

Open Rendering Benchmark: на пути к оценке систем рендеринга

За десятилетия существования фотореалистичного рендеринга не было предложено способа объективного сравнения многочисленных систем его реализации – сегодня ранжирование более 200 программных продуктов до сих пор субъективно, что мешает выбору систем, адекватных конкретным задачам. Открытый стандартный тестовый набор сцен – Open Rendering Benchmark предоставляет возможность объективной оценки существующих рендер-систем.

In this paper, an open set of 3D scenes (the so-called "Benchmark") has been created to evaluate the performance of rendering systems in various cases and, thus, to demonstrate advantages and disadvantages of the algorithms used by them in various conditions. We recreated many scenes from well-known computer graphics publications and added some industrial scenes based on our experience. Our goal is to make the coverage as complete as possible, using as few scenes and render system functionalities as possible so that such a comparison can be easily reproduced in any existing system. Over the decades of photorealistic rendering, a method of objective comparison of its numerous implementations has not been proposed. Today, the choice of more than 200 software products for photo-realistic image synthesis is still largely subjective, which does not contribute to the use of systems that are optimal for solving specific problems. At the same time in the scientific works on computer graphics, the authors are engaged in the so-called "cherry-picking". That is, the scene is very carefully selected to show the advantages of the developed method and get a positive review. But this greatly reduces the value of research in the field as a whole, since it cannot be said whether the proposed method will really work well in other scenes or not. In such areas as machine learning or computer vision of the open popular test datasets created a great many. Only in the graphics we continue to "select cherries" so that the reviewer does not see "rot". It's time to end this. We set ourselves the task of creating a common denominator for all algorithms and all render systems in the world by creating an open standard test set of scenes - Open Rendering Benchmark or "ORB".

3ds Max, Рендер-системы, V-Ray, сравнение, расчёт без допущений.
3ds Max, Render systems, V-Ray, comparison, unbiased rendering.

Vladimir Frolov (vova@frolov.pp.ru) – research scientist, Keldysh Institute of Applied Mathematics RAS and Lomonosov Moscow State University, Russian Federation, Moscow
Pavlov Denis (denis.pavlov@graphics.cs.msu.ru), Ph.D student, Lomonosov Moscow State University, Russian Federation, Moscow; Trofimov Maxim (max.trofimov@graphics.cs.msu.ru), Kazbeev Pavel (silentprayercg@gmail.com), 3D Artist, Lomonosov Moscow State University, Russian Federation, Moscow; Vladimir Galaktionov (vlgal@gin.keldysh.ru) – head of computer graphics department, Keldysh Institute of Applied Mathematics RAS, Russian Federation, Moscow.

Полноценная симуляция реальности – это лишь вопрос времени, причем большого машинного времени. Фотореалистичный рендеринг – это долго. Конечно, фундаментальные проблемы визуализации заключены не только в скорости, но именно она может быть объективно измерена в отличие, например, от глубины проработки физических явлений и субъективной оценки реалистичности сформированного изображения. Удивительно, но за десятилетия существования фотореалистичного рендеринга так и не было предложено объективного способа оценки скорости обработки, даже несмотря на то, что время расчета было и остается большой проблемой – критичным параметром для всех систем рендеринга.

Для имитации реальности в VR-шлеме необходимо порядка 90 кадров в секунду, что пока недостижимо для фотореалистичного рендеринга – в индустрии для отрисовки только одного кадра может требовать десятки часов счета. Сегодня имеется более 200 различных программных продуктов, предназначенных для фотореалистичного синтеза изображений, но отсутствуют методы их объективного сравнения, что не позволяет выбрать наилучшую систему для каждого конкретного случая. Одновременно с этим, чтобы показать преимущества именно своей системы, разработчики тщательно подбирают сцены ("cherry-picking"), позволяющие в наилучшем свете показать преимущества используемых в их продукте методов. Однако, такие ухищрения сильно снижают ценность исследований в

области в целом, поскольку не позволяют сказать, действительно ли предложенный метод будет хорошо работать на других сценах. В машинном обучении и в компьютерном зрении сегодня имеется ряд открытых тестовых наборов для объективной оценки работоспособности используемых моделей и только в фотореалистичном рендеринге разработчик продолжат заниматься «показухой», призванной скрыть недостатки системы.

