

Моделирование теней в 3D сценах с помощью каскадных теневых карт в режиме реального времени¹

А.В. Мальцев

Аннотация. В работе предлагаются методы и алгоритмы моделирования теней в протяженных трехмерных виртуальных сценах, таких как обширные местности и ландшафты, освещаемых направленными источниками света. Предлагаемые методы и алгоритмы основаны на применении каскадных теневых карт. Описываются различные методы нахождения положений разделяющих плоскостей: равномерное, логарифмическое и смешанное разбиения. Рассматриваются вопросы построения и использования теневых карт для выделенных областей пирамиды видимости в режиме реального времени.

Ключевые слова: моделирование теней, каскадные теневые карты, шейдерная визуализация, рендеринг в реальном времени.

Введение

Одним из наиболее существенных графических эффектов, повышающих реалистичность визуализируемых трехмерных виртуальных сцен, являются тени. Они также способствуют правильному восприятию виртуальной среды, например, дают визуальную информацию о взаимном расположении объектов, позволяют лучше ощутить объем пространства и т.д. В [1, 2] были описаны методы и алгоритмы моделирования в реальном времени теней для направленных и всенаправленных источников света. Они основывались на использовании теневых карт. Напомним, что под “теневогой картой” для направленного источника понималась двумерная текстура, несущая информацию из буфера глубины ортографической или перспективной камеры, находящейся в точке расположения источника света и смотрящей в направлении свечения. Для всенаправленных

(точечных) источников освещения предлагалось создавать теневые карты, состоящие из двух текстур. Каждая из них генерировалась с помощью параболического отображения, при котором на плоскость двумерной текстуры отображалось целое полупространство.

В том случае, если имеется довольно протяженная виртуальная сцена, например, обширная местность или ландшафт, и требуется осветить все (или значительную часть) ее пространство направленным источником света, описанные в [1, 2] способы моделирования теней будут давать изображения с достаточно сильным алиасингом (эффектом «ступенчатости») тени. Это происходит по нескольким причинам. Во-первых, теневая карта имеет определенное разрешение в пикселах, следовательно, чем больший участок сцены она охватывает, тем меньше становится ее детализация. То есть мелкие, по сравнению с общим размером сцены, детали отображаются на такой тек-

¹ Работа выполняется при поддержке РФФИ, грант № 13-07-00674

стуре в один или несколько текселов, что в свою очередь приводит к очень грубой границе тени от таких объектов или вовсе отсутствию тени. Повысить детализацию можно путем увеличения разрешения текстуры для теневой карты. Однако современные графические адаптеры ограничивают максимальное разрешение текстур. К тому же, время создания теневой карты с большим разрешением для высокополигональной виртуальной сцены не позволит обеспечить визуализацию такой сцены в режиме реального времени. Во-вторых, при просмотре сцены из виртуальной камеры, имеющей перспективную проекцию, алиасинг тени резко возрастает от дальних объектов к ближним, если детализация для них в карте теней одинакова. Это происходит, поскольку объекты, находящиеся около ближней плоскости отсечения камеры, занимают на изображении большее число пикселей, чем те, которые располагаются около дальней. Следовательно, при одной и той же детализации теневой текстуры одному пикселу дальнего объекта будет соответствовать несколько текселов в этой текстуре, а нескольким пикселям ближнего объекта – один тексел текстуры.

Решить данную проблему для направленных источников света типа «прожектор», освещающих параллельными лучами, можно с помощью каскадных теневых карт. Подход основан на том, что для объектов, находящихся на различных расстояниях от наблюдателя, требуются теневые текстуры с различной детализацией. А именно, для близких объектов детализация текстуры должна быть выше, чем для дальних. Реализация такого решения состоит в разбиении пирамиды видимости виртуальной камеры (наблюдателя) плоскостями (параллельными ближней и дальней плоскостям отсечения) на некоторое число усеченных пирамид, для каждой из которых строится собственная теневая карта (Рис. 1). При этом получается, что, во-первых, мы генерируем текстуру теней не для всей сцены, а только для той части, которую видит наблюдатель. Во-вторых, деление на подобласти позволяет обеспечить для каждой из них теневую карту с требуемой детализацией. Соответственно, даже для небольших объектов в ближнем поле зрения наблюдателя будут визуализироваться тени без артефактов.

