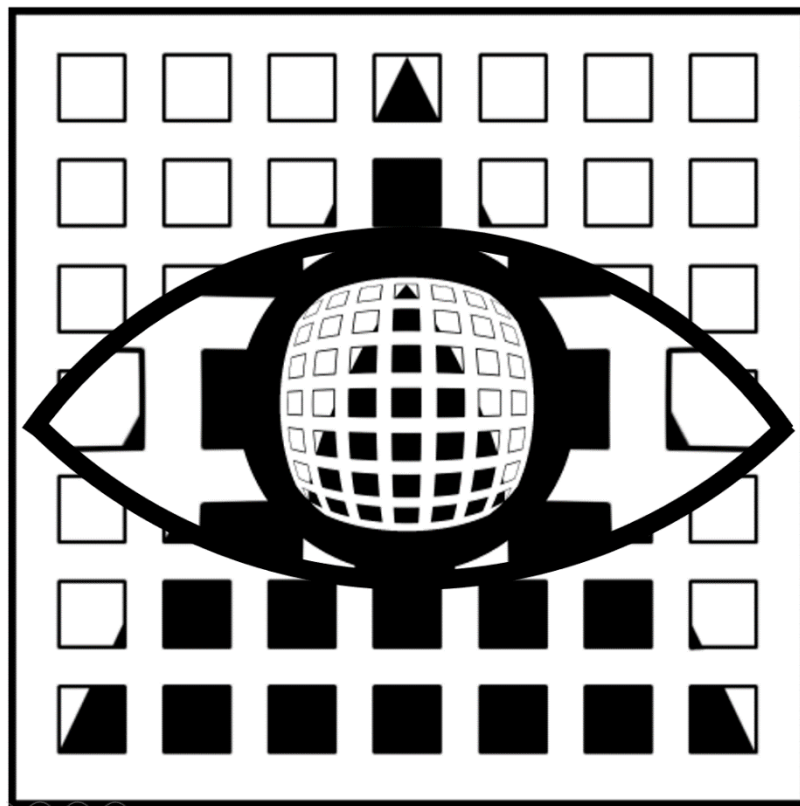


# ИТМО

Д.Д. Жданов, И.С. Потемин, А.Д. Жданов

## МЕТОДЫ РАСЧЕТА ГЛОБАЛЬНОЙ ОСВЕЩЕННОСТИ



Санкт-Петербург  
2023

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

**Д.Д. Жданов, И.С. Потемин, А.Д. Жданов**  
**МЕТОДЫ РАСЧЕТА ГЛОБАЛЬНОЙ**  
**ОСВЕЩЕННОСТИ**

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ  
ИТМО

по направлению подготовки 09.04.04 Программная инженерия  
в качестве Учебно-методическое пособие для реализации основных  
профессиональных образовательных программ высшего образования  
магистратуры

**ИТМО**

Санкт-Петербург  
2023

Жданов Д.Д., Потемин И.С., Жданов А.Д., Методы расчета глобальной освещенности – СПб: Университет ИТМО, 2023. – 160 с.

Рецензент(ы):

Сорокин Максим Игоревич, кандидат технических наук, заведующий лабораторией факультета программной инженерии и компьютерной техники, Университета ИТМО.

Учебно-методическое пособие посвящено решению задач расчета глобального освещения трехмерных сцен. Представлены четыре основных раздела, касающиеся определения основных радиометрических величин и соотношений между ними, детерминистических методов решения проблемы глобального освещения, методов Монте-Карло и способов формирования псевдослучайных величин с заданной плотностью распределения вероятности, и стохастических методов решения проблемы глобального освещения. Пособие предназначено для студентов, обучающихся по направлению подготовки магистрантов 09.04.04 "Системное и прикладное программное обеспечение".

The logo of ITMO University, consisting of the letters 'ИТМО' in a bold, black, sans-serif font.

**Университет ИТМО** – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2023

© Жданов Д.Д., Потемин И.С., Жданов А.Д., 2023

## Содержание

1. Введение .....	6
1.1 Физически корректный рендеринг .....	7
1.2 Проблема глобального освещения .....	8
2. Радиометрия .....	10
2.1 Основные радиометрические величины .....	10
2.1.1 Яркость .....	10
2.1.2 Световой поток .....	14
2.1.3 Двухнаправленная функция отражения (ДФО или BRDF) .....	16
2.2 Уравнение рендеринга .....	19
2.3 Уравнение потенциала .....	21
2.4 Глобальная функция отражения .....	25
2.4.1 Выражение светового потока в терминах GRDF .....	26
2.4.2 Выражения, определяющие GRDF .....	27
2.5 Каустика .....	30
2.6 Заключение .....	31
2.7 Вопросы для самоконтроля усвоенных знаний .....	33
3. Детерминистические методы решения уравнения рендеринга .....	35
3.1 Интегральное уравнение рендеринга .....	35
3.2 Решение уравнения рендеринга для диффузного освещения .....	36
3.3 Расчет глобального освещения для фотометрического шара .....	46
3.3.1 Обмен излучения внутри фотометрического шара .....	46
3.3.2 Яркость на внутренней поверхности фотометрического шара .....	48
3.3.3 Применение фотометрического шара для измерения оптических характеристик .....	53
3.4 Заключение .....	55
3.5 Вопросы для самоконтроля усвоенных знаний .....	57
4. Метод Монте-Карло .....	58
4.1 Интегрирование методом Монте-Карло .....	59



4.2	Стратифицированная выборка .....	62
4.3	Выборка по значимости .....	67
4.4	Многokратная выборка по значимости .....	72
4.5	Метод выделения главной части .....	82
4.6	Решение интегральных уравнений методом Монте-Карло .....	84
4.6.1	Русская рулетка .....	85
4.6.2	Комбинация русской рулетки с методами выборки по значимости и методом выделения главной части .....	87
4.6.3	Применение русской рулетки для решения интегральных уравнений .....	89
4.6.4	Оценка следующего события.....	90
4.7	Формирование случайных величин с заданным PDF .....	91
4.7.1	Равномерное распределение случайной координаты на плоских фигурах .....	92
4.7.2	Равномерное распределение точек на поверхности сферы .....	97
4.7.3	Распределение точек с заданной плотностью вероятности на плоскости .....	101
4.7.4	Распределение лучей с заданной плотностью вероятности .....	107
4.8	Заключение.....	116
4.9	Вопросы для самоконтроля усвоенных знаний.....	120
5.	Применение методов Монте-Карло для решения уравнения рендеринга – трассировка путей .....	122
5.1	Базовый алгоритм.....	122
5.2	Стратифицированная выборка .....	127
5.3	Выборка по значимости .....	128
5.4	Оценка следующего события .....	132
5.5	Многokратная выборка по значимости .....	139
5.6	Выделение главной части.....	142
5.7	Улучшение выборки по значимости и выделению главной части.....	145
5.8	Заключение.....	148
5.9	Вопросы для самоконтроля усвоенных знаний.....	150

6. Применение методов Монте-Карло к уравнению потенциала – трассировка лучей.....	151
6.1 Базовый алгоритм.....	151
6.2 Выборка по значимости.....	154
6.3 Оценка следующего события.....	156
6.4 Заключение.....	160
6.5 Вопросы для самоконтроля усвоенных знаний.....	161
7. Литература.....	161

## 1. Введение

В последнее время алгоритмические и аппаратные решения в компьютерной графике претерпели значительные изменения. На смену растровой графике, используемой в старых игровых приложениях или для визуализации моделей, подготовленных в САД системах, пришла лучевая реалистичная графика. Современная компьютерная графика – это попытка реалистично сформировать изображения реального мира с учетом физически корректных законов распространения света в сцене. Физическая корректность алгоритмов компьютерной графики вывела игровые приложения на новый реалистичный уровень, в киноиндустрии позволила создавать фильмы, подобные Аватару, а в системах виртуальной реальности – создавать миры, неотличимые от реальных. Кроме того, алгоритмы и программные решения реалистичной компьютерной графики нашли применение в инженерных задачах, например, при проектировании сложных оптических устройств, когда реальное оптическое устройство заменяется его виртуальным прототипом, физически корректно определяющим распространение светового излучения в этом устройстве.

Целью настоящего учебного пособия является введение в основы методов расчета глобального освещения в сценах, содержащих объекты со сложными оптическими свойствами. Учебное пособие имеет теоретически характер, и для усвоения теоретических знаний каждый раздел содержит вопросы для самоконтроля, имеющие как теоретический, так и практический характер, ориентированный на применение теоретических знаний для решения практических задач.

Цель авторов – предложить читателям изложение основных концепций и методов расчета яркости глобального освещения в расчете на самостоятельное изучение пособия в течение двух-трех месяцев. При использовании пособия в технических вузах рекомендуется приведенный материал использовать в качестве односеместрового учебного курса в рамках подготовки студентов по направлению «Программная инженерия», образовательной программе «Системное и прикладное программное обеспечение» и блоку дисциплин «Технологии компьютерной визуализации» («Теоретические основы компьютерной графики и вычислительной оптики», «Методы расчета глобальной освещенности», «Фотореалистичная визуализация трехмерных сцен»). Кроме того, настоящее пособие будет полезно при изучении дисциплин не только в областях, смежных с направлением подготовки «Программная инженерия», но и по направлениям, связанным со светотехникой, светодизайном и прикладной оптикой.

## 1.1 Физически корректный рендеринг

Основной проблемой физически корректного рендеринга является проблема глобального освещения. Результат расчета глобального освещения сцены выражается в естественности визуального восприятия ее изображения и обеспечивается в первую очередь физически корректными моделями взаимодействия светового излучения с объектами сцены. Взаимодействие, как правило, представляет собой сложную комбинацию отражения, пропускания и рассеивания света. С точки зрения физических величин результатами расчета глобального освещения являются функции распределения яркости, освещенности или интенсивности излучения, покидающего сцену. Первым подходом к решению проблемы расчета глобального освещения был алгоритм *radiosity*, предложенный Горалом и др. [1] в 1984 году. Алгоритм *radiosity* использует конечно-элементную модель рассеянного освещения в сцене. Это был первый алгоритм, основанный на физической модели распространения светового излучения в сцене. Несмотря на то, что этот метод формирует изображение, свободное от шума, он не получил широкого распространения в связи с большим числом ограничений данного алгоритма. основополагающим решением проблемы реалистичного рендеринга в компьютерной графике стало уравнение рендеринга, предложенное Каджией [2] в 1986 году. Это обеспечило более общую физическую основу для решения проблемы рендеринга, как правило, методами стохастической трассировки лучей.

Другой подход к классификации методов физически корректного рендеринга заключается в определении пространства, в котором алгоритм рендеринга пытается найти решение. Решение может быть ориентировано на поиск распределения световых характеристик в пространстве сцены независимо от условий наблюдения (решение основывается на так называемых объектно-ориентированных алгоритмах) и на поиске распределения световых характеристик на изображении сцены.

Объектно-ориентированные алгоритмы рендеринга пытаются решить глобальную проблему освещения независимо от условий наблюдения, таких как точка обзора и поле зрения. Для решения этой проблемы необходим расчет глобального освещения по всей сцене. Самое простое решение – это метод *radiosity*, который позволяет вычислить глобальную освещенность сцены, если она содержит объекты, оптические свойства которых являются ламбертовыми. Следующий шаг в сторону физической корректности – это алгоритмы, основанные на интегрировании освещенности методами Монте-Карло, например, расчет глобальной освещенности методом стохастической трассировки световых лучей [3]. Эти решения позволяют физически корректно вычислить глобальную

освещенность, однако не учитывают локальные условия освещения, т. е. полученный результат не может быть корректно пересчитан в яркость сцены для заданных условий наблюдения. И, наконец, физически корректные алгоритмы расчета глобального освещения — это карты яркости сцены или фотонные карты [4-6]. Данные алгоритмы основаны на методах интегрирования Монте-Карло и обеспечивают физически корректный пересчет результата в видимую яркость сцены. Основными недостатками данных методов являются смещенность полученного результата, вызванная конечной областью интегрирования яркости, и очень большой объем данных, требуемых для хранения результата рендеринга.

Алгоритмы рендеринга, основанные на синтезе двумерных изображений сцены для заданных условий наблюдения, в ряде случаев используют алгоритмы, аналогичные объектно-ориентированным алгоритмам, однако размерности результата рендеринга принципиально отличны. Если при отсутствии поляризации объектно-ориентированные алгоритмы требуют шести размерностей (три размерности для координат в пространстве, две размерности для ориентации в пространстве и одна размерность для цвета) для хранения данных, то алгоритмы рендеринга, формирующие двумерное изображение, требуют только трех размерностей (две размерности для координат точки на экране и одна размерность для цвета). Снижение размерности на три порядка позволяет использовать несмещенные алгоритмы [7, 8]. Кроме того, проблема алгоритмов расчета глобального освещения, использующих смещенные решения [7, 8], становится не столь острой, поскольку позволяет значительно уменьшить область интегрирования яркости и, в ряде случаев, свести ее практически к нулю. Основным недостатком данной группы алгоритмов является то, что они не решают проблему глобального освещения, а решают только проблему расчета яркости для некоторых заданных условий наблюдения.

## **1.2 Проблема глобального освещения**

Физически корректный рендеринг представляет собой задачу моделирования, которая обычно решается в три этапа:

1. Определение физической проблемы.
2. Формализация проблемы в математической модели.
3. Разработка алгоритма для решения модели.

В этой главе мы сосредоточимся на первых двух шагах. При решении проблемы физически корректного рендеринга основной интерес вызывает свет и его взаимодействие с окружающей средой. Свет излучается от источников света и распространяется в сцене. Свет, распространяющийся в направлении наблюдателя, формирует изображение сцены. Физически корректный рендеринг

— это программная имитация физических процессов распространения света в сцене. Основной проблемой физически корректного рендеринга является проблема расчета глобального освещения, т. е. моделирования процессов распространения и взаимодействия светового излучения с оптически сложными объектами сцены.

В рамках данного курса решение проблемы глобального освещения будет рассматриваться только с точки зрения лучевой теории и радиометрии. То есть когерентность светового излучения, явления интерференции и дифракции не будут рассматриваться. Также не будут рассматриваться вопросы непрямолинейного распространения света в средах с градиентом показателя преломления, распространение света в рассеивающих средах, включая подповерхностное рассеивание. Кроме этого, эффекты поляризации, флуоресценции и двойного лучепреломления рассматриваться не будут. Основные предположения радиометрической лучевой модели распространения света:

1. Линейность – суммарный эффект двух входных сигналов всегда равен сумме эффектов каждого сигнала по отдельности
2. Сохранение энергии – рассеиваемая энергия не может выдавать больше энергии, чем изначальная
3. Прямолинейное распространение света – между двумя объектами свет распространяется по прямой линии.
4. Устойчивость состояния – распределение световой энергии не зависит от времени.

В первую очередь будут определены две основные радиометрические величины, необходимые для расчета глобального освещения. Это яркость и световой поток. Далее будет введено понятие двунаправленной функции рассеивания ДФР (отражения ДФО или пропускания ДФП), определяющей локальный характер рассеивания светового излучения поверхностью сцены. Эти физические величины позволят рассмотреть несколько альтернативных математических моделей глобального освещения. Наиболее известная модель формулируется в виде уравнения рендеринга. Это интегральное уравнение определяет функцию яркости. С ним тесно связано уравнение потенциала. Все это позволяет ввести новое понятие – глобальную функцию распределения отражательной способности, а также два эквивалентных интегральных уравнения, которые определяют ее поведение. Различные модели обеспечивают различные взгляды на решение проблемы глобального освещения, что приводит к появлению различных алгоритмов его расчета.

## 2. Радиометрия

Радиометрические величины являются основными составляющими уравнений глобального освещения. Как правило, они рассматриваются в видимой области спектра в диапазоне 380-780 нм. Каждая радиометрическая величина имеет аналог соответствующей спектральной плотности на единицу длины волны. Спектральная составляющая может быть указана в формулах в виде нижнего индекса, однако большинство формул абстрагируется от спектральных зависимостей, и нижний индекс опускается. Использование фотометрических величин, связанных с интегральной чувствительностью человеческого глаза, при расчете глобального освещения может привести к расхождению с результатами радиометрического расчета, поскольку фотометрические величины не имеют однозначного преобразования в радиометрическое представление, а корректный результат расчета глобального освещения зависит от спектральных свойств источников света и объектов сцены. Фотометрические величины, как правило, используются для анализа конечных результатов вычислений, чтобы оценить воздействие светового излучения на человека. Кроме того, в тексте при определении таких величин, как яркость и световой поток, отсутствует слово «энергетический». Это показывает, что фотометрия и радиометрия используют практически одинаковые определения световых величин с единственным различием: каким образом происходит интегрирование световой величины в спектральном диапазоне.

### 2.1 Основные радиометрические величины

#### 2.1.1 Яркость

Яркость  $L$  для данной точки  $x$  в направлении  $\Theta_x$  определяется как дифференциальная световая энергия, излучаемая с единичной площади поверхности в единицу времени из точки  $x$  в направлении  $\Theta_x$ , см. рис. 2.1.

$$L(x, \Theta_x) = \frac{\partial^2 \Phi}{\partial A_{x\perp} \partial \omega_x}, \quad (2.1)$$

где:

- $\partial A_{x\perp}$  – дифференциальная область поверхности вокруг точки  $x$  и под прямым углом к направлению  $\Theta_x$ ,
- $\partial \omega_x$  – дифференциальный телесный угол вокруг направления  $\Theta_x$ ,
- $\partial^2 \Phi$  – дифференциальный поток с площадки  $\partial A_{x\perp}$  в телесном угле  $\partial \omega_x$ .

Единицы яркости  $\text{Вт}/\text{м}^2\text{ср}$ . Необходимо обратить внимание, что ср на самом деле не являются физическими единицами, но они, как правило, включаются для наглядности. Поскольку значение яркости определено для каждой точки трехмерного пространства и для каждого направления в двумерном пространстве, то функция яркости является пятимерной функцией. Если учесть ее спектральный характер, то она становится шестимерной функцией.

Яркость является фундаментальной физической величиной, поскольку на ее основе формируется большинство радиометрических и фотометрических величин. Кроме того, яркость важна, поскольку именно она воспринимается наблюдателем как изображение реального мира.

Яркость непосредственно перед попаданием на поверхность называется входящей яркостью или яркостью поля. Она будет обозначаться как  $L_i$ . Яркость сразу после покидания поверхности называется исходящей яркостью или яркостью поверхности. Она будет обозначаться как  $L_o$ . Относительно поверхности яркости могут быть определены следующим образом:

$$L_i(x, \Theta_x) = \frac{\partial^2 \Phi}{|\Theta_x \cdot N_x| \partial \mu_x \partial \omega_x}, \quad (2.2)$$

$$L_o(x, \Theta_x) = \frac{\partial^2 \Phi}{|\Theta_x \cdot N_x| \partial \mu_x \partial \omega_x}, \quad (2.3)$$

где:

- $|\Theta_x \cdot N_x|$  – абсолютное значение косинуса угла между направлением  $\Theta_x$  и направлением локальной нормали к поверхности  $N_x$  в точке  $x$ ,
- $\partial \mu_x$  – дифференциальная область поверхности вокруг точки  $x$ ,
- $\partial \omega_x$  – дифференциальный телесный угол вокруг направления  $\Theta_x$ ,
- $\partial^2 \Phi$  – дифференциальный поток с площадки  $\partial \mu_x$  в телесном угле  $\partial \omega_x$ .

Кроме того, существует такое физическое явление, как самосвечение поверхности, которое может быть результатом некоторого физического или химического процесса, например, люминесценции.



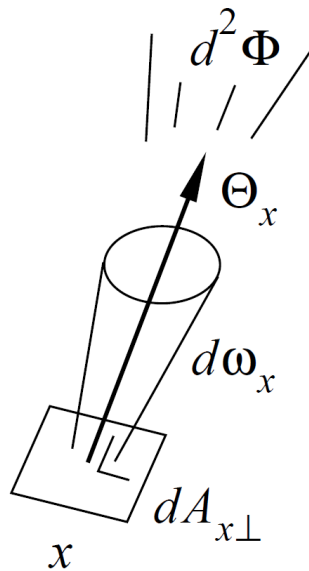


Рис. 2.1. Символы, используемые в определении яркости. Яркость определяет световой поток, излучаемый из данной точки  $x$  в направлении  $\Theta_x$

Наиболее важным свойством яркости для задач глобального освещения является ее постоянство вдоль направления наблюдения при условии отсутствия поглощения или рассеивания в среде, см. рис. 2.2. Это свойство следует из более общего закона сохранения энергии. Данное приближение справедливо для большинства окружающих сцен, которые не содержат сильно рассеивающих, поглощающих или излучающих сред, таких как дым, пар, туман или огонь.

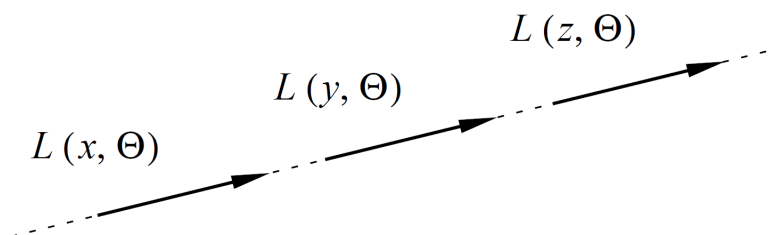


Рис. 2.2. При отсутствии поглощения яркость остается постоянной в направлении излучения  $\Theta$

Постоянство яркости подразумевает, что вдоль заданного направления поверхностная яркость в одной точке сцены равна яркости поля в произвольной точке сцены вдоль этого направления. Эта точка может быть формальным образом определена функцией смещения  $r$ , которая отображает точку  $x$  по направлению  $\Theta_x$  на ближайшую точку поверхности сцены, см. рис. 2.3. Данный инвариант записывается следующим образом:

$$L_o(y, \Theta_y) = L_i(r(x, \Theta_x), \Theta_x). \quad (2.4)$$

В наших выражениях часто будет неявным образом использоваться принцип обратимости лучей:

$$y = r(x, \Theta_x) \Leftrightarrow x = r(y, \Theta_x^{-1}),$$

где направление, противоположное направлению  $\Theta_x$  обозначается как  $\Theta_x^{-1}$ . В большинстве уравнений, которые будут представлены ниже, основное направление переноса предполагает перенос светового излучения, что соответствует физике процесса.

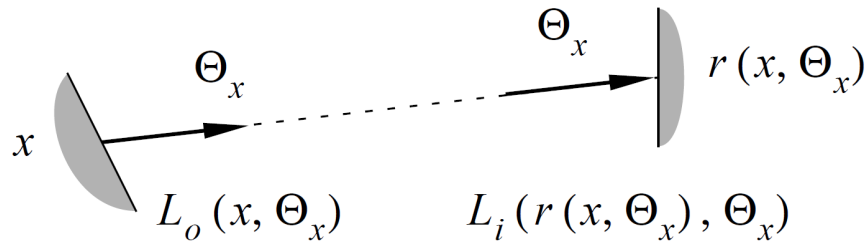


Рис. 2.3. Функция смещения лучей  $r$  отображает точку  $x$  и направление  $\Theta_x$  на ближайшую точку  $x$ , видимую по направления  $\Theta_x$ . Вследствие неизменности яркости по направлению  $\Theta_x$  поверхностная яркость  $L_o(y, \Theta_y)$  равна полевой яркости  $L_i(r(x, \Theta_x), \Theta_x)$

Более общим выражением закона сохранения яркости является инвариант Штраубеля [7], который говорит, что при отсутствии поглощения яркость сохраняется после произвольного числа идеальных событий «зеркального» отражения и преломления, если источник яркости и наблюдатель находятся в среде с одним показателем преломления. Если среды показателей преломления различны, то выходная яркость отличается в квадрат раз относительного показателя преломления:

$$L_o(y, \Theta_y) = L_i(x, \Theta_x) \frac{n_o^2(y)}{n_i^2(x)}. \quad (2.5)$$

Величины  $\frac{L_i(x, \Theta_x)}{n_i^2(x)}$  и  $\frac{L_o(y, \Theta_y)}{n_o^2(y)}$  называют приведенной яркостью, которая сохраняется в пространстве при отсутствии поглощения или рассеивания в среде,

а также на границах сред. В дальнейшем можно будет говорить о приведенной яркости, опуская величину показателя преломления (полагая его равным единице), при этом функция смещения  $r$  не является смещением по прямой линии, а осуществляет перенос излучения из точки  $o$  в точку  $i$  в соответствии с уравнением эйконала [7]. При этом при отсутствии поглощения или рассеивания совершенно неважно, что происходит на пути света из точки  $x$  в точку  $y$ , свет может пройти через среды с градиентом показателя преломления, отразиться или преломиться на гладких поверхностях. Если необходим учет ослабления света в среде, то выражение (2.4) преобразуется к виду:

$$L_o(y, \Theta_y) = \tau(x, \Theta_x, y) L_i(r(x, \Theta_x), \Theta_x), \quad (2.6)$$

где  $\tau(x, \Theta_x, y)$  – функция поглощения световой энергии, возникшая на трассе луча, вышедшего из точки  $x$  в направлении  $\Theta_x$  и дошедшего до точки  $y$ . В дальнейших выражениях функция поглощения не будет использоваться, однако ее можно будет легко добавить во все уравнения, поскольку в большинстве случаев (случай среды с градиентом показателя преломления не рассматривается) это просто некоторый коэффициент (значение которого меньше или равно единице), изменяющий яркость поля, смещенного относительно поверхности на функцию  $r(x, \Theta_x)$ .

### 2.1.2 Световой поток

Световой поток через некоторую область точек пространства и направлений  $S$  определяется как общее количество световой энергии за единицу времени, излучаемой с этой области точек пространства и направлений  $S$ . Световой поток выражается в Вт. По определению – это интеграл яркости по всей области точек пространства и по всем направлениям  $S$ .

$$\Phi = \iint_S L(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x. \quad (2.7)$$

Для поверхностной яркости выражение может быть записано как интеграл по всем точкам поверхности и возможным направлениям излучения с помощью некоторой функции  $W_e(x, \Theta_x)$ .

$$\Phi = \int_A \int_{\Omega_x} L(x, \Theta_x) W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x, \quad (2.8)$$

где:

- $A$  – область излучающей поверхности,
- $\Omega_x$  – полусфера вокруг точки поверхности  $x$ ,

- $W_e(x, \Theta_x)$  – есть единица для всех пар точек и направлений, принадлежащих области излучения, и ноль в противном случае (смысл функции  $W_e$  будет понятен позже),
- $d\mu_x$  – дифференциал области поверхности вокруг точки  $x$ ,
- $d\omega_x$  – дифференциал телесного угла вокруг направления  $\Theta_x$ .

Выражение светового потока может быть представлено в виде скалярного произведения:

$$\Phi = \langle L, W_e \rangle, \quad (2.9)$$

где скалярное произведение определяется как

$$\langle f, g \rangle = \int_A \int_{\Omega_x} f(x, \Theta_x) g(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x.$$

Область точек пространства и направлений – это точки поверхности вместе со всеми возможными направлениями по полусфере, см. рис. 2.4. Классические методы определения яркости вычисляют эти потоки для заданных участков сцены. Однако больший интерес будет представлять набор точек и направлений области сцены, соответствующих пикселю виртуальной камеры, см. рис. 2.5. Методы, основанные на синтезе двумерных изображений, вычисляют именно величину этого потока.

В качестве альтернативы можно определить средневзвешенное значение  $\bar{L}$  функции яркости по некоторой области поверхности в направлении наблюдения, а не общий световой поток, вводя дополнительное определение весовой функции и определяя  $W_e(x, \Theta_x)$ , которая может иметь значения, отличные от нуля или единицы. Эта весовая функция нормализуется следующим образом:

$$\int_A \int_{\Omega_x} W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x = 1.$$

Тогда выражение средней яркости принимает вид светового потока для поверхностной яркости:

$$\bar{L} = \int_A \int_{\Omega_x} L(x, \Theta_x) W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x.$$

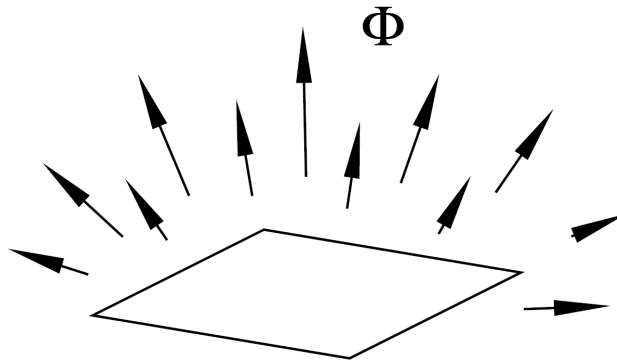


Рис. 2.4. Световой поток — это количество световой энергии, излучаемой в единицу времени с некоторой областью точек пространства и направлений. В методах radiosity область обычно состоит из всех ее точек со всеми возможными направлениями по полусфере

Это выражение представляет интерес, например, при вычислении средневзвешенного значения функции яркости по пикселю. Поскольку вычисление потока и вычисление средневзвешенного значения яркости, по существу, одно и то же, дальнейшее рассуждение будет вестись в терминах световых потоков, неявно предполагая альтернативную возможность средневзвешенного распределения яркости.

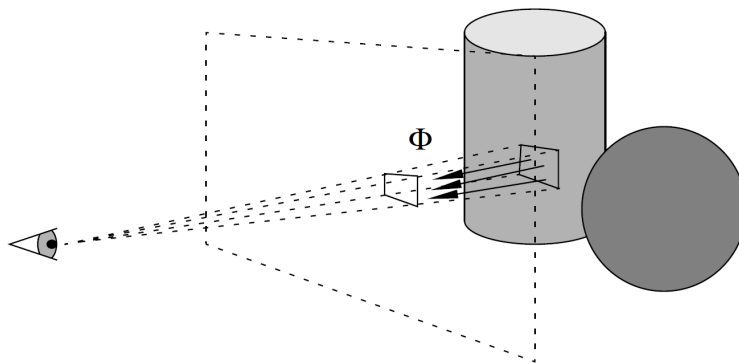


Рис. 2.5. Набор точек и направлений области сцены, соответствующих пикселю виртуальной камеры, используемый в методах рендеринга двумерных изображений сцены

### 2.1.3 Двухнаправленная функция отражения (ДФО или BRDF)

Двухнаправленная функция отражения  $f_r$  для точки на поверхности  $x$ , направления освещения  $\Theta_i$  и направления наблюдения  $\Theta_o$  представляет собой меру количества излучения, падающего в точку  $x$  на поверхности по

направлению  $\Theta_i$ , которое отражается вдоль направления  $\Theta_o$ , см. рис. 2.6. Формально это выглядит так:

$$f_r = \frac{\partial L_{output}(x, \Theta_o)}{L_{input}(x, \Theta_i) |\Theta_i \cdot N_x| \partial \omega_i} \quad (2.10)$$

где:

- $\partial L_{output}(x, \Theta_o)$  – дифференциал яркости поверхности в точке  $x$  и направлении  $\Theta_o$ , полученной в результате отражения света, падающего на поверхность в точке  $x$ , направлении  $\Theta_i$  и в телесном угле  $\omega_i$ .

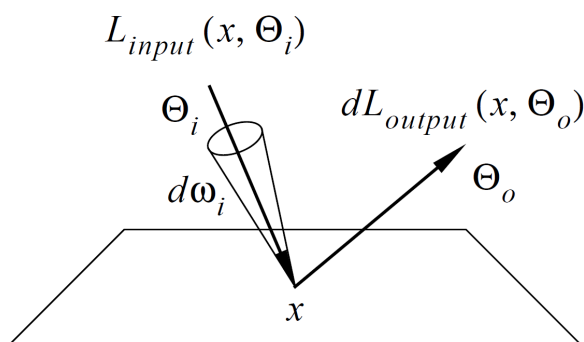


Рис. 2.6. Визуальное представление обозначений, используемых для определения ДФО

Единицы измерений ДФР  $1/\text{ср}$ . ДФР определяет локальные свойства отражения поверхности. Для диэлектрических или металлических поверхностей с шероховатостью, вариация высоты которой значительно превышает длину волны видимого излучения, ДФО есть результат взаимодействия светового излучения с микрофасетами шероховатой поверхности, которая может быть вычислена по формулам Френеля и законам Снелла. Модели таких зеркальных микрофасетных ДФО представлены в работах [9-11]. Для других типов объектов характеристики отражения поверхностей могут быть результатом более сложного процесса взаимодействия светового излучения с материалом, например, быть результатом подповерхностного рассеивания. В общем случае ДФО может быть измерена на спектрогониофотометре и представлена в виде таблицы [12]. В практических задачах часто используются эмпирические модели [12-14]. Окончательно параметры этих моделей могут быть аппроксимированы методом наименьших квадратов по результатам реальных измерений. Обзор моделей ДФО представлен в работе [15]. На рис. 2.7 приведен пример простой модели ДФО.

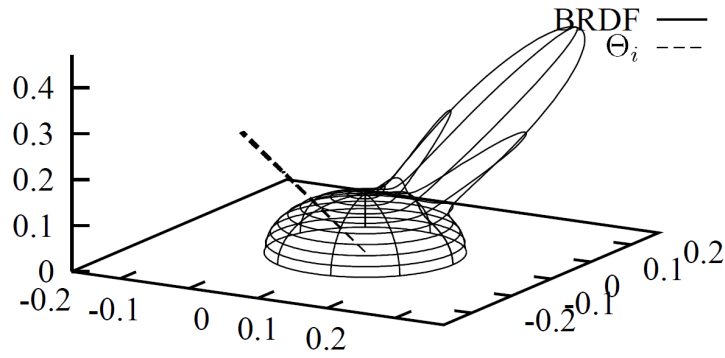


Рис. 2.7. Пример представления ДФО в полярных координатах для модели Фонга при заданном направлении освещения. ДФО есть сумма постоянной диффузной составляющей и зеркального блика вокруг направления отражения

В дальнейших рассуждениях предполагается, что ДФО всегда может быть представлена моделью, подходящей для вычислений яркости или светового потока. Особый случай ДФО – это идеальный ламбертовский отражатель, для которого ДФО является константой по всей полусфере направлений освещения и наблюдения.

$$f_r(x, \theta_i, \theta_o) = f_r, \quad \forall \theta_i \in \Omega_x, \forall \theta_o \in \Omega_o.$$

Определение ДФО может быть легко расширено для того, чтобы включить эффект преломления. В этом случае функция будет называться двунаправленной функцией пропускания (ДФП или BTDF). В общем случае можно рассматривать рассеивание по всей сфере (над и под поверхностью). Тогда эта функция называется двунаправленной функцией рассеивания (ДФР или BSDF). Далее это обобщение будет неявно использоваться.

### Сохранение энергии

Физически корректная ДФО не должна рассеивать большее количество энергии, чем энергия, пришедшая на поверхность в произвольном направлении. Любая энергия, которая не отражается, будет поглощаться и преобразоваться в тепло. Интегральная величина отраженного единичного светового излучения называется коэффициентом направленного отражения по полусфере. Эта величина должна быть меньше или равна единице:

$$\rho(x, \theta_i) = \int_{\Omega_o} f_r(x, \theta_i, \theta_o) |\theta_o \cdot N_x| d\omega_o, \quad \forall \theta_o \in \Omega_o. \quad (2.11)$$

Коэффициент отражения безразмерен.

## Обратимость Гельмгольца

Другим важным свойством ДФО является обратимость Гельмгольца. Поскольку отражающие свойства поверхности не зависят от направления распространения света, то роли входящего и исходящего направлений могут быть изменены, см. рис. 2.8.

$$f_r(x, \Theta_i, \Theta_o) = f_r(x, \Theta_i^{-1}, \Theta_o^{-1}). \quad (2.12)$$

Модели ДФО, которые сохраняют энергию и являются обратимыми, называются физически корректными [16, 17].

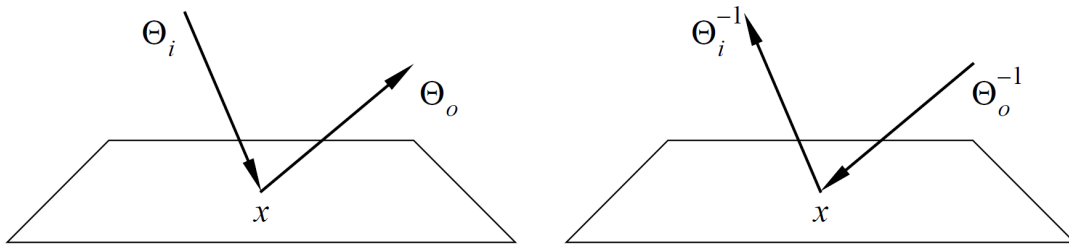


Рис. 2.8. Обратимость Гельмгольца: при замене направлений входящего и исходящего излучений значение ДФО для конкретной точки не меняется

## 2.2 Уравнение рендеринга

В предыдущих разделах было показано, как яркость поля и яркость поверхности вдоль одного направления связаны законом неизменности яркости и как они связаны с локальной ДФО. Объединение двух этих законов дает уравнение рендеринга, которое описывает глобальное отражение света для всей сцены. Впервые это уравнение было введено в компьютерную графику Каджией (в терминах величины, называемой интенсивностью или силой света) [2]. Суть этого уравнения сводится к тому, что излучение, исходящее от поверхности, представляет собой сумму яркости ее самосвечения и отраженной от нее яркости. ДФО описывает отношение между полевой яркостью и отраженной поверхностной яркостью. Используя определение ДФО (2.10), общая яркость поверхности может быть записана следующим образом:

$$\begin{aligned} L_o(x, \Theta_o) &= L_e(x, \Theta_o) + L_r(x, \Theta_o) \\ &= L_e(x, \Theta_o) + \int_{\Omega_i} L_i(x, \Theta_i) f_r(x, \Theta_i, \Theta_o) |\Theta_i \cdot N_x| d\omega_i. \end{aligned}$$



Поскольку каждому значению яркости поля соответствует значение яркости поверхности, выраженное уравнением (2.4), то последнее уравнение можно переписать как

$$L_o(x, \Theta_o) = L_e(x, \Theta_o) + \int_{\Omega_i} L_o(y, \Theta_i) f_r(x, \Theta_i, \Theta_o) |\Theta_i \cdot N_x| d\omega_i,$$

где  $y$  – точка на источнике яркости, т. е.  $y = r(x, \Theta_i^{-1})$ , см. рис. 2.9. Далее будем говорить только о яркости поверхности, поэтому индекс « $o$ » будет опущен.

Кроме того, полусфера входного освещения будет обозначаться через « $\Omega^{-1}$ ». Тогда последнее уравнение преобразуется к виду:

$$L(x, \Theta_x) = L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} L(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y. \quad (2.13)$$

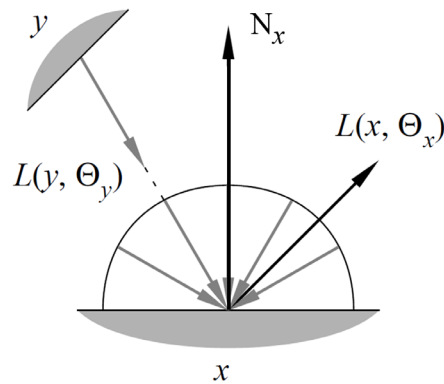


Рис. 2.9. Принципиальная схема формирования яркости в уравнении рендеринга и символы, используемые в этом уравнении

Это уравнение рендеринга, выраженное в терминах яркости, иногда называемое уравнением яркости. Это линейное интегральное уравнение Фредгольма второго рода. Однако, как отметил Арво [18], это особенно важно в том смысле, что область интегрирования – это не вся область функции яркости. Область интегрирования состоит из набора точек (областей), и для каждой из точек имеется единственное направление.

Формально уравнение рендеринга решает глобальную проблему освещения. Как правило, источники света определяются яркостью их самосвечения, отражающими свойствами поверхностей сцены, описываемыми ДФО, и геометрией сцены, описываемой локальными нормальными в точках формирования яркости. Все это позволяет однозначно определить результирующую функцию яркости поверхности.

Уравнение рендеринга может быть записано с использованием оператора интегрирования:

$$L = L_e + TL, \quad (2.14)$$

где интегральный оператор  $T$  определяется следующим образом:

$$(Tf)(x, \Theta_x) = \int_{\Omega_x^{-1}} f(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y.$$

Оператор преобразует функцию поверхностной яркости  $L(y, \Theta_y)$  в функцию поверхностной яркости  $L(x, \Theta_x)$ , которая является результирующей яркостью после однократного отражения. Гершбейн и др. [19] и Арво [18] показали, что оператор переноса может быть разделен на два оператора: оператор локального отражения, который учитывает локальное рассеивание падающего светового излучения, и оператор полевой яркости, который учитывает геометрию сцены.

Для определенных задач интеграл (2.13) удобно представить в виде интеграла по всем поверхностям, являющимися источниками яркости, вместо интеграла по все полусфере. Это достигается с помощью простого преобразования:

$$\begin{aligned} L(x, \Theta_x) &= L_e(x, \Theta_x) \\ &+ \int_A L(y, \Theta_{y \rightarrow x}) f_r(x, \Theta_{y \rightarrow x}, \Theta_x) v(x, y) \frac{|\Theta_{y \rightarrow x} \cdot N_x| |\Theta_{y \rightarrow x} \cdot N_y|}{\|x - y\|^2} d\mu \\ &= L_e(x, \Theta_x) + \int_A L(y, \Theta_{y \rightarrow x}) f_r(x, \Theta_{y \rightarrow x}, \Theta_x) G(x, y) d\mu, \end{aligned} \quad (2.15)$$

где:

- $\Theta_{y \rightarrow x}$  – направление от точки  $y$  к точке  $x$ ,
- $v(x, y)$  – функция видности, равна единице, если точка  $x$  видна непосредственно из точки  $y$ , иначе она равна нулю,
- $\|x - y\|$  – расстояние от точки  $y$  до точки  $x$ ,
- $G(x, y) = v(x, y) \frac{|\Theta_{y \rightarrow x} \cdot N_x| |\Theta_{y \rightarrow x} \cdot N_y|}{\|x - y\|^2}$  – так называемый геометрический форм фактор.

### 2.3 Уравнение потенциала

В этом разделе будет представлено определение потенциала и указаны его наиболее важные свойства. Из определения потенциала вытекает выражение для светового потока в терминах потенциала. Затем будет выведено уравнение

потенциала. Структура изложения данного раздела аналогична изложению понятий яркости и уравнения рендеринга, поскольку они являются аналогами понятиям потенциала и уравнения потенциала. В результате будет показано, что уравнение рендеринга и уравнение потенциала являются сопряженными.

Потенциалом  $W$  для заданной точки  $x$  и направления  $\Theta_x$  является световой поток, проходящий через заданную область  $S$ , полученный в результате излучения света источником единичной яркости с элементарной области вокруг точки  $x$  и в направлении  $\Theta_x$ , см. рис. 2.10. Формально это можно записать как

$$W(x, \Theta_x) = \frac{\partial^2 \Phi_{output}}{L_{input(x, \Theta_x)} \partial \omega_x \partial A_{x\perp}}, \quad (2.16)$$

где:

- $\partial^2 \Phi_{output}$  – дифференциальный световой поток через заданную область  $S$ , сформированный как результат входной яркости  $L_{input(x, \Theta_x)}$  через дифференциальную область  $\partial A_{x\perp}$  в точке  $x$ , телесном угле  $\partial \omega_x$  и направлении  $\Theta_x$ .

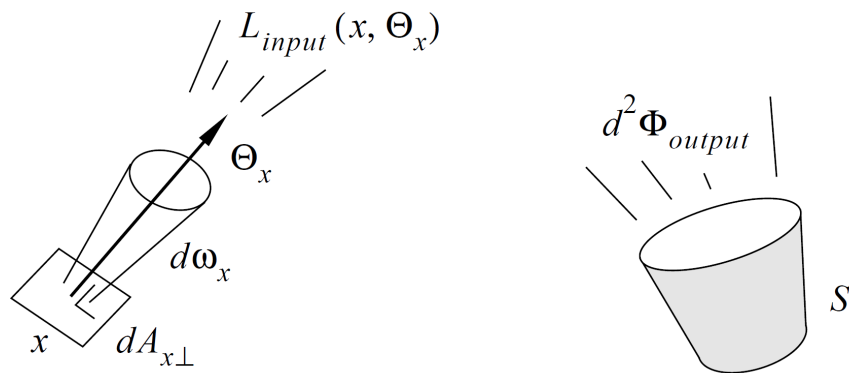


Рис. 2.10. Визуальное представление символов, используемых в определении потенциала. Область пространства и направлений  $S$  может рассматриваться как детектор дифференциального потока, который собирает яркость, излучаемую дифференциальной областью вокруг точки  $x$  и в направлении  $\Theta_x$ . Полученное отношение является потенциалом для точки  $x$  в направлении  $\Theta_x$

Потенциал безразмерен. С точки зрения физического смысла потенциал – это доля светового потока, испущенного с элементарной площадки  $\partial A_{x\perp}$  в направлении  $\Theta_x$  из точки  $x$  в элементарном телесном угле  $\partial \omega_x$ , которую поймала область пространства  $S$ . Как и в случае с функцией яркости, функция потенциала постоянна вдоль заданного направления при распространении света в вакууме.

Следует обратить внимание, что функция потенциала не является потенциальной функцией в обычном физическом смысле, но ради преемственности это название будет использоваться далее.

Особый интерес функция потенциала представляет для областей (и направлений) на излучающих поверхностях сцены. В этом случае уравнение потенциала может быть записано следующим образом:

$$W(x, \Theta_x) = \frac{\partial^2 \Phi_{output}}{L_{input(x, \Theta_x)} |\Theta_x \cdot N_x| \partial \omega_x \partial \mu_x}, \quad (2.17)$$

Используя это определение, световой поток через данную область  $S$ , пришедший от самосветящихся поверхностей, имеющих яркость  $L_e$ , может быть выражен в терминах потенциала:

$$\Phi = \int_A \int_{\Omega_x} L_e(x, \Theta_x) W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x, \quad (2.18)$$

Используя аналогичный оператор скалярного произведения, см. выражение (2.9) для функции яркости, можно написать:

$$\Phi = \langle L_e, W \rangle. \quad (2.19)$$

Точно так же, как функция яркости определяется уравнением рендеринга, функция потенциала определяется интегральным уравнением, называемым уравнением потенциала. Потенциал для точки  $x$  и направления  $\Theta_x$  также является суммой двух слагаемых. Если  $(x, \Theta_x)$  принадлежит заданной области  $S$ , любое излучение будет вносить непосредственный вклад в поток. Это может быть учтено с помощью первого члена  $W_e(x, \Theta_x)$ , который в случае попадания точки в область будет равен единице, а в противном случае нулю. Данное условие было определено в уравнении (2.8). Точка  $x$  и направление  $\Theta_x$  также могут вносить косвенный вклад в поток через отражение, которое может быть выражено в терминах значения потенциала в ближайшей точке от точки  $x$  вдоль направления  $\Theta_x$ , см. рис. 2.11. Формально:

$$\begin{aligned} W(x, \Theta_x) &= W_e(x, \Theta_x) + W_r(x, \Theta_x) \\ &= W_e(x, \Theta_x) + \int_{\Omega_y} W(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y, \end{aligned} \quad (2.20)$$

где  $y = r(x, \Theta_x)$ . По аналогии с яркостью самосветящейся поверхности  $L_e(x, \Theta_x)$  функция  $W_e(x, \Theta_x)$  может рассматриваться как потенциал самосвечения.

Уравнение потенциала может быть записано с помощью оператора интегрирования:

$$W = W_e + T^*W, \quad (2.21)$$

где интегральный оператор  $T^*$  определяется следующим образом:

$$(T^*f)(x, \Theta_x) = \int_{\Omega_y} f(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y.$$

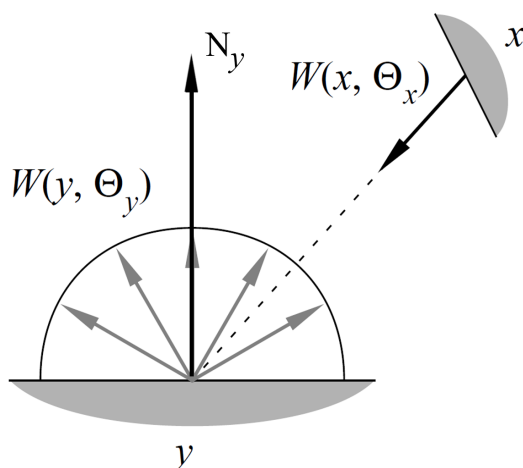


Рис. 2.11. Принципиальная схема, объясняющая уравнение потенциала и символы, используемые в этом уравнении

Аналогично уравнению рендеринга, уравнение потенциала может быть сведено к интегрированию по поверхностям вместо интегрирования по всем направлениям полусферы.

$$\begin{aligned} W(x, \Theta_x) &= W_e(x, \Theta_x) \\ &+ \int_A W(y, \Theta_{y \rightarrow z}) f_r(y, \Theta_x, \Theta_{y \rightarrow z}) v(y, z) \frac{|\Theta_{y \rightarrow z} \cdot N_y| |\Theta_{y \rightarrow z} \cdot N_z|}{\|y - z\|^2} d\mu_z \\ &= W_e(x, \Theta_x) + \int_A W(y, \Theta_{y \rightarrow z}) f_r(y, \Theta_x, \Theta_{y \rightarrow z}) G(y, z) d\mu_z, \end{aligned} \quad (2.22)$$

### **Сопряженность уравнения рендеринга с уравнением потенциала**

Для двух операторов  $T$  и  $T^*$  были введены соответствующие уравнения:

$$\begin{aligned} f &= f_e + Tf \\ g &= g_e + T^*g \end{aligned}$$

Они называются сопряженными относительно скалярного произведения  $\langle \dots \rangle$  если

$$\langle f, g_e \rangle = \langle f_e, g \rangle$$

для любого  $f_e$  и  $g_e$ . Следовательно, уравнение рендеринга и уравнение потенциала (или, точнее: их соответствующие операторы) сопряжены, поскольку их соответствующие выражения (2.8) для яркости и (2.18) для светового потока  $\Phi$  должны давать одинаковые результаты для любого множества  $S$  (т.е. для любой функции  $W_e(x, \Theta_x)$ ) и для любой функции яркости самосветящейся поверхности  $L_e(x, \Theta_x)$ :

$$\Phi = \langle L, W_e \rangle = \langle L_e, W \rangle.$$

## 2.4 Глобальная функция отражения

В предыдущей секции было показано, что функция яркости  $L(x, \Theta_x)$  определяется через функцию яркости самосвечения  $L_e(x, \Theta_x)$ , а функция потенциала  $W(x, \Theta_x)$  определяется через фиксированную функцию  $W_e(x, \Theta_x)$ , определенную в некоторой области  $S$ . Соотношение между функциями  $L(x, \Theta_x)$  и  $W(x, \Theta_x)$  не было определено математическими выражениями, что затрудняет их объединение в единую модель или алгоритм. В работе [20] было предложено понятие глобальной функции отражения – *global reflectance distribution function* (GRDF), которая определяется для данной сцены, но не зависит ни от конкретных функций излучения, таких как  $L_e(x, \Theta_x)$ , ни от фиксированной области, такой как  $S$ .

Аналогично терминам яркости и потенциала, рассмотренным в предыдущих разделах, в этом разделе будет представлено определение GRDF и определяющего интегрального уравнения. Кроме того, будет представлено выражение для светового потока в терминах GRDF.

### Определение

Это определение имеет некоторое сходство с подходом, часто используемым в теории систем, где система изучается на основе ее реакции на некоторый входной сигнал. GRDF или  $F_r$  для точек и направлений  $(x, \Theta_x)$  и  $(y, \Theta_y)$  выражает дифференциальную величину яркости, покидающей точку  $y$  по направлению  $\Theta_y$ , появившуюся в результате излучения источника единичной яркости с поверхности в точке  $x$  по направлению  $\Theta_x$ , см. рис. 2.12. Формально:

$$\begin{aligned}
F_r(x, \Theta_x, y, \Theta_y) &= \frac{\partial^2 L_{output}(y, \Theta_y)}{L_{input}(x, \Theta_x) \partial \omega_x \partial A_{x\perp}} \\
&= \frac{\partial^2 L_{output}(y, \Theta_y)}{L_{input}(x, \Theta_x) \partial \omega_x |\Theta_i \cdot N_x| \partial \mu_x}.
\end{aligned}
\tag{2.23}$$

Глобальная функция отражения может рассматриваться как общий случай ДФО. Последняя функция определяет только локальное поведение света, отраженного от отдельной поверхности, а GRDF определяет глобальные эффекты освещения, являющиеся результатом отражения от всей сцены. Единицами измерения GRDF являются  $1/m^2 sr$ . Поскольку мы предполагаем, что среда между точками  $x$  и  $y$  отсутствует, GRDF постоянна при перемещении точки  $x$  вдоль направления  $\Theta_x$  и точки  $y$  вдоль направления  $\Theta_y$ . Это предположение исходит из того, что функции входного и выходного излучений инвариантны относительно этих направлений.