Традиционно сравнения систем проводится на нескольких сценах, выбранных случайно, что не позволяет проанализировать характеристики устойчивости рендеринга к различным сценариям: массивная геометрия, оптически сложные материалы, сложное освещение, наличие трудновычисляемых феноменов освещенности. Время расчета, которое при этом указывается в подобных сравнениях ровным счетом ничего не демонстрирует, поскольку качество изображения оценивается «на глаз», хотя едва заметная разница в шуме на изображении может означать разницу в скорости в разы. Это особенно верно при приближении к «финальному качеству», когда изображение становится максимально чистым, поскольку качество в физически корректных рендерах, как правило, пропорционально квадрату времени расчета.

В работе [1] впервые был предложен более точный анализ производительности для нескольких систем под Autodesk 3ds Max на семи различных упрощенных сценариях. Для решения проблемы отличающихся эталонов предлагалось для каждой системы использовать свой собственный эталон, получаемый именно этой системой за достаточно большое время и не содержащий шума. При этом считается, что изображения эталонов визуально совпадают – рендер-системы отличаются лишь в «реализации отдельных незначительных моментов». Общий уровень яркости, геометрическая форма объектов на изображениях и характер освещения должны совпадать. Таким образом, в работе [1] фактически измеряется то, насколько быстро алгоритм/реализация/рендер-система сходится к своему собственному эталону.

Хотя сцен при таком подходе было существенно больше, чем в существующих ранее сравнениях, данный набор вряд ли можно считать объективным. Кроме того, среди сцен не было задействовано ни одной из популярных в научном сообществе сцен (например, L1.4, L1.5, L1.6, L2.2), применяемых для выполнения хоть какого-либо сравнения. Все это делает такое сравнение не информативным для специалистов, не позволяющим даже приблизительно сопоставить существующие реализации.

Сравнение систем рендеринга освещения по многим причинам является нетривиальной задачей: в современном фотореалистичном рендеринге не существует стандартов – все рендер-системы вычисляют в пределах разные эталонные изображения; коммерческие продукты зачастую реализуют множество «физически некорректных» функций, причем, каждая система делает это по-своему, например по-разному используя карты нормалей (bump mapping); отсутствует единая метрика качества для изображений, определяющая качество изображения.

0. Подготовка тестовых сцен

Для выполнения всестороннего сравнения систем, в различных условиях воспроизводящих принципиально разные сцены необходим специальный набор тестовых сцен, однако на пути создания такого множества имеется ряд проблем.

Первая проблема – отсутствие стандарта, согласно которому интерпретируются параметры моделей и ведутся вычисления внутри физически корректных рендер-систем. Например, в некоторых из них могут быть реализованы математические модели материалов или источники света, отсутствующие в других. Более того, даже имея идентичный набор моделей и параметров, можно получить отличающиеся изображения, поскольку разные системы могут по-разному интерпретировать данные и параметры

моделей. Простейший пример – степень “матовости” отражения (glossiness), записанная в текстуре. Этот параметр, задаваемый в текстуре от 0 до 1 при конвертации в модель Фонга должен отобразиться на интервал от 1 до бесконечности (0 переходит в единицу, а 1 в бесконечность). Но таких отображений существует бесконечно много, и коммерческие рендер-системы не описывают, какое же именно отображение они используют.

Для минимизации различий в изображениях и обеспечения для каждой рендер-системы возможности в принципе отобразить эти сцены вводится набор ограничений, позволяющих минимизировать используемый набор функций. Выбор таких ограничений – это компромисс между упрощением для исследователей и пользой полученных результатов для индустрии. Для иллюстрации можно привести характерные примеры.

Карты нормалей существенно повышают реалистичность и используются повсеместно. Такие алгоритмы как фотонные карты и однонаправленный Kelemen Metropolis Light Transport теряют свои преимущества при наличии микрорельефа, что снижает их практическую ценность (отчасти именно поэтому они редко используются в индустрии). Поэтому системы и алгоритмы, которые плохо справляются с микрорельефом, нужно «штрафовать» при тестировании, а значит, эта функциональность должна быть включена. Тем не менее, карты нормалей не следует использовать слишком часто – они представляют собой отдельную функциональность рендера и, строго говоря, нарушают физическую корректность интерпретации сцены, поскольку изменяют нормаль к поверхности, не меняя при этом геометрию поверхности.

