ДОПОВІДІ
НАЦІОНАЛЬНОЇ
АКАДЕМІЇ НАУК
УКРАЇНИ

5 • 2010

УДК 004.8

Академік НАН України **Ю. Г. Кривонос, Ю. В. Крак, О. В. Бармак, А. С. Тернов**

Розпізнавання міміки губ при промовлянні слів українською мовою

Пропонується метод розпізнавання міміки губ при промовлянні слів українською мовою на основі синтезованої математичної моделі станів губ конкретної людини. Новизна і практична цінність полягає у створенні систем навчання правильній артикуляції при промовлянні слів українською мовою.

Однією з проблем при спілкуванні людей з вадами слуху з іншими людьми є вміння розпізнавати артикуляцію розмовної мови, тобто вміння "читати по губах". З огляду на це задача читання по губах є альтернативою мовного спілкування. Специфіка артикуляції української мови потребує розробки власних методів для розпізнавання міміки. В той же час, фонетичний принцип української мови дозволяє побудувати загалом однозначний зв'язок між розпізнаною мімікою та відповідною фонемою, з якої складається слово [1].

У даному повідомленні для створення системи навчання правильній артикуляції при промовлянні слів українською мовою вперше пропонується синтезована математична модель станів губ конкретної людини. Для цього зроблено перехід від простору фотографічних зображень облич людини [2] (з процесом промовляння) до векторного простору характеристичних ознак. Цей перехід проходить у декілька етапів.

Етап 1. Виділення на зображенні внутрішніх контурів губ

$$\operatorname{Im} L \to D. \tag{1}$$

Тут $\operatorname{Im} L = \{I_k \colon I_k \in FSV\}$ — впорядкована множина ключових кадрів з відеопотоку FSV (Face Speech Video), сформованого при зйомці мімічних проявів, а саме станів губ, на обличчі людини при промовлянні слів українською мовою (індекс $k = \overline{1,N}$ відповідає за порядковий індекс кадру у вибраній послідовності, де N — кількість ключових кадрів); $I_k = \{\mathrm{col}_{ij}^k\}_{i,j=1}^{m,n}$, $i = \overline{1,m}$, $j = \overline{1,n}$, — зображення розміру $m \times n$ обличчя з мімічним станом губ при промовлянні слів українською мовою, де m та n — відповідно довжина та ширина зображення; $\mathrm{col}_{ij}^k = I_k(i,j)$ — колір пікселя в системі RGB з координатами (i,j) на зображенні I_k; $D = \{D_k \colon D_k = \{d_{\mathrm{top}}^k, d_{\mathrm{bot}}^k\}\}$ — множина контурів губ, де D_k — пара точкових кривих — контурів губ (верхній d_{top}^k та нижній d_{bot}^k) для k-го кадру.