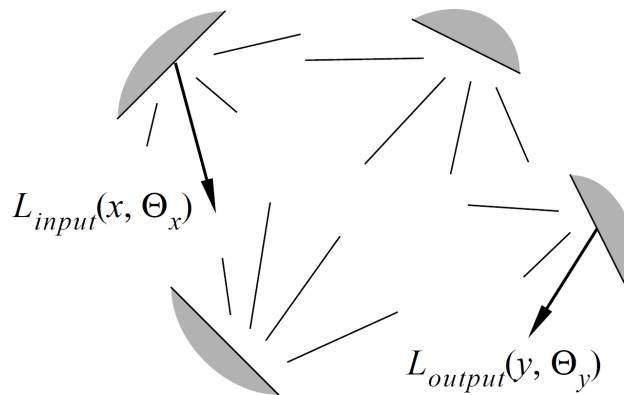


Рис. 2.12. Глобальная функция отражения дает меру доли яркости от яркости некоторой поверхности в точке  $x$  вдоль направления  $\Theta_x$ , которая будет отражена некоторой поверхностью в точке  $x$  по направлению  $\Theta_y$  после всех возможных переотражений, возникающих в сцене

#### 2.4.1 Выражение светового потока в терминах GRDF

Теперь, когда GRDF была определена, ее можно использовать для определения светового потока через заданную область пространства координат и направлений  $S$  (определенную функцией  $W_e(x, \Theta_x)$ ) и функцию самосвечения  $L_e(x, \Theta_x)$ . Из выражения (2.23) и выражения для светового потока в терминах яркости (2.8) можно получить:

$$\Phi = \int_{A_x} \int_{\Omega_x} \int_{A_y} \int_{\Omega_y} L_e(x, \Theta_x) W_e(y, \Theta_y) F_r(x, \Theta_x, y, \Theta_y) \cdot |\Theta_x \cdot N_x| |\Theta_y \cdot N_y| d\mu_y d\omega_y d\mu_x d\omega_x. \quad (2.24)$$

Необходимо отметить, что только в этом выражении функции  $L_e(x, \Theta_x)$  и  $W_e(y, \Theta_y)$  появляются вместе. При этом сама GRDF полностью не зависит от них. При необходимости можно выразить функцию яркости относительно заданной  $L_e(x, \Theta_x)$  и функцию потенциала относительно заданного  $W_e(y, \Theta_y)$  через GRDF:

$$L(x, \Theta_x) = \int_{A_x} \int_{\Omega_x} L_e(x, \Theta_x) F_r(x, \Theta_x, y, \Theta_y) |\Theta_x \cdot N_x| d\mu_x d\omega_x,$$

$$W(y, \Theta_y) = \int_{A_y} \int_{\Omega_y} W_e(y, \Theta_y) F_r(x, \Theta_x, y, \Theta_y) |\Theta_y \cdot N_y| d\mu_y d\omega_y.$$

## 2.4.2 Выражения, определяющие GRDF

GRDF задается набором из двух интегральных уравнений. Во-первых, можно посмотреть на поведение функции при фиксированной паре  $(x, \Theta_x)$ . Аналогично выводу уравнения рендеринга, яркость для фиксированной пары  $(y, \Theta_y)$  является результатом вкладов двух составляющих. Если обе пары точек и направлений равны, то входная яркость вносит непосредственный вклад в выходную яркость. В этом случае выходная яркость не дифференцируема, но ограничена, поэтому ее доля в общем интеграле есть импульс Дирака. Вторая составляющая представляет собой вклад отражений света в точке  $y$ , который зарождается из точек  $z$ , например, в результате многократного отражения света в сцене, см. рис. 2.13. Формально – это выражение может быть представлено в виде:

$$F_r(x, \Theta_x, y, \Theta_y) = \int_{\Omega_y^{-1}} f_r(y, \Theta_z, \Theta_y) F_r(x, \Theta_x, z, \Theta_z) |\Theta_z \cdot N_y| d\omega_z + \delta(x, \Theta_x, y, \Theta_y), \quad (2.25)$$

где  $z = r(y, \Theta_z^{-1})$ . Функция  $\delta(x, \Theta_x, y, \Theta_y)$  – импульс Дирака, которая равна нулю в области  $(x, \Theta_x) \neq (y, \Theta_y)$ , но ее интеграл равен единице по всей области  $(x, \Theta_x)$  для заданных значений  $(y, \Theta_y)$  и наоборот:

$$\int_A \int_{\Omega_x} \delta(x, \Theta_x, y, \Theta_y) |\Theta_x \cdot N_x| d\omega_x d\mu_x = 1,$$

$$\int_A \int_{\Omega_y} \delta(x, \Theta_x, y, \Theta_y) |\Theta_y \cdot N_y| d\omega_y d\mu_y = 1.$$



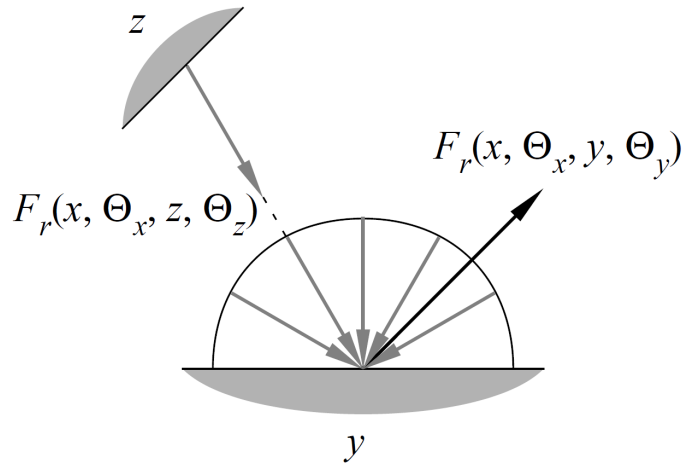


Рис. 2.13. Символы, используемые в уравнении (2.25), определяющем GRDF. Это уравнение аналогично уравнению рендеринга

В качестве альтернативного определения функции GRDF можно посмотреть на ее поведение при фиксированном  $(y, \Theta_y)$ . Это поведение напоминает вид уравнения потенциалов. Вопрос в том, как яркость для  $(x, \Theta_x)$  может повлиять на выходную яркость. Напомним, что если обе пары точек и направлений равны, то входная яркость вносит непосредственный вклад в выходную яркость, которая выражается импульсом Дирака. Входная яркость также может вносить вклад в яркость вторичного излучения, посредством отражения в произвольных направлениях вокруг точки  $z$ , которая видна из точки  $x$  в направлении  $\Theta_x$ , см. рис. 2.14. В результате мы имеем второе рекурсивное уравнение:

$$F_r(x, \Theta_x, y, \Theta_y) = \int_{\Omega_z} f_r(z, \Theta_x, \Theta_z) F_r(z, \Theta_z, y, \Theta_y) |\Theta_z \cdot N_z| d\omega_z + \delta(x, \Theta_x, y, \Theta_y), \quad (2.26)$$

где  $z = r(y, \Theta_z^{-1})$ .

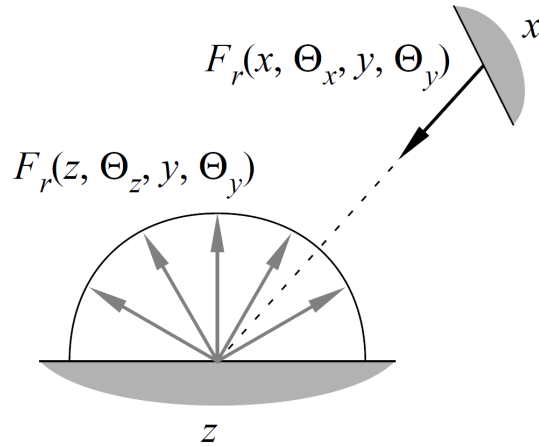


Рис. 2.14. Символы, используемые в уравнении (2.26), определяющем GRDF. Это уравнение аналогично уравнению потенциалов

Видно, что уравнения (2.25) и (2.26) полностью эквивалентны в силу двунаправленной природы распространения света (принцип обратимости световых лучей). Доля излучения, испускаемого из одной точки сцены, которая принимается другой точкой сцены, равна доле, которая была бы получена в этих точках, если бы роли излучателя и приемника поменялись местами. Таким образом, замена траекторий прямых лучей на обратные и наоборот сохраняет коэффициент передачи излучения. Это свойство может быть выражено в терминах GRDF:

$$F_r(x, \Theta_x, y, \Theta_y) = F_r(\tilde{y}, \Theta_y^{-1}, \tilde{x}, \Theta_x^{-1}),$$

где  $\tilde{x} = r(x, \Theta_x)$  и  $\tilde{y} = r(y, \Theta_y)$  см. рис. 2.15. Это выражение тождественно выражению (2.12), определяющему принцип обратимости Гельмгольца для ДФО.

Подстановка этой зависимости в обе части второго интегрального уравнения (2.26) дает:

$$\begin{aligned} F_r(\tilde{y}, \Theta_y^{-1}, \tilde{x}, \Theta_x^{-1}) &= \delta(x, \Theta_x, y, \Theta_y) \\ &+ \int_{\Omega_z} f_r(z, \Theta_z^{-1}, \Theta_x^{-1}) F_r(\tilde{y}, \Theta_y^{-1}, \tilde{z}, \Theta_z^{-1}) |\Theta_z \cdot N_z| d\omega_z. \end{aligned}$$

Соответствующая замена переменных приводит к первому интегральному уравнению (2.25). Это говорит о том, что применение принципа обратимости Гельмгольца к любому из уравнений этой пары будет достаточным условием для однозначного определения GRDF.

Использование операторов позволит представить интегральные уравнения в короткой форме. Использование оператора  $T$ , примененного к уравнению рендеринга (2.14), позволяет представить уравнение (2.25) в виде:

$$F_r(\tilde{y}, \Theta_y^{-1}, \tilde{x}, \Theta_x^{-1}) = \delta(x, \Theta_x, y, \Theta_y) + TF_r(x, \Theta_x, y, \Theta_y), \quad (2.27)$$

где оператор  $T$  применяется к переменным  $y$  и  $\Theta_y$ , рассматривая  $x$  и  $\Theta_x$  как константы. Используя аналогичный оператор  $T^*$  к уравнению потенциала (2.21), выражение (2.26) может быть записано в виде:

$$F_r(\tilde{y}, \Theta_y^{-1}, \tilde{x}, \Theta_x^{-1}) = \delta(x, \Theta_x, y, \Theta_y) + T^*F_r(x, \Theta_x, y, \Theta_y), \quad (2.28)$$

где, в данном случае, оператор  $T^*$  применяется к переменным  $x$  и  $\Theta_x$ , рассматривая  $y$  и  $\Theta_y$  как константы.

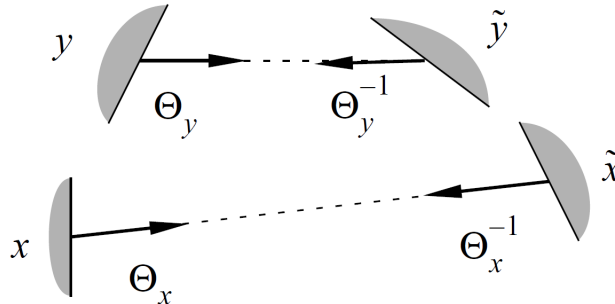


Рис. 2.15. Символы, используемые в принципе обратимости GRDF

Импульс Дирака, появляющийся в двух интегральных уравнениях, может показаться немного неудобным для практической реализации в алгоритмах. Однако надо помнить, что задача интегрирования GRDF сводится к вычислению некоторого светового потока, и это является скорее преимуществом, чем недостатком, поскольку вычисление интеграла импульса Дирака достаточно простое в силу своего определения.

В результате GRDF была определена в терминах исходящего входного излучения и исходящего выходного излучения. Четыре альтернативные комбинации входящего/исходящего входного/выходного излучения приводят к различным определениям и различным наборам двух основных интегральных уравнений. Для конкретных задач некоторые из них могут иметь большую практическую значимость, чем другие.

## 2.5 Каустика

До настоящего момента мы рассматривали случаи непосредственного наблюдения и освещения поверхностей сцены. Однако сцена может содержать «зеркальные» поверхности отражения и пропускания, а также среды с градиентом показателя преломления. Это поверхности, на которых не происходит рассеивание света, и яркость излучающего объекта передается на выходной объект непосредственно, как если бы источник яркости оказался в некоторой смещенной позиции. Кроме того, яркость исходного объекта снижается в соответствии с коэффициентом отражения (пропускания) «зеркального» объекта или за счет поглощения в среде (2.6).

Принципиальных изменений уравнения рендеринга и потенциала не претерпевают. Просто необходимо помнить, что яркость, используемая в этих выражениях, есть приведенная яркость, т.е. яркость, приведенная к вакууму (2.5). Кроме того, перенос яркости из точки  $y$  в точку  $x$  сопровождается некоторой потерей  $\tau(y, \Theta_y, x)$ . Поэтому уравнение рендеринга (2.13) принимает следующий вид:

$$L(x, \Theta_x) = L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} \tau(y, \Theta_y, x) L(y, \Theta_y) f_r(x, \Theta_{y'}, \Theta_x) |\Theta_{y'} \cdot N_x| d\omega_y, \quad (2.29)$$

где  $\Theta_{y'}$  – направление светового луча, вышедшего из локального источника света в точке  $y$  по направлению  $\Theta_y$  и пришедшего в точку  $x$  с новым направлением  $\Theta_{y'}$ . Изменение направления возможно в результате ряда зеркальных преобразований или искривления траектории луча в градиентной среде.

При переходе от приведенных яркостей к яркостям в вакууме уравнение рендеринга примет следующий вид:

$$\frac{L(x, \Theta_x)}{n^2(x)} = \frac{L_e(x, \Theta_x)}{n^2(x)} + \int_{\Omega_x^{-1}} \tau(y, \Theta_y, x) \frac{L(y, \Theta_y)}{n^2(y)} f_r(x, \Theta_{y'}, \Theta_x) |\Theta_{y'} \cdot N_x| d\omega_y, \quad (2.30)$$

где  $n(x)$  – показатель преломления среды в точке сбора яркости,  $n(y)$  – в точках, формирующих яркость (яркость поля).

## 2.6 Заключение

Наиболее важными радиометрическими величинами являются яркость и световой поток. Выражения для этих величин могут иметь как интегральное, так и спектральное представление в зависимости от того, рассматривается ли весь видимый (или заданный) спектр длин волн или только одна длина волны. Функция яркости описывает долю светового потока, излучаемого с области пространства в заданном направлении, отнесенного к этой элементарной области

пространства и элементарному телесному углу вокруг заданного направления. По большей части интерес представляет поверхностное излучение (выходящее из поверхностей) и полевое излучение (падающее на поверхности). Исходя из предположения, что только поверхности преобразуют световое излучение и что среда не рассеивает свет, яркость (в общем случае приведенная яркость) инвариантна вдоль ее направления. Этот инвариант напрямую связывает поверхностную и полевую яркости.

Базовым положением для описания проблемы глобального освещения является локальное освещение. Локальные отражающие свойства поверхности описываются функцией двунаправленного отражения (ДФО). Эта функция определяет соотношение между светом, падающим на поверхность, и светом, выходящим с поверхности в заданном направлении, т.е. эта функция связывает полевую и поверхностную яркости.

Объединяя выражения для яркости поля и поверхностной яркости, получаем уравнение рендеринга (2.13). Это интегральное уравнение определяет значение поверхностной яркости как функции самосветящейся поверхности, геометрии сцены и двунаправленной функции отражения. Вместе с выражением для светового потока в терминах яркости (2.8) это дает первую математическую формализацию проблемы глобального освещения:

$$\Phi = \int_A \int_{\Omega_x} L(x, \Theta_x) W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x. \quad (2.8)$$

$$L(x, \Theta_x) = L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} L(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y. \quad (2.13)$$

Функция потенциала дает меру доли света, излучаемого элементарной областью пространства в заданном направлении, которая в итоге достигает некоторого детектора, собирающего свет с некоторой области и направлений (в телесном угле). Как и в случае с яркостью, потенциал может быть описан с помощью уравнения потенциала (2.20), которое имеет вид, аналогичный уравнению рендеринга. Вместе с выражением для светового потока в терминах потенциала (2.18) это дает альтернативное описание проблемы глобального освещения:

$$\Phi = \int_A \int_{\Omega_x} L_e(x, \Theta_x) W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x. \quad (2.18)$$

$$\begin{aligned} W(x, \Theta_x) &= W_e(x, \Theta_x) + W_r(x, \Theta_x) \\ &= W_e(x, \Theta_x) + \int_{\Omega_x} W(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y. \end{aligned} \quad (2.20)$$

Была введена функция глобального отражения (GRDF), позволяющая оценить долю света, излучаемого объектами сцены во всех направлениях, который выходит с заданной области пространства (поверхности) в заданном направлении. Это десятимерная функция пар всех комбинаций точек и направлений в сцене. GRDF зависит только от геометрии сцены и ее отражающих свойств. Она не зависит ни от функции яркости самосветящейся поверхности (несмотря на то, что функция яркости зависит от нее), ни от функции потенциала самосветящейся поверхности (несмотря на то, что функция потенциала зависит от нее). Ее поведение описывают два эквивалентных интегральных уравнения (2.25) и (2.26), которые соответствуют уравнению рендеринга и уравнению потенциала соответственно. Выражение светового потока в терминах GRDF (2.24) содержит коэффициенты, зависящие от яркости самосветящейся поверхности и потенциала. Все вместе это образует третье представление проблемы глобального освещения:

$$\Phi = \int_{A_x} \int_{\Omega_x} \int_{A_y} \int_{\Omega_y} L_e(x, \Theta_x) W_e(y, \Theta_y) F_r(x, \Theta_x, y, \Theta_y) \cdot |\Theta_x \cdot N_x| |\Theta_y \cdot N_y| d\mu_y d\omega_y d\mu_x d\omega_x. \quad (2.24)$$

$$F_r(x, \Theta_x, y, \Theta_y) = \int_{\Omega_y^{-1}} f_r(y, \Theta_z, \Theta_y) F_r(x, \Theta_x, z, \Theta_z) |\Theta_z \cdot N_y| d\omega_z + \delta(x, \Theta_x, y, \Theta_y). \quad (2.25)$$

$$F_r(x, \Theta_x, y, \Theta_y) = \int_{\Omega_z} f_r(z, \Theta_x, \Theta_z) F_r(z, \Theta_z, y, \Theta_y) |\Theta_z \cdot N_z| d\omega_z + \delta(x, \Theta_x, y, \Theta_y). \quad (2.26)$$

Далее будет показано, что эти математические модели представляют не только теоретический интерес. Применение одних и тех же стандартных математических методов к каждой из них приведет к различным алгоритмам рендеринга.

## 2.7 Вопросы для самоконтроля усвоенных знаний

Следующие вопросы позволят оценить степень усвоения знаний по разделу радиометрия.

1. Дайте определение и объясните физический смысл понятиям входной и выходной яркостей. Докажите закон сохранения яркости при передаче светового излучения в непоглощающей среде с постоянным показателем преломления.

2. Что такое приведенная яркость, какой вид имеет закон сохранения яркости в случае приведенной яркости.
3. Определите связь светового потока и яркости. В первом случае необходимо вычислить световой поток  $\Phi$  от плоского прямоугольного источника света размером  $a \times b$  с постоянным распределением яркости  $L_0$  по поверхности источника света. Во втором случае необходимо вычислить световой поток  $\Phi$  от плоского прямоугольного источника света размером  $a \times b$  с постоянным законом распределения яркости  $L = L_0(\cos \alpha)^\beta$  по поверхности источника света, где  $\alpha$  – это угол между направлением измерения яркости и нормалью к поверхности источника света.
4. Решите обратную задачу. Плоский прямоугольный источник света размером  $a \times b$  имеет световой поток  $\Phi$ . Известен угловой закон распределения яркости  $L \sim (\cos \alpha)^\beta$  и предполагается, что яркость распределена по поверхности источника света равномерно. Необходимо найти множитель пропорции  $L_0$  для восстановления закона  $L = L_0(\cos \alpha)^\beta$ .
5. В чем физический смысл ДФО. Как связать ДФО с двунаправленным коэффициентом яркости поверхности, если последний есть отношение яркости в направлении наблюдения к яркости идеального ламбертовского отражателя при тех же условиях освещения.
6. Какие два основных принципа присущи ДФО. Как на основании этих физических принципов можно проверить физическую корректность модели ДФО. Напишите псевдокод программы проверки физической корректности ДФО.
7. В чем физический смысл уравнения рендеринга. Какие основные проблемы возникают при его решении. Что такое яркость прямого видения и яркость прямого освещения, как их можно вычислить, напишите псевдокод программы. Что такое яркость глобального освещения. Какие проблемы возникают при расчете яркости глобального освещения.
8. В чем физический смысл светового потенциала, какова размерность потенциала.
9. Какой вид имеет интегральное уравнение потенциала, в чем его физический смысл. В чем отличие и схожесть уравнений рендеринга и потенциала.
10. В чем физический смысл глобальной функции отражения. В чем отличие и схожесть глобальной функции отражения и ДФО. Как выражается световой поток через глобальную функцию отражения.
11. Что такое каустическое освещение и в чем проблема расчета яркости каустического освещения. В каких случаях и как яркость каустического

освещения может быть вычислена как яркость прямого освещения. Какие искажения возможны при упрощенном расчете яркости каустического освещения.

12. Решите следующую задачу. Наблюдатель находится в воздухе над бассейном с водой (без ряби), имеющей показатель преломления  $n$ . На дне бассейна находится равномерный источник света с яркостью  $L_0$ . Наблюдатель смотрит на источник света вертикально вниз (под углом  $90^\circ$  к поверхности бассейна и источнику света). Какую яркость увидит наблюдатель. Как яркость зависит от положения наблюдателя.

### 3. Детерминистические методы решения уравнения рендеринга

#### 3.1 Интегральное уравнение рендеринга

Уравнение рендеринга представляет собой интегральное уравнение, где искомая функция яркости  $f(y)$  находится под знаком интеграла. Это интегральное уравнение Фредгольма второго рода:

$$f(x) = g(x) + \int_a^b K(x, y)f(y)dy, \quad (3.1)$$

где  $K(x, y)$  является ядром интегрального уравнения, определяющим его свойства и методы его решения.

Для численного решения интегральных уравнений используются следующие методы [7]:

- Метод итераций. Может быть применен, когда  $K(x, y) < 1$ . Реализуется в виде рекуррентной формулы вида:

$$f(x) = \sum_{i=0}^{\infty} K_i g(x), \quad K_i g(x) = \int_a^b K_{i-1} K(x, y) g(x) dy, \quad K_0 g(x) = g(y).$$

Надо отметить, что при расчете глобальной освещенности методом переноса излучения в сцене, содержащей диффузные поверхности, ядро интегрального уравнения меньше единицы.

- Факторизация ядра. Метод применим, если ядро интегрального уравнения вырожденное:

$$K(x, y) = \sum_{i=1}^{\infty} \psi_i(x)\varphi_i(y).$$



Это приводит к следующему виду интегрального уравнения

$$f(x) = g(x) + \sum_{i=1}^{\infty} \psi_i(x) \int_a^b \varphi_i(y) f(y) dy,$$

что в итоге сводится к системе линейных уравнений.

- Замена интеграла квадратурой. Метод является наиболее общим детерминистическим методом решения проблемы глобального освещения. Метод предполагает следующее приближение:

$$\int_a^b K(x, y) f(y) dy = \sum_{i=0}^{\infty} K(x, x_i) f(x_i) w_i,$$

где  $w_i$  – вес соответствующей квадратуры. В результате интегральное уравнение сводится к решению следующей системы уравнений:

$$f_k = g_k + \sum_{i=1}^{\infty} K(x_k, x_i) f(x_i) w_i,$$

где  $f_k \equiv f(x_k)$ ,  $g_k \equiv g(x_k)$ .

### 3.2 Решение уравнения рендеринга для диффузного освещения

Практическое решение проблемы глобальной освещенности детерминистическими методами возможно при слабой зависимости двунаправленной функции отражения от направлений освещения и наблюдения. Идеальный случай – это сцена, содержащая поверхности исключительно с законом отражения Ламберта.

Для поверхностей, имеющих ламбертовы свойства отражения (ДФР является постоянной для всех направлений освещения и наблюдения), уравнение рендеринга (15) может быть переписано в виде:

$$M(x) = M_e(x) + \kappa(x) \int_A M(y) G(x, y) d\mu, \quad (3.2)$$

где:

- $\kappa(x)$  – двунаправленная функция диффузного отражения (ДФО Ламбертовой поверхности с некоторой долей поглощения),
- $G(x, y)$  – геометрический форм-фактор сцены, его смысл будет ясен ниже, см. (3.4),
- $M(x) = \pi L(x, \Theta_x)$  – светимость поверхности.

Уравнение (3.2) есть уравнение излучательности (радиосити). Если применить к этому интегральному уравнению метод решения с заменой подынтегрального выражения квадратурой, то мы приходим к системе линейных алгебраических уравнений (СЛАН):

$$M_i = M_{ei} + \kappa_i \sum_{j=1}^K M_j G_{ij}, \quad (3.3)$$

где  $K$  – число элементов сцены (как правило, треугольники или прямоугольники, формирующие геометрию объектов сцены).

Данная система может быть решена двумя основными способами. Это классические решения, а именно инвертирование матрицы или метод гауссова исключения. Однако сложность этих алгоритмов достаточно высока и оценивается как  $O(K^3)$ . Если число элементов сцены достигает сотен или даже десятков тысяч, что очень немного для современной сцены, то алгоритмическая сложность становится неприемлемо высокой. Поэтому на практике используются другие решения. Основной метод решения таких СЛАУ – это итерационный, его алгоритмическая сложность, хоть и высока  $O(K^2)$ , однако она значительно ниже методов инвертирования или гауссова исключения. Существует два основных итерационных решения: это «сборка» и «бросание». Рис. 3.1 иллюстрирует эти методы.

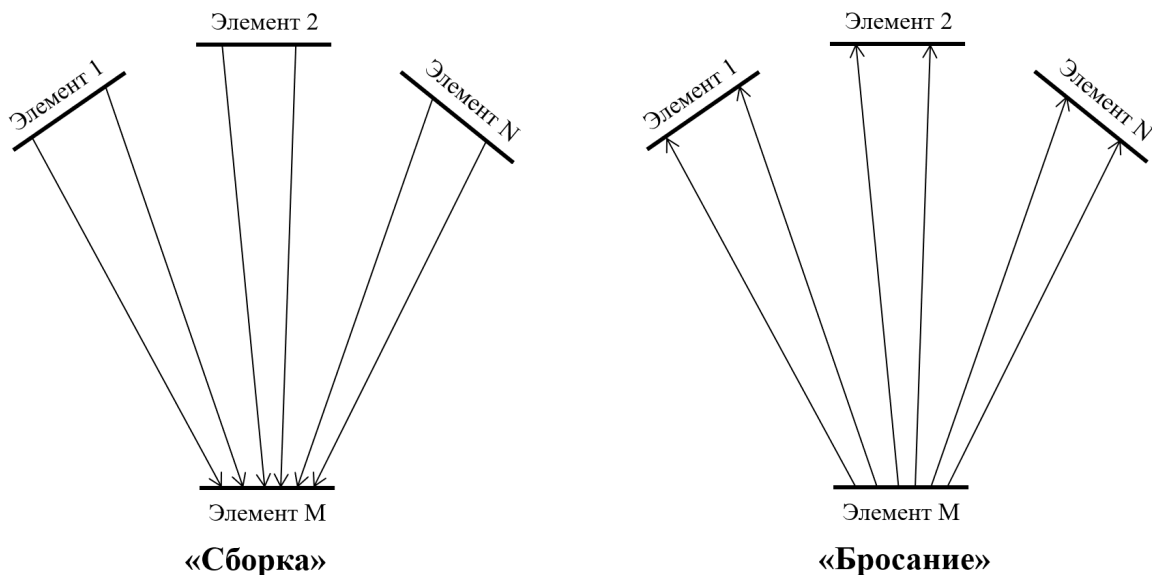


Рис. 3.1. Методы «сборки» (слева) и «бросания» справа

Предполагается, что элементы сцены обладают некоторой яркостью. На первой итерации яркостью обладают только источники света, однако модель может быть сложной, и в сцене могут присутствовать источники, не обладающие яркостью, например, точечные или параллельные источники света. В этой ситуации первичная яркость на данных источниках света вычисляется стандартным образом:

$$L(x) = \kappa(x) \frac{I(\Theta_y)}{|x - y|^2} |\Theta_y \cdot N_x| \text{ – точечный источник света,}$$

$$L(x) = \kappa(x) E(y) |\Theta_y \cdot N_x| \text{ – параллельный источник света,}$$

где:

- $I(\Theta_y)$  – сила света (интенсивность) точечного источника по направлению  $\Theta_y$ ,
- $E(y)$  – плотность светового потока, создаваемая параллельным источником света на площадке, ортогональной направлению излучения.

Поэтому, если сцена содержит параллельные или точечные источники света, то необходим предварительный шаг, создающий дополнительное значение начальной ( $M_{ei}$ ) светимости от данных источников света.

Метод «сборки» – непосредственное использование выражения (3.3). Для каждого элемента  $i$  происходит сбор яркости от всех элементов сцены от 1 до  $K$ , имеющих ненулевую яркость (светимость в данном случае), яркость передается посредством форм-фактора, содержащим значение видности между  $i$  и  $j$  элементами и ДФР ( $\kappa_i$ )  $i$ -го элемента. Сборка повторяется для всех элементов сцены. В результате первой итерации происходит расчет яркости прямого освещения, которая будет являться источником вторичного излучения на следующем шаге итераций. Следующий шаг повторяет предыдущий, при этом яркость на поверхностях постоянно накапливается. Алгоритмически метод «сборки» можно организовать различными способами и в зависимости от реализации метода его сходимость будет меняться. Примеры двух вариантов организации представлены ниже:

```
#define K ... // Число элементов сцены
#define ITERATIONS ... // Число итераций

double me[K] = {...}; // Начальные значения светимости элементов сцены
double g[K][K] = {{...}, ..., {...}}; // Форм-факторы элементов сцены
double k[K] = {...}; // Значения Ламбертовой ДФР элементов сцены
```

<pre>double m[K] = {...}; // Финальные значения светимости элементов сцены // Временные массивы double mc[K] = {...}; double mo[K] = {...};</pre>	
Вариант № 1	Вариант № 2
<pre>for (int i = 0; i &lt; K; i++)     m[i] = me[i]; for (int n = 0; n &lt; ITERATIONS; n++) {     for (int i = 0; i &lt; K; i++)     {         mc[i] = 0;         for (int j = 0; j &lt; K; j++)             if (i != j)                 mc[i] += g[i][j] * m[j];         m[i] = me[i] + k[i] * mc[i];     } }</pre>	<pre>for (int i = 0; i &lt; K; i++)     mo[i] = m[i] = me[i]; for (int n = 0; n &lt; ITERATIONS; n++) {     for (int i = 0; i &lt; K; i++)     {         mc[i] = 0;         for (int j = 0; j &lt; K; j++)             if (i != j)                 mc[i] += g[i][j] * mo[j];     }     for (int i = 0; i &lt; K; i++)     {         mo[i] = k[i] * mc[i];         m[i] += mo[i];     } }</pre>

Метод «бросания» представляет собой обратный процесс по сравнению с методом сбора. Каждый светящийся элемент передает свою яркость всем остальным элементам сцены и тем самым создает освещение, а затем и яркость (посредством ДФР ( $k_i$ ) преобразования освещенности в яркость на каждом  $i$ -ом освещаемом элементе сцены). Результат первой итерации близок к результату первой итерации, полученной методом «сбора». Однако, если последующие операции метода «сбора» достаточно «однообразны», то метод «бросания» более вариативен. Например, можно изменить порядок бросания яркости, начиная с самых ярких элементов сцены. Т.е. в процессе итераций выбирать элемент сцены с наибольшей локальной яркостью, надеясь на то, что наиболее яркий источник создаст большую освещенность и, следовательно, яркость всей сцены. В случае интерактивной визуализации сцены такой подход может оказаться более эффективным. Как и метод «сборки», метод «бросания» можно организовать с помощью различных алгоритмов. Сходимость алгоритмов при этом также будет меняться в зависимости от способов их организации. Примеры двух вариантов организации представлены ниже:

```

#define K ... // Число элементов сцены
#define ITERATIONS ... // Число итераций

double me[K] = {...}; // Начальные значения светимости элементов сцены
double g[K][K] = {{...}, ..., {...}}; // Форм-факторы элементов сцены
double k[K] = {...}; // Значения Ламбертовой ДФР элементов сцены
double m[K] = {...}; // Финальные значения светимости элементов сцены
// Временные массивы
double mc[K] = {...};
double mo[K] = {...};

```

Вариант № 1	Вариант № 2
<pre> for (int i = 0; i &lt; K; i++)     mo[i] = m[i] = me[i]; for (int i = 0; i &lt; K; i++)     {         for (int n = 0; n &lt; ITERATIONS; n++)             {                 int io = n % K;                 // io, например, может быть выбрана                 // из условия максимальной локальной                 // яркости элементов сцены                 for (int j = 0; j &lt; K; j++)                     {                         if (io != j)                             {                                 double dm = k[j] * g[j][io] * mo[io];                                 mo[j] += dm;                                 m[j] += dm;                             }                     }                 mo[io] = 0;             }     } </pre>	<pre> for (int i = 0; i &lt; K; i++)     mo[i] = m[i] = me[i]; for (int n = 0; n &lt; ITERATIONS; n++)     {         for (int i = 0; i &lt; K; i++)             mc[i] = 0;         for (int i = 0; i &lt; K; i++)             {                 for (int j = 0; j &lt; K; j++)                     if (i != j)                         mc[j] += g[j][i] * mo[i];             }         for (int i = 0; i &lt; K; i++)             {                 mo[i] = k[i] * mc[i];                 m[i] += mo[i];             }     } </pre>

Важным моментом является определение числа итераций, необходимых для достижения заданной точности. Корректная оценка точности невозможна, можно только использовать приближенные оценки. Например, можно оценить вариацию

между текущим и предыдущим шагом итерации. Если вариация превышает некоторое заданное значение, то итерации необходимо продолжать. Вариация может оцениваться различными способами, например, среднеарифметическое отклонение между яркостями всех элементов сцены, или соответствующее среднеквадратичное отклонение, или максимальное отклонение и т.п. Другое решение – это оценка недораспределенного потока. Его можно оценить по приближенной формуле. Если средняя коэффициент диффузного отражения сцены можно оценить как

$$\kappa_{ave} = \frac{\sum_{i=1}^k \kappa_i A_i}{\sum_{i=1}^k A_i},$$

где  $A_i$  – площадь  $i$ -го диффузного элемента, то остаточную освещенность можно оценить по формуле:

$$\Delta E = \frac{\sum_{i=1}^k A_i (M_i^{(n)} - M_i^{(n-1)})}{\pi(1 - \kappa_{ave}) \sum_{i=1}^k A_i},$$

где  $M_i^{(n)}$  и  $M_i^{(n-1)}$  – светимости  $i$ -го диффузного элемента на  $(n)$ -ом и  $(n - 1)$ -ом шагах итерации. Оба этих метода являются приближенными и не дают гарантии, что точность была достигнута.

Следующим моментом, снижающим точность расчета глобального освещения, является усреднение яркости на элементах сцены конечного размера. Несмотря на то, что яркость может интерполироваться в пределах элементов сцены и храниться, например, в вершинах соответствующих треугольников или прямоугольников, резкие перепады яркости (например, тени) могут создавать артефакты на элементах большого размера, см. рис. 3.2.

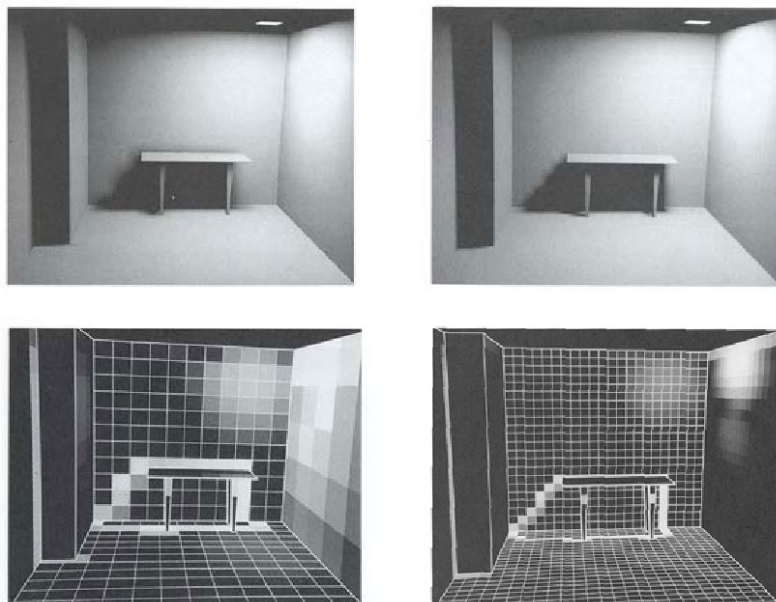


Рис. 3.2. На левом верхнем рисунке видна ступенчатая структура яркости в области тени, на правом верхнем рисунке «ступеньки» значительно меньше. На нижних рисунках показана соответствующая сетка элементов сцены

Повышение детализации элементов сцены (уменьшение размера ее элементов) может повлиять на качество изображения, однако это значительно замедляет расчет и увеличивает требования к оперативной памяти, необходимой для хранения распределения яркости сцены. Поэтому можно использовать адаптивное разбиение геометрии. Рис. 3.3 демонстрирует адаптивное решение для представления геометрии сцены.

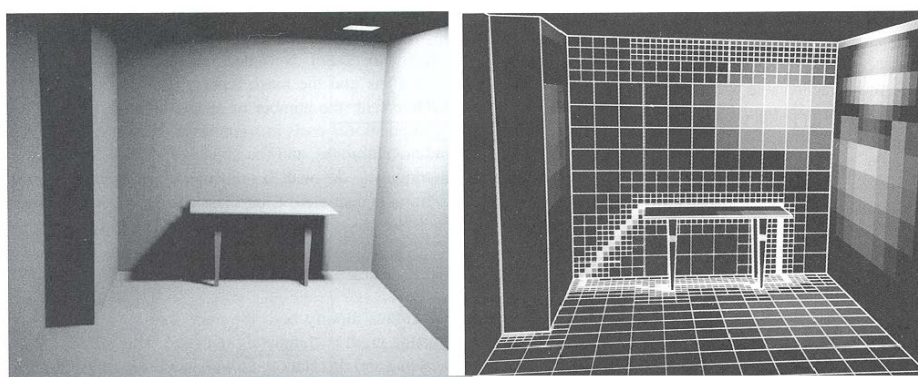


Рис. 3.3. Адаптивное разбиение геометрии сцены на отдельные элементы

Основная проблема – это невозможность определить заранее, какие области сцены требуют более детального разбиения, а какие могут быть огрублены. Очевидно, что процесс разбиения следует сделать адаптивным. Т.е. выбирается начальное разбиение и после небольшого числа итераций анализируется

распределение яркости по поверхностям сцены. В областях с большим градиентом яркости осуществляется разбивка элемента на более мелкие элементы, с низким градиентом яркости – элементы объединяются. При этом яркость соответствующим образом интерполируется на новые элементы меньшего или большего размера. Данная процедура может выполняться без пересчета с начального шага, качество изображения будет несколько хуже за счет ряда интерполяций. Поэтому, если требуется высокое качество изображения, то расчет надо проводить с нулевой итерации, используя оптимальное разбиение, полученное ранее.

Последний момент, который необходимо рассмотреть, – это расчет форм-факторов. Форм-фактор является важным компонентом уравнения глобального освещения. Важно, что форм-фактор зависит только от геометрии сцены. При этом предполагается, что источники света являются либо геометрическими объектами сцены, либо, при наличии искусственных источников света, например, параллельных или точечных, их интенсивность должна быть перенесена на геометрию сцены на предварительной фазе расчета. Таким образом, расчет яркости глобального освещения сводится к переносу яркости между геометрическими объектами сцены, определяемой форм-факторами и геометрическими свойствами объектов сцены. Собственно геометрия сцены при этом в расчете не участвует. Формула для расчета форм-фактора между двумя элементами  $i$  и  $j$  имеет следующий вид:

$$G_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{\cos \alpha_i \cos \alpha_j}{\pi |P_i - P_j|^2} v_{ij} dA_j dA_i. \quad (3.4)$$

Визуальное представление параметров данного уравнения представлено на рис. 3.4.

Следует отметить следующие свойства форм-факторов:

- Обратимость:  $A_i G_{ij} = A_j G_{ji}$ ,
- Аддитивность:  $G_{i(j \cup k)} = G_{ij} + G_{ik}$ ,
- В замкнутой системе  $\forall i, \sum_{j=1}^k G_{ij} = 1$ , т.е. вся световая энергия, покинувшая любую площадку, должна куда-то прийти.



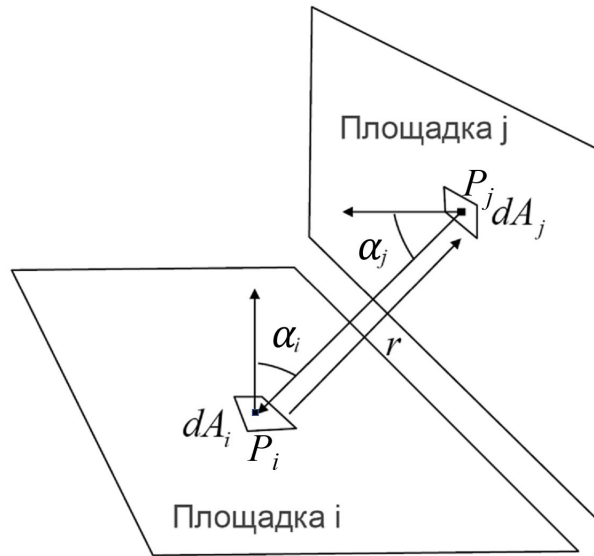


Рис. 3.4. Принцип расчета форм-фактора между двумя элементами  $i$  и  $j$

Технически для расчета форм-факторов можно использовать ряд подходов, показанных на рис. 3.5. Это аналогия Нуссельта, метод полукуба и трассировка лучей.

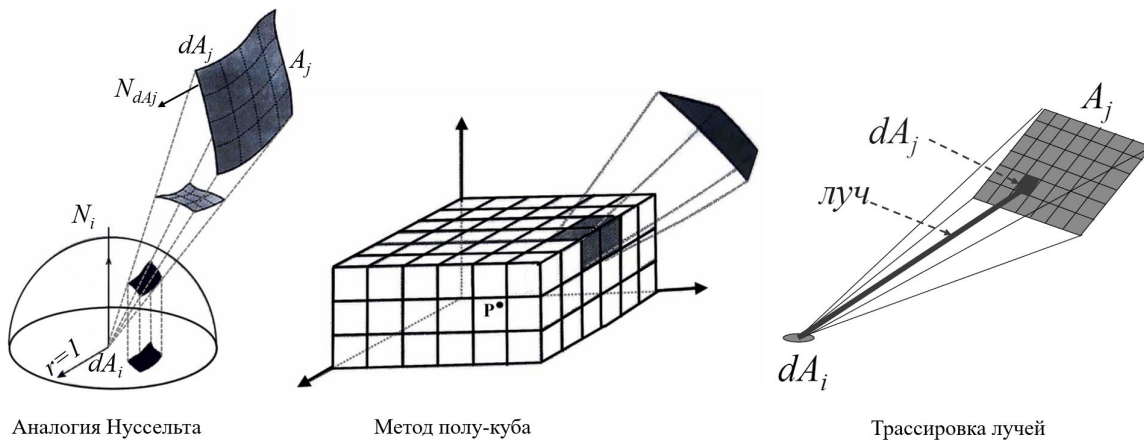


Рис. 3.5. Методы расчета форм-фактора между двумя элементами

Для численных расчетов может оказаться эффективным аналогия Нуссельта. Геометрически форм-фактор – это проекция части полусферы единичного радиуса, вырезаемая конусом, образованным лучами из центра полусферы по границе элемента поверхности  $A_j$ , на экваториальную плоскость сферы  $dA_i$ . Учитывая затенение  $v_{ij}$  от  $dA_j$  до  $dA_i$ , можно написать:

$$dG_{dA_i j} = \int_{A_j} \frac{\cos \alpha_i \cos \alpha_j}{\pi |P_i - P_j|^2} v_{ij} dA_j. \quad (3.5)$$

Далее, проводя аналогичное интегрирование для всей области  $A_i$ , можно получить выражение (3.4). Численный метод интегрирования по аналогии Нуссельта можно сформулировать следующим образом. Вся полусфера единичного радиуса разбивается на участки, образующие равные проекции на экваториальную плоскость. Из точки  $dA_i$  формируются лучи виденья на элементы  $dA_j$  и вычисляется, сколько участков полусферы «увидели» лучи (при этом можно учитывать эффект затенения). Полученное число, отнесенное к общему числу участков, даст дифференциальное значение форм-фактора (3.5).

С точки зрения вычислительной эффективности данное решение не является оптимальным. Более оптимальное решение – это использование метода полукуба, см. рис. 3.5., в котором полусфера заменяется полукубом. Данное решение может быть реализовано на графических ускорителях и выполнено одновременно для всех элементов сцены, см. рис. 3.6.

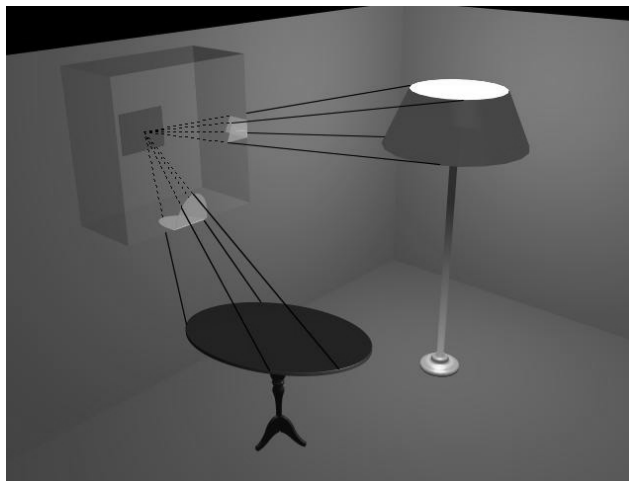


Рис. 3.6. Метод полу-куба

Все элементы поверхностей сцены проецируются на пять граней полукуба. При этом каждому элементу придается свой цвет, что позволяет выделить их на изображении на полукубе. Использование систем растровой графики с Z буфером позволяет определить видимость элементов, т.е.  $v_{ij}$ . Суммирование элементарных форм-факторов (пикселей соответствующего цвета и весовыми коэффициентами, связанными с гранями полу-куба) дает форм-фактор (3.5).

Развитие технологий, ориентированных на лучевые методы расчета, позволило использовать лучевые методы расчета форм-факторов. Как показано на рис. 3.5, достаточно просто сканировать с площадки  $dA_i$  область  $A_j$  и при этом вычислять видимость  $v_{ij}$  областей  $dA_j$  до  $dA_i$ . С алгоритмической точки зрения

данное решение достаточно просто, а современные графические ускорители могут обеспечить очень высокую скорость трассировки лучей.

### 3.3 Расчет глобального освещения для фотометрического шара

Фотометрический шар — это простое устройство для измерения оптического излучения. Данное устройство представляет собой полый шар, на внутренней сферической поверхности которого нанесено неселективное матовое покрытие с высоким коэффициентом отражения. Функция фотометрического шара заключается в пространственном интегрировании светового потока. Поэтому второе название этого устройства – интегрирующая сфера. Прежде чем использовать конструкцию шара для конкретного приложения, важно понять, как работает фотометрический шар. Вопрос о том, как свет распространяется внутри сферической полости фотометрического шара, начинается с рассмотрения диффузных отражающих свойств поверхностей. Исходя из этого, выводится энергетическая яркость внутренней сферической поверхности фотометрического шара и рассматриваются два связанных параметра: множитель сферы и средний коэффициент отражения.

#### 3.3.1 Обмен излучения внутри фотометрического шара

Теория фотометрического шара берет начало в теории метода переноса светового излучения в сцене, содержащей диффузные поверхности. Хотя общая теория достаточно сложна, сфера является уникальным, но простым объектом, для которого существует аналитическое решение.

Рассмотрим обмен излучения между двумя дифференциальными элементами диффузных поверхностей, см. рис 3.7.

Поскольку все поверхности имеют отражение, определяемое законом Ламберта, то яркость поверхности  $dA_1$  и ее световой поток связаны следующим соотношением:

$$L_1 = \frac{\Phi_1}{\pi dA_1}. \quad (3.6)$$

Принимая во внимание, что в данном случае яркость передается без потерь ( $L_1 = L_2$ ), и используя основные формулы радиометрии, можно найти долю энергии, покинувшую площадку  $dA_1$  и пришедшую на сопряженную площадку  $dA_2$  (это так называемый коэффициент обмена):

$$\frac{d\Phi_2}{\Phi_1} = \frac{\cos \alpha_1 \cos \alpha_2}{\pi |P_2 - P_1|^2} dA_2, \quad (3.7)$$

где углы  $\alpha_1$  и  $\alpha_2$  отсчитываются от соответствующих нормалей к поверхностям.

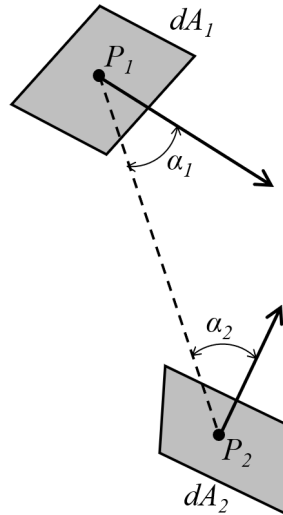


Рис. 3.7. Перенос излучения от элементарной поверхности  $dA_1$  к поверхности  $dA_2$

Рассмотрим два дифференциальных элемента:  $dA_1$  и  $dA_2$  внутри диффузной сферической поверхности, см. рис. 3.8. Поскольку расстояние между точками  $P_1$  и  $P_2$  равно  $|P_2 - P_1| = D \cos \alpha_1 = D \cos \alpha_2$ , то коэффициент обмена будет равен:

$$\frac{d\Phi_2}{\Phi_1} = \frac{dA_2}{\pi D^2}. \quad (3.8)$$

Важно, что коэффициент обмена не зависит от углов и расстояния между точками. Поэтому доля потока, получаемого  $dA_2$ , одинакова для любой точки излучения на поверхности сферы.