Повсеместно используются материалы, смешивающие несколько элементарных компонентов. Это увеличивает размерность пространства интегрирования, поскольку интегратору необходимо делать выбор между компонентами и использовать под это еще одно случайное число, еще одну размерность. Запрет такой функциональности был бы слишком сильным упрощением. Однако, полноценное дерево компонентов выглядит чрезмерно сложным для оценки производительности. Поэтому имеет смысл ограничить такое смешивание двумя компонентами.

Транслюценция (просвечивание) для реалистичной имитации сцен с растительностью. С точки зрения производительности заменять просвечивающие материалы полностью диффузными было бы неправильно – это существенно меняет траектории путей света в сцене

Вторая проблема заключается в составлении объективного набора сцен, наиболее полно покрывающего как типичные случаи из индустрии, так и известные проблемы в науке. Решением может быть расширение подхода из работы [1] в сторону большего внимания сценам, часто используемым в научных работах и демонстрирующим известные проблемы в области интегрирования освещения (light transport challenges). Такие сцены можно группировать по номерам – каждому номеру соответствует некоторая группа сценариев из индустрии либо некоторая известная проблема, обсуждаемая в научных работах по компьютерной графике.

1. Cornell Box с несколькими вариациями. Геометрически простые сцены, содержащие тем не менее почти все основные эффекты: шумное первичное освещение и мягкие тени, зеркальные блики от источника освещения, отраженные каустики.
 - 1.1 Базовый вариант, содержащий поровну все эффекты.
 - 1.2 Базовый вариант + зеркало, создающее широкий каустик, плюс SDS каустик, видимый в отражении. Тривиальное изменение сцены, кардинально меняющее, тем не менее, характер освещения.
 - 1.3 Вариант, ориентированный на многократные глянцевые отражения.

- 1.4 Оригинальный Cornell Box Эрика Вича, используемый им для демонстрации Metropolis Light Transport. Ярко выраженные трудновычислимые феномены освещенности. SDS каустики присутствуют в небольшом количестве (в стеклянном яйце).
2. Сильное вторичное диффузное освещение и геометрия средней тяжести.
 - 2.1 *Dabrovic Sponza* и вторичный свет от направленного источника. Часто используется в статьях по компьютерной графике, однако недостаток оригинальной сцены в том, что все текстуры в ней однотонны и вторичное освещение получается в значительной степени равномерным. Поэтому практически любой сколь угодно плохой алгоритм глобального освещения показывает на этой сцене высокую точность. По этой причине *данную сцену следует исключить из тестового набора*.
 - 2.2 *Crytek Sponza* – усложненный вариант предыдущей сцены, с более детальной геометрией, треугольниками с текстурой непрозрачности на листьях и (opacity для имитации растительности) и разноцветными шторами. В различных демонстрационных программах и статьях шторы (как и растительность) могут частично пропускать свет или полностью отражать его. Последний вариант более распространен.
 - 2.3 *San Miguel* – тяжелая геометрическая сцена с растительностью, пропускающей свет. По-прежнему яркое солнце и сильное вторичное освещение. Более сложная с геометрической точки зрения (частично из-за opacity, частично просто из-за более сложной геометрии).
 - 2.4 Ночной вариант сцены *San Miguel*. Несколько источников света, более сложное первичное освещение.
3. Сцены на сложную геометрию и простые по освещению.
4. Тестирование освещения для простой геометрии и сильно неравномерном первичном освещении.
5. Тест на небесные порталы – источники света в окнах, имитирующие рассеянный свет с улицы. Данный сценарий проверяет систему на наличие реализации многократной выборки по значимости (Multiple Importance Sampling), поскольку при наличии больших площадных источников явная стратегия сэмплирования (теневые лучи) становится мало эффективной.
6. Тесты на вторичное освещение от солнца, образованное относительно узкой, но яркой полоской света.
7. Трудновычислимые феномены освещенности (light transport challenge), вызванные многократными отражениями от поверхностей с различными свойствами.
8. Трудновычислимые феномены освещенности (light transport challenge), вызванные SDS каустиками (бассейн с водой).
9. Трудновычислимые феномены освещенности (light transport challenge), вызванные отражениями от глянцевых поверхностей в сцене, содержащей карты нормалей. Данный вид сцен может представлять проблему для однонаправленных алгоритмов на основе Марковских цепей, работающих в первичном пространстве путей.
10. Трудновычислимые феномены освещенности, оригинальная сцена Эрика Вича с приоткрытой дверью и еще одна сцена, ее вариация с глянцевыми отражениями и

некоторые менее известные сцены, проверяющие трудновычислимые феномены освещенности.

