

Санкт-Петербургский Государственный Университет

Математико-механический факультет

Кафедра системного программирования

Анимирование речи

Курсовая работа студента 371 группы

Брыксина Матвея Александровича

Научный руководитель: к. ф.-м. н., доцент А.Т. Вахитов

Санкт-Петербург

2015

Оглавление

Введение	3
Постановка задачи	4
Обзор подходов к анимированию лиц	4
Полигональная модель Парке	5
Подходы к созданию переходов	5
Морфинг промежуточных состояний.....	6
Подход Осипа к анимированию лиц.....	7
Реализация анимации 3D-модели	8
Входные данные	8
Детектирование точек на лице	9
Общая архитектура приложения.....	9
Морфинг состояний и вычисление коэффициентов.....	10
Рендеринг модели в реальном времени.....	11
Заключение.....	12
Список литературы.....	13

Введение

Реалистичное отображение человеческого лица в динамике со всеми его тонкостями – одна из самых сложных проблем в компьютерной графике. Люди достаточно хорошо научились детектировать и распознавать положение лица, но распознавание детальных выражений лица более сложная задача, так как зависит от особенностей личности. Именно детализированные выражения лица и их качественная динамика позволяют передать индивидуальность и интенсивность модели.

Переходы между положениями лица может быть охарактеризовано как множественные геометрические растяжения и перемещения во времени. Коротковременные локальные изменения могут возникнуть за доли секунды. Более глобальные, такие как движения щек при открытии рта – более продолжительные и плавные изменения. Такие тонкости человеческого лица, как поры или родинки тяжело подвергать локальным деформациям, поэтому принято считать статичными.

Хотя индустрия мультимедиа и кино продолжает развивать направление компьютерной графики, но она до сих пор вынуждена прибегать к существенному вмешательству квалифицированных художников для построения моделей или анимации. Увеличение уровня детализации приводит к росту количества ручного труда, что повышает стоимость разработки.

Анимация лиц включает в себя ряд общих задач, автоматизация которых уменьшает количество ручной работы художников:

- Построение трехмерной модели головы по входным изображениям лица человека;
- Реалистичный рендеринг трехмерной модели головы;
- Построение и отрисовка плавных переходов между состояниями модели.

Отрисовка анимации в реальном времени кроме вышеперечисленных включает в себя задачу детектирования положения лица, основных областей и ключевых точек на

лице. Анимирование лиц в реальном времени разделяется на задачу анимирования модели и задачу проекционного наложения изображения на лицо актера (3D-mapping). Анимирование трехмерных моделей широко распространено в киноиндустрии и было неотъемлемой частью при создании сцен из фильма Аватар.

Постановка задачи

Целью данной курсовой работы является создание реалистичной анимации трехмерной модели головы, повторяющей движение лица актера при разговоре.

Чтобы реализовать поставленную цель были поставлены следующие задачи:

- Провести исследование существующих подходов к отрисовке лиц в динамике;
- Детектирование ключевых точек на лице человека;
- Рендеринг 3D-модели головы по описанию модели и набору текстур;
- Построение анимации переходов между состояниями 3D-модели;

Обзор подходов к анимированию лиц

Основных подходов к созданию анимации лиц не так много. Основные концепции этих подходов небольшие и простые для понимания. Они сильно отличаются в своих деталях: свойствах и особенностях входных данных, возможностями, контекстом применения, а также спецификой полученной анимации.

При исследовании подходов необходимо было найти компромисс между высокой точностью отрисовки для реалистичности моделей и скоростью работы для возможности отрисовки в реальном времени.

Полигональная модель Парке

Фредерик Парке (Frederik Parke) создал первую 3D-модель человеческого лица и описал это в своей диссертации 1974 года [6]. Он создал иерархию объектов на лице и описал параметры, которые влияют на отображение этих объектов в модели.

