

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МНОГОПРОЦЕССОРНЫХ СИСТЕМ ФОРМИРОВАНИЯ ГРАФИЧЕСКИХ ИЗОБРАЖЕНИЙ

А.Г.Додонов¹, А.Н.Мельников², Я.В.Резник³

¹Институт проблем регистрации информации Национальной академии наук Украины
ул. Шпака, 2, Киев, 252113, Украина, тел.: (8044) 454-21-58

²Винницкий социально-экономический институт Университета «Украина»
ул. Фрунзе, 4, Винница, 21007, Украина, тел.: 53-47-18, E-Mail: anmel07@mail.ru

³Винницкий национальный медицинский университет
ул. Пирогова, 56, Винница, 21018, Украина, тел.: (0432) 27-96-83

Аннотация

Рассмотрены пути повышения эффективности и качества формирования графических изображений на основе современных технологий. Проведен анализ предложений ведущих компаний в области компьютерной графики NVIDIA и ATI Technologies, разработки которых направлены на повышение мощности графических систем. Детально проанализированы технологии SLI (NVIDIA) и CrossFire (ATI). Рассмотрены методы конвейерной обработки и распараллеливания, которые направлены на ускорение процесса формирования изображений с использованием аппаратных средств.

Постановка проблемы

В последние годы существенно расширилось применение средств компьютерной графики в самых разнообразных сферах деятельности человека. В современном мире, где роль информации становится все более важной, вполне естественно, что пространственные изображения, которые обладают высокой степенью информативности и реалистичности, превращаются в ее основной носитель. На данном этапе развития компьютерной графики разработка новых высокопродуктивных методов и средств формирования пространственных реалистических изображений является приоритетной задачей, поскольку традиционные методы и средства не обеспечивают необходимую производительность и реализм. Повышение требований к качеству изображений предусматривает повышенные требования к производительности систем формирования трехмерных изображений. Чем точнее и качественнее представляется объект, тем больший объем данных потребует для его описания. При этом игровые приложения и являются движущей силой развития высокой производительности компьютерных систем. Поэтому актуальным является определение оптимальных решений проблемы повышения качества и скорости обработки графических изображений.

Анализ последних исследований и публикаций

Анализируя последние научные исследования и публикации, следует отметить большую заинтересованность ученых, разработчиков и изготовителей в решении проблем, которые возникают при создании новейших технических средств и программ обработки графических изображений. Формирование реалистических изображений в реальном масштабе времени относят к одной из основных проблем компьютерной графики [1], поскольку традиционные подходы к формированию трехмерных изображений не обеспечивают требуемой производительности. Повышение производительности графических видеокарт достигают за счет увеличения тактовой частоты работы составных блоков, усовершенствования архитектурных решений, использования новой алгоритмической базы, замены программной реализации этапов графического конвейера на аппаратную, распараллеливания процесса формирования трехмерных изображений. Вопросам повышения производительности графических средств посвящены работы [1-7].

Одним из основных направлений, которое интенсивно развивается в последнее время, связано с использованием многопроцессорного подхода к формированию трехмерных изображений.

Цель статьи

Целью статьи является анализ основных подходов к многопроцессорной реализации трехмерных изображений, перспективных направлений повышения эффективности формирования графических изображений.

Изложение основного материала

Распространение многоядерных процессоров в последние годы привело к тому, что возможности многопроцессорных систем начали использоваться не только при решении сложных научных проблем, они стали применяться в системах графической обработки информации, в играх и других приложениях. Так, например, в индустрии видеоускорителей растет популярность решений для сопряжения нескольких дискретных акселераторов, повышающих производительность графических подсистем.

На рынке графических процессоров наметилась также новая тенденция – широкое распространение видеокарт, оснащенных несколькими видеочипами. Интенсивно начали использоваться 2-ядерные процессоры, на подходе 4-ядерные, значительно повышающие быстродействие формирования графических изображений.

Сегодня 3D-графика и связанные с ней приложения, т.е. компьютерные игры, стали мощнейшим стимулом прогресса в области персональных компьютеров. При этом поколения графических процессоров сменяют друг друга ежегодно, а дорогие 3D-карты устаревают.

