитектуру вычислительных систем". Эта проблема была обозначена академиком Г.И. Марчуком как фундаментальное научное направление, кратко называемое "проблемой отображения".

В настоящее время наиболее актуальной представляется решение проблемы отображения вычислительных задач на параллельную архитектуру вычислительных систем (ВС), поскольку основным направлением повышения эффективности использования вычислительных средств является использование параллельных методов организации вычислений.

Основным подходом к решению проблемы отображения является анализ вычислительной задачи, выявляющий параллелизм и возможности использования распределенных данных, выполняемый на основе *математически эквивалентных преобразований* модели информационной структуры алгоритма решения исследуемой задачи, или, более обобщенно, модели информационной технологии решения задачи. Построение моделей алгоритмов, разработка методов их преобразования и анализа применительно к задачам линейной алгебры, математической физики рассматривались в работах В.А. Вальковского, В.В. Воеводина, Вл.В. Воеводина, А.П. Ершова, В.П. Иванникова, В.А. Крюкова, В.Е. Котова, А.Л. Ластовецкого, В.Э. Малышкина, А.Г. Марчука, Н.Н. Миренкова, L.J. Baer, M.I. Cole, J. Dongarra, I. Foster, L. Lamport и др.

В работах В.В. Воеводина была выдвинута гипотеза о том, что в конкретных вычислительных областях *типовых информационных структур немного*. Разработка моделей информационных технологий в конкретных прикладных областях является актуальной задачей, поскольку открывает новые подходы к разработке специализированных программных средств, эффективно решающих задачи определенного класса на распределенных ВС, и спецпроцессоров, реализующих быстрое выполнение целевых алгоритмов.

В частности, это актуально в области компьютерной обработки изображений. В этой области исследований следует отметить ведущие отечественные школы В.А. Сойфера, В.В. Сергеева, В.С. Киричука, Я.А. Фурмана, Ю.А. Брюханова, К.К. Васильева и др. Разработка систем параллельной обработки изображений ведется В.П. Пяткиным. В области создания распределенных систем параллельной обработки изображений можно отметить работы групп исследователей из университетов США, Испании (University of Extremadura), Италии (University of Naples "Parthenope"), Нидерландов (Delft University of Technology, University of

Amsterdam), Франции (Université Paris Sud, University of Burgundy) и Японии (Kyushu University). В работах, связанных с параллельной обработкой изображений, в основном рассматриваются вопросы построения параллельных алгоритмов обработки изображений, адаптации последовательных алгоритмов применительно к параллельным архитектурам ВС.

Однако вопросы создания моделей информационных технологий обработки изображений, задачи построения математически эквивалентных преобразований этих моделей и решения на этой основе проблемы отображения задач обработки изображений на архитектуры параллельных ВС до настоящего времени практически не исследовались.

Естественный при обработке изображений параллелизм, основанный на декомпозиции данных, позволяет достаточно легко адаптировать последовательные реализации широкого спектра задач обработки для их выполнения на многоядерных процессорах. Однако с увеличением размеров изображений время их обработки все более определяется временем, затрачиваемым на операции ввода/вывода. В таких условиях при обработке крупноформатных изображений с использованием относительно простых в вычислительном отношении задач эффективность использования многоядерных процессоров резко падает. Существующие тенденции развития вычислительной техники еще более обостряют проблему: все увеличивающийся разрыв между производительностью процессоров и быстродействием устройств постоянного хранения данных существенно снижает показатели общей эффективности ВС при решении прикладных задач. Вместе с тем скорость передачи данных в локальных сетях растет существенно быстрее, чем аналогичный показатель для дисковых запоминающих устройств (ЗУ). В настоящее время скорость обмена данными в сетях на базе технологии Gigabit Ethernet (1GbE) находится приблизительно на одном уровне с дисками типовых конфигураций компьютера, но при переходе на технологию 10GbE ситуация кардинально изменится – визуализация крупноформатного изображения будет выполняться быстрее при удаленном, а не локальном хранении данных. В таких условиях использование распределенных систем для хранения и обработки крупноформатных изображений становится актуальным.

Другим важным фактором при обработке крупноформатных изображений является размер оперативной памяти. Несмотря на то, что появление 64-битных операционных систем отодвинуло ограничение на размер оперативной памяти, существенное наращивание оперативной памяти компьютера относительно типовых решений достаточно дорого. Эффективным решением в данном случае является обработка изображения по частям на разных компьютерах. В сочетании с возможностью параллельной организации таких вычислений в распределенных системах, особенно при организации их на основе кластерных вычислительных систем, этот подход представляется перспективным.

Обоснованное решение проблемы отображения задач обработки изображений на архитектуру распределенных вычислительных систем с различными типами параллелизма должно опираться на моделирование и исследование различных

вариантов организации хранения и обработки изображений применительно к целевой архитектуре аппаратно-программных средств. Например, централизованное размещение крупноформатных изображений на специализированных хранилищах данных является узким местом параллельных систем обработки изображений, с другой стороны, при организации распределенного хранения необходимо решить проблемы, связанные с выбором наилучшей декомпозиции (разбиения на фрагменты) изображения в условиях отсутствия априорной информации о параметрах запускаемых в системе задач обработки, сбалансированностью загрузки компьютеров, участвующих в обработке при заранее выполненной декомпозиции данных, отказоустойчивостью распределенного хранения фрагментов изображений, обеспечением интерактивности системы при визуализации распределенных изображений.

В диссертации на основе разработанных моделей технологий обработки изображений и методов их преобразования, выявляющих параллелизм, предложены алгоритмы динамической организации хранения и обработки данных, которые снимают эти проблемы, предложена архитектура и принципы реализации распределенной системы хранения и обработки изображений, создан ряд программных комплексов, решающих практические задачи обработки изображений.

Цель и задачи исследования

Целью работы является решение проблемы отображения наиболее распространенных информационных технологий обработки изображений на архитектуру распределенных вычислительных систем с различными типами параллелизма и связанной с ней задачи декомпозиции данных при построении распределенных систем обработки и хранения крупноформатных изображений.

В соответствии с поставленной целью определены *основные задачи диссертации*:

1. Анализ типовых информационных технологий обработки изображений и методов их отображения на архитектуру распределенных вычислительных систем с различными типами параллелизма.
2. Выбор методологии и конструирование формальных моделей, описывающих информационную структуру основных технологий обработки изображений.
3. Разработка методов эквивалентного преобразования модели технологий обработки изображений, выявляющих потенциальный параллелизм.
4. Создание методов декомпозиции данных изображений при распределенном хранении и параллельной обработке.
5. Разработка методов обеспечения сбалансированной нагрузки при параллельной обработке изображений в условиях гетерогенной вычислительной среды.
6. Реализация комплексов программ прикладной обработки изображений с использованием распределенных ВС.

Методы исследования

В диссертационной работе используются элементы теории множеств, математической логики, алгебры изображений и теории взаимодействующих

последовательных процессов. Результаты теоретических исследований подтверждены вычислительными экспериментами с имитационными моделями и на тестовых задачах, а также реализацией основных алгоритмов в рамках распределенных систем обработки изображений, работающих в условиях реального производства.

Научная новизна работы

1. Разработана информационная модель наиболее распространенных технологий обработки изображений, новизна которой состоит в том, что она объединяет формальные модели алгоритмов обработки, а также организации данных при хранении и коммуникации, строится на основе теории взаимодействующих последовательных процессов, использует итераторы преобразования данных и включает предиктивную оценку времени выполнения задач обработки.

2. Введена формальная модель итераторов преобразования данных и правила эквивалентного преобразования композиции итераторов, которые используются при построении информационной модели и анализе потенциального параллелизма.

3. Предложен метод формирования децентрализованных структур данных, определяющих распределенные изображения, на основе декомпозиции перекрывающихся фрагментов в условиях отсутствия априорной информации о параметрах перекрытия, необходимого для реализуемых в системе задач обработки. Новизна предлагаемого подхода заключается в том, что размер необходимого перекрытия фрагментов определяется не параметрами задач обработки, а необходимостью обеспечения отказоустойчивости распределенного хранения фрагментов изображений и сбалансированности загрузки компьютеров при обработке.

4. Разработан алгоритм динамической балансировки многопроцессорных систем при распараллеливании выполнения операций обработки изображений в условиях гетерогенной вычислительной среды, новизна которого определяется применением предложенного метода декомпозиции распределенного изображения.

5. Предложена сервис-ориентированная архитектура распределенных систем обработки изображений, основанная на новом способе организации распределенных изображений в виде набора динамически взаимодействующих сервисов хранения и децентрализованной организации вычислений в процессе обработки изображений, ориентированной на данные.

6. Реализованы новые эффективные комплексы программ для решения практических задач обработки изображений: сервис-ориентированная распределенная система обработки крупноформатных изображений, кластерная система формирования и параллельной обработки крупноформатных изображений, программная система передачи крупноформатных изображений по сети Internet, система технического зрения для регистрации железнодорожных составов цистерн, система технического зрения для определения количества гель-частиц в растворе полимера.

На защиту выносятся

1. Информационная модель наиболее распространенных технологий обработки изображений, объединяющая формальные модели алгоритмов обработки на основе итераторов преобразования данных, модели организации данных при хранении изображений и процессов коммуникации при параллельной обработке, модель предиктивной оценки времени выполнения задачи обработки.

2. Формальная модель итераторов преобразования данных при обработке изображений и правила эквивалентного преобразования композиции итераторов. Алгоритмы пересчета времени выполнения задачи при эквивалентных преобразованиях моделей в терминах итераторов.

3. Метод формирования децентрализованных структур данных, определяющих распределенные изображения, на основе декомпозиции перекрывающихся фрагментов, обеспечивающий отказоустойчивость распределенного хранения фрагментов изображений и позволяющий динамически балансировать загрузку компьютеров.

4. Алгоритм децентрализованной динамической балансировки много-процессорных систем при распараллеливании выполнения операций обработки изображений, основанный на предложенном методе декомпозиции распределенного изображения.

5. Сервис-ориентированная архитектура распределенных систем обработки изображений, на основе нового способа динамической организации распределенных изображений в виде набора взаимодействующих сервисов хранения и децентрализованной организации обработки изображений.

6. Комплексы программ, созданные при решении практических задач обработки изображений: сервис-ориентированная распределенная система обработки крупноформатных изображений, кластерная система формирования и параллельной обработки крупноформатных изображений, программная система передачи крупноформатных изображений по сети Internet, программное обеспечение созданных систем технического зрения.

Практическая ценность работы

Практическая значимость работы состоит в том, что разработанные модели и методы отображения технологий обработки изображений на архитектуру параллельных вычислительных систем открывают возможности создания специализированных систем обработки крупноформатных изображений, получаемых при дистанционном зондировании земной поверхности, а также в электронной микроскопии при анализе больших образцов наноструктур.

Результаты, полученные в ходе выполнения работы, использовались при создании сервис-ориентированной распределенной системы обработки крупноформатных изображений, кластерной системы формирования и параллельной обработки крупноформатных изображений, программной системы передачи крупноформатных изображений по сети Internet, системы технического зрения для регистрации железнодорожных составов цистерн, системы технического зрения для определения количества гель-частиц в растворе полимера.