11. Сцена с самосветящимися поверхностями.

12. Тест на большое количество источников света.

13. Известная по курсу SIGGRAPH 2014 сцена красной кухни на трудновычислимые феномены освещенности, содержащая в себе элемент “зеркального коридора”.

С учетом данных рекомендаций был создан специальный набор сцен (бенчмарк) Open Rendering Benchmark (ORB), позволяющий оценить производительность рендеринга в различных ситуациях и, таким образом, оценить возможности рендер-систем и алгоритмов в различных условиях. Данный бенчмарк обладает максимальным покрытием при минимуме сцен, способных вызвать затруднение при визуализации у какой-либо из более двух сотен рендер-систем.

На некоторых из созданных сцен (рис. 1) было проведено пилотное сравнение среди четырех популярных продуктов одной из наиболее популярных систем 3D-моделирования, анимации и визуализации 3D Studio Max: V-Ray

(www.chaosgroup.com/vray/3ds-max), Corona (corona-renderer.com/), Octane (home.otov.com/render/octane-render) и Hydra Renderer (<http://www.raytracing.ru/>).

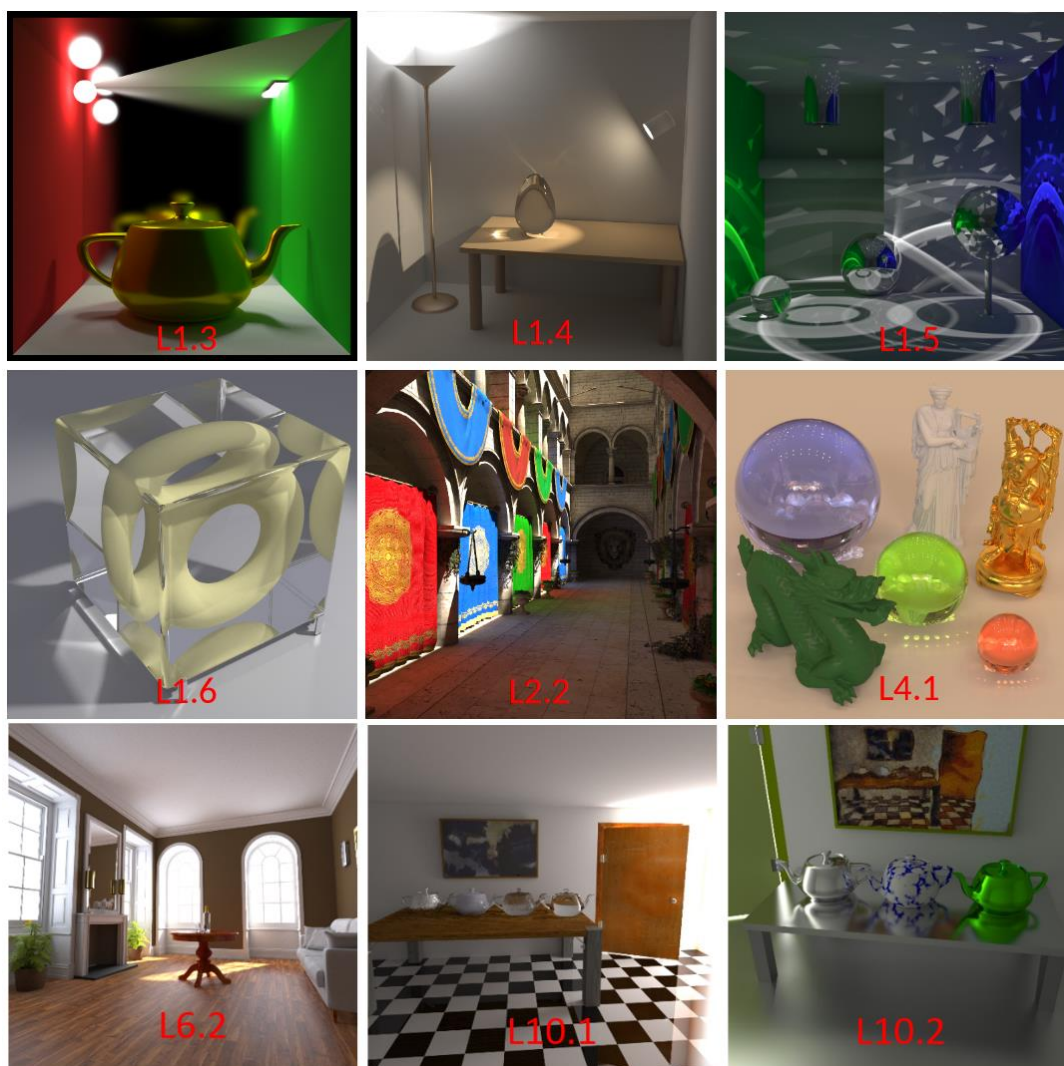


Рис. 1. Изображения некоторых сцен из созданного набора

0.Методика тестирования

Очень часто сравнения рендер-систем выполняются без понимания предметной области и не позволяют сказать ничего полезного об используемом программном обеспечении кроме того, что на данной конкретной сцене один продукт немного быстрее другого (хотя не очень понятно, на сколько именно).

Особенно популярные в России продукты “Corona” и “V-Ray” являются адептами “старой школы” – они работают на CPU и для ускорения используют так называемые алгоритмы со смещенным решением, в которых нет акцент на численную точность и они не обладают способностью достигать правильного решения в пределе, что можно считать недостатком. С другой стороны, типичный пользователь рендер-системы в Autodesk 3ds Max беспокоится не столько о правильности решения, сколько о реалистичности и красоте. Однако здесь есть нюанс: смещенные алгоритмы зачастую дают артефакты, неприемлемые для пользователя.