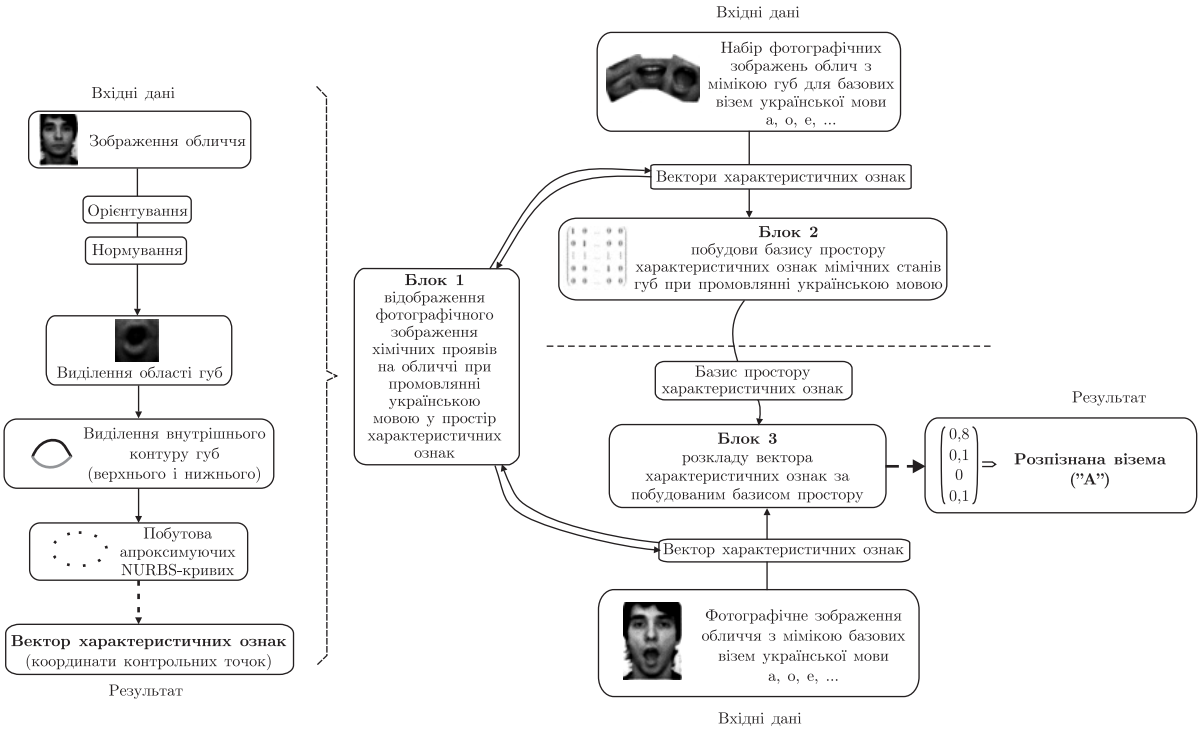

Рис. 1

Етап 2. Апроксимація отриманої точкової кривої *з* внутрішнім контуром губ за допомогою нерівномірних базисних сплайнів (NURBS) — отримання вектора характеристичних ознак

$$D \to P, \tag{2}$$

де $P = \{v_k \colon v_k^i \in H,\ i = \overline{1, M}\}$ — простір характеристичних ознак; H — характеристичні ознаки об'єкта дослідження; v_k — характеристичний вектор; v_k^i — його координати; M — розмірність простору P.

Таким чином, математичною моделлю мімічних проявів губ при промовлянні буде векторний простір керуючих точок NURBS-кривих:

$$\begin{aligned} &P = \{v \colon v = (v^{1,(*)}, v^{2,(*)}, v^{3,(*)}, v^{4,(*)})\},\\ &v^{1,(*)} = (x_0^{p_{\text{top}}}, \ldots, x_{n_{\text{top}}-1}^{p_{\text{top}}}), \qquad v^{2,(*)} = (x_0^{p_{\text{bot}}}, \ldots, x_{n_{\text{bot}}-1}^{p_{\text{bot}}}),\\ &v^{3,(*)} = (y_0^{p_{\text{top}}}, \ldots, y_{n_{\text{top}}-1}^{p_{\text{top}}}), \qquad v^{4,(*)} = (y_0^{p_{\text{bot}}}, \ldots, y_{n_{\text{bot}}-1}^{p_{\text{bot}}}), \end{aligned} \tag{3}$$

де $v \in P$ — це вектор координат опорних точок $p_j^{p_{\text{bot}}}$ та $p_j^{p_{\text{top}}}$ апроксимуючих NURBS кривих $p_{\text{bot}}(u)$, $p_{\text{top}}(u)$, а n_{bot} і n_{top} — кількість контрольних точок для NURBS кривих $p_{\text{bot}}(u)$, $p_{\text{top}}(u)$ відповідно. Тоді розмірність простору P визначається як $M = 2(n_{\text{top}} + n_{\text{bot}})$.

Схема розпізнавання міміки при промовлянні слів українською мовою наведена на рис. 1.

На рис. 1 блок 1 відповідає за попередню обробку вхідної візуальної інформації та перетворення її у простір характеристичних ознак (3). Блок 2 містить у собі алгоритми побудови