Если бесконечно малая область  $dA_1$  обменивается излучением с конечной площадью  $A_2$ , то уравнение (3.8) приобретает вид:

$$\frac{d\Phi_2}{\Phi_1} = \frac{1}{\pi D^2} \int_{A_2} dA_2 = \frac{A_2}{\pi D^2}. \quad (3.9)$$

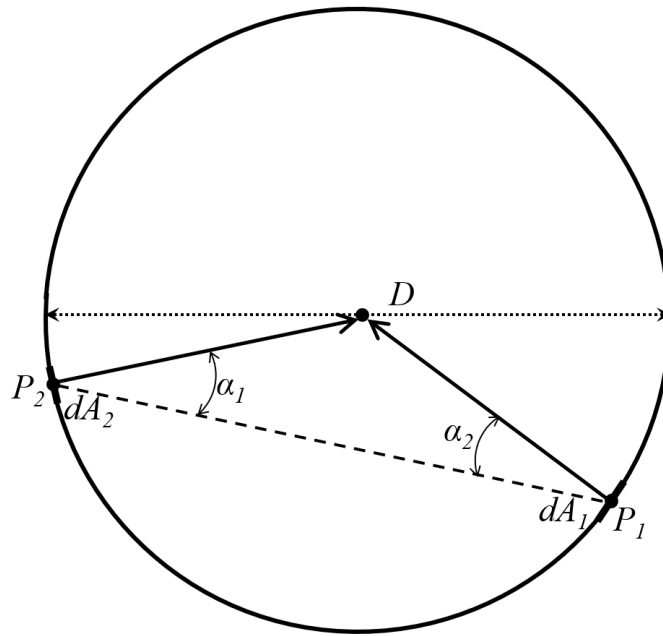


Рис. 3.8. Перенос излучения внутри фотометрического шара от элементарной поверхности  $dA_1$  к поверхности  $dA_2$

Поскольку этот результат также не зависит от  $dA_1$ , то

$$\frac{d\Phi_{A_2}}{\Phi_1} = \frac{A_2}{\pi D^2} = \frac{A_2}{A_S}, \quad (3.10)$$

где  $A_S$  - площадь поверхности всей сферы. Следовательно, доля лучистого потока, получаемого  $A_2$ , представляет собой площадь части поверхности, которую он занимает в пределах сферы. Если излучение передается на всю сферу, то коэффициент обмена равен 1.

### 3.3.2 Яркость на внутренней поверхности фотометрического шара

Свет ( $\Phi_i$ ), падающий на диффузную поверхность, отражаясь, создает виртуальный (вторичный) источник света. Вывод формулы яркости, освещенной изнутри фотометрического шара, начинается с выражения яркости  $L$  диффузной поверхности для входного потока,  $\Phi_i$ .

$$L = \frac{\Phi_i k}{\pi A}, \quad (3.11)$$

где  $k$  – коэффициент диффузного отражения, связанный с ДФО ламбертовой поверхности  $k$  следующим образом:  $k = \pi k$ , и определяющий долю отраженного света по отношению к падающему,  $A$  – освещаемая площадь сферы.

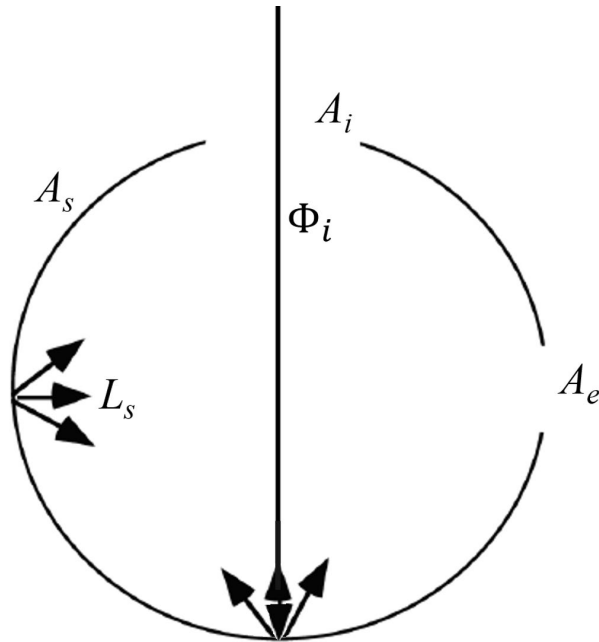


Рис. 3.9. Перенос излучения внутри фотометрического шара от элементарной поверхности  $dA_1$  к элементарной поверхности  $dA_2$

Для фотометрического шара уравнение яркости должно учитывать как множественные отражения от поверхности, так и потери через отверстия портов, необходимые для приема входного потока  $\Phi_i$ , а также для просмотра результирующего излучения. Рассмотрим сферу с входным портом  $A_i$  и выходным портом  $A_e$ .

Входной поток полностью рассеивается при начальном отражении. Величина потока, падающего на всю внутреннюю поверхность сферы или рассеянного с внутренней поверхности сферы:

$$\Phi_{i+1} = \Phi_i k \left( \frac{A_s - A_i - A_e}{A_s} \right), \quad (3.12)$$

где величина в скобках означает долю потока, получаемого поверхностью сферы, которая не задействована под отверстия портов. Этот член удобнее записать как  $(1 - h)$ , где  $h$  - доля порта, и  $h = (A_i + A_e)/A_s$ . Если существует более двух портов,  $h$  вычисляется как сумма всех областей портов.

По аналогичным соображениям величина потока, падающего на поверхность сферы после второго отражения или рассеянного после первого вторичного отражения, равна

$$\Phi_{i+2} = \Phi_i k^2 (1 - h)^2. \quad (3.13)$$

Третье отражение дает величину потока, равную

$$\Phi_{i+3} = \Phi_i k^3 (1 - h)^3. \quad (3.14)$$

Отсюда следует, что после  $n$  отражений суммарный поток, падающий на всю внутреннюю сферическую поверхность фотометрического шара:

$$\Phi_{i+n} = \Phi_i k (1 - h) \{1 + k(1 - h) + \dots + k^{n-1} (1 - h)^{n-1}\}. \quad (3.15)$$

Данный степенной ряд представляет собой убывающую геометрическую прогрессию, показатель которой  $k(1 - h) < 1$ . Суммарный поток можно выразить через формулу суммы бесконечной убывающей геометрической прогрессии:

$$\Phi = \frac{\Phi_i k (1 - h)}{1 - k(1 - h)}. \quad (3.16)$$

Уравнение (3.16) показывает, что общий поток, падающий на внутреннюю сферическую поверхность, выше, чем входной поток, из-за многократных отражений внутри сферы. Также видно, что два потока: падающий на поверхность и отраженный от поверхности диффузного шара, – будут равны после бесконечного числа отражений, поэтому яркость внутренней поверхности сферы можно определить как

$$L = \frac{\Phi}{\pi A_s (1 - h)} = \frac{\Phi_i}{\pi A_s} \cdot \frac{k}{(1 - k(1 - h))}. \quad (3.17)$$

Это уравнение используется для прогнозирования яркости внутренней сферической поверхности для данного входного потока как функции диаметра сферической полости фотометрического шара, коэффициента отражения и доли порта. Обратите внимание, что яркость уменьшается с увеличением диаметра сферы.

### ***Множитель сферы***

Уравнение (3.17) было намеренно разделено на две части. Первая часть равна яркости диффузной поверхности, излучающей поток  $\Phi_i$ . Вторая часть

уравнения представляет собой безразмерную величину, которую можно назвать множителем (или коэффициентом) сферы.

$$M = \frac{k}{1-k(1-h)}. \quad (3.18)$$

Этот множитель объясняет увеличение яркости, вызванное многократным отражением от поверхности сферы. График на рис. 3.10 иллюстрирует зависимость множителя сферы  $M$  от области портов  $h$  и коэффициента отражения поверхности сферы  $k$ .

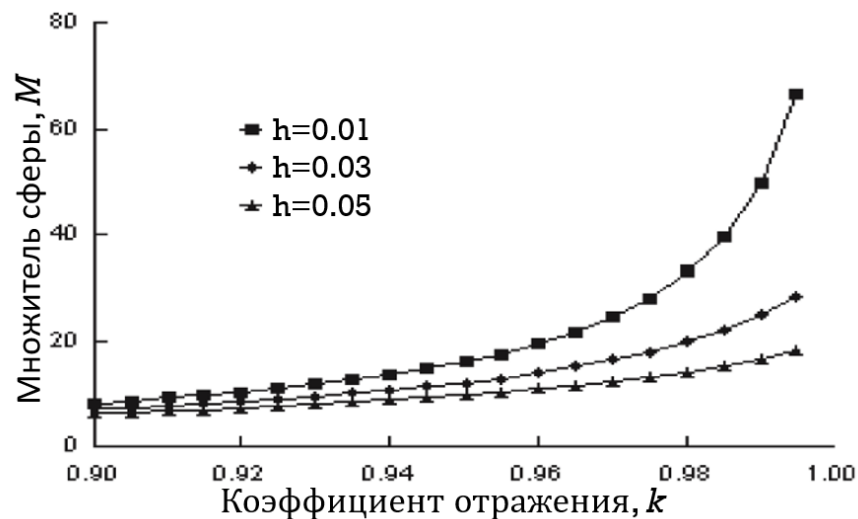


Рис. 3.10. Зависимость множителя сферы  $M$  от области портов  $h$  и коэффициента отражения поверхности сферы  $k$

Упрощенный интуитивный подход к прогнозированию плотности светового потока внутри фотометрического шара может быть сведен к простому делению входного потока на общую площадь поверхности сферы. Однако эффект множителя сферы состоит в том, что яркость внутренней поверхности фотометрического шара, по крайней мере, на порядок больше, чем дает этот простой интуитивный подход. Удобное эмпирическое правило заключается в том, что для большинства реальных фотометрических шаров ( $0.94 < k < 0.99$ ;  $0.02 < h < 0.05$ ) множитель сферы находится в диапазоне 10 - 30.

### **Средний коэффициент отражения**

Множитель сферы в формуле (3.17) определяет случай, когда падающий поток попадает на внутреннюю поверхность фотометрического шара, коэффициент отражения равномерно распределен по всей его внутренней



поверхности, а коэффициент отражения всех областей портов равен нулю. Общее выражение выглядит как

$$M = \frac{k_0}{1 - k_w(1 - \sum_{i=0}^n h_i) - \sum_{i=0}^n k_i h_i}, \quad (3.19)$$

где:

- $n$  – число портов,
- $k_0$  – начальный коэффициент отражения для падающего потока,
- $k_w$  – коэффициент отражения поверхности сферы,
- $k_i$  – коэффициент отражения  $i$ -го порта,
- $h_i$  – доля площади  $i$ -го порта.

Величина  $\bar{k} = k_w(1 - \sum_{i=0}^n h_i) - \sum_{i=0}^n k_i h_i$  может быть также описана как средний коэффициент отражения  $\bar{k}$  для всей внутренней поверхности фотометрического шара. Следовательно, множитель сферы можно переписать с точки зрения как начального, так и среднего коэффициента отражения:

$$M = \frac{\rho_0}{1 - \bar{k}}. \quad (3.20)$$

### ***Пространственное интегрирование***

Точный анализ распределения излучения внутри конкретного фотометрического шара будет зависеть от распределения падающего потока, геометрических деталей реальной конструкции шара и функции распределения коэффициента отражения покрытия внутренней сферической поверхности шара, а также от оптических свойств поверхностей устройств, установленных в открытых портах или внутри фотометрического шара. Рекомендации по проектированию для оптимальных пространственных характеристик основаны на максимизации как коэффициента отражения покрытия, так и диаметра сферической полости относительно требуемых отверстий портов и системных устройств.

Влияние коэффициента отражения и доли порта на пространственное интегрирование можно проиллюстрировать, рассматривая число отражений, необходимых для достижения общего потока, падающего на поверхность сферы, заданного уравнением (3.16). Общий поток на стенке сферы после только  $n$  отражений можно записать как

$$\Phi_n = \Phi_i \sum_{n=1}^n k^n (1 - h)^n. \quad (3.21)$$

Излучение, полученное после  $n$  отражений, можно сравнить с условием устойчивого состояния. Рис. 3.11 иллюстрирует процесс сходимости яркости на внутренней поверхности шара в зависимости от его параметров.

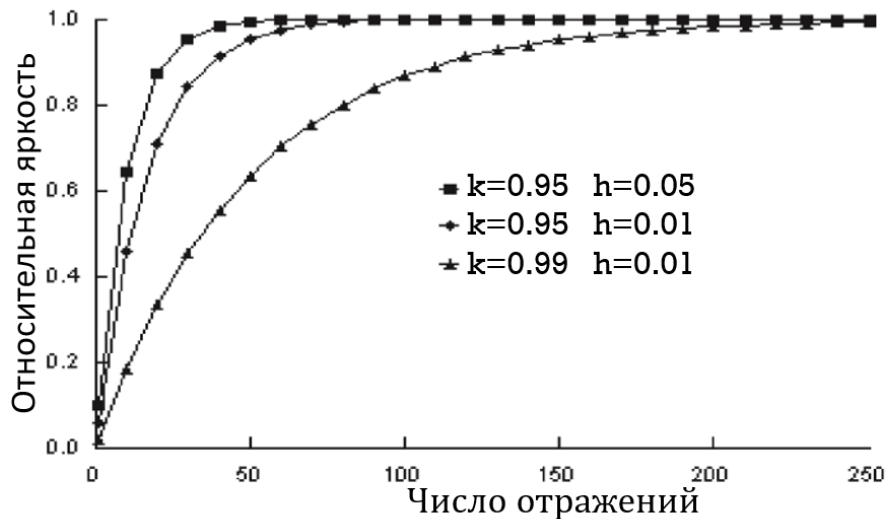


Рис. 3.11. График зависимости сходимости яркости на внутренней поверхности шара от его параметров

Поскольку фотометрический шар чаще всего используется в установившемся режиме, большее число отражений создает излучение в установившемся режиме, который получается как при увеличении  $k$ , так и при уменьшении  $h$ . Следовательно, в конструкции фотометрических шаров должны быть оптимизированы оба параметра для лучшей пространственной интеграции светового потока.

### 3.3.3 Применение фотометрического шара для измерения оптических характеристик

Фотометрические сферы могут использоваться в различных оптических измерениях, требующих определения интегральных характеристик оптических элементов, например, для измерения интегрального светового потока от источников света (как правило, небольшого размера), измерения интегральных свойств отражающих и пропускающих оптических материалов.

#### *Измерение светового потока источников света*

Наиболее естественное применение фотометрической сферы – это измерение светового потока источников света. Источник света помещается в один из портов сферы  $A_i$ , а приемник излучения – в другой порт сферы  $A_e$ , см. рис. 3.9.

При этом необходимо обеспечить экранирование порта  $A_e$  от прямой засветки источником света. Измеряется выходной поток через порт  $A_e$  и пересчитывается в исходный по формуле (3.16). Если измерительное устройство, находящееся в порту  $A_e$ , не калибровано на измерение абсолютного потока, то измерения можно выполнить сравнением результатов измерения с эталоном. Для этого производится измерение с эталоном, для которого  $\Phi_i^{et}$  известен, а затем – с измеряемым образцом. В случае линейной зависимости сигнала на приемном устройстве от величины падающего потока искомый поток определяется простой пропорцией:

$$\Phi_i = \frac{\Phi \Phi_i^{et}}{\Phi^{et}},$$

где  $\Phi^{et}$  – сигнал, измеренный для эталонного источника света.

Если зависимость сигнала на приемнике излучения нелинейно зависит от падающего потока, то вычисление потока выполняется по соответствующей калибровочной кривой приемника излучения.

### ***Измерение коэффициентов отражения и пропускания материалов***

Важным применением использования интегрирующих сфер является измерение коэффициента отражения и пропускания диффузных или рассеивающих материалов. Измерения почти всегда выполняются в монохроматическом свете для заданных длин волн. Единственным исключением может быть измерение коэффициентов отражения или пропускания с использованием фотометрического детектора, калиброванного на кривую видности человеческого глаза.

При измерениях коэффициента отражения образец помещается в отверстие порта напротив входного порта. Падающий поток отражается образцом. Суммарный интегральный коэффициент отражения (диффузной и зеркальной компонент) собирается интегрирующей сферой. Схема измерений представлена на левой части рис. 3.12. Если приемное устройство, расположенное в одном из портов шара, откалибровано на абсолютные значения светового потока, то коэффициент отражения может быть вычислен по формулам (3.16), (3.17) и (3.19), где неизвестной величиной  $k_i$  будет коэффициент отражения  $i$ -го порта, содержащего образец, т.е. коэффициент отражения образца.

Угол падения  $\alpha_i$  при измерениях коэффициента отражения обычно отклоняется от нормали до  $10^\circ$ . Зеркальная составляющая может быть исключена

из измерения путем использования нормального ( $0^\circ$ ) падения или путем подгонки другого порта на пути зеркального отражения и использования черной поглощающей световой ловушки для тушения зеркального потока.

Измерения коэффициента отражения при больших или переменных углах падения выполняются путем помещения образца в центр сферы и вращения его вокруг фиксированного входного пучка. Фотометрический шар может содержать дефлекторы, цель которых заключается в том, чтобы фотоприемник не мог непосредственно просматривать освещенный образец в любом измерении.

При измерении коэффициента пропускания прозрачный образец помещают во входной порт сферы, см. правую часть рис. 3.12, при этом приемник излучения может находиться в одном из портов сферы. В этом случае необходимо учитывать коэффициент отражения прозрачного образца (он может быть ненулевой), который должен быть измерен заранее.

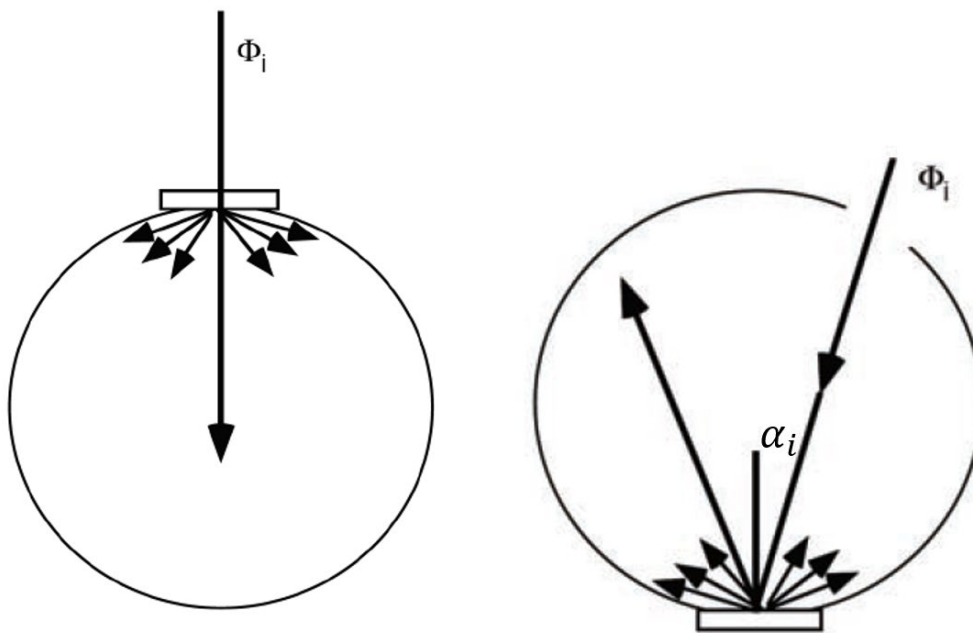


Рис. 3.12. Сема для измерения коэффициента отражения (левый рисунок) и пропускания (правый рисунок)

### 3.4 Заключение

Впервые метод излучательности был реализован в 1991 году в программе Lightscape 3.2. Это был первый метод визуализации трехмерных сцен с учетом глобального освещения. В дальнейшем компания Autodesk приобрела права на данное программное обеспечение и прекратила его продажу. Затем примитивная

версия визуализации глобального освещения, полученного средствами излучательности, была реализована в продукте компании 3D Max, в то время как существующие функциональности Lightscape были просто исключены.

Основными преимуществами методов излучательности можно считать: приемлемое качество визуализации за доступное время и возможность физически корректно решить проблему расчета глобального освещения в пространстве трехмерной сцены. Методы излучательности позволяют найти распределение глобального освещения в сцене, после чего эта сцена может быть визуализирована (как двумерное распределение яркости для заданных условий наблюдения) простейшими средствами визуализации, например, с помощью OpenGL.

Основной недостаток детерминистических методов в том, что они не могут решить проблему глобального освещения для произвольных оптических свойств поверхности. Физически корректно задача может быть решена только для объектов сцены, оптические свойства которых являются идеально диффузными (рассеивание подчиняется закону Ламберта). Как правило, детерминистические методы используют численные решения уравнения излучательности (3.2), которое может быть сведено к СЛАУ (3.3). Для решения уравнения (3.3) используются два основных итерационных метода: «бросания» и «сборки», суть которых заключается либо в передаче своего светового излучения от каждого объекта сцены на все остальные объекты сцены, либо в сборе светового излучения каждым объектом сцены от всех остальных объектов сцены соответственно.

Важным моментом расчета глобального освещения является определение форм-факторов элементов сцены, т.е. долей светового излучения элементов сцены, которые оно может передать остальным элементам сцены. Форм-факторы зависят только от геометрической конфигурации сцены, что позволяет независимо изменять оптические характеристики сцены (параметры источников света и коэффициенты отражения поверхностей) без достаточно трудоемкого перерасчета форм-факторов.

Существуют только два известных аналитических решения уравнения излучательности. Это расчет распределения глобального освещения (распределение освещенности, потока или яркости) внутри фотометрического шара и распределение глобального освещения, сформированного равноинтенсивным точечным источником света, расположенным между двумя параллельными идеально диффузными поверхностями. Аналитическое решение проблемы распределения глобального освещения внутри фотометрического шара

имеет большое практическое значение, поскольку фотометрический шар широко используется в задачах измерения оптических характеристик источников света (определение светового потока) и измерения интегральных коэффициентов отражения и пропускания оптических материалов.

### 3.5 Вопросы для самоконтроля усвоенных знаний

Следующие вопросы позволят оценить степень усвоения знаний по разделу детерминистические методы решения уравнения рендеринга.

1. Какая особенность поверхностей с ламбертовскими оптическими свойствами позволяет преобразовать уравнение рендеринга с систему линейных алгебраических уравнений (уравнение излучательности).
2. Напишите псевдокод для решения уравнения излучательности для сцены в виде куба с равноинтенсивным источником света, находящимся в центре куба. Для решения уравнения используйте метод «бросания». Световой поток источника света равен  $\Phi$ , грани куба имеют ламбертовские оптические свойства с коэффициентом отражения (коэффициентом яркости)  $kd$ .
3. Напишите псевдокод для решения уравнения излучательности для сцены в виде куба с равноинтенсивным источником света, находящимся в центре куба. Для решения уравнения используйте метод «сборки». Световой поток источника света равен  $\Phi$ , грани куба имеют ламбертовские оптические свойства с коэффициентом отражения (коэффициентом яркости)  $kd$ .
4. Какие способы вы можете предложить для оценки числа итераций, необходимых для решения уравнения излучательности с заданной точностью.
5. Предложите алгоритм и напишите псевдокод для адаптивного разбиения плоских прямоугольных поверхностей сцены, повышающий точность расчета пространственного распределения яркости.
6. Предложите алгоритм и напишите псевдокод для расчета формфакторов между двумя плоскими прямоугольными поверхностями.
7. Опишите основные преимущества и недостатки метода излучательности. Какие физические эффекты этот метод поддерживает, а какие – нет. Какие есть возможности улучшения этого метода и за счет чего.
8. В чем особенность модели фотометрического шара. Как можно добиться равномерной освещенности всей поверхности фотометрического шара.
9. Докажите, что если фотометрический шар будет освещен тонким параллельным источником света (например, лазером), то его яркость

будет равномерной по всей поверхности, исключая зону прямой засветки.

10. Каким образом можно использовать фотометрический шар для измерения светового потока источника света. Представьте схему эксперимента и расчетные формулы.
11. Каким образом можно использовать фотометрический шар для измерения коэффициента отражения поверхности образца с ламбертовскими свойствами. Представьте схему эксперимента и расчетные формулы.
12. Для чего нужны порты в фотометрическом шаре. Как меняется яркость поверхности фотометрического шара при добавлении портов (отверстий). Сохранится ли равномерность яркости поверхности фотометрического шара после добавления ряда портов. Представьте аналитическое решение.

#### **4. Метод Монте-Карло**

В этом разделе будут представлены основные принципы методов Монте-Карло, которые потребуются для решения проблемы глобального освещения. Более подробную историю развития методов Монте-Карло можно найти в книгах [7, 21-25]. Методы Монте-Карло имеют долгую историю, однако практическую значимость они приобрели только с появлением компьютерной техники в 50-60-е годы. Позже методы Монте-Карло подверглись большой критике из-за их повального применения ко всем математическим задачам.

Методы Монте-Карло вычисляют результаты на основе набора случайных чисел. Исторически эти методы часто применялись к задачам вероятностного характера. В этом случае методы Монте-Карло могут быть простым моделированием реальных случайных процессов. Например, в ядерной физике поведение нейтронов случайно, а макроскопические потоки нейтронов и другие величины затем могут быть вычислены с помощью так называемого аналогового моделирования. Отдельные частицы испускаются, отражаются, рассеиваются и поглощаются стохастически, что соответствует реальности. По аналогии с действительностью, вычисленные параметры распределения частиц также подвержены некоторой вероятностной неопределенности. Неопределенность вычислений может быть уменьшена, например, путем выполнения большого числа симуляций.

Методы Монте-Карло также могут решать проблемы детерминированного характера. В этом случае задача формализуется с помощью математической модели, которую можно рассматривать как описание случайного процесса.

Компьютерное моделирование случайного процесса позволяет решить исходную задачу. Таким образом, методы Монте-Карло могут быть полезны для решения задач, которые не решить другими способами (аналитическими или численными). Даже задачи вероятностного характера обычно формулируются как математические задачи, для которых сперва должны быть построены математические модели, для которых метод Монте-Карло будет одним из возможных решений.

Метод решения проблем, в которых методы Монте-Карло используются для решения детерминистических процессов, иногда называют сложным методом Монте-Карло. Такие решения являются результатом анализа и абстрагирования проблемы и в основном направлены на повышение эффективности моделирования. Проблема глобального освещения является одним из примеров таких подходов. Стохастическое поведение фотонов формализуется эквивалентными математическими моделями, представленными в первом разделе. В результате методы Монте-Карло не имитируют непосредственное излучение, отражение и поглощение фотонов, а имитируют лежащие в основе этих моделей случайные процессы.

#### 4.1 Интегрирование методом Монте-Карло

Все методы Монте-Карло можно прямо или косвенно рассматривать как методы численного интегрирования. Рассмотрим определенный интеграл от функции  $f(x)$  в интервале  $[0,1]$  и его оценку:

$$I = \int_0^1 f(x)dx. \quad (4.1)$$

Предположим, что  $f(x)$  является функцией в  $L^2(0,1)$ , т.е.  $\int_0^1 f^2(x)dx$  существует. Интеграл можно оценить, используя случайное число, равномерно распределенное в интервале  $[0,1]$ , и оценивая  $f(\xi)$ .  $f(\xi)$  – это так называемая первичная оценка интеграла, которую обозначим угловыми скобками:

$$\langle I \rangle_{prim} = f(\xi). \quad (4.2)$$

Необходимо отметить, что при интегрировании в интервале, отличном от  $[0,1]$ , например  $[a,b]$ , первичная оценка интеграла будет отмасштабирована на величину этого интеграла, поэтому выражение (4.2) примет следующий вид:

$$\langle I \rangle_{prim} = (b - a)f(\xi).$$



Аналогичная масштабировка будет применяться ко всем последующим выражениям, в которых область интегрирования отлична от  $[0,1]$ .

Вычисленное значение для конкретной выборки называется оценкой. Она равна площади поверхности прямоугольника под оцениваемой точкой, см. рис. 4.1 Оценка называется несмещенной, поскольку ее ожидаемое значение  $E(\langle I \rangle)$  равно фактическому значению интеграла (в данном случае функция плотности вероятности равна 1):

$$E(\langle I \rangle_{prim}) = \int_0^1 f(x) dx = I.$$

Предположим, что можно взять совершенно случайные, т.е. некоррелированные, однородные выборки. На практике обычно приходится прибегать к сгенерированным компьютером псевдослучайным числам, которые прошли определенные статистические тесты. Более подробную информацию о генерации случайных и псевдослучайных чисел можно найти в [26].

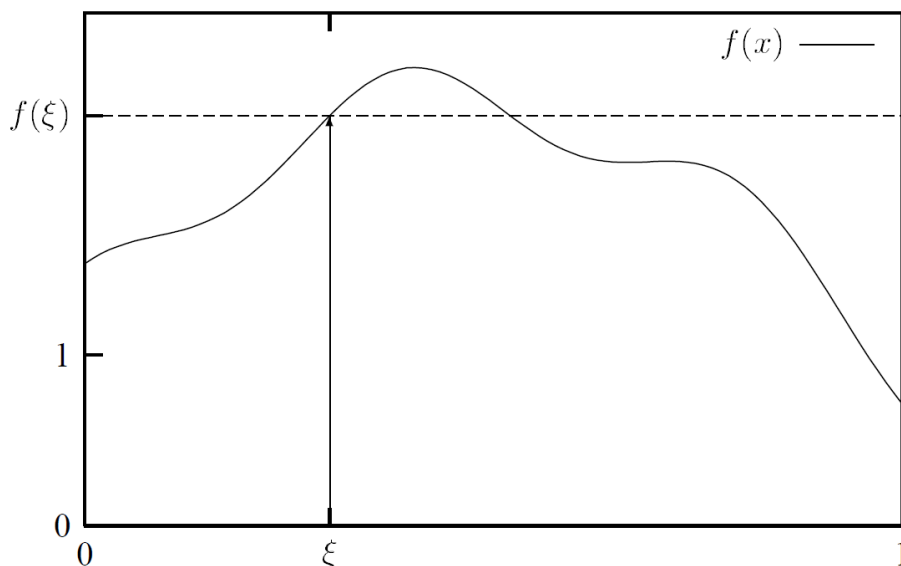


Рис. 4.1. Простой первичной оценкой определенного интеграла функции  $f(x)$  по интервалу является площадь поверхности прямоугольника, определяемая оценкой  $f(\xi)$  в равномерно выбранной случайной точке  $\xi$

Существует неопределенность результата вычислений, которая выражается дисперсией  $V$  и стандартным или среднеквадратичным отклонением  $\sigma$ :

$$V(\langle I \rangle_{prim}) = \sigma_{prim}^2 = \int_0^1 [f(x) - I]^2 dx = \int_0^1 f^2(x) dx - I^2. \quad (4.3)$$

С точки зрения практического применения это выражение не очень полезно, поскольку ни один из интегралов не известен на практике, однако это выражение может служить теоретическим ориентиром для сравнения оптимизаций.

Дисперсия обычно неприемлемо велика при использовании единственной выборки. Основой методов Монте-Карло является уменьшение неопределенности за счет использования большего количества, например,  $N$ , независимых выборок  $\xi_i$  и усреднение их, преобразуя ряд первичных оценок во вторичную оценку. Это эквивалентно переписыванию исходного интеграла (4.1) в виде суммы интегралов, см. рис. 4.2:

$$I = \int_0^1 f(x)dx = \sum_{i=1}^N \frac{\int_0^1 f(x)dx}{N} = \sum_{i=1}^N I_i.$$

Вторичная оценка есть сумма первичных оценок:

$$\langle I \rangle_{sec} = \sum_{i=1}^N \langle I_i \rangle_{prim} = \frac{1}{N} \sum_{i=1}^N f(\xi_i). \quad (4.4)$$

Это средняя несмещенных первичных оценок также является несмещенной оценкой. Рассматривая набор из  $N$  выборок как единственную  $N$ -мерную выборку, ожидаемое значение может быть вычислено следующим образом:

$$E(\langle I \rangle_{sec}) = \int_0^1 \dots \int_0^1 \frac{1}{N} \sum_{i=1}^N f(x_i) dx_1 \dots dx_N = I.$$

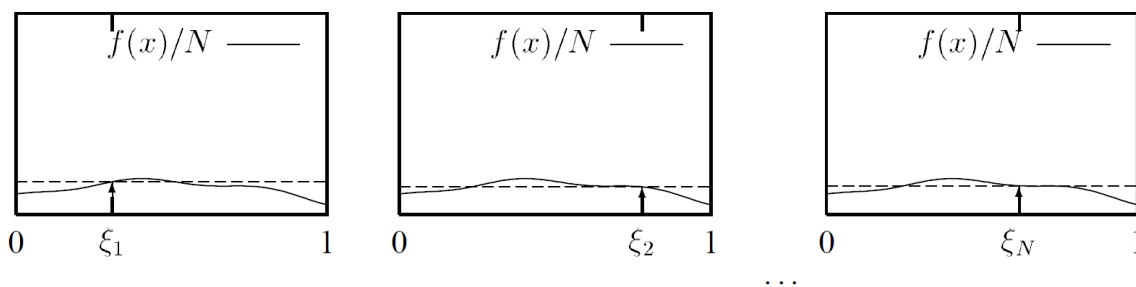


Рис. 4.2. Вторичная оценка интеграла – это среднее значение ряда первичных оценок

При этом среднеквадратичное отклонение уменьшается в  $N$  раз.

$$\begin{aligned} \sigma_{sec}^2 &= \int_0^1 \dots \int_0^1 \left[ \frac{1}{N} \sum_{i=1}^N f(x_i) \right]^2 dx_1 \dots dx_N - I^2 \\ &= \frac{1}{N} \int_0^1 f^2(x) dx - \frac{1}{N} I^2 = \sigma_{prim}^2 / N. \end{aligned} \quad (4.5)$$

Это фундаментальная закономерность. Это означает, что стандартная ошибка любой вторичной оценки такого типа (использующего равномерное распределение случайной величины) пропорциональна  $1/N$ , что не является впечатляющим результатом по сравнению с другими численными методами. Таким образом, исследования методов Монте-Карло в основном сосредоточены на уменьшении другого фактора в этом выражении, а именно среднеквадратичного отклонения первичной оценки значения интеграла. Ряд важных методов, уменьшающих стандартное отклонение, таких как стратифицированная выборка, выборка по значимости и т.п., будут рассмотрены в следующих разделах. Большинство из этих методов используют дополнительную информацию, которая может быть известна об интегрируемой функции.

Для простоты изложения будут рассмотрены только одномерные интегралы. Однако те же самые принципы Монте-Карло могут быть легко распространены на многомерные интегралы. Очевидное решение заключается в представлении многомерных интегралов как множества одномерных интегралов. Однако, как и в случае с детерминированными методами, количество оценок функций, которые мы пытаемся минимизировать, экспоненциально увеличивается с увеличением размерности интеграла. Вычисление  $d$ -мерного интеграла при  $N$  выборках на измерение требуется  $Nd$  выборок. Это увеличение становится неприемлемым для таких задач, как проблема глобальная освещения, когда выборки пикселей изображения, источников света и точек на полусфере, а также такие эффекты, как глубина резкости или размытие при движении, соответствуют разным измерениям. Этот эффект известен как «проклятие» размерности.

Альтернативный подход может рассматривать многомерные интегралы в целом (не по отдельным измерениям) с последующим отбором по области интегрирования. При этом все определения и свойства одномерного интегрирования методом Монте-Карло остаются в силе. Области интегрирования, имеющие сложную форму, в ряде случаев представляют проблему для практической реализации. Возможное решение – это выборка отклонения. Решение простое, но в большинстве случаев неэффективное, что связано с тем, что выборка и расчет функции в точке может многократно повторяться, пока точка не окажется в области интегрирования. Более сложные решения используют выборку по значимости, которая меняет область интегрирования и закон распределения случайных величин в выборке.

## 4.2 Стратифицированная выборка

Можно представить, что «неоправданное» скопление точек выборки (т.е. точки выбираются без понимания поведения интегрируемой функции) может значительно снизить эффективность интегрирования методом Монте-Карло. Без каких-либо предварительных знаний о подынтегральном выражении наиболее эффективным распределением выбора точек является равномерное распределение. Стратифицированная выборка пытается достичь этой цели путем разбиения области интегрирования на несколько подобластей и оценки интегралов в каждой из этих подобластей методом Монте-Карло. Такой подход гарантирует, что каждая подобласть получит заданное количество выборок. В простейшей реализации все подобласти имеют одинаковый размер и каждый из них получает значение одной выборки, см. рис. 4.3. Исходный интеграл (4.1), оцениваемый с использованием  $N$  выборок, может быть записан, как:

$$I = \int_0^1 f(x)dx = \sum_{i=1}^N \int_{A_i} f(x)dx = \sum_{i=1}^N I_i,$$

где  $A_i$  –  $i$ -ая подобласть в области интегрирования. В случае равномерного распределения областей  $i$ -ая подобласть находится в диапазоне  $[(i-1)/N, i/N]$ . Если выборка  $x_i$  осуществляется с равномерной плотностью вероятности  $N$  в подобласти области  $A_i$ , то первичными оценками для  $I_i$  являются  $f(\xi_i)/N$ . Общая оценка становится суммой частичных первичных оценок:

$$\langle I \rangle_{strat} = \sum_{i=1}^N \langle I_i \rangle_{prim} = \frac{1}{N} \sum_{i=1}^N f(\xi_i). \quad (4.6)$$

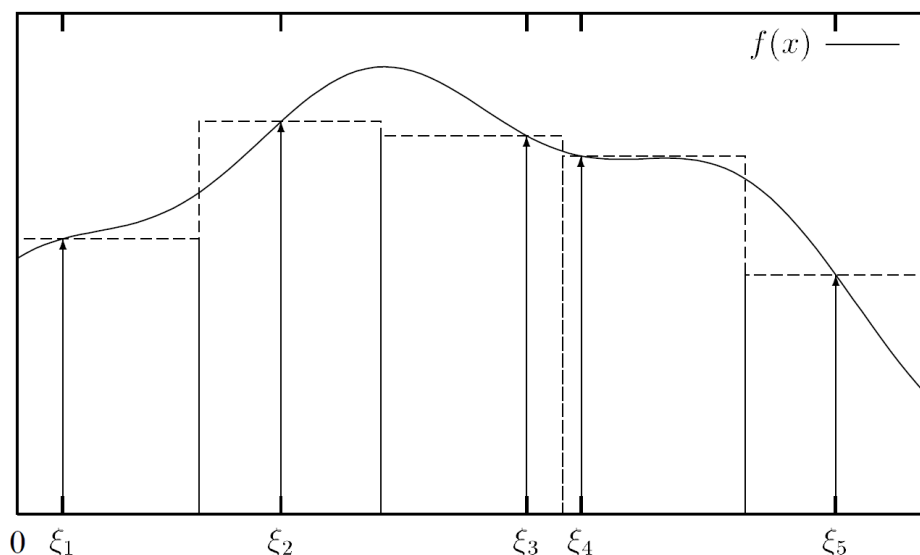


Рис. 4.3. Стратифицированная выборка представляет собой выборку одного или более образцов точек в подобластях общей области (в данном случае рассматривается равномерное разбиение на подобласти общей области)

Это выражение идентично выражению (4.4) для обычной вторичной оценки, но выборка осуществляется иначе. В то время, как ожидаемое значение остается точным интегралом для любого числа выборок, дисперсия отличается от дисперсии обычной вторичной оценки. Это сумма отклонений частичных оценок:

$$\sigma_{strat}^2 = \sum_{i=1}^N \left[ \int_{A_i} \left[ \frac{f(x)}{N} \right]^2 N dx_i - I_i^2 \right] = \frac{1}{N} \int_0^1 f^2(x) dx - \sum_{i=1}^N I_i^2. \quad (4.7)$$

Из-за того, что последний член всегда больше или равен  $I^2$ -члену выражения (4.5), дисперсия стратифицированной выборки всегда меньше или равна дисперсии обычной вторичной оценки с тем же количеством выборок:

$$\sigma_{strat}^2 \leq \sigma_{sec}^2. \quad (4.8)$$

Поскольку затраты на распределение выборок по слоям, как правило, незначительны, это делает стратифицированную выборку одной из фундаментальных основ оптимизации процесса интегрирования. Стратификация отличается от большинства других методов снижения дисперсии тем, что каждая выборка используется для оценки отдельного интеграла подобласти. Оценка в целом является несмещенной, поскольку среднее значение из  $M$  оценок с  $N$  слоями в каждой сходится к точному значению для  $M \rightarrow \infty$ . Однако на практике вместо этого может быть построено  $N \times M$  слоев. С этой точки зрения стратифицированная выборка потенциально может работать лучше, чем сходимость со скоростью  $1/\sqrt{N}$ , типичная для классического метода Монте-Карло.

Успех снижения дисперсии частично зависит от выбора страт. Если не известна дополнительная информация о подынтегральном выражении, то подынтервалы одинакового размера являются наиболее естественной альтернативой. Выбор различных слоев, таких как объединения небольших случайно распределенных подынтервалов, может снизить эффективность вычислений, поскольку улучшения можно достичь только за счет согласованности подынтегральных выражений. Предполагается, что выборки в подобластях будут представительными для всей функции в этой области. Для подынтегральных функций, изменяющихся достаточно быстро, стратифицированная выборка приводит к незначительному уменьшению дисперсии по сравнению с обычной вторичной оценкой.

Для многомерных интегралов действуют те же правила. Однако выбор оптимальных слоев гораздо менее очевиден. Они должны быть “компактными” в

некотором смысле, так, чтобы подынтегральное выражение было более или менее постоянным в этих областях. Предварительные знания о подынтегральном выражении будут полезны для определения параметров слоев. Для квадратной многомерной области интегрирования разбивка каждого интервала, соответствующего измерению на  $N$  подинтервалов, дает  $N^d$  слоев и выборок, как и в случае обычной вторичной оценки. Этот подход страдает от того же «проклятия» размерности, которое обсуждалось в классическом методе Монте-Карло интегрирования.

$N$ -rooks выборка – это дешевая в вычислительном отношении альтернатива расширению стратифицированной выборки до более высоких измерений. Техника названа по аналогии с двумерной шахматной задачей. Двумерная область интегрирования квадратной формы разбивается  $N$  на  $N$  квадратов.  $N$  из этих квадратов выбираются случайным образом по некоторой маске  $N$ , см. рис. 4.4. Затем в каждом из выбранных квадратов выбирается точка. Этот метод гарантирует, что каждая строка и каждый столбец квадратов получат ровно один образец.

$N$ -rooks выборка может быть расширена до произвольного числа измерений  $d$ . Для этой цели каждый интервал интегрирования  $d$  измерений разбивается на  $N$  подинтервалов. Для каждого измерения  $i$  на этапе предварительной обработки создается случайная маска перестановкой  $P_{i1}, \dots, P_{iN}$  индексов  $1, \dots, N$ . Для каждой выборки  $s$  интервалы, которые должны быть выбраны в последовательности измерений, обозначены масками  $(P_{1s}, P_{2s}, P_{2s}, \dots)$ , см. рис. 4.5. Чтобы избежать корреляции между результатами расчета отдельных интегралов, маски должны быть случайными.

	2			
4				
				5
		1		
			3	

Рис. 4.4. Стратификация выборок в одном измерении может быть расширена до двух измерений с использованием  $N$ -rooks выборки. Двумерная область разделена на сетку из  $N \times N$  квадратов.  $N$  квадратов, промаркированных цифрой, выбираются случайным образом по  $N$ -rooks маске. Для каждого выбранного квадрата выполняется ровно одна выборка

*N*-roks выборка особенно полезна, если подынтегральное выражение можно разделить по отдельным измерениям (например, если подынтегральное выражение постоянно по одному из измерений), поскольку выборки идеально стратифицированы в каждом отдельном измерении. При расчете глобального освещения наиболее подходящим примером может служить метод интегрирования освещенности от протяженного источника света. Яркость источника света – это четырехмерная функция (два измерения по пространственным координатам, два по угловым координатам). Как правило, угловое распределение яркости не зависит от пространственной координаты, и наоборот, пространственное распределение яркости не зависит от направления наблюдения. В этом случае область интегрирования естественным образом разбивается на пространственную и угловую. Однако можно ожидать, что прирост эффективности будет уменьшаться по мере роста числа выборок, поскольку они могут начать «неудачным» образом собираться вместе по всей области интегрирования. В этом случае другие методы могут стать более привлекательными, несмотря на большую алгоритмическую сложность.

Интервал #	Размерность #				
	1	2	3	4	...
Выборка #1	3	4	3	1	...
#2	2	1	4	5	...
#3	4	5	2	2	...
#4	1	2	1	3	...
#5	5	3	5	4	...

Рис. 4.5. *N*-roks выборка может быть расширена до более высоких измерений путем создания случайной маски индексов 1, ..., *N* для каждого измерения (показано в виде столбцов). На этапе предварительной обработки и выбора индекса для каждой выборки (показанной здесь в виде строк) и каждого измерения. Первые два измерения в этом примере соответствуют двумерной *N*-roks маске, показанной на рис. 4.4

Методы квази-Монте-Карло могут представлять интересную альтернативу классическим методам стратификации. Целью стратификации является более эффективная выборка подынтегрального выражения путем «правильного» (равномерного) распределения выборок по предметной области. Для методов Монте-Карло выборки должны быть случайными. Однако при генерации их на компьютере они получаются псевдослучайными. Для квази-случайных методов Монте-Карло «случайность» не является определяющей характеристикой, эти методы концентрируются на формировании равномерного распределения

выборок, например [27-29]. Основное свойство несмещенности неприменимо для этих распределений, поскольку они являются скорее псевдо-детерминированными, а не стохастическими. Однако было доказано несколько теорем, которые связывали ошибки, возникающие при использовании этих распределений в задачах интегрирования при условии, что подынтегральное выражение имеет ограниченную вариацию, например [27]. По этой причине в ряде случаев разрабатываются решения, основанные на методах Монте-Карло, а затем они дополняются распределениями квази-Монте-Карло в ожидании лучшей сходимости.

### 4.3 Выборка по значимости

Еще одним мощным решением для оптимизации интегрирования методом Монте-Карло является выборка по значимости. Посмотрим снова на базовый интеграл (4.1). Некоторые области подынтегральной функции могут быть более важными, чем другие, поскольку они имеют более высокие значения функции. Даже несколько выборок в этих областях могут оказать большое влияние на результат интегрирования. Очевидно, что в этих случаях более эффективным решением будет увеличить число выборок в этих областях, например, с помощью введения соответствующей коррекции в виде функции плотности вероятности PDF  $p(x)$ , которая близка к подынтегральной функции и не является постоянной, см. рис. 4.6. После введения такой коррекции функция оценки интеграла должна быть модифицирована, чтобы не создавать смещенного решения. Интеграл может быть переписан следующим образом:

$$I = \int_0^1 f(x)dx = \int_0^1 \frac{f(x)}{p(x)} p(x)dx.$$

Если случайная величина выбирается в соответствии с PDF  $p(x)$ , первичная оценка интеграла будет иметь вид:

$$\langle I \rangle_{prim} = \frac{f(\xi)}{p(\xi)}. \quad (4.9)$$

Ожидаемое значение по-прежнему является фактическим значением интеграла. Среднеквадратичное значение первичной оценки интеграла будет иметь вид:

$$\sigma_{imp}^2 = \int_0^1 \left[ \frac{f(x)}{p(x)} \right]^2 p(x)dx - I^2 = \int_0^1 \frac{f(x)^2}{p(x)} dx - I^2. \quad (4.10)$$



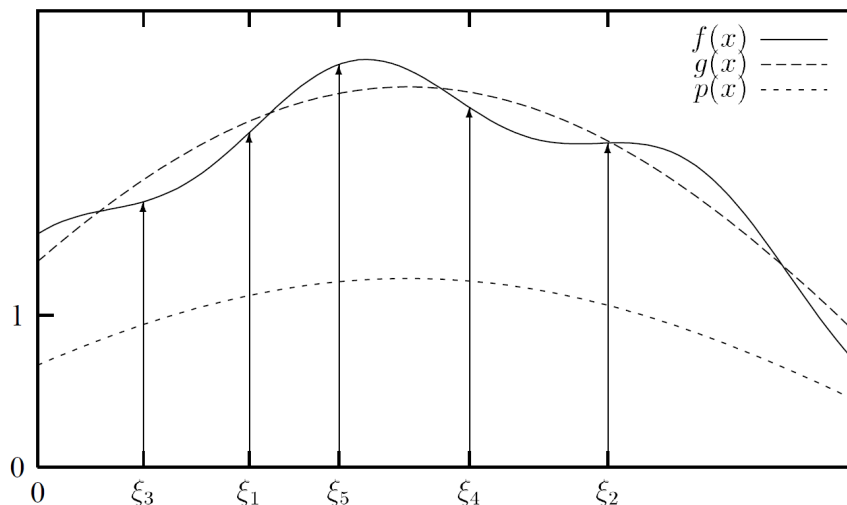


Рис. 4.6. Выборка по значимости стремится сделать больше выборок в областях высоких значений подынтегральной функции  $f(x)$  путем выбора случайных значений в соответствии с функцией плотности вероятности  $p(x)$  и деления полученной оценки интеграла на  $p(x)$ . Функция плотности вероятности является нормализованным вариантом аппроксимации  $g(x)$  для подынтегрального выражения

Любая функция  $p(x)$ , которая в области интегрирования удовлетворяет следующим условиям, может быть использована как функция плотности вероятности:

- $p(x) > 0$  для всех  $x \in [0,1]$ , для которых  $f(x) \neq 0$ ,
- $\int_0^1 p(x)dx = 1$ ,
- Необходимо создавать выборку  $x$  такой, чтобы плотность вероятности ее выбора была равна  $p(x)$ . Для этого должна быть вычислена обратная  $P^{-1}(x)$  функция распределения вероятности  $P(x)$ . Функция распределения вероятностей  $P(x)$  является кумулятивной функцией функции плотности вероятности  $p(x)$ :

$$P(x) = \int_0^x p(t)dt. \quad (4.11)$$

В идеале для неотрицательных подынтегральных функций PDF должна быть пропорциональна подынтегральному выражению:  $p(x) = f(x)/I$ . В этом случае среднеквадратичное отклонение будет равно нулю даже для одной выборки. Единственный минус данного решения – это то, что коэффициент нормализации является тем интегралом, который необходимо вычислить. Поэтому данный вариант не имеет практического значения. Однако любая другая функция, имеющая форму, близкую к подынтегральному выражению, и которая

может быть нормализована, интегрирована, а затем инвертирована, даст PDF, которая уменьшает дисперсию оценки значения интеграла. Если подынтегральное выражение отрицательно, то PDF должна быть пропорциональна абсолютному значению подынтегральной функции. Это дает минимальную ненулевую дисперсию. Если подынтегральное выражение имеет сингулярность, то аналогичная сингулярность должна быть и в PDF. Если это невозможно выполнить, то дисперсия может быть очень высокой и, возможно, даже бесконечно большой, поскольку эта сингулярность будет возведена в квадрат при расчете дисперсии.

Выбор случайной величины  $\xi$  в соответствии с  $p(x)$  обычно выполняется путем выбора случайной величины  $\tau$ , равномерно распределенной в интервале  $[0,1]$ , а затем трансформацией этой случайной величины обратной функцией плотности вероятности:  $\xi = P^{-1}(\tau)$ , см. рис. 4.7. Таким образом, оценка (4.9) приобретает вид:

$$\langle I \rangle_{prim} = \frac{f(P^{-1}(\tau))}{p(P^{-1}(\tau))}. \quad (4.12)$$

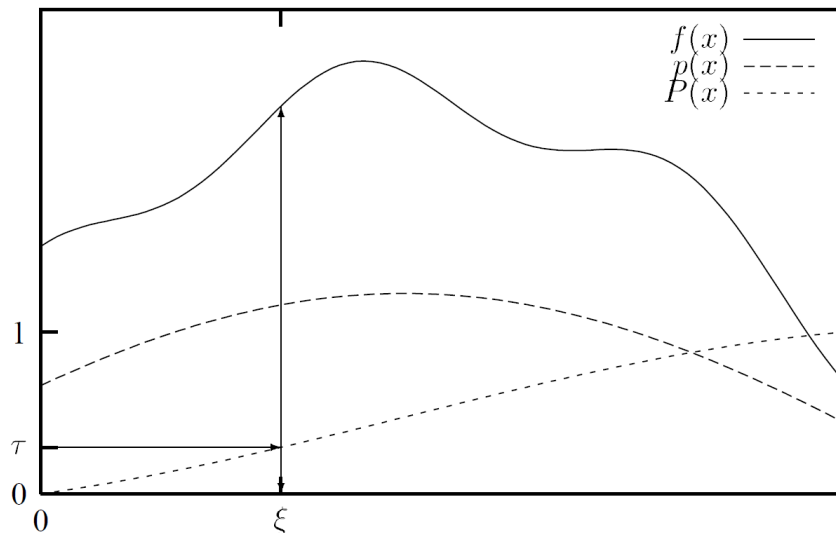


Рис. 4.7. Выборка в соответствии с PDF  $p(x)$  обычно выполняется путем взятия равномерной выборки  $\tau$  и преобразования ее в  $\xi$  обратной функцией распределения вероятностей  $P(x)$ :  $\xi = P^{-1}(\tau)$ . Неоднородная выборка  $\xi$  затем может быть использована при оценке интеграла от функции  $f(x)$

Это приводит к альтернативной интерпретации выборки по значимости как преобразования исходного интеграла:

$$\begin{aligned}
 I &= \int_0^1 f(x)dx = \int_0^1 f(P^{-1}(t)) \frac{dx}{dt} dt \\
 &= \int_0^1 f(P^{-1}(t)) \frac{dP^{-1}(t)}{dt} dt = \int_0^1 \frac{f(P^{-1}(t))}{p(P^{-1}(t))} dt .
 \end{aligned}$$

Равномерная выборка последнего интеграла эквивалентна выборке исходного интеграла в соответствии с  $p(x)$ . Рис. 4.8 иллюстрирует эту идею. Снижение дисперсии происходит вследствие того, что подынтегральное выражение становится более постоянной функцией, и ошибка интегрирования снижается. Выборка по значимости может быть объединена со стратифицированной выборкой, поскольку равномерные функции легче стратифицировать.