В настоящее время в графических видеокартах наибольшего распространения получила тайловая архитектура. Тайловое строение карт используется в подавляющем большинстве 2-х мерных и 3-х мерных игр. Тайл имеет вид картинка фиксированного размера. Тайлы могут быть 2-х видов: прямоугольные и квадратные; изометрические (ощущение глубины и объема). Самый простой тайл – квадратное изображение, симметричное по горизонтали и вертикали. Созданное с помощью совокупности тайлов изображение не имеет заметных «швов».

Достоинства тайловой архитектуры. Построение изображения фрагментами - тайлами - разрешает эффективно использовать кеш графического процессора, поскольку при этом используются меньшие объемы данных и они однородны. Сформировав тайл, графический процессор больше не возвращается к нему в процессе построения кадра. Поскольку тайл имеет небольшие размеры, то в кеш-памяти графического процессора возможно сохранять фрагменты буфера кадра и Z-Буфера. Таким образом, все построения графический процессор проводит, обращаясь не к памяти, а к кешам. После формирования тайла содержимое тайлового буфера кадра и Z-Буфера записывается в оперативную память. Кеширование буфера кадра и Z-Буфера разрешает существенно понизить нагрузку на шину памяти, что особенно актуально при интегрированном выполнении, когда графическое ядро вынуждено делить шину памяти с центральным процессором.

Основным недостатком тайловой архитектуры являются присущие ей задержки, связанные с необходимостью создания списков объектов. Тайловая архитектура разрешает создать очень эффективную реализацию полноэкранный сглаживания с использованием мультиплексирования. Добавление настоящего мультиплексированного сглаживания со случайными сдвигами (джитером) обеспечивает более высокое качество изображения, чем другие решения, которые просто проводят рендеринг в высшем разрешении и фильтруют результат в низком для каждого тайла. К недостаткам разбиения изображения можно отнести неизбежность многократного обращения к одним и тем же данным, находящимся в системной памяти. Размер карты должен быть по возможности квадратным и по размерностям делиться на размер составляющих «блоков».

Разработка новых чипов, базирующихся на тайловой архитектуре, уже ведется многими компаниями, что свидетельствует о ее перспективности.

В настоящее время ведущие компании в области компьютерной графики по разработке и производству графических процессоров (видеоускорителей) NVIDIA и ATI Technologies предлагают наращивать мощь графической подсистемы за счет установки второй графической карты и применения технологий, соответственно SLI (NVIDIA) и CrossFire (ATI), обеспечивающих решение этой задачи.

SLI (Scalable Link Interface) — масштабируемый интерфейс. Для построения подобной системы используются две видеокарты с шиной PCI Express x16, мост, объединяющий их, и системная плата с двумя PCIe x16. Одна из карт является ведущей, а вторая – ведомой, при этом, первая отвечает за сборку финального кадра и вывод его на экран монитора. Объединение в SLI возможно как с использованием специального моста, так и без него, то есть программным путем. В последнем случае возрастает нагрузка на шину PCI, что негативно сказывается на производительности. SLI поддерживается всеми игровыми приложениями, обеспечивая увеличение производительности через режимы рендеринга AFR (Alternate Frame Rendering) или SFR (Split Frame Rendering). Для работы со старыми играми также предлагается так называемый режим совместимости Symmetric Multi-Rendering with Dynamic Load Balancing (SMR), то есть симметричный мультирендеринг с динамическим распределением загрузки, когда задействуется лишь одна графическая карта.

SLI имеет несколько режимов рендеринга.

1. Split Frame Rendering (SFR) – это разделение кадра на две части, за рендеринг каждой из которых отвечает отдельный видеоадаптер. Кадр может делиться на две равные части или размер этих частей может варьироваться динамически в зависимости от сложности сцены.

2. Alternate Frame Rendering (AFR) – поочередный рендеринг кадров. В этом режиме формируется последовательность кадров, которые сменяют друг друга: одна карта просчитывает все четные кадры, а вторая — нечетные.

3. Symmetric Multi-Rendering with Dynamic Load Balancing (SMR) - позволяет добиться лучшего КПД и обеспечить максимальный выигрыш в производительности, так как нагрузка на каждую карту в этом случае распределяется равномерно. Применение специальных высокоинтеллектуальных алгоритмов

динамического распределения нагрузки позволяет получить в режиме SLI увеличение производительности, достигающее 80-90%.

