

Конференция «Большие Данные в национальной экономике»

Москва, 22 октября
ЦВК Экспоцентр



Проблема больших данных в развитии национальной медицины и здравоохранения

Турлапов Вадим Евгеньевич

vadim.turlapov@cs.vmk.unn.ru

д.т.н., проф. каф. МО ЭВМ, ННГУ им.Н.И.Лобачевского
Лаборатория компьютерной графики (www.graph-unn.ru)
Центр биомедицины и вычислительных технологий
(www.biomedcompcenter.ru)

Охрана здоровья как задача национальной экономики

1. По данным за период 2010-2012 уровень смертности (на 100 тысяч человек): в Израиле – 533.4 умерших; в США – 642; в Бельгии – 635; в Норвегии – 549; в Италии – 511.6; в России - 1430; в Украине - 1520.
2. Особенно высока в России смертность в трудоспособном возрасте. Вероятность умереть в возрасте 15–60 лет в РФ почти в два раза выше, чем в среднем по Европе



Охрана здоровья как задача национальной экономики

1. По данным за период 2010-2012 уровень смертности (на 100 тысяч человек): в Израиле – 533.4 умерших; в США – 642; в Бельгии – 635; в Норвегии – 549; в Италии – 511.6; в России - 1430; в Украине - 1520.
2. Особенно высока в России смертность в трудоспособном возрасте. Вероятность умереть в возрасте 15–60 лет в РФ почти в два раза выше, чем в среднем по Европе

Число больничных коек в РФ (в тысяч единиц, <http://fedstat.ru>)

	1985	1990	1996	2003	2004	2005	2006	2007	2008	2009
Всего	1 938	2 038	1 813	1 597	1 601	1 575	1 554	1 522	1 398	1 373

Число амбулаторно-поликлинических учреждений (<http://fedstat.ru>)

	1992	1996	2000	2004	2005	2006	2007	2008	2009
Россия	20 697	22 102	21 254	22 084	21 783	18 792	18 301	15 625	15 342

Индекс развития информационно-коммуникационных технологий (ICT Development Index) — комбинированный показатель, характеризующий достижения стран мира с точки зрения развития информационно-коммуникационных технологий (ИКТ). Индекс разработан в 2007 году на основе 11 показателей, которыми Международный союз электросвязи оперирует в своих оценках развития ИКТ. Россия на 40 месте из 157 стран.



Охрана здоровья

как задача национальной экономики

Потери. Анализ СМАЖ (журнал канадской медицинской ассоциации) :

- ▶ **Экономические потери от рака** прогнозируются на уровне \$100 млрд в странах с низким уровнем дохода и **\$2800 млрд** - со средним уровнем дохода (**РФ чуть выше среднего**).
- ▶ Заболеваемость раком растет во всем мире, а прогнозируемый прирост до 2030 года может составить 82% в странах с низким уровнем дохода и 70% в странах со средним уровнем дохода. Двадцать процентов всех случаев смерти от рака регистрируются в странах с низкими доходами и 50% в странах со средним уровнем дохода.
- ▶ Хотя эти страны не обладают средствами и ресурсами или **достаточно опытными медицинскими специалистами** для решения этих проблем, примерно одну треть от этих показателей можно предотвратить.
- ▶ **В России** доля больных, у которых рак был диагностирован на первой-второй стадии, составляет на 1 янв. 2012 50.5 процента (<http://fedstat.ru>). В странах Европы более половины всех случаев заболевания выявляется на ранних стадиях.
- ▶ В европейских странах пять лет и более, после установления диагноза, проживают от 55 до 75 процентов онкологических больных. В России этот показатель в 2010 году составил 51 %, в 2012 – 51.1%.

Решение: Новые технологии для здравоохранения на основе БОЛЬШИХ ДАННЫХ

1. Создание, наполнение, обслуживание ХРАНИЛИЩ БОЛЬШИХ ДАННЫХ (медицинских, биологических) на вычислительных кластерах (DATA-центризм), создание технологий, web и облачных сервисов на их основе → **необходима постоянная статья в бюджете РФ**
 2. 3D/4D- визуализация и реконструкция в медицине и биологии в РЕАЛЬНОМ ВРЕМЕНИ, высокопроизводительная многомасштабная визуализация. Обеспечение **диагностики**, теле-консультирования, эмуляторов, **ВИРТУАЛЬНЫХ АНАТОМИЧЕСКИХ СТОЛОВ** и образовательных 3D (в т.ч. стерео) пособий для медицинского и биологического образования
 - А. 3D-реконструкция, морфометрия и геометрическое исследование органов, систем и тела в целом, 3D-моделирование и печать для хирургии и травматологии
 - В. Слияние в 3D-модели человека всех методов исследования (прежде всего УЗИ, томографии, акустотермометрии). Гладкая сшивка 3D-моделей с томограммами.
 - С. Моделирование и планирование облучения в онкологии. Моделирование распространения света и излучений в биотканях и рассеивающих материалах (Монте-Карло)
-

Решение: Источники БОЛЬШИХ ДАННЫХ

Закупки томографов в России. Объемы хранения

Российский рынок диагностического оборудования в 2009 году, по оценке Philips, составил 32,3 млрд рублей, в 2010 год — 37,1 млрд рублей. Продажи КТ составляют 1–1,5 млрд рублей в год. Этот сегмент примерно поровну делят ведущие мировые производители медтехники: Philips, General Electric, Siemens и Toshiba. Их дилеры ежегодно продают в России около 100–150 томографов.

МРТ в РФ – 450шт, в США - 30тыс.шт. Потребности российской медицины Philips оценивает в 3,5 тыс. КТ http://www.dp.ru/a/2010/05/18/Philips_otkril_v_Peterbur/.

Основным покупателем КТ является государство (по данным сайта госзаказа РФ, в середине 2010 года медучреждения объявили о закупке пяти томографов на сумму 708 млн рублей (по 140млн.р.)).

Российский томографы (низкополевые, 0.15-0.3Т):

<http://az-mri.com/ms/index1.php>; народный томограф "Юнитом" (<http://unitom.ru/>)

Если учитывать то, что каждый россиянин должен проходить диспансеризацию хотя бы раз в два года, то томографов в России, должно быть как минимум 10000 штук. → 150млн.чел./600дн./10000том.=25чел/день /томограф

Ресурс хранения для РФ, 2 года: $0.15 \cdot 10^9 \text{ч.} \cdot 10^9 \text{Б} = 0.15 \text{ЭкзБ}$
Р.х. для крупного района (200000чел): 0.2ПБ. **Города 1млн. – 1ПБ.**

Поток данных с одного томографа:

Интенсивность в сеансе - 1ГБ за 10сек (100МБ/сек);

1ГБ/чел*30чел/день = 30ГБ/день; в год 9000чел – 9ТБ. С 1000 томографов: 30ТБ/день; 9ПБ/год

Таб.1. TOP-10 моделей томографов по объемам импорта в Россию за период 1-2 кв. 2010 г. в денежном и натуральном выражении

№	Модель	Производитель	Объем импорта, млн. USD	Объем импорта, ед.	Динамика импорта 1-2 кв. 2010/1-2 кв. 2009, %
1	AQUILION ONE	TOSHIBA	8,7	1	-30
2	AQUILION 32	TOSHIBA	5,9	3	267
3	EXCELART VANTAGE TITAN	TOSHIBA	4,5	1	-
4	APERTO	HITACHI	4,0	3	-
5	EXCELART VANTAGE ATLAS-X	TOSHIBA	3,5	1	-
6	MAGNETOM AVANTO	SIEMENS	3,3	1	283
7	EXCELART VANTAGE	TOSHIBA	3,0	1	-28
8	BRIGHTSPEED ELITE	GE	2,9	5	113
9	INTERA	PHILIPS	2,7	4	998
10	SOMATOM SENSATION 64	SIEMENS	2,6	2	-2

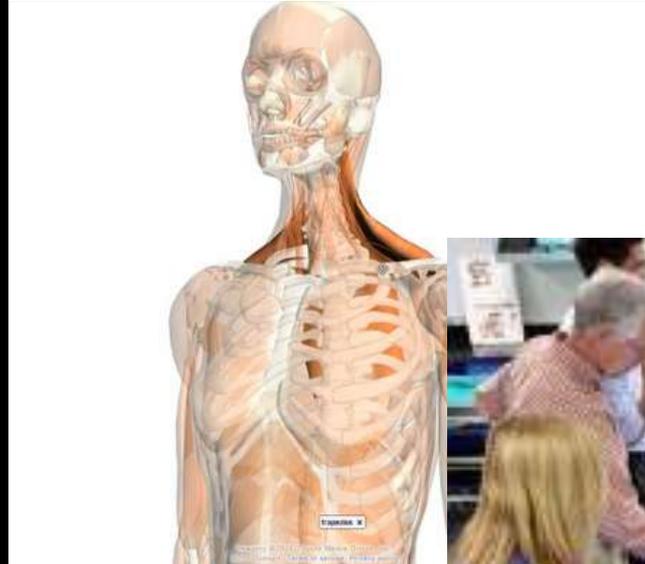
ReVin

Актуальные задачи работы с большими данными

1. **Объемы данных, связанные с современными методами диагностики, прежде всего с методами томографии, и методами лечения в медицине (КТ + ПЭТ, КТ + ОФЭКТ, ...). Будут ежегодно пополняться:**
 - a) **на персональном уровне:** 1-10 гигабайт для хранения двух томограмм пациента в каждом сеансе и регулярного наблюдения хода лечения – в период лечения (Google выделяет 15GB под почту).
 - b) **на уровне клиники (1-2 томографа):** 1ГБ/чел*30чел/день/томограф => 30-60ГБ/день; **в год** 9000чел – 9-18ТБ; N лет = 18*N ТБ
 - c) **на уровне территорий:** район (200000чел), ресурс хранения - 0.2ПБ * N/2 лет; **город** (1млн.чел) - 1ПетаБ * N/2 лет; **РФ** - 0.15*10⁹чел * 10⁹Б = 0.15ЭкзаБ * N/2 лет
 2. **Условия и возможности превращения больших медицинских данных в технологический и стратегический ресурс, работающий на развитие страны и здоровья нации:**
 - a) Классификация и резервирование хранимых данных
 - b) Data centric политика в организации обработки данных – совмещение хранилищ с вычислительными кластерами (дооснащение) для исполнения сервисов на больших данных
 - c) Создание программного обеспечения задач многопоточковой массивнопараллельной обработки больших данных на кластерах как сервисов (SaaS)
 - d) Обеспечение высокопроизводительных коммуникаций для массового доступа к данным
-

Web и облачная 3D визуализация томограмм и анализ больших данных в медицине

3D реконструкция человека как современная медицинская карта
(проект IBM с 2008)



Виртуальный анатомический стол (<http://anatomage.com>)

Существующие компьютерные анатомические атласы

<http://www.zygotebody.com/>

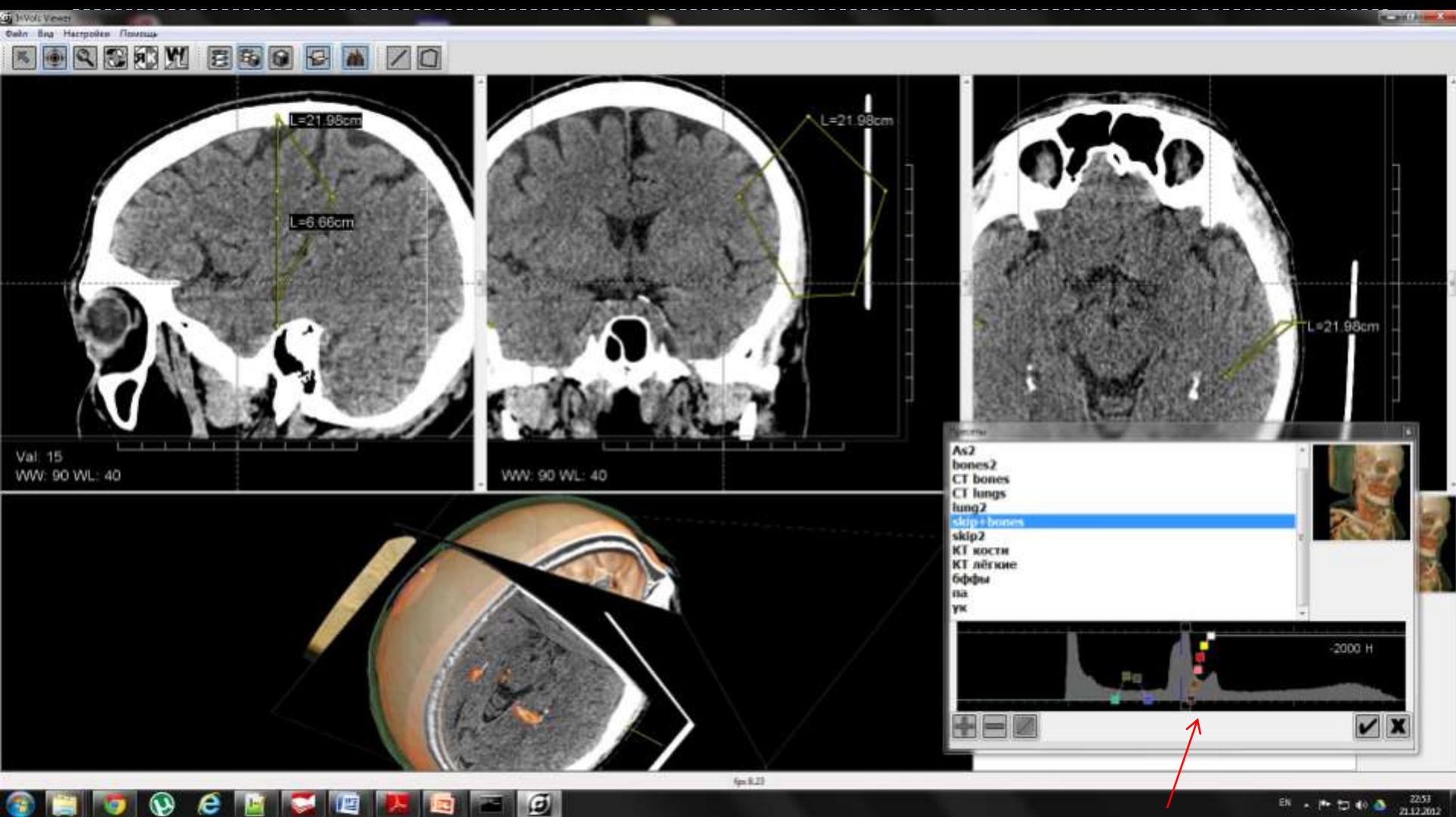
<http://www.gizmonews.ru/2010/12/17/google-body-browser-webgl/>