Системы Octane и Hydra ориентированы на расчет без допущений (то есть просто честно оценивают интеграл методами Монте-Карло) и работают с GPU. Кроме того, они используют методы рендеринга на основе Марковских цепей. Программа Hydra использует алгоритм Multiplexed MLT (MMLT) [2]) и базируется на OpenCL, а Octane PMC/ERPT [3] и CUDA. Octane сегодня одно из самых дорогих и популярных коммерческих решений в мире, а в открытой системе Hydra Renderer применяются алгоритмы расчета глобального освещения, целиком реализованные на GPU через OpenCL. Отличительной особенностью является использование двунаправленной трассировки путей и двунаправленного алгоритма MMLT. Система позиционируется как ПО, соединяющее промышленность с научными разработками в области рендеринга и параллельных вычислений на GPU.

Серьезная проблема тестирования – отсутствие общепризнанных эталонов. Различные программные системы в действительности вычисляют различные изображения. Отличия вызваны отсутствием общего стандарта для фотореалистичного рендеринга и огромным количеством деталей в реализации. Более того, даже одна и та же система может по-разному визуализировать эталонные изображения в зависимости от разных алгоритмов. Это может быть как следствием ошибок в реализации (которые, вообще говоря, не всегда имеет смысл учитывать, если в тестировании нас интересует только скорость интегрирования), так и некоторыми объективными факторами – невозможность расчета определенных феноменов в одном из алгоритмов либо намеренное упрощение вычислений (например, “clamping”, когда выбросы ограничиваются некоторым максимальным значением).

При создании ORB был выбран компромиссный критерий оценки:

- для различных систем и алгоритмов требуется *лишь визуальное совпадения эталонов*. Форма освещения и приблизительный уровень яркости должны совпадать. Допускаются незначительные отличия, не влияющие на суть вычисляемых феноменов освещенности;
- *для каждой системы и каждого алгоритма* вычисляется свой эталон, получающийся при достаточно большом времени расчета но именно *этим алгоритмом и именно в этой системе*. Единственно исключение – для алгоритмов со смещенной оценкой в качестве эталона выбирается Path Tracing (но в той же самой системе);
- если для какой-либо системы первый критерий недостижим для разных алгоритмов, то за эталон принимается трассировка путей (Path Tracing), реализованная в этой системе. Алгоритм Path Tracing – базовый и реализован везде.

