

Анализ и визуализация “больших данных”

В.Л. Авербух, Д.В. Манаков

Институт математики и механики им. Н.Н.Красовского УрО РАН

Термин “большие данные” (big data) активно используется в литературе, но его единого понимания пока не сформировалась. Цель данной статьи попытаться структурировать разрозненную информацию о больших данных. Для нас представляют интерес данные, полученные в результате вычислений, их анализ и визуализация. Кроме различия в источниках существуют и различные методики обработки больших данных. Среди них сжатие, кластеризация, слайсинг, фильтрация. В нашем понимании большие данные - это предельный случай, когда универсальные подходы к анализу данных не работают. Поэтому многомерные данные, данные большого объема, данные с неполной информацией (модель с неопределенностью) являются частными случаями больших данных. Для визуализации больших объемов данных необходимо использование таких аппаратных и программных средств компьютерной графики, как “большие экраны” и среды виртуальной реальности. Требуется анализ и моделирование поведения пользователей, оценка и измерение качества, визуальных и интерактивных систем. Поэтому важны теоретические основания визуализации или формализация, а также верификация и валидация визуализации.

1. Введение

Визуализация «больших данных» (“Big Data Visualization”, “Large Data Visualization”) относится к областям как научной, так и информационной визуализации. В первом случае «большие данные» возникают в результате сложного компьютерного моделирования различных объектов и процессов. Во втором - имеет место визуальное описание и представление абстрактной информации, получаемой в результате процесса сбора и обработки многокатегориальных данных, для анализа которых необходимо применение нескольких количественных и качественных мер оценки.

Работы, описывающие результаты по визуализации больших данных, появились сравнительно недавно. Среди них можно выделить «Белую книгу» компании Intel, посвященную визуализации результатов «большого счета» [1], “установочную” публикацию известного специалиста в области компьютерной визуализации и человеко-компьютерного взаимодействия Б. Шнейдермана [2] и введение к спецвыпуску, посвященному визуализации больших данных [3]. Отметим, что один из соавторов работы [3], профессор Калифорнийского Университета в Дэвисе К.-Л. Ма (Kwan-Liu Ma), в настоящее время является ведущим специалистом по визуализации больших данных. Он является соавтором большого числа научных работ, а также организатором конференций, семинаров и лекций по данному направлению.

Нас в этой работе интересуют, прежде всего, визуализация результатов параллельных и распределенных вычислений. В связи с «большими данными» уместно также вспомнить понятие «большого счета» (многочасового, многосуточного или даже многонедельного процесса научных вычислений). Современные вычислительные системы в ходе компьютерного моделирования способны генерировать очень большие (large or huge) по объему файлы с данными. В тоже время оценка размера файла, прежде всего, связана с возможностями «супервычислений», мощности которых непрерывно возрастают. Усложнение структуры данных связано не только с архитектурой вычислительного комплекса, но и с возможностью решения более сложных в математическом плане задач, например, задач, приводящих к многомерным решениям.

Можно увязать понятие «большие данные» с некоторым предельным (на данный момент) случаем обработки данных, при котором универсальные подходы к анализу и визуализации не работают. Тогда в качестве больших данных могут рассматриваться многомерные и многокатегориальные данные, данные большого объема, данные с неполной информацией. Предельный случай формирует вызовы, на которые необходимо ответить, чтобы двигаться дальше. Решение возникающих проблем приводит к тому, что сегодняшние «большие данные» завтра становятся нормой.

Среди задач визуализации “больших данных” рассматриваются следующие:

- визуализация потоков данных;
- визуальный интеллектуальный анализ данных (Visual data mining);
- визуальный поиск и рекомендации (Visual search and recommendation);
- описание ситуаций на основе больших данных с использованием визуализации (Big data storytelling using visualization);
- масштабируемые методы параллельной визуализации;
- современные аппаратные средства и архитектуры для анализа и визуализации данных;
- человеко-компьютерный интерфейс и визуализация больших данных;
- приложения визуализации больших данных, включая кибер разведку и контрразведку, бизнес-анализ (бизнес разведку), электронную коммерцию, анализ научной информации, образование и т.д.

В плане научной визуализации работы по визуализации больших данных затрагивают проблемы визуализации объемных данных, включая параллельный объемный рендеринг и визуализацию объемных данных без прерывания работы системы (In situ volume visualization).