<http://anatomya-atlas.ru/>

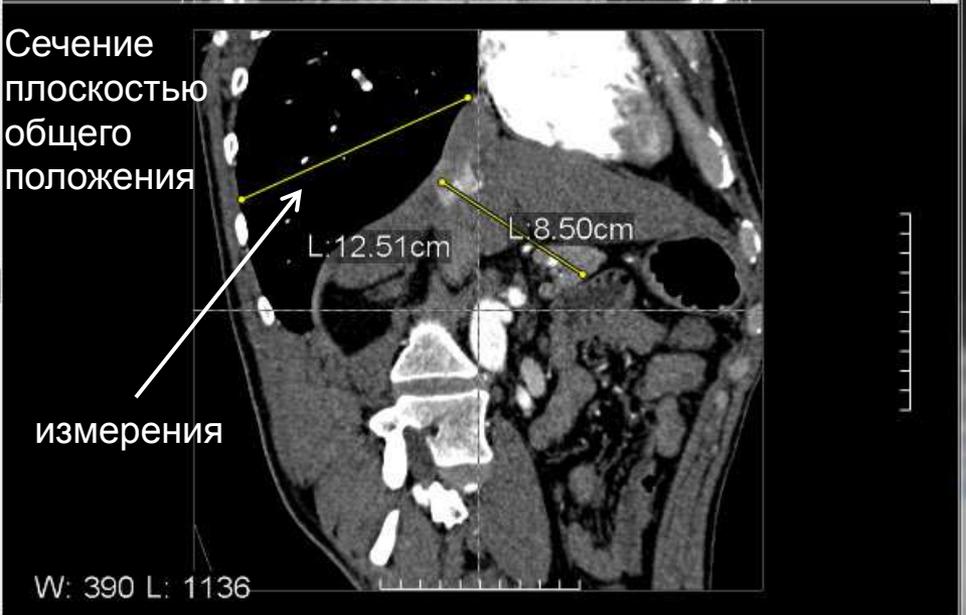
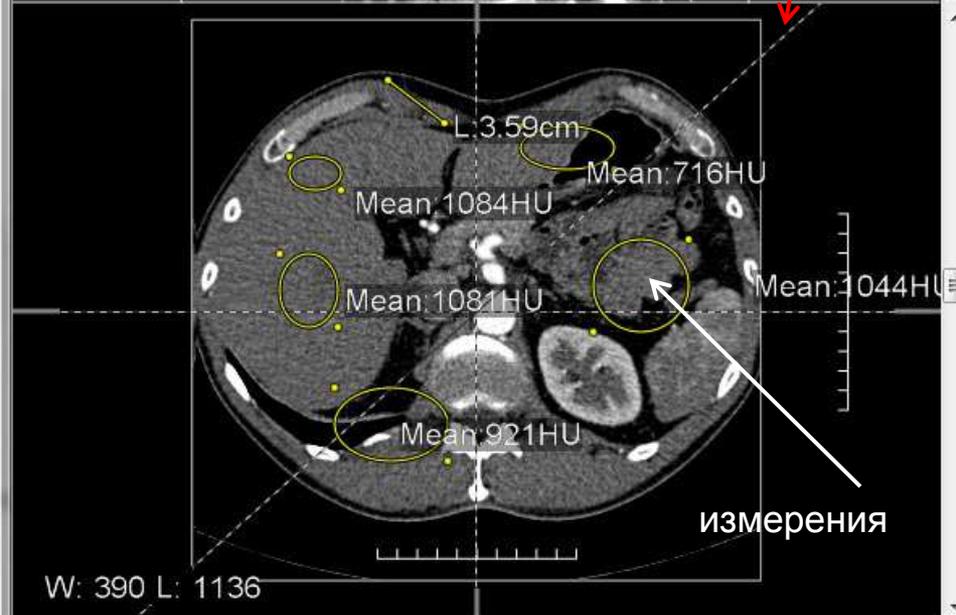
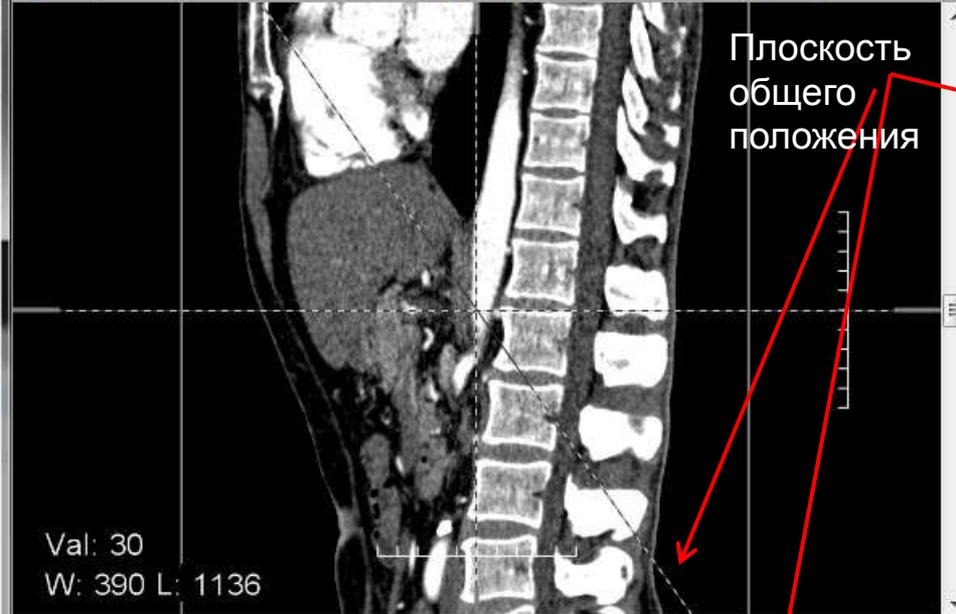
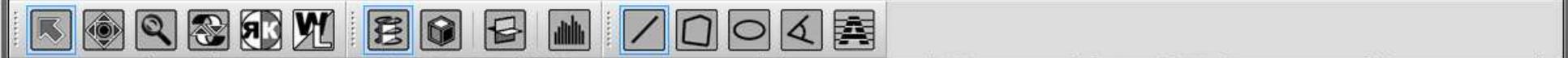
<http://pulsplus.ru/medcare/medatlas/>

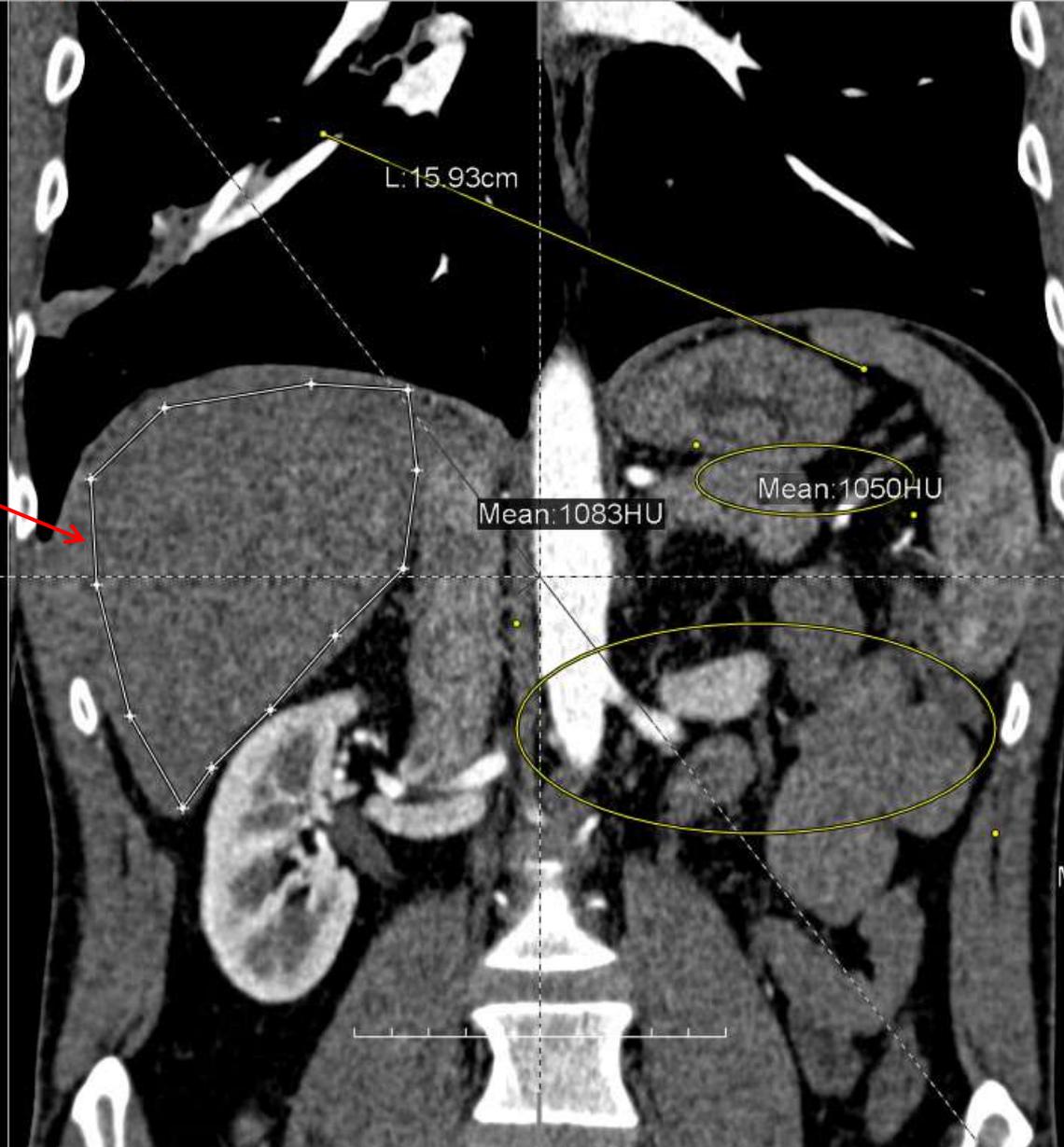


1. Сервис для 3D-визуализации, анализа и диагностики на основе томограмм



Transfer Function





Измерение
объемов

Оконтуривание

Val: 826
W: 390 L: 1136

Mean:1074HU

Mean:1083HU

Mean:1050HU

L:15.93cm

Визуализация с использованием одновременно нескольких Transfer Function

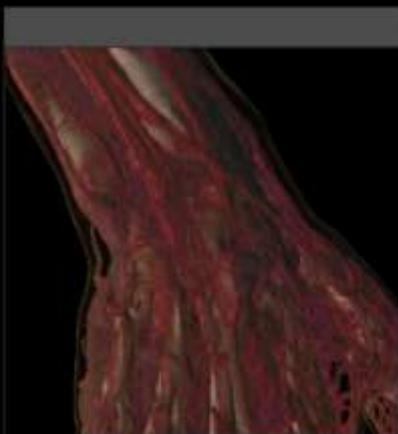
STS (semitransparent surfaces)



MIDA (Maximum intensity differences accumulation)



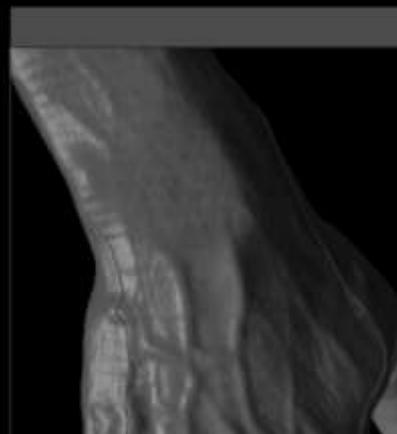
MIP (maximum intensity projection)



sDVR (shaded Direct Volume Rendering)