В данной работе предлагаются методы и алгоритмы построения и использования каскад-

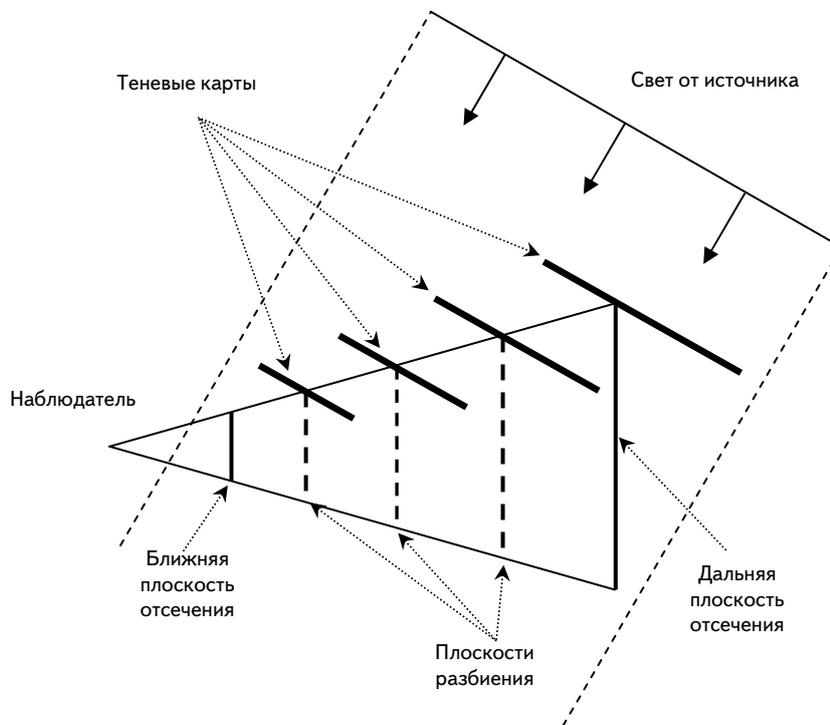


Рис. 1. Каскадные теневые карты

ных теневых карт для направленных источников света типа «прожектор» в режиме реального времени.

1. Разбиение пирамиды видимости на подобласти

Вначале рассмотрим задачу разбиения усеченной пирамиды видимости F_{cam} виртуальной камеры плоскостями, параллельными ближней и дальней плоскостям отсечения, на N частей. Тогда общее число разделяющих плоскостей, включая ближнюю и дальнюю плоскость отсечения, будет равно $N+1$. Возможны различные подходы к решению этой задачи: равномерное, логарифмическое и смешанное разбиения.

При равномерном разбиении F_{cam} разделяющие плоскости располагаются так, чтобы расстояния по оси Z видовой системы координат (СК) VCS между ними были одинаковыми. Тогда расстояние от наблюдателя до i -ой разделяющей плоскости будет равно

$$d_i = d_n + \frac{(d_f - d_n)}{N} i, \quad i \in [0, N], \quad (1)$$

где d_n , d_f – расстояния соответственно до ближней и дальней отсекающих плоскостей. Равномерное разбиение, как показано в [3], отличается сильным увеличением перспективного алиасинга тени на объектах, находящихся близко к наблюдателю.

Логарифмическое разбиение [3] основано на идее достижения константного значения ошибки ρ перспективного алиасинга для объектов, которые находятся в F_{cam} между ближней и дальней плоскостями отсечения. При этом $\rho = \ln(d_f / d_n)$. Тогда i -ая разделяющая плоскость будет находиться на расстоянии

$$d_i = d_n \left(\frac{d_f}{d_n} \right)^{i/N}, \quad i \in [0, N]. \quad (2)$$

Недостатком такого разбиения являются слишком узкие области между ближайшими к наблюдателю разделяющими плоскостями. Они содержат в себе мало объектов, что приводит к снижению эффективности работы алгоритма моделирования теней.

