



Ю.Г. КРИВОНОС, Ю.В. КРАК

УДК 004.93

МОДЕЛИРОВАНИЕ ДВИЖЕНИЙ РУК, МИМИКИ И АРТИКУЛЯЦИИ ЛИЦА ЧЕЛОВЕКА ДЛЯ СИНТЕЗА И ВИЗУАЛИЗАЦИИ ЖЕСТОВОЙ ИНФОРМАЦИИ

Ключевые слова: *моделирование, жестовый язык, информационная технология.*

ВВЕДЕНИЕ

Моделирование движения рук, мимики и артикуляции лица человека является актуальной задачей, поскольку ее решение обуславливает создание инновационных технологий, которые могут существенно повлиять на такие прикладные направления, как компьютерная техника, медицина, робототехника, компьютерная графика, системы виртуальной реальности. Как отмечал Стив Баллмер, генеральный директор Microsoft Corporation, «...мы начинаем видеть появление интерфейсов, которые управляют прикосновением и жестами» [1]. Важным практическим применением исследуемой проблемы есть создание компьютерных систем общения для людей с дефектами слуха, которые общаются как между собой, так и со слышащими людьми в формах и образах, близких и понятных для всех. Развитие современной науки, компьютеризация общества, использование мультимедийных и интернет-технологий создали достаточные условия для разработки таких систем коммуникации [2].

Люди с дефектами слуха для общения между собой применяют жестовый язык, в котором для передачи информации используются движения рук, мимика и артикуляция лица. Отметим, что основными визуальными средствами для воспроизведения этого языка есть фотографические и видеоизображения жестов, на базе которых разрабатывать современные учебные и коммуникационные компьютерные системы [3] достаточно проблематично, поскольку фотографическое изображение не отражает нужной динамики жестов, а в видеоизображении невозможно получить необходимую информативность жеста (нельзя увидеть воспроизведение жеста с разных ракурсов). Эти жесткие ограничения существующих средств воспроизведения жестового языка приводят к необходимости создания комплексной информационной технологии, которая включала бы функциональность по синтезу движений жестового языка на пространственной модели человека.

Цель настоящей статьи — создание соответствующих информационных и математических моделей для реализации возможности анимации процесса общения с помощью жестового языка, используя виртуальную модель человека. Разработана модель для фиксации морфем (минимально значимых единиц) жестового языка; в рамках этой модели создана технология и соответствующее программное обеспечение для получения, сохранения и воспроизведения жестов. Предложены алгоритмические решения для расчета свойственных человеку траекторий движения рук и корпуса модели при переходах от одного жеста к другому, а также для анимации мимики и артикуляции лица человека.

© Ю.Г. Кривонос, Ю.В. Крак, 2011

СТРУКТУРА ПРОСТРАНСТВЕННОЙ МОДЕЛИ ЧЕЛОВЕКА ДЛЯ ФИКСАЦИИ ЕДИНИЦ ЖЕСТОВОГО ЯЗЫКА

Построение процесса воспроизведения жеста на пространственной модели человека можно считать анимацией с соответствующей частотой различных состояний скелетной модели человека, которая упрощенно воссоздает скелет живого человека. Ее можно формализовать как иерархическую структуру, состоящую из звеньев, соединенных между собой кинематическими парами соответствующего класса [4], которые моделируют основные кости человеческого скелета. Современные пакеты трехмерного моделирования (Poser, 3D Studio Max) позволяют генерировать анимацию с помощью виртуальной статической модели с использованием информации об изменении соответствующих углов поворотов костей скелета.

Для формального описания процесса моделирования жеста можно использовать множество, отображающее упрощенный скелет H человека, изменения значений углов Эйлера M и порядка их применения для построения соответствующих костей этого скелета с изменением времени (дискретно, с соответствующей частотой (например, $1/30$ с)):

$$H = \{H_i: H_i = \{k, d_i, M_i \in M\}\}, \quad (1)$$

где H_i — i -я кость в скелете ($i=0, \dots, N-1$, N — количество костей в скелете); k — индекс кости-предка; $d_i = [x_i, y_i, z_i]^T$ — координаты точки конца кости в системе координат, связанной с началом этой кости;

$$M = \{M_i: M_i = (\text{order}_i, \theta_i)\}, \quad (2)$$

где M_i — значения углов Эйлера и порядок применения поворотов с течением времени для i -й кости; $\text{order}_i \in \{1, \dots, 6\}$ — порядок применения поворотов вокруг соответствующих координатных осей для i -й кости (1-XYZ, 2-XZY, 3-YXZ, 4-YZX, 5-ZXY, 6-ZYX); $\theta_i = (\theta_i^j)$, $\theta_i^j = \{(\varphi_{iX}^j, \varphi_{iY}^j, \varphi_{iZ}^j)\}$ — множество углов Эйлера для i -й кости ($j=0, \dots, K-1$, K — количество кадров для воспроизведения движения с заданной частотой).

