

От Монте-Карло К Квази-Монте-Карло

Кирилл А. Дмитриев

Институт Прикладной Математики им. М.В.Келдыша РАН

Москва, Россия

Аннотация

В последнее время исследователи в области генерации физически аккуратных изображений все чаще используют метод квази-Монте-Карло. Это зачастую позволяет достичь более высокой скорости вычислений, чем при применении метода Монте-Карло [1, 2]. Применение метода квази-Монте-Карло в алгоритмах трассировки фотонов имеет ряд интересных особенностей, некоторые из которых рассмотрены в этой статье. Рассмотрен вопрос понижения конструктивной размерности алгоритма, что позволяет повысить эффективность квази-Монте-Карло. Также решена задача моделирования источников света и рассеивающих свойств поверхностей со сложными функциями яркости и рассеивающей способности соответственно. Традиционно в алгоритмах Монте-Карло для этого применяется метод отказов. Применение этого метода в алгоритмах квази-Монте-Карло ведет к потере производительности [3], поэтому предлагается применять метод обратных функций. В статье выводятся формулы для применения этого метода к функциям яркости и отражающей способности общего вида, заданным в табличном виде.

Ключевые слова: глобальное освещение, квази-Монте-Карло, физически аккуратные изображения.

1. ВВЕДЕНИЕ

Задача автоматического построения изображений по заданным исходным данным является весьма актуальной в настоящее время. Исходные данные представляют из себя описание геометрии, рассеивающих свойств поверхностей, и источников света некоторой виртуальной сцены. Задача ЭВМ состоит в том, чтобы автоматически построить изображение заданной сцены с заданного положения наблюдения. Для этого необходимо найти яркость по направлению к наблюдателю каждой видимой им точки поверхности.

Модель распространения света, принятая в большинстве методов генерации физически аккуратных изображений, описывается уравнением рендеринга [4]. Это уравнение часто записывают следующим образом:

$$L(x, w_{out}) = L_e(x, w_{out}) + \tau(L(x, w_{out})) \quad (1)$$

где $L_e(x, w_{out})$ - излученная, $L(x, w_{out})$ - искомая (полная) яркость поверхности в точке x по направлению w , а τ - интегральный оператор, который зависит от геометрии сцены и ДФРС (Двулучевой Функции Рассеивающей Способности) поверхности в точке x . Уравнение (1) описывает интуитивно понятный факт, что яркость поверхности в точке x в направлении w_{out} равна сумме яркостей излученного света $L_e(x, w_{out})$ и света, пришедшего в точку x со всех направлений и отраженного в соответствии с ДФРС поверхности в направлении w_{out} .

Уравнение (1) является известным интегральным уравнением Фредгольма второго рода. В общем виде оно не решается аналитически. Для изучения свойств этого уравнения его записывают в виде бесконечных серий Ньюмана [2]. Для этого вместо $L(x, w_{out})$ внутри интегрального оператора в правой части (1) подставляют выражение $L_e(x, w_{out}) + \tau(L(x, w_{out}))$, которое следует опять же из (1). Воспользовавшись линейностью интегрального оператора, раскрывают скобки и получают:

$$L(x, w_{out}) = L_e(x, w_{out}) + \tau(L_e(x, w_{out})) + \tau^2(L(x, w_{out}))$$

Такую подстановку можно произвести бесконечное количество раз, что в конце концов приводит к исчезновению члена $L(x, w_{out})$ из правой части выражения:

$$L(x, w_{out}) = \sum_{i=0}^{\infty} \tau^i(L_e(x, w_{out})) \quad (2)$$

где под $\tau^0(L_e(x, w_{out}))$ понимается просто $L_e(x, w_{out})$. Каждое слагаемое интегрального ряда в правой части (2) имеет интуитивно понятный физический смысл. $\tau^1(L_e(x, w_{out}))$ представляет из себя отраженную энергию, пришедшую напрямую от источника света), $\tau^2(L_e(x, w_{out}))$ - отраженную энергию, пришедшую после одного переотражения от другой поверхности, $\tau^3(L_e(x, w_{out}))$ - после двух переотражений и т.д.