Смешанное разбиение сочетает в себе два предыдущих способа. Для нахождения d_i в этом случае возьмем линейную комбинацию правых частей равенств (1) и (2) с некоторым вещественным весовым коэффициентом $\lambda \in (0, 1)$:

$$d_i = \lambda d_n \left(\frac{d_f}{d_n} \right)^{i/N} + (1 - \lambda) (d_n + (i/N)(d_f - d_n)), \quad i \in [0, N]. \quad (3)$$

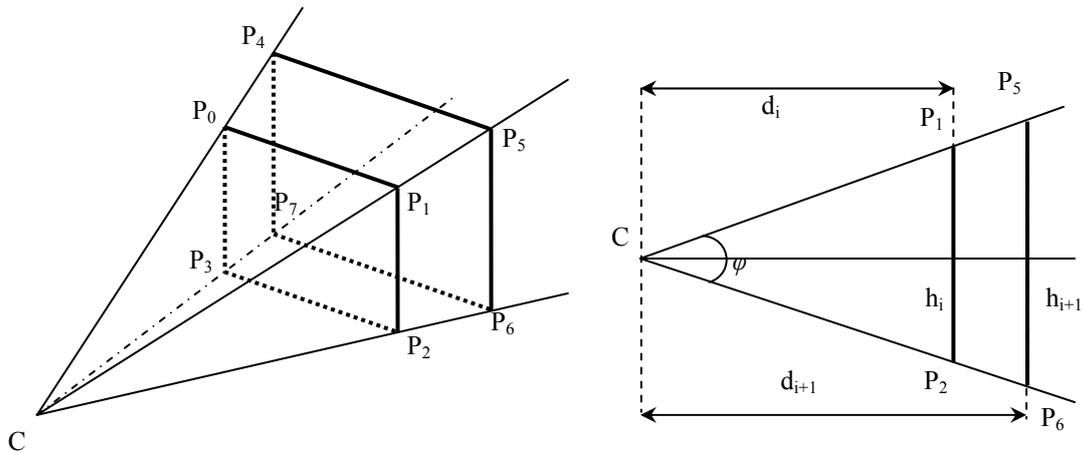
Коэффициент λ для большинства сцен можно задавать равным 0.5. Однако для получения лучшего результата разбиения в некоторых сценах его надо подбирать опытным путем.

Выбирая количество областей N , на которое будет разделена усеченная пирамида видимости F_{cam} , следует придерживаться приемлемого для графического приложения соотношения скорости визуализации и качества получаемых теней. Заметим, что при $N < 3$ результат моделирования теней не будет значительно отличаться от результата с одной теневой картой. При слишком же больших значениях N скорость визуализации виртуальной сцены будет существенно падать, поскольку в этом случае потребуется генерировать множество теневых карт, а, следовательно, и производить множество визуализаций сцены. Для получения качественных теней в режиме реального времени мы рекомендуем устанавливать N в пределах 3-5.

Расстояния до всех плоскостей разбиения, посчитанные по одной из формул (1)-(3), следует запомнить в массиве D расстояний (так, что $D[i] = d_i$), поскольку они потребуются в дальнейшем, как на этапе генерации теневых карт, так и на этапе визуализации сцены с применением этих карт.

2. Подготовка текстур каскадных теневых карт

Найденные ранее разделяющие плоскости разбивают пирамиду видимости наблюдателя на N усеченных пирамид, для каждой из которых необходимо сгенерировать теневую карту. Рассмотрим процесс создания такой карты для одной из усеченных пирамид F_i , образованной разделяющими плоскостями i и $i+1$. Предполо-


 Рис. 2. Определение параметров i -ой усеченной пирамиды

жим, что виртуальная сцена содержит один направленный источник света типа «прожектор».

Разместим в точке расположения источника фиктивную виртуальную камеру с видовой системой координат LVCS и направлением взгляда (ось $-Z$), совпадающим с направлением свечения. Визуализируя с помощью нее трехмерную сцену, мы будем получать в буфере глубины теневую карту. Поскольку рассматриваемый источник освещает параллельными лучами, выберем для созданной фиктивной камеры ортографическое проецирование. Далее необходимо правильно задать параметры этого проецирования.

