

КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ

Юрій В. Крак, Юрій Г. Кривонос, Олександр В. Бармак

Інститут кібернетики ім. В.М. Глушкова НАН України,
03680, МСП, Київ, проспект Академіка Глушкова, 40
krak@unicyb.kiev.ua

Abstract

A mathematical model and its software implementation for sign language speech obtaining and reproduction using 3d human model has been suggested. The suggested technology is proven to be capable of realistic 3d human representation as the data is obtained from the real-world sign language speaker.

1. Вступ і постановка задачі

Дослідження закономірностей і анімації рухів є актуальною задачею. Розв'язок цієї проблеми у формі інноваційних технологій може суттєво вплинути на такі сфери людської діяльності як медицина, робототехніка, комп'ютерна графіка і системи віртуальної реальності.

В Україні більше півмільйона дітей з вадами слуху, а кількість глухих людей, для яких потрібно розробляти сучасні засоби навчання і спілкування – становить багато сотень тисяч. Розвиток сучасної науки, комп'ютеризація суспільства, використання мультимедійних та Інтернет технологій створили достатні умови для розробки комп'ютерних систем комунікації цих людей у формах і образах близьких і зрозумілих для них і для оточуючого середовища.

Люди з вадами слуху для спілкування між собою використовують жестову мову. Основними візуальними засобами які відтворюють цю мову є фотографічні та відео зображення жестів. Створювати за допомогою цих засобів сучасні навчальні та комунікаційні комп'ютерні системи досить проблематично. Фотографічне зображення не відтворює потрібної динаміки жестів а відео зображення досить громіздке і в ньому відсутня потрібна інтерактивність (не можливо подивитися на людину яка відтворює жест з інших ракурсів крім того в якому проведений відеозапис).

Ці суттєві обмеження існуючих засобів відтворення жестової мови спонукають до розробки більш гнучких технологій з допомогою яких можна було б створювати нові комп'ютерні системи навчання та комунікації для людей з вадами слуху. В розвиток цього запропонована концепція [1,2] інформаційної технології невербального спілкування людей з вадами слуху. Комплексна інформаційна технологія включає в себе крім іншої і функціональність по синтезу рухів жестової мови глухих та дактильної абетки на тривимірній моделі людини.

Реалізація можливості генерації анімації процесу мовлення за допомогою жестової мови з використанням віртуальних моделей людей потребує розробки відповідних інформаційних та математичних моделей. Виходячи з цього сформульована наступна постановка задачі:

- 1) потрібно розробити інформаційну та математичну модель для фіксації морфем (мінімально значимих одиниць) жестової мови;
- 2) в рамках цієї моделі потрібно розробити технологію та відповідне програмне забезпечення для отримання, збереження та відтворення жестів;
- 3) потрібно запропонувати алгоритмічне рішення для розрахунку людиноподібних траєкторій руху при переходах від жесту до жесту.

2. Модель для фіксації морфем мови жестів

Процес відтворення жесту трьохмірною моделлю людини можна вважати анімацією з відповідною частотою різних станів скелетної моделі людини.

Скелетна модель людини спрощено відтворює скелет живої людини. Її можна формалізувати як ієрархічну структуру яка складається з поєднаних кінематичних пар, які відтворюють основні кістки скелету людини. Сучасні пакети трьохмірного моделювання (Poser, 3D Studio Max) вміють генерувати анімацію з допомогою віртуальної статичної моделі та інформації про зміну відповідних кутів скелета. Отже для формального опису процесу фіксації жесту можна використовувати множину, що відображає спрощений скелет людини (рис. 1) та зміни значення кутів Ейлера і порядок їх застосування для відповідних кісток цього скелету з плином часу (дискретно, з відповідною частотою (1/30 сек тощо):

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

де H_i – i -та кістка скелету ($i = 0, \dots, N - 1$, N – кількість кісток у скелеті); k – індекс кістки-предка;

$d_i = [x_i, y_i, z_i]^T$ – координати точки – кінця кістки у системі координат, яка пов'язана з початком цієї кістки;

$$M = \{M_i : M_i = \{order_i, \theta_i^j\}\}, \quad (2)$$

де M_i - для i -ї кістки значення кутів Ейлера та порядок застосувань обертань для кістки з плинном часу; $order_i \in \{1, \dots, 6\}$ - порядок застосування обертання