Апробация работы

Основные результаты работы докладывались на 27 научных конференциях и семинарах: Второй Всесоюзной конференции «Методы и средства обработки сложной графической информации», г. Горький, 1985; Всесоюзной конференции «Методы и средства обработки сложной графической информации». г. Горький, 1988; Втором Республиканском семинаре "Проблемы создания систем обработки, анализа и распознавания изображений". г. Ташкент, 1989; Всесоюзном симпозиуме "Методы и применение голографической интерферометрии". г. Куйбышев, 1990; Первой Международной конференции «International Conference On Information Technologies For Image Analysis and Pattern Recognition (ITIAPR'90)», г. Львов, 1990; Четвертой Всесоюзной конференции «Методы средства обработки сложной графической информации». г. Горький, 1991; Второй Всероссийской с участием стран СНГ конференции «Распознавание образов и анализ изображений: Новые информационные технологии», г. Ульяновск, 1995; Пятом Международном семинаре «Распределенная обработка информации», г. Новосибирск, 1995; Третьей конференции «Распознавание образов и анализ изображений: Новые информационные технологии», г. Нижний Новгород, 1997; Пятой международной конференции «Распознавание образов и анализ изображений: новые информационные технологии» (РОАИ-5-2000), г. Самара, 2000; Международной конференции «Математическое моделирование – 2001», г. Самара, 2001; Научно-методической конференции «Телематика 2002», г. С-Петербург, 2002; Шестой Международной конференции «Распознавание образов и анализ изображений: новые информационные технологии» (РОАИ-6-2002), г. Великий Новгород, 2002; Втором Международном научно-практическом семинаре «Высокопроизводительные параллельные вычисления на кластерных системах», г. Нижний Новгород, 2002; Третьем Международном научно-практическом семинаре «Высокопроизводительные параллельные вычисления на кластерных системах», г. Нижний Новгород, 2003; Международной конференции «Распределенные вычисления и Грид-технологии в науке и образовании», г. Дубна, 2004; Седьмой Международной конференции «International Conference on Pattern Recognition and Image Analysis: New Information Technologis» (PRIA-7-2004), г. Санкт-Петербург, 2004; Всероссийской научной конференции «Научный сервис в сети Интернет: технологии распределенных вычислений», г. Новороссийск, 2005; Второй Международной конференции «Distributed Computing and Grid-Technologies in Science and Education», г. Дубна, 2006; Всероссийской научной конференции «Научный сервис в сети Интернет: технологии параллельного программирования», г. Новороссийск, 2006; Четырнадцатой Всероссийской научно-методической конференции «Телематика'07», г. Санкт-Петербург, 2007; Всероссийской научной конференции «Научный сервис в сети Интернет: многоядерный компьютерный мир», г. Новороссийск, 2007; Тринадцатой Всероссийской конференции «Математические методы распознавания образов», Ленинградская обл., г. Зеленогорск, 2007; Седьмом Международном научно-практическом семинаре «Высокопроизводительные параллельные вычисления на кластерных системах»,

г. Нижний Новгород, 2007; Всероссийской научной конференции «Научный сервис в сети Интернет: решение больших задач», г. Новороссийск, 2008; Третьей Международной конференции «Distributed Computing and Grid-Technologies in Science and Education», г. Дубна, 2008; Всероссийской суперкомпьютерной конференции «Научный сервис в сети Интернет: масштабируемость, параллельность, эффективность», г. Новороссийск, 2009.

Связь с государственными и международными программами

Основные результаты диссертации получены в рамках научно-исследовательских работ (НИР) по международным, государственным, межвузовским и региональным научно-техническим программам: грантам Российского фонда фундаментальных исследований №№ 02-01-00119-а, 02-01-08055-инно, 03-01-00109-а, 04-07-90149-а, 04-07-96500-р2004поволжье_в, 05-01-08043-офи_а, 06-08-01024-а, 07-07-00210-а, 10-07-00553-а; программе фундаментальных исследований Президиума РАН «Проблемы создания национальной научной распределенной информационно-вычислительной среды на основе развития GRID технологий и современных телекоммуникационных сетей» по направлению № 4 "Оптимизация вычислительных архитектур под конкретные классы задач, информационная безопасность сетевых технологий, применение распределенных информационно-вычислительных систем"; гранту Президента России по поддержке ведущих научных школ – НШ-3086.2008.9 «Разработка теоретических основ и создание информационных технологий обработки изображений и компьютерной оптики»; Российско-американской программе "Фундаментальные исследования и высшее образование" ("BRHE" CRDF Project SA-014, 2002-2010 гг.); государственной научно-технической программе «Перспективные информационные технологии» (1994-1999 гг.).

Публикации

По теме диссертации опубликовано 62 работы, в том числе 3 монографии, 22 статьи в ведущих рецензируемых журналах и изданиях, входящих в перечень ВАК, 36 статей в сборниках, трудах конференций и тезисов докладов, получено 6 свидетельств о регистрации программы для ЭВМ и патент на изобретение.

Структура и объем работы

Диссертация состоит из введения, четырех разделов, заключения, списка использованных источников и приложений. Она содержит 258 страниц основного текста, 56 рисунков, 11 таблиц. Библиографический список включает 208 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность работы, дана ее общая характеристика, сформулированы цель и задачи исследования.

В первой главе формулируются основные математические модели, связанные с обработкой изображений, и решаются задачи отображения типовых информационных технологий в этой области на архитектуру распределенных вычислительных систем с различными типами параллелизма.

В начале первой главы диссертации анализируется типология основных задач обработки изображений, представленная на рис. 1.

Подчеркнуто, что именно выполнение некой последовательности операций низкого уровня над набором исходных изображений является отправной точкой для многих типичных приложений в области обработки изображений и систем технического зрения. Основные затраты в рассматриваемых приложениях связаны с низкоуровневыми операциями. Эти операции требуют значительных вычислительных ресурсов, ресурсов хранения. Время реакции всего приложения в целом определяется в большинстве случаев именно временем, затрачиваемым на выполнение низкоуровневых операций. Основные исследования в области распараллеливания обработки изображений сосредоточены именно на этом уровне обработки.

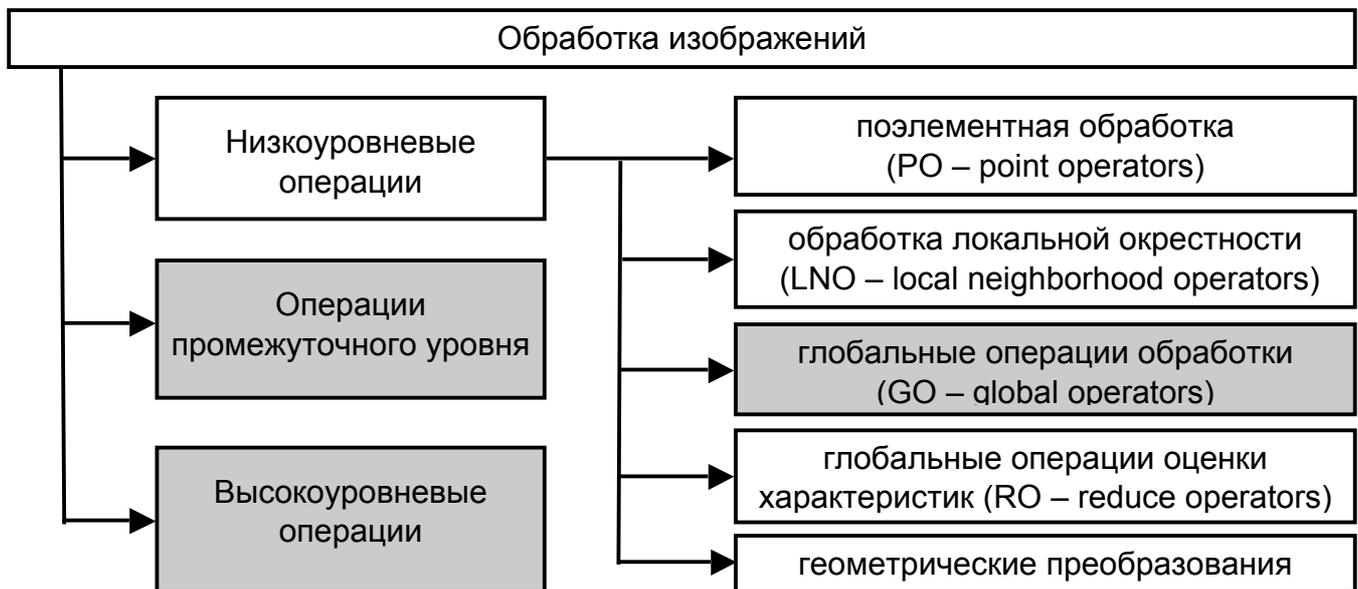


Рисунок 1 – Типология операций обработки изображений

Следует отметить, что глобальные операции обработки изображений – это отдельная обширная область исследований, особенно в части их параллельных реализаций. В диссертации данный класс преобразований не рассматривался. Геометрические преобразования сложно отнести к глобальным операциям, поскольку для вычисления некоторой точки результирующего преобразования необходимо не более четырех точек исходного изображения. Но и локальными их нельзя назвать, так как расположение этих четырех точек в общем случае не находится в локальной окрестности формируемой точки. Из геометрических преобразований в диссертации исследовались практически важные операции масштабирования и поворота на небольшой угол, которые по своей структуре близки к локальным.

Рассмотрение объекта исследования – вычислительных задач обработки изображений, показало, что эти задачи имеют следующую, присущую им особенность – большинство задач может быть решено путем последовательного применения к обрабатываемым изображениям некоторого набора типовых операций обработки. Именно эту особенность эксплуатирует большинство систем обработки изображений общего назначения, т.е. они создаются не под конкретную технологию

обработки, а для решения широкого спектра задач обработки изображений интерактивным образом или на уровне макропрограммирования.

В соответствии с этим в диссертации предложены следующие этапы решения проблемы отображения задач обработки изображений на параллельные ВС:

- формирование моделей основных типовых операций обработки изображений;
- разработка методов формирования моделей сложных информационных технологий обработки, реализуемых как программная сеть, составленная из типовых операций;
- разработка методов анализа и преобразования моделей, выявляющих потенциальный параллелизм;
- решение задач согласования потенциального параллелизма с распределением данных путем формирования возможных вариантов декомпозиции изображений при распределенном хранении и параллельной обработке и последующего сравнительного анализа распределения данных, выявляющего тот вариант, который обеспечивает минимизацию вычислительных и коммуникационных издержек в процессе обработки.

Модель информационной технологии обработки изображений строится как композиция итераторов обработки изображений на основе нотации алгебры CSP и алфавита алгебры изображений Г. Риттера.

В диссертации из множества возможных формальных моделей, применяемых для описания информационной структуры программ, с учетом требований наличия средств описания параллелизма и учета времени работы программы, выбрана теория взаимодействующих последовательных процессов Ч. Хоара (алгебра CSP). Выразительная мощность моделей на основе этой алгебры процессов хорошо соответствует последовательной природе обычной программы, существуют механизмы явного описания взаимодействия процессов при параллельной обработке, которые хорошо соответствуют реальным реализациям в программных системах.

В качестве алфавита, описывающего элементы модели объекта обработки – изображения, выбрана алгебра изображений Г. Риттера. Достоинством используемой в ней нотации является явное разделение структур данных на структуру пространственной организации (взаимозависимости) пикселей изображения – области определения, и структуру области значений изображения, т.е. структуру, определяющую типологию изображения, начиная с бинарного, кончая цветными (с использованием различных цветовых моделей) и многоканальными. Таким образом, использование данного алфавита в формальных моделях технологий обработки изображений позволяет выделить все аспекты реализации операций обработки изображений, связанные с пространственной организацией данных изображений, т.е. структурой представления этих данных.

Основные элементы алфавита модели информационной технологии обработки изображений представлены в таблице 1.

В диссертации предлагается в качестве основного элемента модели рассматривать формальную модель итератора преобразования данных.