В плане информационной визуализации рассматривается визуализация данных о вычислительных сетях и их структурах, а также визуализация изменяющихся по времени данных, включая визуальную кластеризацию и методики графического представления динамики социальных сетей.

Многие задачи визуализации программного обеспечения, возникающие, например, при рассмотрении трасс выполнения параллельных программ, также связаны с большими и очень большими объемами данных. Отметим, что методы визуализации в этом случае, как правило, заимствуются из методов, используемых в информационной визуализации.

Научные интересы нашего исследовательского коллектива связаны с исследованиями и разработками в области научной визуализации результатов параллельных и распределенных вычислений. Накоплен определенный опыт в области визуализации “больших данных”. Выделим в этой связи ряд задач “большого счета” и соответственно анализа и визуализации “больших данных”.

Обеспечение доступа к данным в том числе, решение задач хранения и передачи данных, создание средств удаленной и он-лайн визуализации;

Фильтрация и рендеринг;

Интерпретация визуализируемых данных;

Верификация и валидация визуализации.

В свою очередь по каждому пункту необходимо проведение целого цикла исследований и разработок. При этом нельзя получить решение, раз и навсегда удовлетворяющее пользователей и разработчиков.

Подходы к организации хранения и передачи данных зависят от архитектуры вычислительных комплексов и возможностей вычислительных сетей. Здесь многие проблемы решаются не только за счет увеличения мощности и скорости соответствующих вычислительных средств, но и за счет их стандартизации.

Удаленная визуализация позволяет расширить круг пользователей систем «большого счета». Средства он-лайн визуализации, во-первых, позволяют, вмешиваясь в работу программы в процессе счета, оперативно оценивать промежуточные результаты и принимать решения по изменению параметров вычисления, а во-вторых, на их базе возможна разработка систем визуальной отладки параллельных программ. Имеет место большое количество публикаций по данным вопросам. Разработки нашего исследовательского коллектива в этом направлении описаны, в частности, в работах [4], [5], [6]. В связи с огромными объемами данных необходимо распараллеливать алгоритмы фильтрации и рендеринга. Некоторые наши подходы к решению этих задач описаны в [7], [8], [9]. Отметим, что решаемые нами задачи сходны с теми задачами, которые рассматриваются в рамках научной визуализации больших данных. (Например, визуализация параллельных вычислений “in situ” в какой-то мере включает в себя удаленную и он-лайн визуализацию.)

В данной работе основное внимание уделяется проблематике анализа и интерпретации визуализируемых данных, полученных в результате научного счета.

2. Оценка видов отображения визуализации

В [1] в качестве одной из задач визуализации больших данных указывается на задачу превращения «больших данных» в «большой инсайт». (Insight – озарение, понимание. В работе [10] наличие инсайта рассматривается как критерий при оценке качества визуализации.) Задачу получения «большого инсайта», по мнению авторов, следует решать в рамках визуальной аналитики, которая является инструментом, прежде всего, информационной визуализации. Именно информационная визуализация является инструментом работы при анализе и интерпретации, «деловых» данных, различного происхождения. Следует отметить, что зачастую такие данные, несмотря на «естественное» происхождение (например, сбор и обработка личных данных или результаты анализа социальных сетей), не имеют естественной образности. Проектировщику приходится опираться на традиционные методики визуализации (например, статистические диаграммы различных типов) или придумывать какие-то новые виды отображения. Полученные виды отображения могут оцениваться с применением критерия качества информационной визуализации, основанного на весьма популярной схеме анализа визуальных данных, предложенной Б. Шнейдерманом – “*Overview first, zoom and filter, then details-on-demand*”. Эта схема предполагает анализ больших объемов визуализированных данных [11]. Критерий качества предусматривает проверку системы визуализации на возможность последовательного повторение действий по обзору, изменению масштаба, фильтрации, детализации, вывода зависимостей, получения истории поиска и извлечения подмножества из рассматриваемых данных [12].

Информационная визуализация больших данных традиционно связывается с методиками Data Mining, то есть с методиками обнаружения значимых корреляций, зависимостей и тенденций в результате анализа хранимой информации, выявления отношений между данными различного типа, таких как ассоциации, последовательности аналогии, кластеры и т.п. Применяются различные (в том числе и связанные с искусственным интеллектом) методы выделения и извлечения информации, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий.