навколо відповідних координатних осей для i -ї кістки (1-XYZ, 2-XZY, 3-YXZ, 4-YZX, 5-ZXY, 6-ZYX);

$\theta = (\theta^j), \theta^j = \{\theta_i^j : \theta_i^j = \{\varphi_{iX}^j, \varphi_{iY}^j, \varphi_{iZ}^j\}\}$ -

множина зміни кутів Ейлера для i -ї кістки з плинном часу ($j = 0, \dots, K - 1$, K – кількість кадрів відтворення руху з заданою частотою).

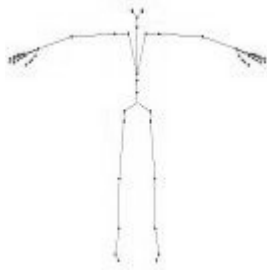


Рис.1. Спрощений скелет людини

3. Технологія для отримання та збереження жестів

Історія тримірної анімації нараховує вже не одне десятиліття, і прогрес у цій області очевидний. В даній роботі, для отримання та збереження жестових одиниць використовується Motion Capture технологія захвату рухів. Типова система Motion Capture є набором пристроїв-показників які прикріплені до тіла людини (або інтегровані в костюм). Основним недоліком існуючих Motion Capture-систем є досить велика вартість як самого обладнання так і відповідних послуг по відцифровці рухів. Авторами запропонована більш проста реалізація цієї технології.

Для отримання множини кутів, які характеризують зміни положення кісток від початкового стану скелета запропоновано наступну технологічну схему:

1. За допомогою трьох камер проводимо зйомку людини, яка відтворює певний жест. Камери встановлені на однаковій фокусній відстані (L) від об'єкту зйомки та фіксують проекції – фронтальну, ліву та праву. Стани скелету характеризують N кадрів зйомки (з заданою частотою кадрів на секунду).

2. Для фіксації i-го стану (i=1..N) скелету відтворимо відповідні кадри зйомки у вигляді відображення на трьох дзеркальних поверхнях типу «трюмо», які розташуємо на задньому плані трьохмірної сцени (рис.2). Скелет розташуємо перед цими дзеркалами і спроектуємо на них ті кістки, кути зміни для яких потрібно визначити (рис. 2). Співставимо точки проекцій кісток скелета з відповідним їм зображенням реальної людини (для фронтального, лівого та правого кадрів), тобто отримаємо нові координати.

3. Для j-го (j=1..N) кадру на парі зображень (фронтальне і ліве або фронтальне і праве) вказуємо точки зчленувань кісток, які змінили положення ($x_{i,new}^{front}, y_{i,new}^{front}$) та ($x_{i,new}^{left\ or\ right}, y_{i,new}^{left\ or\ right}$) та визначаємо їх трьохмірні координати (у системі координат, пов'язаній з кореневою точкою скелету).

Далі, маючи трьохмірні координати нового положення точок зчленування кісток визначимо значення кутів Ейлера, повороти на які переводять кожну кістку з попереднього стану до поточного.

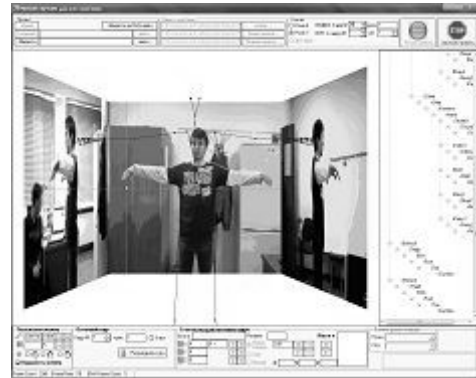


Рис.2. Сцена для фіксації змін кутів у скелета

4. Синтез процесу анімації жестової мови

Для синтезу процесу анімації жестів та міміки трьохмірною моделлю людини запропоновано наступний формальний опис, який використовує відповідні множини параметрів та алгоритми роботи з ними. Трьохмірна модель людини, на якій реалізуватиметься процес анімації жестів та міміки, має наступні атрибути:

$V = \{v_i : v_i = \{x, y, z\}\}$ - множина вершин трикутників для триангуляції поверхні трьохмірної моделі людини;

$N = \{n_i : n_i = \{x, y, z\}\}$ - множина нормалей до вершин;

$T = \{t_i : t_i = \{t, s\}\}$ - множина текстурних координат до вершин;

$V^{ind} = \{V_i^{ind} : V_i^{ind} = \{k_1, k_2, k_3\}\}$ - множина індексів, які вказують порядок побудови трикутників з множини вершин;

$I = \{I_i : I_i = \{img\}\}$ - множина фотографічних зображень елементів моделі – текстури.

