

АЛГОРИТМ ВИЗНАЧЕННЯ ФОРМИ ГУБ ПІД ЧАС АРТИКУЛЯЦІЇ ДЛЯ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ

© Давидов М.В., Нікольський Ю.В., Тиханський С.М., 2010

Розглянуто проблему розпізнавання артикуляції на зображенні обличчя. Пропонується метод розпізнавання артикуляції, оснований на методі активної моделі форми ASM. Метод протестований на відеозображеннях української жестової мови та на базі облич “BioID”.

Ключові слова: жестова мова, артикуляція, розпізнавання образів, розпізнавання обличчя, активна модель форми.

The problem of articulations recognition on the face images is considered. Articulation recognition method that is based on the active shape model (ASM) is proposed. It was tested with the Ukrainian sign language video records and with the samples from the face database “BioID”.

Keywords: sign language, articulation, image recognition, face recognition, active shape model.

Постановка проблеми

В Україні гостро стоїть проблема нестачі ресурсів для навчання жестовій мові як людей з вадами слуху, так і тих, в кого виникає потреба у спілкуванні жестовою мовою. У цій проблемі можна виділити два аспекти: по-перше, – проблему навчання людей із вадами слуху та їхніх близьких жестовій мові, по-друге, – проблему спілкування таких людей із людьми, які не володіють жестовою мовою. Ці проблеми залишаються актуальними впродовж тривалого часу, проте сучасний розвиток інформаційних технологій дає змогу створювати нові засоби для навчання жестовій мові.

Автори статті розробляють спеціальний тренажер, який можна використати для навчання жестовій мові. Цей тренажер містить спеціальні інструменти, за допомогою яких можна оцінювати якість виконання жестів людиною, що дає змогу контролювати процес навчання жестовій мові.

В українській жестовій мові важливу роль відіграє артикуляція. Артикуляція як елемент мови допомагає розрізнити різні слова, які показують одним жестом, надавати словам жестової мови емоційних відтінків, сприймати жестову мову людям, які краще читають по губах, ніж розуміють рухи рук.

У статті наведено методи та загальний підхід, який використано для виявлення та ідентифікації елементів артикуляції на відеозображенні реального часу, та результати застосування розробленого алгоритму для автоматичного розпізнавання форми губ на зображенні обличчя людини за артикуляцією.

Аналіз останніх публікацій

Способи використання елементів зображення, або візем, для навчання людей з вадами слуху в Україні досліджують в Інституті кібернетики ім. В. М. Глушкова НАН України. У праці [1] описано методи *синтезу* артикуляції під час промовляння слів. На основі експериментальних даних виділено 16 класів візем української мови для голосних та приголосних звуків. Деякі віземи відповідають кільком звукам, що дає змогу стверджувати про неповноту інформації, яку отримують лише візуально. Під віземою тут розуміємо зображення форми губ, якого вони набувають під час промовляння літери.

У праці [2] наведено методи визначення параметрів обличчя. Ці методи використовують активні моделі вигляду (ААМ) та активні моделі форми (ASM), які призначені для знаходження ключових точок як характеристик обличчя. У методі ААМ для обчислення величини відмінності між моделлю та фактичним зображенням використано повне зображення обличчя, а в методі ASM –

лише частини зображення, які прив'язані до ключових точок моделі. Також у праці [2] описано загальні принципи побудови систем на основі ASM або AAM. Конкретні задачі можуть мати істотні відмінності від загальних вимог до методів. Тому для таких задач ці методи треба адаптувати із урахуванням вимог до якості розпізнавання, швидкодії, умов зовнішнього середовища, параметрів обличчя, які необхідно отримати.

У праці [3] описано метод покращання результатів опрацювання кадру відеозображення обличчя на основі використання інформації з попередніх кадрів. У цьому методі для кожного кадру визначають межі обличчя на зображенні та порівнюють їх з межами обличчя на попередньому кадрі. На підставі цих даних визначають зміну положення обличчя, що дає змогу покращити якість роботи алгоритму ASM. Такий підхід виявляється ефективним для однорідного тла, колір якого відмінний від кольору шкіри.

Формулювання цілей статті

Автоматичне візуальне розпізнавання жестової мови та перекладу її на текст вимагає розв'язання низки задач, пов'язаних із ідентифікацією положення тіла людини в просторі, форми кисті руки, артикуляції, міміки, окремих жестів, семантичного та прагматичного лінгвістичного аналізу, синтезу письмового перекладу тощо.