В свою очередь научная визуализация связана с анализом научных данных, который включает такие операции, как идентификация, локализация, категоризация, кластеризация, ассоциация, валидация, корреляция [13]. Очевидно, что этот набор можно расширить. В различных приложениях операции из данного набора используются в различных сочетаниях.

В системах научной визуализации используется самые различные виды отображения. Как правило, для данной прикладной области существует свой набор традиционных, привычных методик представления данных. В целом ряде случаев необходимо придумать новые, специализированные методы, как например, в случае четырехмерных множеств сложной структуры, полученных в результате «большого счета» [14]. При выборе и/или поиске набора видов отображения, особенно в случае работы с новыми компьютерными моделями явлений и процессов, необходима оценка пригодности (адекватности в визуализации).

Аналогично случаю информационной визуализации оценка видов отображения научной визуализации может быть построена на проверке того, как в них поддерживаются операции научного анализа данных. Однако такая оценка является проверкой только лишь необходимой, но не достаточных условий качества визуализации.

Рассматривая виды отображения, используемые в научной визуализации, можно оценивать их естественность (точнее привычность для пользователей), устойчивость к масштабированию, возможности для представления сложных структур, а также объектов особого интереса, особых точек, аттракторов, сингулярностей и пр.

Можно оценивать информационную емкость вида отображения, то есть возможность по представлению того или иного объема данных. Например, емкость распределенного графического вывода (scatterplots - набор точек на плоскости; также переводится как диаграмма рассеяния). ограничена разрешением экрана. Реальная информационная емкость всегда ограничена способностью пользователя к различению отдельных объектов их формы, текстуры, цвета и пр. Масштабируемость вида отображения зависит от его потенциальной информационной емкости.

В число критериев оценки видов отображения, используемых при визуализации «больших данных» следует ввести скорость и надежность восприятия, уровень усилий при восприятии и интерпретации визуальной информации, представленной тем или иным способом. Также необходимо оценивать достоверность визуального представления.

В работе [15] рассматриваются четыре взаимозависимых меры оценки больших данных: *вариативность, скорость, объем, достоверность* (равная **1-[неопределенность]**). Говорится о модели '4 Vs' (variety, velocity, volume, veracity).

Возможность выбора различных подходов к выбору критериев оценки и стремление исключить при этом субъективность оценок диктует необходимость формализации предлагаемых подходов. Задачи верификации и валидации визуализации рассматриваются в литературе, в том числе и на регулярно проводимых семинарах по воспроизводимости, верификации и валидации визуализации (EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization). В [16] нами были рассмотрены некоторые подходы к постановке этих задач. Отметим, что не существует общепринятой модели формальной верификации визуализации, включая общепринятую терминологию. Это вызвано тем, что как подходы, так и терминология наследуются из разных математических дисциплин. Например, визуализацию часто рассматривают с позиции общей теории информации [17] или теории монотонной меры [18].

Естественно существование ряда подходов к обработке и визуальному представлению больших данных. В связи с этим возникает необходимость оценки их адекватности конкретным целям и задачам (то есть валидации визуализации), а также общего сравнительного анализа и относительной эффективности.

3. Обработка и визуализация больших данных в параллельных и распределенных вычислениях

Как уже отмечалось, задачи научной визуализации больших данных близки к задачам, решаемым при визуализации в параллельных и распределенных вычислениях. Разделение параллельных и распределенных вычислений часто основывается на том, что в первом случае характерен параллелизм по данным, а во втором - по задачам. Это разделение достаточно условно. Значительно важнее их общность с точки зрения применяемых технологий программирования. Сформировалось направление, связанное с визуализацией в параллельных вычислениях, названное визуальными супервычислениями. Была выработана терминология, разработан ряд модельных и прототипных систем [19]. На разработку новых технологий были направлены первоначальные усилия разработчиков программного обеспечения в области супервычислений, причем обязательными условиями являлись: поддержка интерактивного режима и возможность обработки больших данных, в том числе и их визуализация. Облачные вычисления рассматриваются, как правило, в рамках распределенных вычислений. Анализ результатов в этом случае проводится с применением веб-технологий ("тонкий" клиент).

В связи с параллельными и распределенными вычислениями рассмотрим ряд программных технологий, которые также применимы в задачах обработки, анализа и визуализации «больших данных».

