

УДК 616-073.75:004.9

## ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ПОИСКА ОБЛАСТЕЙ ИНТЕРЕСА В ТРЕХМЕРНЫХ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЯХ

Стромов Г.Г.<sup>1</sup>, Рыжков Д.В.<sup>1</sup>, Фокин В.А.<sup>2</sup><sup>1</sup> *Национальный исследовательский Томский политехнический университет, г. Томск*<sup>2</sup> *Сибирский государственный медицинский университет, г. Томск*

### РЕЗЮМЕ

Цель исследования – реализация программного комплекса для организации исследований по поиску областей интереса в трехмерных медицинских изображениях и реконструкции морфологического субстрата и его апробация на примере анализа большого объема модельных МРТ-снимков.

**Материал и методы.** Программный комплекс апробирован на модельных МРТ-снимках головного мозга, содержащих морфологический субстрат (проявление патологии – рассеянного склероза в тяжелой стадии), предоставляемых ресурсом BrainWeb. Технологический стек, выбранный для реализации предложенной схемы, базируется на кроссплатформенных решениях: для организации долговременного хранилища данных и поддержания информации в согласованном состоянии использована система управления базами данных MariaDB (open-source ветвь MySQL) и процедурное расширение SQL. Для автоматизации рутинных инфраструктурных задач использован Python версии 2.7. Расчетный модуль написан на языке программирования Java 7 с использованием библиотеки классов Spring Framework 3 и MongoDB как средства обмена данными между узлами в кластере. Версионирование кодовой базы основано на git, в качестве сборщика использован Maven 3.

**Результаты.** Полученные при тестировании программы в инфраструктуре кафедры медицинской и биологической кибернетики СибГМУ результаты исследований доказывают принципиальную возможность применения технологии автоматизированного поиска областей интереса и реконструкции морфологического субстрата в трехмерных изображениях МРТ на основе обобщенного анализа различий между референтной и оцениваемой группой снимков.

**Заключение.** Автоматизация процесса диагностики состояния организма на основе анализа медицинских изображений позволяет, с одной стороны, нивелировать в процессе принятия решения негативные моменты, связанные с субъективным компонентом, а с другой – повысить доступность высокотехнологичной медицинской помощи современными телекоммуникационными средствами. Представленное программное обеспечение комплексно решает задачу организации исследований по поиску областей интереса и реконструкции морфологического субстрата в модельных трехмерных медицинских изображениях в автоматизированном режиме с возможностью одновременной обработки нескольких исследуемых групп объектов.

**КЛЮЧЕВЫЕ СЛОВА:** МРТ, рассеянный склероз, область интереса, BrainWeb, java, nosql.

### Введение

Анализ трехмерных медицинских изображений является важным фактором повышения эффективности диагностики и планирования лечебных вмешательств для реализации пациент-ориентированного подхода. Основным методом, позволяющим выделять

области интереса в применяемых методах компьютерной реконструкции медицинских изображений, в настоящее время являются визуальные экспертные оценки. Это существенно снижает оперативность принятия решений и в ряде случаев вносит субъективность в их качество, особенно в условиях экстренных ситуаций, удаленности пациентов от высокотехнологических центров оказания специализированной медицинской помощи и т.п. То есть в тех

✉ *Рыжков Дмитрий Владимирович*, тел. 8-923-412-0762; e-mail: ryzhkoff.d.v@gmail.com

случаях, когда привлечение знаний и опыта высококвалифицированных медицинских специалистов бывает затруднено. В то же время современные компьютерные информационные технологии, такие как телемедицина, высокопроизводительные кластерные системы, параллельная обработка больших массивов данных, являются основой для разработки и реализации методов формализованного выделения областей интереса на медицинских изображениях.

В настоящей статье описывается реализация программного комплекса для организации исследований по поиску областей интереса в трехмерных медицинских изображениях и реконструкции морфологического субстрата и его апробация на примере анализа большого объема модельных МРТ-снимков.

## Материал и методы

Ранее нами была продемонстрирована принципиальная возможность применения методики интегральной оценки состояния биосистем [1, 2] в задачах анализа трехмерных медицинских изображений на примере выделения областей интереса на модельных МРТ-снимках головного мозга, содержащих морфологический субстрат (проявление патологии – рассеянного склероза в тяжелой стадии). В качестве источника данных используется ресурс BrainWeb, разработанный в McConnell Brain Imaging Centre (Montreal Neurological Institute, Канада) [3]. Для расчетов формируются две группы объектов, референтная и оцениваемая, состоящие из объемных снимков на основе нормального и патологического фантомов соответственно. Изображения из каждой группы разбиваются на множество мелких областей равного размера, разворачиваемых регулярным образом в вектор значений, представляющих величины яркостей пикселей, пронормированных на интервале  $0 \div 1$ ; оценка различий между группами проводится для блоков с одинаковыми координатами. Установлено, что в гистограммах распределения интегральных оценок для фрагментов изображений, содержащих морфологические изменения, наблюдается смещение в сторону больших значений по сравнению с интегральными оценками для изображений без морфологических изменений. Задавая определенное пороговое значение для совокупности интегральных оценок, можно выделить фрагменты изображения, содержащие проявление патологии, что является основой для реконструкции морфологического субстрата.