Для получения множества углов (2), характеризующих изменение положения костей от начального состояния скелета, используется технология захвата движения (motion capture) [5]. На основании предложенного подхода создано соответствующее программное обеспечение [6], с помощью которого осуществляется фиксация и оцифровка жестов.

МОДЕЛИРОВАНИЕ МИМИКИ И АРТИКУЛЯЦИИ ЛИЦА ПРИ ВОСПРОИЗВЕДЕНИИ СЛОВ

Для отображения мимики и артикуляции лица на пространственной модели необходимо построить математическую модель, в рамках которой возможно выявить различия в положении контура губ человека для построения системы обучения правильной артикуляции при воспроизведении слов. Математическая модель должна включать в себя реализацию возможностей создания визуального алфавита языка, анализа визуального отображения фонем речи, применения полученных результатов для произвольных моделей лица человека.

Для синтеза математической модели перейдем от пространства фотографических изображений лица человека в процессе проговаривания к векторному пространству характеристических параметров с целью отображения лица человека на пространственной модели. Такой переход будет осуществляться за два этапа.

1. Выделение на изображении лица человека внутреннего контура губ

$$\text{Im } L \rightarrow D, \quad (3)$$

где $\text{Im } L = \{I_k: I_k \in FSV\}$ — упорядоченное множество ключевых кадров видео-

потока FSV (Face Speech Video), сформированного при съемке мимических проявлений на лице человека, а именно положения губ при проговаривании слов ($k=1, K$ — порядковый индекс кадра в выбранной последовательности, где K — количество ключевых кадров); $I_k = \{col_{ij}^k\}$, $i=\overline{1, m}$, $j=\overline{1, n}$, — изображение размера $m \times n$ лица с мимическим положением губ при проговаривании слов, m и n — длина и ширина изображения I_k соответственно; $col_{ij}^k = I_k(i, j)$ — цвет пикселя в системе RGB с координатами (i, j) на изображении I_k ; $D = \{D_k : D_k = \{d_{top}^k, d_{bot}^k\}\}$ — множество верхнего d_{top}^k и нижнего d_{bot}^k контуров губ, D_k — пара точечных кривых — контуров губ для k -го кадра.

2. Аппроксимация полученной точечной кривой внутреннего контура губ с помощью неравномерных базисных сплайнов (NURBS) [7] — получение вектора характеристических признаков:

$$D \rightarrow P, \quad (4)$$

где $P = \{p_k : p_k^i \in F, i=\overline{1, s}\}$ — пространство характеристических признаков, F — характеристические признаки объекта исследования, p_k — характеристический вектор, p_k^i — координаты этого вектора, s — размерность пространства P .

С использованием преобразования (3),(4) создана информационная технология [8], с помощью которой осуществляется моделирование мимики и артикуляции лица человека.

МОДЕЛИРОВАНИЕ ПРОЦЕССА АНИМАЦИИ ЖЕСТОВОГО ЯЗЫКА

Для синтеза процесса анимации жестов и мимики на пространственной модели человека предложены следующее формальное описание соответствующего множества параметров и алгоритмы работы с ними. Трехмерная модель человека, на которой будет реализовываться процесс анимации жестов и мимики, имеет следующие составляющие: $V = \{v_i : v_i = (x, y, z)\}$ — множество вершин треугольников для триангуляции поверхности трехмерной модели человека; $N = \{n_i : n_i = (x, y, z)\}$ — множество нормалей к вершинам; $T = \{t_i : t_i = (t, r)\}$ — множество текстурных координат к вершинам; $V^{ind} = \{V_i^{ind} : V_i^{ind} = (k_1, k_2, k_3)\}$ — множество индексов, указывающих на порядок построения треугольников из множества вершин; $I = \{I_i : I_i = \{img\}\}$ — множество фотографических изображений элементов модели — текстуры; $i=\overline{1, S}$, S — определенное число.