Фильтрация данных

Фильтрация данных включается в стандартный графический конвейер, состоящий из фильтрации данных, их геометрической обработки и растеризации (рендеринга). Любая из этих трех частей может быть реализована как параллельно, так и последовательно. Однако для классификации остается важным то, где параллелизм заканчивается, а также выяснение визуализацию каких данных поддерживает клиент. В случае параллельной фильтрации данных, только фильтрация данных осуществляется параллельно, и, следовательно, на клиент передаются отфильтрованные математические данные. Параллельная фильтрация востребована для двух классов задач - визуализации данных большого объема и в многопараметрических задачах, требующих активного взаимодействия пользователя и системы в процессе визуального анализа с целью уточнения этих параметров или рефакторинга. В частности, в разработанном конструкторе визуализации [20] поддерживается применение фильтров и изменение параметров. Последняя возможность определяется, как императивный подход [21].

Приведем ряд определений, связанных с фильтрацией данных.

Фильтр – это любая операция над данными, изменяющая их количество. При этом исключаются случаи хеширования. Исключение хеширования рассматривается, как необходимое условие установление предпорядка: Формально можно говорить о построение идеала. Исходя из этого определения, сжатие изображения с потерей точности и рендеринг можно рассматривать, как частные случаи фильтрации данных.

Параллельная фильтрация данных – это интерактивный процесс над распределенными данными, где на каждом шаге применяется параллельный фильтр, происходит передача отфильтрованных данных и их визуализация, имеющая вложенный цикл стандартного (непосредственного) взаимодействия.

Фильтрация данных предполагает в результате получение необходимой (интересующей) информации за минимальное время. Возможна другая минимаксная постановка задачи при которой целью фильтрации данных является получение максимума полезной информации с минимальными затратами (например, затратами времени на вычисления, взаимодействия, интерпретации и т.п.). Данные большого объема определяются, как данные, которые невозможно отобразить как, в полном объеме, так и за приемлемое время. Следовательно, важным вопросом задачи обработки данных является вопрос о метриках оценки эффективности.

Слайсинг

Слайсинг – построение срезов. Частный случай фильтрации данных, когда функция от данных равна константе. Кроме сечения плоскостью – это такие стандартные виды отображения, как изолинии и изоповерхности. Условие – равенство константе фактически сокращает размерность данных на единицу. Изменение константы приводит к построению фазового пространства. В качестве константы может выступать идентификатор функции или максимальная длина графа. В качестве такого примера приведем работу [22], в которой параметры статического анализа (графа) уточняются после проведения динамического анализа. Данный подход можно определить, как робастная модель статического анализа с рефакторингом.

Брашинг

Общепринятый визуальный подход закрашивание (“brushing”) [23] интерактивно выделяет подмножества данных, чаще всего цветом. Он поддерживает визуальное связывание разнородных объектов и поэтому направлен на проблему визуальной фрагментации. В отличие от фильтрации, которая удаляет данные с дисплея, закрашивание дает добавочную информацию, накладывая уточненное изображение на существующую структуру. Однако закрашивание требует собственной кодирующей величины, то есть графический атрибут, используемый для выделения выступающих точек, должен быть сохранен, будь то цвет, форма или текстура. Эти графические атрибуты уже стандартно применяются в визуализации со многими параметрами (множественной визуализации). В частности, цвет удобен для категоризации. Таким образом, брашинг – это категоризация или выделение цветом. Он не приводит к сокращению объема данных, а только уточняет структуру данных.

Кластеризация

Кластеризация – это тематическая категоризация или классификация. Данные принадлежат тому кластеру, расстояние до которого минимальное. При кластеризации также возможно сжатие данных - в кластере оставляют элемент с наибольшей мощностью. Одно из клише кластеризации – раздели и властвуй. В связи с развитием интернет-технологий направление вектора формализации сместилось от методов линейного программирования в сторону нечетких множеств, с точки зрения программиста являющимися ассоциативными массивами. В данном случае фундаментом кластеризации является алгоритм Map-Reduce.

Модель потока данных

Эффективность в модели потока данных достигается за счет возможности конвейерной обработки данных (один из методов распараллеливания). Система программирования, основанная на бинарной коммуникативной модели, описана, например, в работе [24].