Вначале вычислим координаты восьми вершин F_i в видовой СК LVCS камеры, привязанной нами к источнику света. Для этого найдем половины длин сторон прямоугольников $P_0P_1P_2P_3$ и $P_4P_5P_6P_7$, являющихся пересечениями плоскостей i и $i+1$ с пирамидой видимости наблюдателя (Рис. 2). Используя известные значения угла φ вертикального раствора камеры и отношения k_{aspect} ширины кадра к его высоте, а также расстояние d_i , найденное в п. 1, получим для $P_0P_1P_2P_3$:

$$h_i = \frac{P_1P_2}{2} = d_i \tan \frac{\varphi}{2}, \quad w_i = \frac{P_0P_1}{2} = h_i k_{aspect}.$$

Аналогично рассчитываются h_{i+1} и w_{i+1} для $P_4P_5P_6P_7$. Тогда искомые координаты вершин F_i в видовой системе VCS наблюдателя будут: $P_0(-w_i, h_i, -d_i)$, $P_1(w_i, h_i, -d_i)$, $P_2(w_i, -h_i, -d_i)$, $P_3(-w_i, -h_i, -d_i)$, $P_4(-w_{i+1}, h_{i+1}, -d_{i+1})$ и т.д. Минус в z -компоненте появляется потому, что взгляд

камеры (наблюдателя) направлен в отрицательном направлении оси Z ее системы координат, а из равенств (1)-(3) было рассчитано абсолютное значение расстояния d_i . Перевод получившихся координат в систему LVCS осуществляется по формуле $P_j' = M_{light} * M_{cam} * P_j$, где M_{light} – матрица преобразования из мировой СК (WCS) в LVCS, M_{cam} – матрица преобразования из СК наблюдателя (VCS) в WCS.

Далее для задания ближней и дальней отсекающих плоскостей области видимости фиктивной камеры, привязанной к источнику света, нужно выбрать из полученных восьми точек в СК LVCS максимальное и минимальное значения z -координат. При этом $|\max P_{j,z}'|$ будет соответствовать ближней отсекающей плоскости, а $|\min P_{j,z}'|$ – дальней (поскольку взгляд камеры направлен вдоль оси $-Z$ ее СК). Однако необходимо учесть, что между полученной ближней отсекающей плоскостью и источником света, могут находиться объекты, невидимые наблюдателю (Рис. 3). Информация о таких объектах должна попасть в создаваемую теневую карту, поскольку они будут отбрасывать тени в область, ограниченную усеченной пирамидой F_i . Следовательно, нужно расширить область видимости камеры, привязанной к источнику, путем смещения ближней плоскости отсечения к источнику до такого значения z_{near} , пока все объекты между источником и усеченной пирамидой F_i не попадут в поле видимости фиктивной камеры.

Остальные плоскости отсечения (левая, правая, верхняя, нижняя) определяются исходя из