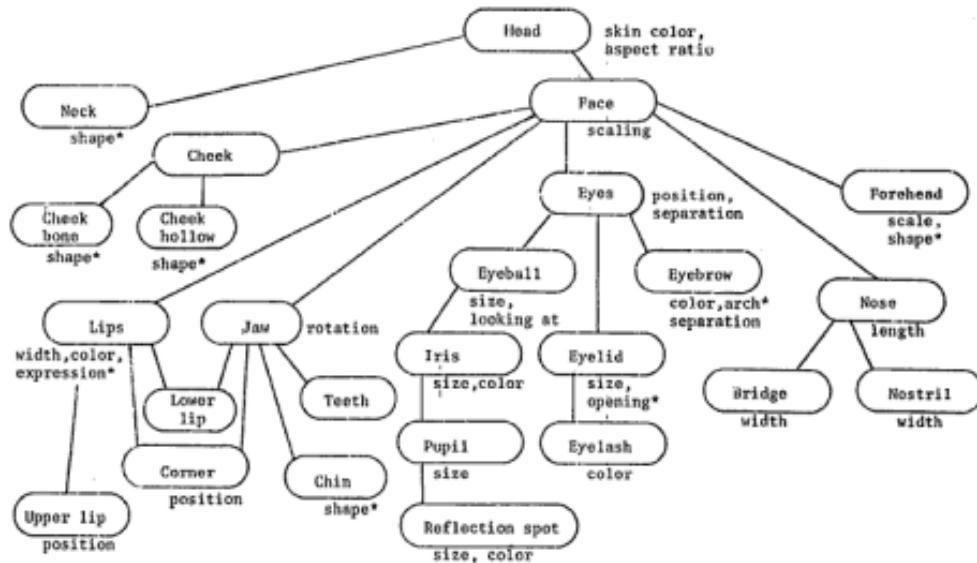


Рисунок 1. Параметризованная иерархия Парке областей на лице

Трёхмерная модель Парке задавалась набором координат вершин и плоскостями (полигонами), проведенными через эти вершины. Количество вершин в моделях не превышало 500, поэтому модели не выглядели сглаженными.

Подходы к созданию переходов

Тони де Пельтри (Tony de Peltrie) в 1985 году создал первую компьютерную анимацию лица. Всем выражениям лица соответствовали фотографии актера, которые проецировались в набор точек на 3D сцене в соответствии с моделью Парке. Дальнейшее развитие алгоритмов в этом направлении было направлено на уменьшение количества входных данных – задание положений модели соответствующих фонемам языка (минимальным звуковым единицам) и создание алгоритмов переходов между положениями модели.

Концептуально другой подход к построению анимации и описанию моделей изложил Кейт Уотерс (Keith Waters) в 1987 году. Трехмерная модель задавалась набором мускулов, заданных геометрическими фигурами, а выражения лица (состояния модели) строились как математические преобразования над геометрическими фигурами. Первый компьютерно-анимированный фильм “Tin Toy” (1988 г.) студии Pixar, получивший Оскара в 1988 году за лучший короткометражный анимированный фильм, использовал именно мускульный подход к описанию моделей Уотерса.

Морфинг промежуточных состояний

Состояния полигональных моделей Парке задаются набором вершин и поверхностей. Если состояния моделей будут иметь одинаковое число вершин, то морфинг состояний (смешивание точек) можно описать математическими уравнениями. Тогда промежуточные состояния будут интерполироваться, что позволит уменьшить количество входных состояний.



Рисунок 2. Состояния в центре получены интерполированием между крайними

Линейная интерполяция

Промежуточное состояние рассчитывается как сумма двух состояний с одним коэффициентом. Подход хорошо масштабируется на несколько состояний.

$$M'(\alpha) = \alpha M_1 + (1 - \alpha)M_2 \quad (1) , \text{ где } M'(\alpha) \text{ — состояние заданное весом } M_1 \text{ и } M_2 \text{ — исходные ключевые состояния}$$

Нелинейная интерполяция

Усовершенствование алгоритма линейной интерполяции. Коэффициент α в уравнении 1 рассчитывается как полином высших степеней. Это позволяет добавлять ускорение в начале анимации и торможение в конце. Коэффициенты в полиноме требуют ручного подбора в зависимости от специфики анимации.