Использование PDF для функций больших размерностей выглядит более громоздко. Для каждого измерения должна быть сформирована такая выборка, чтобы в результате получилось требуемое многомерное распределение. Рассмотрим двумерную PDF  $p(x_1, x_2)$  над областью  $[0,1] \times [0,1]$ . Чтобы произвести выборку пары  $(\xi_1, \xi_2)$  в соответствии с этим распределением, сперва надо выполнить выборку в соответствии со следующим одномерным распределением:

$$p_1(x_1) = \int_0^1 p(x_1, x_2) dx_2.$$

Как только стала известна  $\xi_1$ ,  $\xi_2$  выбирается в соответствии с распределением:

$$p_1(x_2 | x_1 = \xi_1) = \frac{p(\xi_1, x_2)}{\int_0^1 p(\xi_1, x_2) dx_2}.$$

Результирующее распределение соответствует требуемому распределению, поскольку его PDF является продуктом одномерных PDF:  $p(x_1, x_2) = p_1(x_1)p_1(x_2|x_1)$ . В работе [30] представлен ряд преобразований PDF, которые часто используются в компьютерной графике.

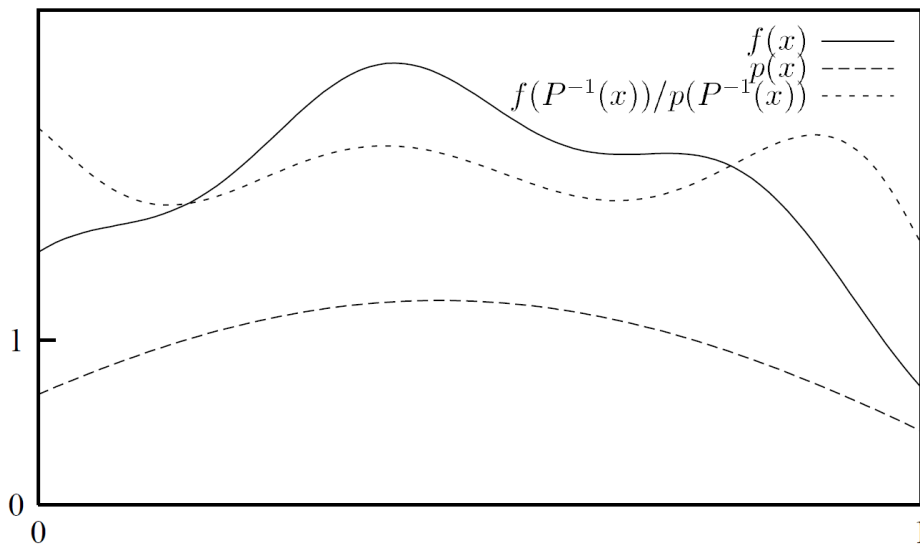


Рис. 4.8. Выборка значимости в соответствии с PDF  $p(x)$  для вычисления интеграла функции  $f(x)$  эквивалентна вычислению интеграла преобразованной функции  $f(P^{-1}(t))/p(P^{-1}(t))$  с использованием равномерной выборки. Если PDF выбрана в соответствии с  $f(x)$ , преобразованная функция будет близка к постоянной и, следовательно, эффективность интегрирования будет выше. При этом значение интеграла остается прежним, а дисперсия уменьшается

Выборка по значимости, как правило, может быть достигнута с помощью различных альтернативных преобразований. Одним из наиболее сложных примеров, где необходимо сформировать сложное многомерное распределение случайной величины, является двунаправленная функция отражения поверхности. В общем случае это одиннадцатимерная функция, которая учитывает направление освещения, направление наблюдения, цвет, координаты точки наблюдения, координаты точки освещения, поляризацию и флуоресценцию. Как правило, четыре последних эффекта не моделируются в большинстве систем компьютерной графики, также в большинстве случаев рассматриваются изотропные поверхности, поэтому число измерений сводится к четырем. При формировании функций плотности одно из направлений (наблюдения или освещения, в зависимости от прямой или обратной техники трассировки лучей) переводится в один слой, а другое направление вместе с цветом – в другой слой. Поэтому сначала происходит рассеивание луча, а потом приведение результата рассеивания к нужному направлению. Другой пример сочетания стратификации и выборки по значимости – это модель излучения протяженного источника света с переменной яркостью, например, экран дисплея. В данном случае предполагается формирование двух слоев. Первый слой отвечает за распределение светимости точек поверхности, а второй слой – за угловое распределение силы света с каждой точки поверхности. Поэтому сперва можно выбрать точку на поверхности в

соответствии с PDF светимости, а затем направление излучения в соответствии с PDF силы света для заданной точки поверхности.

#### **4.4 Многократная выборка по значимости**

В большинстве случаев достаточно сложно создать одну PDF, которая повторяла бы вид всего подынтегрального выражения. В ряде случаев гораздо проще создать несколько PDF, каждая из которых аппроксимирует некоторую определенную часть подынтегрального выражения. Одним из примеров может служить подынтегральное выражение, являющееся произведением двух функций. Часто каждая из функций может быть преобразована в PDF, но не результат их произведения. Тем не менее, обе функции могут с определенной долей точности определять форму подынтегрального выражения в некоторых областях ее интегрирования. Есть несколько вариантов решения для формирования PDF: выбор одной PDF, объединение PDF и выборка из объединенных PDF или выборка из каждой PDF с последующим объединением результата выборки.

В работе [31] была рассмотрена типичная проблема рендеринга, возникающая при расчете яркости прямого освещения, а именно, следует ли выполнять выборку в соответствии с ДФР поверхности или в соответствии с распределением яркости источников света. Предлагается идентифицировать различные факторы, как имеющие низкую, высокую или неизвестную вариацию. Затем делается единичная выборка в соответствии с фактором, имеющим наибольшую вариацию. При выборе источников света формируется линейная зависимость PDF в соответствии со световыми потоками отдельных источников света. Источник света выбирается в соответствии с PDF. Большинство этих методов интуитивно понятны, и для выбора оптимальных параметров полагаются на эмпирические правила. В работках [32, 33] представлен подход в контексте вычислений яркости прямого освещения и двунаправленной трассировки путей. Могут быть построены более сложные и общие схемы для определения важности каждой из отдельных PDF. Такие схемы, как правило, требуют предварительного расчета определенных характеристик сцены, которые могут быть получены либо специальной процедурой предобработки сцены, либо в процессе выполнения рендеринга и применяться по мере накопления соответствующей статистики.

Если имеется несколько первичных оценок, и все они вычислены с использованием разных PDF, то задача объединения их в единую оценку весьма неочевидна. Простое их усреднение в целом не будет оптимальным, поскольку некоторые оценки могут быть более значимыми, чем другие, и их значимость может меняться от области к области. В работках [32, 33] было предложено создать средневзвешенное значение оценок, где веса зависят от положения в

области интегрирования. Выведем условие того, чтобы результат был несмещенным для упрощенного случая с одной выборкой на PDF. Предположим, что  $n$  случайных величин  $\xi_1, \xi_2, \xi_3, \dots, \xi_n$  выбираются в соответствии с рядом PDF  $p_1(x), p_2(x), p_3(x), \dots, p_n(x)$ . Взвешенная сумма первичных оценок интеграла (4.1) может быть записана в виде:

$$\langle I \rangle_{comb} = \sum_{i=1}^n w_i(\xi_i) \frac{f(\xi_i)}{p_i(\xi_i)}. \quad (4.13)$$

где  $w_i(\xi_i)$  – весовая функция. При этом оценка должна оставаться несмещенной:

$$\begin{aligned} E(\langle I \rangle_{comb}) &= \sum_{i=1}^n \int_0^1 \left[ w_i(x) \frac{f(x)}{p_i(x)} \right] p_i(x) dx \\ &= \int_0^1 \left[ \sum_{i=1}^n w_i(x) \right] f(x) dx \equiv \int_0^1 f(x) dx \end{aligned}$$

Это условие выполняется для всех возможных подынтегральных выражений  $f(x)$  тогда и только тогда, когда

$$\sum_{i=1}^n w_i(x_i) \equiv 1. \quad (4.14)$$

Дисперсия объединенных оценок представляет собой сумму дисперсий отдельных оценок, поскольку они некоррелированные:

$$\begin{aligned} \sigma_{comb}^2 &= \sum_{i=1}^n \left[ \int_0^1 \left[ w_i(x) \frac{f(x)}{p_i(x)} \right]^2 p_i(x) dx - \left[ \int_0^1 w_i(x) \frac{f(x)}{p_i(x)} p_i(x) dx \right]^2 \right] \\ &= \int_0^1 \left[ \sum_{i=1}^n \frac{w_i^2(x)}{p_i(x)} \right] f^2(x) dx - \sum_{i=1}^n \left[ \int_0^1 w_i(x) f(x) dx \right]^2. \end{aligned} \quad (4.15)$$

Выбор весовых функций достаточно свободен, необходимое условие для выбора весовых коэффициентов определяется условием (4.14). Рассмотрим ряд возможных решений:

- $w_i(x) = 1/n$ , где  $n$  число оценщиков:

$$\langle I \rangle_{aver} = \frac{1}{n} \sum_{i=1}^n \frac{f(\xi_i)}{p_i(\xi_i)}. \quad (4.16)$$

- $w_i(x) = 1$ , для фиксированного  $i$ , равного  $k$ ,  $w_i(x) = 0$  для всех остальных  $i$ . В этом случае выборка осуществляется для единственного  $k$ :

$$\langle I \rangle_{aver} = \frac{f(\xi_k)}{p_k(\xi_k)}. \quad (4.17)$$

Этот случай представляет главным образом теоретический интерес и не используется на практике, поскольку он исключает все выборки, кроме одной. Тем не менее, этот выбор может быть оптимальным, если его оценка идеальна, а у других нет. Очевидно, что дисперсия равна дисперсии выбранной оценки.

- $w_i(x) = 1$ , если  $p_i(x) > p_j(x)$  для любого  $j$ , не равного  $i$ , иначе  $w_i(x) = 0$ . Это так называемая максимальная эвристика, и ее оценка будет иметь вид:

$$\langle I \rangle_{aver} = \sum_{i=1}^n (p_i(\xi_i) > p_j(\xi_i), \forall i \neq j) \frac{f(\xi_i)}{p_i(\xi_i)}; 0. \quad (4.18)$$

- $w_i(x) = \frac{p_i(x)}{\sum_{j=1}^n p_j(x)}$  – сбалансированная эвристика, для которой

$$\langle I \rangle_{bal} = \sum_{i=1}^n \frac{f(\xi_i)}{\sum_{j=1}^n p_j(\xi_i)}. \quad (4.19)$$

При этом дисперсия равна

$$\sigma_{bal}^2 = \int_0^1 \frac{1}{\sum_{j=1}^n p_j(x)} f^2(x) dx - \sum_{i=1}^n \int_0^1 \left[ \frac{p_i(x)}{\sum_{j=1}^n p_j(x)} f(x) dx \right]^2. \quad (4.20)$$

Этот выбор оптимален в том смысле, что он минимизирует сумму моментов второго порядка отдельных членов уравнения оценки интеграла, т.е. первого члена выражения для дисперсии в уравнении (4.15). Для более общего случая в [33] доказано, что дисперсия, полученная с использованием других весовых функций, может улучшить эвристику баланса лишь на ограниченную величину:

$$\sigma_{comb}^2 \geq \sigma_{bal}^2 - \left(1 - \frac{1}{n}\right) I^2.$$

- $w_i(x) = \frac{p_i^\beta(x)}{\sum_{j=1}^n p_j^\beta(x)}$  – степенная эвристика, являющаяся комбинацией максимальной и сбалансированной эвристик. Ее оценка имеет следующий вид:

$$\langle I \rangle_{pow} = \sum_{i=1}^n \frac{p_i^{\beta-1}(\xi_i)}{\sum_{j=1}^n p_j^\beta(\xi_i)} f(\xi_i). \quad (4.21)$$

Показатель степени  $\beta$  может быть любым числом. При  $\beta = 1$  данная эвристика вырождается в сбалансированную, а при  $\beta = \infty$  – в максимальную. Очевидно, что данная стратегия более эффективна, когда PDF хорошо согласуются с подынтегральным выражением в областях, где значения этого выражения достаточно велики. Показатель степени позволяет выделить значимость PDF в этих областях, а в областях, где значение PDF мало, значимость PDF экспоненциально падает.

Как и в случае выборки по значимости и других методов уменьшения дисперсии, мы можем интерпретировать комбинацию оценок как преобразование исходного интеграла. Формально можно написать:

$$I = \int_0^1 f(x)dx = \sum_{i=1}^n \int_0^1 w_i(x)f(x)dx,$$

при том же самом условии, что  $\sum_{i=1}^n w_i(x_i) \equiv 1$ . С этой точки зрения объединение различных оценок одного и того же подынтегрального выражения с весовыми функциями фактически эквивалентно разбиению подынтегрального выражения на весовые функции и суммированию оценок. Выборка каждого из  $n$  интегралов с соответствующей им PDF и суммирование оценок дают объединенную оценку (4.13). Выборка по значимости соответствует следующим дополнительным преобразованиям:

$$I = \sum_{i=1}^n \int_0^1 \frac{w_i(P_i^{-1}(t))}{p_i(P_i^{-1}(t))} f(P_i^{-1}(t)) dt.$$

Рис. с 4.9 по 4.14 иллюстрируют возможные варианты выбора двух весов PDF для интегрирования функции, являющейся произведением двух функций с явно выраженными максимумами. На каждом рисунке верхний график – это подынтегральная функция. Второй и третий графики показывают, как разбивается подынтегральное выражение: на две равные части (рис. 4.9), на саму функцию и 0 (рис. 4.10), на 0 и саму функцию (рис. 4.11), на части, соответствующие максимальной эвристике (рис. 4.12), на части, соответствующие сбалансированной эвристике (рис. 4.13), и на части, соответствующие степенной эвристике (рис. 4.14). Выборка в соответствующих частях интервала осуществляется по PDF, определяющей пики двух первоначальных функций. Четвертый и пятый графики каждого рисунка показывают функции, полученные после преобразований, полученных в результате выборки по значимости.



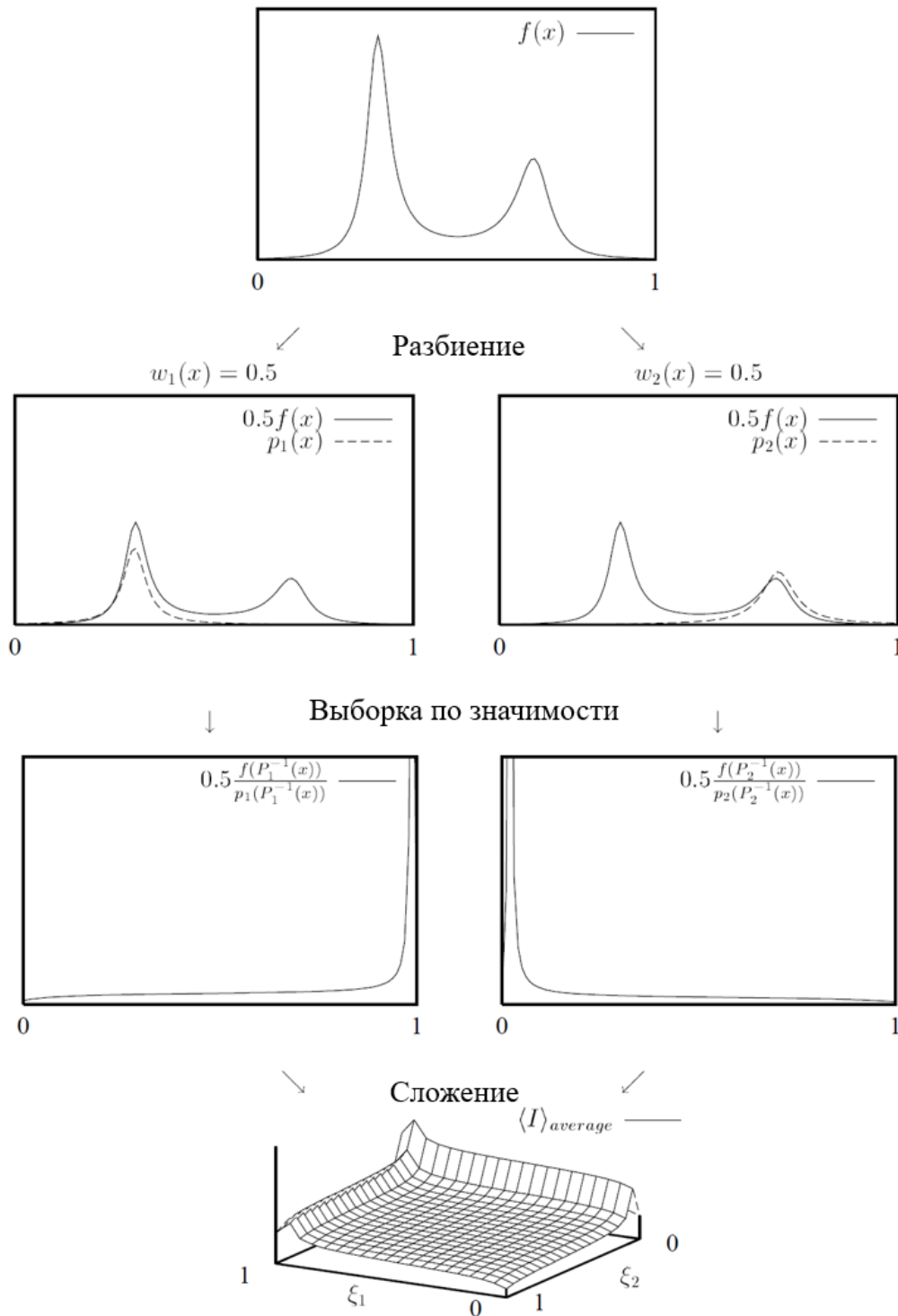


Рис. 4.9. Пример подынтегрального выражения  $f(x)$ , являющегося произведением двух функций с явно выраженными максимумами. Подынтегральное выражение разбивается путем деления его пополам. Обе части выбраны в соответствии с их собственными PDF, что эквивалентно показанным преобразованиям. Финально, оценки суммируются

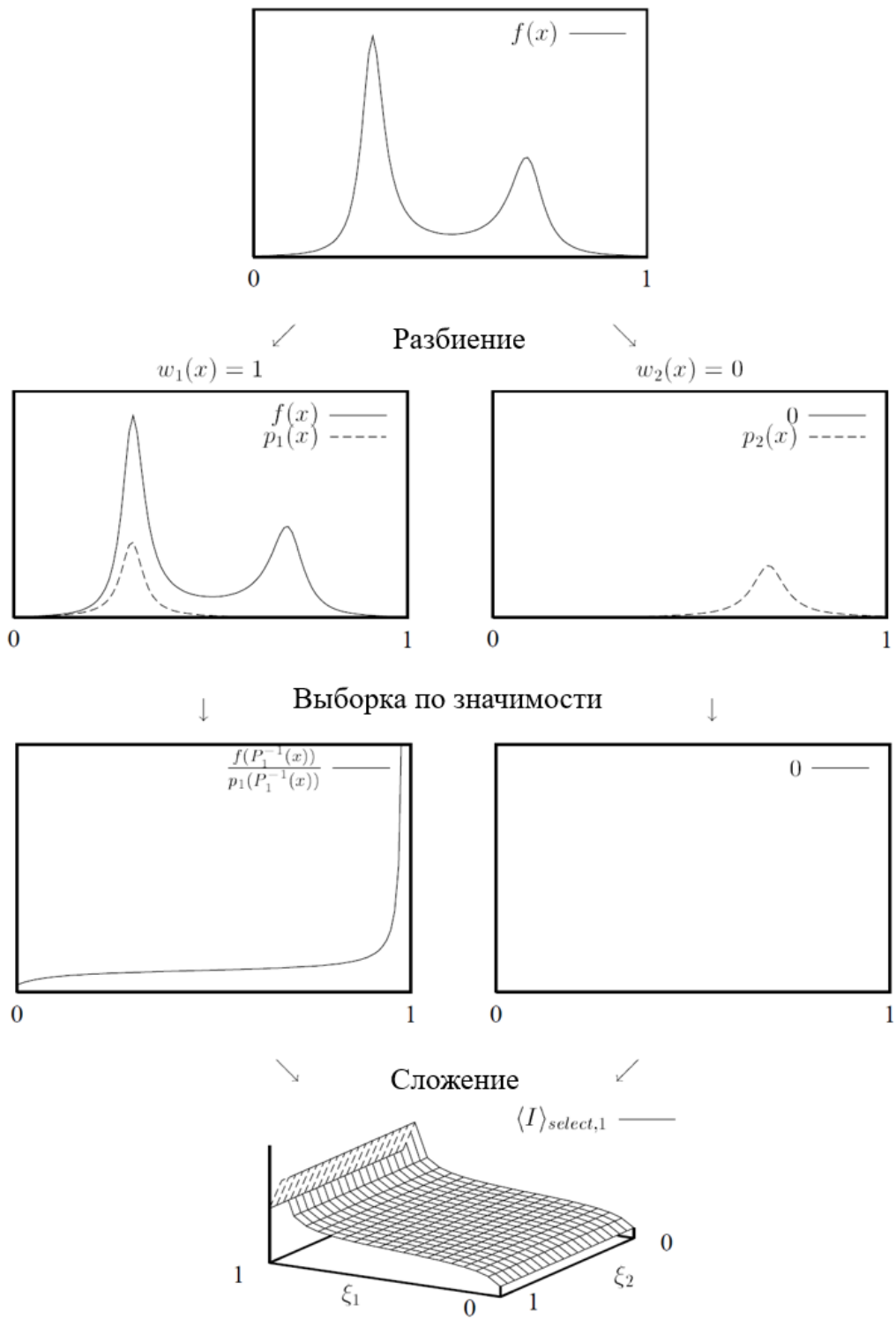


Рис. 4.10. То же самое подынтегральное выражение разбивается на себя и ноль. Выборка для отдельных частей осуществляется с помощью их собственных PDF и их оценки суммируются. Весовая функция выбирается в соответствии с первой оценкой

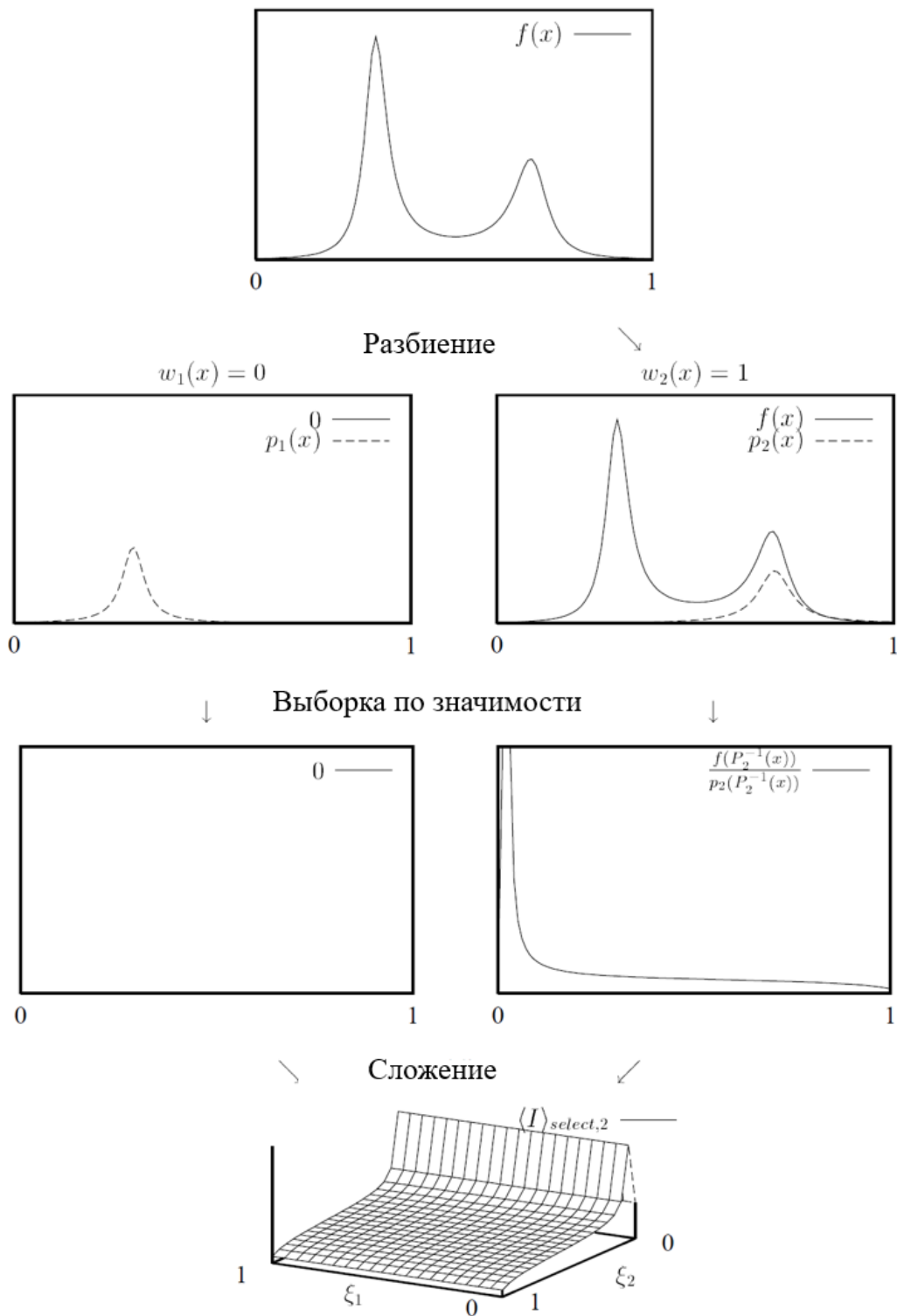


Рис. 4.11. То же самое подынтегральное выражение разбивается на ноль и себя. Выборка для отдельных частей осуществляется с помощью их собственных PDF и их оценки суммируются. Весовая функция выбирается в соответствии со второй оценкой

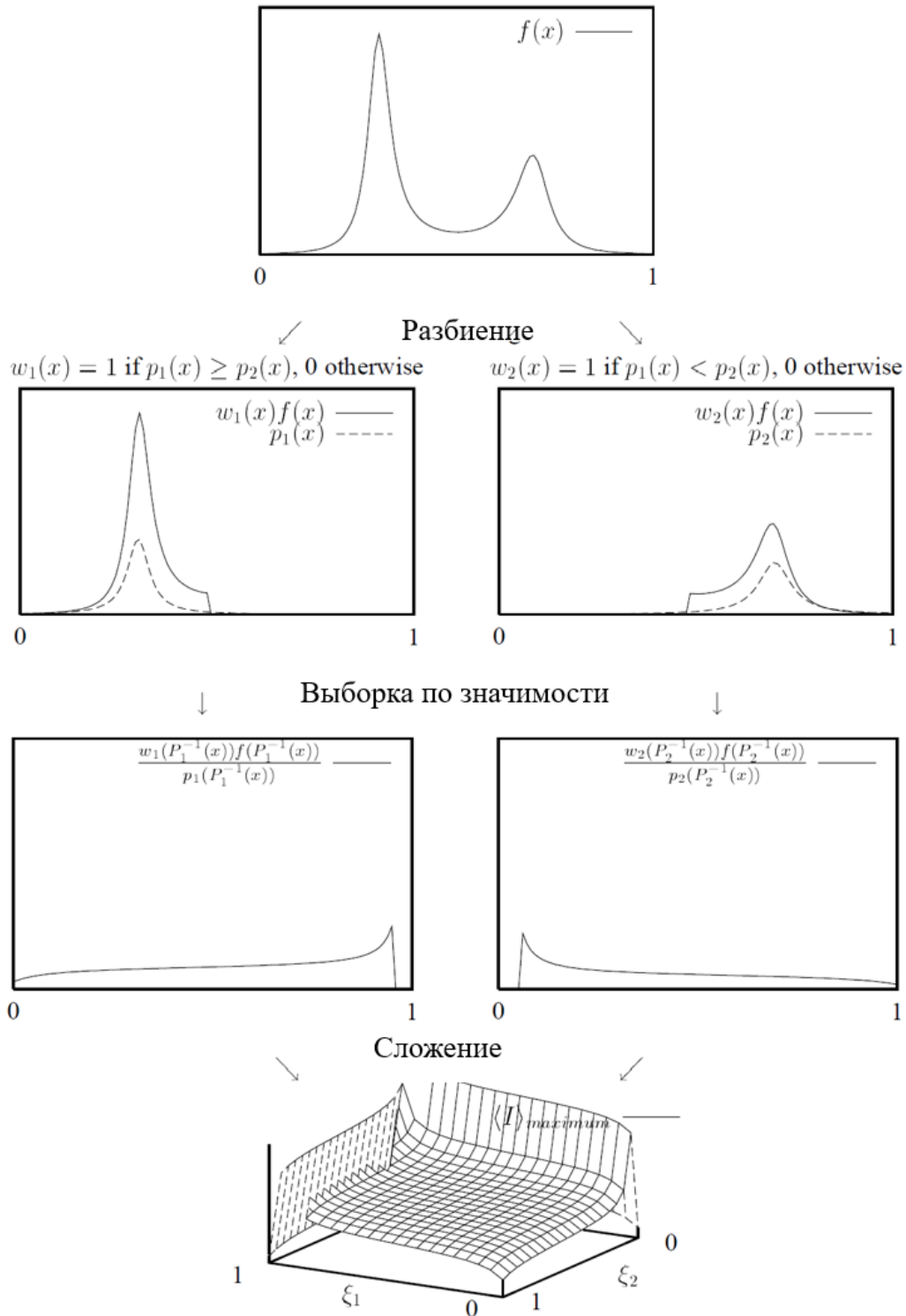


Рис. 4.12. То же самое подынтегральное выражение разбивается на основе максимальной эвристики. Весовые функции равны единице в точках, где соответствующая PDF принимает максимальное значение. Выборка осуществляется для обеих частей, результаты складываются

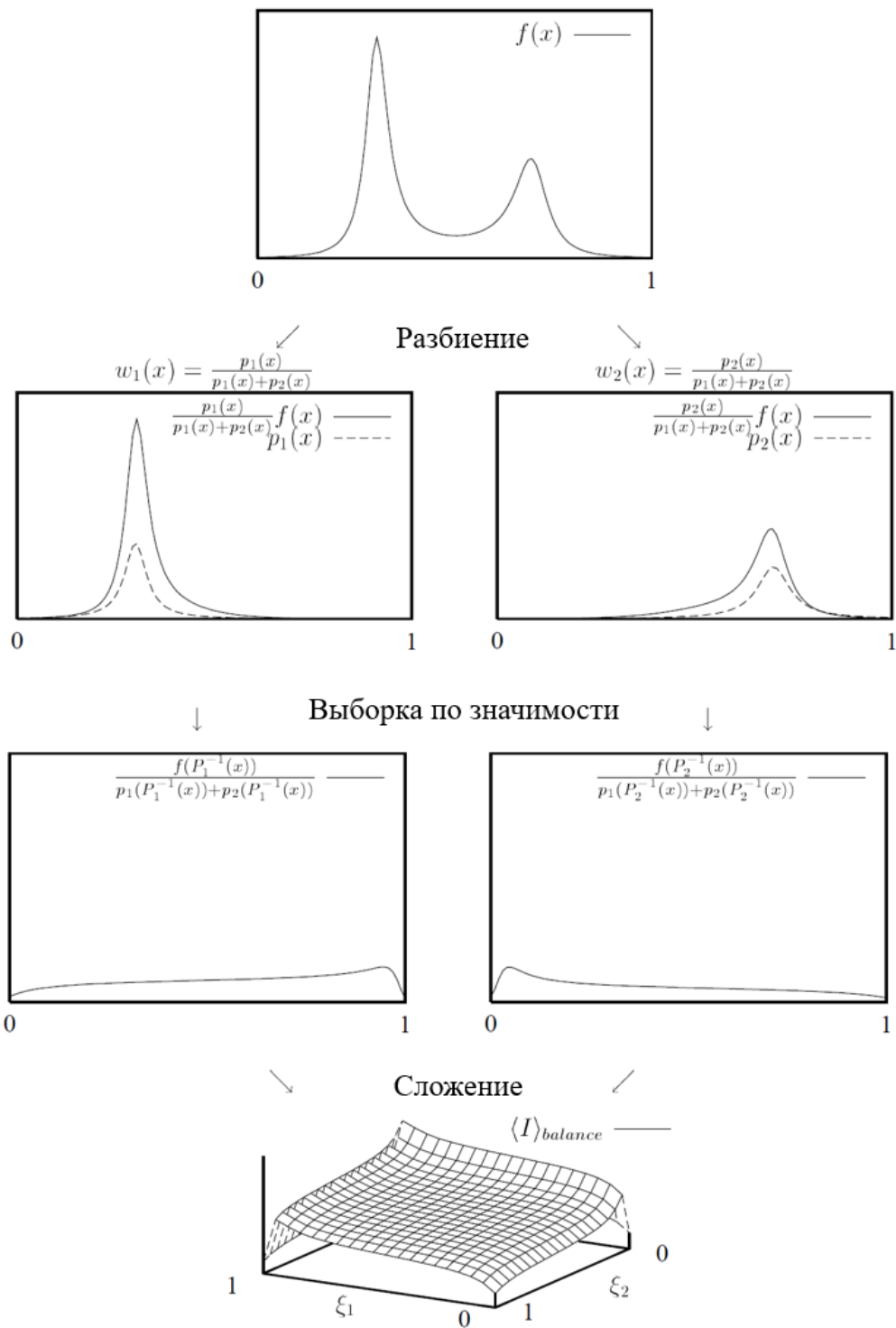


Рис. 4.13. То же самое подынтегральное выражение разбивается на основе сбалансированной эвристики. Весовые функции имеют большее значение в областях, где соответствующая PDF принимает большее значение. Выборка осуществляется для обеих частей, результаты складываются

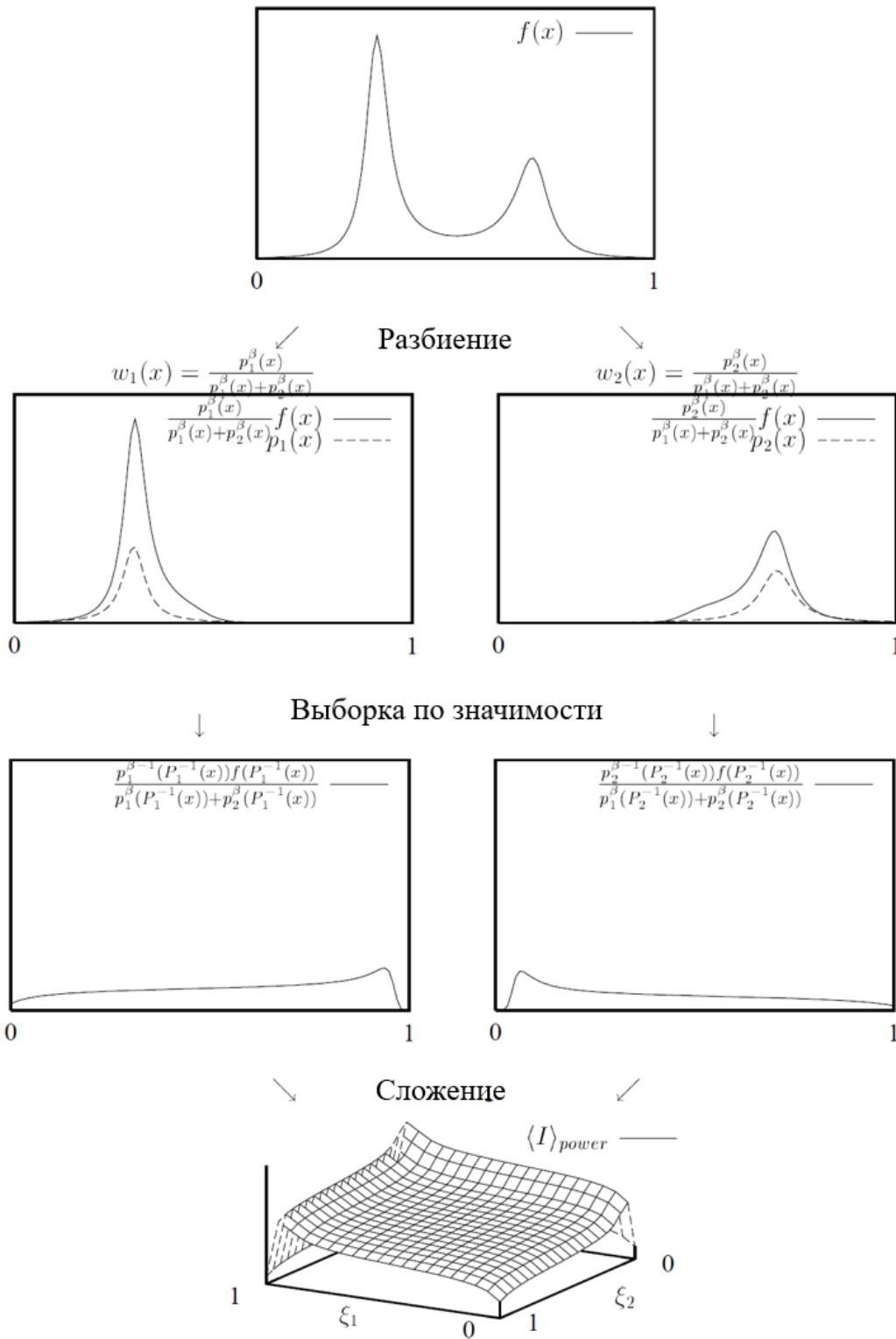


Рис. 4.14. То же самое подынтегральное выражение разбивается на основе степенной эвристики с параметром  $\beta = 2$ . Весовые функции равны, имеют большее значение в областях, где соответствующая PDF принимает большее значение (немного больше чем в случае сбалансированной эвристики)

Оба преобразованных интеграла оцениваются с использованием одной выборки, и их оценки суммируются с указанными весами. Это эквивалентно взятию одной выборки двумерной функции, представленной в нижней части каждого рисунка.

Необходимо отметить, что если выборки будут коррелированными, то результирующая оценка становится непредсказуемой, поэтому при использовании метода многократной выборки по значимости необходимо использовать только некоррелированные выборки.

Дисперсия в данном случае является суммой отклонений, поскольку используются некоррелированные выборки. Чем более постоянными являются обе преобразованные функции  $\frac{f(x)}{p(x)}$ , тем ниже будет дисперсия. В этом примере максимальная эвристика создает функции, которые намного более постоянны, чем исходная функция. Однако в некоторых подинтервалах значения функции падают до нуля, тем самым существенно увеличивая дисперсию. Эффект аналогичен выборке "попадание или промах", где подынтегральное выражение оценивается путем присвоения выборке 1 или константы, если она находится в пределах объема, подлежащего интегрированию, и 0 в противном случае. Этот метод может быть очень неэффективен и может оказать негативное влияние на максимальную эвристику. Сбалансированная эвристика создает более постоянные функции и, следовательно, результаты ее оценки лучше. Степенная эвристика в данном примере лежит где-то между обеими эвристиками и в данном случае оказалась малоэффективной. Однако в большинстве задач рендеринга, использующих метод многократной выборки по значимости, степенная эвристика оказывается более эффективной, чем сбалансированная.

#### 4.5 Метод выделения главной части

Требованиям для формирования PDF из функции, аппроксимирующей подынтегральное выражение, достаточно сложно удовлетворить. Однако, если известна функция  $g(x)$ , аппроксимирующая подынтегральную функцию  $f(x)$ , которая может быть проинтегрирована аналитически, то ее можно использовать в качестве главной части исходного интеграла, см. рис. 4.15. Преобразуем исходный интеграл (4.1) к виду:

$$\begin{aligned} I &= \int_0^1 f(x)dx = \int_0^1 (f(x) - g(x))dx + \int_0^1 g(x)dx \\ &= \int_0^1 (f(x) - g(x))dx + J = \int_0^1 (f(x) - g(x) + J)dx \end{aligned}$$

Если случайная величина  $\xi$  равномерно распределена в интервале  $[0,1]$ , то первичная оценка для последнего выражения становится равной

$$\langle I \rangle_{cont} = (f(\xi) - g(\xi) + J). \quad (4.22)$$

Ожидаемое значение полученной оценки равно интегралу. При этом дисперсия будет равной

$$\sigma_{cont}^2 = \int_0^1 (f(x) - g(x) + J)^2 dx - I^2. \quad (4.23)$$

Как и в случае выборки по значимости, в идеале  $f(x) = g(x)$ , данное решение свело бы дисперсию к нулю, но для этого потребовалось бы знать точное значение интеграла, что на практике практически невозможно. Можно отметить, что метод выделения главной части, который только аппроксимирует подынтегральную функцию, оказывает на нее сглаживающее действие и тем самым уменьшает дисперсию оценки интеграла.

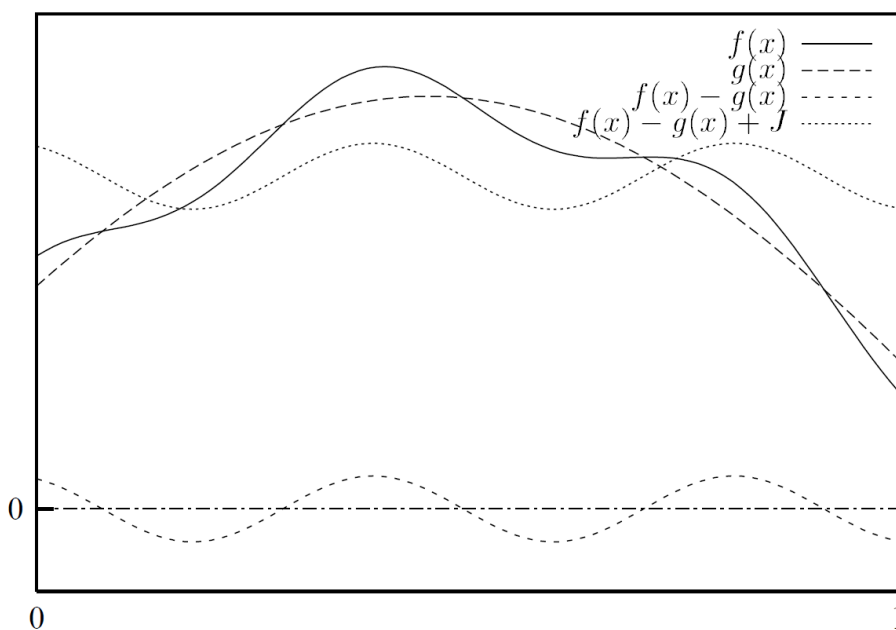


Рис. 4.15. Функция  $g(x)$ , аппроксимирующая  $f(x)$  и интегрируемая аналитически, может служить главной частью для оценки интеграла. Оценка интеграла разницы достаточно близких функций с последующей добавкой известного значения интеграла  $J$  является более эффективным решением, чем оценка интеграла для исходной функции  $f(x)$ . Как и в случае выборки по значимости, снижение дисперсии происходит в результате того, что преобразованная функция является более постоянной



В отличие от выборки по значимости, использование метода выделения главной части не требует расчета кумулятивной функции вероятности и нахождения обратной функции. К недостаткам последнего решения можно отнести необходимость не только достаточно точного понимания поведения подынтегральной функции, но и знания ее точного значения. Метод выборки по значимости свободен от последнего ограничения, поскольку функция плотности вероятности нормализуется. Если функция  $g(x)$  используется для выборки по значимости, то применять ее далее в качестве главной части не имеет смысла. В работах [34, 35] проводится краткое сравнение эффективности выборки по значимости и эффективности использования метода выделения главной части. Интуитивный вывод заключается в том, что для функций, являющихся приближением подынтегральной функции с постоянной относительной ошибкой, более эффективны решения на основе метода выборки по значимости, однако, если приближение к подынтегральной функции происходит с малыми абсолютными ошибками, то более эффективным становится метод выборки с выделением главной части.

#### 4.6 Решение интегральных уравнений методом Монте-Карло

Рассмотрим интегральное уравнение Фредгольма второго рода:

$$f(x) = g(x) + \int_0^1 K(x, y)f(y)dy, \quad (4.24)$$

где  $f(x)$  – неизвестная функция,  $g(x)$  – известная функция,  $K(x, y)$  – ядро интегрального уравнения. Если это уравнение не может быть решено аналитически, что происходит практически во всех задачах реалистичного рендеринга, то оно может быть решено численно. Метод конечных элементов является одним из способов численного определения значений функции  $f(x)$  как суммы базисных функций. С технической точки зрения восстановление всей функции не всегда возможно, поскольку, например, эта функция может оказаться слишком громоздкой для хранения. Однако, если нет необходимости в определении всей функции, а надо вычислить ее значения только в ряде точек выборки, например, задача рендеринга заключается в вычислении этой функции только в точках изображения, то альтернативным подходом может служить метод Монте-Карло. Этот метод вычисляет  $f(x)$  для фиксированного  $x$ , оценивая правую часть уравнения так, как если бы это был обычный интеграл. Поскольку это, в свою очередь, требует оценки некоторого значения  $f(y)$ , то процесс повторяется рекурсивно. Если все последующие интегралы оцениваются для значения случайной величины  $\xi_i$  в интервале  $[0,1]$  с плотностью вероятности  $p_i(x)$ , то результирующая первичная оценка будет выглядеть как

$$\begin{aligned}
\langle f(x) \rangle_{rec} &= g(x) + \frac{K(x, \xi_1)}{p_1(\xi_1)} \langle f(\xi_1) \rangle_{rec} \\
&= g(x) + \frac{K(x, \xi_1)}{p_1(\xi_1)} \left[ g(\xi_1) + \frac{K(\xi_1, \xi_2)}{p_2(\xi_2)} \langle f(\xi_2) \rangle_{rec} \right] \\
&= g(x) + \frac{K(x, \xi_1)}{p_1(\xi_1)} g(\xi_1) + \frac{K(x, \xi_1) K(\xi_1, \xi_2)}{p_1(\xi_1) p_2(\xi_2)} g(\xi_2) + \dots \\
&= \sum_{i=0}^{\infty} \left[ \prod_{j=1}^i \frac{K(\xi_{j-1}, \xi_j)}{p_j(\xi_j)} \right] g(\xi_i),
\end{aligned} \tag{4.25}$$

где  $x = \xi_0$ . Последовательность случайных чисел  $\xi_1, \xi_2, \dots$  называется марковскими цепочками, если PDF  $p_i(\xi_i)$  в каждой точке  $\xi_i$  зависит от предыдущего значения  $\xi_{i-1}$ . Этот процесс можно также называть случайным блужданием, и визуально он представляет собой серию случайных переходов от одной точки к области интегрирования к другой. Интегральное уравнение можно записать с помощью интегрального оператора:

$$f = g + Tf$$

Решение можно записать в виде ряда Неймана:

$$f = g + Tf + T^2f + T^3f + T^4f + \dots$$

Первичную оценку решения интегрального уравнения (4.24) можно рассматривать как оценку этой суммы интегралов методом Монте-Карло.

#### 4.6.1 Русская рулетка

Бесконечное число членов в оценке интеграла (4.24) представляет проблему для практической реализации. Часть членов ряда необходимо в какой-то момент отрезать. Если ряд Неймана сходится, то его конечные члены будут стремиться к 0. Однако усечение ряда в некоторой фиксированной точке эквивалентно тому, что все последующие члены ряда приравниваются нулю. С теоретической точки зрения это приводит к некорректному результату. С практической точки зрения никогда нельзя быть уверенным, что отбрасываемые члены ряда не важны по отношению к накопленной сумме. Более надежная техника отбрасывания членов ряда – это русская рулетка. Ни один из членов ряда не может быть отсечен детерминированным образом, но каждый член и вся последующая цепочка может быть обрезана вероятностным образом. Выполнив операцию деления оценки интеграла на правильную вероятность, компенсирующую вероятностное усечение, мы получаем несмещенную оценку. Первичная оценка простого интеграла (4.1) может быть получена простым преобразованием, см. рис. 4.16:

$$I = \int_0^1 f(x)dx = \int_0^R \frac{1}{R} f\left(\frac{t}{R}\right) dt = \int_0^1 \frac{1}{R} f\left(\frac{t}{R}\right) u\left(\frac{t}{R}\right) dt,$$

где:

- $R$  – некоторое фиксированное число в диапазоне  $(0,1]$ ,
- $u(x)$  – ступенчатая функция, которая равна 1 в диапазоне  $x \leq 1$  и 0 в диапазоне  $x > 1$ .

Первичная оценка интеграла, вычисленная данным методом при равномерном распределении случайной величины  $\xi$  в диапазоне  $[0,1]$ , примет следующий вид:

$$\langle I \rangle_{Rus} = \begin{cases} \frac{1}{R} f\left(\frac{\xi}{R}\right), & \xi \leq R, \\ 0, & \text{иначе.} \end{cases} \quad (4.26)$$

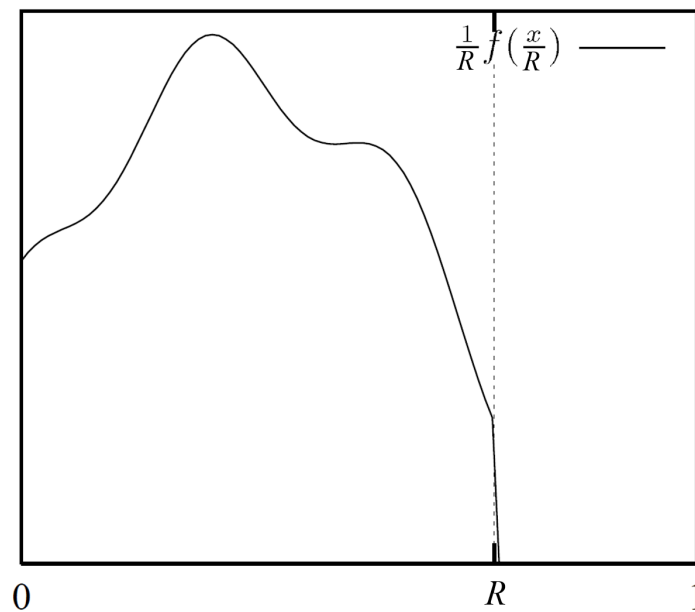


Рис. 4.16. Русскую рулетку можно рассматривать как трансформацию исходного интеграла. Подынтегральная функция уменьшена до значения  $R$  по оси  $x$  и увеличена в  $1/R$  раз по оси  $y$ . Значение интеграл остается прежним. Оценка, полученная методом Монте-Карло, преобразованной функции в интервале  $[0,1]$  дает то же самое математическое ожидание, но с большей дисперсией. Данная техника используется в основном для завершения бесконечной рекурсии при решении интегрального уравнения

Оценка, приведенная в (4.26), показывает, что  $R$  фактически является вероятностью того, что подынтегральная функция  $f(x)$  будет реально оценена. Вероятность того, оценка будет нулевой, равна  $1 - R$ . Поэтому русская рулетка часто реализуется путем выбора двух случайных величин, первая из которых используется для определения того, равна ли оценка нулю или нет, а вторая случайная величина используется для фактической оценки интеграла, если была выбрана ненулевая оценка. Можно показать, что использование русской рулетки увеличивает дисперсию, если метод применяется в комбинации со стратифицированной выборкой ( $N$ -roks выборкой). Таким образом, методы оценки, рассмотренные ранее, являются более предпочтительными, если не возникает необходимости прерывать бесконечную рекурсию при решении интегрального уравнения.