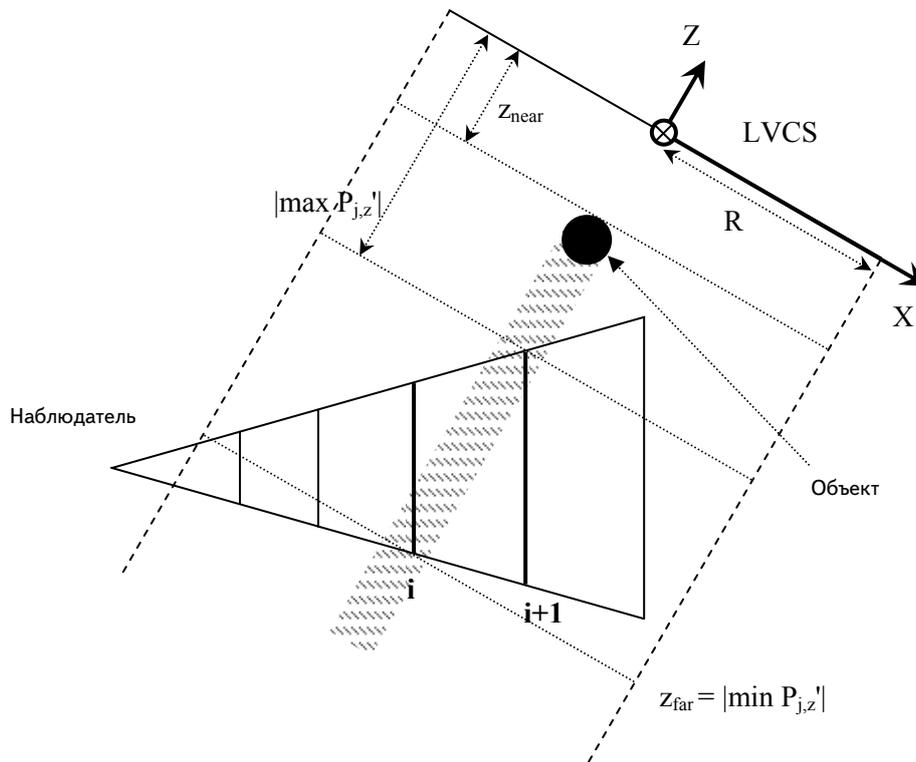


Рис. 3. Коррекция ближней плоскости отсечения для камеры, "привязанной" к источнику света

радиуса R (Рис. 3) цилиндрического светового потока от источника света (подробное описание параметров направленных источников света типа «прожектор» можно найти в [1]). Зная положения отсекающих плоскостей, мы можем получить матрицу $M_{Lprj,i}$ проекционного преобразования для камеры, привязанной к источнику света, например, с помощью функции `glOrtho` графической библиотеки `OpenGL`.

Далее необходимо выделить из спроецированной области участок, охватывающий рассматриваемую область F_i . Для этого умножим рассчитанные ранее в системе `LVCS` вершины P_j' ($j \in [0,7]$) усеченной пирамиды F_i на матрицу $M_{Lprj,i}$, получая точки P_j'' в СК нормализованного объема видимости. Найдем для получившихся точек $x_{\min} = \min P_{j,x}''$, $x_{\max} = \max P_{j,x}''$, $y_{\min} = \min P_{j,y}''$, $y_{\max} = \max P_{j,y}''$. Чтобы перейти от всего объема видимости, задаваемого матрицей $M_{Lprj,i}$, к выделяемому участку применим последовательно преобразования переноса и растяжения пространства, задаваемые соответственно матрицами

$$M_{T,i} = \begin{pmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ и } M_{S,i} = \begin{pmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

где $T_x = -0.5(x_{\max} + x_{\min})$, $T_y = -0.5(y_{\max} + y_{\min})$, $S_x = 2/(x_{\max} - x_{\min})$, $S_y = 2/(y_{\max} - y_{\min})$. Тогда матрица перехода будет иметь вид

$$M_{crop,i} = M_{S,i} M_{T,i} = \begin{pmatrix} S_x & 0 & 0 & S_x T_x \\ 0 & S_y & 0 & S_y T_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

В результате общая матрица проекционного преобразования для усеченной пирамиды F_i будет $M_{proj,i} = M_{crop,i} * M_{Lprj,i}$. Установим ее в качестве текущей проекционной матрицы и визуализируем сцену из фиктивной камеры, привязанной к источнику света, с записью только буфера глубины и отключенным выводом в буфер кадра. Матрицу $M_{proj,i}$ следует сохранить

в памяти, поскольку она потребуется при использовании подготовленной теневой карты.

Итак, мы получили в буфере глубины теневую карту для усеченной пирамиды F_i . Карты для остальных усеченных пирамид создаются аналогичным образом. При генерации теневых карт целесообразно применение так называемого прямого рендеринга в текстуру с использованием современных программно-аппаратных средств, а именно технологии FBO (framebuffer objects [4]). FBO – это расширение OpenGL, которое обеспечивает простой интерфейс для отрисовки в контексты, отличные от буферов GL, предоставленных оконной системой.

Заметим, что в том случае, если виртуальная сцена содержит несколько направленных источников света, то каждый из них будет иметь свою собственную карту теней для каждой из N пирамид.