Для моделирования скелетной анимации необходимо рассчитывать новые значения вершин V треугольников. Для этого используется механизм скининга — алгоритм привязки множества вершин треугольников, определяющих поверхность модели, к значениям углов поворотов скелета. Тогда модель скелетной анимации можно формализовать следующим образом: $MH = \{MH_i : MH_i = \{k, \{l_1, \dots, l_m\}, d_i, Glb_i, Order_i\}$ — описание упрощенного скелета человека (иерархия костей) для реализации скелетной анимации, где MH_i — i -я кость скелета ($i=0, N-1, N$ — количество костей в скелете); k — индекс кости-предка; $\{l_1, \dots, l_m\}$ — множество индексов дочерних костей, $d_i = [x_i, y_i, z_i]^T$ — координаты точки конца кости в системе координат, связанной с началом этой кости; Glb_i — вектор для определения координат кости в глобальной системе координат, $Order_i$ — порядок применения поворотов; $Skin = \{Skin_i : Skin_i = \{(IndexVertex_1, Weight_1), \dots\}$ — множество вершин, влияющих на текущую вершину при изменении углов.

Скининг рассчитывается для каждой вершины V следующим образом:

$$v' = \sum_{i=0}^n \{(v * IM_{H_i} * TM_{H_i}) * w_{H_i}\}, \quad (5)$$

где n — количество костей, относящихся к вершине V ; IM_{H_i} — инверсная матрица для кости H_i ; TM_{H_i} — матрица перемещения относительно кости H_i ; w_{H_i} — весовой коэффициент влияния точек кости H_i на вершину V .

Для моделирования анимации проговаривания и эмоционального окрашивания используется механизм морфинга — плавного перехода от одного состояния объекта к другому. При морфинге используются только опорные состояния, с помощью которых рассчитываются промежуточные состояния и моделируется процесс анимации.

Модель морфинга мимики проговаривания слов опишем следующим образом. Воспроизведение или построение мимики на лице трехмерной модели достигается путем применения относительного (сегментного) морфинга к модели лица и мимического процесса. Выражение для относительного морфинга, состоящего из R морфов, в формализме модели имеет вид

$$V' = V + \sum_{m=1}^R w_m \cdot RM_m, \quad (6)$$

где w_m — весовые коэффициенты, V — множество вершин базовой модели, RM_m — входной морф для смешивания (blending). Результат операции — линейная комбинация множества вершин модели и мимики.

Метод сегментного морфинга дает возможность формировать несколько выражений лица на основе небольшого количества морфов и изменять состояние лица при анимации проговаривания. Преимуществом сегментного морфинга является возможность анимирования челюсти независимо от выражения губ и глаз, а также от эмоциональных проявлений на лице.

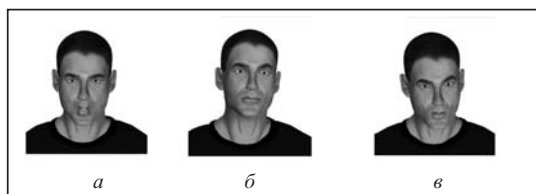


Рис. 1. Положение лица человека при проговаривании губами информации: *a* — визема «у»; *b* — вопросительное выражение лица; *v* — морфинг виземы «у» и вопросительного выражения лица

Рассмотрим пример синтеза состояний лица человека посредством морфинга визем украинской речи (рис. 1) в процессе проговаривания губами передаваемой словесной информации. Морфинг с использованием базисных визем осуществляется с учетом весовых коэффициентов:

- 1) вычисляется количество кадров для анимации;
- 2) определяется набор эмоций, присутствующих в слове;

- 3) рассчитывается набор визем для визуализации процесса артикуляции;
- 4) вычисляются продолжительности этапов анимации и точки появления визем.

Для синхронизации анимации жеста и мимики необходимо, чтобы начало и конец анимации мимики артикуляции совпадали с началом и концом анимации жеста.