Таблица 1 – Основные элементы алфавита модели обработки изображений

Обозначение	Определение
$\mathbf{X} = \{(x_1, x_2) \in \mathbb{Z}^2 : x_1 \in \mathbb{Z}_{n_1}, x_2 \in \mathbb{Z}_{n_2}\}$	Область определения изображения, где $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$
$\mathbb{F} \subset \{\mathbb{Z}^m, \mathbb{R}^m, \mathbb{C}\}$	Область значений изображения определяющая типологию изображения
$\mathbf{a} \in \mathbb{F}^{\mathbf{X}} = \{(\mathbf{x}, \mathbf{a}(\mathbf{x})) : \mathbf{x} \in \mathbf{X} \subset \mathbb{Z}^2, \mathbf{a}(\mathbf{x}) \in \mathbb{F}\}, \mathbf{a}(\mathbf{x})$	Изображение, значение пиксела изображения в точке $\mathbf{x} \in \mathbf{X}$
$\mathbf{X} _{x_1=i} = \{(x_1, x_2) : x_1 = i, x_2 \in \mathbb{Z}_{n_2}\}$	Множество точек, принадлежащих одной строке
$\mathbf{X} _{x_2=j} = \{(x_1, x_2) : x_1 \in \mathbb{Z}_{n_1}, x_2 = j\}$	Множество точек, принадлежащих одному столбцу
$Q(\mathbf{x}) = \{\mathbf{x} + \mathbf{y} : \forall \mathbf{y} \in \mathbf{H}\}$	Произвольная окрестность точки \mathbf{x} , задаваемое множеством точек $\mathbf{H} \subset \mathbb{Z}^2$
$\mathbf{a}_{p i}$	Частичное изображение, часть изображения \mathbf{a} , размещенное на вычислительном узле i
$\mathbf{X}_{p i} : \bigcup_{i=0}^{n-1} \mathbf{X}_{p i} = \mathbf{X}$	Область определения частичного изображения, размещенного на вычислительном узле i
$\mathbf{X}_d = \prod_{i=1}^n \mathbf{X}_{p i}$	Область определения распределенного изображения – дизъюнктивное объединение областей определения всех его частичных изображений
$\mathbf{a}_d = \prod_{i=1, P} \mathbf{a}_{p i}, \prod_{i=1, P} \mathbf{a}_{p i}^{\pm k}, \prod_{i=1, P; j=1, P} \mathbf{a}_{p i, j}, \prod_{i=1, P; j=1, P} \mathbf{a}_{p i, j}^{\pm k, l}$	Распределенное изображение с различными способами декомпозиции: одномерная без пересечения, с частичным перекрытием фрагментов, то же в двумерном случае
$\mathbf{a}^b \in \mathbb{F}^{\mathbf{X}+\mathbf{H}}, \mathbf{a}^b(\mathbf{x}) = \begin{cases} \mathbf{a}(\mathbf{x}), \forall \mathbf{x} \in \mathbf{X} \\ \mathbf{b}(\mathbf{y}), \forall \mathbf{y} \in (\mathbf{X} + \mathbf{H}) \setminus \mathbf{X} \end{cases}$	Расширение изображения $\mathbf{a} \in \mathbb{F}^{\mathbf{X}}$ на отсчеты за пределами области определения исходного изображения $\mathbf{b} \in \mathbb{F}^{(\mathbf{X}+\mathbf{H}) \setminus \mathbf{X}}$
$\mathbf{Z}_d = \mathbf{X}_d + \mathbf{H} = \prod_{i \in \mathbb{Z}_p} (\mathbf{X}_{p i} + \mathbf{H})$	Область определения расширения распределенного изображения
$\mathbf{a}_d^b = \prod_P^k \mathbf{a}_{p i}^{\pm k}{}^b$, где $k = (\text{card}(p_1(\mathbf{H})) - 1)/2$	Расширение распределенного изображения
$\mathbf{a}_{p i}^{\pm k}{}^b = \begin{cases} \mathbf{a}_{p i}(\mathbf{x}), \forall \mathbf{x} \in \mathbf{X}_{p i} \\ \mathbf{a}_{p i+j}(\mathbf{x}), \forall \mathbf{x} \in (\mathbf{X}_{p i} + \mathbf{H}) \cap \mathbf{X}_{p i+j}, j = -1, j = 1 \\ \mathbf{b}(\mathbf{y}), \forall \mathbf{y} \in (\mathbf{X}_{p i} + \mathbf{H}) \setminus (\mathbf{X}_{p i} \cup \mathbf{X}_{p i-1} \cup \mathbf{X}_{p i+1}) \end{cases}$	Расширение частичного изображения Значения $\mathbf{b}(\mathbf{y})$ или задаются константой, или доопределяются значениями пикселей изображения \mathbf{a} одним из трех способов: доопределение ближайшим отсчетом; циклическое доопределение; зеркальное доопределение.

В общем случае итератор – это совокупность трех составляющих (X, F, T) , где X – множество итерируемых объектов, F – оператор преобразования объекта; T – оператор упорядочивания на множестве объектов.

1. Итератор применяется к некоторому множеству объектов X . Если множество пустое, то никаких действий не выполняется.
2. Итератор выполняет однотипные действия над некоторой упорядоченной последовательностью объектов: $\forall x \in X, F(x)$.

3. Упорядоченность обеспечивается наличием некоторого оператора упорядочивания, который последовательно перебирает элементы множества.
4. То, каким образом задано множество, влияет на реализацию оператора упорядочивания, но не влияет на работу итератора в целом.

Аналогично конструкции CSP $b * Q$, которая обозначает традиционный цикл **while** b **do** Q , введено обозначение итератора для случая произвольной реализации оператора упорядочивания $\{x : X\} * F(x)$, которое соответствует оператору **foreach** (x) **do** $F(x)$, реализованному в большинстве современных языков программирования, $\Theta(X, F(x)) = \{x : X\} * F(x)$.

В диссертации сформулированы основные правила преобразований итераторов, где $var(F(x))$ – множество изменяемых переменных процесса, $acc(F(x))$ – множество используемых переменных процесса.

L1. Расщепление итератора. Если $Z = X \cup Y$ таким образом, что $X \cap Y = \emptyset$, то

$$\{x : Z\} * F(x) = \{x : X\} * F(x); \{x : Y\} * F(x) \text{ или } \Theta(Z, F(x)) = \Theta(X, F(x)); \Theta(Y, F(x)).$$

L2. Слияние итераторов по множествам.

$$\text{Если } X \cap Y = \emptyset, \text{ то } \{x : X\} * F(x); \{x : Y\} * F(x) = \{x : Z\} * F(x), \text{ где } Z = X \cup Y.$$

L3. Слияние итераторов по преобразованиям. Если $\forall x, y : x \succ y$ выполняется

$$var(F(x)) \cap acc(G(y)) = \emptyset, \text{ то } \{x : X\} * F(x); \{x : X\} * G(x) = \{x : X\} * (F(x); G(x)).$$

L4. Параллельная декомпозиция итератора по данным. Если $Z = X \cup Y$ и $X \cap Y = \emptyset$, то

$$\{x : Z\} * F(x) = \{x : X\} * F(x); \{x : Y\} * F(x) = \{x : X\} * F(x) \parallel \{x : Y\} * F(x), \text{ при условии, что } \forall x \in X, \forall y \in Y \text{ выполняется } var(F(x)) \cap acc(F(y)) = var(F(y)) \cap acc(F(x)) = \emptyset.$$

L5. Параллельная декомпозиция итератора по преобразованиям.

$$\{x : X\} * (F(x); G(x)) = \{x : X\} * F(x) \parallel \{x : X\} * G(x), \text{ при условии, что}$$

$$\forall x, y \in X \text{ выполняется } var(F(x)) \cap acc(G(y)) = var(G(x)) \cap acc(F(y)) = \emptyset.$$

L6. Вложенные итераторы.

$$\text{Пусть } X = \bigcup_{i=1}^n X_i \text{ и } X_i \cap X_j = \emptyset, \forall i, j : i \neq j, \text{ тогда } \{x : X\} * F(x) = \{X_i : X\} * (\{x : X_i\} * F(x)).$$

Правая часть данного равенства определяет итератор по подмножествам $\Theta(X, F(X_i))$.

L7. Параллельная декомпозиция вложенных итераторов.

$$\text{Пусть } X = \bigcup_{i=1}^n X_i \text{ и } X_i \cap X_j = \emptyset, \forall i, j : i \neq j, \text{ тогда } \{x : X\} * F(x) = \parallel_{i=1}^n (\{x : X_i\} * F(x)),$$

при условии, что $var(\Theta(X_i, F(x))) \cap acc(\Theta(X_j, F(x))) = \emptyset, \forall i, j : i \neq j$.

Эквивалентность и корректность данных правил преобразований доказана в рамках теории CSP.

Итераторы, реализующие основные операции обработки изображений, определены в диссертации следующим образом.

Пусть $\mathbf{a} \in \mathbb{F}^{\mathbf{X}}$ – произвольное изображение, имеющее область определения \mathbf{X} , и $F(\mathbf{a}(\mathbf{x}))$ – процесс, использующий значение пиксела изображения в точке $\mathbf{x} \in \mathbf{X}$, тогда итератор, обеспечивающий применение данного процесса ко всем точкам изображения, определяется следующим образом $\Theta(\mathbf{a}, F(\mathbf{a}(\mathbf{x}))) = \{\mathbf{x} : \mathbf{X}\} * F(\mathbf{a}(\mathbf{x}))$.

В таблице 2 представлены модели основных *типовых* операций обработки изображений и их параллельные композиции на основе распределения данных. В таблице предполагается, что $\mathbf{a} \in \mathbb{F}^{\mathbf{X}}$, $\mathbf{c} \in \mathbb{F}^{\mathbf{X}}$ – произвольные изображения, имеющие одинаковую область определения \mathbf{X} , соответствующие им распределенные изображения \mathbf{a}_d и \mathbf{c}_d определены с помощью одинаковой декомпозиции области определения \mathbf{X}_d , задана операция редукции γ , формирующая некоторую характеристику изображения, такая что $s = s\gamma f(\mathbf{a}(\mathbf{x}))$, $s \in \mathbb{F}$, $\exists e: s = s\gamma e = e\gamma s$.

Таблица 2 – Модели основных типовых операций обработки изображений

Типовая операция обработки изображений	Формальная модель операции
Операция поэлементного преобразования в общем виде	$PO(\mathbf{a}, \mathbf{c}, \Phi) = \{\mathbf{x} : \mathbf{X}\} * (\mathbf{c}(\mathbf{x}) := \Phi(\mathbf{a}(\mathbf{x})))$
Операция линейного поэлементного преобразования	$PO_Lin(\mathbf{a}, \mathbf{c}, \alpha, \beta) = \{\mathbf{x} : \mathbf{X}\} * (\mathbf{c}(\mathbf{x}) := \alpha \cdot \mathbf{a}(\mathbf{x}) + \beta)$
Операция табличного поэлементного преобразования	$PO_LUT(\mathbf{a}, \mathbf{c}, LUT) = \{\mathbf{x} : \mathbf{X}\} * (\mathbf{c}(\mathbf{x}) := LUT(\mathbf{a}(\mathbf{x})))$
Операция поэлементного преобразования двух изображений	$PO_Lin2(\mathbf{a}, \mathbf{b}, \mathbf{c}, \alpha, \beta, \gamma, \delta) =$ $\{\mathbf{x} : \mathbf{X}\} * (\mathbf{c}(\mathbf{x}) := \alpha \cdot \mathbf{a}(\mathbf{x}) + \beta \cdot \mathbf{b}(\mathbf{x}) + \gamma \cdot \mathbf{a}(\mathbf{x}) \cdot \mathbf{b}(\mathbf{x}) + \delta)$
Параллельная декомпозиция поэлементного преобразования	$PO(\mathbf{a}_d, \mathbf{c}_d, \Phi) = \left\ \left\ _{i=1}^P \left(\{\mathbf{x} : \mathbf{X}_{p i}\} * (\mathbf{c}(\mathbf{x}) := \Phi(\mathbf{a}(\mathbf{x}))) \right) \right\ $
Операции оценки характеристик изображения в общем виде	$RO(\mathbf{a}, s, \gamma, f) = (s := e); \{\mathbf{x} : \mathbf{X}\} * (s := s\gamma f(\mathbf{a}(\mathbf{x})))$
Операция суммирования значений всех пикселей изображения	$RO(\mathbf{a}, s, +, 1) = (s := 0); \{\mathbf{x} : \mathbf{X}\} * (s := s + \mathbf{a}(\mathbf{x}))$
Параллельная декомпозиция операции оценки характеристик изображения	$RO(\mathbf{a}_d, s, \gamma, f) = \left\ \left\ _{i=1}^P RO(\mathbf{a}_{p i}, s_i, \gamma, f); \left(s := \prod_{i=1}^P s_i \right) \right\ $
Операция локальной обработки скользящим окном в общем виде	$LNO(\mathbf{a}^b, \mathbf{c}, \mathbf{H}, F) = \{\mathbf{x} : \mathbf{X}\} * \Theta(\mathbf{H}, F(\mathbf{x}, \mathbf{y})) =$ $= \{\mathbf{x} : \mathbf{X}\} * (\{\mathbf{y} : \mathbf{H}\} * F(\mathbf{x}, \mathbf{y}))$
Операция локальной обработки скользящем окном в терминах операции редукции над отсчетами окрестности обработки	$LNRO(\mathbf{a}^b, \mathbf{c}, \mathbf{H}, \gamma, h, f) =$ $\{\mathbf{x} : \mathbf{X}\} * \left(s := e; \{\mathbf{y} : \mathbf{H}\} * (s := s\gamma h(\mathbf{y}, \mathbf{a}^b(\mathbf{x} + \mathbf{y}))); \right.$ $\left. \mathbf{c}(\mathbf{x}) := f(s) \right)$
Операция обработки КИХ-фильтром	$FIR(\mathbf{a}^b, \mathbf{c}, \mathbf{h}, q) =$ $\{\mathbf{x} : \mathbf{X}\} * \left(s := 0; \{\mathbf{y} : \mathbf{H}\} * (s := s + \mathbf{h}(\mathbf{y})\mathbf{a}^b(\mathbf{x} + \mathbf{y})); \right.$ $\left. \mathbf{c}(\mathbf{x}) := s / q \right)$
Параллельная декомпозиция операции локальной обработки скользящим окном в общем виде	$LNO(\mathbf{a}_d^b, \mathbf{c}_d, \mathbf{H}, F) = \left\ \left\ _{i=1}^P LNO(\mathbf{a}_{p i}^b, \mathbf{c}_{p i}, \mathbf{H}, F) \right\ $