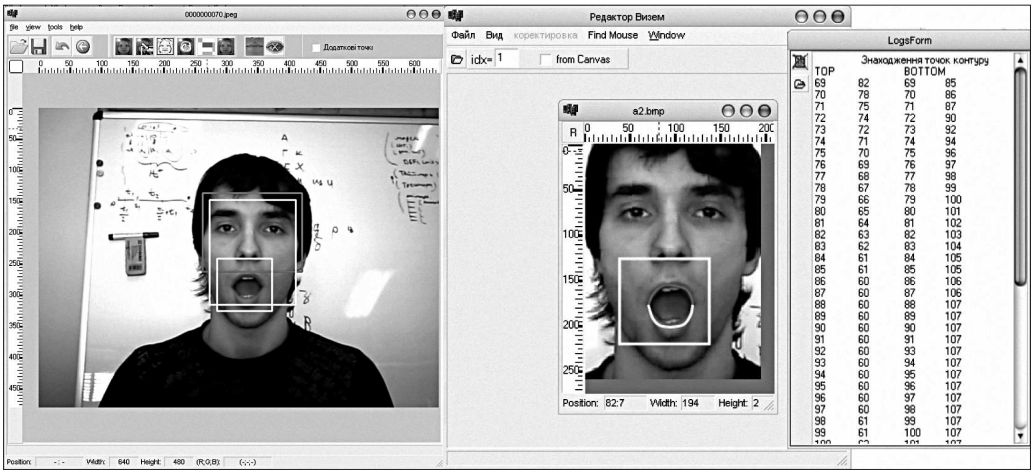

Рис. 2

базису простору характеристичних ознак та оцінки його якості. На виході будується базисна матриця A розміром $M \times L$ простору характеристичних ознак P

$$A = \begin{pmatrix} v_2^{\text{base}_1} & v_2^{\text{base}_2} & \dots & v_2^{\text{base}_L} \\ v_2^{\text{base}_1} & v_2^{\text{base}_2} & \dots & v_2^{\text{base}_L} \\ \dots & \dots & \dots & \dots \\ v_M^{\text{base}_1} & v_M^{\text{base}_2} & \dots & v_M^{\text{base}_L} \end{pmatrix},$$

де $v_j^{\text{base}_i} \in P$, $j = \overline{1, M}$, $i = \overline{1, L}$, L — кількість базисних векторів. В цьому випадку під базисом простору P розуміється набір характеристичних векторів базових мімік або базових візем з $\operatorname{Im} L$. Усього для української мови таких візем шістнадцять, враховуючи стан спокою.

У третьому блоці відбувається розклад вектора характеристичних ознак b, побудованого для вхідного зображення, яке розпізнається, за отриманим базисом. Задача розкладу зводиться до задачі знаходження всіх векторів x, для яких виконується

$$Ax = b. \tag{4}$$

При невиконанні умови $\det(A^T A) > \varepsilon > 0$ найбільш надійним методом для розв'язання подібних задач є метод сингулярного розкладу SVD [3]. На практиці для використання SVD вводять поріг τ близькості до нуля сингулярних чисел, який відображає помилки в початкових даних та обчисленнях. Тоді наближений розв'язок задачі (4) шукається так:

$$x = A^+ b = V \Sigma' U^T b, \tag{5}$$

де $\Sigma' = \operatorname{diag}(\sigma'_1, \sigma'_2, \dots, \sigma'_n)$, $\sigma'_j = 1/\sigma_j$, для $\sigma_j \geqslant \tau$ і $\sigma'_j = 0$ для $\sigma_j < \tau$; V, U^T — матриці з ортонормованими стовпцями.

Результатом роботи пропонованого алгоритму є вектор розкладу, на основі якого приймається рішення про відповідність вхідного вектора конкретним базовим мімікам при промовлянні українською мовою.

Таким чином, для реалізації запропонованого підходу до розпізнавання міміки губ було створене оригінальне програмне забезпечення (рис. 2). Проведені дослідження підтвердили

ефективність і дієвість такого підходу, де, окрім висновку про належність досліджуваної віземи до відповідної базисної віземи, виконується структурний аналіз вхідних даних (зображень губ людини при промовлянні слів української мови), змістом якого є визначення вкладу кожної базової віземи. Запропонована технологія має практичну цінність для створення систем навчання правильній артикуляції при промовлянні слів українською мовою.

1. *Українська* мова. Енциклопедія / Під ред. В. М. Русанівського. – Київ: Українська енциклопедія ім. М. П. Бажана, 2000. – 750 с.
2. *Крак Ю. В., Кривонос Ю. Г., Тернов А. С.* Локалізація і врахування особливостей обличчя людини для задачі розпізнавання за портретною фотографією // Штучний інтелект. – 2007. – № 3. – С. 229–236.
3. *Форсайт Дж.* Машинные методы математических вычислений: Пер. с англ. Х. Д. Икрамова. – Москва: Мир, 1980. – 277 с.

Інститут кібернетики ім. В. М. Глушкова НАН України, Київ

Надійшло до редакції 20.10.2009

Academician of the NAS of Ukraine **Yu. G. Kryvonos, Yu. V. Krak, O. V. Barmak, A. S. Ternov**

Lips-reading recognition during Ukrainian language pronunciation

A method for lips-reading recognition during the Ukrainian language pronunciation has been proposed. The method is based on a mathematical model synthesized from lips positions of a specific man. The novelty and the practical value consist in the creation of a proper articulation learning system.