Так как размерность интегралов в правой части (2) в общем случае не ограничена, классические методы интегрирования (напр. прямоугольников, Симпсона) неэффективны - скорость их сходимости падает слишком быстро при возрастании размерности интеграла. Поэтому стандартным подходом вычисления (2) является метод Монте-Карло, скорость сходимости которого от размерности интеграла не зависит.

2. ИНТЕГРИРОВАНИЕ МОНТЕ-КАРЛО

Для интегрирования с помощью метода Монте-Карло необходимо придумать некоторую случайную величину η , мат. ожидание которой равно значению искомого интеграла. Тогда это значение может быть оценено с любой наперед заданной точностью с помощью статистического среднего величины η по N испытаниям, где N достаточно велико.

Пусть, например, требуется вычислить интеграл: $\int_{I^s} f(x) dx$,

где I^s - единичный s -мерный куб. Пусть областью значений случайной величины η является I^s и ее плотность распределения равна $p_\eta(x)$. Легко показать, что мат. ожидание

случайной величины $\frac{f(\eta)}{p_\eta(\eta)}$ равно искомому интегралу.

Действительно, по определению мат. ожидания:

$$M\left(\frac{f(\eta)}{p_\eta(\eta)}\right) = \int_{I^s} \frac{f(x)}{p_\eta(x)} p_\eta(x) dx = \int_{I^s} f(x)$$

Таким образом, искомый интеграл может быть оценен как статистическое среднее величины $\frac{f(\eta)}{p_\eta(\eta)}$ по N испытаниям:

$$\int_{I^s} f(x) dx \approx \frac{1}{N} \sum_{i=0}^{N-1} \frac{f(\eta_i)}{p_\eta(\eta_i)} \quad (3)$$

где η_i – значение величины η в i -м испытании. Заметим, что если в качестве η используется равномерно распределенная в I^s случайная величина, то оценка (3) является просто средним арифметическим значений функции в случайно выбранных точках.

3. ИНТЕГРИРОВАНИЕ КВАЗИ-МОНТЕ-КАРЛО

Известно, что случайность точек η_i не является необходимым условием сходимости суммы в правой части (3) к искомому интегралу. Для этого достаточно, чтобы эти точки были распределены с плотностью $p_\eta(\eta_i)$. Поэтому был построен целый ряд неслучайных последовательностей точек, позволяющий производить интегрирование в соответствии с (3). Такие последовательности получили название квазислучайных. Обычно строятся равномерно распределенные последовательности точек, а последовательности с требуемым распределением $p_\eta(\eta_i)$ получают из них путем некоторой трансформации. Примером квазислучайных последовательностей точек могут служить ЛЛ_F-последовательности, построенные в работе [5]. Обычно демонстрация большей равномерности квазислучайных последовательностей точек по сравнению с псевдослучайной производится путем проектирования точек на единичный квадрат. Единичный квадрат на Рисунке 1 равномерно разбит на 64 подквадрата и на него нанесены 64 псевдослучайных точки (слева) и квазислучайных точки ЛЛ_F-последовательности (справа). Видно, что в каждый подквадрат на правом рисунке попало ровно по одной точке, в то время как для левого рисунка это далеко не так.