Подход Осипа к анимированию лиц

В своей книге «Stop Staring»[1] Джейсон Осипа (Jason Osipa) отказывается от прорисовки состояний модели, соответствующим каждой фонеме языка, а описывает универсальный алгоритм, позволяющий проинтерполировать все фонемы из четырех ключевых состояний модели.



Рисунок 3. Основные положения губ. Закрытое, открытое, узкое, широкое

Концептуальный подход Осипа заключается в том, что речь человека можно описать циклом из переходов между состояниями закрытого/открытого и узкого/широкого рта. При наличии качественной трехмерной модели к ней можно применить алгоритм линейной интерполяции. Это позволит просчитывать все промежуточные состояния для любых фонем. Создав базу коэффициентов интерполяции для каждой фонемы, можно строить анимацию по входному набору фонем.

Аналогично Осипа описывает циклы бровей и глаз, чтобы добавить эмоции в модель и делать анимацию более реалистичной.

Реализация анимации 3D-модели

Чтобы выполнить поставленную цель, за основу был взят подход к анимированию лиц Джейсона Осипа. Такой алгоритм требует наименьшее количество входных данных (состояний модели), а линейное интерполирование позволяет быстро просчитывать коэффициенты, что важно для отрисовки в реальном времени.

Всю работу приложения можно разделить на несколько подготовительных этапов, необходимых для анимации, и цикл детектирования ключевых точек на лице и отрисовка трехмерной модели в реальном времени.

Входные данные

На подготовительном этапе приложение считывает четыре ключевых состояния модели открытое/закрытое и узкое/широкое. Трехмерная модель описана в формате obj, который поддерживается большинством 3D-редакторов, и содержит в себе информацию о вершинах, нормалях, плоскостях и координатах в файле текстуры.

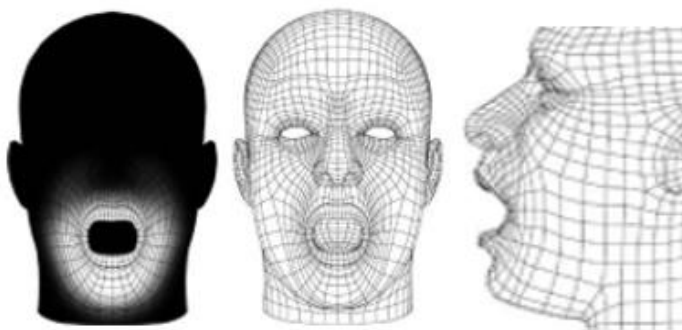


Рисунок 4. Объектная модель в состоянии открытого рта

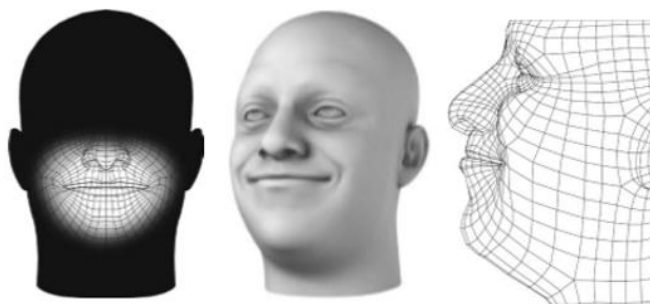


Рисунок 5. Объектная модель в состоянии широкого рта

Для создания реалистичного изображения модели кроме состояний модели на подготовительном этапе также считывается текстура в формате .jpg, .png, .tga, которая во время цикла отрисовки будет накладываться на модель.