Для моделювання скелетної анімації потрібно вміти розраховувати нові значення вершин трикутників (V). Для цього пропонується використовувати механізм скінінга. Скінінг можна означити як алгоритм прив'язки множини вершин трикутників, які означають поверхню моделі до значень кутів скелету. Тоді модель скелетної анімації можна формалізувати наступним чином:

$MH = \{MH_i : HM_i = \{k, \{l_1, \dots, l_m\}, d_i, Glb_i, Order_i\}\}$

– опис спрощеного скелету людини (ієрархія кісток) для реалізації скелетної анімації, де MH_i – i-та кістка скелету (i=0, ..., N-1, N – кількість кісток у скелеті); k – індекс кістки-предка; $\{l_1 \dots l_m\}$ –

множина індексів нащадків кістки, $d_i = [x_i, y_i, z_i]^T$ – координати точки – кінця кістки у системі координат, яка пов'язана з початком цієї кістки; Glb - вектор для визначення координат кістки у глобальній системі координат, $Order_i$ - порядок застосування обертання.

$$Skin = \{Skin_i : Skin_i = \{(IndexVert_i, Weight_i), \dots\}\}$$

- множини вершин, які впливають на поточну вершину при зміні кутів.

Скінінг розраховується для кожної вершини V наступним чином:

$$v_j = \sum_{i=0}^N \{ (v_j * IBM_{H_i} * JM_{H_i}) * JW_{H_i} \},$$

де: n : кількість кісток, пов'язаних з вершиною v ;

IBM_{H_i} : інверсна bind-pose матриця для кістки H_i ;

JM_{H_i} : матриця переміщення для кістки H_i ; JW_{H_i}

: ваговий коефіцієнт для застосування впливу точок кістки H_i на вершину v .

Для моделювання анімації промовляння та емоційного забарвлення запропоновано використовувати механізм морфінгу. Морфінг можна означити як алгоритм плавного переходу від одного стану об'єкта до іншого. При використанні морфінгу використовуються лише опорні стани, з допомогою яких розраховуються проміжні стани і моделюється процес анімації. Модель морфінга міміки промовляння можна формалізувати наступним чином:

Відображення або побудова міміки на обличчі трьохмірної моделі досягається при застосуванні відносного (сегментного) морфінгу до моделі голови і мімічного процесу. Формула відносного морфінгу для M морфів у формалізмі моделі матиме наступний вигляд:

$$V' = V + \sum_{m=1}^M w_m \cdot TM_m,$$

де w_m - вагові коефіцієнти, V – меш базової моделі без морфінгу, TM_m - вхідний меш (морф) для блендингу (blending). Результатом операції є лінійна комбінація мешів моделі і мімік.

Метод сегментного морфінгу дає змогу: формувати декілька виразів обличчя на основі невеликої кількості морфів та змінювати стан обличчя при анімації промовляння. Додатковою перевагою сегментного морфінгу є те, що можна анімувати щелепу незалежно від губ та очі (кліпанія), незалежно від емоційного прояву на обличчі.

5. Синтез проміжних траєкторій-переходів між жестами

Лексичною одиницею у звичайній мові є слово, стале словосполучення або інша конструкція, що спроможна означувати предмети, явища, їх ознаки тощо. У жестовій мові такою лексичною одиницею є жест. Він має вигляд або знака або завершеного руху. В запропонованій моделі жесту-знаку буде відповідати один стан спрощеного скелету людини H з відносно сталим значенням кутів M , а жесту-руху – послідовна зміна значення кутів та порядку застосувань обертань для кожної кістки з плином часу.