Технология SLI компании NVIDIA позволяет объединить возможности сразу двух видеокарт и повысить производительность в полтора-два раза при отсутствии проблем с совместимостью и стабильностью. В 2008 году компания NVIDIA реализовала технологию SLI на трех видеокартах.

Преимущества: значительно возрастает производительность; системные платы на nForce 4 SLI пользуются большим спросом у пользователей (продано более миллиона плат).

Недостатки: в некоторых играх увеличение скорости не всегда оправдывает высокую цену подобной системы; карты (обязательно одного производителя), их тактовые частоты и прошивки должны быть одинаковыми; поддержка слишком малого числа игр (менее 20 названий).

CrossFire (ATI) — технология объединения видеокарт ATI Technologies, разработанная для высокопроизводительных игровых систем. CrossFire поддерживается всеми картами модельного ряда Radeon X1000.

Достоинством является то, что платы могут быть не идентичными — они должны принадлежать только одной серии, но по функциональности чипов могут различаться. Суммарное быстродействие связки видеокарт определяется характеристиками наименее производительного чипа. Например, если ядро на одной из плат содержит 36 пиксельных конвейеров, а на другой — 48, то второй чип будет использовать только 36 конвейеров. То же правило распространяется и на частоты видеокарт. Отсюда следует, что рациональным является объединение двух одинаковых плат.

Отличительной главной особенностью видеокарт серии CrossFire Edition — является наличие дополнительного чипа Compositing Engine. Эта микросхема берет частично сформированную картинку из стандартной карты и смешивает с такой же промежуточной картинкой из карты CrossFire Edition. В результате получается полноценное изображение, обработанное с большей скоростью. Чип Compositing Engine — представляет собой программируемое устройство с гибкой поддержкой различных видеокарт, в том числе будущих продуктов.

Особенностью технологии CrossFire является поддержка самых разнообразных методов распределения задачи по обработке каждого кадра между двумя картами. Система ATI CrossFire обеспечивает три режима работы — SuperTiling, Scissor, Alternate Frame Rendering. Выбор режима рендеринга выбирается автоматически с помощью Catalyst A.I.

Режимы CrossFire.

1. SuperTiling — стандартный режим. Он совместим со всеми Direct3D-играми и приложениями.

Картинка разделяется на небольшие квадраты (quad), которые раскладываются графическим процессором по разным конвейерам. Метод SuperTiling заключается в разделении всего экрана на небольшие квадратные участки размером 32x32 пикселя, обработка которых производится каждой из карт в шахматном порядке. Метод SuperTiling обеспечивает сбалансированное распределение нагрузки между двумя видеокартами, а также наибольшую сбалансированность и лучшую производительность системы с нынешним поколением графических чипов. В перспективе планируется улучшить этот параметр таким образом, чтобы размер обрабатываемых участков стал динамически изменяющимся.

Преимущества: совместим со всеми играми, обеспечивает оптимальное распределение нагрузки.

Недостатки: в некоторых играх скорость возрастает не значительно, обе карты просчитывают геометрию сцены, этот режим не работает с OpenGL-играми.

2. В режиме Scissor ("Разрезание") картинка делится на две части по горизонтали. Этот метод известен как Split-Frame Rendering или рендеринг разделенного кадра. Соответственно, одна карта вычисляет верхнюю часть, а вторая — нижнюю (как в случае SLI). Динамическая балансировка нагрузки позволяет равномерно нагрузить обе видеокарты. Такой режим является стандартным для приложений OpenGL. При методе Split-Frame Rendering, рендеринг разделенного кадра, каждая часть кадра может делить кадр по горизонтали (хотя можно и по вертикали), поровну (теоретически можно и не поровну). Все зависит от производительности спаренных в единую связку видеокарт: если одна из них выполнена на более мощном чипе, ей достанется большая нагрузка. Лучшее соотношение деления кадра определяется автоматически для каждого конкретного случая, каждой конкретной игры. Такой метод рендеринга поддерживается в D3D и OGL приложениях.

Преимущества: работает с большинством игр, делит и геометрическую и пиксельную загрузку, высокая степень асинхронности работы VPU, ускоритель полностью владеет своей подотчетной зоной изображения результата.