Для оценки ошибки и последующего вычисления баллов используется известная метрика MSE (Mean Squared Error, квадрат ошибки – среднеквадратичное отклонение определяемой величины от ее математического ожидания, рис. 4). Данная метрика обладает линейной зависимостью от времени расчета для систем с расчётом без допущений. Таким образом, чтобы вдвое уменьшить MSE необходимо вдвое дольше рассчитывать изображение. Для рендер-систем на GPU была использована видеокарта Nvidia GTX2070, а для CPU рендер-систем аналогичный по стоимости процессор Intel Xeon E5 2690v2 3ГГц (10 ядер, 20 потоков).

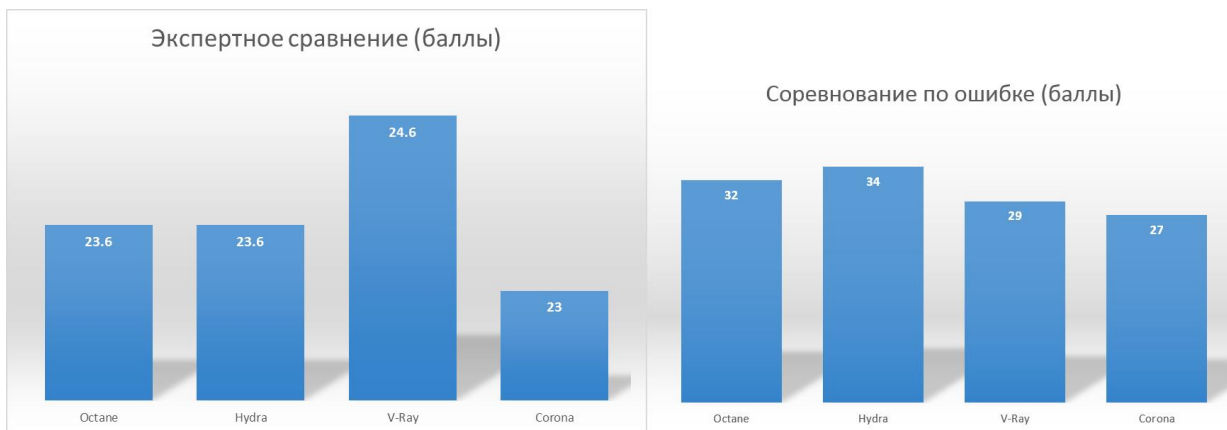


Рис. 2. Экспертное сравнение (слева) и метрике MSE – соревнование по точности/ошибке (справа)