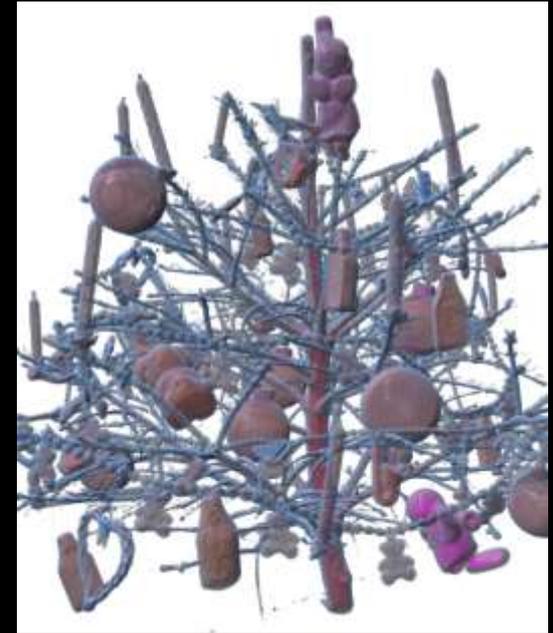
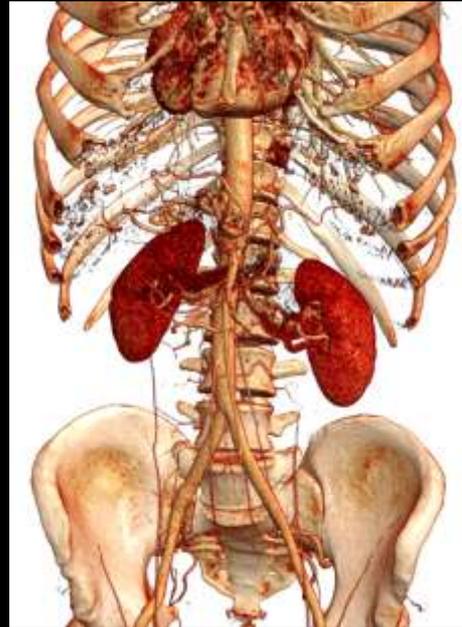
Opaque isosurface



sDVR via the 2D transfer function



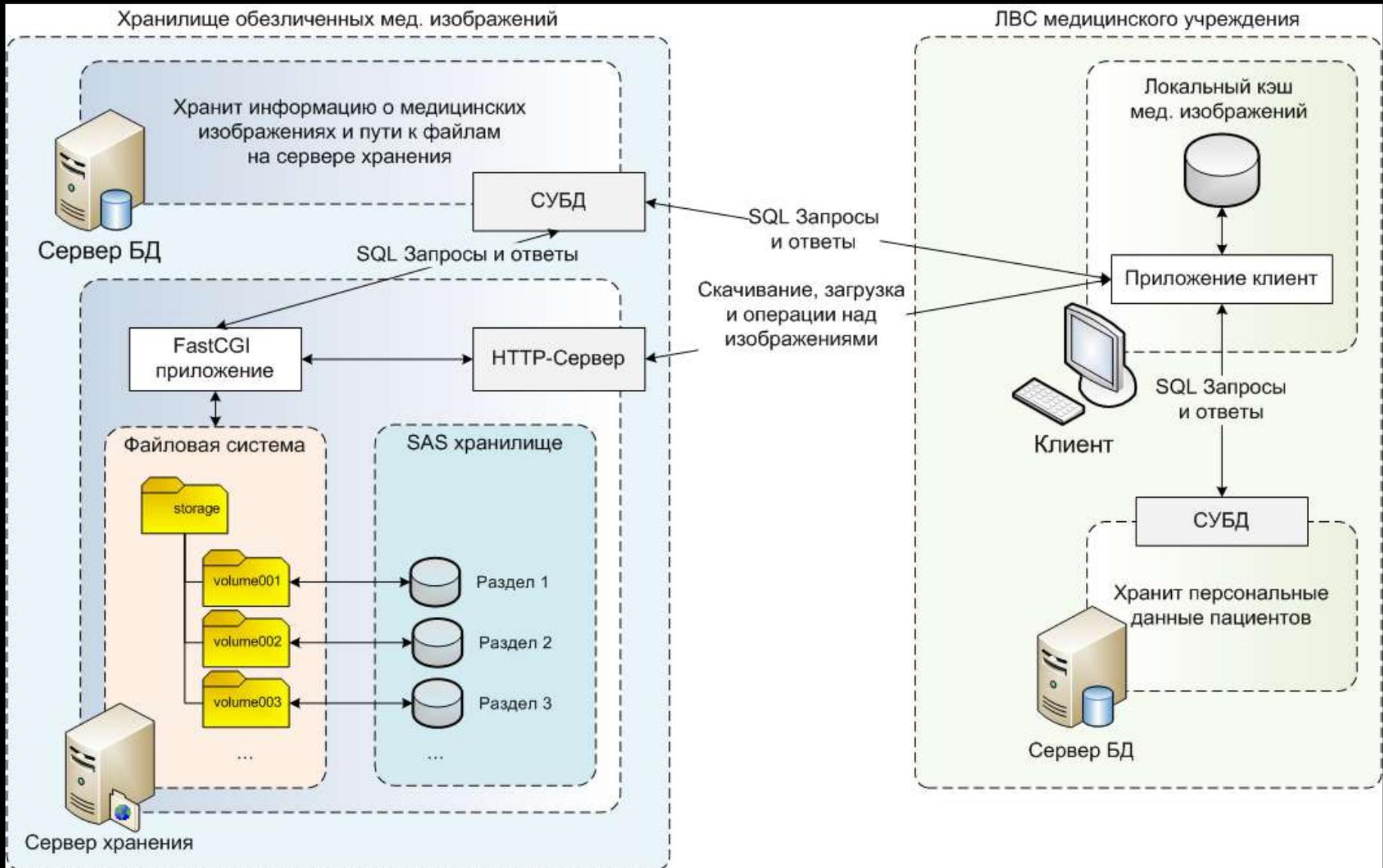
Совмещение высокого качества визуализации на GPU с рекордной производительностью



5382 x 512 x 512 x12bit

[CT_5000.avi](#)

2. Хранилище медицинских данных на вычислительном кластере (в формате DICOM)

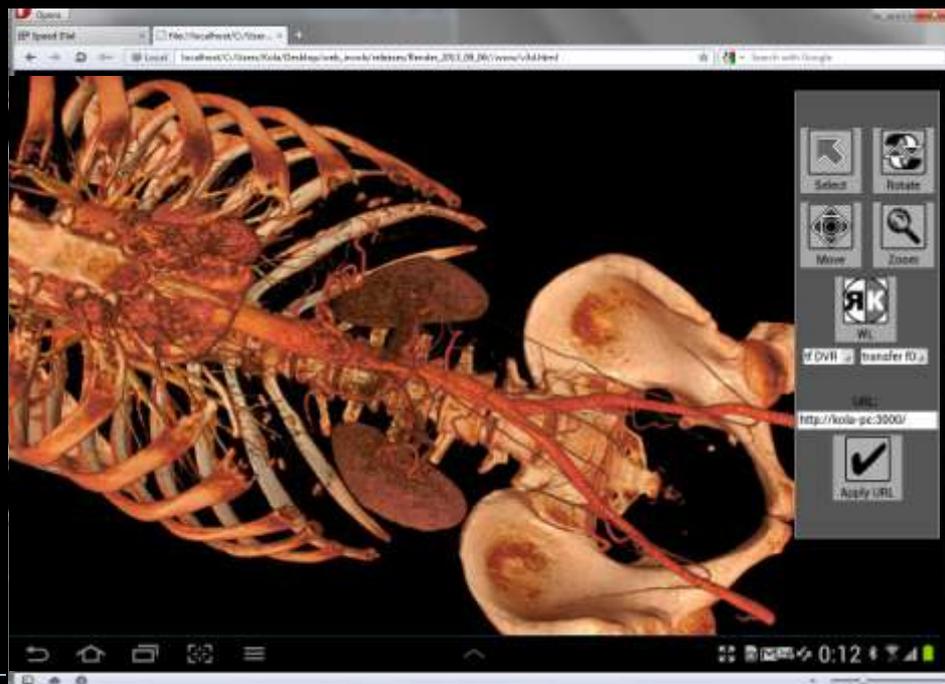


3. Web-интерфейс и облачный сервис 2D- и 3D-визуализации на удаленном хранилище (10000 томограмм на кластере ННГУ)

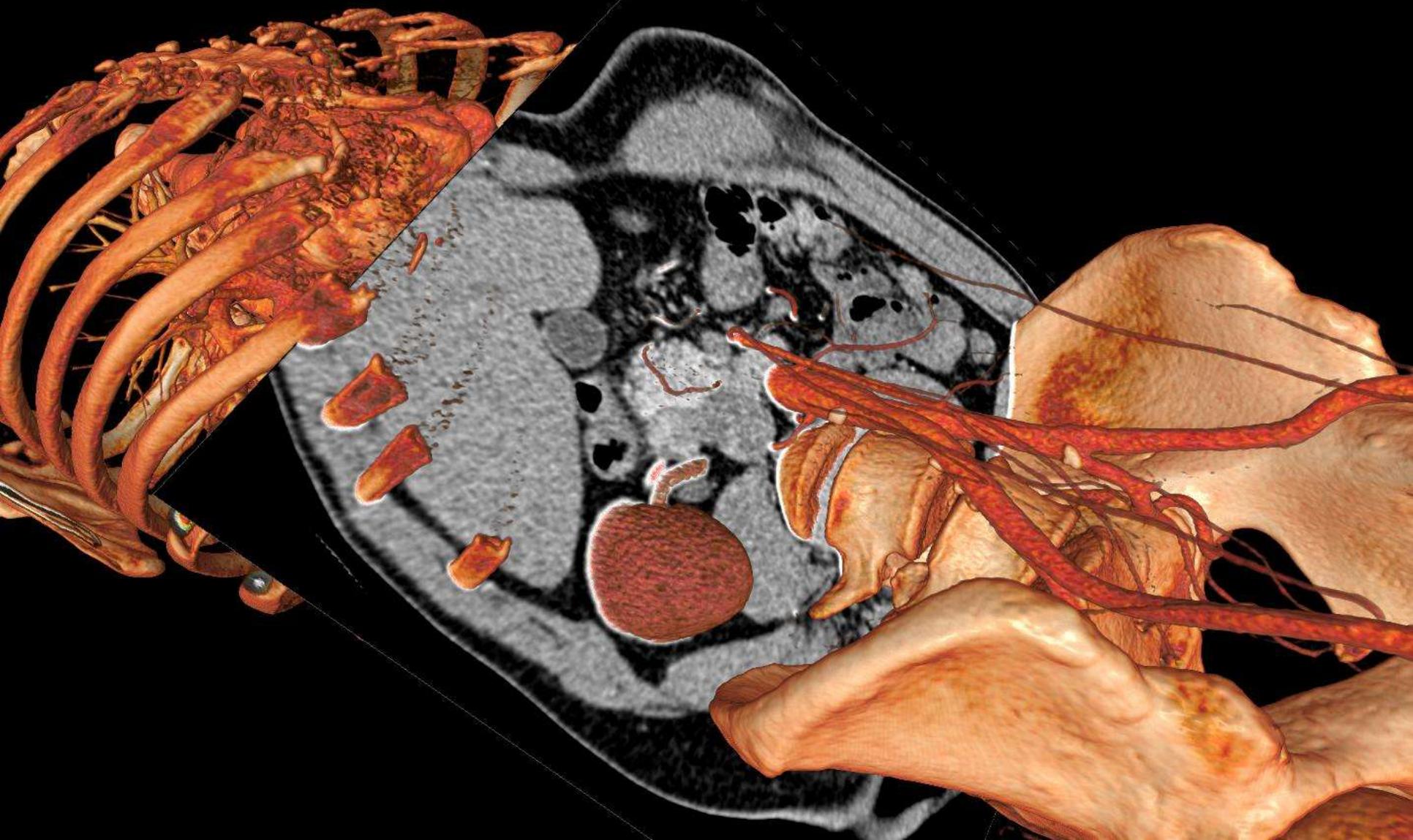
Будет экспонироваться на международной выставке-форуме «ОТКРЫТЫЕ ИННОВАЦИИ» в Москве с 31 октября по 2 ноября (<http://www.forinnovations.ru/>)

Особенности реализации:

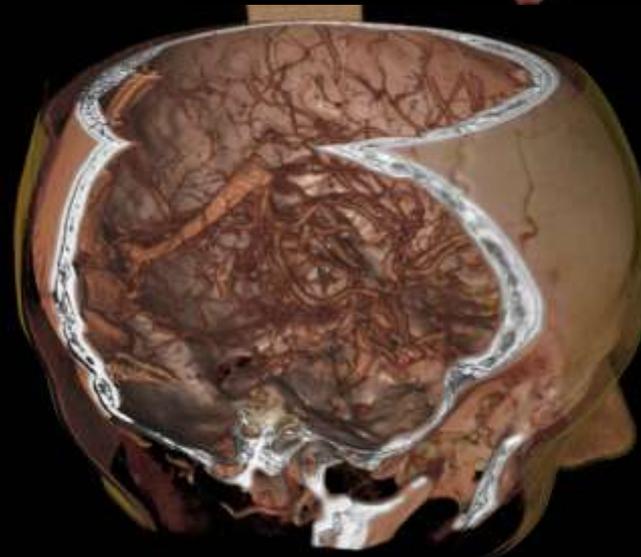
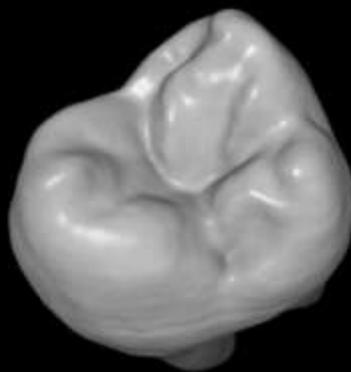
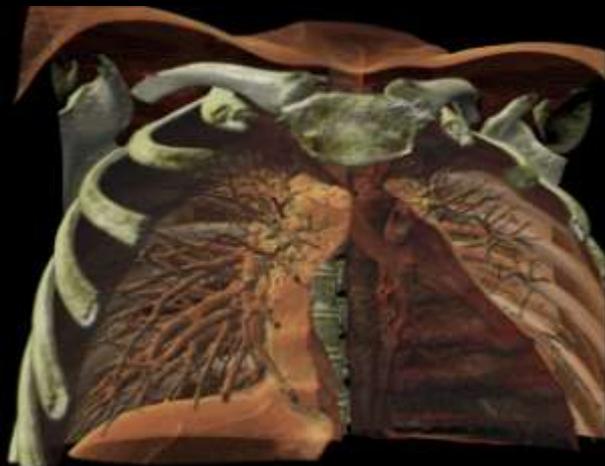
- 1) Работа с хранилищем и вычислениями на гетерогенном кластере.
- 2) “Тонкий клиент” (в качестве клиента - web-браузер). Использование HTML5
 - a) технология multi touch на мобильных устройствах (ОС «Android»)
 - b) быстрая растеризация получаемых изображений



