

УДК 004.925.83

**Вяткин С.И.**

Институт автоматизи́ки и электрометрии СО РАН

**Павлов С.В.**

Винницкий национальный технический университет

**Романюк С.А.**

Винницкий национальный технический университет

## РЕЙКАСТИНГ ОБЪЕМНЫХ ДАННЫХ И ФУНКЦИОНАЛЬНО ЗАДАНЫХ ПОВЕРХНОСТЕЙ ДЛЯ МЕДИЦИНСКИХ ПРИЛОЖЕНИЙ С ПРИМЕНЕНИЕМ ГРАФИЧЕСКИХ УСКОРИТЕЛЕЙ

*Рейкастинг на основе GPU стал важным инструментом визуализации данных для медицинских изображений. Хотя метод и формирует высококачественные изображения, его основным недостатком является высокая вычислительная сложность.*

*В биомедицинской визуализации необходима визуализация 3D-наборов данных, полученных от КТ или МРТ-сканеров, в виде наглядного двумерного изображения в процессе, называемом объемным рендерингом.*

*В результате популярности и наглядности представления объемных данных появился широкий класс методов объемного рендеринга. Рейкастинг является одним из таких методов. Он позволяет производить высококачественный объемный рендеринг, но это ресурсоемкий метод, который при современных технологиях не обеспечивает интерактивность при визуализации больших наборов данных при обработке на центральном процессоре. Появление эффективных графических процессоров, доступных практически для всех современных рабочих станций, в сочетании с их высокой степенью программируемости открывает широкую область новых приложений для видеокарт. В этих приложениях используется рейкастинг, использующий внутренний параллелизм при обработке полностью независимых световых лучей, что позволяет использовать преимущества потоковой параллельной архитектуры графического процессора.*

*В статье предложен метод рейкастинга объемных данных и функционально заданных поверхностей, реализованный с применением графических ускорителей. В методе предусмотрена визуализация большого числа объемов, комплекс полупрозрачных и функционально заданных объектов, сложных полупрозрачных объемов, в том числе пересечений объемов в конструктивном твердотельном моделировании. Метод, включая стадию растеризации, реализован на CUDA, что позволяет полностью контролировать иерархию памяти, в частности доступ к высокой пропускной способности и низкое время ожидания распределенной памяти. Обеспечивается интерактивная частота кадров при одновременном отображении более ста произвольно перекрывающихся объемов.*

**Ключевые слова:** функционально заданные поверхности, объемный рендеринг, рей-кастинг, CSG, CUDA, GPU.

**Постановка проблемы.** Рейкастинг – самый распространенный подход для рендеринга объема благодаря своей гибкости при генерации изображения высокого качества [1–4]. Большой объем однородных данных, содержащихся в объемной модели, очень хорошо распараллеливается при вычислениях. Сегодня даже обычный графический процессор (GPU) способен отобразить один объем данных высокого разрешения с высокой частотой кадров с помощью программно-аппаратных шейдеров ускорителя.

Однако отображение одного однородного объема не подходит для многих приложений. Одно-

временный рендеринг нескольких объемов необходим при обработке нескольких наборов данных. В медицинской визуализации такие методы получения анатомических изображений, как компьютерная томография (КТ), магнитно-резонансная томография (МРТ), УЗИ, и методы получения функциональных изображений, например, позитронная эмиссионная томография (ПЭТ), однофотонная эмиссионная компьютерная томография, функциональная МРТ, используются для обследования пациента [5–9]. Кроме того, модели медицинских инструментов и совмещение поверхностей с объемными данными необходимы для

компьютерной хирургии. Виртуальные хирургические операции являются одними из наиболее перспективных направлений в медицине, в которых применяются трехмерная компьютерная графика и технологии виртуальной реальности. Системы виртуальной ортопедической хирургии, которые могут работать на общедоступных персональных компьютерах, становятся актуальными. Например, во время обучения общей ортопедической хирургии врачи исследуют переломы на пластических костях с помощью хирургических инструментов и имплантатов. Хирурги заинтересованы, во-первых, в экономии стоимости костей, так как синтетические кости хорошего качества стоят дорого; во-вторых, появляется возможность работать с нужными костями, которые могут быть недоступны на рынке в данный момент. Кроме того, полезной возможностью является исследование на виртуальных моделях реальных переломов, полученных из данных КТ или МРТ. Конечно, это виртуальное обучение не заменит полностью обязательного обучения на синтетических костях, но может позволить студентам выполнять начальную рутинную работу полностью в виртуальной среде, тем самым экономя расходы и время обычной подготовки. А также можно будет выполнить хирургическое планирование операции. Рендеринг нескольких объемов одновременно в композиции с поверхностями осуществить труднее, чем визуализацию отдельных объемов, так как требует обработки перекрытий объемов и смешивания цветов. Пересчет объемов в единой системе координат не желательно, так как это приводит к потере качества или увеличению требований к памяти. Обязательное требование в виртуальной хирургии – сочетание объемов с поверхностями. Функциональные модели хорошо подходят для встраивания посторонних объектов (инструментов, имплантатов и пр.) в объемные данные.