Алгоритмы с внешней памятью

Несомненно, для научной визуализации значительный интерес представляют внеядерные (“out-of-core”) подходы или алгоритмы с внешней памятью, минимизирующие накладные расходы дискового ввода-вывода [25]. Двумя наиболее часто используемыми подходами являются организация данных со многими разрешениями или применение мультиразрешения (используется для уровня детализации) и организация данных, зависящая от точки зрения. Для этих случаев характерно применение k-дерева при реструктуризации данных, в прочем, как и для объемного рендеринга [26]. Реструктуризация данных по k-дереву обеспечивает возможность быстрого поиска и извлечения данных.

Автономные вычисления

Считается, что решением возрастающей административной сложности вычислительных инфраструктур являются автономные вычисления. Автономные вычисления обращаются к

вычислительным системам, которые обладают способностью самопознания и самоуправления. Такая система может характеризоваться одним или несколькими из нижеследующих атрибутов:

Самоконфигурирование – система может объединять новые и существующие компоненты без вмешательства администратора;

Самооптимизация – система может непрерывно пытаться изменить конфигурацию, чтобы определить, является ли текущая оптимальной;

Самовосстановление (самооздоровление) – система может отслеживать ошибки и восстанавливаться после неправильных аппаратных или программных компонент;

Самозащита – система может отслеживать попытки взлома и реагировать соответствующим образом [19].

Существует достаточно широкий пласт работ по адаптивному интерфейсу, который можно рассматривать, как частный случай автономных вычислений, связанный с самооптимизацией под конкретного пользователя. Примеры из области параллельных вычислений – это планирование очереди задач и оптимизация коллективных обменов.

4. Заключение и выводы

После постановки проблемы «больших данных» и задач их визуализации по этим вопросам опубликовано большое количество работ. Есть понимание того, что формируется новая предметная область. Сформировалась структура новой дисциплины, хотя и требующая дальнейшей детализации. (См. рисунок 1, который представляет собой адаптацию Figure 1 [15].)

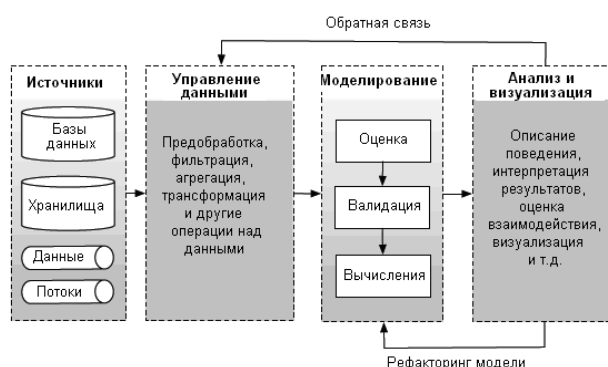


Рис. 1. Структура обработки “больших данных”.

Интерес к новому направлению проявляется и среди отечественных специалистов. Сравнительно недавно издательство “Открытые системы” провело специализированную конференцию по большим данным. (См. предварительную программу [27]).

Как мы уже говорили выше, о новом направлении можно говорить и в связи с тем, большие данные формируют новые вызовы, на которые необходимо ответить. Например, формальная верификация и валидация моделей и подходов, связанных с обработкой больших данных, получают первостепенное значение. В частности, это связано с тем, что методы высокого порядка точности при достаточно мелком разбиении начинают расходиться. Одно из объяснений этого факта может быть следующим: в предельном случае (при достаточно больших данных) вычислительная погрешность и погрешность метода становятся взаимозависимыми. Важно, чтобы предлагаемые подходы рассматривались, как инфраструктурные технологии.

Анализ и интерпретация «больших данных» требуют использования новых технологий компьютерной графики, сред виртуальной и расширенной реальности. Отсюда следует необходимость анализа влияния «человеческого фактора», проведения комплексных исследований не только с точки зрения компьютерных наук и математики, но и когнитивной психологии. Поэтому среди серьезных проблем нового направления можно указать на проблему “больших картинок” (*big pictures*), отображающих большие данные. Отметим, что работы в этом направлении появились еще до появления направления «больших данных» [28,29,30]. Основные трудности здесь связаны с восприятием и интерпретацией сверхбольших объемов графической

информации [31]. Интересные результаты в этом плане принадлежат Б. Тверски и ее коллегам (См., например, [32], [33].) Кстати, в области компьютерной графики сложилась ситуация, когда ее возрастающие возможности, (например, в области фотореалистической графики, виртуальных и расширенных сред) зачастую превышают традиционные потребности в визуализации.