Описанный алгоритм реализован в программном комплексе, спроектированном по принципу модульной архитектуры и включающем несколько базовых

подсистем. Блок долговременного хранения данных обеспечивает содержание исходных материалов, справочных данных и результатов исследований, а также промежуточной информации, необходимой системе на этапе выполнения расчетов и реконструкции субстратов, в согласованном состоянии. Справочные данные представляют собой так называемые словарные сущности, содержащие перечни возможных параметров генерации изображений (тип фантома, на основе которого производились файлы, уровень наложенного шума, режим сканирования, толщина срезов и т.д.), а также их форматы и разрешения, которые являются критериями поиска и формирования сравниваемых групп объектов. Отобранный по ряду признаков объект включается в одну или несколько групп исследования. На уровне подсистемы управления исследованиями две группы объектов (например, совпадающие по уровню шума, режиму сканирования, но отличающиеся фантомом) сопоставляются между собой как референтная и оцениваемая по методике интегральной оценки с заданными параметрами статистического моделирования и восстановления морфологического субстрата на основе полученного распределения оценок.

Модуль загрузки файлов изображений отделен от подсистемы хранения, что обеспечивает необходимую гибкость в управлении данными, позволяя осуществлять загрузку файлов как в ручном режиме по востребованию с файловой системы в случае работы с единичными изображениями, так и в автоматическом режиме с использованием настраиваемого планировщика операционной системы при больших объемах исследуемых групп. В последнем случае предусмотрена передача данных через сеть с удаленной рабочей станции, производящей модельные изображения с заданными параметрами. Данная операция требует значительных аппаратных ресурсов, поэтому подготовка большой группы изображений занимает существенное время. Использование планировщика в этой ситуации позволяет оптимальным образом задействовать вычислительные мощности и минимизировать время простоя системы.

Расчетный модуль принимает на вход сведения о конкретном исследовании и обрабатывает группы объектов в рабочем потоке, состоящем из трех последовательных этапов.

1. Предобработка: разбиение снимков из сравниваемых групп на области заданных размеров (как правило, плоские прямоугольные или кубические блоки) и разворачивание их в векторы нормализованных значений яркостей пикселей с записью в подсистему долговременного хранения данных с привязкой к координатам.

натам блока в пространстве изображения и системного идентификатора обрабатываемого файла.

2. Расчет интегральных оценок: подготовленные на первом этапе векторы значений объединяются в так называемые матрицы состояний, строки которых соответствуют объектам сравниваемых групп, а столбцы – элементам вектора. В процессе расчета интегральных оценок различий между референтным и оцениваемым состояниями происходит статистическое моделирование референтной выборки [2]. Полученное среднее значение и другие параметры распределения интегральных оценок для каждого блока записываются в блок долговременного хранения данных.

3. Реконструкция субстрата: происходит анализ распределения полученных интегральных оценок, выбор пороговых значений на основе предустановленного для конкретного исследования перцентиля и реконструкция морфологического субстрата на их основе.

Системой обеспечивается автоматическое переключение между стадиями рабочего потока, приостановка расчета и его возобновление с точки останова, а также безопасное завершение процесса обработки в случае сбоя, не затрагивающее другие расчеты.

Реконструкция субстрата происходит в нескольких режимах. В режиме термокарты весь диапазон интегральных оценок линейно квантуется соответственно разрядности исходных изображений (например, для 8-битовых изображений диапазон будет разбит на 256 интервалов). Фрагментам в реконструированном изображении, соответствующим координатам блоков, на которые разбивались объекты из анализируемых групп, присваивается яркость, соответствующая интервалу, в который попадает проквантованная величина интегральной оценки для блока. В режиме битовой карты блоки, содержащие оценки выше заданного порогового значения, закрашиваются белым цветом, блоки с оценкой ниже порога – черным. В совмещенном режиме блоки с оценкой выше порогового значения закрашиваются пропорционально величине интегральной оценки, как в случае с режимом термокарты, а блоки с оценкой ниже порогового значения – черным, как в случае с битовой картой.

Технологический стек, выбранный для реализации предложенной схемы, базируется на кроссплатформенных решениях. Блок долговременного хранения данных реализован с использованием системы управления базами данных (СУБД) MariaDB (open-source ветвь MySQL) и процедурного расширения SQL. Блок загрузки данных и отгрузки результатов по электронной почте написан на языке Python версии 2.7. Расчетный модуль написан на языке программирования Java 7 с использованием библиотеки классов Spring

Framework 3, предоставляющей механизмы организации кода отдельных задач в единый рабочий поток методом декларативного описания с помощью языка XML. Кроме того, технология Java предоставляет возможности гибкого конфигурирования приложения по расходу вычислительных ресурсов, что обеспечивает предсказуемое поведение программы на различных аппаратных платформах и операционных системах. Для сборки и развертывания приложения на рабочей станции задействован сборщик Maven 3 с подключаемыми расширениями; управление кодовой базой осуществляется через распределенную систему контроля версий git и закрытый репозиторий на ресурсе GitHub. Вычислительное ядро программы покрыто серией модульных тестов на основе библиотеки TestNG 6.