**Анализ последних исследований и публикаций.** Метод рейкастинга [1–4] (англ. raycasting – бросание лучей) основывается на вычислении пересечения лучей, испускаемых камерой через проекционный экран, и примитивов сцены. Расстровое изображение строится на основе замеров пересечения лучей с визуализируемой поверхностью в пространстве

**Постановка задачи.** Цель статьи – разработать метод рейкастинга объемных данных и функционально заданных поверхностей, реализованный с применением графических ускорителей.

**Изложение основного материала исследования.** В основе рейкастинга стоит идея испускать

лучи из «глаз» наблюдателя, один луч на пиксель, и находить самый близкий объект, который блокирует путь распространения этого луча. Используя свойства материала и эффект света в сцене, алгоритм рейкастинга может определить затенение данного объекта. Предположение в упрощении основано на том, что если поверхность размещена перед светом, то свет достигнет поверхности и не будет заблокирован или находиться в тени. Затенение поверхности вычисляется с помощью алгоритмов затенения традиционной компьютерной трёхмерной графики. Одним из преимуществ рейкастинга является способность легко обрабатывать неплоские поверхности и сплошные тела. Если математическая поверхность может быть пересечена лучом, то ее изображение может быть получено с помощью рейкастинга. Сложные объекты могут быть созданы с использованием методов моделирования сплошных тел.

Для описания поверхностей часто используются функции отклонения (второго порядка) от базовой квадрики [10]. Поверхности строятся с помощью квадрик и представляются композицией базовой квадрики и возмущений:

$$F'(x, y, z) = F(x, y, z) + \sum_{i=1}^N f_i R_i(x, y, z), \quad (1)$$

Где  $f_i$  – форм-фактор;  $R(x, y, z)$  – возмущение:

$$R_i(x, y, z) = \begin{cases} Q_i^3(x, y, z), & \text{если } Q_i(x, y, z) \geq 0, \\ 0, & \text{если } Q_i(x, y, z) < 0, \end{cases} \quad (2)$$

где  $Q(x, y, z)$  – возмущающая квадрика.

Чтобы поверхность была гладкой, степень должна быть больше двух. Это условие гарантирует непрерывность функции и её производной.

Геометрическая модель создаёт условия для конструирования объектов и их композиций различной сложности. Для этого используется множество геометрических операций  $\Phi$ , определяемое математически следующим образом [11]:

$$F_j: M^1 + M^2 + \dots + M^n \rightarrow M. \quad (3)$$

Для формирования моделей сложных объектов на базе функций возмущения используются теоретико-множественные, осуществляемые с применением булевых операций объединения и пересечения. Бинарная операция (n=2) (3) объектов  $G_1$  и  $G_2$  означает операцию  $G_3 = F_j(G_1, G_2)$  с определением:

$$f_3 = \psi(f_1(X), f_2(X)) \geq 0, \quad (4)$$

где  $\psi$  – непрерывная вещественная функция двух переменных.

В работе представлен метод, направленный на разработку недорогой альтернативы обычному

ортопедическому хирургическому обучению и прочих медицинских приложений.

Рейкастинг объемных данных. Метод позволяет визуализировать несколько объемов в комбинации с функциональными поверхностями с применением сортировки по глубине всех фрагментов сцены [12; 13; 14] (рис. 1). Для этого применяется тайловая технология визуализации [15]. Суть ее заключается в следующем. Процесс растривания разбивается на два этапа. В общем случае первый этап заканчивается, когда рассматриваемой части соответствует клетка определенного размера (тайл-tile). Примитивами промежуточного описания являются фрагменты пересечения геометрических объектов с клетками. Второй этап вычислений – обработка списка объектов, определение видимости и цвета пикселей. Применяется рейкастинг изображения со сложными пересечениями объемных данных с поверхностями. Все вычисления выполняются на графическом акселераторе (GPU). Описание состоит из простого графика сцены.