“Большие данные” порождают предельные на данный момент случаи, что требует комплексного решения проблем компьютерной графики, визуализации, визуальной аналитики и когнитивной психологии. Анализ зарубежных работ по визуализации больших данных за последние годы показывает, что они включают в себя целый ряд направлений компьютерной визуализации, среди которых научная и информационная визуализация, визуализация программного обеспечения, визуальный анализ данных, верификация и валидация визуализации, изучение восприятия и когнитивной составляющей визуализации при использовании «больших экранов» и сред виртуальной реальности. Необходим учет этого опыта при разворачивании аналогичных исследований в нашей стране.

Литература

1. Big Data Visualization: Turning Big Data into Big Insights. The Rise of Visualization-based Data Discovery Tools. White Paper. Intel IT Center. March 2013.
2. <http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/big-data-visualization-turning-big-data-into-big-insights.pdf>
3. Shneiderman. The big picture for big data: Visualization. Science, 343:730, February 2014.
4. Keim D. Qu H., Ma K.-L. Big-Data Visualization // IEEE Computer Graphics and Applications. July/August 2013. Pp. 50-51.
5. Авербух В.Л., Байдалин А.Ю., Васев П.А., Исмагилов Д.Р., Зенков А.И., Манаков Д.В., Перевалов Д.С., Шагубаков М.Р. Задачи визуализации параллельных вычислений. // Вопросы атомной науки и техники. Сер. Математическое моделирование физических процессов. 2002. Вып. 3. Стр. 40-52.
6. Бахтерев М.О., Васёв П.А., Казанцев А.Ю., Манаков Д.В. Система удалённой визуализации для инженерных и суперкомпьютерных вычислений // Вестник ЮжУрГУ, N 17 (150), 2009, серия «Математическое моделирование и программирование», Выпуск 3. Стр. 4-11.
7. Васёв П.А. Среда поддержки интерактивной визуализации для суперкомпьютерных вычислений // Вопросы атомной науки и техники. Серия: Математическое моделирование физических процессов. 2009. Выпуск 4. Стр. 67-77.
8. Горбашевский Д.Ю., Казанцев А.Ю., Манаков Д.В. Параллельная фильтрация в системе визуализации параллельных вычислений // ГрафиКон'2006, 1-5 июля 2006. Труды Конференции. Новосибирск. 2006. С. 333-336.
9. Manakov D., Mukhachev A., Shinkevich A. Visualization of the Distributed Data of Huge Volume. Assembly, Filtration, Sorting // Proceedings of the 13-th International Conference on Computer Graphics and Vision Graphicon-2003 Moscow, September 5-10, 2003. Pp. 198-201.
10. Михайлов И.О. Параллельный рендеринг воксельной графики // Международная научная конференция Параллельные вычислительные технологии 2013. Челябинск. Сборник Трудов. Том 2. 2013. Стр. 608.
11. North Ch. Toward Measuring Visualization Insight // IEEE Computer Graphics and Applications May/June 2006, Volume: 26, Issue: 3, pp. 20-23.
12. Shneiderman B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations // Proceedings of the IEEE Conference on Visual Languages, September 3-6, 1996. Pp. 336-343.
13. Maletic J.I., Marcus A., Collard M.L. A task oriented view of software visualization // International Workshop on Visualizing Software for Understanding and Analysis. 2002. Pp. 32-40.