Таким образом, реализация формальной модели программного комплекса позволяет проводить исследования на любой поддерживаемой платформой Java десктопной или серверной системе без поправок в кодовой базе.

Процедура расчета интегральной оценки – ресурсозатратная операция. В зависимости от размера блоков, на которые разбиваются анализируемые изображения, может проводиться от нескольких десятков тысяч до нескольких миллионов расчетных операций, обрабатывающих суммарно от нескольких до нескольких десятков ГБ данных. Задача обсчета одного исследования может быть решена за приемлемое время на единичной рабочей станции. Так, расчет различий между референтной группой, состоящей из 50 изображений нормы, и одним изображением с патологией по всему объему на максимальной детализации при количестве расчетных циклов около 1 млн и объеме обрабатываемых данных в 17 ГБ занимает около 48 ч на 12 вычислительных потоках Intel Xeon E5645. Естественно, что организация нескольких серий исследований требует принципиально иного способа построения вычислительного цикла.

Существует несколько подходов к оптимизации такого рода задач. Одним из популярных решений является перенос вычислений на графическую карту (GPU) с использованием технологии CUDA или OpenCL, что в ряде случаев может существенно (на несколько порядков) повысить скорость расчетов [4–11]. Проблемы использования данных технологий связаны с необходимостью разработки принципиально отличающейся от традиционных приемов программирования модели данных и методов ее обработки, а также написания дополнительного кода для прямой и обратной трансляции данных на прикладной уровень, что может оказаться нетривиальной задачей. Учитыва-

вая, что современные графические карты имеют объем оперативной памяти в несколько гигабайт, а объем обрабатываемой информации может быть больше в разы или на порядки, потребуется серия последовательных запусков расчетов на GPU с неизбежными при этом затратами на поддержание массивов обрабатываемых данных в согласованном состоянии. Таким образом, накладные расходы на инфраструктурные операции, жесткая привязка к аппаратным решениям и отказ от кроссплатформенности в контексте данной задачи интегральной оценки многомерных данных нивелируют достоинства рассматриваемой технологии.

Выходом может стать горизонтальное масштабирование существующего приложения на несколько вычислительных узлов с равномерным распределением нагрузки на задействованные процессоры. Реализация данного подхода требует организации очереди заданий и общего для всех узлов хранилища промежуточных данных (как сформированных матриц для расчета, так и результатов вычислений). Для медико-биологических исследований разработано множество решений для организации совместного доступа к данным в распределенной системе, часть из которых основана на NoSQL-базах данных – системах с высокой отказоустойчивостью и хорошей горизонтальной масштабируемостью, но обеспечивающих при этом слабую согласованность данных (weak consistency) (например, [12–17]). Такой класс систем хорошо подходит для организации так называемого вычислительного кэша – временного хранилища для расчетов, в котором операции записи и чтения совершаются единожды для одного набора данных. Отсутствие связей между хранящимися записями обеспечивает быстроту операций с ними. В текущей реализации программного комплекса нами была выбрана документо-ориентированная база данных MongoDB, хранящая записи в виде коллекций JSON-подобных структур, позволяя, с одной стороны, производить операции чтения-записи быстро, что свойственно нереляционным базам данных, но сохраняя при этом минимально-структурированное представление данных (что определяется самим форматом хранения данных JSON) – с другой, являясь своего рода компромиссом между реляционными СУБД и базами, хранящими данные в виде пар ключ-значение. В качестве системы для организации очереди выбрано приложение RabbitMQ – популярное промышленное решение, простое в настройке и эксплуатации. Принципиальная схема работы распределенной вычислительной системы показана на схеме (рис. 1).

## Результаты

Тестирование программы проводилось на кафедре медицинской и биологической кибернетики СибГМУ с использованием связанных в локальную сеть рабочих

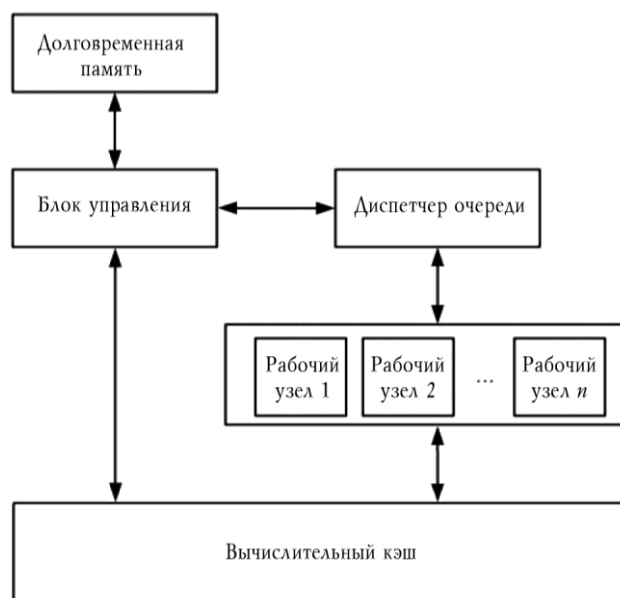


Рис. 1. Принципиальная схема устройства распределенной системы вычисления интегральных оценок различий между многомерными массивами данных с использованием NoSQL-базы в качестве временного хранилища и реляционной СУБД как долговременного хранилища