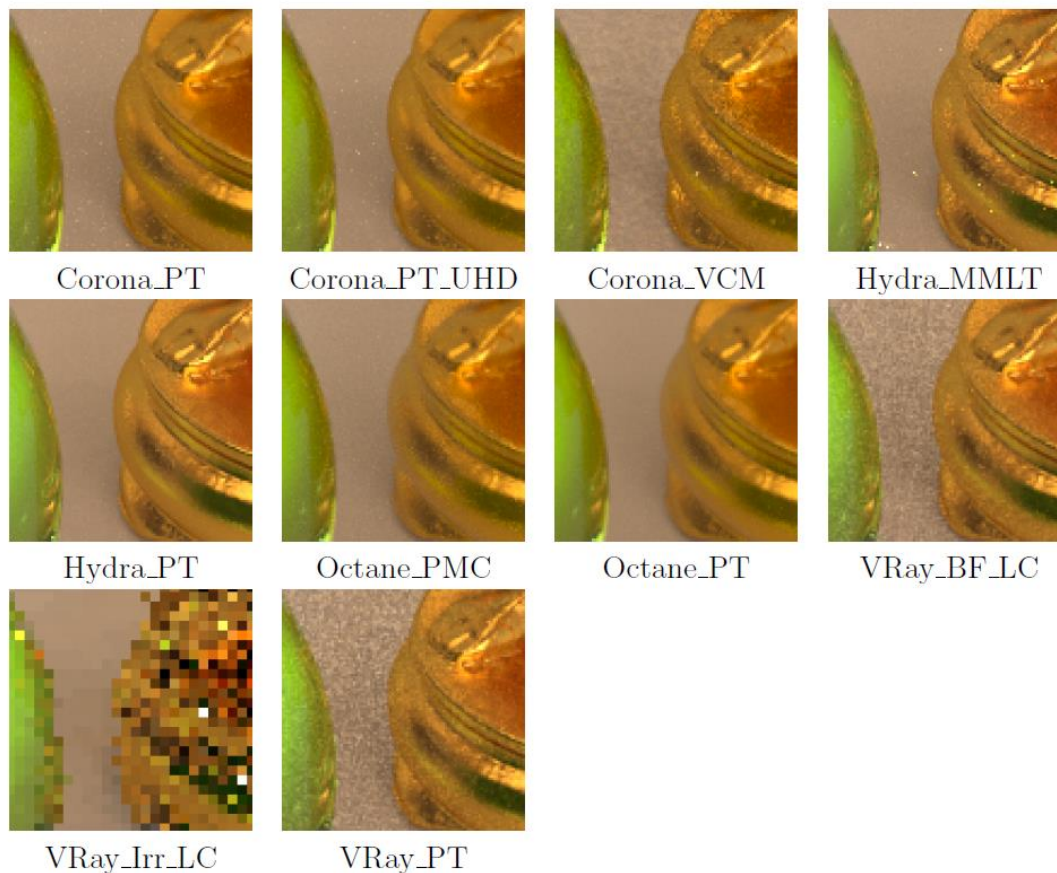


Рис. 3. Пример сопоставления рендер-систем и их алгоритмов на сцене ORB L4.1

Поскольку результаты численного сравнения по метрике MSE не всегда наглядны для 3D художников, поэтому для иллюстрации возможностей ORB для индустрии было использовано еще и экспертное сравнение. Хотя экспертная оценка и субъективна, но в ряде случаев экспертам важна не точность расчета, а такой субъективный параметр, как реалистичность изображения, отсутствие шумов и артефактов.

В экспертном тестировании приняли участие десять профессиональных 3D-художников, каждому предлагалось оценить 17 сопоставлений отдельных фрагментов изображений и провести ранжирование по местам – 1, 2 и 3. За первое место – 3 балла, за второе 2, а за третье 1 (рис. 3). Максимум баллов = $3 \cdot 17 = 51$. При этом для каждой системы выбирался только наиболее удачный алгоритм. Эксперт мог поставить ноль баллов системе, если считал, что изображение некорректно либо результат, “не тянет” на призовое место. Результаты экспертного ранжирования приведены на рис. 2 слева.

Для сравнения результатов по MSE в один ряд с экспертным мнение использовалась соревновательная форма, в которой единственным “экспертом” была метрика MSE, а баллы ставились тем же образом от 3 за первое место до 1 или 0 за последнее (рис. 2 справа).

$$MSE = \frac{1}{W * H} \sum_{i=1}^{W*H} ((R_i - R_i^{ref})^2 + (G_i - G_i^{ref})^2 + (B_i - B_i^{ref})^2)$$

Рис. 4. Вычисление MSE между измеряемым изображением (R,G,B) и эталоном (R^{REF},G^{REF},B^{REF}), получаемым за большое время. W — размер изображения по горизонтали. H — по вертикали.

0. Результаты

Результаты сравнения (пример на рис.3) оказались достаточно неожиданными. С одной стороны, по метрике MSE победили системы с расчетом без допущений на GPU (Hydra и Ostone), что, в принципе, ожидаемо – эти системы ориентированы на точный расчет. С другой стороны, система V-Ray победила в экспертном сравнении. Однако разница между победителем и проигравшим в экспертном сравнении незначительна (1,5 балла), а по метрике MSE – 7 баллов. Измерения MSE – более чувствительный способ сравнения и более объективный, позволяющие непосредственно измерить ошибку.

Неожиданным оказалось также, что эксперты не заметили большой разницы в производительности рендер-систем – благодаря применению разных подходов, алгоритмов и вычислителей (CPU и GPU) все продукты, за исключением отдельных сцен, достигают паритета. Из этого можно сделать как минимум два вывода.