Для проведения анализа вариантов организации вычислений при обработке изображений вводится расширение информационной модели, позволяющее оценить время работы процесса, описывающего обработку изображений.

Введем функцию $t = \vartheta(n)$, определяющую время выполнения итератора $\Theta(X, F(x))$ в зависимости от количества элементов множества X , и ее обозначение при записи модели $\Theta(X, F(x))[\vartheta(n)]$, где $n = \text{card}(X)$.

В диссертации сформулированы правила формирования функций времени выполнения итератора при выполнении преобразований.

$$\text{Для L1 и L2: } \Theta(Z, F(x))[\vartheta(n+k)] = \Theta(X, F(x))[\vartheta(n)]; \Theta(Y, F(x))[\vartheta(k)],$$

где $n = \text{card}(X)$, $k = \text{card}(Y)$ и $\text{card}(Z) = n + k$.

Для L3:

$$(\{x : X\} * F(x))[\vartheta_1(n)]; (\{x : X\} * G(x))[\vartheta_2(n)] = (\{x : X\} * (F(x); G(x)))[\vartheta(n)],$$

где $n = \text{card}(X)$ и, в общем случае, $\vartheta(n) \neq \vartheta_1(n) + \vartheta_2(n)$.

$$\text{Для L4, L5 и L7: } (\Theta(X, F(x))[\vartheta_F(n)] \parallel \Theta(Y, G(x))[\vartheta_G(k)])[\max(\vartheta_F(n), \vartheta_G(k))],$$

где $n = \text{card}(X)$, $k = \text{card}(Y)$.

Для L6: Если $X = \bigcup_{i=1}^n X_i$, $\forall i \ k = \text{card}(X_i)$ и $X_i \cap X_j = \emptyset, \forall i, j : i \neq j$, то

$$\Theta(X, F(x))[\vartheta(n \cdot k)] = \{X_i : X\} * (\{x : X_i\} * F(x))[\vartheta(k)].$$

В ходе экспериментальных исследований было установлено, что для итераторов, в теле которых нет операций чтения/записи данных с внешних запоминающих устройств (ВЗУ), операций коммуникации, хорошей моделью функции времени выполнения итератора является линейная функция $\Theta(X, F(x))[\vartheta(n) = \vartheta_F \cdot n] = \{x : X\} * (F(x)[\vartheta_F])$.

Так же в ходе вычислительных экспериментов показано, что для параллельных систем с общей памятью при параллельной декомпозиции вложенных итераторов, если $X = \bigcup_{i=1}^P X_i$, $\forall i \ k = \text{card}(X_i)$, $n = \text{card}(X) = P \cdot k$, $X_i \cap X_j = \emptyset, \forall i, j : i \neq j$, модель функции времени выполнения итератора определяется следующим образом

$$(\{x : X\} * F(x))[\vartheta(n) = \vartheta_0 + \vartheta_F \frac{n}{P}] = \parallel_{i=1}^P ((\{x : X_i\} * F(x))[\vartheta_F \cdot k]),$$

где ϑ_0 – накладные расходы на формирование параллельных ветвей итератора.

Экспериментально подтверждено, что при слиянии итераторов по преобразованиям модель функции времени выполнения объединенного итератора, в общем случае, не может быть определена через модели функций времени выполнения составляющих его итераторов, т.е.

$$(\{x : X\} * F(x))[\vartheta_F \cdot n]; (\{x : X\} * G(x))[\vartheta_G \cdot n] = (\{x : X\} * (F(x); G(x)))[\vartheta(n) \neq \vartheta_F \cdot \vartheta_G \cdot n],$$

где $n = \text{card}(X)$.

Это связано с особенностями оптимизирующих компиляторов и конвейерной организацией выполнения команд в современных процессорах. Поэтому при слиянии итераторов по преобразованиям параметры модели функции времени выполнения объединенного итератора должны оцениваться в ходе самостоятельных вычислительных экспериментов.

В диссертации предложена методика оценивания параметров модели функции времени выполнения итератора в различных режимах работы используемых программно-аппаратных средств вычислительной системы. Данная методика учитывает гетерогенный характер функционирования прикладной программы в условиях конкурентного выполнения пользовательских и служебных процессов в

операционной системе. Методика включает в себя порядок проведения тестовых экспериментов, методы согласования экспериментальных значений времени, полученных в различных метриках, методы оценивания параметров модели функции времени выполнения итератора, основанные на использовании принципа согласованного оценивания, предложенного в работах В.А. Фурсова.

Завершают первую главу диссертации примеры построения информационных моделей типичных задач обработки изображений, которые часто используются как при исследовании различных реализаций систем обработки изображений, так и в специализированных тестах (фирмы Intel, портала www.ixbt.com) при сравнении производительности вычислительных систем. В каждом примере сначала строится модель, формируемая как простая последовательность типовых операций, а затем с помощью эквивалентных преобразований – модель, допускающая отображение на параллельные вычислительные системы. Рассмотрены следующие примеры: 1) эквализация гистограммы; 2) линейное повышение контрастности; 3) выделение контуров с помощью оператора Собела; 4) каскадная медианная фильтрация; 5) последовательная обработка изображения набором КИХ-фильтров с произвольной маской.

Проведены вычислительные эксперименты по оценке достоверности разработанных формальных моделей. На рис. 2 и 3 представлены результаты исследования на примере задачи выделения контуров.

На рис. 3 показано сравнение оценок эффективности параллельной обработки на 8 процессорах, полученных на тестовых реализациях задач и при моделировании.

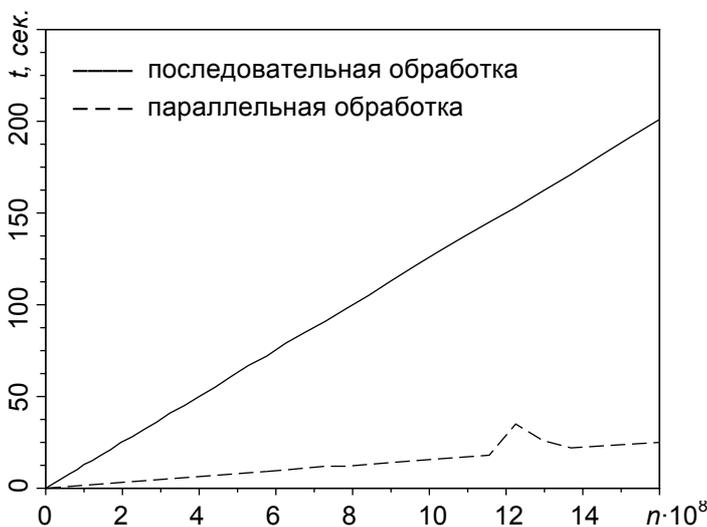


Рисунок 2 – Экспериментальные зависимости времени выполнения обработки от размера изображения



Рисунок 3 – Зависимость эффективности параллельной обработки от размера изображения

На графиках времени выполнения параллельной реализации задачи и экспериментальной эффективности заметно влияние конкурентного выполнения процессов пользователя и служебных процессов операционной системы. Поэтому при параллельной реализации задач обработки на кластерных ВС нецелесообразно полностью загружать вычислительные узлы процессами обработки, необходимо оставлять резерв для выполнения задач операционной системы.

Во второй главе рассматриваются методы решения задачи согласования потенциального параллелизма с распределением данных, в частности, влияние различных вариантов представления и размещения данных в вычислительной системе при хранении, а также организации коммуникаций в процессе параллельной обработки на общую эффективность решения типовых задач обработки изображений.

Для сравнительного анализа различных вариантов организации хранения изображений разработаны модели, позволяющие оценивать время доступа к данным в зависимости от выбранной архитектуры системы и отражающие основные аспекты взаимодействия программы с устройствами хранения.

В диссертации представлена модель доступа к данным на ВЗУ, которая учитывает особенности конкурентного выполнения процессов пользователя и служебных процессов операционной системы. Модели операций чтения/записи данных с диска определены следующим образом:

$$\begin{aligned} I(\mathbf{a}) &= (\{\mathbf{x} : \mathbf{X}\} * (c? \mathbf{a}(\mathbf{x}) \rightarrow SKIP)) \parallel SYSTEM_READ, \\ SYSTEM_READ &= (-eof) * (READ(B)[\vartheta_r \cdot n_B]; \{b : B\} * (c!b \rightarrow SKIP)), \\ O(\mathbf{a}) &= ((\{\mathbf{x} : \mathbf{X}\} * (c! \mathbf{a}(\mathbf{x}) \rightarrow SKIP)); (c!eof \rightarrow SKIP)) \parallel SYSTEM_WRITE, \\ SYSTEM_WRITE &= (-eof) * ((\{b : B\} * ((c?b \rightarrow SKIP)|(c?eof \rightarrow SKIP))); WRITE(B)[\vartheta_w \cdot n_B]). \end{aligned}$$

В общем случае (как это отмечено в работах Ластовецкого и Донгара) точная модель времени выполнения операций ввода/вывода может быть получена лишь в виде некоторой экспериментальной зависимости от объема передаваемых данных

$$I(\mathbf{a})[\vartheta(n)], n = card(\mathbf{a} \in \mathbb{F}^X) \text{ и } O(\mathbf{a})[\vartheta(n)], n = card(\mathbf{a} \in \mathbb{F}^X).$$

Для большеформатных изображений такая зависимость может приближаться к линейной $I(\mathbf{a})[\approx \vartheta_r \cdot n], n = card(\mathbf{a} \in \mathbb{F}^X)$ и $O(\mathbf{a})[\approx \vartheta_w \cdot n], n = card(\mathbf{a} \in \mathbb{F}^X)$,

что подтверждается результатами экспериментов, показанных на рисунке 4.