Общая архитектура приложения

Весь цикл отрисовки анимации трехмерной модели можно разделить на несколько последовательных этапов работы приложения:

- Детектирование положение рта на входном изображении с камеры;
- Определение состояния и положения рта по сдетектированным точкам;
- Расчет коэффициентов интерполяции и морфинг текущего состояния между конечными состояниями модели;
- Отрисовка сцены с трехмерной моделью;

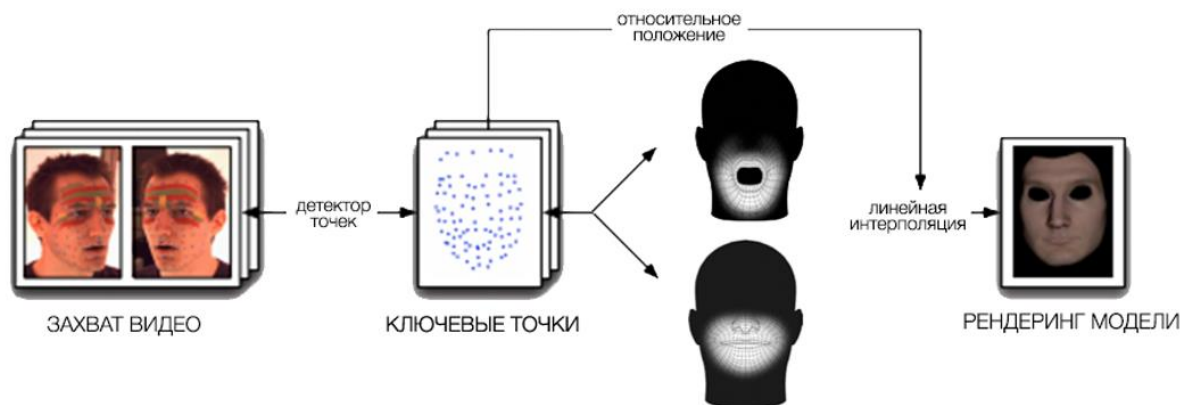


Рисунок 6. Этапы цикла отрисовки анимации

Детектирование точек на лице

Чтобы сдетектировать точки на лице использовались возможности библиотеки открытого доступа для отслеживания лиц CSIRO Face Analysis SDK [7], которая базируется на библиотеке открытого доступа OpenCV [3]. Библиотека позволяет сдетектировать 66 ключевых точек на лице (рисунок 7).

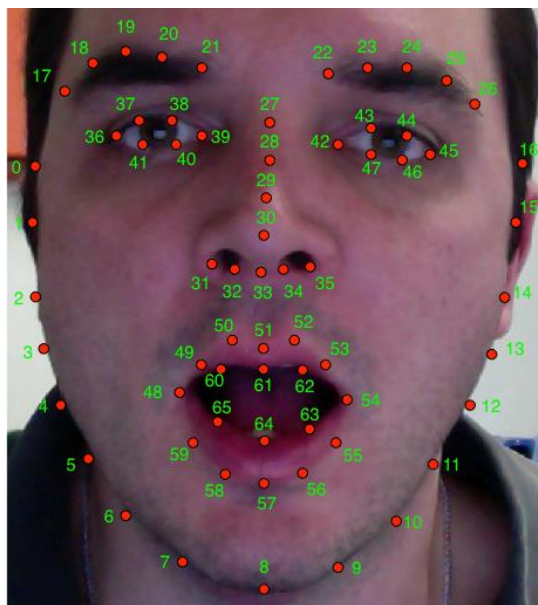


Рисунок 7. Порядок сдетектированных точек на лице

Определение положения рта происходит из анализа положения точек с номерами с 48 по 65 (рисунок 6). На подготовительном этапе перед анимацией считываются начальные координаты положения рта, соответствующие четырем входным ключевым состояниям модели. За основу ключевых состояний берется разница между крайними точками слева/справа и снизу/сверху. Такая инициализация необходима для просчета коэффициентов линейной интерполяции в реальном времени.

Морфинг состояний и вычисление коэффициентов

Морфинг промежуточных состояний осуществляется алгоритмом линейной интерполяции. Применять подход нелинейной интерполяции с плавной анимации в случае отрисовки в реальном времени не имеет смысла, так изменение состояний происходит слишком быстро.