Первый – борьба между разными подходами, методами и системами рендеринга еще далека от завершения и нет явного лидера. А если, так то без ущерба для производительности можно использовать бесплатное открытое решение.

Второй – скорость расчета без допущений на GPU почти сравнялась со скоростью “приближенных” методов на CPU, использующих смещенную или состоятельную оценку. Это означает, что за исключением небольшого количества случаев (вроде искусственного примера L1.5, рис. 1) “приближенные” методы теряют свою актуальность в фотореалистичном рендеринге. Кроме этого методы расчета без допущений, как правило, универсальней и проще в поддержке (добавление новых материалов, источников и т. д.). Расчет без допущений делает осмысленным усложнение моделей взаимодействия света с веществом для все большего погружения рендер-систем в физику явлений и углубления моделирования вплоть до квантового уровня.

Сравнение коммерческой и открытой системы на GPU показывает – существенной разницы в производительности между ними нет. В одних сценариях побеждает открытая Hydra, а в других Octane. Таким образом, для фотореалистичного рендеринга имеется, как минимум, одна открытая рендер-система, сравнимая по производительности с коммерческими. Благодаря ORB можно показать, что закрытые разработки не показывают преимуществ над открытой реализацией хорошо известных методов.

Сравнивая Octane и Hydra, можно косвенно сравнить и технологии CUDA и OpenCL, поскольку – на сцене с тяжелой геометрией, но простым освещением (сцена L2.2, рис. 1, 2) ощутимой разницы в производительности трассировки лучей нет, поскольку здесь скорость интегрирования будет зависеть исключительно от скорости трассировки лучей. Следовательно, налицо пример, когда закрытым коммерческим решениям, использующим аппаратные особенности конкретных GPU вполне без ущерба производительности можно предпочесть открытую и аппаратно-независимую технологию OpenCL.

Дальнейшее развитие бенчмарка ORB будет продолжаться в направлении расширения набора сцен с целью создания тестовой платформы, позволяющей провести максимально объективное сравнение современных рендер-систем. Открытость исходных кодов Hydra Renderer позволяет всем заинтересованным исследователям воспроизвести результаты тестирования и самостоятельно сопоставить изображения, генерируемые различными коммерческими и бесплатными рендер-системами.

Литература

1. В. А. Фролов., В. А. Галактионов., М. А. Трофимов. Сравнение промышленных систем расчета освещения (по состоянию на 2014 год) // Приволжский научный журнал. – 2014 - №4, с. 79-85.
2. Toshiya Nachisuka, Anton S. Kaplanyan, and Carsten Dachsbacher. 2014. Multiplexed metropolis light transport. // ACM Trans. Graph. – 2014. Vol. 33, 4, Article 100, 10p. DOI: <https://doi.org/10.1145/2601097.2601138>
3. Yu-Chi Lai, Shao Hua Fan, Stephen Chenney, and Charcle Dyer. Photorealistic image rendering with population monte carlo energy redistribution. // Proceedings of the 18th Eurographics conference on Rendering Techniques (EGSR'07, 2007). - Eurographics Association: Aire-la-Ville, Switzerland, - p.287-295. DOI: <http://dx.doi.org/10.2312/EGWR/EGSR07/287-295>
4. Ray Tracing Systems. Hydra Renderer. Отечественная рендер-система на GPU с открытыми исходными кодами, создаваемая при поддержке института Прикладной математики имени М. В. Келдыша РАН и лаборатории компьютерной графики и мультимедиа ВМК МГУ. URL = <http://www.raytracing.ru/>

Владимир Фролов (vova@frolov.pp.ru), Владимир Галактионов (vlgal@gin.keldysh.ru) – сотрудники, ИПМ им. М. В. Келдыша РАН, ВМК МГУ им. М.В. Ломоносова (Москва); Денис Павлов (denis.pavlov@graphics.cs.msu.ru) – аспирант, ВМК МГУ им. М.В. Ломоносова (Москва); Максим Трофимов (max.trofimov@graphics.cs.msu.ru), Павел Казбеев (silentprayercg@gmail.com) – 3D-художники, ВМК МГУ им. М.В. Ломоносова (Москва).