Недостатки: обе карты просчитывают геометрию сцены, что требует балансировки на лету зон для равномерного распределения загрузки; появляются проблемы с антиалиазингом на стыке зон; требует заметного вмешательства в драйвер, поэтому высока вероятность неверной работы некоторых приложений.

3. Режим Alternate Frame Rendering (AFR) - осуществляет поочередный рендеринг кадров.

Достоинство такого способа заключается в том, что оба графических чипа работают совершенно

независимо друг от друга, в результате чего метод поочередного рендеринга кадров (AFR) потенциально должен показывать высокую производительность во всех режимах работы. В то же время метод AFR работает практически во всех современных играх, за исключением тех, где информация о текущем кадре закладывается в предыдущем кадре. В таком случае подойдут два других типа рендеринга. Рендеринг методом AFR поддерживается в Direct 3D-играх и OGL приложениях.

Преимущества: карты оптимально распределяют нагрузку, AFR обеспечивает полное ускорение геометрии; делит пиксельную и геометрическую нагрузку, причем геометрия не дублируется по шине — разные ускорители получают разные наборы данных; ускоритель полностью отвечает за свой кадр, отсутствуют следы стыковки, даже в случае сложного постпроцессинга, нет ограничений на метод построения кадра.

Недостатки: не работает в играх, где используется функция *gender-to-texture*; неравномерное чередование кадров и распределение загрузки; КПД сильно зависит от CPU и системы, а также от характера сцены и падает с увеличением числа кадров в секунду; проблема со значительной задержкой между кадром, который демонстрируется и кадром, который в данный момент формируется.

Преимущества CrossFire: ATI CrossFire работает со всеми играми на основе API DirectX и API OpenGL; можно использовать разные видеокарты (от разных производителей, с разными частотами); CrossFire-версии карт работают со старыми моделями Radeon X800/X850; возможность работы с несколькими мониторами (до 5); CrossFire обеспечивает работу всех существующих игр благодаря системе рендеринга.

Недостатки CrossFire: стоимость ведущей (master) видеокарты CrossFire заметно выше, чем у ведомой, в то время как стоимость обеих карт NVIDIA одинакова; малая доступность технологии на рынке.

На основе выше изложенного можно сделать выводы.

Компания ATI Technologies предложила более гибкую и перспективную CrossFire технологию объединения двух игровых видеокарт в одном системном блоке.

В перспективе технологии SLI и CrossFire придут на помощь центральному процессору при обработке физических процессов в играх. Одна из карт тандема CrossFire/SLI будет заниматься обработкой физики, а другая — графики. Здесь компании ATI и NVIDIA идут схожими путями. ATI предлагает использовать материнские платы с тремя разъемами PCIe x16 и три видеокарты для максимальной производительности. Две отвечают за графическое изображение, третья — за физику.

При сравнительно одинаковых параметрах видеокарт Nvidia и ATI - ATI показывает более высокую производительность за счет использования тайловой системы.

При среднем ценовом показателе (что немаловажно для развития технологий) графических процессоров Nvidia и ATI, последняя компания реализовала возможность работы с картами, не рассчитанными изначально на парное подключение, а значит большую совместимость.

Поддержка многопроцессорности – одно из перспективных направлений аппаратного ускорения. Многопроцессорность - это способ дать пользователям возможность повышения производительности за счет масштабируемости систем в переходный период до выпуска следующего, более производительного поколения процессоров.

Конвейерная обработка и распараллеливание

Для повышения ускорения процесса обработки информации с помощью аппаратных средств используются два метода: конвейерная обработка и распараллеливание.

Метод конвейерной обработки применяется в том случае, когда этапы задачи упорядочены и выполняются последовательно один за другим. Специализированные устройства выполняют операции для каждого этапа, передает результаты на следующий этап, и сразу же начинают выполнять новую работу, не ожидая окончания всего процесса. При конвейерной обработке теоретически процесс ускоряется пропорционально числу этапов конвейера.

Метод распараллеливания используется тогда, когда задача может быть разделена на почти независимые кванты. При этом для выполнения задачи создается несколько параллельных процессоров. Каждый квант работы обрабатывается свободным процессором.

Как конвейерная обработка, так и распараллеливание широко используются на этапе геометрических преобразований, эти два метода также полезны повсюду в универсальном графическом конвейере.