Посчитанный коэффициент интерполяции из текущего положения точек на лице во время цикла отрисовки относительно проинициализированных ключевых состояний на лице подается как коэффициент состояний модели на этапе отрисовки.

Рендеринг модели в реальном времени

Для отрисовки трехмерной модели использовались инструменты открытой библиотеки OpenGL [2] (GLUT, GLEW) и открытой математической библиотеки GLM.

Упрощенная схема графического конвейера OpenGL представлен на рисунке 7.

Для отрисовки трехмерной модели были написаны программы вершинных и фрагментов шейдеров на языке GLSL.

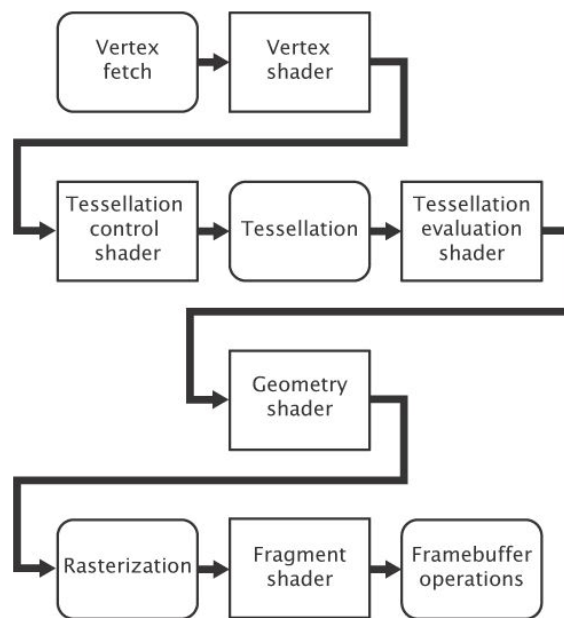


Рисунок 8. Упрощенная схема графического конвейера OpenGL

Заключение

В рамках проведенной работы был проведен анализ существующих подходов к описанию трехмерных моделей головы и способов построения анимации. Исходя из особенностей и специфики существующих алгоритмов, был выбран подход позволяющий сократить количество входных данных до минимума, что позволяет уменьшить количество подготовительной работы.

Для создания реалистичной анимации трехмерной модели головы, повторяющей движение губ на лице актера

- сделано детектирование и распознавание состояния рта на лице на входном изображении с камеры при помощи CSIRO SDK,
- реализован алгоритм линейной интерполяции для анимации переходов между состояниями 3D-модели,
- реализован рендеринг трехмерной модели головы по входному описанию модели в формате obj и файлам текстур в формате jpg, png, tga.

Дальнейшие перспективы

Для построения более реалистичной анимации речи планируется добавить поддержку состояний, необходимых для анимирования бровей и глаз. Это позволит добавлять эмоции в анимацию.

Чтобы добавить динамики в модель и придать речи большую естественность, планируется добавить отслеживание изменения положения головы на камере и отрисовку соответствующего положения модели головы.

Список используемой литературы

- [1] Jason Osipa “Stop Staring: Facial Modeling and Animation Done Right”, 3th edition, 2010.
- [2] URL: <https://www.opengl.org/> дата обращения 26.05.2015.
- [3] URL: <http://opencv.org/> дата обращения 26.05.2015.
- [4] URL: <http://face.ci2cv.net/> дата обращения 26.05.2015.
- [5] Parke, F.I. (Nov. 1982) “Parameterized Models for Facial Animation” IEEE Computer Graphics and Applications, 2 (9) 61- 68.
- [6] Parke, F.I. (Dec. 1974) “A Parametric Model for Human Faces” PhD. dissertation, University of Utah.
- [7] M. Cox, J. Nuevo, J.Saragih and S. Lucey, “CSIRO Face Analysis SDK“, AFGR 2013.
- [8] J. Saragih, S. Lucey and J. Cohn, “*Deformable Model Fitting by Regularized Landmark Mean-Shift*“, IJCV 2011.