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ДЛЯ ВОСПРОИЗВЕДЕНИЯ ПРОЦЕССА АНИМАЦИИ

Для воспроизведения процесса анимации жестов и мимики трехмерной моделью человека создано программное обеспечение, реализующее скелетную анимацию для воспроизведения жеста (рис. 2, *a*) и морфемную анимацию для воспроизведения мимики проговаривания и эмоций (рис. 2, *b*). Реализована соответствующая программная функциональность, которая, используя трехмерное API OpenGL, воспроизводит по заданным атрибутам 3D модель человека. На рис. 3 приведена модель анимации жеста с использованием алгоритмов скининга и морфинга.

На базе созданной информационной технологии реализованы украинский жестовый язык и методика преподавания жестового языка в специализированных общеобразовательных школах для незрячих детей.

Функционально технология состоит из трех информационных блоков (темы, слова и предложения) и блока воспроизведения жеста виртуальной моделью, который имеет определенное предназначение — с его помощью возможно продемонстрировать динамику жеста в учебном процессе. Поскольку жесты оцифровываются носителями жестового языка, они, по сути, становятся стандартами воспроизведения жеста. Реализованная в блоке функциональность для покадровой демонстрации жеста служит средством, с помощью которого станет возможным изучение жеста без конкретного преподавателя. Это может послужить основой создания стандарта жестового языка для детей из школ разных регионов. Фактически при таком подходе жестовый язык (в основном своем множестве) может стать стандартным на территории Украины.

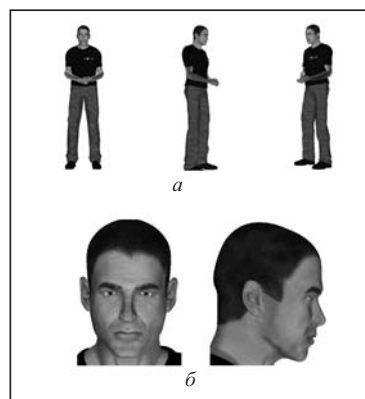


Рис. 2. Пространственная модель человека (а) и морфемная модель лица человека (б)



Рис. 3. Модель процесса анимации жеста «ты»

ЗАКЛЮЧЕНИЕ

При использовании рассмотренных моделей для фиксации движений, которые воспроизводят украинский жестовый язык, было оцифровано несколько десятков жестов. Благодаря разработанной технологии воспроизведения жестов из этого множества (с помощью пространственной модели) появилась возможность моделировать движения, полученные с видеоизображения конкретного человека — носителя жестового языка.

Реализована технология, с помощью которой воспроизводятся уроки украинского жестового языка. Созданная технология и программное обеспечение позволяют унифицировать изучение жестового языка и в этом случае могут стать основой создания стандарта жестового языка.

Дальнейшие исследования направлены на усовершенствование предлагаемой технологии:

- наполнение базы данных жестов;
- создание средств для семантического связывания предложений обычного языка с предложениями на жестовом языке;
- усовершенствование пространственной модели человека с учетом специфики реализации жестового языка.

СПИСОК ЛИТЕРАТУРЫ

1. Ballmer S. The fifth revolution — keynote script // CeBIT, 4–9 March 2008. Интернет-ресурс. — www.cebit.de/7514?pm=cb08-168-e.
2. Кривонос Ю.Г., Крак Ю.В., Бармак О.В., Тернов А.С. Информационная технология невербального общения людей с проблемами слуха // Искусственный интеллект. — 2008. — № 3. — С. 325–331.

3. Беликов В.И. Жестовые системы коммуникация // Семиотика и информатика. — 1973. — Вып. 20. — С. 127–148.
4. Кириченко М.Ф., Крак Ю.В., Сорока Р.О. Оптимізація маніпуляційних роботів. — К.: Либідь, 1990. — 144 с.
5. Lander J. Working with motion capture file formats. Game developer. — Miller Freeman Inc. USA, 1998. — P. 30–37.
6. Кривонос Ю.Г., Крак Ю.В., Бармак О.В., Тернов А.С., Троценко Б.А. Информационная технология для моделирования украинского жестового языка // Искусственный интеллект. — 2009. — № 3. — С. 186–198.
7. Piegl L., Tiller W. The NURBS book, 2nd Edition. — Berlin, Germany: Springer-Verlag, 1996. — 645 p.
8. Крак Ю.В., Бармак О.В., Тернов А.С. Информационная технология для автоматического чтения по губам украинской речи // Компьютерная математика. — 2009. — № 1. — С. 86–95.

Поступила 10.03.2011