Суттєвою відмінністю жестового мовлення є його аморфність. Мовленевий жест містить поняття, але не виражає форму числа, рід, відмінок, час тощо. Тобто із досить обмеженої кількості жестів утворюються їх прості поєднання шляхом аглютинації (склеювання) у відомому порядку:

- діяова особа, предмет – дія ([Я] – [ПРАЦЮВАТИ]),

- дія – заперечення ([ХОТІТИ] – [НІ]),
- предмет – якість,
- стан ([ДИТИНА] – [ХВОРА] [ТЯЖКО])

тощо. Беручи до уваги, що однією з важливих ознак жестової одиниці є локалізація (місце виконання жесту) виникає проблема розрахунку траєкторій-переходів від одного жесту до іншого, що подібні до рухів людей. У моделі жести отримуються шляхом оцифрування жестів реальних людей – носіїв жестової мови. Як екстенсивний, можна запропонувати підхід для отримання траєкторій-переходів шляхом фіксації та оцифрування всіх можливих траєкторій між областями-локалізаціями жестів. При такому підході залишиться задача згладжування стиків між жестом і початком траєкторії переходу. Це пов'язано з тим, що локалізація жеста досить відносне поняття – для групи жестів з однією локалізацією не можливо точно зафіксувати стан спрощеного скелету.

Виходячи з цього сформульована наступна постановка задачі: потрібно, в рамках запропонованої моделі, розв'язати задачу розрахунку траєкторій-переходів між двома довільними жестами.

У загальному випадку, задача планування рухів маніпуляційної системи у області з обмеженнями є досить складною проблемою. Розв'язок оберненої задачі (по відомому значенню трьохмірних координат точки-цілі визначити значення узагальнених координат маніпуляційної системи) є задачею мінімізації функціоналу у просторі з обмеженнями (т.я. таких значень узагальнених координат – множина і потрібно знайти оптимальне значення) і вона не має загального аналітичного розв'язку[3]. Існуючі чисельні підходи [4] пропонують оптимальний пошук траєкторії із множини можливих. В рамках моделі яка використовується потрібно знайти функціональний перехід від кінцевого стану спрощеного скелету (вектор узагальнених координат – значення кутів обертання навколо відповідних осей при кожній ланці (кістці)) одного жесту до початкового стану (інший вектор узагальнених координат) іншого жесту. При цьому цей функціональний перехід повинен породжувати таку траєкторію яка візуально близька до траєкторії людини при аналогічному переході.

Скористаємося тим, що можна організувати планування траєкторій маніпуляційної системи, як аналога процедур навчання й прийняття рішень виконання маніпуляцій і локомоцій вищих організмів [5]. Можна сказати, що вищі організми, для переміщення використовують те керування, для здійснення якого потрібно зробити найменшу роботу тобто найменшу кількість руху. У нашому випадку найменшою кількістю руху буде лінійна зміна кутів для переводу одного стану скелету у інший. Для формалізації жести та з метою подальшого використання механізму аглютинації (для створення жестових конструкцій типу речень) використаємо наступну множину:

$$G = \{G_i : G_i = \{f_i^{beg}, f_i^{end}, S_i^{beg}, S_i^{end}, \theta_i\}\}$$

(3)

де G_i – конкретний жест ($i=0, \dots, \infty$) який відповідає послідовності зміни станів спрощеного скелета з дискретним (з відповідною частотою) плинном часу (1)

(2); f_i^{beg} – номер стану, що відповідає початку жеста;

f_i^{end} – номер стану, що відповідає кінцю жеста;
 $S_i^{beg} = \{(x, y, z)\}$ – координати кінців кісток спрощеного скелета для стану f_i^{beg} (у системі координат, пов'язаний з корневою кісткою спрощеного скелета); S_i^{end} – аналогічно для стану f_i^{end} ; θ_i – множина змін кутів Ейлера (із (2)).

Для розрахунку траєкторії (θ) між жестами G_i та G_m пропонується розрахувати значення узагальнених координат наступним чином:

$$\begin{aligned} \theta_0 &= \theta_{end}(G_i), \\ \theta_i &= \theta_{i-1} + (\theta_{beg}(G_m) - \theta_{end}(G_i)) / k, \quad i = 1, \dots, k-1, \end{aligned} \quad (4)$$

де $\theta_{end}(G_i)$ – значення узагальнених координат для останнього стану (f_i^{end}) жеста G_i ; $\theta_{beg}(G_m)$ – значення узагальнених координат для першого стану (f_i^{beg}) жеста G_m ; k – кількість станів між цими двома жестами.