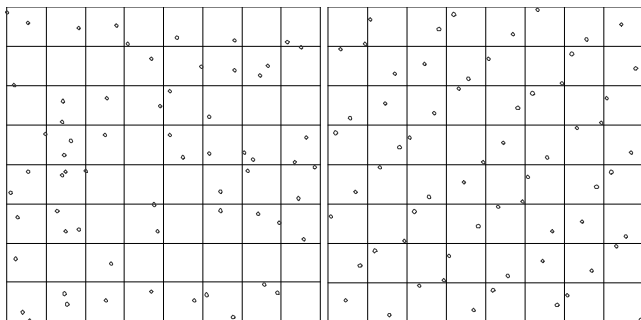


Рисунок 1: Псевдослучайные и квазислучайные точки

Известно, что алгоритмы квази-Монте-Карло, использующие квазислучайные последовательности точек вместо псевдослучайных чисел, сходятся к искомому результату, по

крайней мере не медленнее, а обычно быстрее, чем соответствующие алгоритмы Монте-Карло. Для увеличения производительности алгоритма Монте-Карло обычно достаточно просто формально заменить датчик псевдослучайных чисел на генератор квазислучайных точек. Однако, есть некоторые особенности, которые могут помочь получить еще больший выигрыш от такой замены.

Во-первых, имеет большое значение конструктивная размерность алгоритма [3]. Под конструктивной размерностью алгоритма понимается максимальная размерность точек η , необходимая для его работы. Например, конструктивная размерность интегрирования по формуле (3) равна s . Чем меньше конструктивная размерность, тем больший выигрыш можно надеяться получить от замены псевдослучайных чисел квазислучайными точками.

Во-вторых, имеет большое значение характер интегрируемой функции. Очевидно, что если функция s -мерного аргумента x существенно зависит только от координат с младшими индексами, то по сути интегрирование производится не в s -мерном кубе, а в кубе меньшей размерности. Функция может существенно зависеть и от всех координат, но зависимость от старших координат может быть существенно меньшей, чем от младших. Именно так происходит, например, при решении уравнения рендеринга. Свет поглощается в процессе преотражений, поэтому члены (2) с большими степенями имеют меньшее влияние на значение интегральной суммы. Чем меньше эта зависимость, тем больший выигрыш можно надеяться получить от замены псевдослучайных чисел квазислучайными точками. Разрывы в интегрируемой функции, напротив, ухудшают сходимость квази-Монте-Карло и приближают ее к сходимости Монте-Карло.

4. ПРИМЕНЕНИЕ КВАЗИ-МОНТЕ-КАРЛО ДЛЯ ТРАССИРОВКИ ФОТОНОВ

Существует множество методов Монте-Карло для решения уравнения рендеринга (1) (см. обзор в [2]). Здесь рассматривается использование квазислучайных точек в одном из таких методов – трассировке фотонов. Он состоит в моделировании распространения световой энергии от источников света к глазу наблюдателя путем трассировки фотонов в направлении распространения света. Фотоны испускаются из источников света и трассируются, отражаясь и преломляясь поверхностями сцены. Фотон трассируется до тех пор, пока он либо не будет поглощен одной из поверхностей, либо не вылетит за пределы сцены. При каждом ударе в геометрию сцены энергия фотона помещается в некоторую структуру данных. После того, как накоплена достаточная статистика, можно оценивать искомую величину $L(x, w_{out})$ по данным о попаданиях фотонов в некоторой окрестности точки x .

Проводя параллель с теорией, изложенной с разделом 2, можно считать, что одно испытание состоит в трассировке одного фотона. Фотон с индексом i трассируется в соответствии со значением η_i s -мерной случайной величины η в i -м испытании. Затем результаты всех испытаний усредняются по формуле схожей с (3). Обозначим координаты η_i как $\eta_{i0}, \eta_{i1}, \eta_{i2}, \dots, \eta_{i,s-1}$. Тогда η_{i0} используется для выбора источника света, испускающего фотон, η_{i1}, η_{i2} – для определения направления испускания и т.д. Если вместо

координат величины η_i , определяемых датчиком псевдослучайных чисел, использовать координаты i -й точки некоторой квазислучайной последовательности, формально получим алгоритм квази-Монте-Карло.