**Рис. 1. Воксельные (ткани и кости) и функциональные (имплантаты) данные описания сцены**

Внутренние узлы дерева, связанные с операциями конструктивной твердотельной геометрии (CSG), ссылаются к объекту, а также к объемной текстуре и материалу с его свойствами. Экземпляры можно создавать, ссылаясь на тот же объект и/или объемную текстуру не один раз. Алгоритм рейкастинга лучей включает обычные шаги рендеринга: преобразование объекта, растеризация и обработка фрагментов. Обработка фрагментов выполняется в параллельно-пиксельном режиме для клеток 8x8 пикселей и отслеживает прохождение соседних лучей по объемам, представленным как 3D текстуры. Для этого также используется кэш когерентности текстур. Система рендеринга проходит сцену только один раз. Реализованы все шаги в программном обеспечении CUDA,

поэтому имеется доступ ко всем промежуточным результатам конвейера, которые хранятся в общей памяти для максимальной производительности. Реализованы два отдельных вычислительных конвейера (один для обработки объектов, а другой – для обработки фрагментов). Производится сортировка списка всех соответствующих фрагментов для каждого пикселя, которая завершается после обработки всех объектов. Вычислительный конвейер объектов отвечает за геометрические преобразования и определение принадлежности объектов клеткам. Пиксельный конвейер получает фрагменты в произвольном порядке. Далее они сортируются в Z-направлении, прежде чем они могут быть использованы. Окончательная трассировка лучей и затенение требуют смешивания нескольких пересекающихся объемных объектов и эффективного накопления отдельных объемных вкладов вдоль лучей. В обычном рендеринге сцен преобладают непрозрачные объекты. Высокая производительность определяется возможностью идентификации маскированных частей сцены в начале конвейера. В отличие от классического объемного подхода, предлагаемый метод рендеринга должен растривать все поверхности и передать информацию на экране покрытия поверхностей (маскирование) до обработки пикселей. После преобразования объекта в пространство экрана для каждой клетки (тайла) 8x8 пикселей вычисляется маска. Эта простая стратегия хорошо подходит для аппаратной технологии CUDA и предпочтительнее более сложных подходов. Для метода важно, чтобы каждый фрагмент, охватываемый границей объема, вычислялся один раз с момента инициализации. Поэтому вычисляются покрытые пиксели для каждой поверхности. Блок вычислений, ответственный за клетку, состоит из 64 вычислителей, один для каждого пикселя в тайле. Первая задача для вычислителя – построить представление луча, состоящего из пересечений с поверхностями, отсортированными по глубине. Вычислитель выполняет итерацию по списку поверхностей, записывает и проверяет свой бит в маске покрытия. Если бит установлен, то он вычисляет глубину пересечения с помощью уравнения поверхности. Z-значения и идентификаторы поверхности хранятся в виде массива в быстром доступе общей памяти. Поэтому выделяется только одно 32-разрядное значение каждой записи, 16-бит – на запись Z-значения и 16-бит для идентификатора поверхности. Записи в этом массиве чередуются, чтобы избежать конфликтов. Кроме того, имеется общая память 16 КБ, которая

разделяется на 64 слота, что позволяет каждому вычислителю сохранить максимум 63 записи. Это ограничивает максимальную глубину сложности 63. Массив поддерживается в отсортированном порядке путем вставки новых значений в соответствующее положение, начиная с элементов, ближних к наблюдателю. Получается, что время, затраченное на сортировку, невелико по сравнению с последующим временем прохождения луча, даже для специально разработанного худшего сценария, в котором весь список фрагментов расположен в обратном порядке (т. е. имеет квадратичную временную сложность). Более того, благодаря пространственной когерентности для умеренно сложных сцен, операция вставки используется для многих потоков. Обоснование такого подхода состоит в том, что фрагменты, маскированные 63 более близкими слоями, будут иметь минимальное влияние на окончательное изображение и, следовательно, могут быть замаскированы. Чтобы проиллюстрировать это утверждение, протестированы аппроксимации на объектах с очень высокой глубинной сложностью. После сортировки пиксельный поток выполняет итерацию отсортированного массива в порядке от ближайшего до самого дальнего. Глубина интервала между двумя последовательными записями в массиве определяет однородный сегмент луча относительно пересекающихся объектов. Переходя по массиву, поддерживается информация о пересекающихся объектах сцены, которые являются листьями дерева конструктивной твердотельной геометрии (CSG). Если один или несколько объектов сцены пересекаются, начальная точка луча должна быть преобразована в координаты модели, образуя таким образом набор лучей пространства объекта. Для каждого пространства объекта луч должен использовать одинаковый размер шага в мировой системе координат, чтобы правильно смешивать образцы для каждого объекта сцены. Так как объекты сцены могут иметь различные размер и разрешение, используется минимальный размер шага всех объектов вдоль отрезка луча, чтобы гарантировать, отсутствие пропусков. Так как это