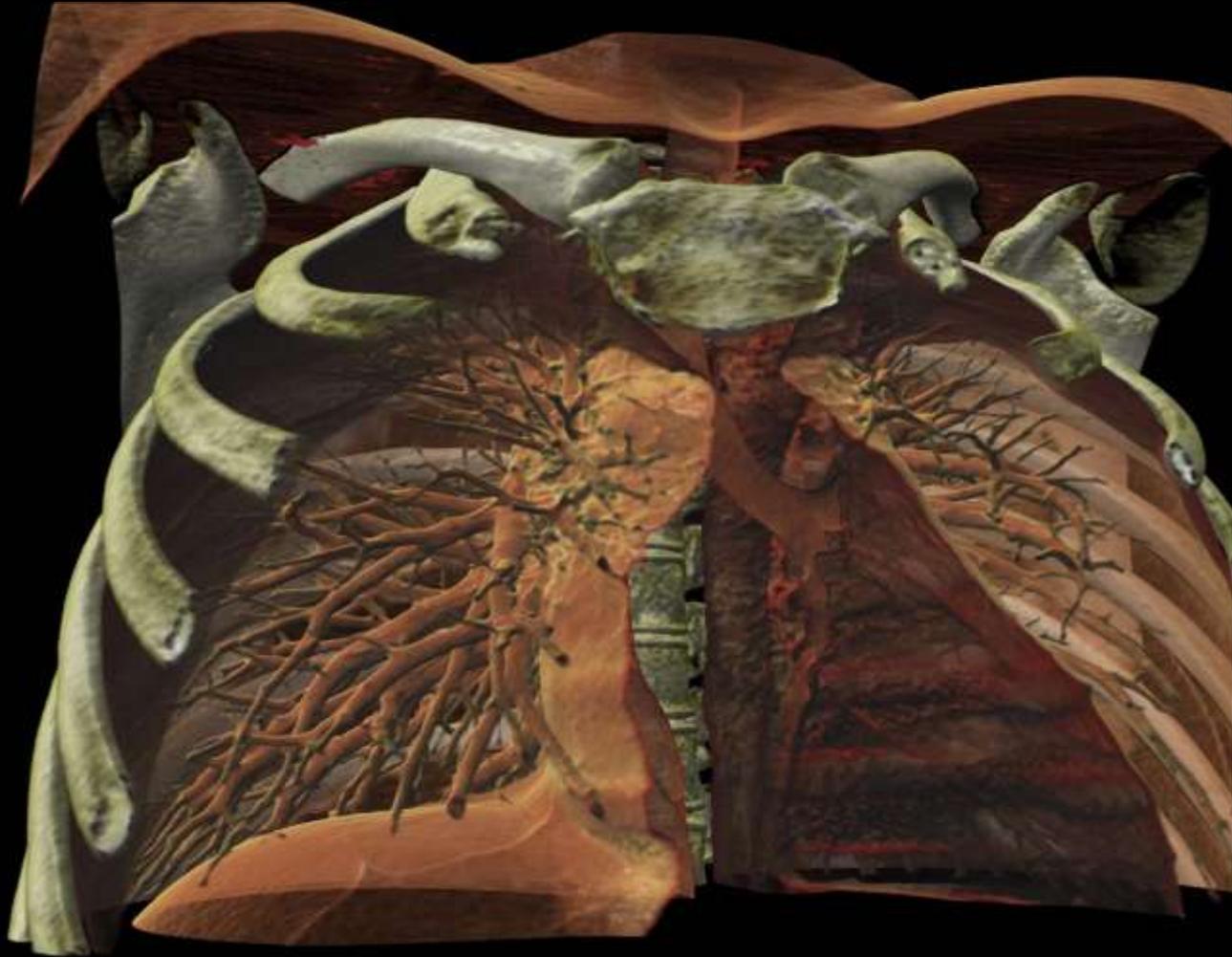
3D + МПР (мультипланарная реконструкция),
сечение произвольной плоскостью



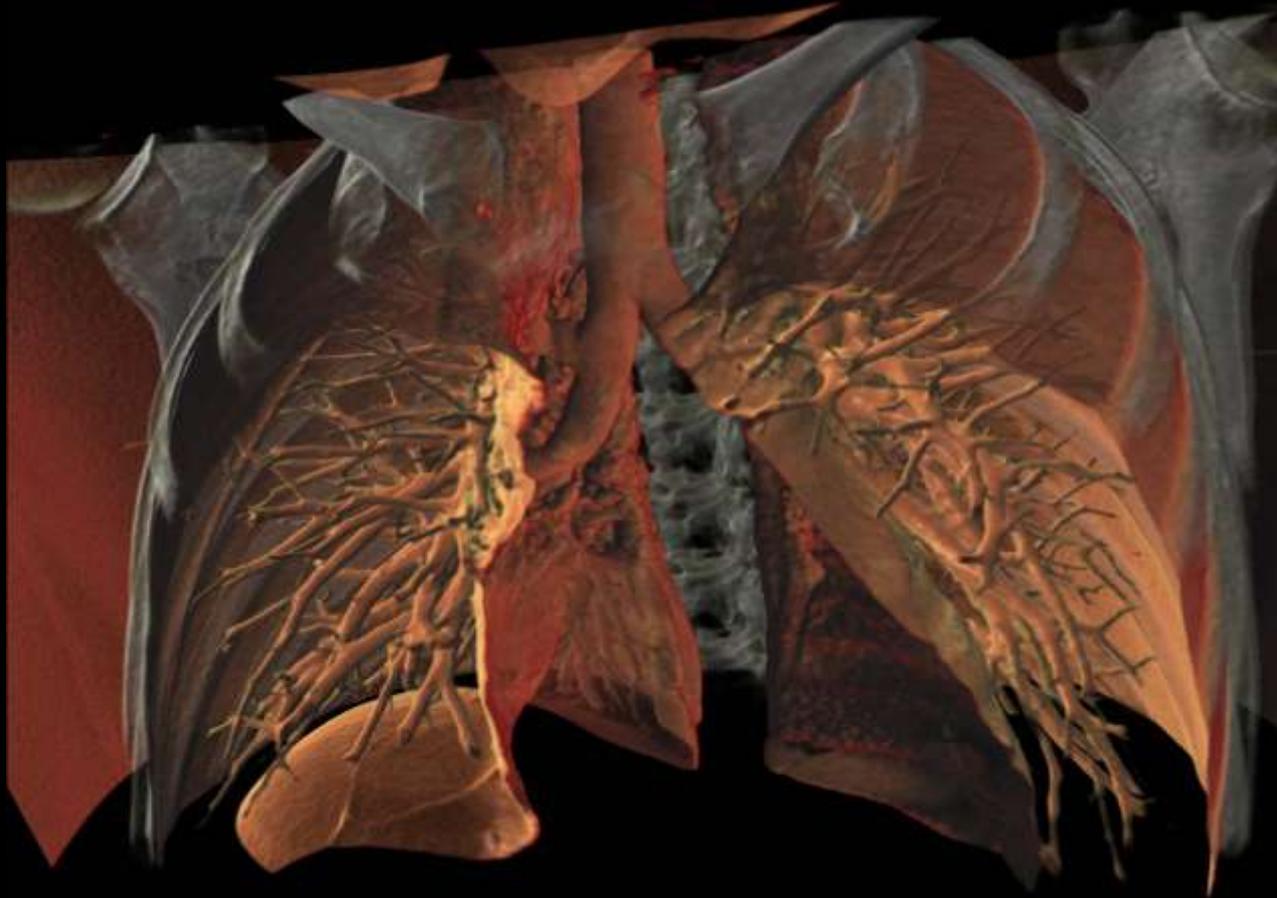
3D реконструкция в хирургии и стоматологии