Русская рулетка формирует несмещенную оценку. Однако ее дисперсия выше дисперсии, которая может быть получена с помощью стандартной оценки (4.3):

$$\begin{aligned} \sigma_{Rus}^2 &= \int_0^1 \left[ \frac{1}{R} f\left(\frac{\xi}{R}\right) u\left(\frac{\xi}{R}\right) \right]^2 d\xi - I^2 = \int_0^R \frac{1}{R^2} f^2\left(\frac{\xi}{R}\right) d\xi - I^2 \\ &= \frac{1}{R} \int_0^1 f^2(\xi) d\xi - I^2 > \sigma_{prim}^2 \end{aligned} \quad (4.27)$$

#### 4.6.2 Комбинация русской рулетки с методами выборки по значимости и методом выделения главной части

Необходимо проанализировать эффекты сочетания русской рулетки, выборки по значимости и метода выделения главной части, поскольку эти результаты будут представлять практический интерес для алгоритмов расчета глобальной освещенности.

Как было отмечено выше, функция, используемая в качестве PDF для выборки по значимости, не дает никаких дополнительных преимуществ для снижения дисперсии, если ее используют в качестве главной части. Однако метод выделения главной части может снизить дисперсию для русской рулетки. Рассмотрим функцию  $g(x)$ , которая аппроксимирует подынтегральное выражение  $f(x)$ , и для которой можно аналитически вычислить интеграл ( $J = \int_0^1 g(x) dx$ ), чтобы создать PDF ( $p(x) = g(x)/J$ ). Если мы применим выборку по значимости после русской рулетки, то первичная оценка будет комбинацией уравнений (4.12) и (4.18):

$$\langle I \rangle_{Rus,imp} = \begin{cases} \frac{1}{R} \frac{f(P^{-1}(\tau/R))}{p(P^{-1}(\tau/R))}, & \tau \leq R, \\ 0, & \text{иначе.} \end{cases} \quad (4.28)$$

На рис. 4.17а показано преобразованное подынтегральное выражение примера, представленного на рис. 4.6. Функция стала более постоянной в области, отличной от нуля. При этом дисперсия равна

$$\sigma_{Rus,imp}^2 = \frac{1}{R} \int_0^1 \frac{f^2(P^{-1}(\tau))}{p^2(P^{-1}(\tau))} d\tau - I^2. \quad (4.29)$$

Если функция  $g(x)$  используется в качестве главной части перед применением русской рулетки, то первичная оценка будет:

$$\begin{aligned} \langle I \rangle_{Rus,imp,con} &= \begin{cases} \frac{1}{R} \frac{f(P^{-1}(\tau/R)) - g(P^{-1}(\tau/R))}{p(P^{-1}(\tau/R))} + J, & \tau \leq R, \\ J, & \text{иначе.} \end{cases} \\ &= \begin{cases} \frac{1}{R} \frac{f(P^{-1}(\tau/R))}{p(P^{-1}(\tau/R))} - \left(\frac{1}{P} - 1\right)J, & \tau \leq R, \\ J, & \text{иначе.} \end{cases} \end{aligned} \quad (4.30)$$

На рис. 4.17б показано финальное подынтегральное выражение. Подынтегральная функция стала более постоянной, поскольку ее часть, отличающаяся от 0, сместилась вниз, а постоянная нулевая часть была поднята выше. Даже если выбирается область за  $R$ , где исходная функция не вычисляется, оценка интеграла будет отличаться от 0. Можно показать, что дисперсия равна

$$\sigma_{Rus,imp,con}^2 = \frac{1}{R} \int_0^1 \frac{f^2(P^{-1}(\tau))}{p^2(P^{-1}(\tau))} d\tau - \left(\frac{1}{P} - 1\right) (2I - J)J - I^2. \quad (4.31)$$

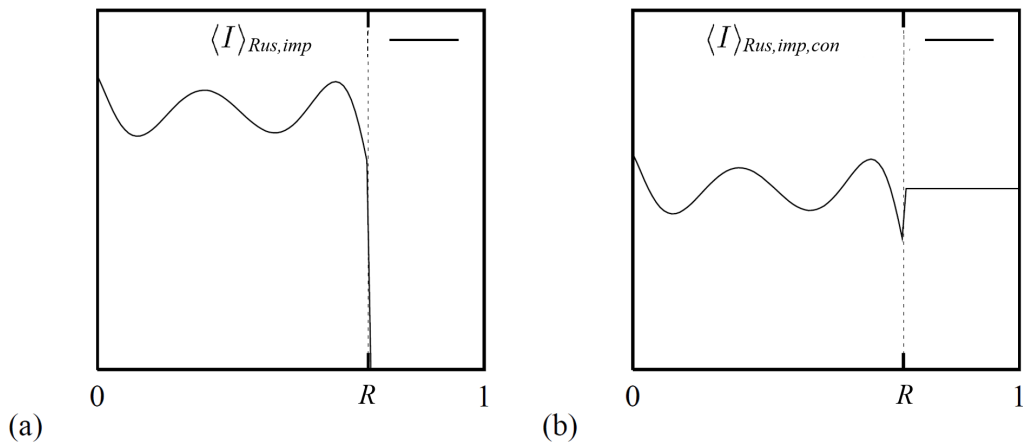


Рис. 4.17. (а) Применение выборки по значимости к подынтегральной функции, показанной на рис. 4.6, после русской рулетки сглаживает подынтегральную функцию. (б) Использование функции, пропорциональной PDF, в качестве главной части обычно не дает никакого дополнительного эффекта. Однако, применение этого метода перед русской рулеткой позволяет сгладить подынтегральное выражение

Сравнение уравнений (4.29) и (4.31) показывает, что дисперсия будет меньше, если интеграл аппроксимирующей функции лежит достаточно близко к его фактическому значению. А именно:

$$\sigma_{Rus,imp,con}^2 < \sigma_{Rus,imp}^2 \Leftrightarrow 0 < J < 2I. \quad (4.32)$$

### 4.6.3 Применение русской рулетки для решения интегральных уравнений

Русская рулетка обычно применяется тогда, когда интеграл фактически является частью интегрального уравнения. Примером такого интегрального уравнения может служить уравнение глобальной освещенности, являющееся уравнением Фредгольма второго рода. Если русская рулетка применяется на каждом уровне рекурсии, случайное блуждание гарантированно завершается. Это становится конечным рядом  $\xi_1, \xi_2, \dots, \xi_k$ , потому что интеграл оценивается нулем в точке  $k$ . Если каждый  $i$  выбирается в соответствии с PDF  $p_i(x)$  с вероятностью  $R_i$ , то первичная оценка (4.25) в случае интегрального уравнения (4.24) становится конечной суммой:

$$\begin{aligned}
\langle f(x) \rangle_{Rus,rec} &= g(x) + \frac{K(x, \xi_1)}{R_1 p_1(\xi_1)} g(\xi_1) + \frac{K(x, \xi_1) K(\xi_1, \xi_2)}{R_1 p_1(\xi_1) R_2 p_2(\xi_2)} g(\xi_2) + \dots \\
&+ \frac{K(x, \xi_1) K(\xi_1, \xi_2)}{R_1 p_1(\xi_1) R_2 p_2(\xi_2)} g(\xi_2) \dots \frac{K(\xi_{k-1}, \xi_k)}{R_k p_k(\xi_k)} g(\xi_k) \\
&= \sum_{i=0}^k \left[ \prod_{j=1}^i \frac{K(\xi_{j-1}, \xi_j)}{R_j p_j(\xi_j)} \right] g(\xi_i),
\end{aligned} \tag{4.33}$$

где  $\xi_0 = x$ .

Для большинства физических задач нормализованное значение ядра  $K(x, y)$  меньше, чем 1, для любого  $x$  в диапазоне  $[0, 1]$ :

$$\int_0^1 K(x, y) < 1.$$

Затем ядро можно использовать для модификации PDF. Это соответствует выбору вероятности выборки  $R_i$ , равной интегралу ядра, и нормализации PDF на эту вероятность:

$$\begin{aligned}
R_i &= \int_0^1 K(\xi_{i-1}, y) dy \\
p_i(x) &= \frac{K(\xi_{i-1}, x)}{R_i}
\end{aligned}$$

Таким образом, использовать полученную PDF для оценок столь же допустимо, как и обычную PDF. В этом случае первичная оценка упрощается до элегантного выражения:

$$\langle f(x) \rangle_{subPDF} = \sum_{i=1}^k g(\xi_k). \tag{4.34}$$

Этот подход полностью соответствует идее выборки по значимости и нашел естественное применение в задачах расчета глобальной освещенности.

#### 4.6.4 Оценка следующего события

В наших приложениях оценка (4.33) страдает от высокой дисперсии, поскольку исходная функция  $g(x)$ , как правило, равна 0 во всей области. Оценка может случайным образом получить большие вклады, что в итоге только ухудшит ситуацию. В этом случае «корректная» оценка следующего события дает большие преимущества. Исходный интеграл интегрального уравнения (4.23) может быть разделен на два отдельных интеграла, каждый из которых решается по-разному:

$$f(x) = g(x) + \int_0^1 K(x, y)f(y)dy = g(x) + h(x),$$

где:

$$\begin{aligned} h(x) &= \int_0^1 K(x, y)f(y)dy = \int_0^1 K(x, y)[g(y) + h(y)]dy \\ &= \int_0^1 K(x, y)g(y)dy + \int_0^1 K(x, y)h(y)dy \end{aligned}$$

Для первого интеграла предпочтительная оценка должна соответствовать PDF, форма которой аналогична доминирующему множителю  $g(y)$  подынтегрального выражения. Для второго подынтегрального выражения интеграл  $h(y)$  неизвестен, поэтому он выбирается, как и раньше. Формально, если на каждом уровне рекурсии  $\zeta_i$  отбирается в соответствии с PDF  $p_i(z)$ , и каждый  $\xi_i$  – в соответствии с модифицированным PDF  $K(\xi_{i-1}, x)$ , то первичная оценка становится равной

$$\langle f(x) \rangle_{nextev} = g(x) + \langle h(x) \rangle_{nextev},$$

где:

$$\begin{aligned} \langle h(x) \rangle_{nextev} &= \frac{K(x, \zeta_1)}{p_1(\zeta_1)} g(\zeta_1) + \langle h(\xi_1) \rangle_{nextev} \\ &= \frac{K(x, \zeta_1)}{p_1(\zeta_1)} g(\zeta_1) + \frac{K(\xi_1, \zeta_2)}{p_1(\zeta_1)} g(\zeta_2) + \langle h(\xi_2) \rangle_{nextev} = \dots \\ &= \sum_{i=1}^k \frac{K(\xi_{i-1}, \zeta_i)}{p_i(\zeta_i)} g(\zeta_i) \end{aligned}$$

Объединяя два уравнения, получаем:

$$\langle f(x) \rangle_{nextev} = g(x) + \sum_{i=1}^k \frac{K(\xi_{i-1}, \zeta_i)}{p_i(\zeta_i)} g(\zeta_i). \quad (4.35)$$

#### 4.7 Формирование случайных величин с заданным PDF

В этом разделе были представлены основные решения методов Монте-Карло, которые важны для реализации алгоритмов расчета глобального освещения. Одной из важных составляющих методов Монте-Карло является формирование распределений в многомерном пространстве с заданной плотностью распределения случайной величины. Далее будет рассмотрен ряд базовых примеров формирования случайных величин.



### 4.7.1 Равномерное распределение случайной координаты на плоских фигурах

В трехмерной сцене геометрические объекты, как правило, формируются плоскими примитивами, такими как прямоугольник, треугольник и круг. В процессе рендеринга, основанного на методе Монте-Карло, возникают задачи формирования распределений точек на поверхности с заданной плотностью вероятности, например, при испускании луча с протяженного источника света. В данном разделе будут рассмотрены простейшие способы формирования равномерного распределения точек.

#### *Прямоугольник*

Рассмотрим прямоугольник, представленный на рис. 4.18.

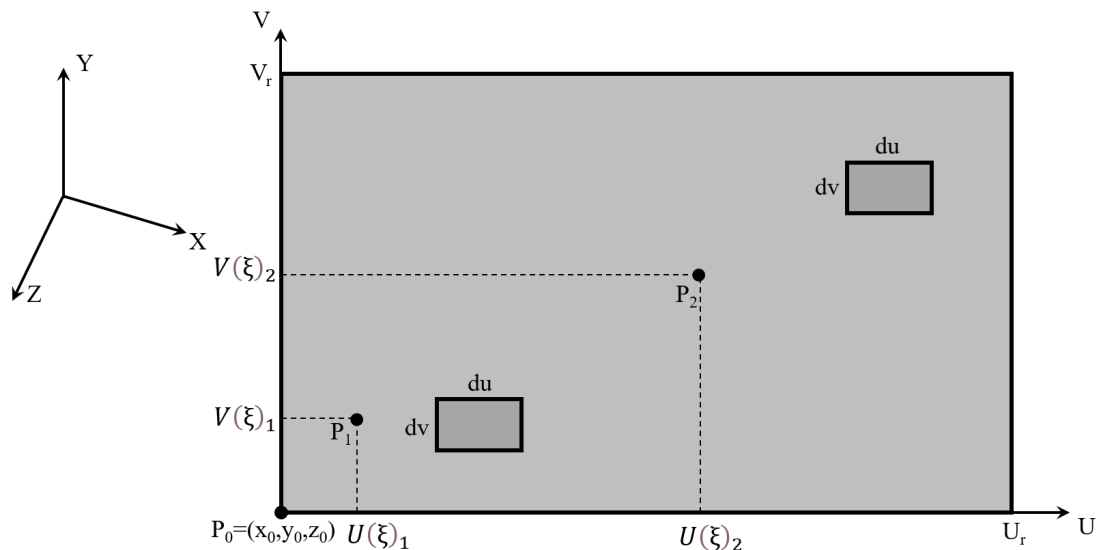


Рис. 4.18. Формирование равномерного распределения точек внутри прямоугольника

Для формирования координат точек, равномерно распределенных внутри прямоугольника, построим систему координат с осями по сторонам прямоугольника  $(U, V)$ . При равномерном распределении случайной величины  $\xi$  по осям  $U$  и  $V$  в диапазоне  $[0, U_r[$  и  $[0, V_r[$  соответственно плотность распределения точек в прямоугольнике пропорциональна элементарным площадям  $du \cdot dv$ . Очевидно, что данная величина постоянна по всей поверхности прямоугольника, поэтому координата точки внутри прямоугольника может быть вычислена как

$$P = V_r \xi_v + U_r \xi_u, \quad (4.36)$$

где  $\xi_u$  и  $\xi_v$  – случайные числа, равномерно распределенные в диапазоне  $[0,1[$ .

Принимая во внимание, что прямоугольник может быть произвольным образом ориентирован в пространстве, положение точки на прямоугольнике, единичные вектора сторон которого в системе координат  $XYZ$  имеют направления  $U$  и  $V$ , а вершина прямоугольника расположена в точке  $P_0$ , можно записать следующим образом:

$$P = P_0 + V(V_r \xi_v) + U(U_r \xi_u). \quad (4.37)$$

### Треугольник

Рассмотрим треугольник, представленный на рис. 4.19.

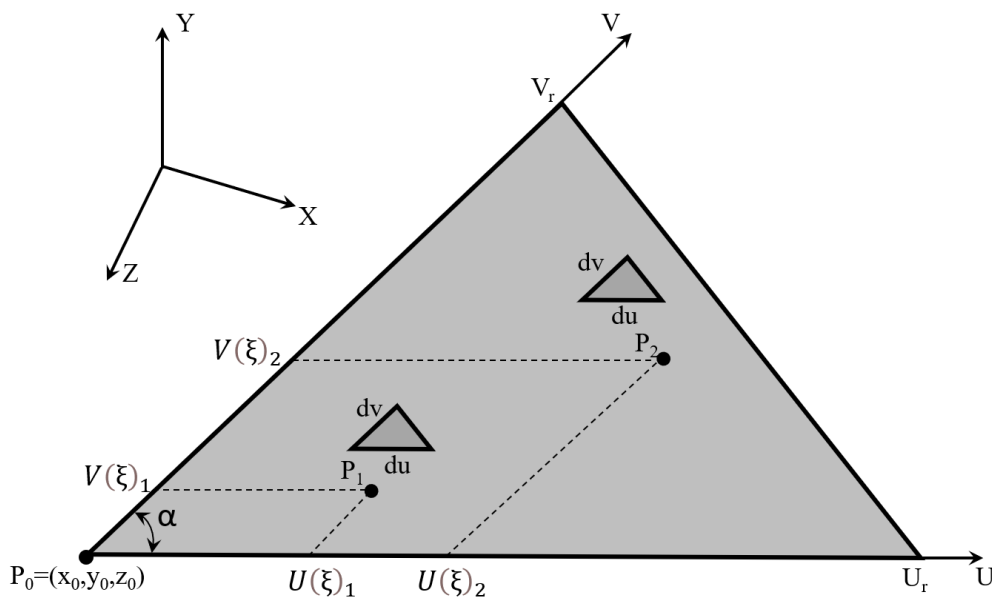


Рис. 4.19. Формирование равномерного распределения точек внутри треугольника

Для распределения координат точек, равномерно распределенных внутри треугольника, построим систему координат с осями по сторонам треугольника  $(U, V)$ . При равномерном распределении случайной величины  $\xi$  по осям  $U$  и  $V$  в диапазоне  $[0, U_r[$  и  $[0, V_r[$  соответственно плотность распределения точек в треугольнике пропорциональна элементарным площадям  $du \cdot dv \cdot \sin \alpha / 2$ . Очевидно, что данная величина постоянна по всей поверхности прямоугольника, поэтому координата точки внутри прямоугольника может быть вычислена как

$$P = V_r \xi_v + U_r \xi_u, \text{ если внутри треугольника,}$$

где  $\xi_u$  и  $\xi_v$  – случайные числа, равномерно распределенные в диапазоне  $[0,1[$ .

Необходимо обратить внимание, что точка формируется не в треугольнике, а в параллелограмме, т.е. половина точек будет находиться вне треугольника, и их надо отбросить. Критерием попадания точки внутрь треугольника может служить условие  $\xi_v + \xi_u \leq 1$ . Далее возможны два варианта. Первый вариант представляет собой выборку с отклонением, суть которой заключается в том, что выборка случайной величины в некоторой подобласти может быть сформирована соответствующей выборкой в области с последующим отбрасыванием образцов, оказавшихся вне подобласти. В данном случае выборка осуществляется в параллелограмме, и точки, не попавшие в треугольник, отбрасываются. Поэтому, если  $\xi_v + \xi_u > 1$ , то точка отбрасывается, формируются новые случайные величины до тех пор, пока  $\xi_v + \xi_u \leq 1$ . Далее формируется точка в треугольнике. Второй вариант: если  $\xi_v + \xi_u > 1$ , то формируются «зеркальные» случайные величины  $1 - \xi_v$  и  $1 - \xi_u$ , для которых условие попадания в треугольник выполняется. В результате точка формируется следующим образом:

$$P = V_r \xi_v + U_r \xi_u, \text{ если } \xi_v + \xi_u \leq 1$$

$$P = V_r(1 - \xi_v) + U_r(1 - \xi_u), \text{ если } \xi_v + \xi_u > 1$$

Принимая во внимание, что треугольник может быть произвольным образом ориентирован в пространстве, положение точки на треугольнике, единичные вектора сторон которого в системе координат  $XYZ$  имеют направления  $U$  и  $V$ , а вершина треугольника расположена в точке  $P_0$ , можно записать следующим образом:

$$P = P_0 + V(V_r \xi_v) + U(U_r \xi_u), \text{ если } \xi_v + \xi_u \leq 1$$

$$P = P_0 + V(V_r(1 - \xi_v)) + U(U_r(1 - \xi_u)), \text{ если } \xi_v + \xi_u > 1 \quad (4.38)$$

### ***Круг***

Рассмотрим круг, представленный на рис. 4.20.

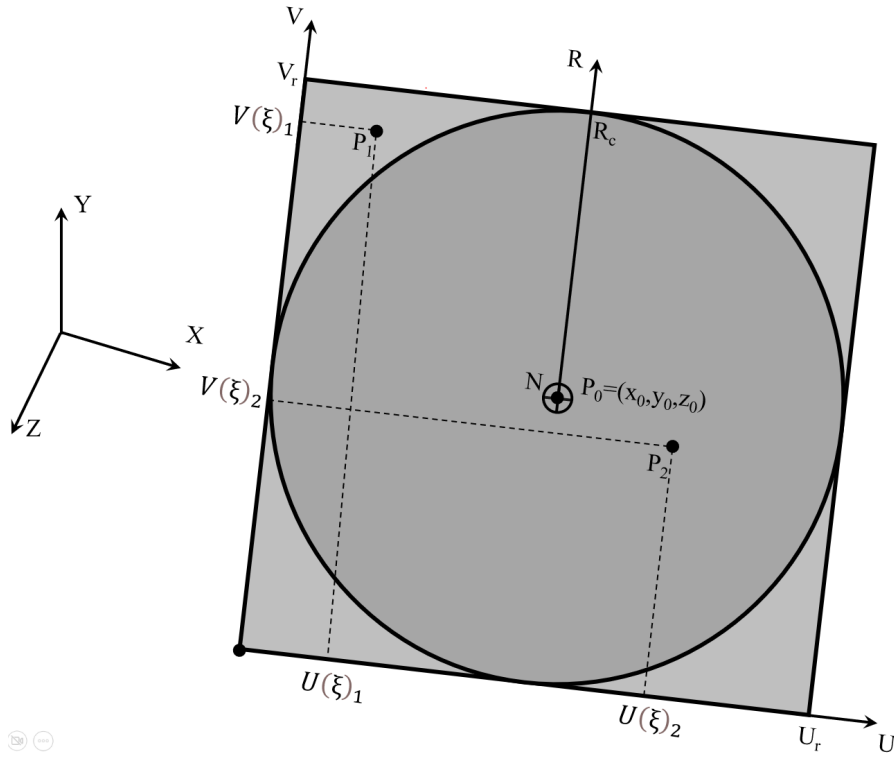


Рис. 4.20. Формирование равномерного распределения точек внутри круга (используется выборка с отклонением)

Для формирования распределения координат точек, равномерно распределенных внутри круга, построим систему координат относительно нормали к поверхности круга (круг можно задать положением его центра  $P_0$  и ориентацией нормали  $N$ ). Одна из осей круга выбирается из условия  $N \cdot V = 0$  (т.е. любой вектор, для которого скалярное произведение равно 0, может быть выбран за ось, например,  $(-N_y / \sqrt{N_y^2 + N_x^2}, N_x / \sqrt{N_y^2 + N_x^2}, 0) = U$ ), вторая ось  $N \times V = U$ . Далее формируется описанный квадрат, внутри которого создается равномерное распределение случайной величины. Используя алгоритм выборки с отклонением, отбираем точки, принадлежащие кругу:

$$P = P_0 - VR_c - UR_c + V(2R_c\xi_v) + U(2R_c\xi_u), \text{ если } \xi_v^2 + \xi_u^2 \leq 1, \quad (4.39)$$

где  $\xi_u$  и  $\xi_v$  – случайные числа, равномерно распределенные в диапазоне  $[0,1[$ .

Второй вариант формирования равномерного распределения точек внутри круга не предполагает использование выборки с отклонением, равномерное распределение точек будет формироваться непосредственно внутри круга, см. рис. 4.21. Выберем полярную систему координат, где положение точки определяется

радиусом и углом поворота относительно некого базового направления  $R$ , ортогонального направлению нормали  $N$ .

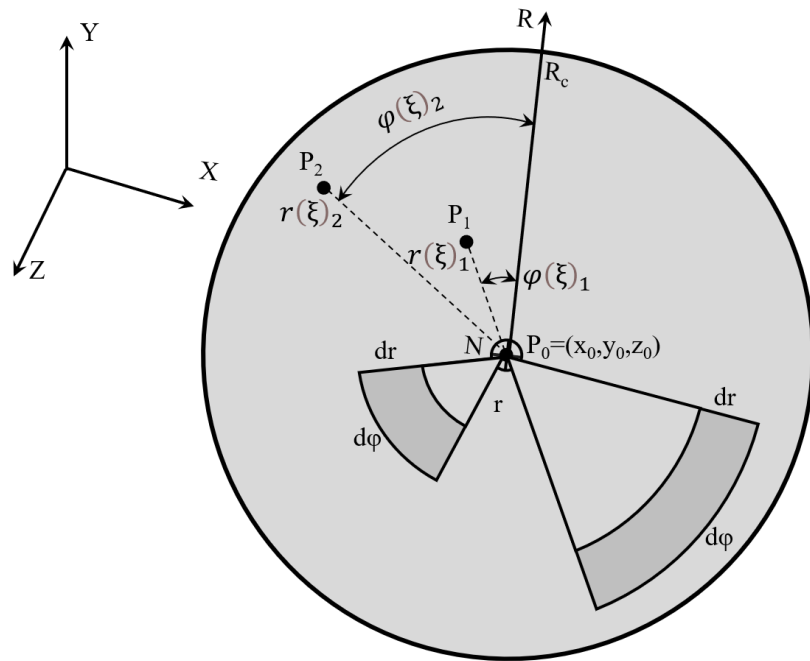


Рис. 4.21. Формирование равномерного распределения точек внутри круга

Для формирования равномерного распределения точек внутри круга необходимо (в полярной системе координат) построить функцию плотности вероятности, соответствующую площади кругового сегмента с фиксированными приращениями координат  $dr$  и  $d\varphi$ .

$$ds = \frac{1}{2} d\varphi ((r + dr)^2 - r^2) = rd\varphi dr. \quad (4.40)$$

Видно, что для координаты  $\varphi$  плотность вероятности постоянна. Поэтому данное двумерное распределение можно представить, как два независимых распределения по  $r$  и  $\varphi$ , причем по координате  $\varphi$  случайная величина  $\xi$  будет иметь равномерное распределение. Поэтому рассмотрим плотность вероятности по координате  $r$ :

$$p(r) = r \cdot$$

Введем нормировку плотности вероятности:

$$\int_0^{R_c} r dr = \frac{1}{2} R_c^2.$$

Вероятность будет иметь следующий вид:

$$P(r) = \frac{2}{R_c^2} \int_0^r t dt = \left(\frac{r}{R_c}\right)^2. \quad (4.41)$$

Обратная функция вероятности имеет следующий вид:

$$P(r')^{-1} = r = R_c \sqrt{r'}, \quad (4.42)$$

где  $r'$  - приведенный радиус, варьирующийся от 0 до 1. Следовательно, для построения равномерного распределения точек на круге необходимо формирование пары чисел:

$$\varphi = 2\pi\xi_\varphi, r = R_c\sqrt{\xi_r},$$

где  $\xi_r$  и  $\xi_\varphi$  – случайные числа, равномерно распределенные в диапазоне  $[0,1[$ .

Точка на круге в локальных координатах  $VUN$  (формирование системы координат было показано в предыдущем примере) будет иметь вид:

$$P_{VUN} = (r \cos \varphi, r \sin \varphi, 0). \quad (4.43)$$

Для формирования распределения в системе координат сцены  $XYZ$  необходимо построить матрицу перехода из системы координат  $VUN$  в  $XYZ$ . Матрица перехода имеет вид:

$$M = \begin{pmatrix} V_x & V_y & V_z \\ U_x & U_y & U_z \\ N_x & N_y & N_z \end{pmatrix}.$$

Координаты точки будут иметь вид:

$$P_{XYZ} = P_0 + P_{VUN} \cdot M. \quad (4.44)$$

#### 4.7.2 Равномерное распределение точек на поверхности сферы

Алгоритм формирования равномерного распределения точек на поверхности сферы может быть построен различными методами, исходя из различных способов определения координат и элементарных областей на сфере. Рассмотрим первый метод формирования точек на сфере, см. рис. 4.22.

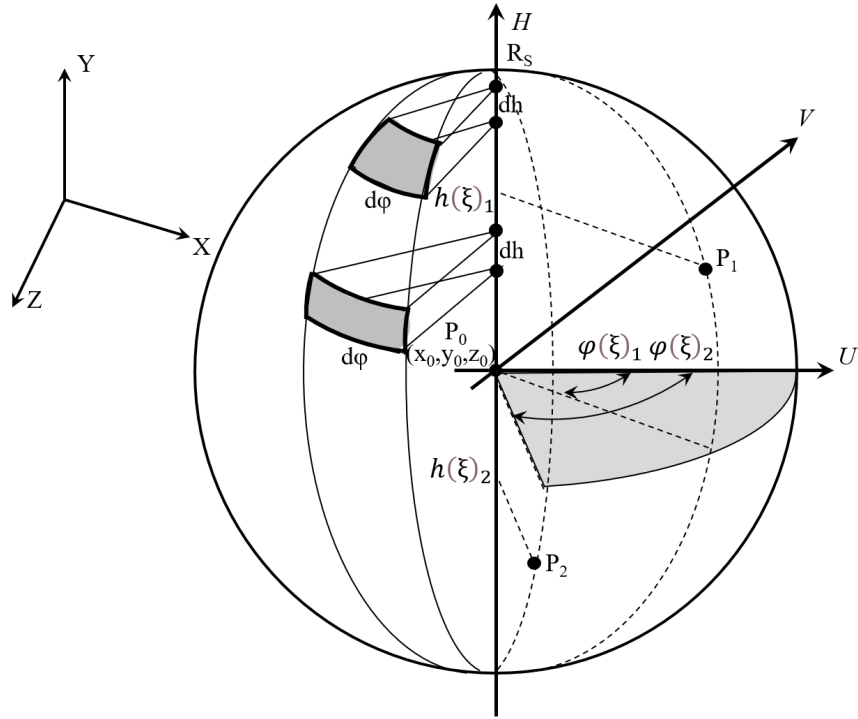


Рис. 4.22. Формирование равномерного распределения точек на поверхности сферы (первый метод)

Представим координаты точки на сфере в полярной системе координат ( $h$  – высота по оси сферы  $H$ ,  $\varphi$  – азимутальный угол и  $R_S$  – радиус сферы, в данном случае постоянная величина) и выразим через них площадь сферического сегмента:

$$ds = R_S \cdot d\varphi \cdot dh.$$

Видно, что площадь сегмента постоянна по поверхности сферы, поэтому в полярной системе координат равномерное распределение по координатам обеспечивает равномерное распределение по поверхности сферы:

$$\varphi = 2\pi\xi_\varphi, h = R_S(2\xi_h - 1), \quad (4.45)$$

где  $\xi_h$  и  $\xi_\varphi$  – случайные числа, равномерно распределенные в диапазоне  $[0,1]$ .

В декартовой системе координат  $UVH$ , где  $U$  – произвольный единичный вектор, ортогональный  $H$ , а  $V = H \times U$ , точка на поверхности сферы запишется следующим образом:

$$P_{UVH} = (\sqrt{1 - h^2} \cos \varphi, \sqrt{1 - h^2} \sin \varphi, h). \quad (4.46)$$

Переход в систему координат сцены производится простым смещением, поскольку поворот сферы вокруг своего центра инвариантен:

$$P_{XYZ} = P_0 + P_{UVH}. \quad (4.47)$$

В ряде случаев необходимо найти единичный вектор, равномерно распределенный по направлению в пространстве. Приравнивая  $R_S = 1$ , получаем

$$s = (\sqrt{1 - h^2} \cos \varphi, \sqrt{1 - h^2} \sin \varphi, h). \quad (4.48)$$

Второй метод решает задачу равномерного распределения точек на сфере в сферической системе координат, см. рис. 4.23. Координата точки на поверхности сферы имеет вид  $(R_S \sin \theta \cos \varphi, R_S \sin \theta \sin \varphi, R_S \cos \theta)$ .

Для формирования равномерного распределения точек на поверхности сферы необходимо (в сферической системе координат) построить функцию плотности вероятности, соответствующую площади сферического сегмента с фиксированным приращением координат  $d\theta$  и  $d\varphi$ .

$$ds = d\varphi d\theta \sin \theta. \quad (4.49)$$

Видно, что для координаты  $\varphi$  плотность вероятности постоянна. Поэтому данное двумерное распределение можно представить как два независимых распределения по  $\theta$  и  $\varphi$ , причем по координате  $\varphi$  случайная величина  $\xi$  будет иметь равномерное распределение. Поэтому рассмотрим плотность вероятности по координате  $\theta$  и введем нормировку плотности вероятности:

$$\int_0^\pi \sin \theta d\theta = 2,$$

где  $\sin \theta$  определяет вариацию площади сферической ячейки в зависимости от угла  $\theta$ .



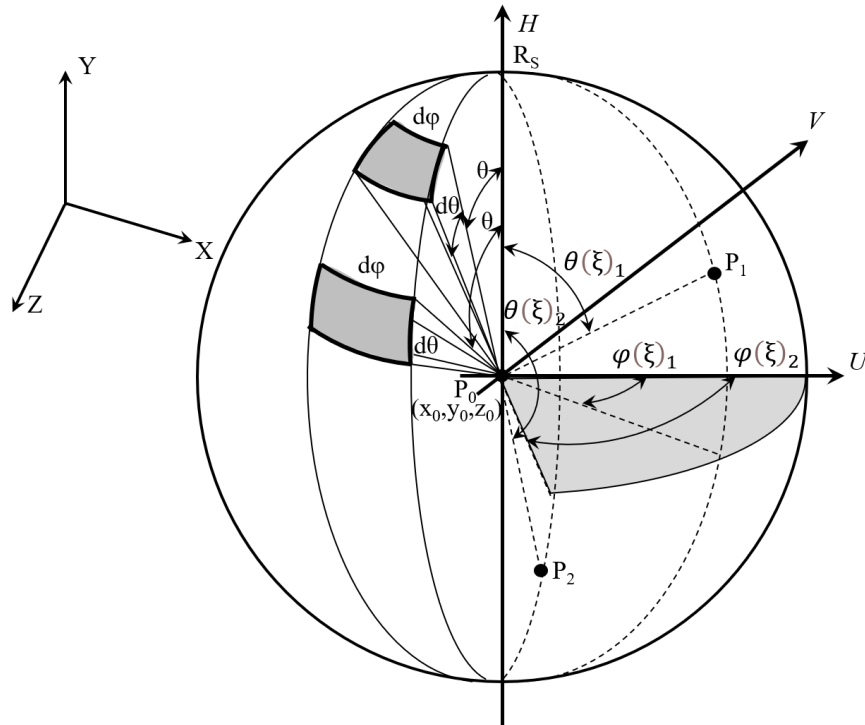


Рис. 4.23. Формирование равномерного распределения точек на поверхности сферы (второй метод)

Вероятность будет иметь следующий вид:

$$P(\theta) = \frac{1}{2} \int_0^\theta \sin t \, dt = \frac{1}{2} (1 - \cos \theta). \quad (4.50)$$

Обратная функция вероятности имеет следующий вид:

$$P(\theta')^{-1} = \theta = \cos^{-1}(2\theta' - 1), \quad (4.51)$$

где  $\theta'$  - приведенный обратный угол, варьирующийся от 0 до 1. Следовательно, для построения равномерного распределения точек на поверхности сферы необходимо формирование пары чисел:

$$\varphi = 2\pi\xi_\varphi, \theta = \cos^{-1}(2\xi_\theta - 1), \quad (4.52)$$

где  $\xi_\theta$  и  $\xi_\varphi$  - случайные числа, равномерно распределенные в диапазоне  $[0,1]$ .

Точка на круге в локальных координатах  $VUH$  (формирование системы координат было рассмотрено на примере построения равномерного распределения в круге) будет иметь вид:

$$P_{VUN} = (R_S \sin \theta \cos \varphi, R_S \sin \theta \sin \varphi, R_S \cos \theta). \quad (4.53)$$

Выражение (4.50) идентично выражению (4.45), поскольку  $h/R_S = \cos \theta = 2\xi_\theta - 1$ .

### 4.7.3 Распределение точек с заданной плотностью вероятности на плоскости

В ряде случаев распределение оптических характеристик на поверхности сцены может варьироваться, и это должно находить отражение в распределении плотности вероятности этих характеристик, используемых для расчета глобального освещения. Простейшее распределение — это двумерная функция, заданная на плоской прямоугольной поверхности, см. рис. 4.24. Как правило, функция плотности вероятности ассоциируется с распределением светового потока, освещенности или светимости на плоской поверхности объекта.

#### *Распределение случайной величины на плоском прямоугольнике*

Если функция плотности вероятности  $p(u, v)$  задана аналитическим образом, и вероятность может быть аналитически вычислена, и обратная функция вероятности также может быть вычислена аналитическим образом, то формирование случайных величин выполняется по аналогии с формулами (4.49) – (4.53).

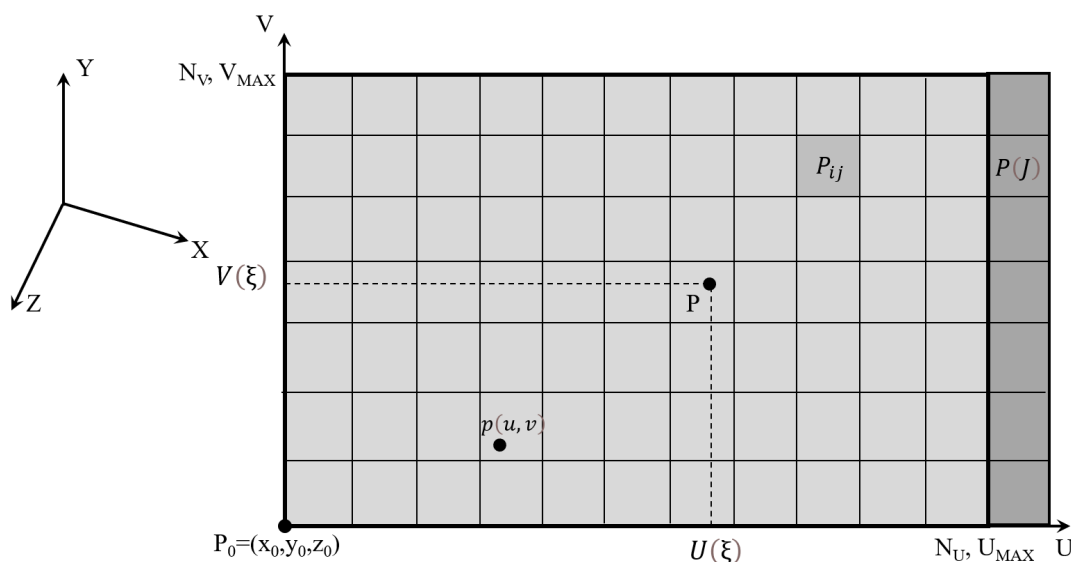


Рис. 4.24. Формирование распределения точек с заданной плотностью вероятности на прямоугольнике

Как правило, вероятность и ее обратная функция не могут быть вычислены аналитическим образом. Поэтому плотность вероятности  $p(u, v)$  задается в виде табулированной функции, в узлах которой задается плотность вероятности, а между узлами выполняется интерполяция (как правило, билинейная) см. рис. 4.24. Также предполагается, что интеграл  $P_{ij}$  в каждой ячейке таблицы может быть вычислен. Исходя из этого, мы можем вычислить интеграл  $P_{\Sigma}$  плотности вероятности в области определения функции:

$$P_{\Sigma} = \int_0^{U_{MAX}} \int_0^{V_{MAX}} p(u, v) du dv = \sum_{i=1}^{N_U} \sum_{j=1}^{N_V} P_{ij}.$$

В случае билинейной интерполяции плотность вероятности в ячейке  $ij$  можно записать в виде:

$$\begin{aligned} p(u, v) &= \left( p(u_i, v_j) \frac{v_{j+1}-v}{\Delta v_j} + p(u_i, v_{j+1}) \frac{v-v_j}{\Delta v_j} \right) \frac{u_{i+1}-u}{\Delta u_i} \\ &+ \left( p(u_{i+1}, v_j) \frac{v_{j+1}-v}{\Delta v_j} + p(u_{i+1}, v_{j+1}) \frac{v-v_j}{\Delta v_j} \right) \frac{u-u_i}{\Delta u_i} \quad , \\ &= d_{ij} + c_{ij}u + b_{ij}v + a_{ij}uv \end{aligned}$$

где:

- $i$  и  $j$  индексы ячеек табулированной функции плотности вероятности с разрешением  $N_u$  и  $N_v$ , где  $N_u$  и  $N_v$  – число ячеек этой табулированной функции по осям  $u$  и  $v$  соответственно,
- $p(u_i, v_j)$ ,  $p(u_i, v_{j+1})$ ,  $p(u_{i+1}, v_j)$  и  $p(u_{i+1}, v_{j+1})$  – значения плотности вероятности в узлах функции  $(u_i, v_j)$ ,  $(u_i, v_{j+1})$ ,  $(u_{i+1}, v_j)$  и  $(u_{i+1}, v_{j+1})$  соответственно,
- $\Delta u_i = u_{i+1} - u_i$ ,
- $\Delta v_j = v_{j+1} - v_j$ ,
- $d_{ij}$ ,  $c_{ij}$ ,  $b_{ij}$ ,  $a_{ij}$  – коэффициенты, вычисляемые следующим образом:

$$\begin{aligned}
d_{ij} &= \frac{1}{\Delta u_i \Delta v_j} [(p(u_i, v_j)v_{j+1} - p(u_i, v_{j+1})v_j)u_{i+1} - \\
&\quad (p(u_{i+1}, v_j)v_{j+1} - p(u_{i+1}, v_{j+1})v_j)u_i] \\
c_{ij} &= \frac{1}{\Delta u_i \Delta v_j} [(p(u_{i+1}, v_j) - p(u_i, v_j))v_{j+1} - \\
&\quad (p(u_{i+1}, v_{j+1}) - p(u_i, v_{j+1}))v_j] \\
b_{ij} &= \frac{1}{\Delta u_i \Delta v_j} [(p(u_i, v_{j+1}) - p(u_i, v_j))u_{i+1} - \\
&\quad (p(u_{i+1}, v_{j+1}) - p(u_{i+1}, v_j))u_i] + \\
a_{ij} &= \frac{1}{\Delta u_i \Delta v_j} [p(u_i, v_j) - p(u_i, v_{j+1}) - p(u_{i+1}, v_j) + p(u_{i+1}, v_{j+1})]
\end{aligned}$$

В результате вероятность в ячейке  $P_{ij}$  можно записать в виде:

$$\begin{aligned}
P_{ij} &= \int_{u_i}^{u_{i+1}} \int_{v_j}^{v_{j+1}} (d_{ij} + c_{ij}u + b_{ij}v + a_{ij}uv) dudv \\
&= (u_{i+1} - u_i)(v_{i+1} - v_i) [d_{ij} + \frac{c_{ij}}{2}(u_{i+1} + u_i) + \frac{b_{ij}}{2}(v_{i+1} + v_i) + \\
&\quad + \frac{a_{ij}}{4}(u_{i+1} + u_i)(v_{i+1} + v_i)]
\end{aligned}$$

Далее можно вычислить дискретную функцию вероятности, соответствующую распределению вероятности по ячейкам таблицы. Эту вероятность можно представить в виде двумерной функции:

$$\begin{aligned}
P(0) &= 0 \\
P(J) &= \sum_{i=1}^{N_U} P_{ij} / P_{\Sigma} \\
P(I)_J &= P(J - 1) + \sum_{i=1}^I P_{ij} / P_{\Sigma}
\end{aligned} \tag{4.54}$$

или свести к одномерной функции:

$$P((J - 1)N_U + I) = P(J - 1) + \sum_{i=1}^I P_{ij} / P_{\Sigma}. \tag{4.55}$$

Формируется обратная функция вероятности, для которой в случае двумерной функции вероятности сначала вычисляется индекс  $j$ , а затем индекс  $i$ :

$$\begin{aligned}
j &= P(\xi_j)^{-1} \\
i &= P(\xi_j - P(J - 1))_j^{-1},
\end{aligned} \tag{4.56}$$

где  $\xi_j$  – случайное число, равномерно распределенное в диапазоне  $[0,1[$ .

Для одномерного распределения индексы вычисляются следующим образом:

$$\begin{aligned} k &= P(\xi_j)^{-1} \\ j &= k/N_U + 1, \quad i = k \% N_U + 1 \end{aligned} \quad (4.57)$$

Поскольку функция вероятности и ее обратная функция имеют табличное представление, то наиболее оптимальное решение для вычисления обратной функции вероятности  $P(\xi_j)^{-1}$  – использование метода бинарного поиска, т.е. поиска интервала, для которого  $P_{j-1} \leq \xi_j < P_j$ . При этом  $j$  становится искомым значением обратной функции.

Это позволяет определить индекс ячейки в таблице. Поскольку цель есть построение непрерывного распределения точек в прямоугольной области, то следующим шагом является построение распределения случайной величины внутри выбранной ячейки с плотностью вероятности  $p(u, v)$ . Для построения необходимо воспользоваться методом выборки с отклонением, см. рис. 4.25.

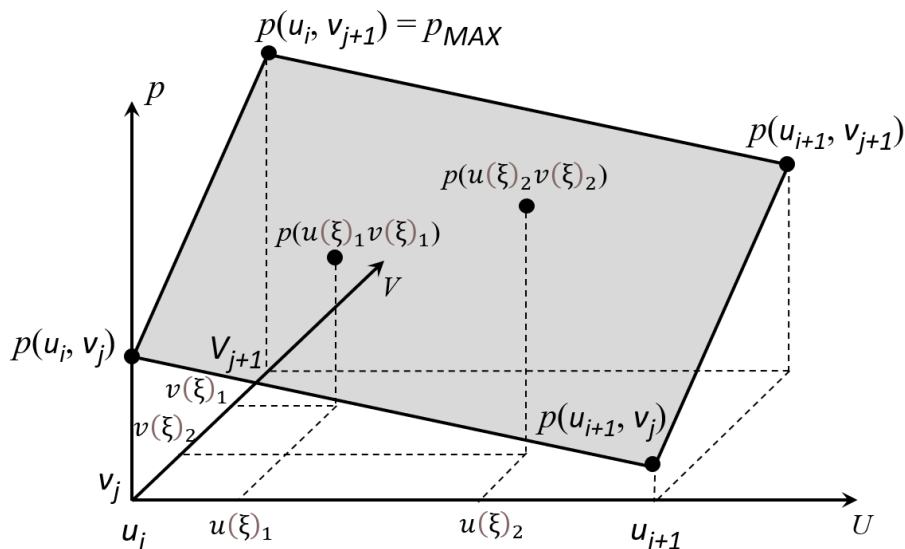


Рис. 4.25. Формирование распределения точек с заданной плотностью вероятности внутри выбранной ячейки прямоугольника

Поскольку функция плотности вероятности внутри ячейки вычисляема, найдем ее максимум  $p_{MAX}$  (как правило, это одна из вершин  $p(u_i, v_j)$ ,  $p(u_{i+1}, v_j)$ ,  $p(u_i, v_{j+1})$ ,  $p(u_{i+1}, v_{j+1})$ ). Сформируем равномерное распределение точек внутри

области  $\left[ [u_i, u_{i+1}[ , [v_j, v_{j+1}[ [ \right]$ , см. выражение (4.36). Точка имеет координаты  $\left( u_i + \frac{\xi_u U_{MAX}}{N_U}, v_i + \frac{\xi_v V_{MAX}}{N_V} \right)$ . Затем формируется второе случайное число  $\xi_M$  в диапазоне  $[0, p_{MAX}[$ . Далее проверяется условие:

$$\begin{aligned}
 p\left(u_i + \frac{\xi_u U_{MAX}}{N_U}, v_i + \frac{\xi_v V_{MAX}}{N_V}\right) &\geq \xi_M, \text{ то точка принимается} \\
 p\left(u_i + \frac{\xi_u U_{MAX}}{N_U}, v_i + \frac{\xi_v V_{MAX}}{N_V}\right) &< \xi_M, \text{ то точка отвергается}
 \end{aligned}
 \tag{4.58}$$

Переход в систему координат сцены осуществляется в соответствии с выражением (4.37).

### *Распределение случайной величины на треугольнике*

Как правило, основной геометрический примитив, формирующий поверхности объектов сцены, – это треугольник. Поэтому, если необходимо сформировать распределение случайной величины на поверхности сцены, например, распределение светимости, заданное текстурой, то это распределение необходимо ограничить вписанным треугольником. Связь двумерного распределения в  $(u, v)$  координатах и координат вершин треугольника определяются текстурными координатами, заданными для каждой вершины треугольника  $(u_1, v_1)$ ,  $(u_2, v_2)$  и  $(u_3, v_3)$ , см. рис. 4.26.

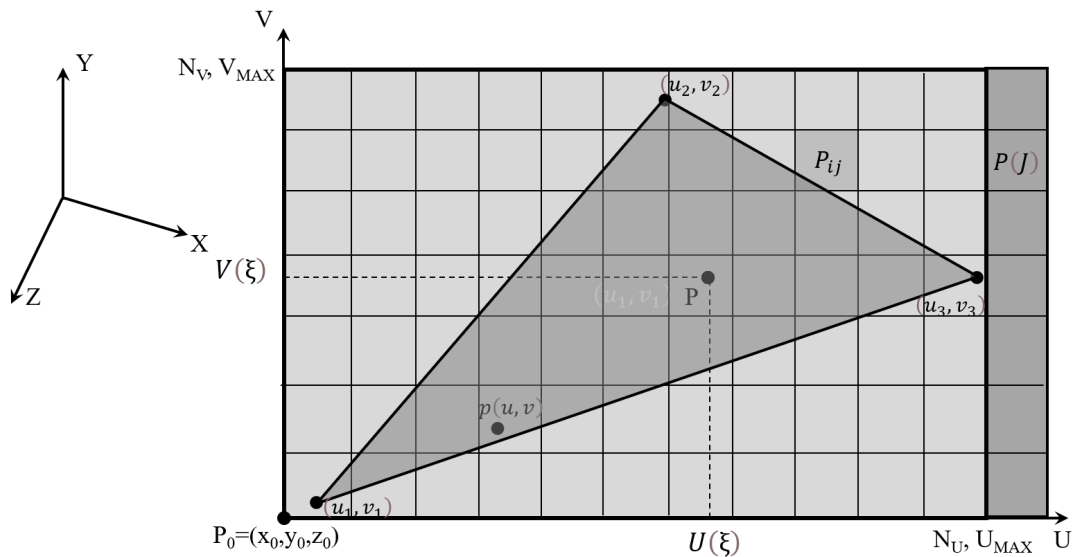


Рис. 4.26. Формирование распределения точек с заданной плотностью вероятности на треугольнике

Даже если функция плотности вероятности  $p(u, v)$  задана аналитическим образом, и ее вероятность может быть аналитически вычислена, то основная проблема возникает на границах треугольника, где сторона треугольника частично отсекает ячейку  $ij$ . Поэтому плотность вероятности  $p(u, v)$  удобнее задать в виде табулированной функции, в узлах которой задается плотность вероятности, а между узлами выполняется интерполяция (как правило, билинейная). Основная задача – это нахождение интегральной величины  $P_{ij}$  для каждой ячейки функции распределения плотности вероятности. Для ускорения расчетов всю область определения функции плотности вероятности можно разбить на 3 подобласти: область ячеек внутри треугольника (т.е. все вершины ячейки лежат внутри треугольника и  $P_{ij}$  вычисляется как для прямоугольной области), область ячеек вне треугольника (т.е. все вершины ячейки лежат вне треугольника и  $P_{ij} = 0$ ) и граничная область (т.е. часть вершин ячейки лежит вне треугольника). Для определения принадлежности узла ячейки области треугольника можно использовать барицентрический тест или, что более просто для плоского треугольника в декартовой системе координат, вычислить направление обхода вершин треугольника от точки  $(u_p, v_p)$ :

$$\begin{aligned} & [(u_2 - u_1)(v_3 - v_1) - (v_2 - v_1)(u_3 - u_1)] \\ & \cdot [(u_p - u_1)(v_3 - v_1) - (v_p - v_1)(u_3 - u_1)] < 0 \\ \vee & [(u_3 - u_2)(v_1 - v_2) - (v_3 - v_2)(u_1 - u_2)] \\ & \cdot [(u_p - u_2)(v_1 - v_2) - (v_p - v_2)(u_1 - u_2)] < 0 \\ \vee & [(u_1 - u_3)(v_2 - v_3) - (v_1 - v_3)(u_2 - u_3)] \\ & \cdot [(u_p - u_3)(v_2 - v_3) - (v_p - v_3)(u_2 - u_3)] < 0 - \text{точка снаружи.} \end{aligned}$$

Скалярные разности вершин треугольников и произведения их разностей могут быть вычислены заранее, что значительно ускорит процесс анализа узлов ячеек. Более сложная процедура – это нахождение вероятностей  $P_{ij}$  для ячеек табулированной функции в области границ треугольника. Наиболее простое решение – это метод выборки с отклонением, когда случайная величина равномерно распределяется внутри ячейки, а затем в вероятность  $P_{ij}$  добавляется плотность  $p(u, v)$  только тех точек  $(u, v)$ , чьи координаты принадлежат треугольнику. Интеграл плотности вероятности вычисляется аналогичным образом.

$$P_{\Sigma} = \sum_{i=1}^{N_U} \sum_{j=1}^{N_V} P_{ij}.$$

Далее решение выполняется аналогично распределению на прямоугольной области (см. выражения (4.54) – (4.58)) за исключением того, что при

формировании распределения точек с заданной плотностью вероятности внутри выбранной ячейки прямоугольника необходимо использовать дополнительную проверку попадания точек в область границ треугольника.