5. Понижение конструктивной размерности алгоритма

Вообще говоря, алгоритм трассировки фотонов, кратко описанный в разделе 4 имеет бесконечную конструктивную размерность. Траектория фотона (теоретически) может содержать бесконечное количество звеньев и заранее нельзя сказать, сколько координат η_{ij} нам потребуется для реализации такой траектории. На практике же фотоны быстро поглощаются и вероятность бесконечно длинной траектории равна 0. Поэтому будем понимать под конструктивной размерностью не максимальное, а среднее число координат, необходимых для трассировки одного фотона. Рассмотрим, какое число координат необходимо кратко рассмотренному в разделе 4 “стандартному” алгоритму Монте-Карло для трассировки одного фотона. Для простоты будем предполагать, что моделируемая сцена содержит только точечные источники света. Тогда:

- Одна координата используется для выбора источника света, излучающего фотона.
- Две координаты используются для выбора направления полета фотона.
- При каждом столкновении фотона с поверхностью одна координата используется для выбора события (поглощение, идеально зеркальное отражение, неидеально зеркальное отражение, и т.д.). Если выбранное событие неидеально зеркальное, две координаты необходимо для выбора направления полета рассеянного фотона.

Таким образом, конструктивная размерность трассировки одного фотона равна:

$$D_r = 3 + B_s + 3B_d + C \quad (4)$$

где B_s и B_d – количество соответственно идеально и неидеально зеркальных рассеиваний фотона, C равно 1 в случае, если фотона вылетел за пределы сцены и 2 в случае поглощения.

Известно, что при выборе дискретных событий можно переиспользовать одно и то же псевдослучайное значение несколько раз, применяя каждый раз процедуру перенормировки. Этот метод носит название модифицированного метода суперпозиции и был первоначально предложен для экономии псевдослучайных чисел в целях уменьшения времени, необходимого для их генерации ([3], стр. 66). Его предлагается применять к квазислучайным значениям в целях понижения конструктивной размерности алгоритма.

Пусть необходимо, используя число η_1 , выбрать одно из n дискретных событий, имеющих вероятности p_0, \dots, p_{n-1} . Будет

выбрано событие k такое, что $\sum_{i=0}^{k-1} p_i \leq \eta_1 < \sum_{i=0}^k p_i$.

Последующая процедура перенормировки числа η_1 задается формулой:

$$\eta_1' = \frac{\eta_1 - \sum_{i=0}^{k-1} p_i}{p_k} \quad (5)$$

Перенормированное значение может вновь использоваться для выбора как дискретного, так и непрерывного события. Такую перенормировку можно использовать для экономии случайных значений после выбора любых дискретных событий: после выбора источника света, события рассеивания фотона и т.д. Применение этого простого алгоритма приводит к значительному снижению по сравнению с (4) числа координат, используемых при трассировке одного фотона:

$$D_r' \leq 2 + 2B_d + C \quad (6)$$

Интересно, что конструктивная размерность (6) модифицированного алгоритма абсолютно не зависит от количества зеркальных рассеиваний фотона. Надо отметить, что подобная перенормировка может использоваться как в алгоритмах Монте-Карло, так и в алгоритмах квази-Монте-Карло. Но в алгоритмах Монте-Карло она не приносит никакого выигрыша кроме незначительной экономии времени на генерацию псевдослучайных чисел. В алгоритмах квази-Монте-Карло выигрыш может оказаться существенным, т.к. улучшается характер сходимости.

6. Моделирование источников света и ДФРС поверхностей, заданных в табличном виде

Для уменьшения дисперсии (а значит, увеличения точности) в алгоритмах Монте-Карло генерации изображений обычно применяется метод существенной выборки [3]. Его особенность состоит в том, что плотность фотонов, излученных/рассеянных из точки в любом направлении, в точности пропорциональна потоку энергии в этом направлении.

Основные два места применения метода существенной выборки в алгоритмах трассировки фотонов - излучение фотонов из источника света и рассеивание фотонов при столкновении их с поверхностью. В первом случае плотность фотонов в каждом направлении определяется функцией силы света источника, во втором - ДФРС поверхности в точке попадания фотона. Наиболее общим представлением для этих функций является задание их на сетке в сферической системе координат. Рассмотрим случай источника света. Одним из промышленных стандартов для задания силы света источника является “IES LM-63- 1986: IES Recommended Standard File Format for Electronic Transfer of Photometric Data”. Система координат и сетка для задания силы света в соответствии с этим стандартом изображены на Рисунке 2.