Высокое качество визуализации. Примеры



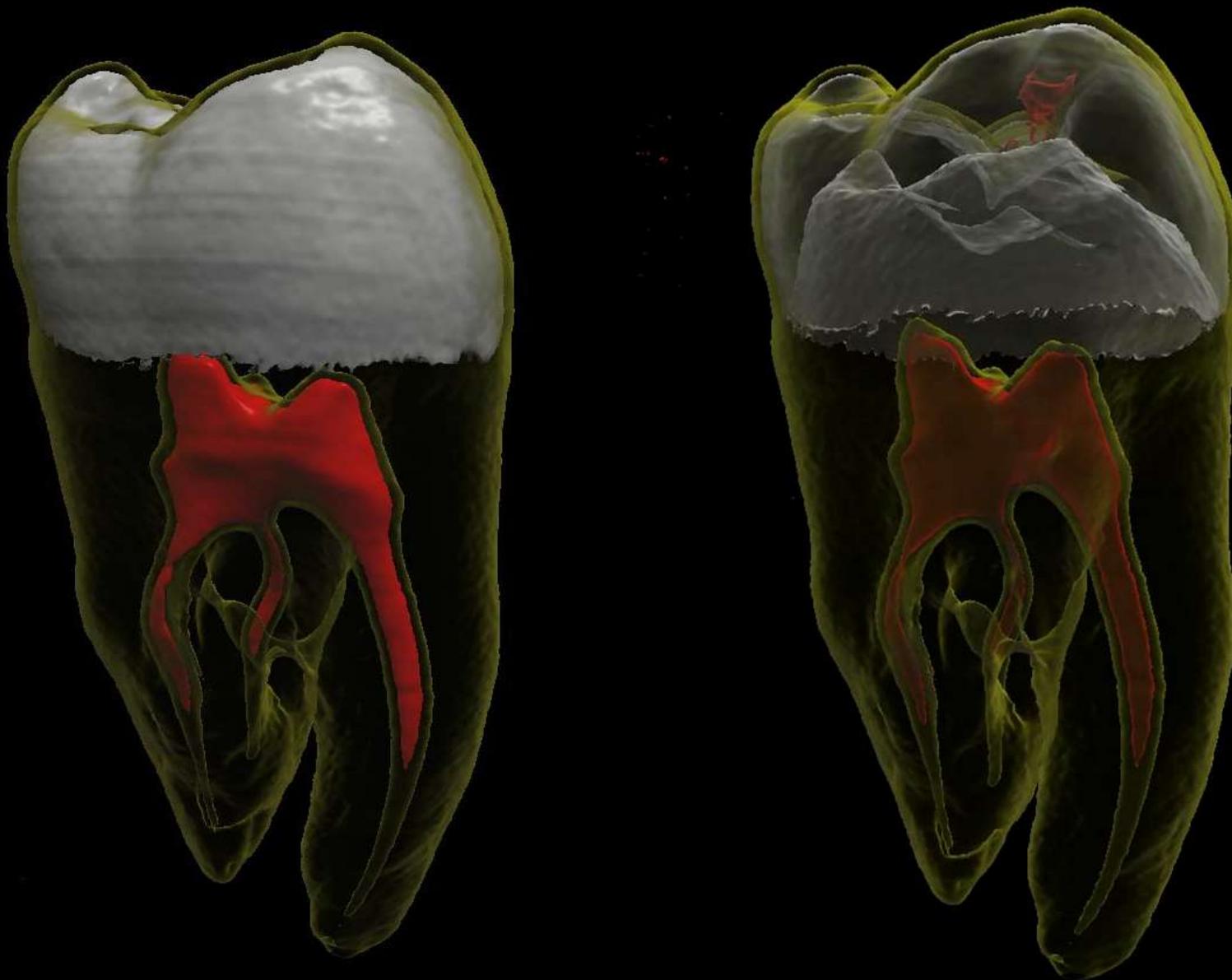
Высокое качество визуализации. Примеры



Высокое качество визуализации. Примеры

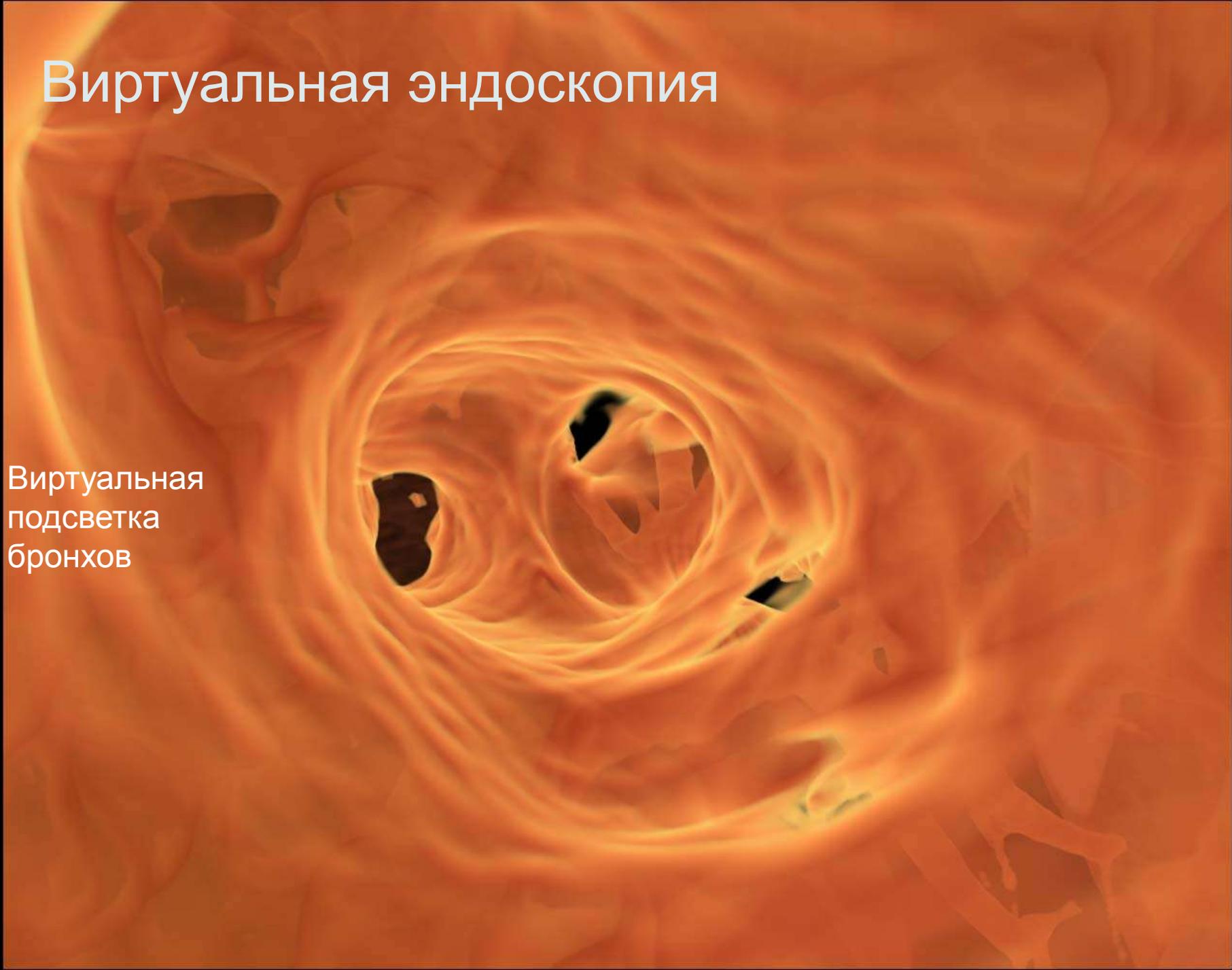


Высокое качество визуализации. Примеры

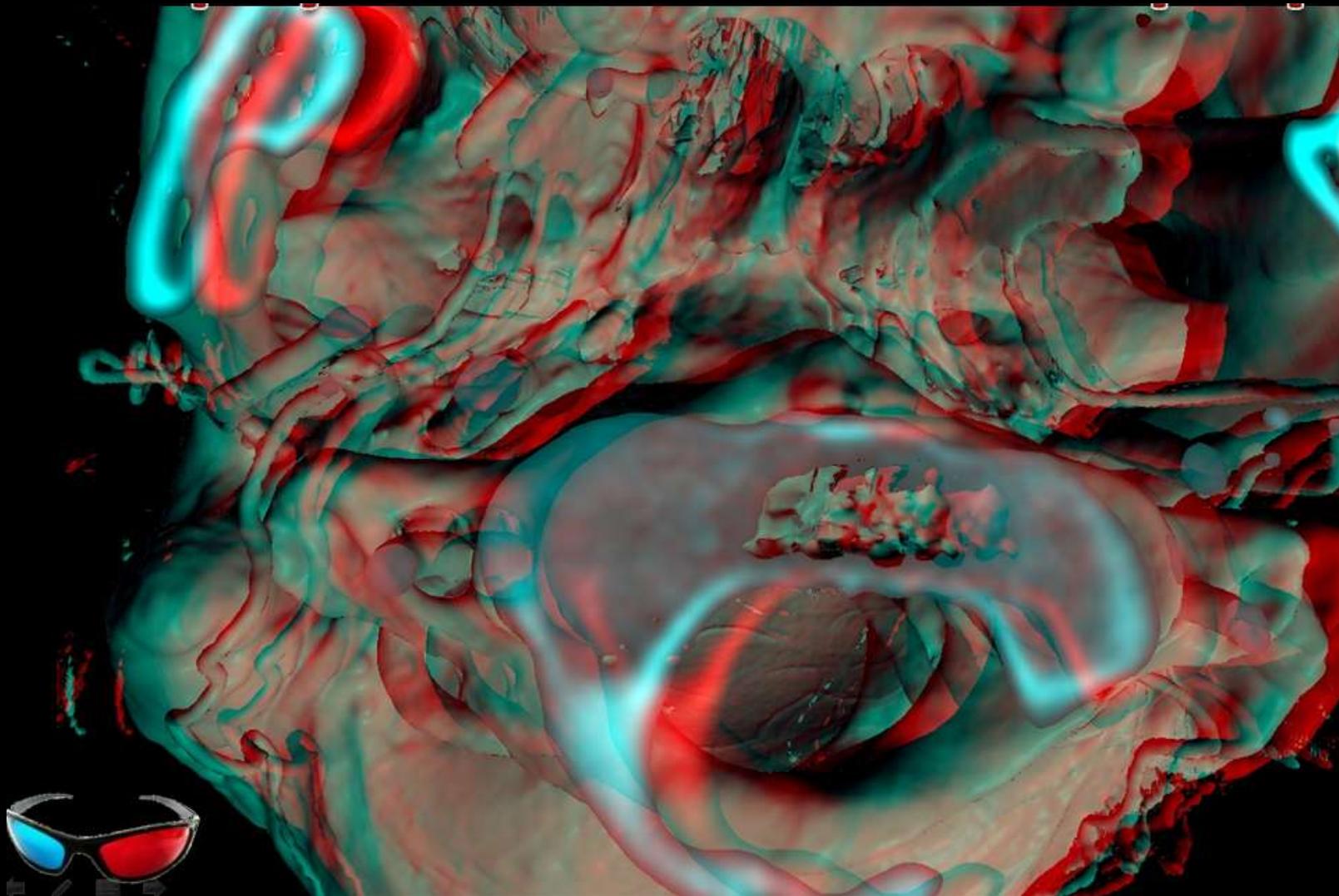


Виртуальная эндоскопия

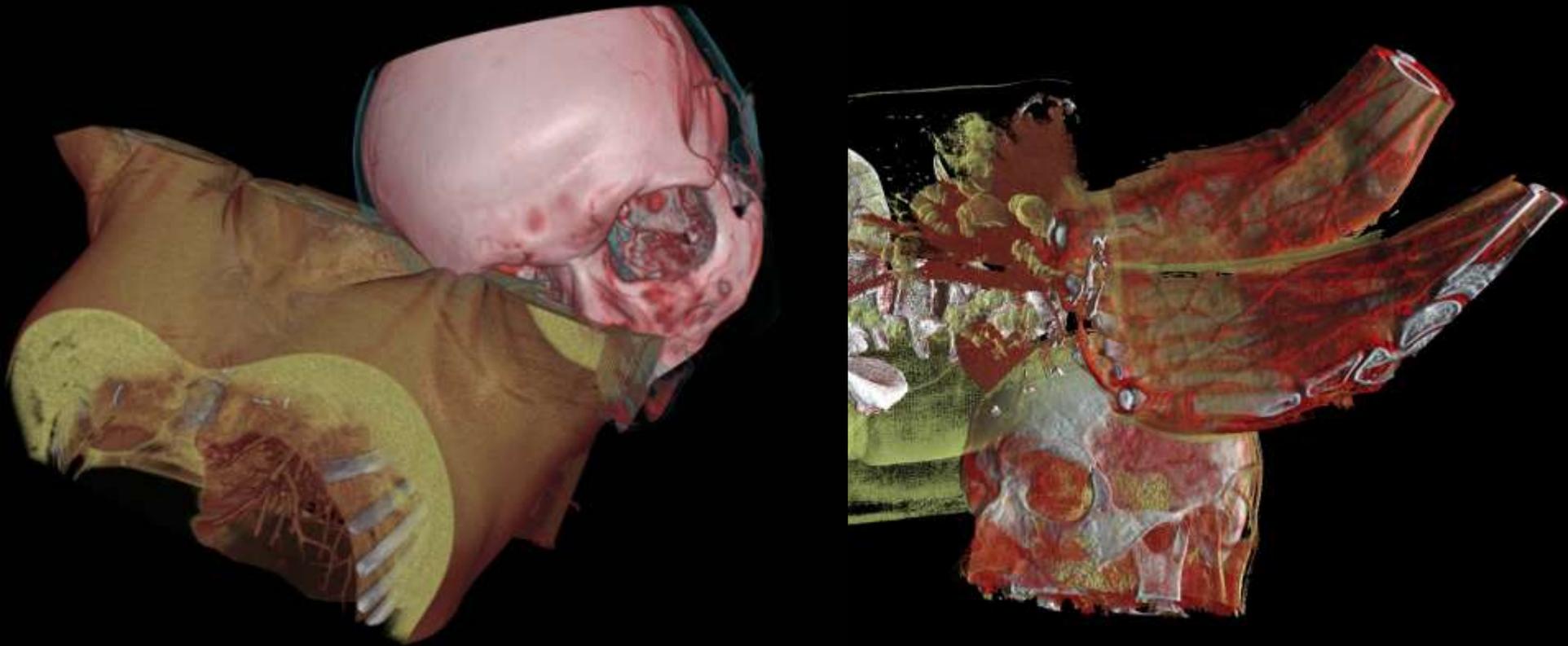
Виртуальная
подсветка
bronхов



Пример стереовизуализации



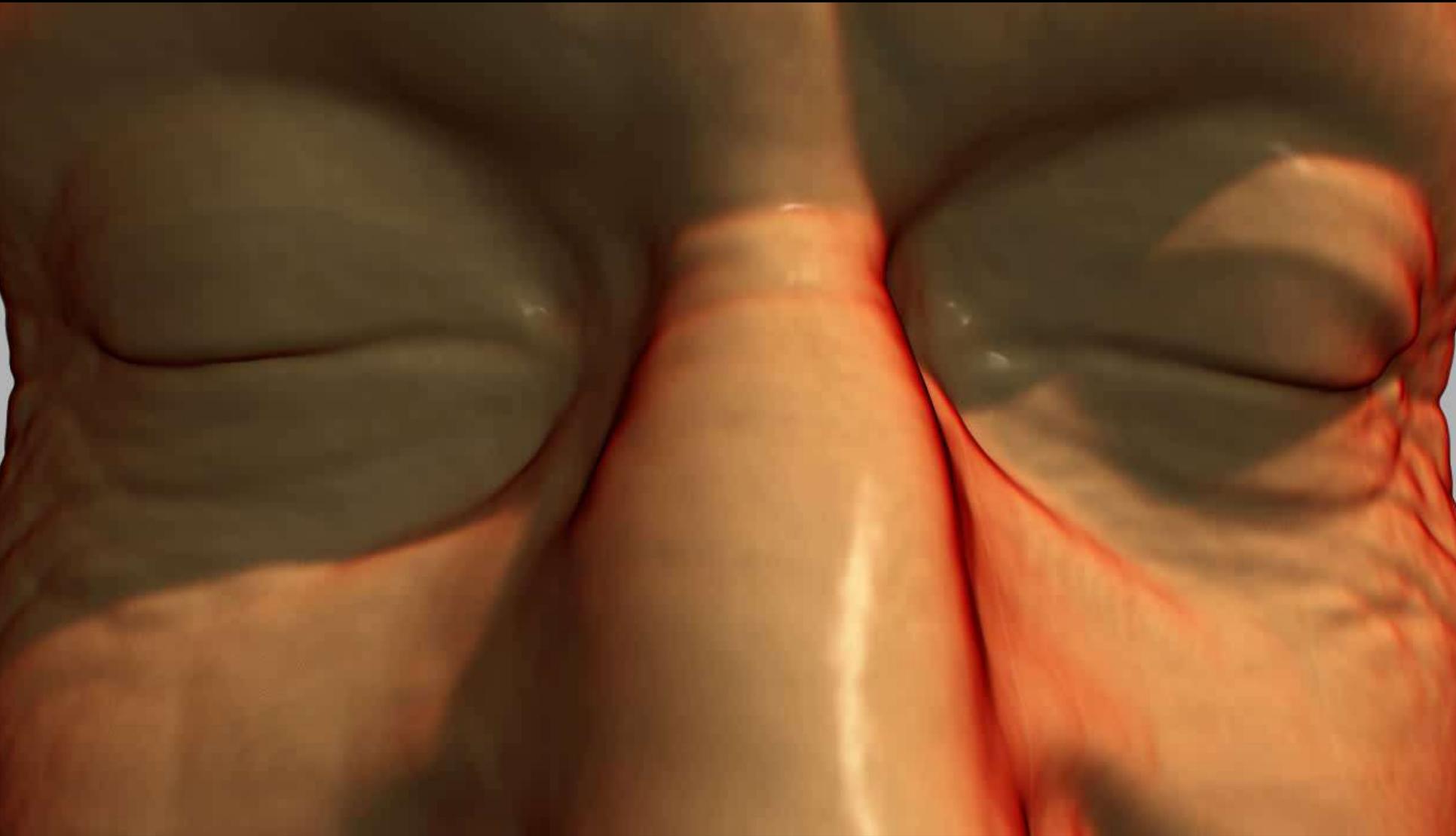
Мульти-объёмный рендеринг



Для получения полной картины исследования разнородные массивы данных могут быть визуализированы совместно