#### 4.7.4 Распределение лучей с заданной плотностью вероятности

Направление распространения лучей, как правило, не является равномерной функцией. В большинстве случаев для моделирования светового излучения, испускаемого источником света, используется модель Ламберта, она же используется в качестве основной модели рассеивания диффузных материалов. Однако существуют и другие модели излучения и рассеивания, которые будут рассмотрены ниже.

##### *Модель Ламберта*

Модель Ламберта – это осесимметричная модель косинусного излучателя. Этот источник света обеспечивает равномерное излучение при косинусном падении интенсивности относительно основного направления излучения:

$$I(\theta) = \cos \theta, \quad (4.59)$$

где  $I(\theta)$  – интенсивность излучения в направлении  $\theta$  относительно единичного вектора  $H$  (основного направления излучения), заданного в системе координат сцены  $XYZ$ , см. рис. 4.27. Это распределение осесимметрично, поэтому распределение по координате  $\varphi$  будет равномерным в диапазоне от 0 до  $2\pi$ . Поскольку интенсивность есть величина светового потока в телесном угле, а телесный угол в сферической системе координат это  $\sin \theta d\varphi d\theta$ , то распределение плотности светового потока в сферической системе координат (в зависимости только от угла  $\theta$ ) на единичной сфере пропорционально

$$\cos \theta \sin \theta. \quad (4.60)$$

Для формирования направления (точек на единичной сфере) можно воспользоваться общим подходом с формированием обратной функции вероятности (данное решение будет рассмотрено ниже для модели Фонга). Недостаток общего решения заключается в том, что, помимо формирования распределения в некоторой локальной системе координат, необходимо еще обеспечить матричное преобразование из локальной системы координат в систему координат сцены, см. выражения (4.43) и (4.44).



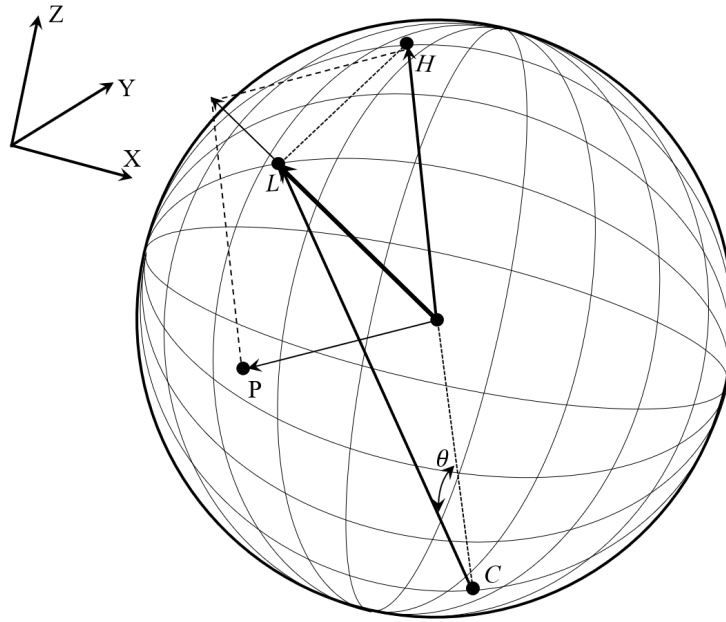


Рис. 4.27. Модель излучения Ламберта. Длина вектора  $CL$  пропорциональна интенсивности излучения  $I(\theta) = \cos \theta$

Для модели Ламберта существует более простое решение, которое заключается в следующей последовательности действий:

- Формирование равномерного распределения точек на поверхности сферы, например, используя выражения (4.45) и (4.48). При этом система координат, относительно которой формируется данное распределение, может быть абсолютно произвольной. В данном случае эта система координат должна быть  $XYZ$ , в которой задан вектор основного направления излучения  $H$ . В результате имеем единичный вектор  $P$  на поверхности сферы в системе координат  $XYZ$ .
- Вычисляется вектор  $L = H + P$ .
- Производится нормировка полученного вектора  $L$ .

$$L = \frac{H+P}{|H+P|}. \quad (4.61)$$

В результате вектор  $L$  получает косинусное распределение интенсивности в пространстве сцены  $XYZ$  относительно основного направления излучения  $H$ .

### **Модель Фонга**

Модель Фонга – это осесимметричная модель степенного косинусного излучателя. Этот источник света обеспечивает узкую диаграмму излучения,

интенсивность которого падает по степени косинуса угла относительно основного направления излучения:

$$I(\theta) = (\cos \theta)^\beta, \quad (4.62)$$

где  $\beta$  – степень «остроты» излучения. Схематично модель Фонга представлена на рис. 4.28. Это распределение осесимметрично, поэтому распределение по координате  $\varphi$  будет равномерным в диапазоне от 0 до  $2\pi$ . Для модели Фонга распределение плотности светового потока в сферической системе координат (в зависимости только от угла  $\theta$ ) на единичной сфере будет пропорционально

$$(\cos \theta)^\beta \sin \theta. \quad (4.63)$$

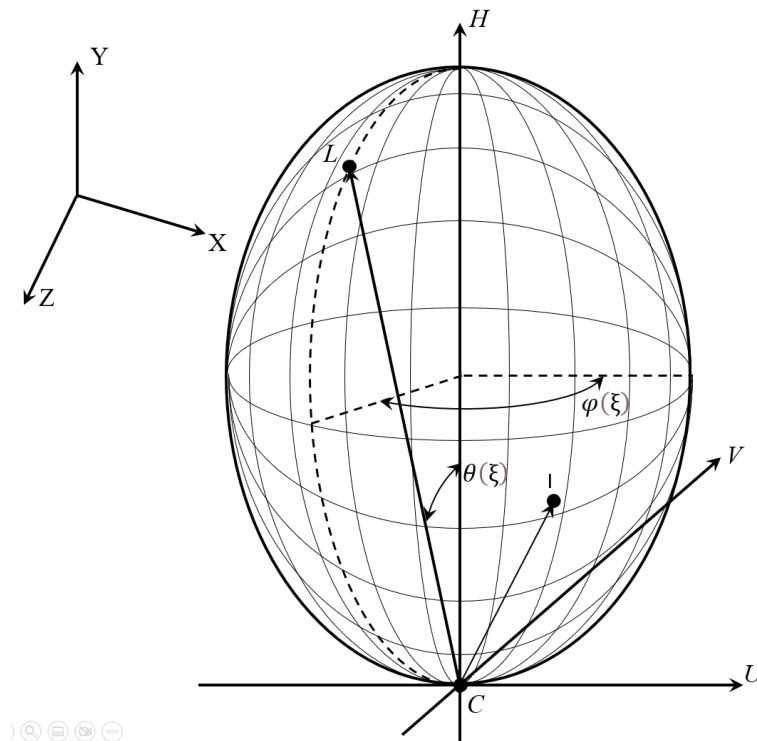


Рис. 4.28. Аналитическое представление одномерной модели излучения Фонга. Длина вектора  $CI$  пропорциональна интенсивности излучения  $I(\theta) = (\cos \theta)^\beta$

Для формирования направления (точек на единичной сфере) можно воспользоваться общим подходом с формированием обратной функции вероятности. В данном случае существует аналитическое решение для построения вероятности и ее обратной функции:

$$\int (\cos \theta)^\beta \sin \theta d\theta = -\frac{(\cos \theta)^{\beta+1}}{\beta+1} + C. \quad (4.64)$$

Введем нормировку плотности вероятности (функция определена в диапазоне  $[0, \pi/2[$ ):

$$\int_0^{\pi/2} (\cos \theta)^\beta \sin \theta d\theta = \frac{1}{\beta+1}.$$

Вероятность будет иметь следующий вид:

$$P(\theta) = (\beta + 1) \int_0^\theta (\cos t)^\beta \sin t dt = 1 - (\cos \theta)^{\beta+1}. \quad (4.65)$$

Обратная функция вероятности имеет следующий вид:

$$P(\theta')^{-1} = \theta = \cos^{-1}((\theta' - 1)^{1/\beta+1}), \quad (4.66)$$

где  $\theta'$  - приведенный обратный угол, варьирующийся от 0 до 1. Следовательно, для построения направления излучения по модели Фонга в сферической системе координат необходимо формирование пары чисел:

$$\varphi = 2\pi\xi_\varphi, \theta = \cos^{-1}((\xi_\theta - 1)^{1/\beta+1}), \quad (4.67)$$

где  $\xi_\theta$  и  $\xi_\varphi$  – случайные числа, равномерно распределенные в диапазоне  $[0,1[$ .

Единичный вектор в локальной системе координатах  $UVN$  (формирование системы координат и матрицы перехода в глобальную систему координат были рассмотрены на примере построения равномерного распределения в круге) будет иметь вид:

$$s_{UVN} = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta). \quad (4.68)$$

Для формирования распределения в системе координат сцены  $XYZ$  необходимо построить матрицу перехода из системы координат  $UVN$  в  $XYZ$ . Матрица перехода имеет вид:

$$M = \begin{pmatrix} V_x & V_y & V_z \\ U_x & U_y & U_z \\ N_x & N_y & N_z \end{pmatrix}.$$

В системе координат  $XYZ$  вектор будет иметь вид:

$$S_{XYZ} = S_{VUN} \cdot M. \quad (4.69)$$

### Модель неравномерного излучения

Общая модель неравномерного излучателя описывает распределение интенсивности излучения как функции двух углов  $I(\varphi, \theta)$ . Данная модель, как правило, не имеет аналитического представления и является табулированной функцией, имеющей возможность интерполяции ее значений между узлами. Схематично данная модель представлена на рис. 4.29. Распределение плотности светового потока в сферической системе координат  $(\varphi, \theta)$  на единичной сфере будет пропорционально

$$I(\varphi, \theta) \sin \theta. \quad (4.70)$$

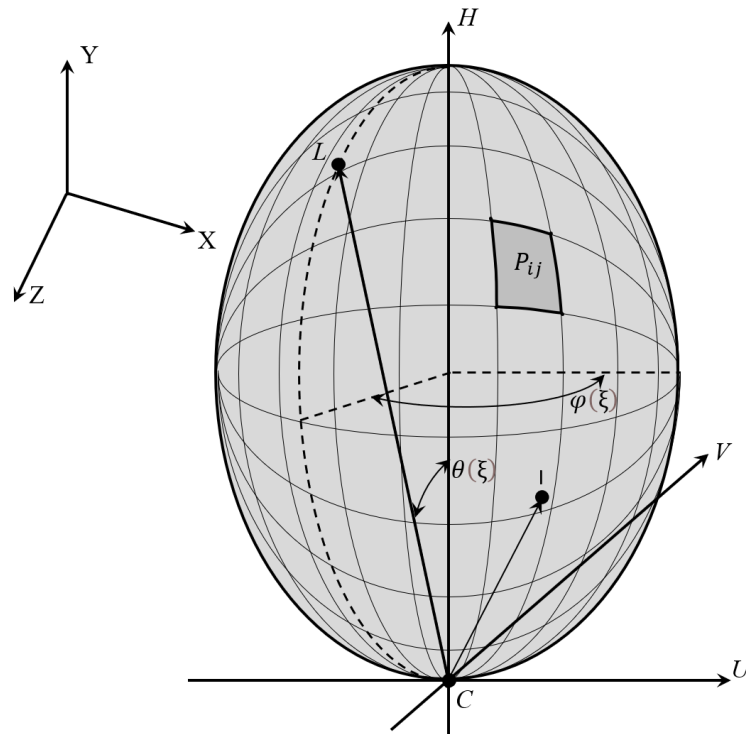


Рис. 4.29. Модель табулированной функции интенсивности излучения. Длина вектора  $CI$  пропорциональна интенсивности излучения  $I(\varphi, \theta)$

Поскольку интенсивность  $I(\varphi, \theta)$  задается в виде табулированной функции, узлы которой содержат интенсивность, то интерполяцию необходимо производить в единицах интенсивности. С другой стороны, интеграл  $P_{ij}$  в каждой ячейке таблицы может быть вычислен исходя из плотности распределения светового потока на единичной сфере. Поэтому мы можем численно вычислить интеграл  $P_{\Sigma}$  плотности вероятности в области определения функции:

$$P_{\Sigma} = \int_0^{2\pi} \int_0^{\pi} I(\varphi, \theta) \sin \theta d\varphi d\theta.$$

В случае билинейной интерполяции интенсивность  $I(\varphi, \theta)$  имеет вид:

$$I(\varphi, \theta) = \left( I(\varphi_i, \theta_j) \frac{\theta_{j+1} - \theta}{\Delta\theta_j} + I(\varphi_i, \theta_{j+1}) \frac{\theta - \theta_j}{\Delta\theta_j} \right) \frac{\varphi_{i+1} - \varphi}{\Delta\varphi_i} + \left( I(\varphi_{i+1}, \theta_j) \frac{\theta_{j+1} - \theta}{\Delta\theta_j} + I(\varphi_{i+1}, \theta_{j+1}) \frac{\theta - \theta_j}{\Delta\theta_j} \right) \frac{\varphi - \varphi_i}{\Delta\varphi_i},$$

что можно привести к виду:

$$\begin{aligned} I(\varphi, \theta) = & \frac{1}{\Delta\varphi_i \Delta\theta_j} [(I(\varphi_i, \theta_j) \theta_{j+1} - I(\varphi_i, \theta_{j+1}) \theta_j) \varphi_{i+1} - \\ & (I(\varphi_{i+1}, \theta_j) \theta_{j+1} - I(\varphi_{i+1}, \theta_{j+1}) \theta_j) \varphi_i] + \\ & \frac{\varphi}{\Delta\varphi_i \Delta\theta_j} [(I(\varphi_{i+1}, \theta_j) - I(\varphi_i, \theta_j)) \theta_{j+1} - \\ & (I(\varphi_{i+1}, \theta_{j+1}) - I(\varphi_i, \theta_{j+1})) \theta_j] + \\ & \frac{\theta}{\Delta\varphi_i \Delta\theta_j} [(I(\varphi_i, \theta_{j+1}) - I(\varphi_i, \theta_j)) \varphi_{i+1} - \\ & (I(\varphi_{i+1}, \theta_{j+1}) - I(\varphi_{i+1}, \theta_j)) \varphi_i] + \\ & \frac{\varphi\theta}{\Delta\varphi_i \Delta\theta_j} [I(\varphi_i, \theta_j) - I(\varphi_i, \theta_{j+1}) - \\ & I(\varphi_{i+1}, \theta_j) + I(\varphi_{i+1}, \theta_{j+1})] \\ & = d_{ij} + c_{ij}\varphi + b_{ij}\theta + a_{ij}\varphi\theta \end{aligned} \quad (4.71)$$

где:

- $i$  и  $j$  индексы ячеек табулированной функции интенсивности с разрешением  $N_{\varphi}$  и  $N_{\theta}$ , где  $N_{\varphi}$  и  $N_{\theta}$  – число ячеек на сферической сетке этой функции по осям  $\varphi$  и  $\theta$  соответственно,
- $I(\varphi_i, \theta_j)$ ,  $I(\varphi_i, \theta_{j+1})$ ,  $I(\varphi_{i+1}, \theta_j)$  и  $I(\varphi_{i+1}, \theta_{j+1})$  – значения интенсивности в узлах функции  $(\varphi_i, \theta_j)$ ,  $(\varphi_i, \theta_{j+1})$ ,  $(\varphi_{i+1}, \theta_j)$  и  $(\varphi_{i+1}, \theta_{j+1})$  соответственно,
- $\Delta\varphi_i = \varphi_{i+1} - \varphi_i$ ,
- $\Delta\theta_j = \theta_{j+1} - \theta_j$ ,
- $d_{ij}$ ,  $c_{ij}$ ,  $b_{ij}$ ,  $a_{ij}$  – коэффициенты, вычисляемые следующим образом:

$$\begin{aligned}
d_{ij} &= \frac{1}{\Delta\varphi_i\Delta\theta_j} [(I(\varphi_i, \theta_j)\theta_{j+1} - I(\varphi_i, \theta_{j+1})\theta_j)\varphi_{i+1} - \\
&\quad (I(\varphi_{i+1}, \theta_j)\theta_{j+1} - I(\varphi_{i+1}, \theta_{j+1})\theta_j)\varphi_i] \\
c_{ij} &= \frac{1}{\Delta\varphi_i\Delta\theta_j} [(I(\varphi_{i+1}, \theta_j) - I(\varphi_i, \theta_j))\theta_{j+1} - \\
&\quad (I(\varphi_{i+1}, \theta_{j+1}) - I(\varphi_i, \theta_{j+1}))\theta_j] \\
b_{ij} &= \frac{1}{\Delta\varphi_i\Delta\theta_j} [(I(\varphi_i, \theta_{j+1}) - I(\varphi_i, \theta_j))\varphi_{i+1} - \\
&\quad (I(\varphi_{i+1}, \theta_{j+1}) - I(\varphi_{i+1}, \theta_j))\varphi_i] + \\
a_{ij} &= \frac{1}{\Delta\varphi_i\Delta\theta_j} [I(\varphi_i, \theta_j) - I(\varphi_i, \theta_{j+1}) - I(\varphi_{i+1}, \theta_j) + I(\varphi_{i+1}, \theta_{j+1})]
\end{aligned} \tag{4.71}$$

В результате интеграл можно записать в виде:

$$\begin{aligned}
P_\Sigma &= \int_0^{2\pi} \int_0^\pi I(\varphi, \theta) \sin \theta \, d\varphi d\theta \\
&= \sum_{i=1}^{N_\varphi} \sum_{j=1}^{N_\theta} \int_{\varphi_i}^{\varphi_{i+1}} \int_{\theta_j}^{\theta_{j+1}} (d_{ij} + c_{ij}\varphi + b_{ij}\theta + a_{ij}\varphi\theta) \sin \theta \, d\varphi d\theta \\
&= \sum_{i=1}^{N_\varphi} \sum_{j=1}^{N_\theta} (\varphi_{i+1} - \varphi_i) [(\cos \theta_j - \cos \theta_{j+1}) \left( d_{ij} + \frac{c_{ij}}{2} (\varphi_{i+1} + \varphi_i) \right) + \\
&\quad (\theta_{j+1} \cos \theta_{j+1} - \theta_j \cos \theta_j + \sin \theta_{j+1} - \sin \theta_j) \left( b_{ij} + \frac{a_{ij}}{2} (\varphi_{i+1} + \varphi_i) \right)].
\end{aligned} \tag{4.72}$$

Далее, аналогично распределению на плоскости, можно вычислить дискретную функцию вероятности, соответствующую распределению вероятности по ячейкам таблицы. Необходимо отметить, что функция вероятности имеет размерность светового потока, а не интенсивности. Эту вероятность можно представить в виде двумерной функции, см. выражение (4.54) или одномерной функции, см. выражение (4.54).

Далее формируется обратная функция вероятности, для которой в случае двумерной функции вероятности сначала вычисляется индекс  $j$ , а затем индекс  $i$ , см. выражение (4.56). В случае одномерного распределения индексы вычисляются по формулам (4.57).

Поскольку функция вероятности и ее обратная функция имеют табличное представление, то наиболее оптимальное решение для вычисления обратной функции вероятности  $P(\xi_j)^{-1}$  – использование метода бинарного поиска, т.е.

поиска интервала, для которого  $P_{j-1} \leq \xi_j < P_j$ . При этом  $j$  становится искомым значением обратной функции.

В результате определяется индекс ячейки в таблице. Поскольку цель есть построение непрерывного распределения сферических координат, то следующим шагом является построение распределения случайной величины внутри выбранной ячейки с интенсивностью, соответствующей заданной  $I(\varphi, \theta)$ . Для построения необходимо воспользоваться методом выборки с отклонением, см. рис. 4.25.

Необходимо помнить, что вероятность была построена в единицах светового потока (аналогично распределению на плоскости), а интерполяция должна быть выполнена в единицах интенсивности (в сферической системе координат). Поэтому для реализации метода выборки с отклонением необходимо сформировать равномерное распределение координат точек на единичной сфере внутри заданной ячейки. Воспользуемся выражениями (4.45) и (4.48).

$$\varphi = \xi_\varphi, h = \xi_h, \quad (4.73)$$

где  $\xi_\varphi$  – случайное число, равномерно распределенное в диапазоне  $[\varphi_i, \varphi_{i+1}[$ ,  $\xi_h$  – случайное число, равномерно распределенное в диапазоне  $[\cos \theta_{j+1}, \cos \theta_j[$ . В результате распределение ограничивается одной ячейкой  $[[\varphi_i, \varphi_{i+1}[$ ,  $[\theta_j, \theta_{j+1}[$  и координаты точки на единичной сфере приобретают вид:

$$s = (\sqrt{1 - h^2} \cos \varphi, \sqrt{1 - h^2} \sin \varphi, h). \quad (4.74)$$

Далее идет расчет по формуле (4.58). Поскольку функция интенсивности внутри ячейки вычисляема, найдем ее максимум  $I_{MAX}$  (как правило, это одна из вершин  $I(\varphi_i, \theta_j)$ ,  $I(\varphi_{i+1}, \theta_j)$ ,  $I(\varphi_i, \theta_{j+1})$ ,  $I(\varphi_{i+1}, \theta_{j+1})$ ). Сформируем равномерное распределение точек на сфере внутри области  $[[\varphi_i, \varphi_{i+1}[$ ,  $[\theta_j, \theta_{j+1}[$ , см. выражение (4.74). Точка имеет координаты  $(\varphi_i + \xi_\varphi, \theta_i + \cos^{-1} \xi_h)$ . Затем формируется второе случайное число  $\xi_M$  в диапазоне  $[0, I_{MAX}[$ . Далее проверяется условие:

$$\begin{aligned} I(\varphi_i + \xi_\varphi, \theta_i + \cos^{-1} \xi_h) \geq \xi_M, \text{ то точка принимается} \\ I(\varphi_i + \xi_\varphi, \theta_i + \cos^{-1} \xi_h) < \xi_M, \text{ то точка отвергается} \end{aligned} \quad (4.75)$$

Угловое распределение световой величины может задаваться не только в терминах интенсивности, но и в терминах яркости. При этом все расчеты, связанные с физическими величинами излучения при расчете вероятности, сохраняются. Вероятность вычисляется в терминах светового потока, и в качестве исходной функции используется интенсивность. Т.е. яркость в узлах табулированной функции пересчитывается в интенсивность излучения (площадная составляющая не учитывается в этом выражении):

$$I(\varphi, \theta) = L(\varphi, \theta) \cos \theta.$$

В этом случае интерполяцию лучше выполнять, используя косинусную зависимость между узлами функции. Это аналогично модели билинейной интерполяции по яркости, когда в узлах хранится интенсивность, соответствующая определенной яркости:

$$I(\varphi, \theta) = \cos \theta \left[ \left( I(\varphi_i, \theta_j) \frac{\theta_{j+1} - \theta}{\Delta \theta_j \cos \theta_j} + I(\varphi_i, \theta_{j+1}) \frac{\theta - \theta_j}{\Delta \theta_j \cos \theta_{j+1}} \right) \frac{\varphi_{i+1} - \varphi}{\Delta \varphi_i} + \left( I(\varphi_{i+1}, \theta_j) \frac{\theta_{j+1} - \theta}{\Delta \theta_j \cos \theta_j} + I(\varphi_{i+1}, \theta_{j+1}) \frac{\theta - \theta_j}{\Delta \theta_j \cos \theta_{j+1}} \right) \frac{\varphi - \varphi_i}{\Delta \varphi_i} \right],$$

что можно привести к виду:

$$\begin{aligned} I(\varphi, \theta) = & \frac{\cos \theta}{\Delta \varphi_i \Delta \theta_j} \left[ \left( \frac{I(\varphi_i, \theta_j) \theta_{j+1}}{\cos \theta_j} - \frac{I(\varphi_i, \theta_{j+1}) \theta_j}{\cos \theta_{j+1}} \right) \varphi_{i+1} - \right. \\ & \left. \left( \frac{I(\varphi_{i+1}, \theta_j) \theta_{j+1}}{\cos \theta_j} - \frac{I(\varphi_{i+1}, \theta_{j+1}) \theta_j}{\cos \theta_{j+1}} \right) \varphi_i \right] + \\ & \frac{\varphi \cos \theta}{\Delta \varphi_i \Delta \theta_j} \left[ \left( \frac{I(\varphi_{i+1}, \theta_j) - I(\varphi_i, \theta_j)}{\cos \theta_j} \right) \theta_{j+1} - \right. \\ & \left. \left( \frac{I(\varphi_{i+1}, \theta_{j+1}) - I(\varphi_i, \theta_{j+1})}{\cos \theta_{j+1}} \right) \theta_j \right] + \\ & \frac{\theta \cos \theta}{\Delta \varphi_i \Delta \theta_j} \left[ \left( \frac{I(\varphi_i, \theta_{j+1})}{\cos \theta_{j+1}} - \frac{I(\varphi_i, \theta_j)}{\cos \theta_j} \right) \varphi_{i+1} - \right. \\ & \left. \left( \frac{I(\varphi_{i+1}, \theta_{j+1})}{\cos \theta_{j+1}} - \frac{I(\varphi_{i+1}, \theta_j)}{\cos \theta_j} \right) \varphi_i \right] + \\ & \frac{\varphi \theta \cos \theta}{\Delta \varphi_i \Delta \theta_j} \left[ \frac{I(\varphi_i, \theta_j)}{\cos \theta_j} - \frac{I(\varphi_i, \theta_{j+1})}{\cos \theta_{j+1}} - \right. \\ & \left. \frac{I(\varphi_{i+1}, \theta_j)}{\cos \theta_j} + \frac{I(\varphi_{i+1}, \theta_{j+1})}{\cos \theta_{j+1}} \right] \\ = & \cos \theta (d_{ij} + c_{ij} \varphi + b_{ij} \theta + a_{ij} \varphi \theta) \end{aligned} \quad (4.76)$$

где интерполяционные коэффициенты яркости будут иметь следующий вид:



$$\begin{aligned}
d_{ij} &= \frac{1}{\Delta\varphi_i\Delta\theta_j} \left[ \left( \frac{I(\varphi_i, \theta_j)\theta_{j+1}}{\cos\theta_j} - \frac{I(\varphi_i, \theta_{j+1})\theta_j}{\cos\theta_{j+1}} \right) \varphi_{i+1} - \right. \\
&\quad \left. \left( \frac{I(\varphi_{i+1}, \theta_j)\theta_{j+1}}{\cos\theta_j} - \frac{I(\varphi_{i+1}, \theta_{j+1})\theta_j}{\cos\theta_{j+1}} \right) \varphi_i \right] \\
c_{ij} &= \frac{1}{\Delta\varphi_i\Delta\theta_j} \left[ \left( \frac{I(\varphi_{i+1}, \theta_j) - I(\varphi_i, \theta_j)}{\cos\theta_j} \right) \theta_{j+1} - \right. \\
&\quad \left. \left( \frac{I(\varphi_{i+1}, \theta_{j+1}) - I(\varphi_i, \theta_{j+1})}{\cos\theta_{j+1}} \right) \theta_j \right] \\
b_{ij} &= \frac{1}{\Delta\varphi_i\Delta\theta_j} \left[ \left( \frac{I(\varphi_i, \theta_{j+1})}{\cos\theta_{j+1}} - \frac{I(\varphi_i, \theta_j)}{\cos\theta_j} \right) \varphi_{i+1} - \right. \\
&\quad \left. \left( \frac{I(\varphi_{i+1}, \theta_{j+1})}{\cos\theta_{j+1}} - \frac{I(\varphi_{i+1}, \theta_j)}{\cos\theta_j} \right) \varphi_i \right] + \\
a_{ij} &= \frac{1}{\Delta\varphi_i\Delta\theta_j} \left[ \frac{I(\varphi_i, \theta_j)}{\cos\theta_j} - \frac{I(\varphi_i, \theta_{j+1})}{\cos\theta_{j+1}} - \frac{I(\varphi_{i+1}, \theta_j)}{\cos\theta_j} + \frac{I(\varphi_{i+1}, \theta_{j+1})}{\cos\theta_{j+1}} \right]
\end{aligned} \tag{4.77}$$

В результате интеграл можно записать в виде

$$\begin{aligned}
P_\Sigma &= \int_0^{2\pi} \int_0^\pi I(\varphi, \theta) \sin\theta \, d\varphi d\theta \\
&= \sum_{i=1}^{N_\varphi} \sum_{j=1}^{N_\theta} \int_{\varphi_i}^{\varphi_{i+1}} \int_{\theta_j}^{\theta_{j+1}} (d_{ij} + c_{ij}\varphi + b_{ij}\theta + a_{ij}\varphi\theta) \sin\theta \cos\theta \, d\varphi d\theta \\
&= \frac{1}{4} \sum_{i=1}^{N_\varphi} \sum_{j=1}^{N_\theta} \int_{\varphi_i}^{\varphi_{i+1}} \int_{2\theta_j}^{2\theta_{j+1}} \left( d_{ij} + c_{ij}\varphi + \frac{1}{2}b_{ij}t + \frac{1}{2}a_{ij}\varphi t \right) \sin t \, d\varphi dt \\
&= \frac{1}{4} \sum_{i=1}^{N_\varphi} \sum_{j=1}^{N_\theta} (\varphi_{i+1} - \varphi_i) \left[ (\cos 2\theta_j - \cos 2\theta_{j+1}) \left( d_{ij} + \frac{c_{ij}}{2}(\varphi_{i+1} + \varphi_i) \right) + \right. \\
&\quad \left. (2\theta_{j+1}\cos 2\theta_{j+1} - 2\theta_j\cos 2\theta_j + \sin 2\theta_{j+1} - \sin 2\theta_j) \left( \frac{b_{ij}}{2} + \frac{a_{ij}}{4}(\varphi_{i+1} + \varphi_i) \right) \right].
\end{aligned} \tag{4.78}$$

Далее расчет выполняется аналогичным образом, см. выражения (4.73) – (4.75) с новой функцией интерполяции по яркости (4.76), (4.77).

## 4.8 Заключение

В этом разделе были представлены основные решения методов Монте-Карло, которые важны для реализации алгоритмов глобального освещения. Все методы Монте-Карло являются методами численного интегрирования. Они вычисляют оценку для определенного интеграла на основе случайных выборок. Первичная оценка – это оценка как функция одной выборки. Оценка является несмещенной, если ее ожидаемое значение равно фактическому значению подынтегрального выражения. Вторичные оценки усредняют несколько некоррелированных результатов первичной оценки и являются более надежными

оценками. Неопределенность результата количественно определяется дисперсией, которая обратно пропорциональна количеству выборок.

Поскольку дисперсия обратно пропорциональна корню из числа выборок, ее уменьшение при увеличении числа выборок сильно замедляется, и исследования методов Монте-Карло направлены на поиск способов, позволяющих уменьшить дисперсию при минимальном числе выборок. Как правило, эти методы нацелены на сглаживание подынтегральной функции и придание ей большей постоянности. Для методов Монте-Карло интегрирование функций, близких к постоянным, означает меньшую дисперсию для того же числа выборок или ту же дисперсию для меньшего числа выборок.

Первой важной оптимизацией является стратификация или расслоение выборок таким образом, чтобы они разбили область определения функции на участки с более или менее равномерным видом подынтегральной функции на каждом участке. Выборки, взятые в таких областях, будут лучше определять форму подынтегрального выражения. Этот метод интегрирования в чем-то напоминает детерминистические методы интегрирования с оптимальным разбиением области интегрирования. Примером такого подхода может служить метод разбиения области интегрирования на подобласти, а затем взятие единственной выборки в каждой из подобластей, оценка частных интегралов и, в итоге, суммирование оценок. Данный метод может быть расширен для интегрирования функций произвольного числа размерностей (вплоть до бесконечности) с использованием выборки  $N$ -roks.

Выборка по значимости предназначена для формирования большего числа выборок в областях, где подынтегральное выражение принимает большое значение. Это достигается за счет использования соответствующей функции плотности вероятности (PDF) и соответствующей модификации подынтегрального выражения. В идеальном случае PDF должна быть пропорциональна подынтегральному выражению. Значимость выбора соответствует равномерной выборке подынтегрального выражения после его преобразования. Для создания такого распределения необходимо вычисление обратной функции распределения вероятностей, т.е. эта функция должна быть практически вычисляемой.

Если подынтегральное выражение достаточно сложное и существует несколько потенциально «хороших» PDF, их можно объединить, используя сбалансированную эвристику, которая обладает доказуемо «хорошими» свойствами. Такое объединение оценок эквивалентно разбиению подынтегрального выражения и суммированию оценок для разных частей.

Главная часть – это функции, аппроксимирующие подынтегральное выражение, интеграл которых может быть вычислен аналитически. Выборка разности между главной частью и подынтегральным выражением с добавлением аналитически вычисленного интеграла дает новую первичную оценку интеграла с меньшей дисперсией.

Методы Монте-Карло также могут использоваться для численных решений интегральных уравнений. Фактическое решение интегрального уравнения оценивается только в определенной точке путем рассмотрения интегрального уравнения как рекурсивного определенного интеграла. Естественно, интеграл должен быть сходящимся. Бесконечную рекурсию можно остановить с помощью русской рулетки. Была продемонстрирована возможность сочетания русской рулетки и выборки по значимости с методом выделения главной части. Также показано, как этот подход позволяет уменьшить дисперсию. Важным методом уменьшения дисперсии, особенно при решении интегральных уравнений, является оценка следующего события.

Было показано, что все рассмотренные методы снижения дисперсии можно рассматривать как преобразования исходного интеграла. Уменьшение дисперсии происходит за счет того, что преобразованное подынтегральное выражение становится более равномерным и его интегрирование требует меньшего числа выборок. Кроме того, различные методы уменьшения дисперсии могут быть объединены, что позволяет дополнительно снизить дисперсию, например, стратифицированная выборка и выборка по значимости, могут быть объединены путем последовательного применения соответствующих преобразований.

Все рассмотренные методы снижения дисперсии являются несмещенными: усреднение большего числа первичных оценок гарантированно сходится к истинному решению. Однако в ряде случаев требуется большое количество выборок для снижения дисперсии до приемлемого уровня. На практике методы Монте-Карло часто имеют точность не более 0,1%. Поэтому основным является максимально возможное использование аналитических решений. Это естественным образом вытекает из предположения, что чем больше информации имеется о проблеме, тем значительно можно уменьшить вероятностный аспект метода Монте-Карло и связанную с ним дисперсию. Как было сказано в работе [36]: “ The only good Monte Carlos are dead Monte Carlos ”.

Остается открытым вопрос, не являются ли детерминированные методы предпочтительнее вероятностных. Неопределенность, присущая результатам методов Монте-Карло, может показаться неудобной. С другой стороны,

детерминированные методы часто имеют строгие границы погрешности. Уменьшить эту погрешность можно только изменив условия интегрирования и выполнив расчет заново. Однако для конкретных приложений гарантия того, что вычисление будет выполнено с требуемой точностью, отсутствует, поскольку физические величины часто имеют сложные разрывы и колебания в пространстве и во времени. Кроме того, методы детерминированного интегрирования обычно страдают от проклятия размерности, что является гораздо меньшей проблемой для методов Монте-Карло. Преимущество методов Монте-Карло заключается в том, что в ряде случаев для повышения точности достаточно просто увеличить число выборок, добавляя их к результатам предыдущего интегрирования. Методы Монте-Карло являются альтернативой детерминистическим методам для задач со сложными и многомерными подынтегральными выражениями.

В ряде других исследований рассматривалась возможность использования смещенных оценок, которые на практике имеют малую погрешность, соизмеримую с погрешностями несмещенных методов Монте-Карло. Другая возможность ускорения сходимости вычисления интегралов – это использование методов квази Монте-Карло. Методы квази Монте-Карло применяют те же методы, что и методы Монте-Карло, только заменяют псевдослучайные числа на квазидетерминированные последовательности чисел, которые обладают рядом желаемых свойств для снижения дисперсии. Хотя для ряда задач их скорость сходимости намного выше, имеются технические проблемы интегрирования в пространстве с заранее неограниченной размерностью.

Также были рассмотрены примеры методов формирования случайных величин (координат точек и направлений) с заданной плотностью вероятности. Эти методы будут полезны для расчета глобального освещения методами трассировки путей и трассировки световых лучей методами Монте-Карло. Были рассмотрены только базовые подходы, которые в дальнейшем могут быть расширены для решения более сложных задач, например, излучение света с поверхности протяженного источника света (с учетом цвета источника света задача становится пятимерной), или рассеивание световых лучей на поверхности, обладающей анизотропной ДФР (с учетом света луча задача становится пятимерной), или формирование распределения с заданной плотностью вероятности на поверхности с прямоугольным ограничением.

Далее будет рассмотрено применение рассмотренных стандартных методов Монте-Карло к проблеме расчета глобального освещения и физически корректному рендерингу.

## 4.9 Вопросы для самоконтроля усвоенных знаний

Следующие вопросы позволят оценить степень усвоения знаний по разделу методы Монте-Карло.

1. В чем суть метода Монте-Карло при расчете интеграла сложной функции. В чем преимущества и недостатки метода Монте-Карло.
2. Используя метод Монте-Карло, напишите программу расчета функции  $y = x^2/2$  на интервале  $[0, 2]$ . Используйте одну, 100, 10000 и 1000000 выборок и сравните точность с аналитическим решением. Все варианты выборок должны быть выполнены внутри одного расчетного цикла, т.е. каждое последующее число выборок начинается с предыдущего числа. Проанализируйте результат.
3. Используя метод Монте-Карло со стратификацией, напишите программу расчета функции  $y = x^2/2$  на интервале  $[0, 2]$ . Используйте одну, 100, 10000 и 1000000 выборок и сравните точность с аналитическим решением. Все варианты выборок должны быть выполнены внутри одного расчетного цикла, т.е. каждое последующее число выборок начинается с предыдущего числа. Для стратификации разбейте интервал на 5 равных частей. Проанализируйте результат.
4. Используя метод Монте-Карло с выборкой по значимости, напишите программу расчета функции  $y = x^2/2$  на интервале  $[0, 2]$ . Используйте одну, 100, 10000 и 1000000 выборок и сравните точность с аналитическим решением. Все варианты выборок должны быть выполнены внутри одного расчетного цикла, т.е. каждое последующее число выборок начинается с предыдущего числа. В качестве функций плотности вероятности используйте  $PDF = x$  и  $PDF = x^2/2$ . Проанализируйте результат.
5. Используя метод Монте-Карло с многократной выборкой по значимости, напишите программу расчета функции  $y = x^2/2$  на интервале  $[0, 2]$ . Используйте одну, 100, 10000 и 1000000 выборок и сравните точность с аналитическим решением. Все варианты выборок должны быть выполнены внутри одного расчетного цикла, т.е. каждое последующее число выборок начинается с предыдущего числа. В качестве функций плотности вероятности используйте  $PDF = x$  и  $PDF = x^2/2$  с одинаковыми весами. Проанализируйте результат.
6. Используя метод Монте-Карло с выделением главной части, напишите программу расчета функции  $y = x^2/2$  на интервале  $[0, 2]$ . Используйте

одну, 100, 10000 и 1000000 выборок и сравните точность с аналитическим решением. Все варианты выборок должны быть выполнены внутри одного расчетного цикла, т.е. каждое последующее число выборок начинается с предыдущего числа. В качестве функций, определяющих главную часть плотности вероятности, используйте  $PDF = x^2/2 - 0.1$  и  $PDF = x^2/2$ . Проанализируйте результат.

7. Используя метод Монте-Карло с русской рулеткой, напишите программу расчета функции  $y = x^2/2$  на интервале  $[0, 2]$ . Используйте одну, 100, 10000 и 1000000 выборок и сравните точность с аналитическим решением. Все варианты выборок должны быть выполнены внутри одного расчетного цикла, т.е. каждое последующее число выборок начинается с предыдущего числа. Русская рулетка предполагает обнуление подынтегральной функции  $R = 1.8$ . Проанализируйте результат.
8. Используя метод Монте-Карло с русской рулеткой и выборкой по значимости, напишите программу расчета функции  $y = x^2/2$  на интервале  $[0, 2]$ . Используйте одну, 100, 10000 и 1000000 выборок и сравните точность с аналитическим решением. Все варианты выборок должны быть выполнены внутри одного расчетного цикла, т.е. каждое последующее число выборок начинается с предыдущего числа. Русская рулетка предполагает обнуление подынтегральной функции  $R = 1.8$ , в качестве функций плотности вероятности используйте  $PDF = x$  и  $PDF = x^2/2$ . Проанализируйте результат.
9. Объясните, как метод русской рулетки позволяет решить уравнение Фредгольма второго рода (уравнение рендеринга).
10. В чем заключается метод оценки следующего события, как этот метод может быть применен к решению уравнения Фредгольма второго рода (уравнение рендеринга).
11. Написать программу, формирующую случайную точку, равномерно распределенную внутри треугольника, заданного вершинами  $v_1(x_{v_1}, y_{v_1}, z_{v_1}), v_2(x_{v_2}, y_{v_2}, z_{v_2}), v_3(x_{v_3}, y_{v_3}, z_{v_3})$ . Доказать, что распределение, построенное с помощью данной программы, формирует равномерное распределение внутри треугольника, и точки не выходят за рамки треугольника.
12. Написать программу, формирующую случайную точку, равномерно распределенную внутри круга с центром  $c(x_c, y_c, z_c)$ , радиусом  $r$  и нормалью  $N(x_N, y_N, z_N)$ . Доказать, что распределение, построенное с помощью данной программы, формирует равномерное распределение внутри круга, и точки не выходят за рамки круга. При написании

программы используйте два варианта распределения: непосредственное формирование точек внутри круга и выборку с отклонением внутри квадрата.

13. Написать программу, формирующую случайную точку, равномерно распределенную внутри сферы с центром  $s(x_c, y_c, z_c)$  и радиусом  $r$ . Доказать, что распределение, построенное с помощью данной программы, формирует равномерное распределение внутри сферы, и все точки находятся на сфере. Используя данную программу сформируйте равномерное распределение в пространстве.
14. Написать программу, формирующую случайную точку, неравномерно распределенную внутри прямоугольника, заданного вершиной  $v_1(x_{v_1}, y_{v_1}, z_{v_1})$  и векторами двух его сторон  $v_2(x_{v_2}, y_{v_2}, z_{v_2}), v_3(x_{v_3}, y_{v_3}, z_{v_3}), v_2 \cdot v_3 = 0$ . Неравномерность распределения определяется за счет переменной равномерной плотности четырех прямоугольных областей прямоугольника одинаковой площади, вписанных в основной прямоугольник. Доказать правильность формирования распределения точек внутри прямоугольника.
15. Написать программу, формирующую случайное направление с косинусной плотностью вероятности (закон Ламберта) относительно заданного направления  $N(x_N, y_N, z_N)$ . Доказать косинусный характер распределения случайных направлений, сформированных данной программой.
16. Написать программу, формирующую случайное направление с плотностью вероятности  $PDF = (\cos \alpha)^\beta$  (закон Фонга) относительно заданного направления  $N(x_N, y_N, z_N)$ . Доказать, что полученная плотность вероятности имеет вид  $PDF = (\cos \alpha)^\beta$ .

## 5. Применение методов Монте-Карло для решения уравнения рендеринга – трассировка путей

### 5.1 Базовый алгоритм

Первый подход для решения проблемы глобального освещения основывается на решении уравнения рендеринга. Для решения этого уравнения был разработан ряд алгоритмов стохастической трассировки лучей. Трассировка лучей была впервые представлена Уиттедом в работе [37] в виде эмпирического и детерминированного алгоритма для моделирования различных эффектов освещения. Кук [38] расширил алгоритм методами стохастической выборки [38, 39]. Каджия [2] был первым, кто предложил стохастический алгоритм для

решения уравнения рендеринга. Далее шло развитие и оптимизация этих алгоритмов для физически корректного решения уравнения рендеринга, например [40-43]. Ряд алгоритмов предлагал несмещенные решения, ряд – смещенные. Далее будут рассмотрены основные идеи несмещенных алгоритмов, позволяющих решить проблему глобального освещения. Базовый алгоритм начинается с выражения потока излучения в терминах яркости (2.8) и интегрального уравнения, описывающего функцию яркости (2.13):

$$\Phi = \int_A \int_{\Omega_x} L(x, \Theta_x) W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x, \quad (2.8)$$

$$L(x, \Theta_x) = L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} L(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y. \quad (2.13)$$

Когда мы говорим о глобальном освещении, мы подразумеваем расчет светового потока  $\Phi$  точки сцены. Выбор случайных переменных  $x_0$  и  $\Theta_{x_0}$  в соответствии с некоторой PDF  $p_0(x, \Theta_x)$  по всем точкам поверхности и направлениям по полусфере позволяет получить следующую первичную оценку интеграла (2.8):

$$\langle \Phi \rangle_{path} = \frac{W_e(x_0, \Theta_{x_0}) |\Theta_{x_0} \cdot N_{x_0}|}{p_0(x_0, \Theta_{x_0})} \langle L(x_0, \Theta_{x_0}) \rangle_{path}.$$

Для алгоритмов, основанных на рендеринге изображений, потенциал самосвечения объектов сцены (для всех точек и направлений) зависит от параметров камеры (направления на входной зрачок и точку изображения) и поэтому известен только косвенно. Это определяется моделью камеры (точечная камера (перспективная/ортогональная), линзовая камера (идеальная/реальная)) или специальным наблюдателем. Обычно точка  $x_{-1}$  на изображении (для точечной камеры) или на входном зрачке (для линзовой камеры) и направление  $\Theta_{x_0}$  указывают на точку изображения (для точечной камеры один из параметров  $x_{-1}$  или  $\Theta_{x_0}$  будет избыточен) и выбираются в соответствии с PDF  $p_{-1}(x, \Theta_x)$ . Затем точка  $x_{-1}$  переносится методом трассировки лучей в точку  $x_0$  на ближайшей поверхности сцены  $x_0 = r(x_{-1}, \Theta_{x_0}^{-1})$  (для линзовой системы далее идет преобразование точки  $x_{-1}$  в точку на изображении). PDF  $p_0(x, \Theta_x)$ , которая неявным образом использовалась для выбора  $x_0$  и  $\Theta_{x_0}$ , может быть легко оценена:

$$p_0(x_0, \Theta_{x_0}) = \frac{p_{-1}(x_{-1}, \Theta_{x_0}) |\Theta_{x_0} \cdot N_{x_0}|}{|x_{-1} - x_0|^2}. \quad (5.1)$$

Функция яркости, входящая в выражение оценки светового потока, неизвестна, но ее конкретное значение  $L(x_0, \Theta_{x_0})$  можно оценить, если



рассмотреть уравнение рендеринга (2.13) как интеграл с бесконечной рекурсией, описанный в разделе 4.6. Применение русской рулетки, описанной в этом разделе, гарантирует завершение рекурсии за конечное число шагов без смещения результата. Если вероятность продолжения рекурсии на каждом из шагов (сегментов трасс луча) равна  $R_i$  и каждое направление сегмента трассы луча  $\Theta_{x_i}$  выбирается в соответствии с плотностью вероятности  $p_i(\Theta_x)$  по полусфере, то общая первичная оценка значения светового потока (4.33) будет:

$$\langle \Phi \rangle_{path} = \frac{W_e(x_0, \Theta_{x_0}) |\Theta_{x_0} \cdot N_{x_0}|}{p_0(x_0, \Theta_{x_0})} \cdot \prod_{j=1}^l \left[ \frac{f_r(x_{j-1}, \Theta_{x_j}, \Theta_{x_{j-1}}) |\Theta_{x_j} \cdot N_{x_{j-1}}|}{R_j p_j(\Theta_{x_{j-1}})} \right] L_e(x_i, \Theta_{x_i}), \quad (5.2)$$

где  $l$  – число случайных сегментов трассы луча  $x_0, \dots, x_l \cdot N_{x_{l-1}}$ .

Каждая точка  $x_i$  вычисляется переносом  $x_i = r(x_{i-1}, \Theta_{x_i}^{-1})$ . Перенос выполняется методом трассировки луча из точки  $x_{i-1}$  в направлении  $\Theta_{x_i}^{-1}$  до ближайшей точки на поверхности сцены  $x_i$ , см. рис. 5.1 (необходимо отметить, что в случае среды с переменным показателем преломления перенос может быть непрямолинейным в соответствии с уравнением эйконала [7]). Приведенное решение является общей формой алгоритма трассировки путей без явной выборки источников света, см. рис. 5.2. Данный алгоритм был впервые представлен Каджией [2] как решение уравнения рендеринга методом Монте-Карло на физической основе. Лучи испускаются из виртуальной камеры, трассируются по сцене и добавляют яркость точкам виртуальной камеры каждый раз, когда они попадают в источник света. Этот подход можно назвать методом сборки. Рис. 5.3 и рис. 5.4. демонстрируют работу данного алгоритма.