Конвейерная обработка проста, но она менее гибкая, чем распараллеливание. Этапы в аппаратном конвейере в основном реализуются схемно, что приводит к большим трудностям при добавлении нового этапа. Производительность системы падает, если нагрузка на некотором этапе становится больше, чем на других, этот этап начинает сдерживать работу всего конвейера.

Так как при распараллеливании каждый квант выполняется в основном независимо, то этот метод является гибче, чем конвейерная обработка. Параллельные процессоры могут выполнить дополнительную работу и не допустить простоя остальной аппаратуры. Эти свойства отсутствуют у аппаратного конвейера.

Однако важным является то, что конвейерная обработка и распараллеливание не противоречат, а дополняют друг друга. Например, внутри каждого параллельного геометрического процессора функционирует конвейер вычислений с плавающей точкой. Существенным моментом оценки графической аппаратуры является то, каким из двух способов, схемным или более общим, микропрограммным, она реализована.

Схемная реализация обычно больше подходит для хорошо определенных задач, которые почти наверняка не будут меняться, схемная аппаратура работает быстрее, но она не очень гибка и проектировать ее труднее. Аппаратура растеризации и вывода на дисплей часто реализуется схемно, а аппаратура для геометрических преобразований обычно микропрограммно.

Типы распараллеливания. Широкое распространение получили два способа реализации микропрограммных процессоров: SIMD (Single Instruction, Multiple Data) - выполнение одной и той же микропрограммы осуществляют все процессоры и MIMD (Multiple Instruction, Multiple Data) - свою собственную микропрограмму выполняет каждый процессор.

SIMD-архитектура дешевле, чем MIMD. Преимущество SIMD-процессоров проявляется при вычислениях однородной природы, примером является обработка треугольников одного типа с одинаковыми режимами OpenGL. MIMD-архитектура более свойственна для менее однородных работ, таких как обработка треугольников с изменяющимися режимами OpenGL.

Конвейер с N этапами или N-кратное распараллеливание не всегда приводят к N-кратному увеличению производительности. Изменение состояния внутри аппаратного конвейера, по всей видимости, потребует наличия отстойников, которые снижают производительность, так как не все этапы будут постоянно загружены. В параллельной системе также может не хватать работы для загрузки всех N процессоров или могут возникать коммуникационные издержки, необходимые для координации параллельной обработки.

Выводы

Реализация многопроцессорной и многопоточковой обработки компьютерной графики является наиболее оптимальным решением для удовлетворения современных требований к продуктивности систем формирования трехмерных изображений. Повышение производительности формирования трехмерных изображений позволяет повысить качество генерации графических сцен, их динамику, использовать новые эффекты, включая и просчет физических процессов. Многопроцессорное формирование графических изображений позволяет достичь комфортной частоты обновления сцен, что обеспечивает плавное перемещение моделей, а, следовательно, повышает реализм восприятия процессов.

Литература:

- [1] Баяковский Ю. М. О некоторых фундаментальных проблемах компьютерной (машинной) графики / Ю. М. Баяковский, В. А. Галактионов // "Информационные технологии и вычислительные системы". — 2004. — № 4. С. 3—24.
- [2] Вяткин С. И. Растеризационные методики и архитектуры систем визуализации реального времени / С. И. Вяткин, Б. С. Долговесов, В. М. Фомичев // Труды 17—й Междунар. конф. "Графикон—2007", 2007. — Р. 164—169.
- [3] Графические адаптеры. Четвертое поколение // Upgrade. — 2005. — №21. — С.45—47.
- [4] Штрассер В., Шиллинг А., Книттель Г. Архитектуры высокопроизводительных графических систем / В. Штрассер, А. Шиллинг, Г. Книттель // Открытые системы. — 1995. — №5. — С. 53—60.
- [5] Куц В. Современные графические процессоры [Электронный ресурс] / В. Куц. — Режим доступа: <http://www.3dnews.ru/video/gpu—2007/>
- [6] Орлов Д. А. Распараллеливание процесса динамической визуализации трехмерных сцен / Д. А. Орлов // Вычислительные сети. Теория и практика. — 2005. — Номер 2 (7).
- [7] Пахомов С. Революция в мире графических процессоров / С.Пахомов // Компьютер пресс. — 2006. — №12.