3. Применение каскадных теневых карт для моделирования теней

Теперь рассмотрим применение полученных теневых карт для моделирования теней. При визуализации виртуальной сцены из позиции наблюдателя во фрагментном шейдере необходимо определить для каждого рассматриваемого фрагмента, к какой из N усеченных пирамид видимости он относится. Для этого сравним расстояние d_{frag} от точки P , соответствующей фрагменту, до наблюдателя (в видовой системе координат VCS) с ранее записанными в массив D расстояниями от наблюдателя до разделяющих плоскостей. Значение d_{frag} определяется по формулам:

$$d_{frag} = |P_{VCS,z}|, P_{VCS} = M_{MV} \cdot P_{OCS},$$

где P_{VCS} , P_{OCS} – координаты точки P в видовой и объектной СК соответственно, M_{MV} – текущая модельно-видовая матрица. Если $D[i] \leq d_{frag} < D[i+1]$, где $i \in [0, N]$, то фрагмент находится в i -ой усеченной пирамиде и для него необходимо использовать соответствующую ей теневую карту. Для проведения теста глубины фрагмента относительно источника найдем координаты P в СК нормализованного объема видимости (LNDCS) для камеры, привязанной к источнику:

$$P_{LNDCS} = M_{proj,i} \cdot M_{light} \cdot M_{model} \cdot P_{OCS},$$

где M_{model} – матрица модельного преобразования, переводящая координаты фрагмента из объектной системы координат (OCS) в мировую (WCS), M_{light} – матрица видового преобразования, переводящая координаты фрагмента из мировой системы координат в видовую систему (LVCS) фиктивной камеры, привязанной к источнику. $M_{proj,i}$ – произведение матриц, задающее проекционное преобразование и сохраненное на этапе построения теневой карты для i -ой усеченной пирамиды. Поскольку координаты точек в системе LNDCS по всем трем осям находятся в отрезке $[-1, 1]$, а для проведения теста затененности фрагмента с использованием теневых карт нужны координаты, соответствующие отрезку $[0, 1]$, выполним преобразование

$$P_{tex} = M_{bias} \cdot P_{LNDCS},$$

$$\text{где } M_{bias} = \begin{pmatrix} 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Пусть $P_{tex} = (x_p, y_p, z_p)$. Заметим, что z_p соответствует расстоянию (глубине) от точки P до источника освещения. В теневой карте записано расстояние от источника освещения до ближайшей точки какого-либо объекта в том же направлении. Поэтому для определения освещенности точки по паре координат (x_p, y_p) из i -ой теневой карты выбирается записанное в ней значение глубины z_{map} и сравнивается со значением z_p . При этом возможны два варианта:

- $z_{map} < z_p$, что соответствует ситуации, когда фрагмент, которому соответствует точка P , закрыт от попадания в него прямых лучей из источника участком поверхности какого-то объекта в сцене (другими словами, фрагмент находится в тени);
- $z_{map} \cong z_p$ (z_p и z_{map} приблизительно эквивалентны), то есть в теневой карте записана с некоторой погрешностью глубина самой точки P , а следовательно, фрагмент освещается источником.

Литература

1. Мальцев А.В., Михайлюк М.В. Моделирование теней в виртуальных сценах с направленными источниками освещения // Информационные технологии и вычислительные системы. – 2010. – № 2.
2. Мальцев А.В., Михайлюк М.В. Реализация теней для всенаправленных источников света в реальном режиме времени // Программные продукты и системы. – 2008. – № 3.
3. Fan Zhang, Hanqiu Sun, Leilei Xu, Lee Kit Lun. Parallel-Split Shadow Maps for Large-Scale Virtual Environments // In Proceedings of ACM International Conference on Virtual Reality Continuum and Its Applications. – 2006. – Pp. 311–318.
4. http://www.opengl.org/wiki/Framebuffer_Object#Multisampling_Considerations (дата обращения: 15.09.2013)

Мальцев Андрей Валерьевич. Старший научный сотрудник Центра визуализации и спутниковых информационных технологий НИИСИ РАН. Окончил Московский государственный институт радиотехники, электроники и автоматики в 2008 году. Кандидат физико-математических наук. Автор 23 печатных работ. Область научных интересов: компьютерная графика, системы виртуальной реальности, информационные технологии. E-mail: avmaltcev@mail.ru