станций с типовой конфигурацией (двухъядерные процессоры Intel iCore3 с 2 ГБ ОЗУ) и сервера (s1366 Intel Xeon E5645, 24 ГБ ОЗУ). Было сформировано несколько групп сравнения с уровнем наложенного шума в 3; 5; 7; 9; 10; 14 и 15% соответственно. В ходе серии расчетов было обработано в общей сложности 10,8 тыс. модельных изображений общим объемом 76 ГБ, при этом объем промежуточных данных, порождаемых внутри системы, составил около 1 ТБ.

На кластере рабочих станций выполнялись поисковые расчеты в выбранных срезах с различными шагами партиционирования снимков и уровня помех, на сервере проводился анализ по всему объему изображений с установленными с помощью кластера оптимальными (сводящими к минимуму уровень случайно классифицированных сегментов) для каждого уровня шума параметрами реконструкции субстрата. Прирост скорости в сети из  $n$  рабочих станций в пределе составляет  $n - 1$ , однако, в зависимости от аппаратного оснащения, конфигурации локальной сети и реализации TCP/IP-стека в конкретной операционной системе может варьировать. Для 20 вычислительных процессов получено 11-кратное ускорение.

При невысоком уровне зашумленности методикой отлично детектируются участки поражения нервной ткани, при этом доля случайно определенных фрагментов незначительна (рис. 2). Как видно, полученная термокарта с шагом партиционирования исходных снимков  $2 \times 2 \times 2$  содержит информацию о структурной неоднородности в области локализации патологии, а ее трансформация в битовую карту по рассчитанному в ходе анализа распределения интегральных оценок порогу четко отображает область интереса как группу компактно локализованных сегментов результирующего изображения.

Увеличение размера блоков, на которые партиционируются снимки, имеет несколько значений. С одной стороны, это позволяет учесть различия по большому объему ткани и уменьшить влияние случайных помех на результат распознавания при относительно невысоком уровне шума, исключив определенное количество ошибочно классифицированных фрагментов изображения, пропорциональное размеру блока. С другой стороны, при достаточно высоком уровне шума можно добиться выявления области интереса, пропадающей на малых размерах блоков. В любом случае неизбежно падает уровень детализации получаемой картины различий в снимках.

Наши расчеты показали, что в зависимости от уровня помех есть несколько зон, или «окон», на которые можно условно разделить чувствительность распознавания различий в снимках:

- зона высокой чувствительности: при уровне до 9% субстрат определяется при любом способе разбиения изображений и во всех срезах;

- зона условной восприимчивости: при уровне шума от 10 до 14% области интереса определяются, но не на всех размерностях блоков, и (или) могут выпадать на соседних срезах;

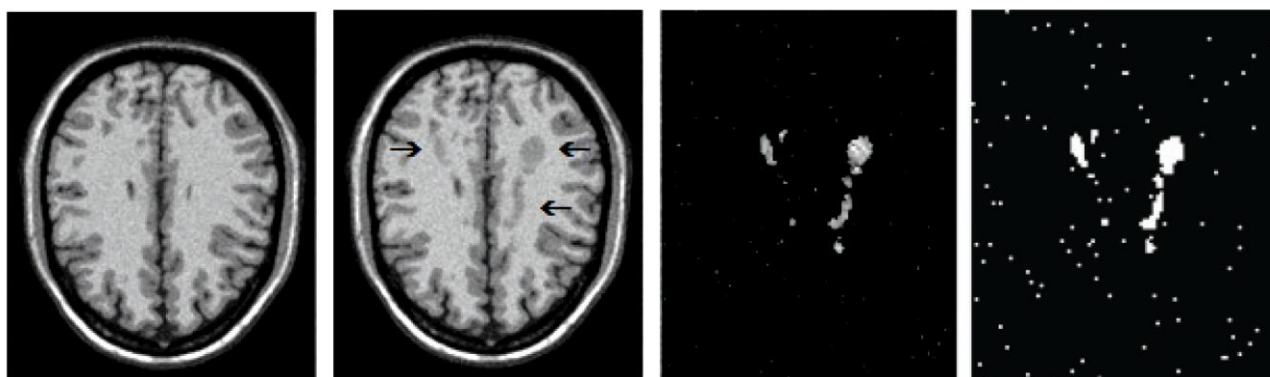
- «слепая» зона: при 15% и выше – различимость к патологии на фоне помех исчезает.

Оптимальность выбора размеров партиционирования снимков определяется исходя из решаемой за-

дачи. Если важно получить подробную информацию о свойствах изучаемой области, лучше выбрать малые размеры блоков. В этом случае гетерогенность изучаемого объекта, если таковая присутствует, обязательно проявится на термокарте в виде прилежащих сегментов с высокой, но различной яркостью.