Модель коммуникационных взаимодействий программ должна, с одной стороны, быть достаточно простой и конструктивной, а с другой стороны, отражать особенности выполнения процессов обработки изображений с помощью итераторов. Существующие модели времени выполнения коммуникационных операций в большинстве случаев не учитывают возможность совмещения процессов передачи данных и их обработки. В диссертации предложена конструктивная модель и методы оценивания ее параметров которые позволяют учесть указанные особенности. Ниже представлены модели коммуникационных взаимодействий, определяющие операции передачи и получения изображения:

$$\begin{aligned} S(\mathbf{a}) &= ((\{\mathbf{x} : \mathbf{X}\} * (c! \mathbf{a}(\mathbf{x}) \rightarrow SKIP)); (c!eof \rightarrow SKIP)) \parallel SYSTEM_SEND \\ SYSTEM_SEND &= (-eof) * ((\{b : B\} * ((c?b \rightarrow SKIP)|(c?eof \rightarrow SKIP))); COMM(B)[\vartheta_0 + \vartheta_1 \cdot n_B]) \\ R(\mathbf{a}) &= (\{\mathbf{x} : \mathbf{X}\} * (c? \mathbf{a}(\mathbf{x}) \rightarrow SKIP)) \parallel SYSTEM_RECV \\ SYSTEM_RECV &= (-eof) * (COMM(B)[\vartheta_0 + \vartheta_1 \cdot n_B]; \{b : B\} * (c!b \rightarrow SKIP)). \end{aligned}$$

Если операции обработки и коммуникации чередуются в теле итератора, то время выполнения такого итератора определяется следующим образом

$\{X_i : X\} * (F(X_i)[\vartheta_F]; S(X_i)) [(\max(\vartheta_F, \vartheta_0 + \vartheta_1 \cdot n)) \cdot k]$, где $X = \bigcup_{i=1}^k X_i$, $n = \text{card}(X_i)$.

Зависимость эффективности параллельной обработки от размера изображения, полученная с учетом операций ввода/вывода (рис. 5), показывает, что вспомогательные этапы чтения/записи изображений являются узким местом систем параллельной обработки изображений, использующих централизованное хранение.

Централизованный доступ к данным в значительной степени связан с тем, как организованы вычисления в большинстве существующих систем параллельной обработки. В этих системах используется программа-менеджер, которая выполняет декомпозицию изображений непосредственно перед обработкой, причем размер перекрытия обрабатываемых фрагментов выбирается в соответствии с параметрами выполняемых операций.

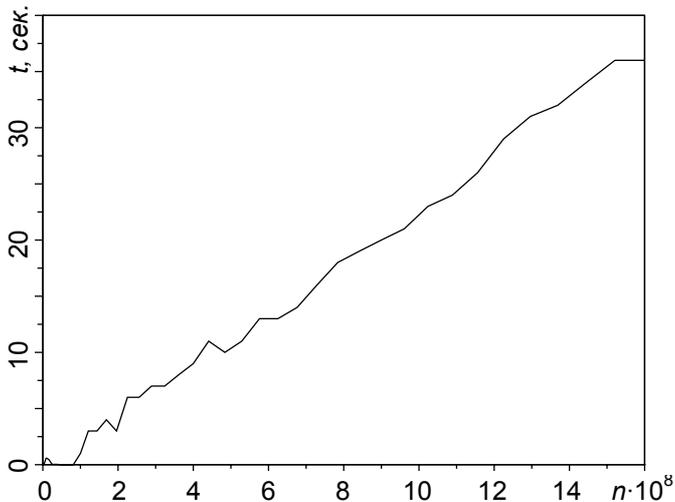


Рисунок 4 – Экспериментальная зависимость времени чтения данных изображения от размера изображения



Рисунок 5 – Зависимость эффективности от размера изображения с учетом операций ввода/вывода

Централизованный доступ к данным в значительной степени связан с тем, как организованы вычисления в большинстве существующих систем параллельной обработки. В этих системах используется программа-менеджер, которая выполняет декомпозицию изображений непосредственно перед обработкой, причем размер перекрытия обрабатываемых фрагментов выбирается в соответствии с параметрами выполняемых операций.

В диссертации предложен альтернативный подход, основанный на концепции распределенного изображения, которая обеспечивает распараллеливание операций доступа к данным, используя принцип "обрабатываем там, где храним".

Распределенное изображение – это структура данных, определяющая способ и параметры разбиения изображения на фрагменты, список компьютеров, где находятся эти фрагменты, место их размещения и формат хранения. Декомпозиция выполняется при создании или импорте изображения в систему и не изменяется в процессе хранения. При этом появляется задача выбора оптимальной декомпозиции (разбиения на фрагменты) изображения в условиях отсутствия априорной информации о параметрах запускаемых в системе задач обработки. Необходимо

также учесть проблемы, связанные с обеспечением отказоустойчивости распределенного хранения фрагментов изображений, сбалансированности загрузки компьютеров, участвующих в обработке при заранее выполненной декомпозиции данных, интерактивности системы при визуализации распределенных изображений.

Анализ вариантов декомпозиции изображений при выполнении различных операций обработки показал, что наиболее целесообразным выбором представляется декомпозиция распределенного изображения в виде перекрывающихся фрагментов.

В диссертации разработан новый метод формирования децентрализованных структур, определяющих размещение данных распределенных изображений, на основе декомпозиции перекрывающихся фрагментов. Новизна предлагаемого подхода заключается в том, что размер необходимого перекрытия фрагментов определяется не параметрами последующей задачи обработки, поскольку она априори неизвестна, а необходимостью решения проблем отказоустойчивости распределенного хранения фрагментов изображений и сбалансированности загрузки компьютеров.

Рассмотрим основные аспекты предлагаемого подхода на примере одномерной декомпозиции. Пусть необходимо разместить изображение, имеющее область определения X , на P узлах хранения.

Выполняется предварительная декомпозиция на $2P$ блоков:

$$X = \bigcup_{i=1}^{2P} X_{q|i}, \quad X_{q|i} \cap X_{q|j} = \emptyset, \quad \forall i \neq j.$$

Фрагмент распределенного изображения $a_{p|m}$ на m -ом узле хранения содержит два блока предварительной декомпозиции с номерами $2m-1$ и $2m$. Эти два блока называются основными и соответствуют варианту разбиения изображения на непересекающиеся фрагменты

$$X_{p|m} = X_{q|2m-1} \cup X_{q|2m}.$$

Декомпозиция в виде перекрывающихся фрагментов формируется путем добавления теневого блока, блока предварительной декомпозиции с номерами $2m-2$ и $2m+1$, т.е. копий примыкающих к данному узлу основных блоков соседних узлов хранения. В свою очередь, младший основной блок узла хранится в качестве одного из теневых блоков на узле с меньшим номером, а старший – на узле с большим номером. Для трех последовательных узлов хранения получаем

$$X_{p|m-1}^{\pm} = X_{q|2m-4} \cup X_{q|2m-3} \cup X_{q|2m-2} \cup X_{q|2m-1}$$

$$X_{p|m}^{\pm} = X_{q|2m-2} \cup X_{q|2m-1} \cup X_{q|2m} \cup X_{q|2m+1},$$

$$X_{p|m+1}^{\pm} = X_{q|2m} \cup X_{q|2m+1} \cup X_{q|2m+2} \cup X_{q|2m+3}.$$

Именно данные основных блоков $X_{q|2m-1} \cup X_{q|2m}$ формируются на m -ом узле в процессе распределенной обработки изображения. По окончании процесса обработки выполняется обмен теневыми данными.

Такая структура данных распределенного изображения позволяет восстановить его при отказе одного из узлов хранения, а также обеспечивает возможность

формировать отсчеты основных блоков нового изображения без передачи данных между соседними узлами при многоэтапном выполнении локальных операций в процессе распределенной обработки изображения.

Предложенный способ организации распределенных изображений по построению обеспечивает отказоустойчивость за счет использования полного перекрытия областей при формировании частичных изображений.

Минимизацию избыточности, необходимой для восстановления информации при отказе одного из узлов хранения, обеспечивает одномерная декомпозиция в виде перекрывающихся на половину своей высоты горизонтальных полос. В данном случае суммарный объем данных, занимаемый изображением в распределенной системе хранения, возрастает в два раза. Использование двумерной декомпозиции с аналогичными размерами перекрытия увеличивает объем хранимой информации в четыре раза, не обеспечивая при этом повышения уровня отказоустойчивости системы в целом.

В диссертации предложен путь снижения степени избыточности распределенной системы хранения за счет использования иерархических методов компрессии изображений, имеющих поддержку мультиразрешения. Существующие варианты таких методов обеспечивают компрессию данных без внесения погрешности, что необходимо для использования в составе распределенной системы обработки изображений. Предложена организация данных для программных средств, реализующих распределенные изображения, которая использует трехуровневую иерархическую структуру: уровень эскиза, обеспечивающий быструю визуализацию всего изображения, промежуточный уровень, уточняющий представление изображения при увеличении степени детальности просмотра и являющийся основой для формирования реальных отсчетов изображения с помощью иерархических методов компрессии, и собственно уровень исходных отсчетов, представленный при хранении упакованной информацией, обеспечивающей точное восстановление исходных отсчетов по отсчетам промежуточного уровня.

В диссертации проведен сравнительный анализ эффективности различных вариантов декомпозиции, обоснован выбор параметров распределения данных в трехуровневой иерархической структуре.

В третьей главе решаются проблемы динамической организации данных при хранении и параллельной обработке изображений, в частности, разработаны методы динамического размещения данных при хранении и динамической балансировки вычислительных процессов при обработке изображений.

Приводится обзор основных методов централизованного и децентрализованного управления процессом вычислений в параллельных и распределенных системах. Рассматриваются основные алгоритмы сбалансированной загрузки процессоров и их применение к задачам параллельной обработки изображений, статические и динамические методы выравнивания нагрузки (балансировки), анализируются границы применимости, преимущества, сложности и недостатки. Исследуются существующие методы динамической

балансировки путем оценки вычислительных издержек и эффективности параллельной обработки на моделях.

В диссертации предложены принципы динамической организации хранения распределенного изображения, которые обеспечивают размещение данных, минимизирующее среднее время реакции на запрос пользователя.

Распределенное изображение в системе может быть создано двумя способами: импорт файла изображения или набора файлов изображений, составляющих панораму; создание нового изображения в процессе обработки существующего распределенного изображения.

При импорте изображения выполняется разбиение изображения на фрагменты в соответствии с предложенным во второй главе методом формирования децентрализованных структур данных изображений на основе декомпозиции в виде перекрывающихся фрагментов. Размещение полученных таким образом фрагментов изображения выполняется на доступных узлах хранения в соответствии с выбранными критериями.

Данные распределенного изображения после его создания не изменяются, при обработке создается новое распределенное изображение. Фрагменты нового изображения, создаваемого при обработке существующего распределенного изображения, размещаются на тех же узлах, где размещались фрагменты исходного изображения.

В процессе работы системы фрагменты не перемещаются. Фрагменты только дублируются или удаляются. Фрагменты дублируются:

- при доступе к распределенному изображению пользователя, у которого на компьютере еще нет ни одного фрагмента данного изображения; пользователю дублируется тот фрагмент, время доступа к которому с компьютера пользователя максимально; если разброс времени доступа ко всем фрагментам лежит в допустимом диапазоне (это параметр системы), то на компьютер пользователя дублируется тот фрагмент, который имеет минимальный уровень репликации в системе, в противном случае – дублируемый фрагмент выбирается случайно;
- при отказе (тайм-ауте) доступа к узлу хранения формируется копия недоступного фрагмента по данным тех узлов хранения, которые хранят теневые копии его основных блоков; новое размещение выбирается из числа доступных узлов в системе, имеющих минимальную загрузку предоставленной системе области хранения данных.