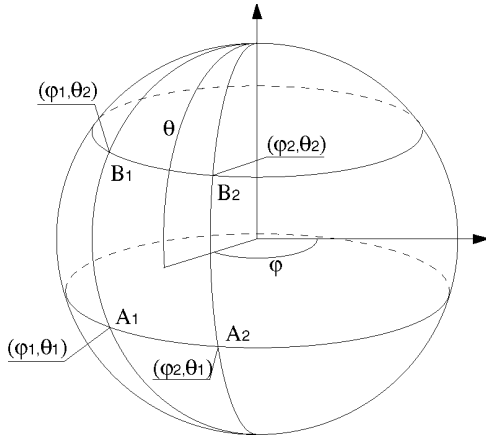


Рисунок 2: Сетка задания значений силы света.

Для определения значения силы света по направлению (φ, θ) , находится ячейка сетки, которой принадлежит это направление. Искомая сила света вычисляется билинейной интерполяцией значений A_1, A_2, B_1, B_2 в углах ячейки (Рисунок 2). Например, сила света внутри изображенной на Рисунок 2 ячейки определяется по формуле:

$$I(\varphi, \theta) = \frac{(\theta - \theta_1)((\varphi - \varphi_1)B_2 + (\varphi_2 - \varphi)B_1)}{(\varphi_2 - \varphi_1)(\theta_2 - \theta_1)} + \frac{(\theta_2 - \theta)((\varphi - \varphi_1)A_2 + (\varphi_2 - \varphi)A_1)}{(\varphi_2 - \varphi_1)(\theta_2 - \theta_1)} \quad (7)$$

Итак, задача состоит в том, чтобы по равномерно распределенным в единичном квадрате значениям (η_1, η_2) выбирать направления излучения фотонов (φ, θ) так, чтобы плотность фотонов в каждом направлении была пропорциональна $I(\varphi, \theta)$.

Первым шагом алгоритма является нахождение ячейки сетки, через которую будет излучен фотон. Для этого на основании значения η_1 выбирается одно из дискретных событий, вероятность каждого из которых пропорциональна потоку через соответствующую ячейку. Выбор события и последующая перенормировка η_1 делается так, как описано в предыдущем разделе. Пусть выбрана ячейка C_k . Необходимо выбрать направление излучения фотона внутри этой ячейки с плотностью вероятности, пропорциональной (7). Из условия равенства 1 интеграла плотности вероятности, она определяется по формуле:

$$p(\varphi, \theta) = \frac{I(\varphi, \theta)}{\int_{C_k} I(\varphi, \theta)}$$

Наилучшим методом для моделирования заданной плотности вероятности считается метод обратных функций [3]. Для его применения в данном случае необходимо вычислить условные плотности вероятности:

$$p_1(\theta) = \int_{\varphi_1}^{\varphi_2} p(\varphi', \theta) d\varphi'$$

$$p_2(\varphi | \theta) = \frac{p(\varphi, \theta)}{p_1(\theta)}$$

Соответствующие p_1 и p_2 функции распределения будут равны:

$$F_1(\theta) = \int_{\theta_1}^{\theta} p_1(\theta') \sin(\theta') d\theta'$$

$$F_2(\varphi | \theta) = \int_{\varphi_1}^{\varphi} p_2(\varphi' | \theta) d\varphi'$$