Несколько увеличив размер блока, можно снизить количество артефактов, обусловленных помехами, в конечном изображении, сохранив при этом достаточную информативность. Высокая зашумленность анализируемых изображений (которую можно трактовать еще и как незначительно отличающуюся от окружающей среды анализируемую область) не позволяет получить детализированную информацию об объекте, и разбиение снимка на крупные фрагменты в такой ситуации делает возможным обнаружить наличие патологии (рис. 3).

Таким образом, общая тенденция, которая наблюдается во всех сериях расчетов: при увеличении размера блока повышается чувствительность методики, но информация о структурных особенностях области интереса снижается. Техника реконструкции субстрата применима только для изображений, попадающих в зону высокой чувствительности, поскольку только в этом случае можно получить непрерывное пространственное описание найденного объекта, и заключается в послойной сшивке битовых карт в единое трехмерное изображение. Несмотря на то, что при этом в результирующем изображении могут остаться артефакты, их распределение носит разреженный характер и они могут быть погашены применением фильтров. В условиях сохранения информации о связанности областей в нескольких слоях анализируемого изображения возможна реконструкция морфологического субстрата. На рис. 4 приведен пример изоповерхностного рендеринга реконструированного морфологического субстрата для уровня помех в 3% и размера блока  $2 \times 2 \times 2$ .



*a**б**в**г*

Рис. 2. Исходные изображения с уровнем шума в 3% на основе нормального (*a*) и патологического (*б*) фантомов и восстановленный морфологический субстрат (указан стрелками): термокарта распределения интегральных оценок (*в*) и фрагменты изображения с оценками выше 99-го перцентиля (*г*)