Фрагменты удаляются в процессе оптимизации системы хранения или по команде пользователя:

- если фрагмент имеет уровень репликации выше, чем заданный уровень репликации в системе (параметр системы) и время последнего доступа (использования его данных) к этому фрагменту превышает заданное время (параметр системы);
- если пользователь дал команду на удаление всего распределенного изображения; в этом случае анализируется история использования этого изображения, если с ним работал только пользователь, издавший запрос на удаление, то все

фрагменты изображения удаляются из системы; если изображение использовалось другими пользователями, то удаляется (или перемещается на другой узел, в зависимости от уровня репликации фрагмента) только фрагмент, размещенный на компьютере пользователя, издавшего запрос.

Соответственно, на компьютере пользователя хранится по одному фрагменту каждого распределенного изображения с которыми он работает. Если пользователь длительное время не использовал какое-либо изображение, т.е. не обрабатывал и не просматривал его, то данный фрагмент распределенного изображения может быть удален с компьютера пользователя с тем, чтобы оптимизировать размещение этого распределенного изображения в системе.

В диссертации на основе предложенного во второй главе метода декомпозиции распределенного изображения разработан оригинальный алгоритм динамического распределения загрузки процессоров при выполнении локальной обработки.

Рассмотрим суть алгоритма на примере взаимодействия двух соседних узлов, m -ого и $(m+1)$ -го, при формировании той части результирующего изображения, которая содержится в блоках предварительной декомпозиции данных с номерами $2m$ и $2m+1$. Заметим, что каждый из узлов имеет всю информацию, чтобы сделать эту работу самостоятельно (или почти всю, в случае выполнения операции локальной обработки скользящим окном).

m -й узел начинает формировать строки $2m$ -го блока, начиная с первой строки в порядке возрастания номеров строк, $(m+1)$ -й узел, в свою очередь, формирует строки $(2m+1)$ -го блока, но, начиная с последней строки блока, в порядке убывания номеров строк. После формирования определенного количества строк каждый узел информирует соседний узел о времени, за которое он выполнил эту работу. На основании этой информации вычисляется прогнозируемый номер строки изображения, на которой узлы завершат такую совместную обработку.

Таким образом, в процессе работы смежные узлы двигаются навстречу друг другу, сообщая о скорости своего процесса вычислений при достижении заранее определенных моментов. При этом прогнозируемый номер строки изображения, на которой эти процессы встретятся, постоянно корректируется в зависимости от текущей загрузки вычислительных узлов. Таким образом, все процессы завершат свою работу практически одновременно. Разница во времени при этом составит не больше, чем время обработки одной строки.

Одновременно каждый узел участвует в двух таких процессах, попеременно формируя строки старшего блока в порядке возрастания номера и строки своего младшего блока в порядке убывания, при необходимости переходя к формированию строк теневого блока.

В результате будет сформировано новое распределенное изображение в полном объеме, но возможно с неравномерным размещением данных по узлам хранения. Далее в фоновом режиме узлы хранения сформированного распределенного изображения обмениваются своими данными с тем, чтобы привести структуру распределенного изображения к необходимому виду. Однако пользователь может получить результат своего запроса сразу по завершении процесса обработки.

В диссертации проведено аналитическое исследование эффективности предложенного алгоритма для идеализированного случая несбалансированной вычислительной системы, у которой производительность одного из узлов выше остальных в $\mu > 1$ раз. Показано, что если распределенная система состоит из четырех узлов ($M = 4$), то предлагаемый алгоритм может равномерно распределить нагрузку даже при трехкратном превышении производительности одного из узлов над остальными. Более производительный узел обеспечит обработку половины строк изображения, а на долю остальных узлов достанется по 1/6 части.

При увеличении числа узлов обработки/хранения разброс производительности, который может быть скомпенсирован без временных потерь, постепенно уменьшается до двухкратного. Относительное сокращение времени обработки изображения от повышения производительности одного узла при различном количестве вычислительных узлов в распределенной системе приведено на рис. 6.

Достоинство предлагаемого алгоритма заключается в том, что он является полностью децентрализованным, и обеспечивает равномерное распределение нагрузки, если производительность соседних узлов не отличается больше чем в два раза.

Разработаны формальные модели, описывающие разработанный алгоритм динамического распределения загрузки процессоров при выполнении операций локальной обработки.

Выполнен анализ данных моделей, получены оценки влияния гетерогенности функционирования параллельной системы на время выполнения задачи обработки изображений.

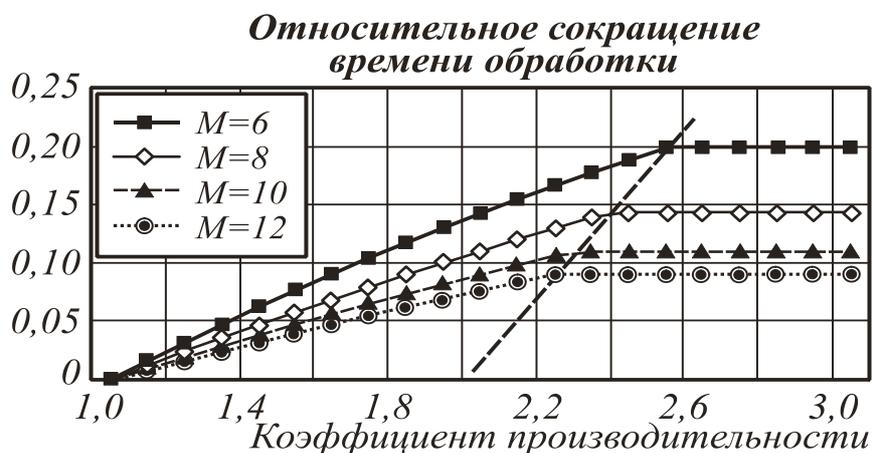


Рисунок 6 – Относительное сокращение времени обработки изображения при повышении производительности одного из узлов

Проведены вычислительные эксперименты по сравнению времени решения задач обработки изображений с использованием предлагаемой динамической балансировки и без нее. Эксперименты проводились на вычислительном кластере HP c3000, имеющем 14 вычислительных узлов. Исследовались режимы монопольного использования и разделяемого доступа, различные модели нагрузки. Эксперименты показали, что при возрастании вычислительной нагрузки вдвое на 7 узлах (из 14) использование предлагаемой динамической балансировки

обеспечивает время решения модельной задачи, пропорциональное общей нагрузке кластера. Результаты экспериментов в режиме монопольного использования хорошо согласуются с модельными оценками времени выполнения задач обработки изображений.

В четвертой главе рассматривается реализация разработанных в диссертации методов организации распределенного хранения и параллельной обработки крупноформатных изображений, описываются программные комплексы, позволяющие решать практические задачи обработки изображений.

Разработана архитектура сервис-ориентированной распределенной системы обработки крупноформатных изображений (рис.7), основанная на методах динамической организации данных распределенных изображений. В системе используется нетрадиционный "ориентированный-на-данные" подход к организации вычислений в процессе обработки изображений, при котором данные изображений, точнее, их фрагменты заранее распределены по компьютерам распределенной системы, а в процессе обработки на эти компьютеры вместо данных изображений рассылаются процедуры их обработки.

Основной структурной единицей системы является распределенное изображение, которое доступно пользователю как набор взаимодействующих сервисов хранения, т.е. каждый фрагмент распределенного изображения представляется в виде отдельного сервиса (фрейм-сервиса) со своим уникальным именем в иерархии изображений-сервисов. Декомпозиция данных при формировании федерации фрейм-сервисов одного изображения выполняется на основе методов, изложенных в главе 2, которые обеспечивают отказоустойчивость распределенного хранения данных изображений и децентрализованное динамическое выравнивание нагрузки при выполнении операций локальной обработки изображений в распределенной системе.

Для осуществления служебных функций реализован системный уровень сервисов, включающий мастер-сервис и узел-сервисы. Мастер-сервис содержит ссылки на доступные системе узел-сервисы и все фрейм-сервисы распределенных изображений, размещенных в системе. Мастер-сервис выполняет импорт и экспорт изображений, создание и размещение их в виде федерации фрейм-сервисов, периодически запускает и контролирует процесс оптимизации размещения распределенных изображений в системе.

На каждом узле, подключенном к системе, запущен узел-сервис, который содержит информацию о конфигурации предоставленных системе ресурсов и ссылки на фрейм-сервисы распределенных изображений, размещенных на данном узле. Узел-сервис создает новый фрейм-сервис по запросу от мастер-сервиса или другого фрейм-сервиса, запущенного на данном узле, при этом он размещает на своем узле хранения данные фрагмента нового распределенного изображения и запускает соответствующий этому фрагменту фрейм-сервис.

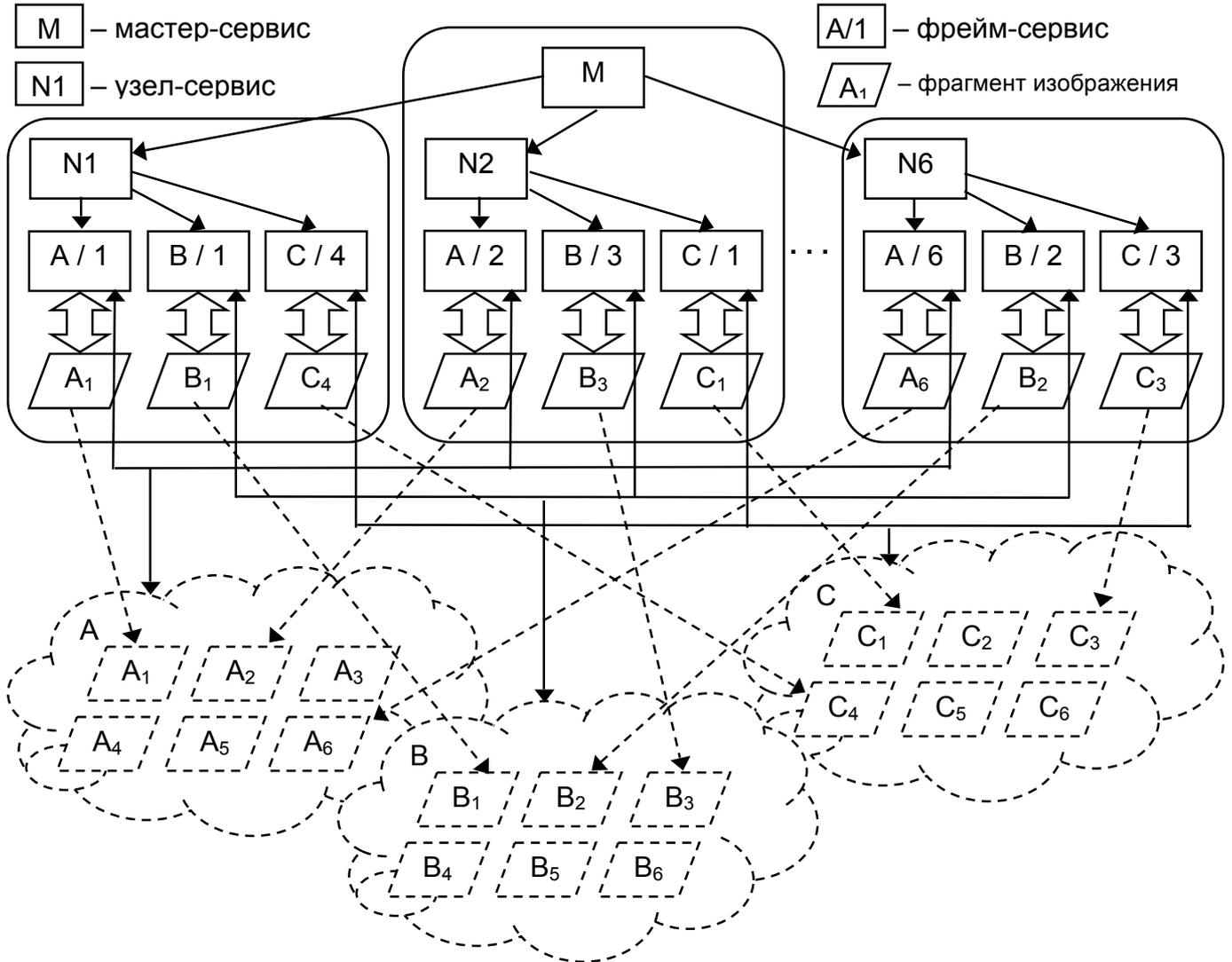


Рисунок 7 – Архитектура сервис-ориентированной распределенной системы обработки изображений

Каждому фрагменту соответствует фрейм-сервис, который содержит информацию о расположении всех других фрагментов. Фрейм-сервис помимо интерфейса доступа к данным имеет интерфейс, обеспечивающий получение и выполнение задания на обработку контролируемых данных. Результатом обработки является новое распределенное изображение, фрагменты которого могут быть не совсем согласованы между собой в части обеспечения последующей сбалансированности распределенной обработки. Но наличие "интеллекта" у фрейм-сервиса, связей с соседями позволяет выполнить все необходимые для такого согласования действия автономно, в фоновом режиме, предоставляя пользователю необходимые ему функции (чаще всего это визуализация) и данные непосредственно сразу после создания.