14. Baker, M. P., Wickens, C. D. Human Factors in Virtual Environments for the Visual Analysis of Scientific Data // Technical Report, NCSA. 1995.
15. Авербух В.Л., Байдалин А.Ю., Бахтерев М.О., Васёв П.А., Казанцев А.Ю., Манаков Д.В. Опыт разработки специализированных систем научной визуализации // Научная визуализация. Квартал 4. Том 2. Номер 4. 2010. Стр. 27-39.
16. Assuncao M.D., Calheiros R.N., Bianchi S., Netto M.A.S., Buyya R. Big Data Computing and Clouds: Challenges, Solutions, and Future Directions // arXiv:1312.4722v2, 22 Aug 2014. <http://arxiv.org/pdf/1312.4722.pdf>
17. Авербух В.Л., Анненкова О.Г., Бахтерев М.О., Манаков Д.В. Задачи визуализации программного обеспечения параллельных и распределенных вычислений // Международная научная конференция Параллельные вычислительные технологии 2014. Южный федеральный университет. Сборник трудов. Стр. 7–18.
18. Dasgupta A., Chen M., Kosara R. Conceptualizing Visual Uncertainty in Parallel Coordinates // Comput. Graph. Forum 31(3). 2012. Pp. 1015-1024.
19. Fout N., Ma K.-L. Reliable Visualization: Verification of Visualization based on Uncertainty Analysis/ Tech. rep., University of California, Davis, 2012.
20. Brodli K., Brooke J., Chen M., Chisnall D., Fewings A., Hughes C., M. W. Jones, Riding M., Roard N. Visual Supercomputing - Technologies, Applications and Challenges // Eurographics 2004, STAR Reports. Pp. 37-68.
21. Бахтерев М.О., Васёв П.А., Один метод визуализации результатов суперкомпьютерных вычислений // Материалы III Всероссийской научно-технической конференции “Суперкомпьютерные технологии” (СКТ-2014), Таганрог, НИИ МВС ЮФУ. 2014. С. 50-55.
22. Васёв П.А., Бахтерев М.О. Декларативно-императивный метод конструирования сцен научной визуализации // XV Международная конференция «Супервычисления и Математическое Моделирование». Тезисы. ФГУП «РФЯЦ ВНИИЭФ». Саров. 2014, стр. 46.
23. Зубов М.В., Пустыгин А.Н., Старцев Е.В. Получение типов данных в языках с динамической типизацией для статического анализа исходного кода с помощью универсального классового представления // Вестн. Астрахан. гос. техн. ун-та. Сер. управление, вычисл. техн. информ., 2013, N 2, Стр. 66–74.
24. Bartram L., Ware C. Filtering and brushing with motion // Information Visualization, V.1 N.1, March 2002. Pp.66-79,
25. Бахтерев М.О., Васёв П.А., Казанцев А.Ю., Альбрехт И.А. Методика распределенных вычислений RiDE // Параллельные вычислительные технологии (ПаВТ’2011): труды международной научной конференции. Челябинск: Издательский центр ЮУрГУ, 2011. С. 418–426.
26. Shih M., Zhang Y., Ma K.-L., Sitaraman J., Mavriplis D. Out-of-Core Visualization of Time-Varying Hybrid-Grid Volume Data // IEEE Symposium on Large Data Analysis and Visualization. 2014. Pp. 93 – 100.
27. Михайлов И.О., Авербух В.Л. Современные методы визуализации больших и сверхбольших объёмных данных // XV Международная конференция «Супервычисления и Математическое Моделирование». Тезисы. ФГУП «РФЯЦ ВНИИЭФ». Саров. 2014, стр. 97-98.
28. <http://www.ospcon.ru/node/107242>
29. Goodall J.R., Lutters W.G., Rheingans P., Komlodi, A. Preserving the big picture: visual network traffic analysis with TNV // Workshop on Visualization for Computer Security, 2005. (VizSEC 05). Pp. 47 – 54.
30. Sedlmair M., Ruhland K., Hennecke F., Butz A., Bioletti S., O'Sullivan C. Towards the Big Picture: Enriching 3D Models with Information Visualisation and Vice Versa // Smart Graphics. Lecture Notes in Computer Science. Volume 5531, 2009, pp 27-39.
31. Steeda Ch.A., Ricciuto D.M., Shipman G., Smith B., Thornton P.E., Wang D., Shi X., Williams D.N. Big data visual analytics for exploratory earth system simulation analysis // Computers & Geosciences 61 (2013) 71–82.
32. Shilpa, Manjit K. Big Data Visualization tool with Advancement of Challenges // International Journal of Advanced Research in Computer Science and Software Engineering 4(3), March - 2014, pp. 665-668.

33. Tversky, B. Spatial cognition: Embodied and situated // *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press. 2009. Pp. 201-217.
34. Kessel A., Tversky B. Visualizing space, time, and agents: production, performance, and preference // *Cognitive Processing*. 2011, 12(1). Pp. 43-52.