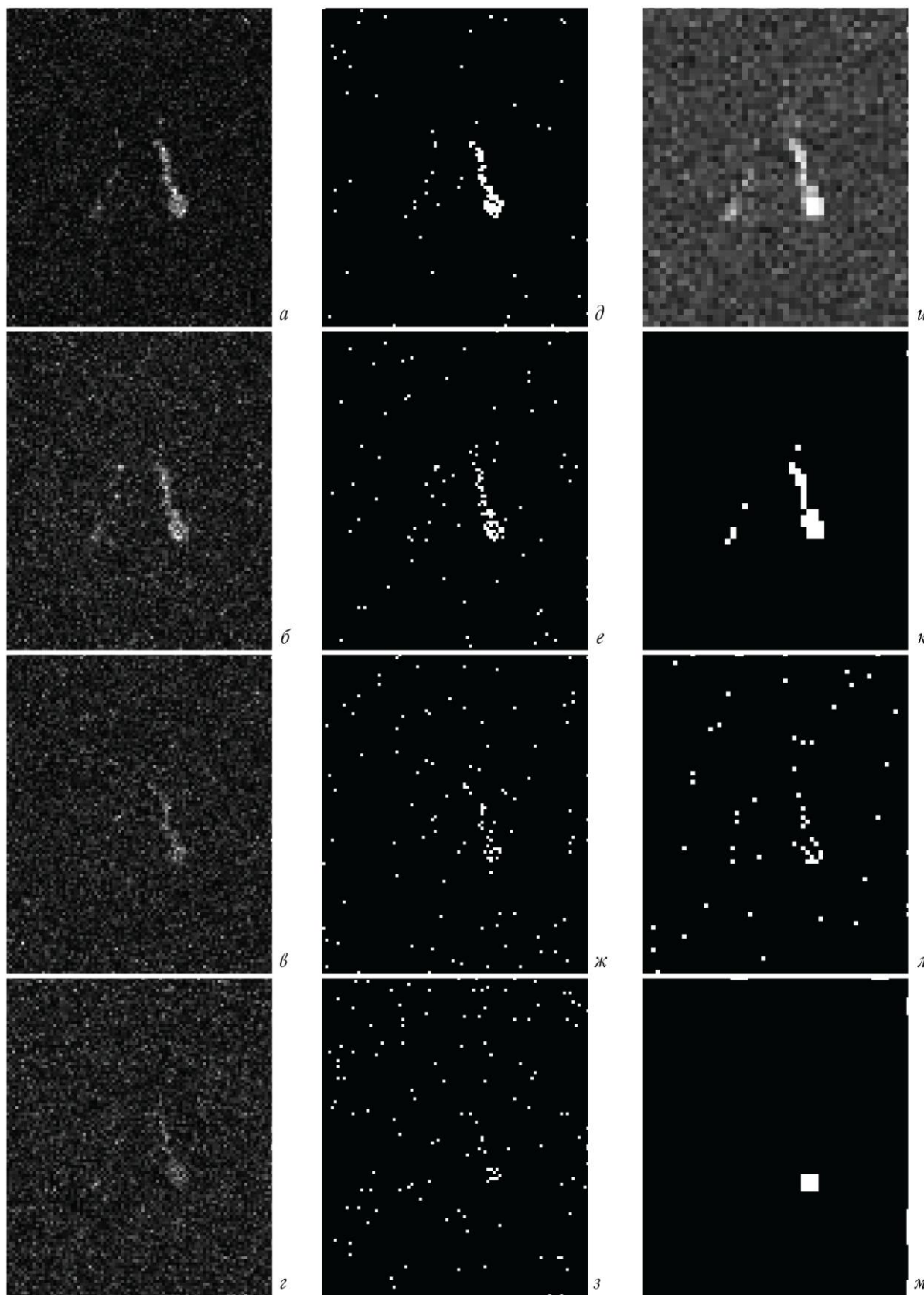


Рис. 3. Примеры термокарт и битовых карт при различном уровне шума и шагов партиционирования исходных изображений на одном срезе. Блок размерностью  $2 \times 2 \times 2$  для 5% (а – термокарта, д – битовая карта), для 7% (б и е соответственно), 9% (в и ж), 10% (г и з); и и к – термокарта и битовая расчеты для блоков  $4 \times 4 \times 3$  при уровне в 5%; л – битовая карта для блоков  $3 \times 3 \times 3$  при 10%-м шуме; м – для блоков  $12 \times 12 \times 1$  и шуме в 10%

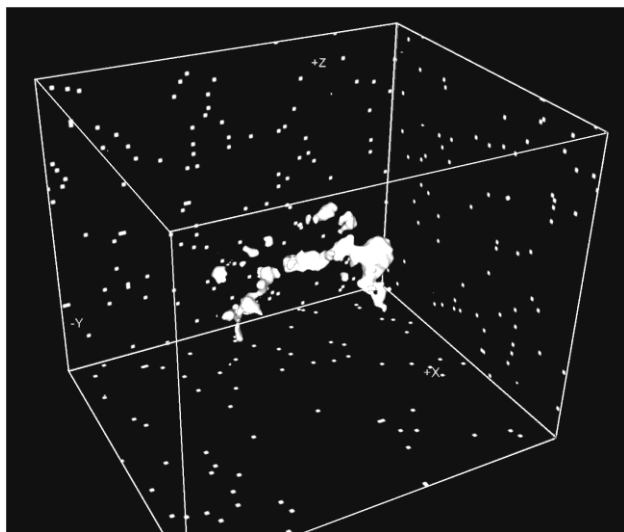


Рис. 4. Воксельное изображение реконструированного морфологического субстрата для шага партиционирования  $2 \times 2 \times 2$  и уровня шума в 3%

## Обсуждение

Как отмечается X. Llado и соавт. [18], при разработке алгоритмов поиска областей интереса возможны два подхода в тестировании: апробация на синтетических данных и на реальных. Использование изображений на основе фантомов дает хорошую основу для количественной оценки алгоритмов [19]. Однако синтетические данные не воспроизводят весь комплекс факторов, связанных с реальными данными, и алгоритмы, работающие в таких средах, при тестировании на реальных данных могут потерпеть неудачу. В данной работе тестирование алгоритма интегральной оценки проводилось с использованием модельных изображений, предоставляемых платформой BrainWeb. Эти наборы данных позволили уточнить граничные условия применимости методики интегральной оценки по шуму, а также выявить закономерности между шагом партиционирования сравниваемых изображений и качеством получаемого реконструированного изображения.

Адаптация методики под обработку реальных данных потребует внесения ряда изменений в процесс предварительной подготовки материала и организации групп сравнения. Согласно данным K.J. Udupa и соавт. [20], человек лучше автоматизированных алгоритмов в плане распознавания и определение тканей головного мозга (белое вещество, серое вещество, ликвор) нужно поручить оператору, а определение границ областей поражения – автоматизированной методике. Возможны также подходы на основе комбинаций группы алгоритмов, к примеру, классификации с учителем на основе РВТ-деревьев [21]. Наиболее вероятным

для используемой методологии является переход к интерактивному взаимодействию, когда экспертом на анализируемом изображении отмечается область сравнения и ставится в соответствие набор референтных изображений, представляющих анализируемую область при отсутствии поражения.

## Заключение

Автоматизация процесса диагностики состояния организма на основе анализа медицинских изображений позволяет, с одной стороны, нивелировать в процессе принятия решения негативные моменты, связанные с субъективным компонентом, а с другой – повысить доступность высокотехнологичной медицинской помощи современными телекоммуникационными средствами. Представленное программное обеспечение комплексно решает задачу организации исследований по поиску областей интереса и реконструкции морфологического субстрата в модельных трехмерных медицинских изображениях в автоматизированном режиме с возможностью одновременной обработки нескольких исследуемых групп объектов. Программный комплекс реализован на решениях с открытым исходным кодом и может быть запущен на любой популярной аппаратной платформе и операционной системе. В случае выполнения ресурсозатратных исследований вычислительная нагрузка может быть равномерно распределена по нескольким рабочим станциям гетерогенной локальной сети. Полученные с помощью комплекса результаты исследований доказывают принципиальную возможность применения технологии автоматизированного поиска областей интереса и реконструкции морфологического субстрата в трехмерных изображениях магнитно-резонансной томографии на основе обобщенного анализа различий между референтной и оцениваемой группой снимков.