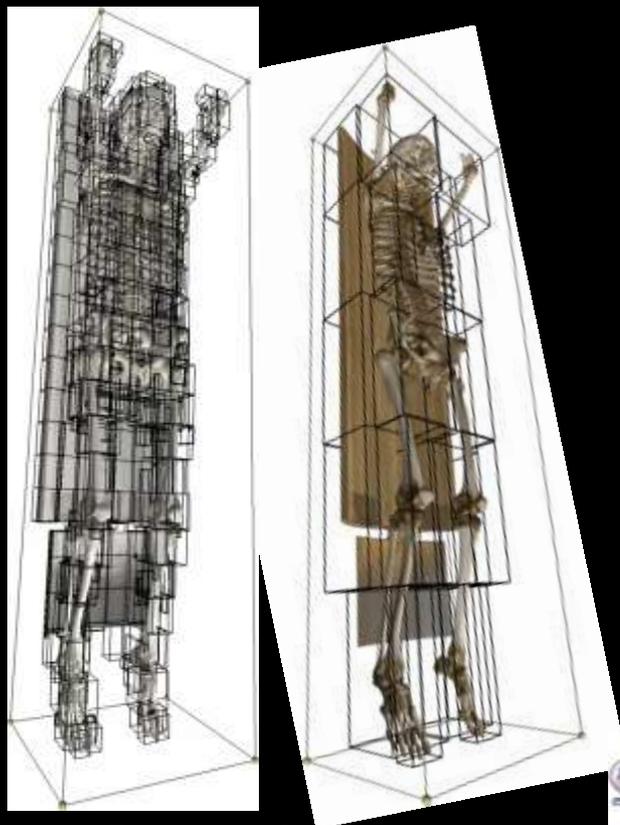
Пример визуализации теней в условиях блочной декомпозиции данных



Пример визуализации теней в условиях блочной декомпозиции данных



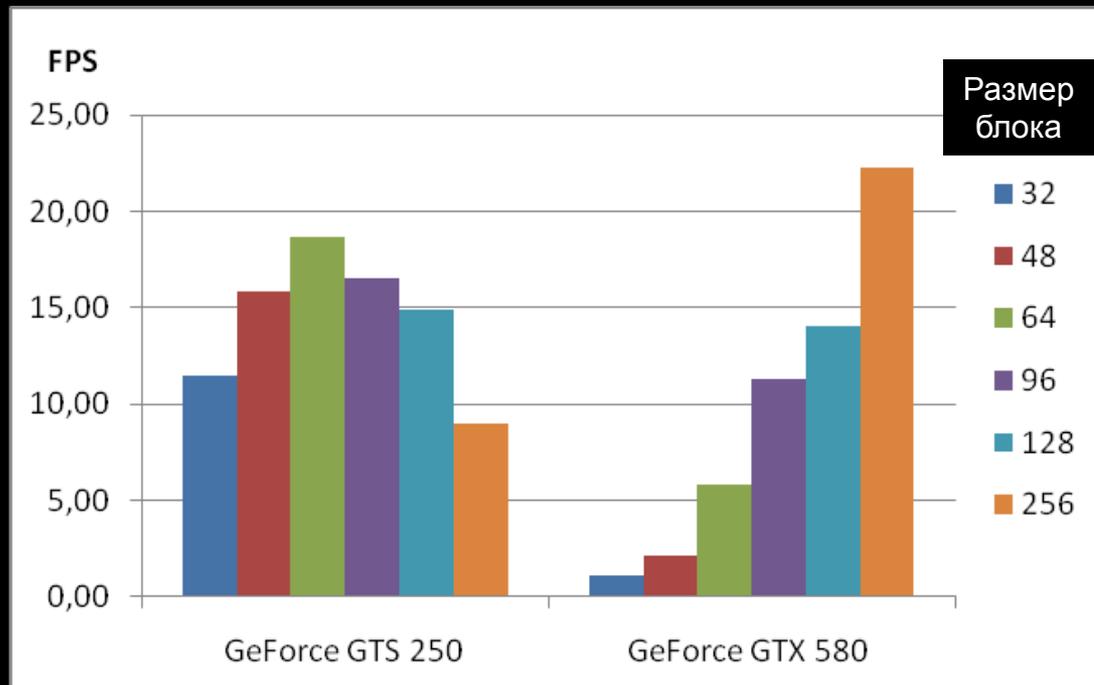
Оптимизация производительности по размеру блока декомпозиции



Размеры блоков:

64x64x64

256x256x256



[CT_5000.avi](#)

пример тестовых данных размером ~2GB:

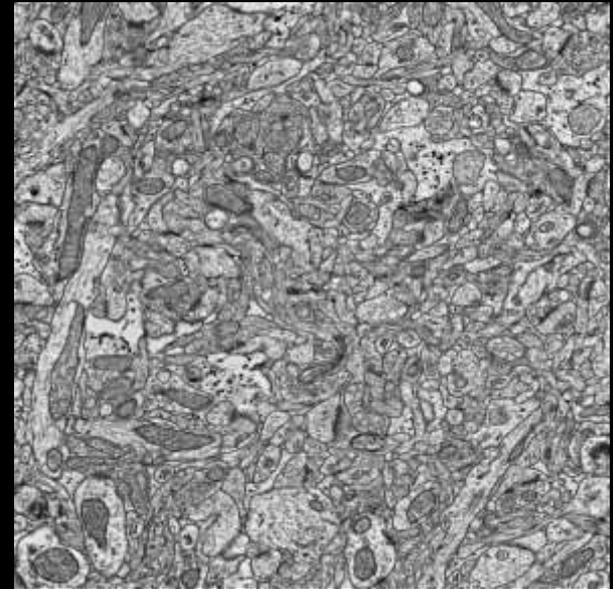
5382x512x512 (12-bit)

[DVR.wmv](#)

Новые возможности на основе облачного 3D-визуализатора и ПО на его основе

1. Хранение всех медицинских исследований пациента в персональном кабинете в виде ссылок на адреса реального хранения.
2. Возможность автоматического совместного анализа всей совокупности объективных данных пациента, изменений во времени. Обнаружение болезни на ранней стадии.
3. Возможность обращения за консультацией и лечением к любому врачу мира.
4. Повышение квалификации врачей и медицинской культуры населения.
5. Новая цифровая медицина

**Возможности
3D-реконструкции
и геометрического
исследования клеток мозга
на GPU**



[1] [Davi D. Bock](#), [Wei-Chung Allen Lee](#), [Aaron M. Kerlin](#), [Mark L. Andermann](#), [Greg Hood](#), [Arthur W. Wetzel](#), [Sergey Yurgenson](#), [Edward R. Soucy](#), [Hyon Suk Kim](#) & [R. Clay Reid](#). Network anatomy and *in vivo* physiology of visual cortical neurons // Nature. march 2011. V.471, P.177–182

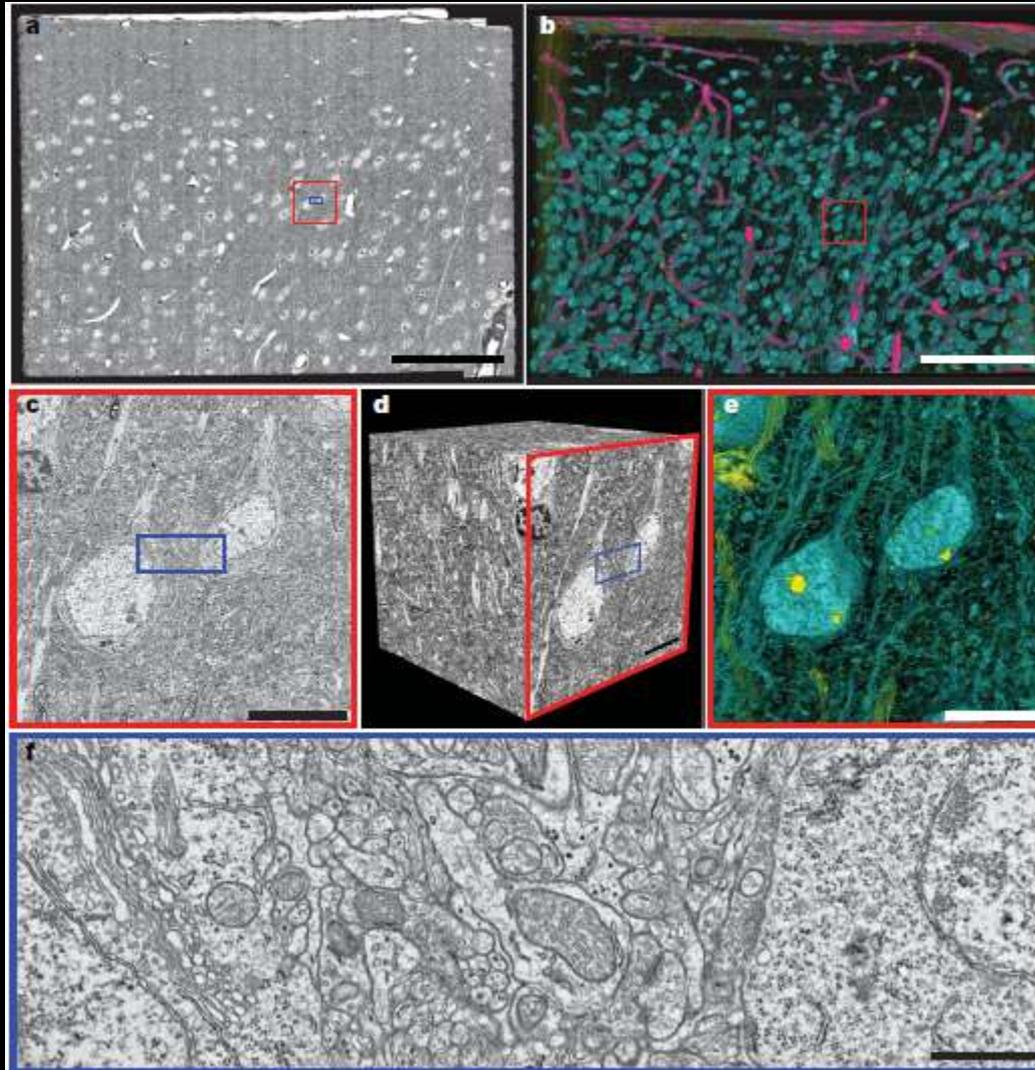
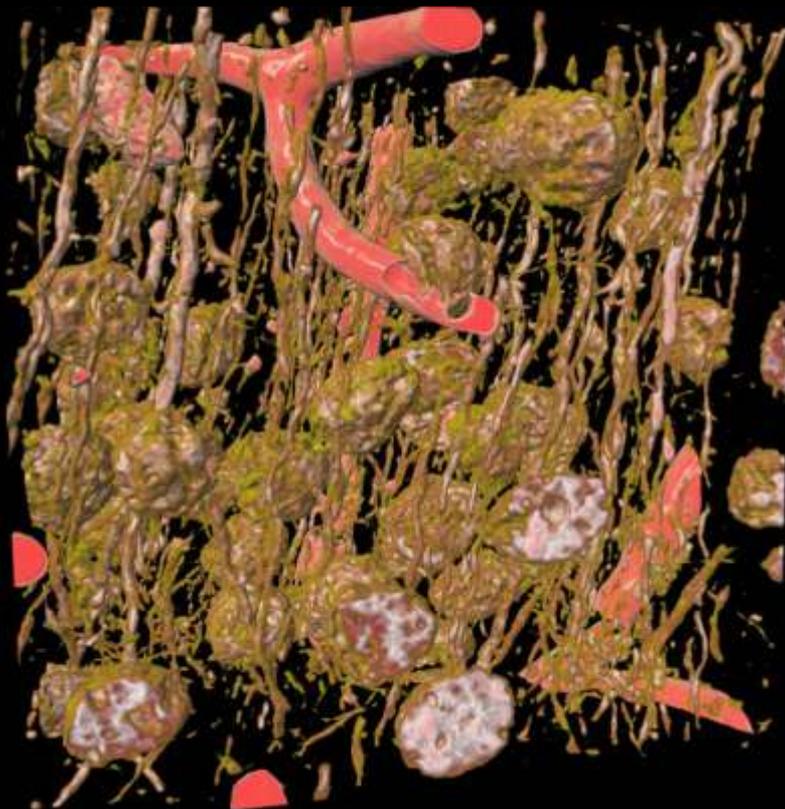
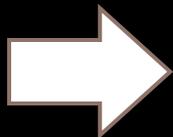
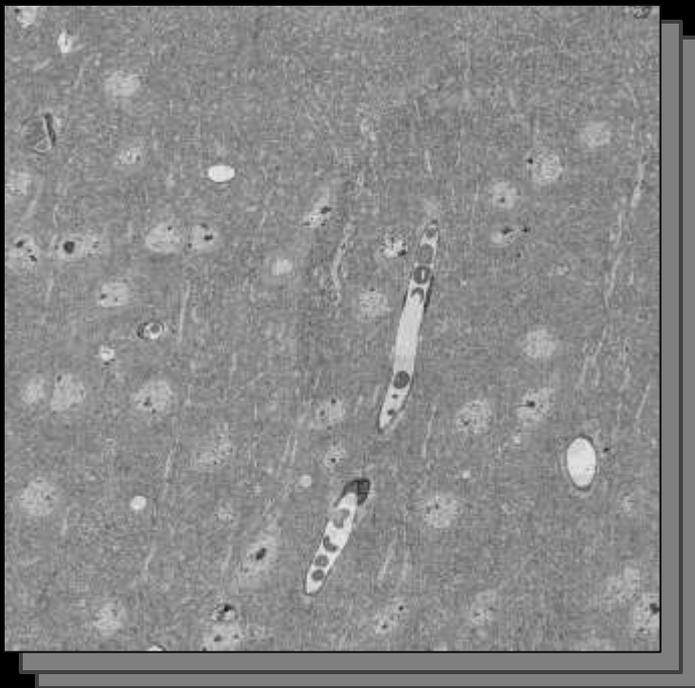