может привести к передискретизации разрешения объектов, обеспечивается вариант регулировки этого параметра. Пропуск пустого пространства является необходимым методом ускорения для объемной визуализации. Эффективная обработка граничной геометрии в нашем методе рендеринга позволяет пропускать внешнее и внутреннее пространство для ускорения вычислений. Это легко может быть достигнуто путем выполнения вычисления параллельно на GPU, тем самым избегая дополнительных медленных передач данных из памяти. Время вычислений для набора данных 5K составляет около одной миллисекунды на NVIDIA 470 GTX. Новые ограничивающие геометрические поверхности вставляются в глобальный список поверхностей в глобальной памяти GPU и связаны с соответствующими объектами сцены. Во время обхода луча с помощью отсортированного списка фрагментов ограничивающие геометрические поверхности позволяют одновременно пропускать внутреннее пространство на основе набора активных объектов, который обновляется всякий раз, когда фрагмент граничной геометрии обрабатывается, при добавлении или удалении определенного объекта. Использование плотно прилегающей ограничительной геометрии увеличивает частоту кадров до 25%, в частности, при передаче функций, которые делают большие части объема прозрачными. В целом на оптимизацию влияют как ядро поверхностей, так и ядро обработки пикселей.

**Выводы.** За счет последних улучшений в многоядерных моделях программирования GPU разработан новый подход к визуализации множества объемов и геометрии. CUDA позволил разработать программный конвейер рендеринга, который выполняется параллельно. Такой подход актуален для широкого спектра применений. Это увеличивает практическую применимость объемного рендеринга к очень сложным сценам, включая пересекающиеся объемы и геометрию для медицинских приложений. Как показано в статье, метод является простым в реализации, надежным и адаптируется под конкретное применение.

#### Список литературы:

1. Yagel R., Machiraju R. Data Parallel Volume Rendering Algorithms. *The Visual Computer*. 1995. № 11 (6). P. 319–338.
2. Krueger J., Westermann J. Acceleration techniques for GPU-based volume rendering. *IEEE Visualization*. 2003. P. 287–292.
3. Volume Rendering / R.A Drebin, L. Carpenter, P. Hanrahan. *Computer Graphics*. 1988. № 22 (4). P. 65–74.
4. A Fast Display Method for Volumetric Data / L. Sobeirajski, D. Cohen, A. Kaufman, R. Yagel, D. Acker. *The Visual Computer*. 1993. № 10 (2). P. 116–124.

5. Herman G.T. Image Reconstruction from Projections. The Fundamentals of Computerized Tomography. Academic Press, 1980.
6. Accelerated Volume Rendering and Tomographic Reconstruction Using Texture Mapping Hardware / B. Cabral, N. Cam, J. Foran. Proceedings 1994 Symposium on Volume Visualization. P. 91–98.
7. A reproducible evaluation of ants similarity metric performance in brain image registration / B.B. Avants, N.J. Tustison, G. Song, P.A. Cook, A. Klein and J.C. Gee. Neuroimage. 2011. № 54 (3). P. 2033–2044.
8. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease / R.J. Killiany et al. Annals of neurology. 2000. № 47 (4). P. 430–439.
9. Multiresolution volume visualization with a texture-based octree / I. Boada, I. Navaza, R. Scopigno. The Visual Computer. 2001. № 17. P. 185–197.
10. Вяткин С.И. Моделирование сложных поверхностей с применением функций возмущения. Автометрия. 2007. Т. 43. № 3. С. 40–47.
11. Вяткин С.И. Преобразования функционально заданных форм. Программные системы и вычислительные методы. 2014. № 4. С. 484–499.
12. Everitt C. Interactive Order-Independent Transparency. Technical report NVIDIA OpenGL applications engineering, 2001.
13. Mammen A. Transparency and Antialiasing Algorithms Implemented with the Virtual Pixel Maps Technique. IEEE Computer Graphics Applications, pp. 43–55, 1989.
14. Porter T. and Duff T. Compositing Digital Images. Computer Graphics. 1984. Volume 18. Number 3 July. P. 253–259.
15. Вяткин С.И. Метод рекурсивного поиска элементов изображения функционально заданных поверхностей. Автометрия. 2017. Т. 53. № 3. С. 53–57.