*Исследовательская группа выражает особую благодарность лектору и старшему инженеру-программисту Университета Макгилла (Монреаль, Канада) Роберту Винсенту (Robert Vincent) за предоставленные ссылки в репозитории с программным обеспечением для генерации модельных изображений с фантомов и оперативные консультации по отладке и устранению проблем в работе с ним в инфраструктуре кластера СибГМУ.*

## Литература

1. Стромов Г.Г., Фокин В.А., Евтушенко Г.С. Интегральная оценка трехмерных биомедицинских изображений с использованием технологии распределенных вычислений // Биотехносфера, 2012. № 3–4. С. 68–72.



2. Фокин В.А. Технология интегральной оценки состояния биомедицинских систем // Системы управления и информационные технологии. 2008. № 1.1 (31). С. 191–194.
3. URL: <http://brainweb.bic.mni.mcgill.ca/brainweb/> (дата обращения: 07.01.2014).
4. Shi and Zhang. Fast network centrality analysis using GPUs // BMC Bioinformatics. 2011. 12. 149.
5. Blazewicz et al. Protein alignment algorithms with an efficient backtracking routine on multiple GPUs // BMC Bioinformatics. 2011. 12. 181.
6. Scharfe et al. Fast multi-core based multimodal registration of 2D cross-sections and 3D datasets // BMC Bioinformatics. 2010. 11. 20.
7. Shi et al. Parallel mutual information estimation for inferring gene regulatory networks on GPUs // BMC Research Notes. 2011. 4. 189.
8. Zhang et al. Developing a multiscale, multi-resolution agent-based brain tumor model by graphics processing units // Theoretical Biology and Medical Modelling. 2011. 8. 46.
9. Wan et al. High-performance blob-based iterative three-dimensional reconstruction in electron tomography using multi-GPUs // BMC Bioinformatics. 2012. 13 (Suppl 10). S4.
10. Bisset et al. High-performance biocomputing for simulating the spread of contagion over large contact networks // BMC Genomics. 2012. 13 (Suppl 2). S3.
11. Su et al. Parallel-META: efficient metagenomic data analysis based on high-performance computation // BMC Systems Biology. 2012. 6 (Suppl 1). S16.
12. Lewis et al. Hydra: a scalable proteomic search engine which utilizes the Hadoop distributed computing framework // BMC Bioinformatics. 2012. 13. 324.
13. Taylor. An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics // BMC Bioinformatics. 2010. 11 (Suppl 12). S1.
14. Srinivasan A., Faruque T., Joshi S. Data and task parallelism in ILP using MapReduce // Machine Learning. 2012. 86. 141–168.
15. Griffith et al. Scenario driven data modelling: a method for integrating diverse sources of data and data streams // BMC Bioinformatics. 2011. 12 (Suppl 10). S17.
16. Saxena et al. Implementation of a secure genome sequence search platform on public cloudleveraging open source solutions // Journal of Cloud Computing: Advances, Systems and Applications. 2012. 1. 14.
17. Borozan et al. CaPSID: A bioinformatics platform for computational pathogen sequence identification in human genomes and transcriptomes // BMC Bioinformatics. 2012. 13. 206.
18. Llado X. et al. Automated detection of multiple sclerosis lesions in serial brain MRI // Neuroradiology. 2012. Vol. 54. P. 787–807.
19. Hakulinen U. et al. Repeatability and variation of region-of-interest methods using quantitative diffusion tensor MR imaging of the brain // BMC Medical Imaging. 2012. 13. 30.
20. Udupa K.J. et al. Multiple sclerosis lesion quantification using fuzzy-connectedness principles // Medical Imaging. 1997. V. 16, Iss. 5. P. 598–609.
21. Wells M. et al. Fully automated segmentation of multiple sclerosis lesions in multispectral MRI // Pattern Recognition and Image Analysis. 2008. V. 18, Iss. 2. P. 347–350.

Поступила в редакцию 11.04.2014 г.

Утверждена к печати 07.05.2014 г.

Стромов Глеб Геннадьевич – аспирант кафедры медицинской и промышленной электроники НИ ТПУ (г. Томск).

Рыжков Дмитрий Владимирович (✉) – аспирант кафедры медицинской и промышленной электроники НИ ТПУ (г. Томск).

Фокин Василий Александрович – д-р техн. наук, профессор кафедры медицинской и биологической кибернетики СибГМУ (г. Томск).

✉ Рыжков Дмитрий Владимирович, тел. 8-923-412-0762; e-mail: ryzhkoff.d.v@gmail.com

## SOFTWARE FOR REGIONS OF INTEREST RETRIEVAL ON MEDICAL 3D IMAGES

Stromov G.G.<sup>1</sup>, Ryzhkov D.V.<sup>1</sup>, Fokin V.A.<sup>2</sup>

<sup>1</sup> National Research Tomsk Polytechnic University, Tomsk, Russian Federation

<sup>2</sup> Siberian State Medical University, Tomsk, Russian Federation

### ABSTRACT

**Background.** Implementation of software for areas of interest retrieval in 3D medical images is described in this article. It has been tested against large volume of model MRIs.

**Material and methods.** We tested software against normal and pathological (severe multiple sclerosis) model MRIs from the BrainWeb resource. Technological stack is based on open-source cross-platform solutions. We implemented storage system on Maria DB (an open-sourced fork of MySQL) with P/SQL extensions. Python 2.7 scripting was used for automatization of extract-transform-load operations. The computational core is written on Java 7 with Spring framework 3. MongoDB was used as a cache in the cluster of workstations. Maven 3 was chosen as a dependency manager and build system, the project is

hosted at Github.