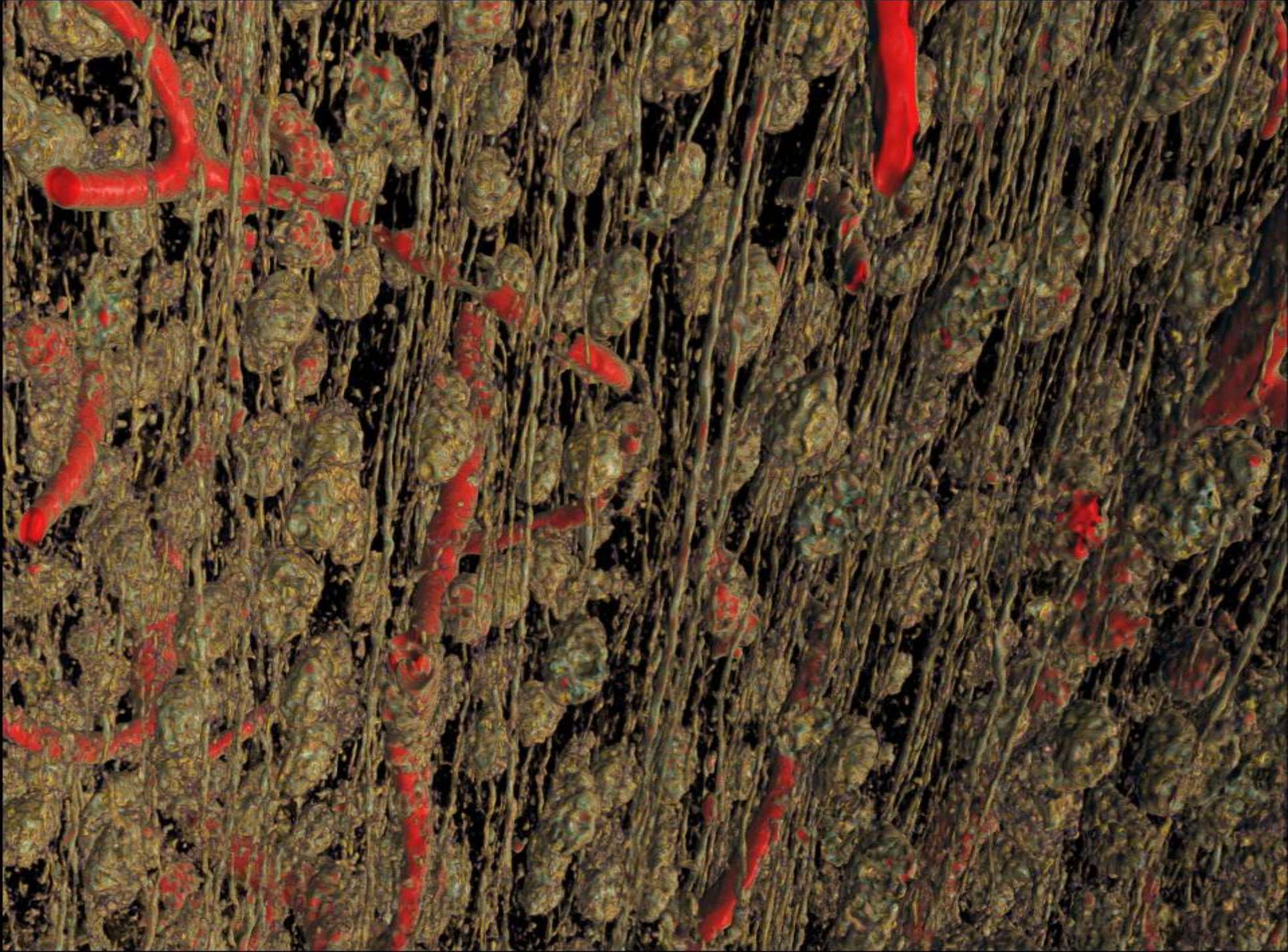
Figure 2 | Large-scale EM. **a**, Electron micrograph of an entire 120,000 × 80,000 pixel thin section, showing the pial surface (top) and cortical layers 1 through to upper 4 (bottom). **b, e**, Three-dimensional renderings of the EM volume through the entire series (**b**, Supplementary Movie 3) and through 50 sections of the cube in **d** (**e**, red outline in **b**, Supplementary Movie 4). **d**, A cube of the EM volume with **c** as one face. **c, f**, Zoomed-in view of two functionally characterized cells (**c**; red outline in **a**), and the neuropile between them (**f**; blue outlines in **a, c** and **d**), illustrating the density of axons and dendrites coursing between cell bodies. Pink represents electron transparent regions (for example, blood vessels), yellow represents regions that are electron dense (such as nucleoli, oligodendrocyte nuclei, and myelin), and aqua denotes regions with pixel values in between (for example, nuclei, cell bodies and dendrites). Scale bars: **a, b**, 100 μm; **c–e**, 10 μm; **f**, 1 μm.

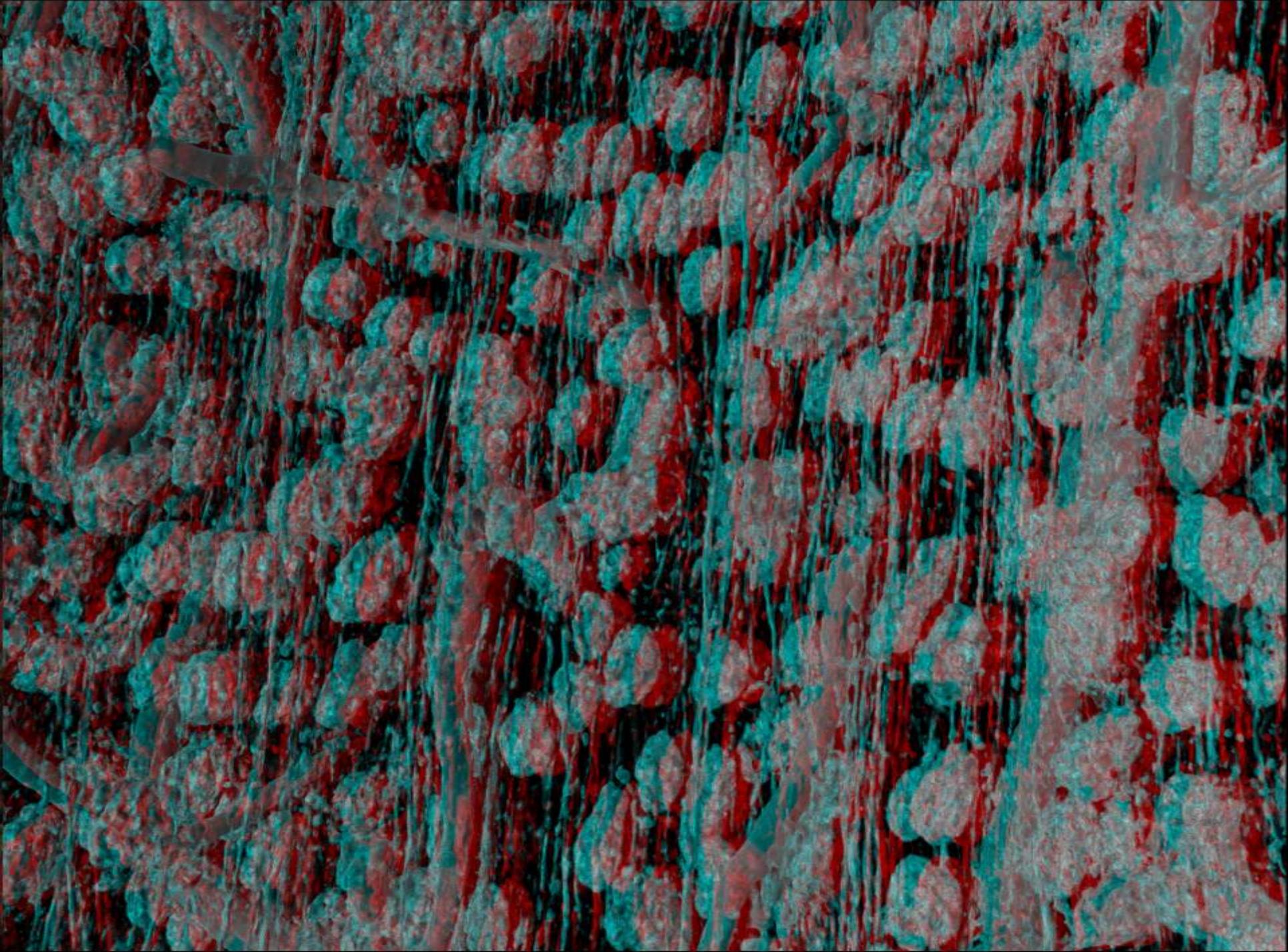
3D-визуализация фрагмента мозга по данным микронного разрешения



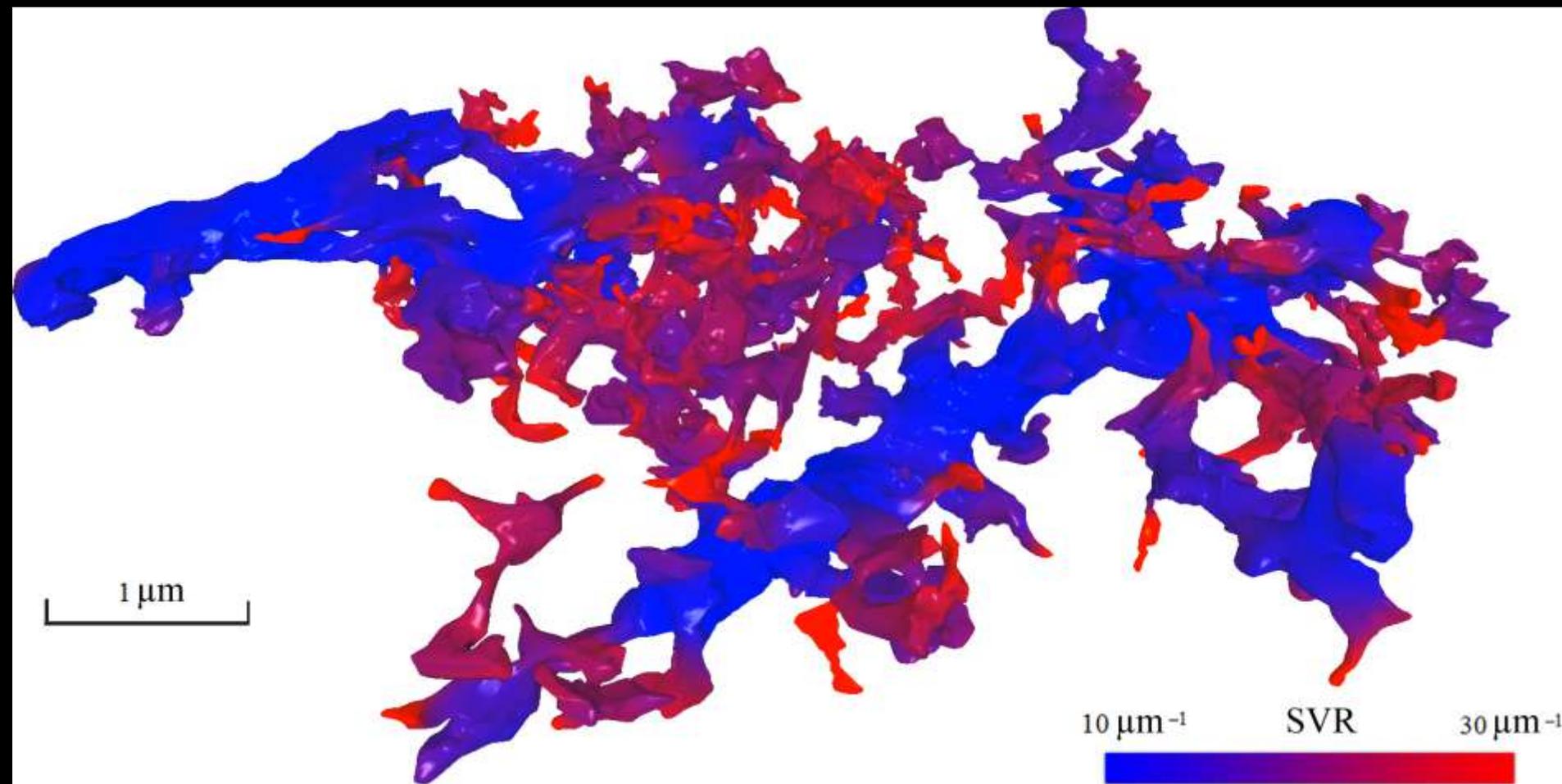
Фрагмент гиппокампа мозга мыши:
255x255x50 мкм, 512x512x102 воксела







3D-реконструкция и геометрическое исследование (морфометрия) клеток мозга



SVR (surf.-to-volume ratio) -раскраска поверхности астроцитов

► методом сфер

Оптимизация размера блока в методе блочной декомпозиции для задачи морфометрии клеток мозга

Таблица 3.1. Зависимость времени работы (в минутах) GPU в алгоритме раскраски модели астроцитов (57 406 вершин, 115 753 тр.) от размеров блока (ширина, высота)

