

Докладчик МУРЗИН Федор Александрович

Тема : ОРГАНИЗАЦИЯ И АНАЛИЗ МНОГОМЕРНЫХ И НЕОДНОРОДНЫХ ДАННЫХ В ЗАДАЧАХ ОБРАБОТКИ ИЗОБРАЖЕНИЙ, ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКЕ, ГЕОФИЗИКЕ И ЛИНГВИСТИКЕ

диссертация на соискание ученой степени
доктора технических наук по специальности
05.13.17 – Теоретические основы информатики

В настоящее время активно развиваются такие направления информационных технологий, как специализированная обработка изображений и сигналов, а также автоматический анализ текстов на естественном языке. Решение подобных задач возможно только с использованием многопроцессорных систем, суперкомпьютеров и облачных технологий. Во всех этих направлениях необходимо выполнять обработку многомерных и неоднородных данных, которые должны специальным способом структурироваться, храниться и обрабатываться. Задачи такого класса относятся к вычислительно трудоемким, требующим распараллеливание по набору процессоров и использующих динамическую память с параллельным доступом. Для их решения целесообразно разрабатывать не только методы, но и схемы вычислительных устройств, которые позволяют наиболее эффективно их реализовать.

В связи с этим возникает необходимость в проведении теоретических исследованиях создания и использования новых архитектур вычислительных систем и методов организации памяти с параллельным доступом, а также создания и анализа для них эффективных алгоритмов обработка изображений, сигналов и автоматического анализа текстов.

В первой главе приводится описание принципов, лежащих в основе организации компьютерной памяти с параллельным доступом к информации.

Доказываются теоремы, представляющие собой обобщения на многомерный случай результатов Д. Ван Воорхиса и Т. Моррина. В результате, описывается модель памяти, предназначенной для хранения произвольного набора конечномерных массивов. При этом в предлагаемой модели возможно осуществление параллельного доступа к сечениям, выделяемым в массивах фиксацией одной из координат, и к большому набору многомерных параллелепипедов, являющихся подмассивами той же размерности в исходных массивах.

Выбор массивов, а не каких-либо других типов данных, обусловлен тем, что для этого случая возникающие задачи принимают наиболее чистый (рафинированный) вид, что облегчает их решение. Кроме того, этот случай является основным при создании компьютеров, ориентированных на численные методы и обработку изображений.

Во второй главе рассмотрены структурные принципы построения системы для анализа динамических образов, содержащих множество подвижных точечных объектов. Основными функциями такой системы являются: преобразование светового потока в двумерную матрицу сигналов; обнаружение объектов, определение их координат, направлений и скоростей движения относительно координатной системы датчика; отслеживание объектов в режиме обратной связи, и выдача полученных данных в удобном для человека виде. Особенностью предлагаемой системы является широкое использование параллелизма на всех этапах: восприятия, хранения и обработки информации.

Доказывается теорема на основе которой обосновывается, что если на фотоприемную матрицу добавить кольцевой регистр, построенный на основе ПЗС (как и сама матрица), то на самой матрице можно осуществить перестановку данных, что позволяет эти данные обрабатывать в параллельном режиме с использованием нескольких модулей памяти.

В третьей главе автор исследует вопросы, связанные с распараллеливанием метода частиц в ячейках (PIC-метода) на различных параллельных архитектурах. Рассматривается задача – модель плазмы без столкновения частиц одного типа (электронов), в которой электрическое поле описывается уравнением Пуассона. Исходный алгоритм отображается на три вида параллельных архитектур: система с коммутатором, гиперкуб и система, состоящая из центрального процессора и одного матричного, типа современных GPU.

Для всех архитектур при соответствующих допущениях выведены формулы коэффициентов ускорения в зависимости от начальных данных задачи: количества ячеек сетки, количества частиц, количества итераций, осуществляемых при решении уравнения Пуассона. Для отображения исходного алгоритма на гиперкуб используется код Грэя. В данной главе также исследуется параллельный алгоритм решения задачи о взаимодействии потоков разреженной плазмы.

Содержание первых трех глав позволяет сделать вывод, что создано новое научное направление, позволяющее описывать, обосновывать и анализировать: принципы организации компьютерной памяти, с параллельным доступом к сегментам, содержащимся внутри многомерных массивов; разрабатывать параллельные архитектуры вычислительных систем, в том числе, для отслеживания множества подвижных точечных объектов; предлагать методы отображения задач вычислительной математики и обработки изображений на различные наиболее важные архитектуры вычислительных систем.

В четвертой главе рассматриваются алгоритмы обработки сигналов, получаемых в процессе радиоактивного каротажа нефтяных скважин. Автором диссертации предложено: 1) автоматическое вычисление чистых спектров гамма-излучения неупругого рассеяния и гамма-излучения радиационного захвата. Разработан и опробован алгоритм, позволяющий в автоматическом режиме вычислять коэффициент вычета фона, что полностью исключает субъективизм при обработке и повышает качество результатов; 2) полуавтоматическая энергетическая привязка загруженных данных энергетических спектров. При этом учитывается нелинейность аппаратных характеристик от энергии регистрируемых гамма-квантов, а также отслеживается температурный дрейф энергетической шкалы прибора от времени; 3) полуавтоматическая обработка временных спектров; 4) новые методы калибровок при расчете коэффициента нефтенасыщенности: метод “Дельта C/O” и метод “Кросс-плот”; 5) новые методы кластеризации каротажных данных.

В пятой главе рассматриваются задачи анализа текстов на естественных языках. Разрабатываются алгоритмы, которые в результате анализа структуры текста, позволяют оценить релевантность текстов поисковому запросу. Такие оценки основываются на контексте поискового запроса, а не ограничиваются только ключевыми словами, их близостью или частотой. При работе алгоритмов предлагается использовать, получаемые на выходе программной системы Link Grammar Parser, семантико-синтаксические отношения между словами предложения.

В данной главе также рассматриваются логические методы определения смысловой близости предложений. Методика определения тем текстов, обобщает ранее выполненные исследования других ученых (Нирадж Кумар и др.). В том числе исследуется вариант метода, использующий размытую логику Заде.

В заключении работы приводятся основные результаты диссертационного исследования.