**Results.** As testing on SSMU's LAN has showed, software has been developed is quite efficiently retrieves ROIs are matching for the morphological substratum on pathological MRIs.

**Conclusion.** Automation of a diagnostic process using medical imaging allows to level down the subjective component in decision making and increase the availability of hi-tech medicine. Software has shown in the article is a complex solution for ROI retrieving and segmentation process on model medical images in full-automated mode.

*We would like to thank Robert Vincent for great help with consulting of usage the BrainWeb resource.*

**KEY WORDS:** MRI, severe multiple sclerosis, interest retrieval, BrainWeb, java, nosql.

*Bulletin of Siberian Medicine, 2014, vol. 13, no. 4, pp. 99–107*

### References

1. Stromov G.G., Fokin V.A., Yevtushenko G.S. Integral'naja ocenka trehmernyh biomedicinskih izobrazhenij s ispol'zovaniem tehnologii raspredelennyh vychislenij [The integral estimation of the three-dimensional biomedical images using distributed computing technologies]. *Biotehnosfera – Biotechnosphere*, 2012, no. 3–4, pp. 68–72.
2. Fokin V.A. Tehnologija integral'noj ocenki sostojanija biomedicinskih sistem [Technology of integrated assessment of biomedical systems]. *Sistemy upravleniya i informatsionnye tekhnologii – Management Systems and information technologies*, 2008, no. 1.1 (31), pp. 191–194.
3. URL: <http://brainweb.bic.mni.mcgill.ca/brainweb/> (accessed 07.01.2014).
4. Shi and Zhang. Fast network centrality analysis using GPUs. *BMC Bioinformatics*, 2011, 12, 149.
5. Blazewicz et al. Protein alignment algorithms with an efficient backtracking routine on multiple GPUs. *BMC Bioinformatics*, 2011, 12, 181.
6. Scharfe et al. Fast multi-core based multimodal registration of 2D cross-sections and 3D datasets. *BMC Bioinformatics*, 2010, 11, 20.
7. Shi et al. Parallel mutual information estimation for inferring gene regulatory networks on GPUs. *BMC Research Notes*, 2011, 4, 189.
8. Zhang et al. Developing a multiscale, multi-resolution agent-based brain tumor model by graphics processing units. *Theoretical Biology and Medical Modelling*, 2011, 8, 46.
9. Wan et al. High-performance blob-based iterative three-dimensional reconstruction in electron tomography using multi-GPUs. *BMC Bioinformatics*, 2012, 13, (Suppl 10), S4.
10. Bisset et al. High-performance biocomputing for simulating the spread of contagion over large contact networks. *BMC Genomics*, 2012, 13 (Suppl 2), S3.
11. Su et al. Parallel-META: efficient metagenomic data analysis based on high-performance computation. *BMC Systems Biology*, 2012, 6 (Suppl 1), S16.
12. Lewis et al. Hydra: a scalable proteomic search engine which utilizes the Hadoop distributed computing framework. *BMC Bioinformatics*, 2012, 13, 324.
13. Taylor. An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. *BMC Bioinformatics*, 2010, 11 (Suppl 12), S1.
14. Srinivasan A., Faruque T., Joshi S. Data and task parallelism in ILP using MapReduce. *Machine Learning*, 2012, 86, 141–168.
15. Griffith et al. Scenario driven data modelling: a method for integrating diverse sources of data and data streams. *BMC Bioinformatics*, 2011, 12 (Suppl 10), S17.
16. Saxena et al. Implementation of a secure genome sequence search platform on public cloudleveraging open source solutions. *Journal of Cloud Computing: Advances, Systems and Applications*, 2012, 1, 14.
17. Borozan et al. CaPSID: A bioinformatics platform for computational pathogen sequence identification in human genomes and transcriptomes. *BMC Bioinformatics*, 2012, 13, 206.
18. Llado X. et al. Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology*, 2012, vol. 54, pp.787–807.
19. Hakulinen U. et al. Repeatability and variation of region-of-interest methods using quantitative diffusion tensor MR imaging of the brain. *BMC Medical Imaging*, 2012, 13, 30.
20. Udupa K.J. et al. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *Medical Imaging*, 1997, vol. 16, iss. 5, pp. 598–609.
21. Wells M. et al. Fully automated segmentation of multiple sclerosis lesions in multispectral MRI. *Pattern Recognition and Image Analysis*, 2008, vol. 18, iss. 2, pp 347–350.

**Stromov Gleb G.**, National Research Tomsk Polytechnic University, Tomsk, Russian Federation.

**Ryzhkov Dmitriy V.** (✉), National Research Tomsk Polytechnic University, Tomsk, Russian Federation.

**Fokin Vasilij A.**, Siberian State Medical University, Tomsk, Russian Federation.

✉ **Ryzhkov Dmitriy V.**, Ph. +7-923-412-0762; e-mail: ryzhkoff.d.v@gmail.com