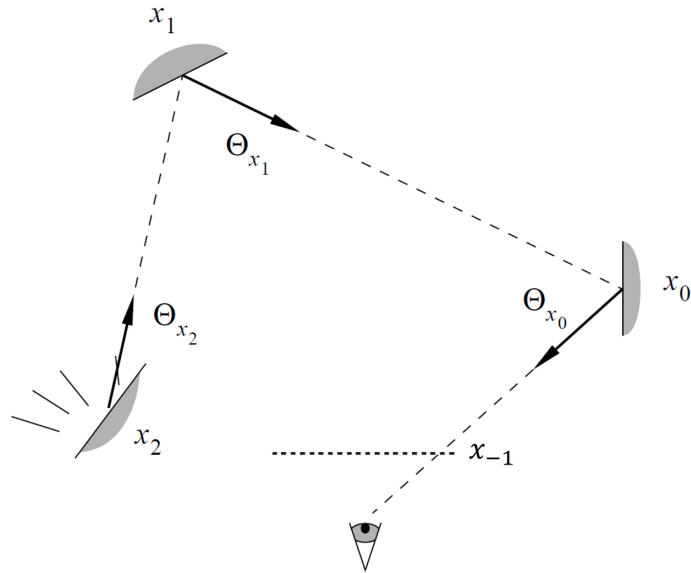


Рис. 5.1. Символы, используемые в методах трассировки путей

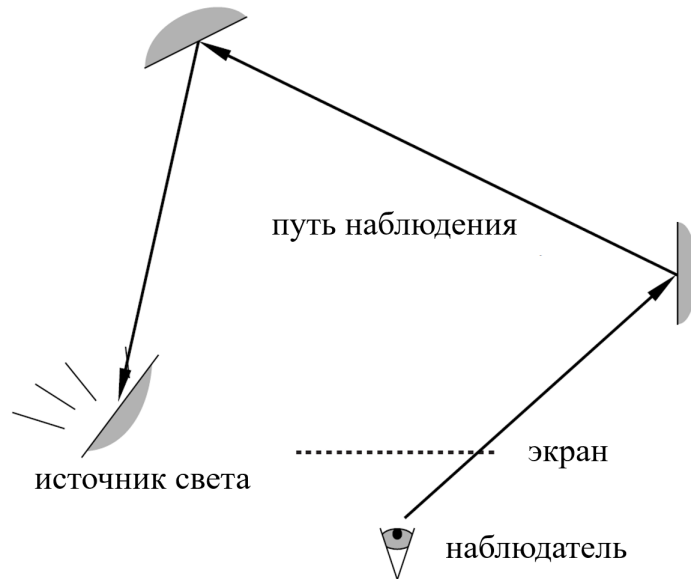


Рис. 5.2. Схематическое представление алгоритма трассировки путей. Задача заключается в оценке светового потока, проходящего через пиксель. Трассировка первого сегмента луча от точки виртуального глаза через пиксель соответствует выборке выражения для светового потока. Последующее случайное блуждание луча по сцене соответствует рекурсивной оценке значений яркости. Каждый раз, когда луч пересекает источник света, к оценке значения яркости добавляется определенный вклад

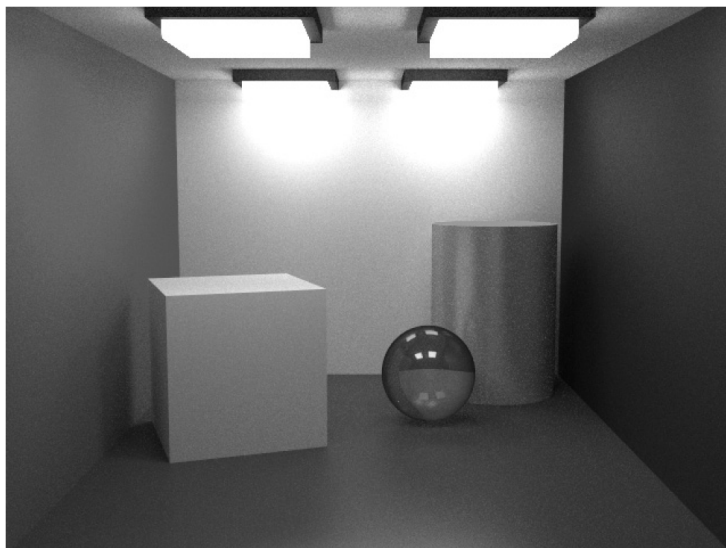


Рис. 5.3. Пример изображения сцены, синтезированной с помощью алгоритма трассировки путей. Алгоритм хорошо работает на сценах, содержащих объекты с зеркальными или квазизеркальными оптическими свойствами, а также протяженные источники света большого размера. На изображении видны типичные эффекты глобального освещения: вторичное освещение, глянцевое отражение, мягкие тени

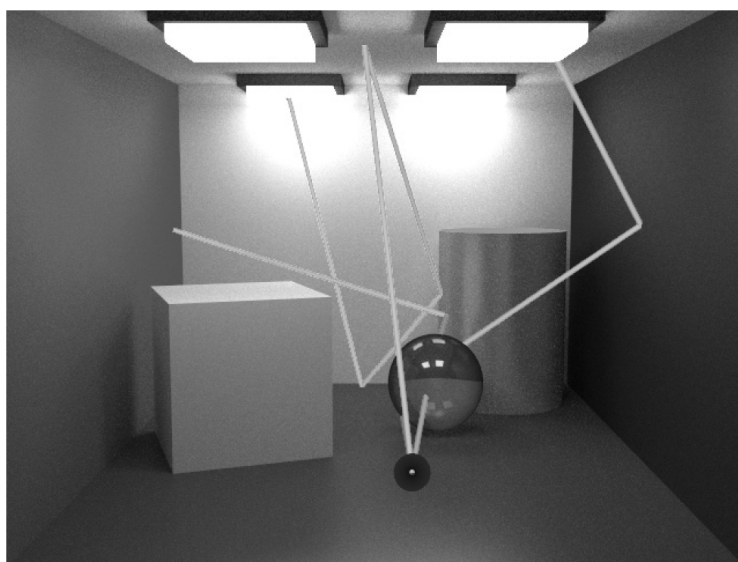


Рис. 5.4. Пример визуализации ряда реальных траекторий трасс лучей. Трассы лучей начинаются с одной точки, находящейся в нижней части изображения внутри черного круга. Каждый раз, когда луч попадает на источник света, точка изображения увеличивается на его яркость. Если источники света имеют малый размер, то на них попадает мало лучей, что приводит к значительному шуму изображения, имеющего вид «снега»

## 5.2 Стратифицированная выборка

Многомерная стратифицированная выборка, представленная в разделе 4.2, также может быть применена к выражению светового потока (2.8) и уравнению рендеринга (2.13). При решении данных уравнений большинство выборок, как правило, многомерны. Грамотно организованная стратифицированная выборка представляет собой случай квазислучайного распределения, что может привести к снижению дисперсии в оценке яркости или светового потока. Однако с практической точки зрения проблема заключается в том, что квазипоследовательности могут быть сконструированы для любого фиксированного числа измерений, но не для произвольного и априори неизвестного числа измерений, как требуется для решения уравнения (5.2). Снижение дисперсии происходит за счет когерентности в подынтегральном выражении, однако подынтегральное выражение, сильно прерывистое по всей полусфере интегрирования за счет прерывистой функции приходящего излучения. Излучение, приходящее в точку наблюдения, определяется геометрией сцены и ее оптическими характеристиками, которые с точки зрения наблюдателя непредсказуемы. Поэтому даже после одного шага рекурсии, т.е. после одного отражения луча, когерентность практически разрушается.

Однако стратифицированная выборка может быть полезна, по крайней мере, для выбора исходных параметров луча (координаты и направления). Этот подход может быть применен, например, к распределению точек на зрачке линзовой камеры. Первые сегменты лучей будут иметь лучшее пространственное распределение для выборки поверхностей, видимых через пиксель. Даже если после такого квазиравномерного распределения лучи попадают на одну поверхность, доля лучей, отраженных после применения алгоритма русской рулетки, может оказаться лучшим приближением к желаемому. Каждый из отраженных лучей вносит свой вклад в оценку. Таким образом, уменьшение «случайности» в этом процессе распространения лучей снижает общую «случайность» результата. Это рассуждение может быть расширено в процессе дальнейшей трассировки. Однако в процессе трассировки лучей в ход могут вступать другие механизмы, например, выборка по значимости в соответствии с ДФР поверхности. В целом, стратифицированная выборка потенциально может улучшить результаты при незначительных затратах.

С технической точки зрения стратификация может сводиться к упрощению методов формирования случайных выборок. Это возникает, когда многомерное распределение можно разбить на группы независимых распределений меньшей размерности. Примером может служить испускание луча с точек изображения

линзовой камеры. Четырехмерное распределение можно разбить на две независимые группы: выбор точки на поверхности изображения и выбор точки на поверхности входного (выходного, в зависимости от техники расчета) зрачка. Другим примером может служить формирование светового луча, испускаемого с поверхности протяженного источника света. Данное распределение четырехмерное, состоящее из выбора точки на поверхности источника света (две размерности) и выбора направления излучения. Как правило, угловая диаграмма излучения постоянна для всех точек поверхности. Поэтому данное четырехмерное распределение разбивается на два двумерных: выбор точки излучения на поверхности источника света (выполняется по распределению светимости по поверхности источника света) и выбор направления излучения (выполняется по распределению интенсивности излучения гониометрической диаграммы).

### 5.3 Выборка по значимости

Выборка по значимости, рассмотренная в разделе 4.3, должна формировать выборку значений в областях большой значимости подынтегрального выражения, например, преобладание испускания световых лучей с областей с высокой светимостью на источниках света или выбор направлений рассеивания лучей на поверхности таким образом, чтобы большая их часть была ориентирована вокруг направления зеркального отражения. В случае детерминированной трассировки лучей алгоритмы, не учитывающие эффекты глобального освещения, могут формировать один луч в направлении зеркального отражения. Такое решение позволяет устранить неопределенность результата и, следовательно, шум изображения. Однако данное решение не является физически корректным. Для достижения физической корректности необходимо решать уравнение рендеринга, основываясь на методах Монте-Карло. Возможны различные варианты формирования случайных выборок. Для решения данной задачи наиболее общим методом будет выборка по значимости, которая рассматривается ниже.

Интегральное уравнение (2.8) имеет известную часть  $W_e(x, \Theta_x) |\Theta_x \cdot N_x|$ , которая индивидуальна для каждой точки сцены. Поэтому наиболее естественная выборка в соответствии с PDF, пропорциональной известной части. Значение PDF для стохастических переменных  $x_0$  и  $\Theta_{x_0}$  будет иметь вид:

$$p_0(x, \Theta_x) = \frac{W_e(x, \Theta_x) |x, \Theta_x|}{W}, \quad (5.3)$$

где  $W$  представляет собой нормализующий коэффициент:

$$W = \int_A \int_{\Omega_x} W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x.$$

Эта PDF гарантирует, что направление и координаты точки луча выбираются таким образом, что вклад в световой поток, вычисляемый для данной точки, пропорционален значимости этого светового потока. Определение  $W_e(x, \Theta_x)$  и PDF соответственно зависит от модели камеры. Если построение такой PDF вызывает определенные трудности, то, как правило, можно построить приближенную PDF и соответствующим образом изменить оценки. Как правило, для сложной сцены построение такой PDF вызывает ряд технических сложностей и напоминает расчет формфакторов геометрических объектов сцены (3.4), который должен быть выполнен перед процессом рендеринга. Кроме того, должна быть выполнена нормировка и обращение интеграла вероятностей. Более простое решение – определять  $W_e(x, \Theta_x)$  для заданной камеры, используя выражение (5.1). В этом случае решение будет зависеть от типа камеры.

Можно отметить, что камера может быть не только точечной, формирующей «идеальное» изображение в единицах яркости, или линзовой, формирующей распределение освещенности на изображении, но и специальным детектором, собирающим световой поток в некотором телесном угле  $\Omega_x$  и площади  $A$ , как показано на рис. 5.5.

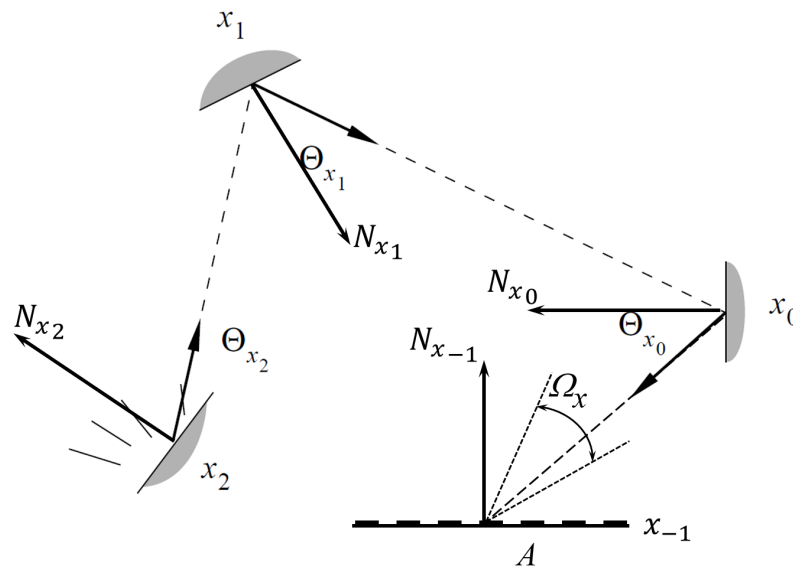


Рис. 5.5. Расчет глобального освещения виртуального детектора светового излучения. Детектор, накапливающий световое излучение, представляет собой поверхность площадью  $A$  и телесным углом  $\Omega_x$ . Структура детектора позволяет накапливать пространственное и(или) угловое распределение светового потока и затем осуществлять его пересчет в единицы интенсивности, освещенности или яркости изображения

В данном случае формула (5.2) приобретает несколько иной вид:

$$\langle \Phi \rangle_{path} = \frac{W_e(x_{-1}, \Theta_{x_0}) |\Theta_{x_0} \cdot N_{x_{-1}}|}{p_0(x_{-1}, \Theta_{x_0})} \cdot \prod_{j=1}^l \left[ \frac{f_r(x_{j-1}, \Theta_{x_j}, \Theta_{x_{j-1}}) |\Theta_{x_j} \cdot N_{x_{j-1}}|}{R_j p_j(\Theta_{x_{j-1}})} \right] L_e(x_i, \Theta_{x_i}), \quad (5.4)$$

Плотность вероятности, с которой выбираются  $(x_{-1}, \Theta_{x_0})$ , похожа на (5.3). Однако величина  $W_e(x, \Theta_x)$  априори неизвестна, поэтому ее надо либо исключить из PDF (установить равной 1), либо оценить с помощью предварительного расчета как функцию видности, например, методом трассировки лучей с поверхности  $A$  в телесном угле  $\Omega_x$ .

Уравнение рендеринга (2.13) имеет известную часть  $f_r(x_{i-1}, \Theta_{x_i}, \Theta_{x_{i-1}}) |\Theta_{x_i} \cdot N_{x_{i-1}}|$  в каждой точке рекурсии. Интегралом этого коэффициента по полусфере будет функция  $\rho(x_{i-1}, \Theta_{x_{i-1}})$  – коэффициент направленного отражения. Согласно закону сохранения энергии (2.9), эта величина всегда меньше 1. Таким образом, эта функция может быть использована в качестве PDF для выборки  $\Theta_{x_i}$  на полусфере, см. рис. 5.6:

$$R_i p_i(\Theta_x) = f_r(x_{i-1}, \Theta_{x_i}, \Theta_{x_{i-1}}) |\Theta_{x_i} \cdot N_{x_{i-1}}|. \quad (5.5)$$

Как показано в выражении (4.34), первичная оценка (5.2) упрощается к виду:

$$\langle \Phi \rangle_{path, imp} = W \sum_{i=0}^l L_e(x_i, \Theta_{x_i}). \quad (5.6)$$

Выборка по значимости, основанная на ДФР, гарантирует, что плотность лучей в направлениях больших значений ДФР будет выше, и это соответствует локальным отражающим свойствам поверхности. В большинстве случаев аналитическое решение задачи построения выборок в соответствии с локальными свойствами поверхности невозможно, поскольку интегрирование PDF, а затем их инвертирование аналитически не решаются. Поэтому возможны два решения. Во-первых, создается некоторая приближительная PDF, имеющая сходство с реальной ДФР, а затем задача решается с помощью более общего подхода, представленного в уравнении (4.2). Это можно решить, используя модифицированную модель Фонга, имеющую аналитическое решение, см. [44] и раздел 4.7.4. Второе решение – это табулирование функции ДФР, численное интегрирование ДФР и

формирование вероятности в виде табулированной функции. Табулированная функция вероятности может быть инвертирована численным образом (данный подход был рассмотрен в разделе 4.7.4) и служить основой для формирования случайных величин с требуемой PDF.

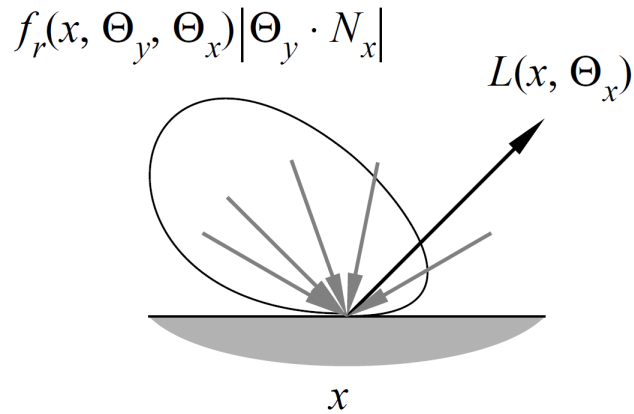


Рис. 5.6. Выборка значений переменных в уравнении рендеринга обычно основана на локальных свойствах отражения поверхности. Большее число выборок происходит вдоль направлений, для которых значение ДФР  $f_r(x_i, \Theta_y, \Theta_x)$ , умноженное на косинус угла между нормалью и направлением освещения  $|\Theta_{x_i} \cdot N_x|$ , велико. Поэтому алгоритм трассировки путей пытается выделить более «важные» траектории луча. Выборка по значимости позволяет в основном уменьшить оценку дисперсии для квазизеркальных поверхностей, где ДФР в большей степени определяет вид подынтегрального выражения

Что еще более важно, PDF может быть не идеальной из-за неизвестной функции яркости в интегральном уравнении. Выборка по значимости, выполняемая по плотности вероятности, пропорциональной  $f_r(x_{i-1}, \Theta_{x_i}, \Theta_{x_{i-1}}) |\Theta_{x_i} \cdot N_{x_{i-1}}|$ , игнорирует фактор яркости, который может оказывать большое влияние на вид подынтегрального выражения. Источники света могут находиться вне области зеркального рассеивания, куда PDF направляет лучи, и яркость, накопленная по этим «оптимальным» направлениям, будет нулевой. Поэтому важно указание на яркие области сцены, куда следует направлять лучи. Однако яркие области сцены также не являются панацеей, обе составляющие PDF ( $L_e(x_i, \Theta_{x_i})$  и  $f_r(x_{i-1}, \Theta_{x_i}, \Theta_{x_{i-1}}) |\Theta_{x_i} \cdot N_{x_{i-1}}|$ ) должны участвовать вместе для формирования оптимальной PDF.

Выборка по значимости необходима для алгоритма трассировки путей, поскольку это единственный способ эффективного сбора яркости в направлении острых пиков квазизеркальных ДФР, которые доминируют в подынтегральном выражении. В этой ситуации функция яркости будет относительно постоянной по



сравнению с ДФР, поэтому ДФР является хорошей основой для формирования PDF. Данный алгоритм лучше всего работает в сценах с квазизеркальными поверхностями и большими протяженными источниками света, например, небо или HDR изображения.

#### 5.4 Оценка следующего события

Для большинства сцен источники света достаточно малы по отношению к размеру сцены. Это приводит к тому, что вероятность попадания лучей в источник света невелика, и большинство случайных траекторий не будет влиять на оценку значения потока. Данный эффект аналогичен выборке «попадание или промах», которая заведомо имеет большую дисперсию. Как обсуждалось в разделе 4.6.3, данная проблема может быть частично исправлена путем применения оценки следующего события к выборке уравнения рендеринга (2.13):

$$\begin{aligned} L(x, \Theta_x) &= L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} L(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \\ &= L_e(x, \Theta_x) + L_r(x, \Theta_x), \end{aligned}$$

где  $L_r(x, \Theta_x)$  соответствует яркости вторичного освещения, полученного в результате диффузного отражения света. Можно записать:

$$\begin{aligned} L_r(x, \Theta_x) &= \int_{\Omega_x^{-1}} [L_e(y, \Theta_y) + L_r(y, \Theta_y)] f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \\ &= \int_{\Omega_x^{-1}} L_e(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \\ &\quad + \int_{\Omega_x^{-1}} L_r(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \\ &= \int_A L_e(y, \Theta_y) f_r(x, \Theta_{y \rightarrow x}, \Theta_x) v(x, y) \frac{|\Theta_{y \rightarrow x} \cdot N_x| |\Theta_{y \rightarrow x} \cdot N_y|}{|x - y|^2} d\mu_y \\ &\quad + \int_{\Omega_x^{-1}} L_r(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \\ &= L_d(x, \Theta_x) + L_i(x, \Theta_x), \end{aligned}$$

где:

$$L_d(x, \Theta_x) = \int_A L_e(y, \Theta_y) f_r(x, \Theta_{y \rightarrow x}, \Theta_x) v(x, y) \frac{|\Theta_{y \rightarrow x} \cdot N_x| |\Theta_{y \rightarrow x} \cdot N_y|}{|x - y|^2} d\mu_y, \quad (5.7)$$

$$L_i(x, \Theta_x) = \int_{\Omega_x^{-1}} L_r(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y. \quad (5.8)$$

Первый интеграл  $L_d(x, \Theta_x)$  определяет яркость прямого освещения. Второй интеграл  $L_i(x, \Theta_x)$  определяет яркость вторичного освещения. Яркость

вторичного освещения формируется аналогичным образом, осуществляя расчет яркости «первичного» освещения для всех точек, в которых осуществляется сбор прямого освещения  $L_d(x, \Theta_x)$ . В результате мы имеем бесконечную рекурсию или интеграл бесконечной размерности.

Яркость прямого освещения обычно оценивают путем случайного выбора точки на источнике света с последующим расчетом и интегрированием яркости по выбранным локальным источникам света (2.15). Простейшим вариантом для выборки точки на источнике света является PDF, которая пропорциональна светимости источника света, т.е. пространственному распределению светового потока или интегралу яркости по полусфере в каждой точке источника света:

$$p(y) = \frac{B(y)}{\Phi}, \quad (5.9)$$

где:

- $B(y) = \int_{\Omega_y} L_e(y, \Theta_y) |\Theta_y \cdot N_y| d\omega_y$  – светимость источника света в точке  $y$   $[\text{Вт}/\text{м}^2]$ , для общности можно использовать обозначение  $\Phi(y)$ ,
- $\Phi = \int_A \int_{\Omega_y} L_e(y, \Theta_y) |\Theta_y \cdot N_y| d\omega_y d\mu_y$  – общая светимость всех источников сцены или световой поток, излучаемый всеми источниками сцены  $[\text{Вт}]$ .

Все источники света в сцене можно рассматривать как единый источник света со сложным распределением яркости по его поверхности. Каждый отдельный источник света формирует разрыв функции плотности вероятности (5.9). Кроме того, удобнее говорить не о светимости источника света  $B(y)$ , а о распределении светового потока (плотности светового потока, обозначая его  $\Phi(y)$ ) по источникам сцены. Это позволит ввести в модель источники света, не обладающие яркостью вследствие своего бесконечно малого линейного размера (например, точечный источник света) или углового размера (например, параллельный источник света). Для таких источников света выражения (5.7) и (5.9) приобретают несколько другой вид. Для источников с бесконечно малой линейной размерностью:

$$L_d(x, \Theta_x) = \int_A I_e(y, \Theta_y) f_r(x, \Theta_{y \rightarrow x}, \Theta_x) v(x, y) \frac{|\Theta_{y \rightarrow x} \cdot N_x|}{|x-y|^2} d\mu_y, \quad (5.10)$$

где  $I_e(y, \Theta_y)$  – интенсивность излучения (сила света) источника с точки  $y$  в направлении  $\Theta_y$  на точку наблюдения  $x$ . Интегрирование по области  $A$  сохраняется, поскольку источник света может иметь нулевую размерность только в одном измерении (например, линейный источник света). Плотность вероятности примет следующий вид:

$$p(y) = \frac{I(y)}{\Phi}, \quad (5.11)$$

где:

- $I(y) = \int_{\Omega_y} I_e(y, \Theta_y) d\omega_y$  – пространственное распределение светового потока в точке  $y$ , для общности можно использовать обозначение  $\Phi(y)$ ,
- $\Phi = \int_A \int_{\Omega_y} I_e(y, \Theta_y) d\omega_y d\mu_y$  – общий световой поток, испускаемый источником света данного типа [Вт].

Для источников с бесконечно малой угловой размерностью выражения (5.7) и (5.9) приобретают следующий вид:

$$L_d(x, \Theta_x) = \int_{\Omega_A} E_e(y, \Theta_y) f_r(x, \Theta_{y \rightarrow x}, \Theta_x) v(x, y) |\Theta_{y \rightarrow x} \cdot N_x| d\omega_y, \quad (5.12)$$

где  $E_e(y, \Theta_y)$  – плотность светового потока с точки  $y$  в направлении  $\Theta_y$  на точку наблюдения  $x$ . Предполагается, что источник света имеет бесконечную линейную размерность, которая в численных расчетах ограничена областью определения сцены  $A_D$ . Интегрирование осуществляется по угловой области источника света  $\Omega_A$ . Данное интегрирование необходимо, поскольку источник света может иметь нулевую размерность только в одном измерении (например, дуга в сферической системе координат). Плотность вероятности в данном случае определяется для направления  $\Theta_y$  и примет следующий вид:

$$p(\Theta_y) = \frac{E_e(\Theta_y)}{\Phi}, \quad (5.13)$$

где:

- $E_e(\Theta_y) = \int_{A_D} E_e(y, \Theta_y) d\mu_y$  – угловое распределение светового потока в точке  $y$ , для общности можно использовать обозначение  $\Phi(y)$ ,
- $\Phi = \int_{\Omega_A} \int_{A_D} E_e(y, \Theta_y) d\mu_y d\omega_y$  – общий световой поток, испускаемый источником света данного типа [Вт].

При наличии в сцене ряда разнородных источников выражение (5.9) не может быть явным образом применено. Поэтому к функции плотности вероятности (5.9) следует добавить еще одну размерность – номер источника света, а выборку точки на источнике света свести к двум последовательным выборкам: выбор номера источника света  $j$  и затем выбор точки на источнике света. Плотность вероятности выбора источника света  $j$  выражается следующим образом:

$$p(j) = \frac{\Phi(j)}{\sum_{j=1}^K \Phi(j)}, \quad (5.14)$$

где  $K$  – число источников света в сцене. Далее, по формулам (5.9), (5.11) или (5.13) выполняется выбор точки на источнике света  $y$  или  $\Theta_y$  в зависимости от типа источника света.

Если все интегралы выбираются в соответствии с PDF, представленными выражениями (5.3), (5.5), (5.9), (5.11) и (5.13), то вид оценки светового потока, представленного выражением (5.6), изменяется на

$$\langle \Phi \rangle_{path,imp,nextev} = W[L_e(x_0, \Theta_{x_0}) + \sum_{i=0}^l \frac{\Phi_\Sigma}{\Phi(j)} f_r(x_i, \Theta_{y \rightarrow x_i}, \Theta_{x_i}) v(x_i, y) \cdot E_e(y, \Theta_{y \rightarrow x_i}) |\Theta_{y \rightarrow x_i} \cdot N_{x_i}|], \quad (5.15)$$

где  $\Phi_\Sigma = \sum_{j=1}^K \Phi(j)$   $E_e(y, \Theta_{y \rightarrow x_i})$  – освещенность, приведенная к световому потоку  $\Phi(j)$ , в точке наблюдения в направлении на точку освещения. Освещенность вычисляется в зависимости от типа источника света (три основных типа источников представлены на рис. 5.7):

- $E_e(y, \Theta_{y \rightarrow x_i}) = L_e(y, \Theta_y) \frac{|\Theta_{y \rightarrow x_i} \cdot N_y|}{|x_i - y|^2}$  – протяженный источник света,
- $E_e(y, \Theta_{y \rightarrow x_i}) = I_e(y, \Theta_y) \frac{1}{|x_i - y|^2}$  – точечный источник света,
- $E_e(y, \Theta_{y \rightarrow x_i})$  – для параллельного источника света данная величина задается непосредственно.

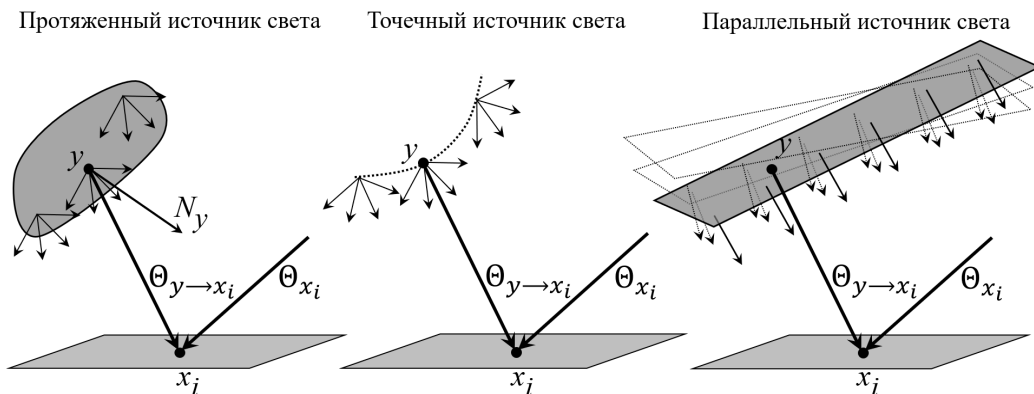


Рис. 5.7. Метод расчета яркости прямого освещения для трех основных типов источников света. Слева направо: протяженный источник света, точечный источник света и параллельный источник света

Обозначения, представленные в уравнении (5.15), представлены на рис. 5.8.

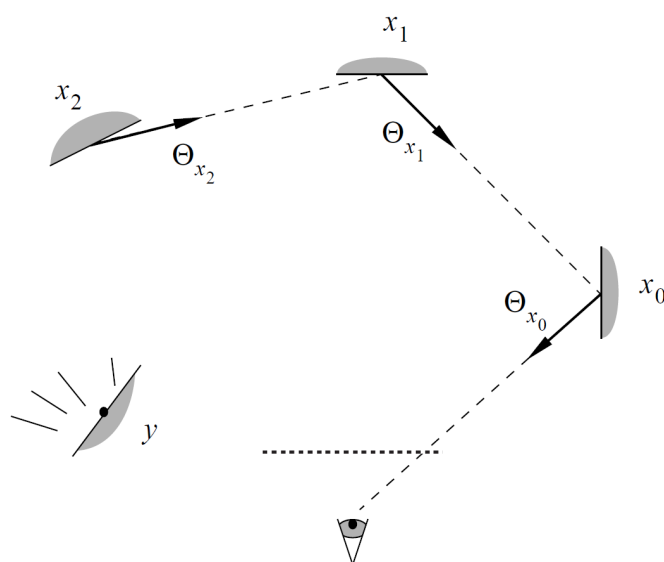


Рис. 5.8. Символы, используемые в алгоритме трассировки путей с оценкой следующего события. Случайная трасса остается такой же, как и без оценки следующего события (см. рис. 5.1), однако, на каждом диффузном событии на источнике света выбирается дополнительная точка  $y$  для оценки яркости прямого освещения

Результирующий алгоритм представляет собой трассировку путей с явной выборкой источников света для вычисления яркости прямого освещения, см. рис. 5.9. Алгоритм трассирует случайные лучи из камеры, которые затем распространяются в сцене. Функция видимости оценивается с помощью теневого луча в каждой точке пересечения луча с диффузной поверхностью.

Особое внимание следует обратить на результат взаимодействия луча с идеально зеркальной и квазизеркальной поверхностями. Отличие идеально зеркальной поверхности от квазизеркальной заключается в том, что первый тип поверхности не формирует собственной яркости, а осуществляет перенос точки наблюдения на следующую поверхность трассы луча. С точки зрения модели рассеивания зеркальная ДФР представляет собой масштабированную дельта-функцию, причем величина масштаба – это коэффициент зеркального отражения. Для диэлектрических поверхностей коэффициент отражения может быть вычислен по формулам Френеля. Формальное применение уравнения (5.15) к ДФР в виде дельта-функции возможно. В соответствии с выражением (5.15) осуществляется расчет яркости прямого освещения по формуле (5.15), которая всегда будет нулевой, и далее происходит перенос луча на следующую

поверхность сцены в соответствии с законом отражения (формально выражение (5.5) может быть применено, поскольку PDF интегрируема, и обратная функция может быть получена). Квазизеркальная поверхность, в отличие от идеального зеркала, формирует узкое рассеивание вблизи зеркальной трассы луча и формирует яркость прямого освещения вокруг этой узкой области наблюдения.

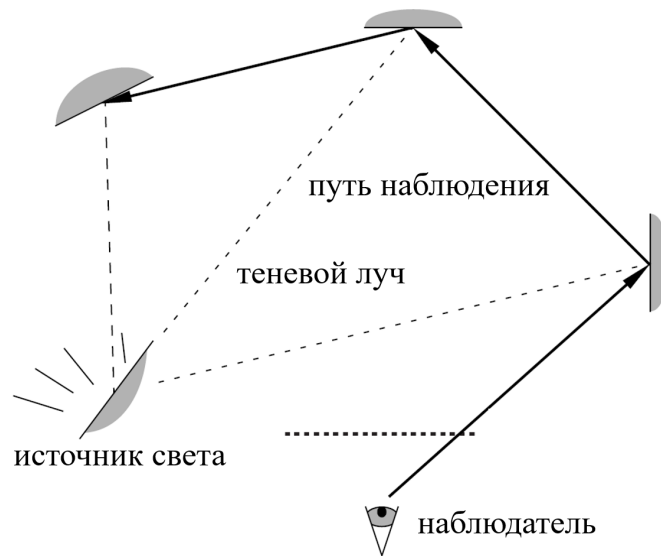


Рис. 5.9. Схема алгоритма трассировки путей с оценкой следующего события. Яркость прямого освещения вычисляется явно в каждой точке случайной трассы луча путем выборки случайной точки на случайном источнике света

PDF, представленная выражениями (5.9), (5.11) и (5.13), позволяет выбрать точки на источниках света для расчета яркости прямого освещения, но не учитывает никаких особенностей интеграла, оцениваемого в (5.15). В результате оценка может стать сколь угодно большой, например, из-за близости выбранной точки наблюдения к точке освещения (для точечного и протяженного источников света в оценке яркости появляется множитель  $\frac{1}{|x_i - y|^2}$  в выражении освещенности) или из-за «неудачного» выбора точки на источнике света, совпадающей с направлением зеркального отражения в точке наблюдения на поверхности с квазизеркальной ДФР. Эти оценки яркости могут на порядки превышать средние оценки яркости после усреднения ряда оценок в соседних точках изображения. В конечном итоге это приводит к появлению свехярких точек на изображении.

В отличие от предыдущего алгоритма, данный алгоритм ориентирован на расчет сцен с преимущественно плавными ДФР и источниками света малого размера, освещающими преимущественно большую часть сцены. В оценке

данного интеграла доминируют PDF, полученные для источников света. Рис. 5.10 и 5.11 иллюстрируют работу данного алгоритма.

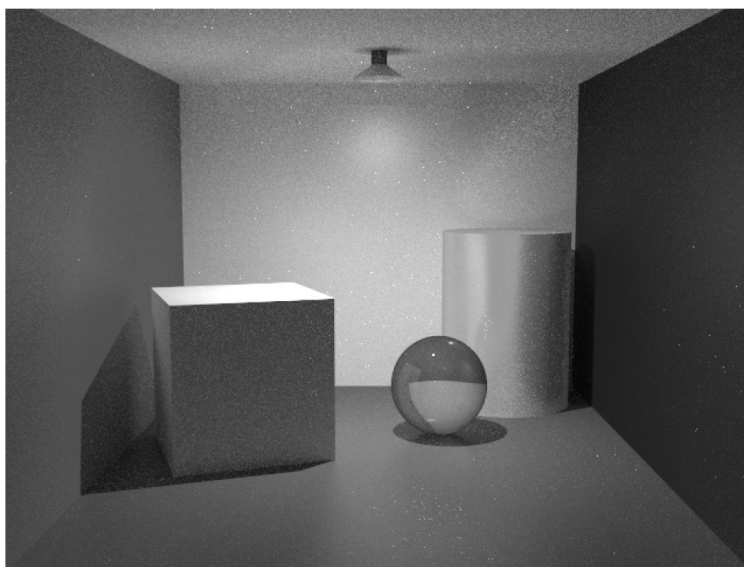


Рис. 5.10. Пример изображения сцены, полученной с помощью алгоритма трассировки пути с оценкой следующего события. Алгоритм хорошо работает с источниками света малого размера

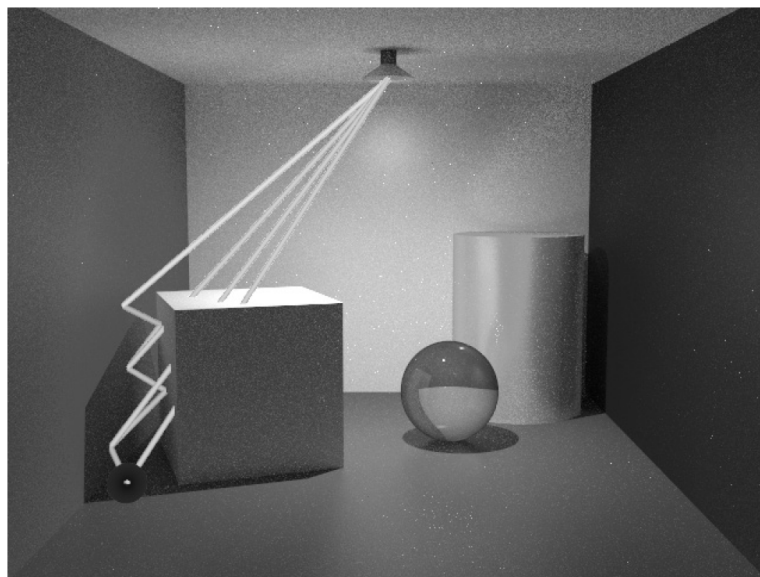


Рис. 5.11. Пример случайной трассы луча, начинающегося с точки в левом нижнем углу изображения. С каждой точки сегмента трассы луча отбрасывается теневой луч на источник света. Если луч не экранируется объектом сцены, то к точке изображения добавляется яркость прямого освещения. В случае экранирования яркость не добавляется

Основным недостатком данного подхода является невозможность физически корректного расчета яркости каустического освещения. Каустическое освещение возникает при наблюдении диффузной поверхности, освещенной свозь ряд поверхностей с идеальными свойствами зеркального отражения. Метод, основанный на выборке по значимости, частично решает эту проблему, поскольку факт попадания луча в источник света не связан с типом события луча. Это может быть как случай каустического, так и диффузного освещения. Основные проблемы этого метода – это низкая эффективность и невозможность осуществить попадание луча в источник света бесконечно малого размера (линейного или углового). При использовании метода оценки следующего события яркость «прямого» видения источника света ушла из выражения (5.15), поскольку пересечение луча с зеркальной поверхностью не формирует яркость, и попадание луча в источник света также не формирует яркость. Поэтому выражение (5.15) должно быть расширено некоторой логикой:

$$\langle \Phi \rangle_{path,imp,nextev} = W[L_e(x_0, \Theta_{x_0}) + \sum_{i=0}^l \frac{\Phi_{\Sigma}}{\Phi_{(j)}} f_r(x_i, \Theta_{y \rightarrow x_i}, \Theta_{x_i}) v(x_i, y) \cdot E_e(y, \Theta_{y \rightarrow x_i}) |\Theta_{y \rightarrow x_i} \cdot N_{x_i}| \quad (5.16)$$

$$+ \left[ \begin{array}{l} L_e(x_i, \Theta_{x_i}), \text{ if } i > 0 \wedge f_r(x_{i-1}, \Theta_{x_{i-1} \rightarrow x_i}, \Theta_{x_i}) \equiv \delta \\ 0, \text{ if } f_r(x_{i-1}, \Theta_{x_{i-1} \rightarrow x_i}, \Theta_{x_i}) \neq \delta \end{array} \right].$$

То есть алгоритм выборки по значимости добавляет свою яркость в суммарную оценку только в том случае, если предыдущее преобразование на поверхности имело идеально зеркальный характер. Однако, данное решение не помогает для случая источников света бесконечно малого размера.

## 5.5 Многократная выборка по значимости

Одним из наиболее удачных примеров приложений для объединения различных оценок является многократная выборка по значимости, используемая для расчета яркости прямого освещения, см. [33]. Если используется выборка по значимости для формирования траектории луча, рассеянного по полусфере (или сфере в случае полной ДФР, включающей как отражение, так и пропускание) в соответствии с ДФР, то яркость формируется на сегментах луча, попадающих на источники света, см. (5.6). Если для оценки яркости применяется метод оценки следующего события, с помощью которого происходит выбор источника света и точки на источнике света в соответствии с распределением плотности светового потока (что также означает выборку по значимости), то мы также оцениваем яркость прямого освещения на каждом из сегментов трассы луча, см. (5.15). В



результате оба подхода позволяют оценить одну и ту же яркость на каждом сегменте луча, и эти локальные яркости могут быть переоценены с учетом их локальной значимости. Необходимо отметить, что данное решение не требует дополнительных вычислений и трассировок лучей по отношению к (5.15), поскольку трассировка лучей в сцене, позволяющая оценить яркость прямого видения, также используется для оценки выбора следующего события.

Для того, чтобы выполнить многократную выборку по значимости, обе выборки должны быть выполнены в одном пространстве параметров, например, в пространстве точек на поверхностях сцены. Поэтому, когда алгоритм выбирает направление луча, а затем трассирует его до точки пересечения с поверхностью сцены, то плотность вероятности для выборки той точки можно определить с помощью простого преобразования. Если  $p(\Theta)$  – это PDF, выбирающая направление  $\Theta_x$  на полусфере  $\Omega_x$ , и если  $y = r(x, \Theta_x)$ , то неявная плотность вероятности для выборки  $y$  равна (см. рис 5.12):

$$p(y)d\mu = p(\Theta_x)d\omega \Rightarrow p(y) = p(\Theta_x) \frac{|\Theta_x \cdot N_y|}{|x_i - y|^2}$$

$$p(\Theta_x) = \frac{B(y)}{\Phi} \frac{|\Theta_x \cdot N_y|}{|x_i - y|^2}$$

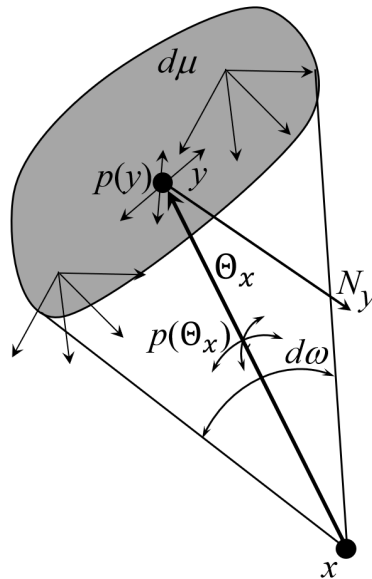


Рис. 5.12. Связь между угловым и пространственным распределениями случайной величины

Алгоритм многократной выборки по значимости может быть построен на сбалансированной или степенной эвристиках (степень, как правило, равна двум). В простейшем случае веса могут быть пропорциональны PDF, приведенных к одному пространству направлений  $p(\Theta_x)$ . Алгоритм имеет следующий вид:

```

Color lum; // оценка яркости в точке изображения
Ray ray; // луч, трассируемый из камеры, содержит начало луча, направление,
// точку попадания на поверхность, новое направление в точке и
// нормали в точках начала и конца

// Испустить луч из камеры
ray = scene.Camera().EmitRay();

// Пока луч не покинет сцену или поглотится
for (int i = 0; (Object object = scene.Intersect(ray)) is OK; i++)
{
Color BRDF_lum, light_lum, light_illum;
double PDF_BRDF, PDF_light;
Vector point;
Light light;

// Первое пересечение со сценой
if (i == 0)
lum = object.TraceLuminance(ray); continue; //  $L_e(x_0, \Theta_{x_0})$ 

// Яркость прямого освещения источником света (light sampling)
// Выбор источника света и точки на источнике света
scene.ChooseLightAndPoint(light, point); //  $p(y) = \frac{B(y)}{\Phi}$ 
// Вычисление PDF
PDF_light = light.PDF(point); //  $p_i(\Theta_x) = \frac{B(y)}{\Phi} \frac{|\Theta_{y \rightarrow x_i} \cdot N_y|}{|x_i - y|^2}$ 
// Расчет PDF при рассеивании луча на точку y
PDF_BRDF = object.PDF(ray, point); //  $p_i(\Theta_x) = f_r(x_{i-1}, \Theta_{x_i}, \Theta_{x_{i-1}}) |\Theta_{x_i} \cdot N_{x_{i-1}}|$ 
// Расчет затенения
double v = scene.Shadow(ray.Point(), point);
// Расчет прямого освещения  $E_e(y, \Theta_{y \rightarrow x_i})$ 
light_illum = light.Illuminance(ray, point); //  $E_e(y, \Theta_{y \rightarrow x_i}) = L_e(y, \Theta_y) \frac{|\Theta_{y \rightarrow x_i} \cdot N_y|}{|x_i - y|^2}$ 
// Расчет яркости прямого освещения
light_lum = object.Luminance(ray, point, light_illum);
//  $L(x_i, \Theta_y) = \frac{B(y)}{\Phi} f_r(x_i, \Theta_{y \rightarrow x_i}, \Theta_{x_i}) v(x_i, y) E_e(y, \Theta_{y \rightarrow x_i}) |\Theta_{y \rightarrow x_i} \cdot N_{x_i}|$ 
// Взвешенная яркость
lum += PDF_light * light_lum / (PDF_light + PDF_BRDF);
}

```

```

// Яркость прямого попадания в источник света (BRDF sampling)
// Новое направление луча, включая проверку на поглощение
if (object.TransformRay(ray) is not OK)
    break;
// Трассировка луча и расчет яркости на объекте, куда попал луч
lum_BRDF = object.TraceLuminance(ray); //  $L_e(x_i, \Theta_{x_{i+1}})$ 
// PDF объекта BRDF
PDF_BRDF = object.BRDFPDF(ray); //  $p_i(\Theta_x) = f_r(x_i, \Theta_{x_{i+1}}, \Theta_{x_i}) |\Theta_{x_i} \cdot N_{x_i}|$ 
// PDF объекта как источника света
PDF_light = object.LightPDF(point); //  $p_i(\Theta_x) = \frac{B(x_{i+1}) |\Theta_{x_{i+1} \rightarrow x_i} \cdot N_{x_{i+1}}|}{\Phi |x_i - x_{i+1}|^2}$ 
// Взвешенная яркость
lum += PDF_BRDF * BRDF_lum / (PDF_light + PDF_BRDF);
}

```

Данный алгоритм является простым, но он не оптимален с точки зрения выбора весов, поскольку он не учитывает общих условий освещения в точке наблюдения, а ориентируется при оценке весов на случайно выбранный источник света. Кроме того, выборка значимости не ориентируется на условия оптимальности с точки зрения глобального освещения.

## 5.6 Выделение главной части

Метод выделения главной части, рассмотренный в разделе 4.5, также может быть применен к построению математической модели глобального освещения. Аппроксимирующая функция, которая используется в качестве PDF для выборки по значимости, не может улучшить результаты сходимости, если ее использовать в качестве главной части. Однако было показано, что выделение главной части может улучшить результат, если PDF является не критичной. Это возможно для уравнения рендеринга (2.13) с использованием PDF в соответствии (5.5). Поэтому в [45] было предложено выделить главную часть, пропорциональную PDF, предполагая, что входящее излучение равно некоторому постоянному окружающему излучению  $L_a$  (рис. 5.13). В результате уравнение рендеринга приводится к следующему виду:

$$\begin{aligned}
L(x, \Theta_x) &= L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} L(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \\
&= L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} L_a f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \\
&\quad + \int_{\Omega_x^{-1}} [L(y, \Theta_y) - L_a] f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \quad , \\
&= L_e(x, \Theta_x) + \rho(x, \Theta_x) L_a \\
&\quad + \int_{\Omega_x^{-1}} [L(y, \Theta_y) - L_a] f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y
\end{aligned}$$

где  $\rho(x, \Theta_x)$  – интегральный коэффициент отражения поверхности в точке  $x$  при ее наблюдении по направлению  $\Theta_x$ .

Если выборка осуществляется по той же PDF, то оценка (5.6) будет иметь вид:

$$\begin{aligned}
\langle \Phi \rangle_{path, impon, c} &= W [L_e(x_0, \Theta_{x_0}) + \rho(x_0, \Theta_{x_0}) L_a \\
&\quad + \sum_{i=1}^l [L_e(x_i, \Theta_{x_i})] - L_a + \rho(x_i, \Theta_{x_i}) L_a] \quad . \quad (5.17) \\
&= W \left[ L_a + \sum_{i=0}^l [L_e(x_i, \Theta_{x_i}) - (1 - \rho(x_i, \Theta_{x_i})) L_a] \right]
\end{aligned}$$

Метод выделения главной части вносит некоторый приближительный вклад в оценку, даже если русская рулетка завершает случайное блуждание в какой-то случайной точке. При этом результат остается несмещенным.

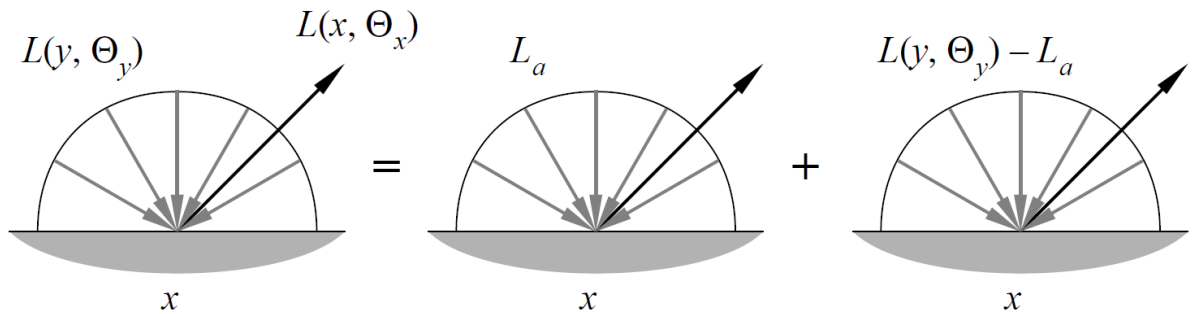


Рис. 5.13. Простую главную часть для уравнения рендеринга можно получить, выбрав постоянную величину входящего излучения окружающего пространства  $L_a$ . Результирующее отраженное излучение равно  $\rho(x, \Theta_x)L_a$ , которое может быть вычислено аналитически. Далее, вместо выборки  $L(y, \Theta_y)$  алгоритм трассировки пути суммирует это приближение с результатом выборки разности яркостей  $L(y, \Theta_y) - L_a$  для получения несмещенной оценки фактического отраженного излучения  $L(x, \Theta_x)$ . При выборе правильного значения  $L_a$  оценка потока будет иметь меньшую дисперсию

Метод выделения главной части также может быть объединен с оценкой следующего события. Он приблизительно оценивает вторичное входящее излучение, поскольку самосвечение вычисляется отдельно. Оценка (5.15) приобретает вид:

$$\begin{aligned} \langle \Phi \rangle_{path,imp,nextev,con} = & W[L_e(x_0, \Theta_{x_0}) + L_a \\ & + \sum_{i=0}^l \left[ \frac{\Phi_{\Sigma}}{\Phi(j)} f_r(x_i, \Theta_{y \rightarrow x_i}, \Theta_{x_i}) v(x_i, y) \cdot E_e(y, \Theta_{y \rightarrow x_i}) \Big|_{\Theta_{y \rightarrow x_i}} \cdot N_{x_i} \right], \\ & - (1 - \rho(x_i, \Theta_{x_i})) L_a ]. \end{aligned} \quad (5.18)$$

Уравнение (4.32) формально выражает, насколько близко интегральная главная часть должна приближаться к фактическому интегральному значению, чтобы уменьшить дисперсию выборки. Для выражения (4.17), не учитывающего оценку следующего события, это означает, что  $L_a$  должно быть таким, чтобы в каждой точке  $x_i$ :

$$0 < \rho(x_i, \Theta_{x_i}) L_a < 2L_r(x_i, \Theta_{x_i}),$$

где  $L_r$  – реально отраженная яркость, см. главу 4.6.4.

Для выражения (4.18), учитывающего оценку следующего события, это означает, что  $L_a$  должно быть таким, чтобы в каждой точке  $x_i$ :

$$0 < \rho(x_i, \Theta_{x_i}) L_a < 2L_i(x_i, \Theta_{x_i}),$$

где  $L_r$  – отраженное излучение, возникшее в результате вторичного освещения.

Для ряда сцен  $\frac{L_r(x_i, \Theta_{x_i})}{\rho(x_i, \Theta_{x_i})}$  и  $\frac{L_i(x_i, \Theta_{x_i})}{\rho(x_i, \Theta_{x_i})}$  могут быть относительно постоянными величинами, поскольку освещение становится более плавным после каждого отражения. Таким образом, если рендеринг выполняется итерационно, как череда последовательных расчетов яркости с последующим усреднением результата, то можно выбрать оптимальный уровень яркости внешнего освещения  $L_a$ . Кроме того, каждая точка изображения может иметь собственные значения  $L_a$ . Необходимо отметить, что значение  $L_a$  может быть выбрано близким к нулю, при этом дисперсия возвращается к дисперсии методов выборки по значимости или оценки следующего события.

## 5.7 Улучшение выборки по значимости и выделению главной части

Для дальнейшего улучшения результатов работы методов выборки по значимости и выделению части можно принять во внимание локальные характеристики отражения поверхностей (ДФР), представляющие собой важный фактор подынтегрального выражения яркости. Яркость поля, которая является другим фактором в подынтегральном выражении, просто аппроксимируется как постоянная. Поэтому подробная информация о яркости поля могла бы помочь улучшить работу методов выборки по значимости и выделению главной части.