Быстрая визуализация крупноформатных распределенных изображений, реализованная в данной системе, основывается на предложенном в главе 2 способе организации данных внутри фрейм-сервисов на основе пирамидальных структур и иерархических методов компрессии. Функция быстрой визуализации изображений большого размера обеспечивается наличием у каждого фрейм-сервиса так называемого эскиза изображения и заранее сформированного слоя данных с

промежуточным уровнем детальности. Визуализирующее приложение, обращаясь к сервису с наиболее быстрым доступом, получает данные, необходимые для предварительного просмотра изображения. При просмотре изображения в полноразмерном режиме пользователю отображается только выбранный им фрагмент, формируемый как результат обращения к нескольким фрейм-сервисам. При прокрутке изображения в окне происходит последовательное обращение к необходимым фрейм-сервисам, которые обеспечивают формирование и подкачку данных. Для обеспечения комфортного уровня интерактивности подкачка выполняется в режиме чтения с упреждением (prepaging). Наличие перекрытий хранящихся на фрейм-сервисах фрагментов делает этот процесс более плавным.

Отметим, что предлагаемый подход позволил реализовать в данной распределенной системе возможность формирования крупноформатного распределенного изображения через объединение нескольких различных перекрывающихся снимков панорамы. При этом взаимодействующие сервисы распределенной системы хранения самостоятельно (или с минимальным вмешательством пользователя) решают вопросы сшивки, т.е. нахождения общих точек, попарное согласование локальных координат каждого фрагмента и отображение их на глобальные координаты полного распределенного изображения.

В диссертации разработаны интерфейсы основных сервисов системы и API для доступа к системе из прикладных программ пользователя. Разработано приложение для итерактивного взаимодействия с распределенной системой: импорта, экспорта и визуализации изображений, формирования заданий обработки изображений.

В диссертации описывается система параллельной обработки крупноформатных изображений, которая является модификацией распределенной системы обработки, реализованной на кластерной вычислительной системе. Данная система имеет следующие отличительные особенности:

- каждому пользователю выделяется вычислительный узел, который является основным для использования системы обработки изображений и упорядоченный список дополнительных вычислительных узлов, которые используются при отказе или загруженности основного; основной узел в кластерной ВС является аналогом пользовательского компьютера в распределенной системе;
- при оптимизации размещения фрейм-сервисов не учитывается время доступа основного узла пользователя к другим узлам хранения – оно считается одинаковым для всех узлов;
- мастер-сервис размещен на управляющем узле вычислительного кластера и дополнительно совмещает функции всех узел-сервисов системы.

Система параллельной обработки крупноформатных изображений реализована на вычислительном кластере HP c3000 на базе 7 сдвоенных блейд серверов HP ProLiant BL2x220c G5 в качестве вычислительных узлов и 1-го сервера HP ProLiant BL260c в качестве управляющего узла. Основные характеристиками кластера: 1 управляющий узел (CPU: Quad-Core Intel Xeon 5430 2.66 GHz, cache 12 Mb, 1333MHz; RAM: 4Gb DDR2-667; HDD: 240Gb) и 14 вычислительных узлов (2 CPU: Quad-Core Intel Xeon 5430 2.33 GHz, cache 12 Mb, 1333MHz; RAM: 8Gb DDR2-667;

HDD: 120Gb). Данная система использовалась в качестве тестовой при проведении вычислительных экспериментов, описанных в диссертации.

В этой главе также представлена инструментальная программная система обработки изображений с модульной динамической структурой, предназначенная для разработки, моделирования и исследования алгоритмов обработки изображений. Система позволяет в условиях ограничений на объем оперативной памяти интерактивно выполнить сложную обработку изображения, задавая ее как программную сеть на множестве модулей, реализующих простые операции над изображением, не создавая при этом промежуточных временных изображений на диске. Заложенные при разработке этой программной системы особенности проектных решений позволили без существенной переработки реализовать версию для распределенной обработки изображений.

Рассмотрена программная система передачи крупноформатных изображений по сети Internet, предназначенная для распределенных офисов, ориентированных на работу с изображениями. В системе для сжатия изображений используется метод иерархической сеточной интерполяции, выбор которого обусловлен особо высокими требованиями к качеству и степени сжатия передаваемой информации в задачах издательской деятельности.

Приведены две системы технического зрения (СТЗ), при разработке которых в качестве исследовательской системы использовалась сервис-ориентированная распределенная система обработки изображений.

СТЗ "Система регистрации железнодорожных составов цистерн" (рис. 8) решает актуальную задачу ввода информации о составах в рамках автоматизации всего технологического процесса наполнения железнодорожных составов цистерн на нефтеналивных терминалах. Рассмотрены особенности распределенной архитектуры программного комплекса данной СТЗ и проектные решения, связанные с особенностями процесса перемещения цистерн при наливе. Описывается дополнительно реализованный распределенный программный комплекс тестирования и настройки параметров СТЗ, который позволил сократить время получения скорректированных значений с нескольких суток до нескольких часов.

СТЗ "Система регистрации железнодорожных составов цистерн" успешно внедрена и работает на ОАО "Самара-терминал" (г. Сызрань) и на Уфимском нефтеперерабатывающем заводе.

СТЗ лабораторного анализа (рис. 9), предназначенная для обнаружения локальных неоднородностей при наблюдении длительного процесса, внедрена и успешно эксплуатируется уже более четырех лет на ЗАО «КуйбышевАзот» (г.Тольяти) при проведении лабораторного анализа для определения количества гель-частиц в растворе полимера.

Использование СТЗ вместо наблюдателя при проведении данного анализа позволило существенно снизить психовизуальную нагрузку на лабораторный персонал и, тем самым, повысить точность и достоверность определения количества гель-частиц в растворе полимера, обеспечить документируемость проводимых лабораторных анализов. Практика показала, что автоматическое распознавание с

использованием описанной СТЗ позволяет выявить в 5-7 раз большее количество разрывов по сравнению с визуальным наблюдением.

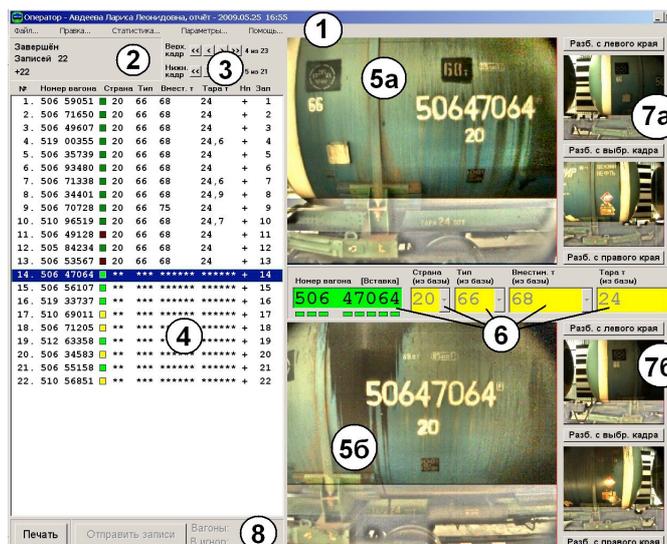


Рисунок 8 – Пользовательский экран рабочего места оператора системы регистрации железнодорожных составов цистерн

СТЗ для обнаружения локальных неоднородностей вполне применима и в таких областях как охранные системы, системы контроля, автоматизированные линии и т.п., где необходимо отслеживать изменения наблюдаемых объектов при мониторинге длительных процессов.

В приложениях с основному тексту диссертации приводятся основные понятия и законы теории взаимодействующих последовательных процессов, используемые в работе, характеристики СТЗ, а также документы, подтверждающие использование результатов диссертации.

ЗАКЛЮЧЕНИЕ

В диссертации на основе методов динамической организации данных, моделей итераторов обработки и правил их эквивалентного преобразования решена задача моделирования и формирования структуры распределенных систем обработки изображений с различными типами параллелизма. Совокупность полученных результатов является решением фундаментальной научной проблемы – "проблемы отображения" для широко распространенного класса задач математической обработки изображений.

В диссертационной работе получены следующие основные результаты:

1. Разработана модель информационных технологий обработки изображений на основе алгебры изображений и теории взаимодействующих последовательных процессов, которая объединяет формальные модели алгоритмов обработки изображений, модели организации данных при хранении изображений и коммуникации, а также модели предиктивной оценки времени выполнения задач обработки. Разработанные модели позволяют реализовать сквозную методологию анализа и эквивалентного преобразования информационных

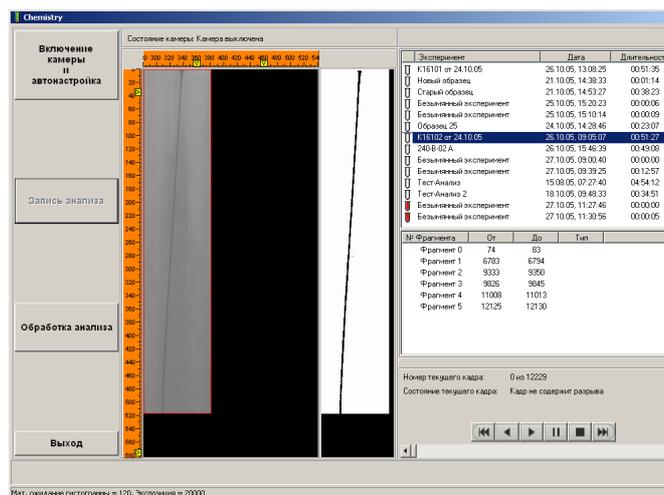


Рисунок 9 – Пользовательский экран СТЗ для обнаружения локальных неоднородностей

моделей процессов параллельной обработки изображений, которая охватывает все уровни иерархии общей формальной модели и позволяет выявить весь ресурс параллелизма программы в терминах параллелизма по данным и параллелизма по вычислениям.