#### **РЕЙКАСТИНГ ОБ'ЄМНИХ ДАНИХ І ФУНКЦІОНАЛЬНО ЗАДАНИХ ПОВЕРХОНЬ ДЛЯ МЕДИЧНИХ ДОДАТКІВ З ВИКОРИСТАННЯМ ГРАФІЧНИХ ПРИСКОРЮВАЧІВ**

*Рейкастинг на основі GPU став важливим інструментом візуалізації даних для медичних зображень. Хоча метод і формує високоякісні зображення, його основним недоліком є висока обчислювальна складність.*

*У біомедичній візуалізації необхідно відображати 3D-набори даних, отриманих від КТ або МРТ-сканерів, у вигляді наочного двовимірного зображення в процесі, який називають об'ємним рендерингом.*

*За рахунок популярності й наочності подання об'ємних даних з'явився широкий клас методів об'ємного рендеринга. Рейкастинг є одним із таких методів. Він дає змогу реалізувати високоякісний об'ємний рендеринг, але це ресурсоємний метод, який за сучасних технологій не забезпечує інтерактивності під час візуалізації великих наборів даних у процесі обробки на центральному процесорі. Поява ефективних графічних процесорів, доступних практично для всіх сучасних робочих станцій, у поєднанні з їх високим ступенем програмованості відкриває широку ділянку нових додатків для відеокарт. У цих додатках використовується рейкастинг, що застосовує внутрішній паралелізм під час обробки повністю незалежних світлових променів, що дає змогу використовувати переваги потокової паралельної архітектури графічного процесора.*

*У статті запропоновано метод рейкастингу об'ємних даних і функціонально заданих поверхонь, реалізований із застосуванням графічних прискорювачів. У методі передбачена візуалізація великого числа об'єктів, комплекс напівпрозорих і функціонально заданих об'єктів, складних напівпрозорих об'єктів, у тому числі перетинів об'єктів у конструктивному твердотільному моделюванні. Метод, включаючи стадію растеризації, реалізовано на CUDA, що дає змогу повністю контролювати ієрархію пам'яті, зокрема доступ до високої пропускної здатності й низький час очікування розподіленої пам'яті.*

*Забезпечується інтерактивна частота кадрів при одночасному відображенні понад ста об'єктів, що доволіно перекриваються.*

**Ключові слова:** функціонально задані поверхні, об'ємний рендеринг, рей-кастинг, CSG, CUDA, GPU.

#### **RAY CASTING OF VOLUME DATA AND FUNCTIONALLY SURFACED SURFACES FOR MEDICAL APPLICATIONS WITH APPLICATION OF GRAPHIC ACCELERATORS**

*Medical image three-dimensional reconstruction technology has been widely used in medical diagnosis and treatment planning.*

*GPU based ray casting has become a valuable tool for the visualization of medical image data. While the method produces highquality images, its main drawback is the high computational load.*

*In biomedical imaging, it might be necessary to visualize 3D datasets obtained from CT or MRI scanners as a meaningful 2D image, in a process called volume rendering. As a result of the popularity and usefulness of volume data, a broad class of volume rendering techniques has emerged. Ray casting is one of these techniques. It allows for high quality volume rendering, but is a computationally expensive technique which, with current technology, lacks interactivity when visualizing large datasets, if processed on the CPU. The advent of efficient GPUs, available on almost every modern workstations, combined with their high degree of programmability opens up a wide field of new applications for the graphics cards. Ray casting is among these applications, exhibiting an intrinsic parallelism, in the form of completely independent light rays, which allows to take advantage of the massively parallel architecture of the GPU. This paper describes the implementation and analysis of a set of shaders which allow interactive volume rendering on the GPU by resorting to ray casting techniques.*

*A method is proposed for the ray-casting of volumetric data and functionally defined surfaces using graphic accelerators. The method provides for visualization of a large number of volumes, a complex of translucent and functionally defined objects, complex translucent volumes, including intersections of volumes in constructive solid modeling. The method, including the rasterization stage, is implemented on CUDA, which allows to fully control the memory hierarchy, in particular, access to high bandwidth and low latency of distributed memory. An interactive frame rate is provided with simultaneous display of more than 100 arbitrarily overlapping volumes.*

*The proposed method can be used for realistic visualization of medical data.*

**Key words:** *functionally defined surfaces, volumetric rendering, ray-casting, CSG, CUDA, GPU.*