Для более корректного определения функции значимости информация о яркости поля по всей сцене должна храниться в соответствующей структуре данных. Возможны различные представления структур данных для хранения информации о распределении яркости по объектам сцены. Методы граничных элементов, например, переноса излучения (radiosity), хранят информацию для каждого из элементов дискретизированных поверхностей, например, треугольников, формирующих поверхность. Как правило, информация ограничивается объектами, рассеивающими световое излучение по закону Ламберта, однако существуют решения для хранения информации о направлении, например, путем дискретизации полусферы [46] или с помощью сферических гармоник [47]. Также можно использовать различные воксельные структуры с различной технологией разбиения пространства (регулярное, октодереве, бинарное дерево, BVH дерево и т. п. [48-50]).

В общем случае необходимо иметь пятимерное дерево, определяющее распределение яркости поля по сцене (учет цвета превратит это представление в шестимерное дерево). Размеры представляют собой три обычных пространственных измерения и два дополнительных угловых измерения. Для удобства точка в пятимерном пространстве, соответствующая точке  $(x, y, z)$  в декартовой системе координат и направлению  $(\varphi, \theta)$ , нормализована так, чтобы значения координат в каждом из измерений находились в диапазоне  $[0,1]$ :

$$(x_n, y_n, z_n, s_n, t_n),$$

где:

$$\begin{aligned}x_n &= \frac{(x-x_{min})}{(x_{max}-x_{min})}, \\y_n &= \frac{(y-y_{min})}{(y_{max}-y_{min})}, \\z_n &= \frac{(z-z_{min})}{(z_{max}-z_{min})}, \\s_n &= \frac{(\cos \theta + 1)}{2},\end{aligned}$$

$$t_n = \frac{\varphi}{2\pi}.$$

Данная параметризация обеспечивает равномерное разбиение пятимерного пространства на воксели одинакового объема. Каждый воксель соответствует постоянному 3D объему и постоянной угловой 2D площади. Пятимерное дерево, когда каждый узел разбивается на 32 дочерних узла, требует огромной оперативной памяти, однако большинство реальных сцен, как правило, состоят из пустого пространства, в котором рассеивающие поверхности занимают небольшой объем.

Данные, хранящиеся в дереве, представляют собой усредненные значения яркости поля. В процессе моделирования алгоритм трассировки путей оценивает значения яркости поля, которая затем усредняется по значениям, уже сохраненным в соответствующих узлах дерева. Всякий раз, когда количество оценок яркости поля, усредненных в узле, превышает заданный максимальный порог, узел разделяется, создавая 32 новых дочерних узла. Данный подход обусловлен значимостью узла. Чем больше лучей попало в узел, тем выше его значимость и тем точнее должна быть оценена яркость поля в этом узле. Можно предложить несколько другую организацию дерева, построенную на бинарном разбиении пятимерного пространства. Пространство разбивается на 2 новых дочерних узла по измерению, которое накопило допустимое количество лучей. В любом случае максимальное разбиение пространства на воксели происходит в областях (измерениях), которые накопили большее число лучей.

Доступ к яркости поля в любой точке пространства может быть получен из листа дерева методом простого спуска по его узлам. Для ускорения доступа могут быть применены различные ускоряющие решения, например, хэш таблицы [51]. Кроме того, для ускорения доступа к данным и привязки к параметрам поверхностей, осуществляющих рассеивание света, деревья могут быть построены для каждого элемента сцены, обладающего уникальными оптическими свойствами. Это может позволить не усреднять параметры яркости поля, сформированной разнотипными объектами сцены. В любом случае дерево имеет абстрактный интерфейс и может быть настроено на любой способ его использования.

При попадании луча в пятимерный воксель мы можем использовать распределение яркости поля для создания PDF, по которой будет осуществляться выборка по значимости и выделение главной части. Исходные нормализованные PDF  $p_i(\theta_i)$  из уравнения (5.5) на основе ДФР неявно ведут к преобразованию в другое пространство параметров, которое рассматривалось в разделе 4.3.

Запишем это пространство параметров явным образом, используя  $(\xi_1, \xi_2) \in [0,1] \times [0,1]$ . Перепишем уравнения рендеринга:

$$\begin{aligned} L(x, \Theta_x) &= L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} L(y, \Theta_y) f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| d\omega_y \\ &= L_e(x, \Theta_x) + R \int_{\Omega_x^{-1}} L(y, \Theta_y) \frac{f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x|}{Rp(\Theta_y)} p(\Theta_y) d\omega_y, \\ &= L_e(x, \Theta_x) + \rho(x, \Theta_x) \int_0^1 \int_0^1 L(y, \Theta_y) d\xi_1 d\xi_2 \end{aligned}$$

где  $\Theta_y$  определяется параметрами  $(\xi_1, \xi_2)$ . Выборка, производимая в соответствии с исходными PDF, соответствует равномерной выборке в этом новом пространстве параметров. Чтобы учесть дополнительную информацию о яркости поля, в каждой точке пространства создается дополнительная PDF и выделяется главная часть в этом пространстве параметров. Для этой цели мы осуществляем повторную выборку сохраненных значений яркости поля на регулярной сетке. Это позволяет сформировать кусочно-постоянную аппроксимацию  $\tilde{L}(\xi_1, \xi_2)$  яркости поля. Далее это приближение можно использовать для построения дополнительной PDF:

$$p'(\xi_1, \xi_2) = \frac{\tilde{L}(\xi_1, \xi_2)}{\tilde{L}}, \quad (5.19)$$

где:

- $\tilde{L} = \int_0^1 \int_0^1 \tilde{L}(\xi_1, \xi_2) d\xi_1 d\xi_2$  – коэффициент нормализации PDF. Он может быть вычислен аналитически, поскольку  $\tilde{L}$  является кусочно-постоянной функцией.

Конечным результатом этой стратегии выбора является то, что больше лучей направляется не только по высокой отражающей способности ДФР (как было и раньше), но по направлениям, создающим наибольшую яркость. Оба фактора (PDF) могут быть важны для результирующие оценки яркости. Если этот процесс повторяется на каждом шаге рекурсии с использованием стратегии русской рулетки (на основе коэффициента отражения), то оценка потока принимает вид:

$$\langle \Phi \rangle_{path, betterimp} = W \sum_{i=0}^l \frac{\tilde{L}_i}{\tilde{L}_i(\xi_{1i}, \xi_{2i})} L_e(x_i, \Theta_{x_i}). \quad (5.20)$$

Выражение может быть адаптировано для метода оценки следующего события аналогично выражениям (5.15 - 5.16).



В работе [52] была предложена подобная методика, использующая несколько другой подход к расчету. Это решение сохраняет все отдельные световые фотоны (формирует фотонную карту), полученные на этапе прямой трассировки световых лучей. Во время процесса расчета яркости происходит извлечение этих данных для улучшения выборки по значимости. Метод расчета яркости аналогичен методу, рассмотренному выше. Кусочно-постоянная аппроксимация функции яркости также полезна для реализации метода выделения главной части. С использованием той же самой стратегии русской рулетки оценка светового потока в итоге принимает вид:

$$\begin{aligned}
 \langle \Phi \rangle_{path, betterimp, con} &= W [L_e(x_0, \Theta_{x_0}) + \rho(x_0, \Theta_{x_0}) \tilde{L}_0 \\
 &\quad \sum_{i=1}^l \left[ \frac{\tilde{L}_i}{\tilde{L}_i(\xi_{1i}, \xi_{2i})} (L_e(x_i, \Theta_{x_i}) - \tilde{L}_i(\xi_{1i}, \xi_{2i})) + \rho(x_i, \Theta_{x_i}) \tilde{L}_i \right] \\
 &= W \left[ \tilde{L}_0 + \sum_{i=0}^l \left[ \frac{\tilde{L}_i}{\tilde{L}_i(\xi_{1i}, \xi_{2i})} L_e(x_i, \Theta_{x_i}) - (1 - \rho(x_i, \Theta_{x_i})) \tilde{L}_i \right] \right].
 \end{aligned} \tag{5.21}$$

Выражение может быть адаптировано для включения метода оценки следующего события, аналогично уравнению (5.18).

## 5.8 Заключение

В этой главе было продемонстрировано, что алгоритмы рендеринга по методу Монте-Карло являются ничем иным, как стандартным подходом Монте-Карло, описанным в главе 4, применяемым к различным математическим моделям для решения проблемы глобального освещения, описанной в главе 2. Несмотря на то, что теоретические модели математически эквивалентны, различные методы оценки этих интегралов методом Монте-Карло приводят к появлению совершенно разных алгоритмов.

Уравнение рендеринга естественным образом приводит к стохастической трассировке лучей или, более конкретно, к трассировке путей, т.е. трасс лучей от камеры к источникам света, формирующим яркость. Алгоритм рендеринга оценивает световой поток, приходящий на каждую точку изображения (или ячейку детектора), исходя из яркости, которую наблюдает точка изображения. Первым шагом работы данного алгоритма является выбор направлений и поверхностей, которые видит точка изображения. Затем происходит рекурсивная оценка интегрального уравнения рендеринга путем выбора направлений рассеивания лучей. Каждый раз, когда выбранное направление приводит к источнику света, имеющему яркость, к оценке интеграла добавляется соответствующий вклад. Таким образом, данный алгоритм обладает достаточной

эффективностью только в том случае, если источники света велики по сравнению с остальной частью сцены.

Существует ряд методов уменьшения дисперсии. Эти методы ориентированы на специфику сцен и порождают специальные алгоритмы рендеринга. Например, различные методы стратифицированной выборки позволяют более равномерно распределять координаты на выходных зрачках оптических систем объективов, формировать более равномерные распределения по поверхностям источников света и т.п.

Выборка по значимости позволяет выбирать направления рассеивания света в соответствии с оптическими характеристиками поверхностей сцены, т.е. направлять большее число лучей в направлении пиков рассеивания. Этот подход обладает большим преимуществом для сцен, содержащих квази-зеркальные поверхности (поверхности с узконаправленной ДФР), поскольку оценка интегральной яркости в направлениях вне пика ДФР будет отмасштабирована на малое значение ДФР и может быть пренебрежимо мала. В результате луч, пошедший по траектории низкого значения ДФР, не принесет ощутимого вклада в световой поток точки изображения.

Оценка следующего события значительно повышает эффективность рендеринга для сцен, содержащих источники света малого размера, и делает возможным рендеринг сцен с источниками света, чей размер в одном из измерений равен нулю. Например, рендеринг сцен, освещаемых точечными или параллельными источниками света, возможен только в случае, если алгоритм использует оценку следующего события. Кроме того, оценка следующего события в алгоритме рендеринга позволяет повысить эффективность вычислений для большинства моделей сцен, даже если источники света в них имеют достаточно большой размер. В данном случае определяющими являются оптические свойства поверхностей сцены, которые должны иметь плавную ДФР без ярко выраженных пиков.

Важную роль играет многократная выборка по значимости. Этот подход позволяет объединить сильные стороны двух предыдущих решений. В каждой точке диффузной трассы луча разыгрываются два способа оценки яркости. Первый способ основан на выборке по значимости трассы луча и ориентирован на сбор яркости от протяженных источников света большого размера поверхностями с узконаправленной ДФР. Второй способ нацелен на оценку следующего события и позволяет собрать яркость источников света малого размера или поверхностей, имеющих гладкую ДФР. Объединение результатов двух оценок с весами, пропорциональными PDF, приведенных к пространству

направлений луча, позволяет значительно снизить дисперсию и повысить эффективность рендеринга. Кроме того, добавление главной части в подынтегральное выражение также позволяет существенно снизить дисперсию.

Была предложена специальная структура данных, которая хранит распределение яркости поля по сцене. Эта пятимерная структура позволяет построить PDF в каждой точке сцены и сформировать трассы лучей по направлению максимальной яркости в каждой точке сцены. Данная стратегия превосходит стратегию выбора направления рассеивания по значимости ДФР в точке встречи луча с поверхностью. В отличие от выбора по значимости ДФР, которая является локальной, выбор по значимости яркости поля выполняется на основе глобальной информации об источниках яркости в сцене, что позволяет направлять лучи в сторону, приносящую больший вклад в яркость изображения. Эта стратегия может легко сочетаться с уже рассмотренными стратегиями расчета яркости и светового потока точек изображения. Единственным недостатком данной стратегии является большой объем данных, необходимый для хранения пятимерной структуры высокого разрешения.

## **5.9 Вопросы для самоконтроля усвоенных знаний**

Следующие вопросы позволят оценить степень усвоения знаний по разделу применение методов Монте-Карло для решения уравнения рендеринга – трассировка путей.

1. Каким образом вычисляется видимая яркость в базовом алгоритме метода трассировки путей. Какова эффективность данного метода. Будет ли алгоритм работать с моделями точечного и параллельного источников света.
2. Написать псевдокод программы базового метода рендеринга, использующего алгоритм трассировки путей.
3. Каким образом стратификационная выборка может использоваться для ускорения сходимости алгоритмов трассировки путей.
4. Как оптические свойства поверхности (ДФР) используются для алгоритмов выборки по значимости. Что в процессе рендеринга позволяет улучшить этот алгоритм, является ли этот алгоритм оптимальным с точки зрения сходимости процесса рендеринга. Обоснуйте ответ.
5. Какая должна быть оптимальная стратегия выборки по значимости направления луча, рассеянного на поверхности, для ускорения процесса сходимости рендеринга.

6. Имеет ли смысл формировать группу рассеянных лучей (более одного) от одного падающего луча в методе трассировки путей с выборкой по значимости. Обоснуйте ответ.
7. Написать псевдокод программы базового метода рендеринга, использующего алгоритм трассировки путей с выборкой по значимости направлений рассеянных лучей.
8. В чем заключается принцип оценки следующего события, используемого в методе трассировки путей. Может ли данный метод позволить вычислить яркость каустического освещения. Какие модели источников света этот метод может поддерживать.
9. Имеет ли смысл выбирать более одного источника света для оценки яркости прямого освещения в методе трассировки путей с оценкой следующего события. Обоснуйте ответ.
10. Напишите псевдокод программы базового метода рендеринга, использующего алгоритм обратной трассировки путей с оценкой следующего события при расчете яркости прямого освещения.
11. В чем преимущество метода многократной выборки по значимости для оценки яркости прямого освещения в алгоритме трассировки путей с оценкой следующего события. Может ли этот метод вычислить яркость каустического освещения, какие ограничения у этого метода.
12. Напишите псевдокод программы базового метода рендеринга, использующего алгоритм трассировки путей с многократной выборкой по значимости направлений рассеянных лучей.

## 6. Применение методов Монте-Карло к уравнению потенциала – трассировка лучей

### 6.1 Базовый алгоритм

Второй подход к решению проблемы глобального освещения начинается с выражения потока излучения в терминах потенциала (2.18) и уравнения потенциала (2.20):

$$\Phi = \int_A \int_{\Omega_x} L_e(x, \Theta_x) W_e(x, \Theta_x) |\Theta_x \cdot N_x| d\mu_x d\omega_x, \quad (2.18)$$

$$\begin{aligned} W(x, \Theta_x) &= W_e(x, \Theta_x) + W_r(x, \Theta_x) \\ &= W_e(x, \Theta_x) + \int_{\Omega_y} W(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y. \end{aligned} \quad (2.20)$$

Как и ранее, задача заключается в вычислении потока или среднего значения яркости для каждой из точек изображения. Выборка стохастических переменных  $x_0$  и  $\Theta_0$  в соответствии с некоторой PDF  $p_0(x_0, \Theta_{x_0})$  по всем точкам поверхности и направлениям по полусфере дает следующую первичную оценку для интеграла (2.18):

$$\langle \Phi \rangle_{light} = \frac{L_e(x_0, \Theta_{x_0}) | \Theta_{x_0} \cdot N_{x_0} |}{p_0(x_0, \Theta_{x_0})} \langle W(x_0, \Theta_{x_0}) \rangle_{light}. \quad (6.1)$$

В этом случае функция потенциала неизвестна, но ее конкретное значение  $W(x_0, \Theta_{x_0})$  можно оценить, если рассмотреть уравнение потенциала (2.20) как бесконечно рекурсивный интеграл. Применяя русскую рулетку с вероятностью  $R_i$  на каждом  $i$ -ом уровне рекурсии и выбирая направление  $\Theta_{x_i}$  в соответствии с PDF  $p_i(\Theta_{x_i})$  по полусфере, можно получить следующую первичную оценку:

$$\langle \Phi \rangle_{light} = \frac{L_e(x_0, \Theta_{x_0}) | \Theta_{x_0} \cdot N_{x_0} |}{p_0(x_0, \Theta_{x_0})} \cdot \sum_{i=0}^l \left[ \prod_{j=1}^l \frac{f_r(x_j, \Theta_{x_{j-1}}, \Theta_{x_j}) | \Theta_{x_j} \cdot N_{x_j} |}{R_j p_j(\Theta_{x_j})} \right] W_e(x_i, \Theta_{x_i}), \quad (6.2)$$

где  $l$  – число случайных переотражений луча в точках трассы  $x_0, \dots, x_{l+1}$ . Каждая точка  $x_{i+1} = r(x_i, \Theta_{x_i})$  находится методом трассировки лучей, см. рис. 6.1. Общий алгоритм представлен на рис. 6.2. Этот алгоритм практически полностью повторяет алгоритм простой трассировки путей. Однако имеется ряд отличий. Одно из отличий заключается в том, что лучи стартуют от источников света, рассеиваются на поверхностях сцены и снова распространяются в сцене. Каждый раз, когда световые лучи проходят через область точки изображения (ячейки приемника детектора), в эту точку добавляется вклад в виде светового потока, см рис. 5.5. Этот алгоритм можно назвать алгоритмом световой съемки.

Другие отличия заключаются в том, как происходит трассировка лучей. Во-первых, случайные трасы независимы от точек на изображении или ячейках приемника излучения. Поэтому вместо того, чтобы привязывать каждый луч к точке изображения, можно «привязать» каждый луч ко всем точкам изображения (ячейкам детекторов) одновременно. Вклад может быть добавлен к оценке любой из точек изображения, через которые проходит луч. При этом с формальной точки зрения ячейка изображения может быть чем угодно. Это может быть точка обычного изображения, или точка изображения, формируемого линзовой камерой, ячейка детектора, собирающего свет в определенной пространственной области внутри заданного телесного угла, сама сцена, где световой поток накапливается на ее геометрических объектах в виде, например, треугольной сетки или фотонной карты, и т.п.

Другое важное отличие метода трассировки лучей от метода трассировки путей заключается в скоростях их сходимости. Все определяется мерой областей, где потенциал самоизлучения и яркость самосвечения отличаются от нуля. Если рассмотреть обычное изображение (или даже изображение, формируемой линзовой камерой с ненулевой апертурой), то мера областей, где потенциал самоизлучения отличается от нуля, обычно намного меньше, чем мера областей, где яркость самосвечения отличается от нуля. Интуитивно это понятно: точка изображения покрывает гораздо меньше точек и направлений, чем точки и направления на реальных источниках света. Таким образом, эффект попадания или промаха луча в точку изображения очень неэффективен с точки зрения процесса сходимости. Большинство случайных трасс лучей вообще не влияют ни на одну из точек изображения. Если модель камеры не содержит реальную апертуру, то вероятность попадания в точку изображения равна нулю. Таким образом, этот алгоритм совершенно бесполезен для вычисления световых потоков точек изображения, если апертура линзовой камеры не слишком велика. В работе для оценки видимости дороги в условиях тумана [53] был применен алгоритм отслеживания световых лучей. Чтобы получить приемлемые результаты, зрачок глаза имел радиус 0.8м.

Основное преимущество метода трассировки световых лучей появляется при расчете глобального освещения сцены, когда необходимо рассчитать распределение светового потока по элементам сцены. В работах [54-59] используется данный алгоритм для расчета светового потока, покидающего сцену, представленную в виде конечных элементов. Поскольку конечные элементы покрывают всю сцену и собирают излучение в пределах полусферы, световой луч имеет большой шанс оставить свой след (в виде светового потока, переносимого лучом) на приемнике излучения такого рода. Проблемы могут возникнуть, если элементы сцены (локальные детекторы, патчи, и т.п.) имеют слишком высокое пространственное или угловое разрешение. Однако этот подход позволяет вычислить глобальное распределение освещенности или яркости поля, что будет полезно для улучшения работы алгоритмов выборки по значимости, рассмотренных в разделе 5.7.

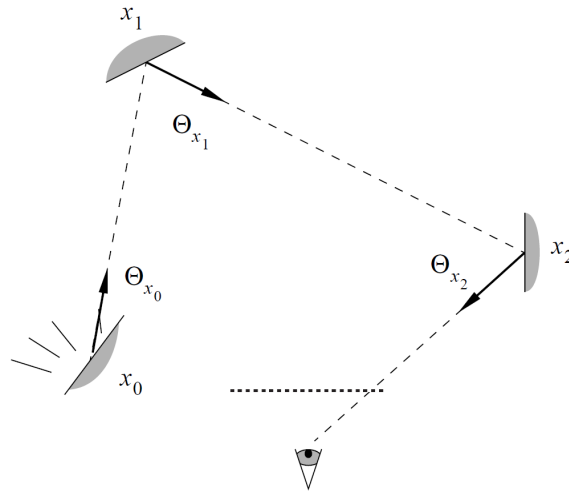


Рис. 6.1. Символы, используемые в алгоритме световых лучей

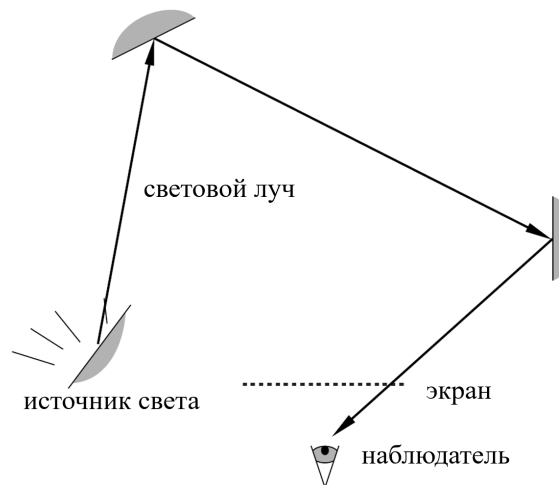


Рис. 6.2. Обзор алгоритма трассировки световых лучей. Необходимо оценить поток излучения, проходящий через каждый пиксель. Трасса первичного луча от источника света соответствует выборке выражения для потока в терминах потенциала. Дальнейшая случайная трасса луча в сцене соответствует рекурсивной оценке значений потенциала. Каждый раз, когда световой луч проходит через пиксель, он добавляет вклад к оценке его светового потока

## 6.2 Выборка по значимости

Данный алгоритм полностью аналогичен алгоритму, построенному на уравнении (5.3). PDF для выборки стохастических переменных  $x_0$  и  $\Theta_{x_0}$  в алгоритме трассировки световых лучей равна:

$$p_0(x, \Theta_x) = \frac{L_e(x, \Theta_x) |x, \Theta_x|}{\Phi}, \quad (6.3)$$

где  $\Phi$  - тот же коэффициент нормализации, что и в уравнении (5.9), т.е. общий световой поток самосветящейся поверхности. Полученная PDF гарантирует, что точки и направления выбраны пропорционально их собственному излучению. Это означает, что большее число случайных трасс лучей начнется от ярких источников света и в направлениях их максимальной яркости, т.е. в областях, которые создают наиболее значимое освещение сцены.

Для источников света, чей размер в одном из измерений равен нулю, выражение (6.3) несколько меняет вид, см. выражения (5.11) и (5.13). Так, для точечного источника света PDF зависит только от направления и будет иметь следующий вид:

$$p_0(\Theta_x) = \frac{I(\Theta_x)}{\Phi}, \quad (6.4)$$

где  $I(\Theta_x)$  – интенсивность (сила света источника света). Для параллельного источника света PDF зависит только от координат на плоскости, освещающей область сцены, и будет иметь следующий вид:

$$p_0(x) = \frac{E(x)}{\Phi}, \quad (6.5)$$

где  $E(x)$  – освещенность (плотность светового потока) области сцены по направлению  $\Theta_x$ .

Аналогично выражению (5.5) случайные переменные  $\Theta_{x_i}$ , которые определяют остальную часть трассы светового луча, могут быть выбраны в соответствии с ДФР, умноженной на косинусный коэффициент. Формально PDF равна:

$$R_i p_i(\Theta_x) = f_r(x_i, \Theta_{x_i}, \Theta_{x_{i-1}}) |\Theta_{x_i} \cdot N_{x_i}|. \quad (6.6)$$

В результате первичная оценка светового потока (6.1) приобретает вид, аналогичный выражению (5.6):

$$\langle \Phi \rangle_{light, imp} = \Phi \sum_{i=0}^l W_e(x_i, \Theta_{x_i}). \quad (6.7)$$

В работах [60] и [61] было предложено дальнейшее усовершенствование методов определения значимости источников света и направлений рассеянного



излучения соответственно. Основные идеи этих методов повторяют идеи, изложенные в разделе 5.7. Оптимизационное решение нацелено на то, чтобы посылать большее число лучей в направлениях с высоким потенциалом. Необходимая информация о потенциале собирается и сохраняется во время расчета. Однако схема, предложенная в этих работах, базируется на других типах структур данных и не учитывает ДФР при выборе значимости.

### 6.3 Оценка следующего события

В работе [62] и разделе 5.4 было показано, каким образом оценка следующего события в базовом алгоритме рендеринга позволяет кардинально улучшить качество синтезируемого изображения. Аналогичное решение может быть применено и к алгоритму трассировки световых лучей, но на этот раз в качестве основы используется уравнение потенциала (2.20):

$$\begin{aligned} W(x, \Theta_x) &= W_e(x, \Theta_x) + \int_{\Omega_y} W(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y, \\ &= W_e(x, \Theta_x) + W_r(x, \Theta_x) \end{aligned}$$

где последний член соответствует потенциалу переотраженного потенциала падающего света.

$$\begin{aligned} W_r(x, \Theta_x) &= \int_{\Omega_y} W(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y \\ &= \int_{\Omega_y} [W_e(y, \Theta_y) + W_r(y, \Theta_y)] f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y \\ &= \int_{\Omega_y} W_e(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y \\ &\quad + \int_{\Omega_y} W_r(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y, \\ &= \int_{A_{apert}} W_e(y, \Theta_{y \rightarrow z}) f_r(y, \Theta_x, \Theta_{y \rightarrow z}) \frac{|\Theta_{y \rightarrow z} \cdot N_y| |\Theta_{y \rightarrow z} \cdot V|}{|y-z|^2} d\mu_z \\ &\quad + \int_{\Omega_y} W_r(y, \Theta_y) f_r(y, \Theta_x, \Theta_y) |\Theta_y \cdot N_y| d\omega_y \end{aligned}$$

где  $V$  – общее направление визирования камеры. Для простой модели камеры точка  $z$  в первом интеграле может быть выбрана равномерно по ее апертуре  $A_{apert}$ :

$$p(z) = \frac{1}{A_{apert}}. \quad (6.8)$$

Для идеальной модели камеры-обскура точка  $z$  всегда является отверстием бесконечно малого размера. Как и при простой трассировке световых лучей,

каждая случайная трасса луча используется для оценки яркости или светового потока для всех точек изображения одновременно. Определить, к какой точке изображения относится сегмент луча или теневой луч, как правило, несложно.

Если для всех остальных интегралов выборка осуществляется в соответствии с PDF (6.3), (6.4), (6.5) и PDF с русской рулеткой (6.6), оценка интеграла (6.7) примет вид:

$$\langle \Phi \rangle_{light,imp} = \Phi [W_e(x_0, \Theta_{x_0}) + A_{apert} \sum_{i=0}^l W_e(x_{i+1}, \Theta_{x_{i+1} \rightarrow z}) \cdot f_r(x_{i+1}, \Theta_{x_i}, \Theta_{x_{i+1} \rightarrow z}) v(x_{i+1}, z) \frac{|\Theta_{x_{i+1} \rightarrow z} \cdot N_{x_{i+1}}| |\Theta_{x_{i+1} \rightarrow z} \cdot V|}{|x_{i+1} - z|^2}]. \quad (6.9)$$

На рис. 6.3 представлены символы, используемые в приведенных выше выражениях. Алгоритм данного метода показан на рис. 6.4. Он является зеркальной копией алгоритма трассировки пути с оценкой следующего события. Оценка следующего события устраняет эффект выборки "попадание или промах" в точку изображения и снижает его шум. Однако эффект все еще сохраняется, поскольку число вкладов, добавляемых к каждой точке изображения, случайно. Множитель ; в оценке светового потока более безопасен, чем в методе трассировки путей, поскольку, как правило, расстояние от камеры до геометрии сцены ограничено снизу, и этот множитель не может стать бесконечно большим.

Данный алгоритм хорошо решает проблемы визуализации каустического освещения (освещения через зеркальные поверхности) от источников света малого размера. Эти эффекты, как правило, не могут быть визуализированы никакими другими методами. Однако для всех методов трассировки путей и световых лучей остается одна главная проблема – визуализация одновременных эффектов каустического освещения и наблюдения, например, визуализация дна бассейна, освещенного источником света малого размера.

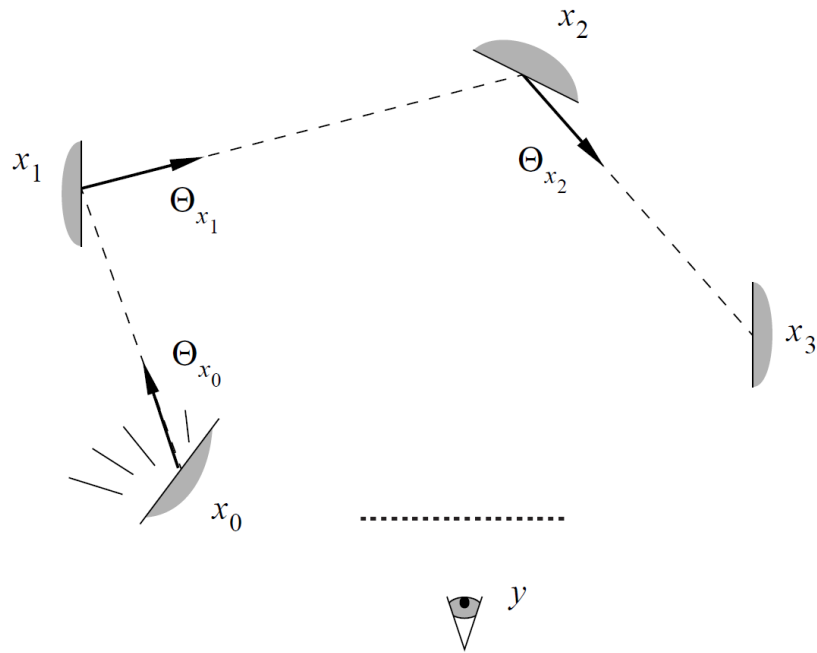


Рис. 6.3. Символы, используемые в алгоритме трассировки световых лучей с использованием оценки следующего события. Трасса случайного луча остается такой же, как и без оценки следующего события, см. рис. 6.1, но теперь выборка дополнительной точки для оценки светового потока производится в соответствии с апертурой камеры

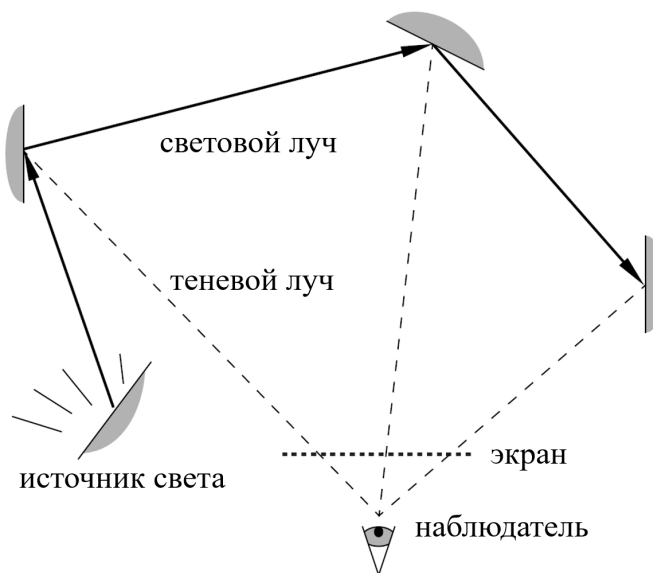


Рис. 6.4. Схематический обзор алгоритма трассировки световых лучей с оценкой следующего события. Прямые вклады потенциала вычисляются в каждой диффузной точке случайной трассы луча путем выбора соответствующих пикселей (если они существуют)

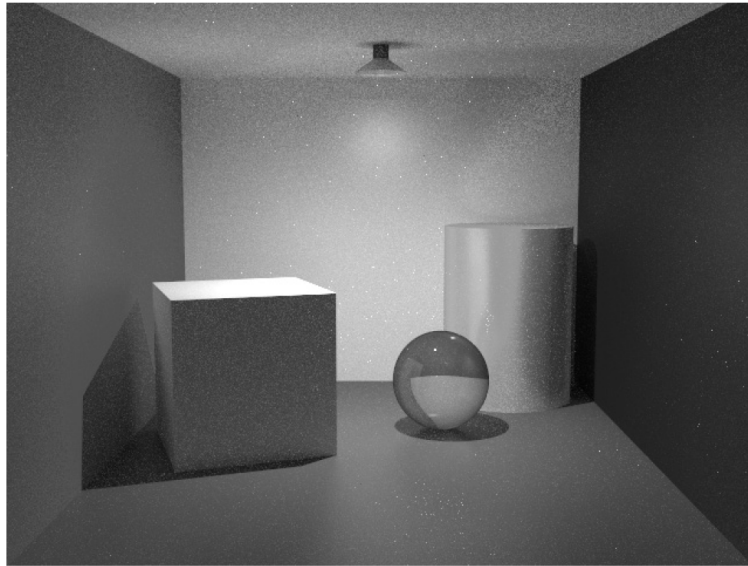


Рис. 6.5. Пример сцены, представленной на рис. 5.10. Рендеринг выполнен методом трассировки световых лучей с оценкой следующего события. Алгоритм трассировки световых лучей хорошо подходит для визуализации каустического освещения от источников света небольшого размера. Качественный рендеринг эффектов, отличающихся от каустического освещения, потребовал бы значительно большее время, чем алгоритмы, основанные на трассировке путей

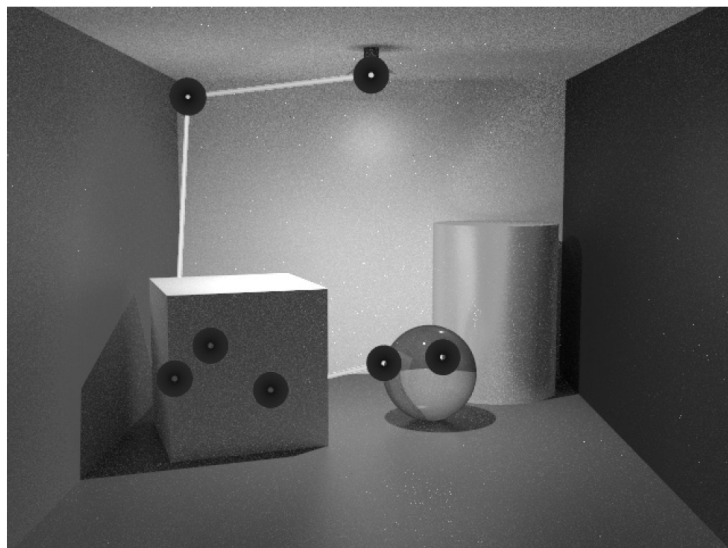


Рис. 6.6. Пример случайной трассы луча, начинающегося от источника света. Из каждой точки пересечения луча с поверхностью сцены в камеру запускается теневой луч, и в случае отсутствия затенения его вклад добавляется к соответствующей точке изображения

## 6.4 Заключение

В этой главе был продемонстрирован другой подход к решению проблемы глобального освещения. Этот подход базируется не на методах трассировки путей, а на его зеркальном аналоге – трассировке световых лучей. Точно так же алгоритмы рендеринга по методу Монте-Карло, являющиеся ничем иным, как стандартным подходом Монте-Карло, описанным в главе 4, были применены к различным математическим моделям для решения проблемы глобального освещения, описанной в главе 2. Как и в методе трассировки путей, различные методы оценки интегралов глобального освещения методом Монте-Карло приводят к появлению совершенно разных алгоритмов методов трассировки световых лучей.

Уравнение потенциала, сопряженное с уравнением рендеринга, естественным образом приводит к алгоритму трассировки световых лучей, который в некотором роде (базируясь на обратимости Гельмгольца) является зеркальной копией алгоритма трассировки путей. Исходя из выражения для светового потока в терминах потенциала, алгоритм выбирает точки и направления на источниках света, т.е. формирует исходные лучи на источниках света. Следующим шагом алгоритм рекурсивно оценивает интегральное уравнение путем выборки направлений отражения и формирования соответствующих световых лучей. Каждый раз, когда световой луч проходит через точку изображения или какой-либо другой детектор, к оценке светового потока детектора добавляется вклад в виде светового потока луча. Выборка по значимости позволяет направить луч в сторону максимальной ДФР, что в общем случае позволяет сделать работу алгоритма оценки светового потока более эффективной. Русская рулетка, как и в случае трассировки путей, позволяет прервать бесконечную рекурсию трассы светового луча.

Алгоритм трассировки световых путей, даже с учетом выборки по значимости ДФР, не может работать с наблюдателями, имеющими малую апертуру. Вероятность попадания луча в зрачок камеры обскура или модели глаза человека практически равна нулю. Поэтому данный алгоритм применим для расчета освещенности на элементах большого размера, например, на треугольниках геометрии сцены или на специальных частях (патчах). Для рендеринга изображений, требующих расчета яркости (светового потока) в точках изображения камеры с малой апертурой, алгоритм рендеринга расширяется методом оценки следующего события. В этом случае в каждой диффузной точке трассы светового луча формируются теневые лучи, соединяющие точку трассы луча со всеми приемниками излучения сцены. Далее световой поток, рассеянный на элементе сцены, передается всем приемникам излучения. Этот алгоритм

позволяет существенно снизить дисперсию изображения, хотя в большинстве случаев она остается выше, чем в методе трассировки путей. Значимым преимуществом данного метода является возможность расчета яркости каустического освещения от источников света малого размера, что практически невозможно в алгоритмах, основанных на трассировке путей.

## 6.5 Вопросы для самоконтроля усвоенных знаний

Следующие вопросы позволят оценить степень усвоения знаний по разделу применение методов Монте-Карло к уравнению потенциала – трассировка лучей.

1. В чем отличие метода трассировки лучей от метода трассировки путей с точки зрения накопления значений яркости (светового потока) в точках изображения сцены.
2. В чем заключается проблема базовых методов прямой трассировки лучей с точки зрения скорости сходимости методов рендеринга.
3. Могут ли базовые методы трассировки лучей, включая методы трассировки лучей с выборкой по значимости, использовать точечную камеру. Какие изменения в алгоритмы трассировки лучей необходимо внести для работы с точечной камерой.
4. Написать псевдокод базового алгоритма программы расчета яркости изображения методом трассировки лучей.
5. Какую физическую величину позволяют вычислить методы трассировки лучей (базовый и с выборкой по значимости) при использовании линзовой камеры. Возможен ли расчет яркости, какие ограничения имеют эти методы. Обоснуйте ответ.
6. Можно ли обеспечить совпадение результатов расчета, полученных методами прямой трассировки лучей и трассировки путей. Если да, то какое необходимое условие, накладываемое на модель ДФР, должно быть выполнено.
7. Возможен ли расчет яркости каустического освещения методами трассировки лучей. Какие ограничения возникают при расчете яркости каустического освещения.
8. Написать псевдокод программы расчета видимой освещенности сцены методом трассировки лучей с выборкой по значимости.
9. Написать псевдокод программы расчета яркости изображения сцены методом трассировки лучей с оценкой следующего события.

## 7. Литература

1. Goral C. M. et al. Modeling the interaction of light between diffuse surfaces // ACM SIGGRAPH computer graphics. – 1984. – Vol. 18. – No. 3. – Pp. 213-222.
2. Kajiya J. T. The rendering equation // Proceedings of the 13th annual conference on Computer graphics and interactive techniques. – 1986. – Pp. 143-150.
3. Lange B. The simulation of radiant light transfer with stochastic ray-tracing // Photorealistic Rendering in Computer Graphics: Proceedings of the Second Eurographics Workshop on Rendering. – Berlin, Heidelberg : Springer Berlin Heidelberg, 1994. – Pp. 30-44.
4. Jensen H. W. Global illumination using photon maps // Rendering Techniques' 96: Proceedings of the Eurographics Workshop in Porto, Portugal, June 17–19, 1996 7. – Springer Vienna, 1996. – Pp. 21-30.
5. Jensen H. W. Realistic image synthesis using photon mapping. – Natick : Ak Peters, 2001. – 364p.
6. Nachisuka T., Jensen H. W. Stochastic progressive photon mapping // ACM SIGGRAPH Asia 2009 papers. – 2009. – Pp. 1-8.
7. Будаков В.П., Макаров Д.Н. Компьютерная графика с приложением в светодизайн: Учебник для высших учебных заведений. – Москва, 2022. – 264 с.
8. Pharr M., Jakob W., Humphreys G. Physically based rendering: From theory to implementation. – MIT Press, 2023.
9. Cook R. L., Torrance K. E. A reflectance model for computer graphics // ACM Siggraph Computer Graphics. – 1981. – Vol. 15. – Np. 3. – Pp. 307-316.
10. Cook R. L., Torrance K. E. A reflectance model for computer graphics // ACM Transactions on Graphics (ToG). – 1982. – Vol. 1. – No. 1. – Pp. 7-24.
11. de Berg M. T. Efficient algorithms for ray shooting and hidden surface removal: Efficiënte algoritmen voor ray shooting en hidden surface removal. – 1992.
12. Ward G. J. Measuring and modeling anisotropic reflection // Proceedings of the 19th annual conference on Computer graphics and interactive techniques. – 1992. – Pp. 265-272.
13. Schlick C. A customizable reflectance model for everyday rendering // Fourth Eurographics Workshop on Rendering. – 1993. – Pp. 73-83.
14. Fournier A. Separating reflection functions for linear radiosity // Rendering Techniques' 95: Proceedings of the Eurographics Workshop in Dublin, Ireland, June 12–14, 1995 6. – Springer Vienna, 1995. – Pp. 296-305.
15. Schlick C. A survey of shading and reflectance models // Computer Graphics Forum. – Edinburgh, UK : Blackwell Science Ltd, 1994. – Vol. 13. – No. 2. – Pp. 121-131.
16. Lewis R. R. Making shaders more physically plausible // Proceedings of the Fourth Eurographics Workshop on Rendering. – Paris, France, 1993. – Pp. 47-62.

17. Lewis R. R. Making shaders more physically plausible // Computer Graphics Forum. – Edinburgh, UK : Blackwell Science Ltd, 1994. – Vol. 13. – No. 2. – Pp. 109-120.
18. Arvo J. The role of functional analysis in global illumination // Rendering Techniques' 95: Proceedings of the Eurographics Workshop in Dublin, Ireland, June 12–14, 1995 6. – Springer Vienna, 1995. – Pp. 115-126.
19. Gershbein R., Schröder P., Hanrahan P. Textures and radiosity: Controlling emission and reflection with texture maps // Proceedings of the 21st annual conference on Computer graphics and interactive techniques. – 1994. – Pp. 51-58.
20. Lafortune E. P., Willems Y. D. A theoretical framework for physically based rendering // Computer Graphics Forum. – Edinburgh, UK : Blackwell Science Ltd, 1994. – Vol. 13. – No. 2. – Pp. 97-107.
21. Hammersley J. Monte carlo methods. – Springer Science & Business Media, 2013.
22. Kalos M. H., Whitlock P. A. Monte carlo methods. – John Wiley & Sons, 2009.
23. Shreider Y. A. (ed.). The Monte Carlo method: the method of statistical trials. – Elsevier, 2014. – Vol. 87.
24. Ermakow S. M. Die Monte-Carlo-Methode und verwandte Fragen. – Berlin: VEB Deutscher Verlag der Wissenschaften, 1975.
25. Davis P. J., Rabinowitz P. Methods of numerical integration. – Courier Corporation, 2007.
26. Knuth D. E. The Art of Computer Programming: Fundamental Algorithms, volume 1. – Addison-Wesley Professional, 1997.
27. Zaremba S. K. The mathematical basis of Monte Carlo and quasi-Monte Carlo methods // SIAM review. – 1968. – Vol. 10. – No. 3. – Pp. 303-314.
28. Spanier J. A new family of estimators for random walk problems // IMA Journal of Applied Mathematics. – 1979. – Vol. 23. – No. 1. – Pp. 1-31.
29. Spanier J., Maize E. H. Quasi-random methods for estimating integrals using relatively small samples // SIAM review. – 1994. – Vol. 36. – No. 1. – Pp. 18-44.
30. Shirley P. Nonuniform random point sets via warping // Graphics Gems III (IBM Version). – Morgan Kaufmann, 1992. – Pp. 80-83.
31. Shirley P., Wang C., Zimmerman K. Monte carlo techniques for direct lighting calculations // ACM Transactions on Graphics (TOG). – 1996. – Vol. 15. – No. 1. – Pp. 1-36.
32. Veach E., Guibas L. Bidirectional estimators for light transport // Photorealistic Rendering Techniques. – Berlin, Heidelberg : Springer Berlin Heidelberg, 1995. – Pp. 145-167.
33. Veach E., Guibas L. J. Optimally combining sampling techniques for Monte Carlo rendering // Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. – 1995. – Pp. 419-428.



34. Halton J. H. On the relative merits of correlated and importance sampling for Monte Carlo integration // *Mathematical Proceedings of the Cambridge Philosophical Society*. – Cambridge University Press, 1965. – Vol. 61. – No. 2. – Pp. 497-498.
35. Halton J. H. A retrospective and prospective survey of the Monte Carlo method // *Siam review*. – 1970. – Vol. 12. – No. 1. – Pp. 1-63.
36. Hammersley J. M. Conditional monte carlo // *Journal of the ACM (JACM)*. – 1956. – T. 3. – No. 2. – Pp. 73-76.
37. Whitted T. An improved illumination model for shaded display // *ACM Siggraph 2005 Courses*. – 2005. – Pp. 4-es.
38. Cook R. L., Porter T., Carpenter L. Distributed ray tracing // *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*. – 1984. – Pp. 137-145.
39. Cook R. L. Stochastic sampling in computer graphics // *ACM Transactions on Graphics (TOG)*. – 1986. – Vol. 5. – No. 1. – Pp. 51-72.
40. Ward G. J., Rubinstein F. M., Clear R. D. A ray tracing solution for diffuse interreflection // *ACM SIGGRAPH 2007 courses*. – 2007. – Pp. 11-es.
41. Shirley P. S. *Physically based lighting calculations for computer graphics*. – University of Illinois at Urbana-Champaign, 1991.
42. Schlick C. An adaptive sampling technique for multidimensional integration by ray-tracing // *Photorealistic Rendering in Computer Graphics: Proceedings of the Second Eurographics Workshop on Rendering*. – Berlin, Heidelberg : Springer Berlin Heidelberg, 1994. – Pp. 21-29.
43. Lange B. The simulation of radiant light transfer with stochastic ray-tracing // *Photorealistic Rendering in Computer Graphics: Proceedings of the Second Eurographics Workshop on Rendering*. – Berlin, Heidelberg : Springer Berlin Heidelberg, 1994. – Pp. 30-44.
44. Lafortune E. P., Willems Y. D. Using the modified phong BRDF for physically based rendering // *Dept. Comput. Sci., Katholieke Universiteit Leuven, Leuven, Belgium, Tech. Rep. CW197*. – 1994.
45. Lafortune E. P., Willems Y. D. The ambient term as a variance reducing technique for Monte Carlo ray tracing // *Photorealistic Rendering Techniques*. – Berlin, Heidelberg : Springer Berlin Heidelberg, 1994. – Pp. 168-176.
46. Immel D. S., Cohen M. F., Greenberg D. P. A radiosity method for non-diffuse environments // *Acm Siggraph Computer Graphics*. – 1986. – Vol. 20. – No. 4. – Pp. 133-142.
47. Sillion F. X. et al. A global illumination solution for general reflectance distributions // *Proceedings of the 18th annual conference on Computer graphics and interactive techniques*. – 1991. – Pp. 187-196.
48. Bentley, J. L. *Data structures for range searching* / J.L. Bentley, J.H. Friedman // *ACM Computing Surveys (CSUR)*. – 1979. – Vol. 11. – No. 4. – Pp. 397-409.

49. Chen, Q. P. Using the k-d tree data structure to accelerate Monte Carlo simulations / Q.P. Chen, B. Xue, J.I. Siepmann // *Journal of Chemical Theory and Computation*. – 2017. – Vol. 13. – No. 4. – Pp. 1556-1565.
50. Hunt, W. Fast kd-tree construction with an adaptive error-bounded heuristic / W. Hunt, W.R. Mark, G. Stoll // *2006 IEEE Symposium on Interactive Ray Tracing*. – IEEE, 2006. – Pp. 81-88.
51. Жданов Д. Д. и др. Использование многоуровневых хэш-таблиц для ускорения процесса рендеринга. / Жданов Д. Д., Лысых А. И., Халимов Р. Р., Кинев И. Е., Жданов А. Д. // *Программирование*. – 2023. – №. 3. – С. 37-48.
52. Jensen H. W. Importance driven path tracing using the photon map // *Rendering Techniques' 95: Proceedings of the Eurographics Workshop in Dublin, Ireland, June 12–14, 1995* 6. – Springer Vienna, 1995. – Pp. 326-335.
53. Rozé C. et al. Evaluations of the sighting distance in a foggy atmosphere by Monte Carlo simulation // *Atmospheric Environment*. – 1994. – Vol. 28. – No. 5. – Pp. 769-775.
54. Pattanaik S. N. The Mathematical Framework of Adjoint Equations for Illumination Computation // *ICCG*. – 1993. – Pp. 123-138.
55. Pattanaik S. N. Computational Methods for Global Illumination and Visualization of Complex 3D Environments : PhD thesis. – BITS Pilani, 1993.
56. Pattanaik S. N., Mudur S. P. The potential equation and importance in illumination computations // *Computer Graphics Forum*. – Edinburgh, UK : Blackwell Science Ltd, 1993. – Vol. 12. – No. 2. – Pp. 131-136.
57. Pattanaik S. N., Mudur S. P. Efficient potential equation solutions for global illumination computation // *Computers & Graphics*. – 1993. – Vol. 17. – No. 4. – Pp. 387-396.
58. Pattanaik S., Bouatouch K. Adjoint equations and particle tracing for global illumination // *Publication Interne*. – 1995. – Vol. 903.
59. Pattanaik S. N., Mudur S. P. Adjoint equations and random walks for illumination computation // *ACM Transactions on Graphics (TOG)*. – 1995. – Vol. 14. – No. 1. – Pp. 77-102.
60. Dutre P., Willems Y. D. Importance-driven monte carlo light tracing // *Photorealistic Rendering Techniques*. – Berlin, Heidelberg : Springer Berlin Heidelberg, 1995. – Pp. 188-197.
61. Dutre P., Willems Y. D. Potential-driven Monte Carlo particle tracing for diffuse environments with adaptive probability functions // *Rendering Techniques' 95: Proceedings of the Eurographics Workshop in Dublin, Ireland, June 12–14, 1995* 6. – Springer Vienna, 1995. – Pp. 306-315.
62. Dutré P., Lafortune E. P., Willems Y. D. Monte Carlo light tracing with direct computation of pixel intensities. – 1993.



Жданов Дмитрий Дмитриевич  
Потемин Игорь Станиславович  
Жданов Андрей Дмитриевич

## **Методы расчета глобальной освещенности**

**Учебно-методическое пособие**

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати

Заказ №

Тираж

Отпечатано на ризографе



**Редакционно-издательский отдел**  
**Университета ИТМО**  
197101, Санкт-Петербург, Кронверкский пр., 49, литер А