У цій статті описано загальний підхід, який використано для розв'язання задачі розпізнавання артикуляції людини, яка розмовляє жестовою мовою. Для цього спочатку визначено роль артикуляції в українській жестовій мові та сформульовано вимоги до системи розпізнавання артикуляції. Після цього запропоновано оптимальний за швидкістю та якістю розпізнавання метод визначення параметрів обличчя, що визначають артикуляцію людини.

Роль міміки та артикуляції у жестовій мові

Міміка – це рухи м'язів обличчя людини. Якщо людина розмовляє жестовою мовою, то міміка відіграє роль, яка аналогічна інтонаціям у словесній мові для надання словам та виразам емоційних відтінків. Іноді можна одними й тими самими жестами показати зовсім різні поняття лише засобами міміки. Міміка також дає змогу передавати емоції тексту. Наприклад, якщо показують слово “сумний”, то міміка повинна передавати цей настрій.

Отже, артикуляція жесту виконує такі функції:

- деякі люди, особливо ті, що втратили слух вже в дорослому віці, краще сприймають інформацію з руху губ, ніж з жестів;
- люди з вадами слуху під час розмови здебільшого дивляться на обличчя співрозмовника;
- рух губ дає змогу розрізняти слова, які показано однаковими жестами.

Отже, розпізнавання артикуляції та міміки може мати вирішальне значення для розпізнавання багатозначних жестів, отримання семантичних даних про текст, або для підвищення точності розпізнавання жестової мови.

Система розпізнавання артикуляції

Визначимо такі терміни:

- Ключові точки – це виділені розробником точки на зображенні обличчя, взаємне розміщення яких характеризує форму обличчя.
- Модель обличчя – це набір координат ключових точок, що визначають форму обличчя.
- Модель зображення обличчя – це набір параметрів, за якими здійснюється порівняння ключових точок моделі обличчя із зображенням обличчя.
- Модель віземи – це модель обличчя з позначенням віземи, якій відповідає ця модель обличчя.
- Навчальний приклад – задані зображення обличчя та параметри моделі обличчя, що відповідає цьому зображенню.

Загальну схему розробленої системи розпізнавання артикуляції показано на рис. 1.

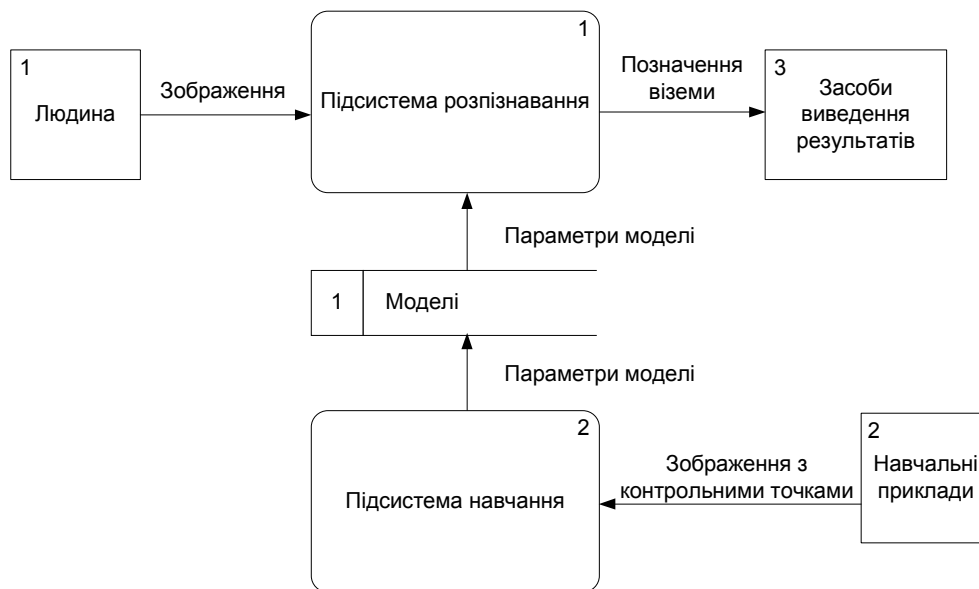


Рис. 1. Загальна схема системи розпізнавання артикуляції

Система розпізнавання взаємодіє з такими зовнішніми сутностями:

- людина, що розмовляє мовою жестів перед відеокамерою;
- навчальні приклади, за допомогою яких побудовано моделі обличчя та моделі зображень обличчя;
- засоби виведення результатів, наприклад, монітор, на який виводиться позначення розпізнаних візем або інша система, яка використовує результати роботи системи розпізнавання.