6. Засіб для відтворення процесу анімації

Для відтворення процесу анімації жестів та міміки трьохмірною моделлю людини створено засіб, який реалізує скелетну (для відтворення жесту) та морфемну (для відтворення міміки промовляння та емоцій) анімацію. Реалізована відповідна програмна функціональність, яка, використовуючи трьохмірне API OpenGL відтворює по означеним атрибутам модель людини (рис. 3) та, з використанням алгоритмів скінінгу та морфінгу відтворює анімацію (рис. 4).



Рис. 3. Відтворення трьохмірної моделі людини



Рис. 4. Кадри анімації жесту «Ти»

Для реалізації української жестової мови було створена програмний засіб (рис. 5), який відтворює методику викладання жестової мови у спеціальних загальноосвітніх школах для глухих дітей. За основу було взято рекомендовану міністерством освіти програму для початкових класів.



Рис. 5. Програма «Українська жестова мова»

Функціонально програма складається з трьох інформаційних блоків (теми, слова та речення) і блоку відтворення жесту віртуальною моделлю. Основним блоком є блок «Теми». У ньому зконцентровані основні методичні відомості, які розглядаються на уроці: що потрібно у цій темі сформувані у учнів, які навички потрібно засвоїти, які особливості морфології жестової мови та які види синтаксичних конструкцій містить матеріал. Приведений список нових жестів по темі що розглядається. Приведені речення з цими словами які містять граматичні та синтаксичні конструкції.

Блок відтворення жесту віртуальною моделлю у цій реалізації несе особливу функцію. Саме з допомогою нього стало можливим демонструвати у навчальному процесі динаміку жесту. Беручи до уваги, що жести оцифровуються носіями жестової мови, вони, по суті, стають стандартами відтворення жесту. Реалізована у блоці можливість покадрового показу жесту слугує засобом, з допомогою якого стане можливим вивчення жесту без особливостей показу конкретним викладачем. Це стане основою того, що вивчені дітьми жести будуть однаковими для різних шкіл, регіонів.

7. Висновки

Використовуючи модель для фіксації рухів, які відтворюють українську жестову мову, була оцифрована множина із 100 жестів. Відтворення жестів із цієї множини показало спроможність запропонованої технології досить реалістично відтворювати на трьохмірній моделі рухи, що отримані з відео-зображення конкретної людини – носія жестової мови.

Було реалізовано програмне забезпечення, яке відтворює декілька уроків української жестової мови.

Розроблене програмне забезпечення може стати основою створення стандарту жестової мови. Стане можливим вирішити проблему відмінностей для одних і тих самих жестів, яка виникає від того, що діти вивчають жест, який містить особливості конкретного викладача.

Подальші дослідження направленні на вдосконалення запропонованої технології:

- наповнення бази даних жестів основною множиною жестів української жестової мови – створення стандарту жестової мови;
- створення засобу для семантичного зв'язування речень українською мовою з реченнями на жестовій мові.

8. Посилання

- [1] Кривонос Ю.Г. Інформаційна технологія невербального спілкування людей з вадами слуху/ Кривонос Ю.Г., Крак Ю.В., Бармак О.В. [та ін.] // Штучний інтелект. – 2008. №3. – ст. 325-331
- [2] Кривонос Ю.Г. Інформаційна технологія для моделювання української мови жестів / Кривонос Ю.Г., Крак Ю.В., Бармак О.В. [та ін.] // Штучний інтелект. – 2009. №3. – ст. 186-198
- [3] Кириченко М.Ф. Оптимізація маніпуляційних роботів / Кириченко М.Ф., Крак Ю.В., Сорока Р.О. – Київ: Либідь, 1990. – 144 с
- [4] Бармак А.В. Информационная компьютерная технология для моделирования и управления манипуляционными и виртуальными системами/ Бармак А.В.// Проблемы управления и информатики. – 2003. №5. – с.143-154
- [5] Бернштейн Н.А. Очерки по физиологии движений и физиологии активности / Бернштейн Н.А. - Медицина, М., 1966. 350 с.