Ширина \ Высота	1	2	4	8	16	32	64	128	256	512
1	31,87	17,43	10,07	6,26	4,48	3,18	2,46	1,99	1,66	1,48
2	16,90	9,71	5,92	4,04	3,01	2,42	1,92	1,67	1,48	1,38
4	9,34	5,81	3,84	2,84	2,29	1,87	1,66	1,50	1,39	x
8	5,50	3,71	2,70	2,21	1,81	1,59	1,48	1,39	x	x
16	3,75	2,64	2,04	1,69	1,53	1,43	1,36	1,30	x	x
32	2,40	1,88	1,51	1,36	1,24	1,15	1,12	x	x	x
64	1,69	1,42	1,21	1,10	1,04	1,01	x	x	x	x
128	1,18	0,95	0,85	0,80	0,78	x	x	x	x	x
256	0,82	0,73	0,70	0,67	0,65	x	x	x	x	x
512	0,87	0,81	0,76	x	x	x	x	x	x	x

Вычислительная трудоемкость задачи составляет $O(n^2)$. Применен блочный метод загрузки тестируемой порции треугольников поверхности биологического объекта. X-зона отказа GPU

Повышение производительности морфометрии клеток мозга при применении GPU

Таблица 3.2. Время расчёта SVR во всех вершинах модели при различных способах вычисления объёма и сравнение с реализацией вычисления SVR на GPU

Примеры моделей (Рис.6) [вершины, треугольники]	CPU _V [мин]	CPU _S [мин]	GPU _S [мин]	Ускорение GPU _S /CPU _S	Ускорение GPU _S /CPU _V
57 406 вершин, 115 753 тр.	1294,89	294,03	4,79	61,40	270,38
408 вершин, 822 тр.	5,47	1,12	0,10	11,39	55,51
364 вершин, 728 тр.	0,86	0,18	0,04	4,33	20,43
1 364 вершин, 2 724 тр.	8,06	1,62	0,17	9,77	48,59

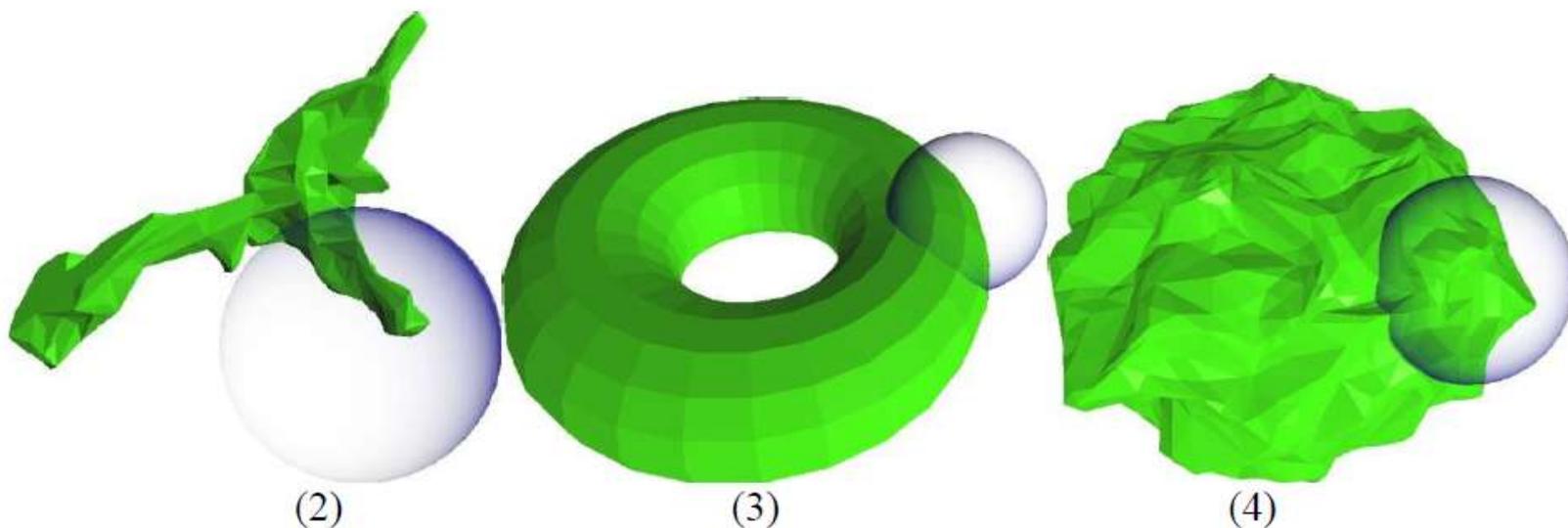
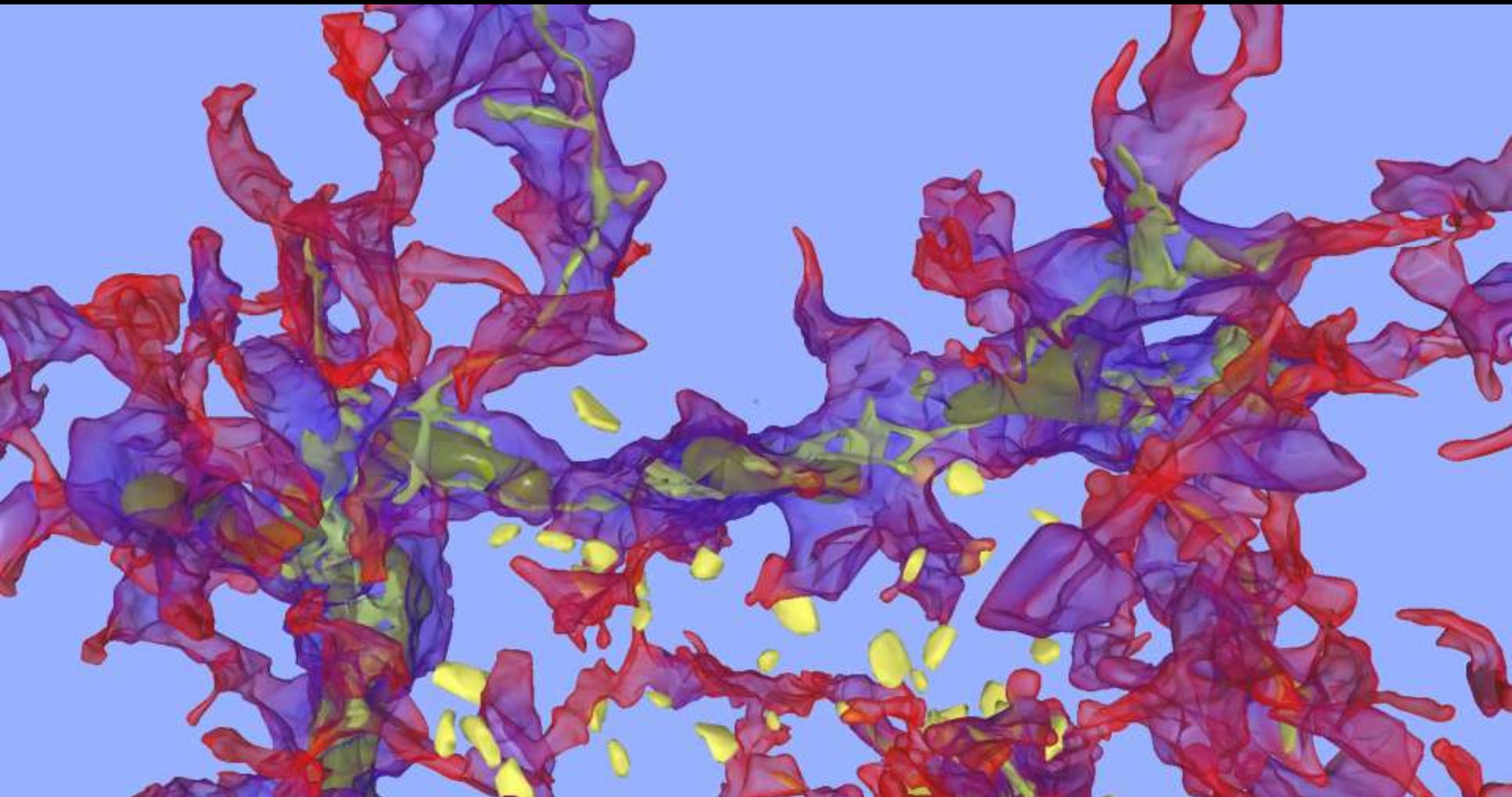
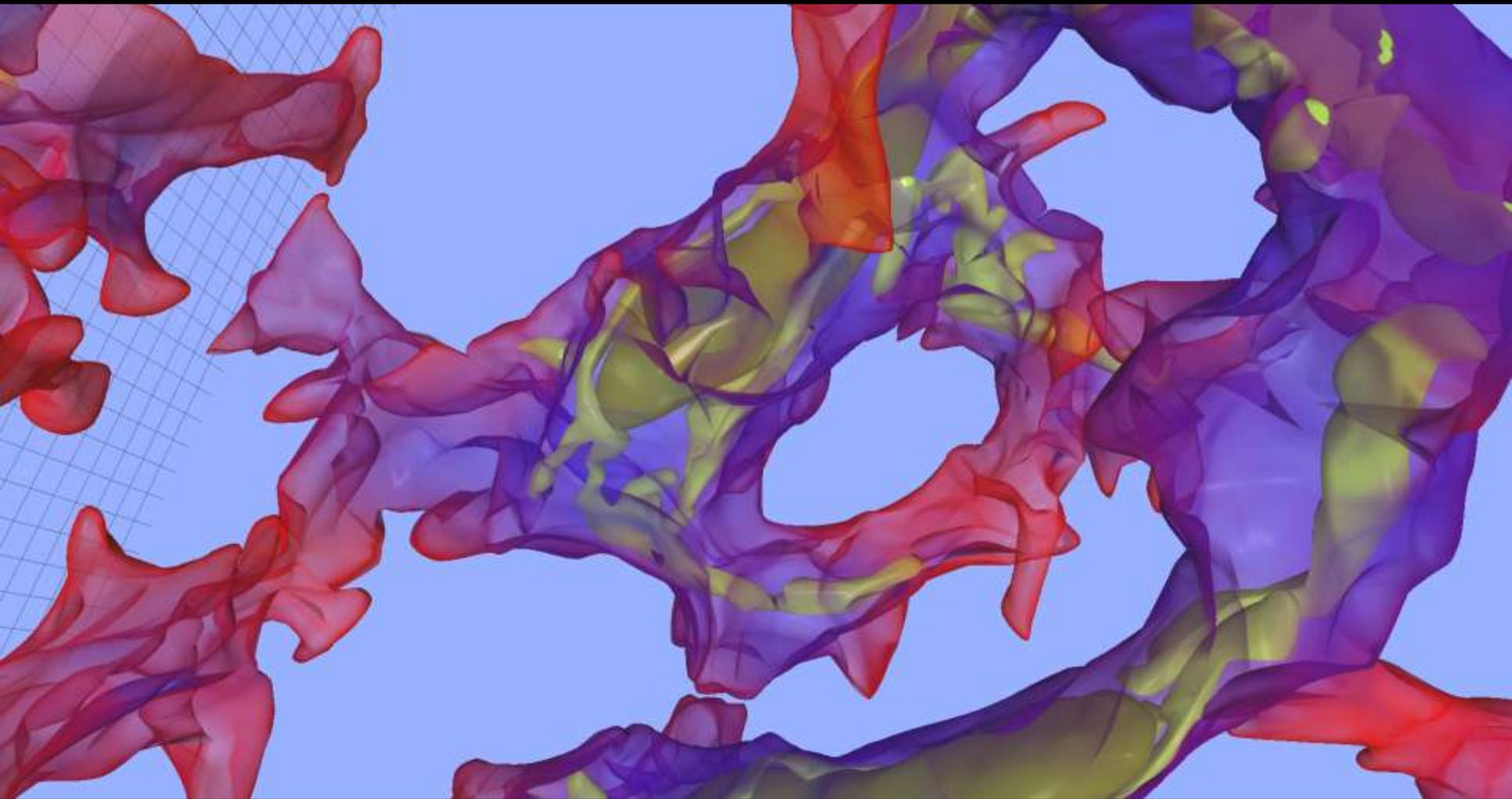


Рис.3.2. Тестовые полигональные поверхности: (1) астроцит - см. на рис.3.3. Сферы рядом с моделями показывают, какой радиус сферы использовался в эксперименте

3D-реконструкция и геометрическое исследование клеток мозга. Пример



Исследование связи SVR с кальцийсодержащими включениями



Литература

1. Гаврилов, Н.И. Турлапов, В.Е. Оценка качества визуализации медицинских и научных данных в величинах отношения пикового сигнала к шуму // Вестн. Нижегородского гос. университета. Информационные технологии. 2012.-№5(2), с. 264–273
2. Гаврилов, Н.И., Турлапов, В.Е., Патрушев, И.В., Семьянов, А.В. Локальное численное исследование морфологии 3d реконструкций биологических объектов в величинах SVR – отношения площади к объему // Вестн. Нижегородского гос. университета. Математическое моделирование и оптимальное управление – 2012.-№5(2), с. 47–56
3. Гаврилов, Н.И., Турлапов, В.Е. Подходы к оптимизации GPU-алгоритма *volume raycasting* для применения в составе виртуального анатомического стола // *Scientific Visualization*, 2012, Кв. 2, Т. 4, № 2, С.21 - 56
4. Крутских Валерий. Народный томограф. <http://unitom.ru/ru/press/65-journals/368-populyarnaya-nauka> авг. 2013