2. Сформулирована формальная модель итераторов преобразования данных при обработке изображений, в том числе правила эквивалентного преобразования композиции итераторов и алгоритмы пересчета времени выполнения задачи при эквивалентных преобразованиях моделей в терминах итераторов. Данные правила и алгоритмы используются при построении информационной модели и анализе потенциального параллелизма.
3. Предложен новый метод формирования децентрализованных структур, определяющих размещение распределенных данных изображений на основе декомпозиции перекрывающихся фрагментов, позволяющий динамически балансировать загрузку компьютеров и обеспечивающий отказоустойчивость распределенного хранения фрагментов изображений. Предложен новый способ представления данных распределенного изображения на основе трехуровневой иерархической структуры, обеспечивающий быструю визуализацию распределенного изображения и реализующий эффективные методы повышения отказоустойчивости распределенного хранения изображений на основе комбинирования специализированных методов компрессии данных и алгоритмов внесения избыточности средствами репликации в рамках предложенного метода декомпозиции распределенного изображения.
4. Разработан оригинальный алгоритм динамической балансировки многопроцессорных систем при распараллеливании операций обработки изображений, который опирается на предложенный метод декомпозиции распределенного изображения. Алгоритм является полностью децентрализованным, что является оптимальным при использовании его в распределенных системах. Алгоритм позволяет обеспечить равномерное распределение нагрузки при параллельной обработке изображений в условиях гетерогенной вычислительной среды, причем возможно полное выравнивание нагрузки при двухкратном различии производительности соседних узлов.
5. Разработан децентрализованный подход к организации вычислений в процессе обработки изображений, ориентированный на данные, при котором фрагменты изображений заранее распределены по компьютерам распределенной системы, а в процессе обработки на компьютеры распределенной системы вместо данных изображений рассылаются процедуры их обработки.
6. Предложена сервис-ориентированная архитектура построения распределенных систем обработки изображений, на основе методов динамической организации распределенных изображений в виде набора взаимодействующих сервисов хранения – федерации фрейм-сервисов.
7. Разработан ряд комплексов программ для систем обработки изображений, позволяющих решать практические задачи и использующих основные результаты диссертационной работы, в частности, для сервис-ориентированной

распределенной системы обработки крупноформатных изображений, кластерной системы формирования и параллельной обработки крупноформатных изображений, программной системы передачи крупноформатных изображений по сети Internet, системы технического зрения для регистрации железнодорожных составов цистерн, системы технического зрения для определения количества гель-частиц в растворе полимера.

Основные результаты диссертации отражены в следующих публикациях

Монографии

1. Мясников, В.В. Теоретические основы цифровой обработки изображений [Текст] / В.В. Мясников, С.Б. Попов, В.В. Сергеев, В.А. Сойфер // Методы компьютерной обработки изображений / под общ. ред В.А. Сойфера; М.В. Гашников [и др.] – 2-ое изд., испр. – М.: Физматлит, 2003. – Часть I. – С.13-298.
2. Гашников, М.В. Компрессия изображений [Текст] / М.В. Гашников, Н.И. Глумов, С.Б. Попов, В.В. Сергеев // Методы компьютерной обработки изображений / под общ. ред В.А. Сойфера; М.В. Гашников [и др.] – 2-ое изд., испр. – М.: Физматлит, 2003. – Часть II, Глава 6. – С.385-458.
3. Computer Image Processing, Part I: Basic concepts and theory [Текст] / edited by Victor A. Soifer; V.V. Myasnikov, S.B. Popov, V.V. Sergeyev, V.A. Soifer – VDM Verlag, 2010. – 286 p.
4. Gashnikov, M.V. Image Compression [Текст] / M.V. Gashnikov, N.I. Glumov, S.B. Popov, V.V. Sergeyev // Computer Image Processing, Part II: Methods and algorithms / edited by Victor A. Soifer; A.V. Chernov [et al.] – VDM Verlag, 2010. – Chapter 6. – P.87-160.

Статьи в ведущих рецензируемых журналах и изданиях, входящих в перечень ВАК

5. Арефьев, Е.Ю. Опыты по реконструктивной томографии с использованием автоматизированной системы обработки изображений [Текст] / Е.Ю. Арефьев, И.Д. Багбая, К.В. Овчинников, С.Б. Попов, И.Н. Сисакян, В.А. Сойфер // Компьютерная оптика. – 1987. – Вып. 2 – С.31-35.
6. Арефьев, Е.Ю. Контроль фазового микрорельефа элементов компьютерной оптики [Текст] / Е.Ю. Арефьев, М.А. Голуб, К.В. Овчинников, С.Б. Попов, И.Н. Сисакян, В.А. Сойфер, Д.Н. Тихонов, А.Г. Храмов, Г.В. Шамалова // Журнал технической физики. – 1990. – Том 60, вып. 6. – С.157-161.
7. Popov, S.B. Investigation of Modifications of Algorithms for Fractal Image Encoding [Текст] / S.B. Popov, I.A. Khasanov // Pattern Recognition and Image Analysis. – 1996. – Vol. 6, No.1. – p.174.
8. Popov, S.B. Architecture of the Software for Image Processing in OS/2 [Текст] / S.B. Popov, V.V. Sergeyev, N.I. Frolova // Pattern Recognition and Image Analysis. – 1996. – Vol. 6, No.2. – p.432.
9. Glumov, N.I. Some Application Shells of Image Processing for IBM PCs [Текст] / N.I. Glumov, V.V. Myasnikov, S.B. Popov, P.V. Raudin, V.V. Sergeyev, N.I. Frolova, A.V. Chernov // Pattern Recognition and Image Analysis. – 1996. – Vol. 6, No. 2. – p.372.

10. Popov, S.B. Scalable Automatic System of Image Processing with the Possibilities of Adaptation and Distributed Processing [Текст] / S.B. Popov // Pattern Recognition and Image Analysis. – 1998. – Vol.8, No.3. – pp.380-381.
11. Gashnikov, M.V. Software System for Transmitting Large-Size Images via the Internet [Текст] / M.V. Gashnikov, N.I. Glumov, S.B. Popov, V.V. Segreyev, E. A Farberov // Pattern Recognition and Image Analysis. – 2001. – Vol. 11, No.2. – pp.430-432.
12. Попов, С.Б. Кластерная технология формирования и параллельной фильтрации больших изображений [Текст] / С.Б. Попов, В.А. Соيفер, А.А. Тараканов, В.А. Фурсов // Компьютерная оптика. – 2002. – Вып. 23. – С.75-78.
13. Никоноров, А.В. Сравнительный анализ моделей цветообразования при офсетной многокрасочной печати [Текст] / А.В. Никоноров, С.Б. Попов // Компьютерная оптика. – 2002. – Вып. 23. – С.79-83.
14. Никоноров, А.В. Принцип согласованности оценок в задаче идентификации моделей цветопроизведения [Текст] / А.В. Никоноров, С.Б. Попов, В.А. Фурсов // Компьютерная оптика. – 2002. – Вып. 24. – С.148-151.
15. Дроздов, М.А. Кластерная технология определения восстанавливающих фильтров и обработки больших изображений [Текст] / М.А. Дроздов, Д.И. Зимин, С.Б. Попов, С.А. Скуратов, В.А. Фурсов // Компьютерная оптика. – 2003. – Вып. 25. – С.175-182.
16. Nikonorov, A. Identifying Color Reproduction Models [Текст] / Nikonorov A., Popov S., Fursov V. // Pattern Recognition and Image Analysis. – 2003. – Vol. 13, No.2. – pp. 315–318.
17. Volotovskii, S.G. Machine Vision System for Registration of Oil Tank Wagons [Текст] / Volotovskii S.G., Kazanskii N.L., Popov S.B., Khmelev R.V. // Pattern Recognition and Image Analysis. – 2005. – Vol. 15, No. 2. – pp.461-463.
18. Волоотовский, С.Г. Система технического зрения для распознавания номеров железнодорожных цистерн с использованием модифицированного коррелятора в метрике Хаусдорфа [Текст] / С.Г. Волоотовский, Н.Л. Казанский, С.Б. Попов, Р.В. Хмелев // Компьютерная оптика. – 2005. – Вып. 27. – С.177-184.
19. Буланов, А.П. Система технического зрения для регистрации железнодорожных составов цистерн [Текст] / Буланов А.П., Волоотовский С.Г., Казанский Н.Л., Попов С.Б., Хмелев Р.В., Шумаков С.М. // Автоматизация в промышленности. – 2005. – № 6. – С. 57-59.
20. Игнатов, Н.А. Моделирование системы управления динамометрическим стендом [Текст] / Н.А. Игнатов, Н.Л. Казанский, Ю.И. Корнев, С.Б. Попов // Вестник Самарского государственного технического университета. Серия Физико-математические науки. – 2005. – Вып. 38. – С.115-121.
21. Волоотовский, С.В. Распознавание номеров железнодорожных цистерн с использованием быстрой локализации и модификации алгоритма сравнения объекта с эталоном по среднеквадратической метрике Хаусдорфа [Текст] / С.В. Волоотовский, Н.Л. Казанский, С.Б. Попов, Р.В. Хмелев // Обзорение прикладной и промышленной математики. – 2005. – Том 12, № 3. – С. 714.
22. Попов, С.Б. Концепция распределенного хранения и параллельной обработки крупноформатных изображений [Текст] / С.Б. Попов // Компьютерная оптика. – 2007. – Том 31, № 4. – С. 77-85.

23. Казанский, Н.Л. Система технического зрения для определения количества гелевых частиц в растворе полимера [Текст] / Н.Л. Казанский, С.Б. Попов // Компьютерная оптика. – 2009. – Том 33, № 3. – С. 325-331.
24. Kazanskiy, N. L. Machine Vision System for Singularity Detection in Monitoring the Long Process [Текст] / N. L. Kazanskiy and S. B. Popov // Optical Memory and Neural Networks (Information Optics). – 2010. – Vol. 19, No. 1. – pp. 23-30.
25. Попов, С.Б. Моделирование информационной структуры параллельной обработки изображений [Текст] / С.Б. Попов // Компьютерная оптика. – 2010. – Том 34, № 2. – С.231-242.
26. Казанский, Н.Л. Использование волноводного резонанса для создания нанооптических спектральных пропускающих фильтров [Текст] / Н.Л. Казанский, П.Г. Серафимович, С.Б. Попов, С.Н. Хонина // Компьютерная оптика. – 2010. – Том 34, № 2. – С.162-168.

Свидетельства о регистрации программ для ЭВМ и патент на изобретение

27. Программное обеспечение обработки изображений IPS 1.0 RSX11M / Бамбулевич К.Э., Васин А.Г., Маслов А.М., Попов С.Б., Сергеев В.В., Соيفер В.А. // Гос. фонд алгоритмов и программ СССР, № 50850000495, 1985.
28. Система регистрации железнодорожных составов цистерн / Волотовский С.Г., Казанский Н.Л. Попов С.Б. // Свидетельство об официальной регистрации программ для ЭВМ № 2004611969 по заявке № 2004611381 от 29 июня 2004 г. Зарегистрировано в Реестре программ для ЭВМ 26 августа 2004 г.
29. Программное обеспечение распознавания номеров на основе анализа топологии контуров/ Волотовский С.Г., Казанский Н.Л., Попов С.Б. // Свидетельство об официальной регистрации программ для ЭВМ № 2004611970 по заявке № 2004611382 от 29 июня 2004 г. Зарегистрировано в Реестре программ для ЭВМ 26 августа 2004 г.
30. Программное обеспечение распознавания номеров на основе анализа взаимных отклонений геометрических форм объекта и эталона / Хмелев Р.В., Казанский Н.Л., Попов С.Б. // Свидетельство об официальной регистрации программ для ЭВМ № 2004611971 по заявке № 2004611383 от 29 июня 2004 г. Зарегистрировано в Реестре программ для ЭВМ 26 августа 2004 г.
31. Спектральный идентификатор моделей цветовоспроизведения / Никоноров А.В., Попов С.Б., Фурсов В.А. // Свидетельство об официальной регистрации программ для ЭВМ № 2007614955 по заявке № 2007613925 от 5 октября 2007 г. Зарегистрировано в Реестре программ для ЭВМ 3 декабря 2007 г.
32. Модуль реализации модели цветовоспроизведения Нойгебауэра-Юла-Нейлсона/ Никоноров А.В., Попов С.Б., Фурсов В.А. // Свидетельство об официальной регистрации программ для ЭВМ № 2008611512 по заявке № 2008610367 от 5 февраля 2008 г. Зарегистрировано в Реестре программ для ЭВМ 25 марта 2008 г.
33. Способ распознавания разрывов струи раствора на изображении / Казанский Н.Л., Козин Н.Е., Попов С.Б., Фурсов В.А. // Патент РФ на изобретение № 2336563 от 20.10.2008 года по заявке № 2006115915/09 от 10.05.2006 года. Бюл. № 29.