Тогда углы (φ, θ) излучения фотона определяются из уравнений:

$$F_1(\theta) = \eta_1'$$

$$F_2(\varphi | \theta) = \eta_2$$

В целях экономии места здесь не приводится вывод этих уравнений для функции силы света, определяемой формулой (7). Дается только окончательный результат. Заметим лишь, что вывод заметно упрощается применением очевидной формулы, справедливой для любой линейной функции $L(x)$:

$$\int_a^b L(x) dx = (b - a)L\left(\frac{a + b}{2}\right).$$

углы (φ, θ) излучения фотона могут быть найдены из уравнений:

$$\eta_1' = \frac{\varphi_2 - \varphi_1}{2(\theta_2 - \theta_1) \int_{C_k} I(\varphi, \theta)} [(B_1 + B_2 - A_1 - A_2)(\sin(\theta) - \theta \cos(\theta) - \sin(\theta_1) + \theta_1 \cos(\theta_1)) - (\cos(\theta_1) - \cos(\theta))(\theta_1(B_1 + B_2) - \theta_2(A_1 + A_2))] \quad (8)$$

$$\eta_2 = \frac{(\varphi - \varphi_1)}{2(\varphi_2 - \varphi_1)(\theta_2 - \theta_1) p_1(\theta) \int_{C_k} I(\varphi, \theta)} [(B_2(\varphi - \varphi_1) + B_1(2\varphi_2 - \varphi_1 - \varphi))(\theta - \theta_1) + (A_2(\varphi - \varphi_1) + A_1(2\varphi_2 - \varphi_1 - \varphi))(\theta_2 - \theta)]$$

Первое уравнение является трансцендентным и его решение аналитически проблематично. Именно поэтому обычно в методах Монте-Карло вместо метода обратных функций используют метод отказов для моделирования этого распределения. Основная идея метода отказов состоит в том, что в соответствии с равномерным распределением генерируются пары углов (φ, θ) до тех пор, пока очередная пара не удовлетворит определенному условию. Известно, что использование этого метода приводит к значительному снижению эффективности алгоритмов квази-Монте-Карло по двум причинам: возрастает конструктивная размерность алгоритма и увеличивается разрывность задачи [3]. Поэтому при переходе на квазислучайные последовательности точек необходимо найти альтернативный способ генерации направлений испускания фотонов из источников света с таблично заданной силой света.

Предлагается моделировать необходимое распределение путем численного решения первого из уравнений (8). Хотя оно выглядит достаточно громоздко, на самом деле большинство его членов могут быть рассчитаны заранее (для каждой ячейки задания функции отдельно). Тогда одна итерация решения этого уравнения методом секущих требует вычисления одного \sin , одного \cos , четыре операции сложения/вычитания и четыре умножения. Кроме метода секущих было реализовано решение этого уравнения методами Ньютона и Риддера [6]. Но в большинстве случаев метод секущих показывает лучший результат ввиду своей

простоты и хорошей обусловленности решаемого уравнения. Второе из уравнений (8), из которого находится угол ϕ , является квадратным и его решение не вызывает труда.

Моделирование ДФРС поверхностей, заданных в табличном виде, можно производить похожим образом. В общем случае ДФРС поверхности является функцией четырех углов: два определяют направления падающего луча, и два – рассеянного. Первые два угла однозначно вычисляются при попадании фотона на поверхность. После этого функция рассеивания фотона поверхностью, по сути, становится функцией двух углов, заданной в сферической системе координат. Так же как для источников света, наиболее общим видом ее задания является аналогичное Рисунку 2 сеточное представление. Повторив для нее все рассуждения, сделанные для источников света, придем к формуле, аналогичной (8) для нахождения направления рассеянного фотона по двум псевдо(квази)случайным значениям.

Заметим, что изложенный алгоритм имеет смысл только в случае использования квази-Монте-Карло, т.к. улучшает характер сходимости. В случае алгоритмов Монте-Карло, сходимость которых не зависит от конструктивной размерности, замена метода отказов на метод обратных функций в данном случае выигрыша не дает.

7. РЕЗУЛЬТАТЫ

В качестве примера применения предложенных алгоритмов, используется простая сцена, состоящая из пластины, освещенной точечным источником света. Функция яркости источника света задана в табличном виде. Между источником света и пластиной находится толстое стекло, искаженное некоторой шумовой функцией. Свет, преломляясь в стекле, образует интересные световые эффекты на пластине (Рисунок 3). Такие эффекты, называемые каустиками, возникают часто и в реальной жизни (например, на дне бассейна). Их моделирование обычно требует долгих вычислений, но необходимо для физически достоверного представления действительности. В таблице приведены данные об относительной погрешности и времени вычислений различных описанных алгоритмов. Все вычисления производились для $4'456'448$ фотонов.

Алгоритм	Погрешность	Время
Монте-Карло	2.3%	161.7 сек
квази-Монте-Карло	1.45%	158.1 сек
+ использование метода обратных функций (раздел 6)	1.19%	162.1 сек
+ понижение конструктивной размерности (раздел 5)	1.05%	157.9 сек

Видно, что в данном случае уже формальная замена псевдослучайных чисел на квазислучайные точки значительно увеличивает точность вычислений. При этом время немного уменьшается, т.к. генерация точек ЛП_r-последовательности немного дешевле генерации псевдослучайных чисел. Использование метода обратных функций вместо метода отказов еще больше увеличивает точность, но немного (приблизительно на 2%) замедляет

трассировку лучей. Это происходит потому, что для испускания каждого фотона численно решается трансцендентное уравнение. Наконец, при понижении конструктивной размерности способом из раздела 5, точность еще повышается. Увеличение скорости трассировки фотонов в этом случае связано с тем, что используется меньшее количество координат квазислучайных точек, а значит, экономится время на их генерацию. Изображения на рисунке 3 демонстрируют уровень шума для первого и последнего алгоритмов из таблицы при одном и том же числе фотонов.

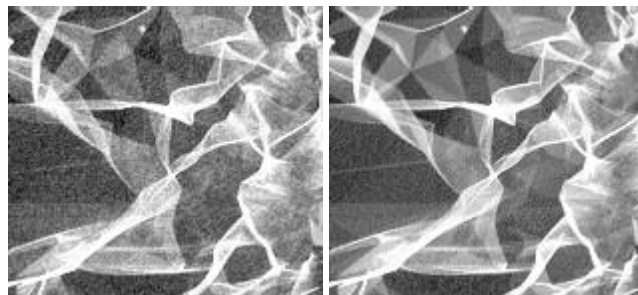


Рисунок 3: Изображения сцены после $4'456'448$ фотонов, полученные алгоритмами Монте-Карло (слева) и квази-Монте-Карло (справа)

Автор благодарит компанию Integra Inc. (<http://www.integra.co.jp>) за предоставление исходных кодов системы генерации физически аккуратных изображений для проведения экспериментов, а также В.А.Галактионова и Ю.М.Баяковского за помощь в проведении исследований.

8. ЛИТЕРАТУРА

- [1] S.Heirich and A.Keller. *Quasi-Monte Carlo Methods in Computer Graphics Part II: The Radiance Equation. Technical Report 243/94, 1994*
- [2] L. Szirmay-Kalos. *Monte-Carlo Methods in Global Illumination. Institute of Computer Graphics, Vienna University of Technology, Vienna, 1999*
- [3] И.М.Соболь. *Численные методы Монте-Карло. "Наука", 1973.*
- [4] J.T.Kajiya. *The rendering equation. In David C. Evans and Rusell J. Athay, editors, Computer Graphics (SIGGRAPH'86 Proceedings), vol. 20, pages 143-150, August 1986.*
- [5] И.М.Соболь. *Многомерные квадратурные формулы и функции Хаара. "Наука", 1969*
- [6] W. Press, B. Flannery, S. Teukolsky and W. Vetterling. *Numerical Recipes: The Art of Scientific Computing, 2nd edition, Cambridge University Press, 1992*

Об авторе

Кирилл А. Дмитриев является аспирантом Института Прикладной Математики им. М.В.Келдыша РАН
E-mail: kadmitr@gin.keldysh.ru