Система розпізнавання поділяється на підсистему навчання та підсистему розпізнавання. Ці підсистеми є незалежними, але використовують спільне сховище даних “Моделі”.

Вхідними даними підсистеми навчання є навчальні приклади із зображеннями обличчя та множиною координат ключових точок. Крім того, для кожного зображення обличчя визначено візему, якій відповідає це зображення, та задано позначення цієї віземи. В результаті роботи цієї підсистеми в сховище даних “Моделі” записано параметри моделі зображення обличчя та моделей візем.

Підсистема навчання обробляє навчальні приклади та визначає параметри моделей, які використовуються під час розпізнавання артикуляції. Перетворення даних з навчальних прикладів на готові моделі відбувається у такій послідовності.

1. Попередня фільтрація вхідних зображень. Застосовується, щоб уникнути впливу факторів зовнішнього середовища на результати розпізнавання. На цьому кроці, наскільки можливо, усувають шуми відеокамери, змінні параметри освітлення тощо.

2. Отримання параметрів зображення. Визначаються параметри моделі обличчя, моделі зображення обличчя та моделі віземи для конкретного навчального прикладу.

3. Узгодження моделей різних навчальних прикладів. Здійснюється побудова універсальної моделі зображення обличчя та універсальних моделей візем (для кожної віземи окрема модель) на основі моделей, одержаних з навчальних прикладів.

Діаграму потоків даних підсистеми навчання показано на рис. 2.

Підсистема розпізнавання безпосередньо реалізує алгоритми, використовуючи моделі зі сховища даних “Моделі”. Вхідними даними цієї підсистеми є зображення з відеокамери, а вихідними даними – позначення розпізнаної віземи.

Ця підсистема отримує зображення з відеокамери, обробляє його та класифікує відповідно до зображеної віземи. Процес класифікації складається з таких кроків.

1. Попередня фільтрація вхідного зображення. Виконується для зменшення впливу умов зовнішнього середовища на результати розпізнавання. Під час попередньої фільтрації усуваються шуми відеокамери, зменшується вплив змінних параметрів освітлення тощо.

2. Визначення моделі обличчя. Знаходять модель обличчя для обличчя на вхідному зображенні. Для цього використовують модель зображення обличчя зі сховища даних “Моделі”.

3. Визначення віземи. Здійснюється класифікація моделі обличчя, отриманої на попередньому кроці. Для цього модель обличчя порівнюють з моделями візем зі сховища даних “Моделі”.

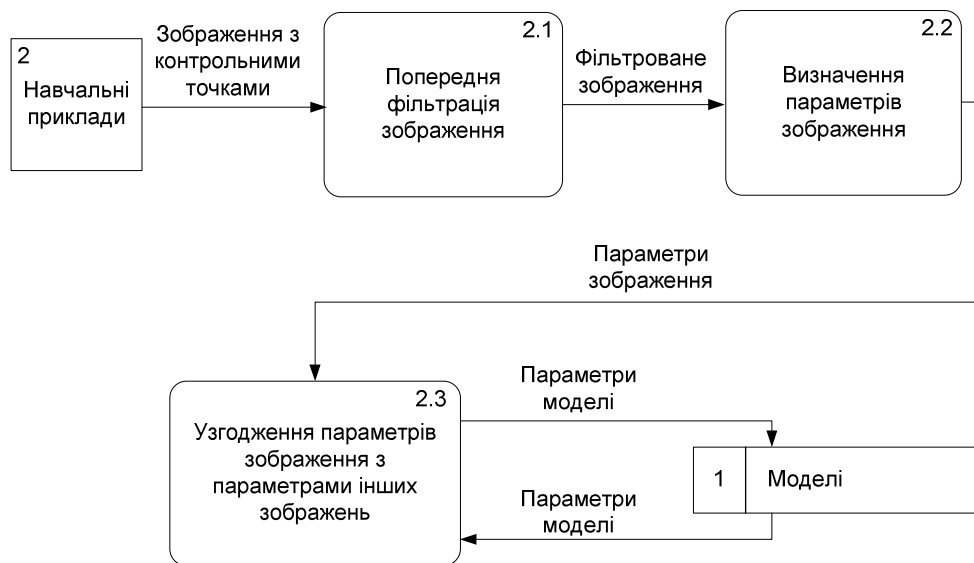


Рис. 2. Діаграма потоків даних підсистеми навчання

Діаграму потоків даних підсистеми розпізнавання артикуляції показано на рис. 3.

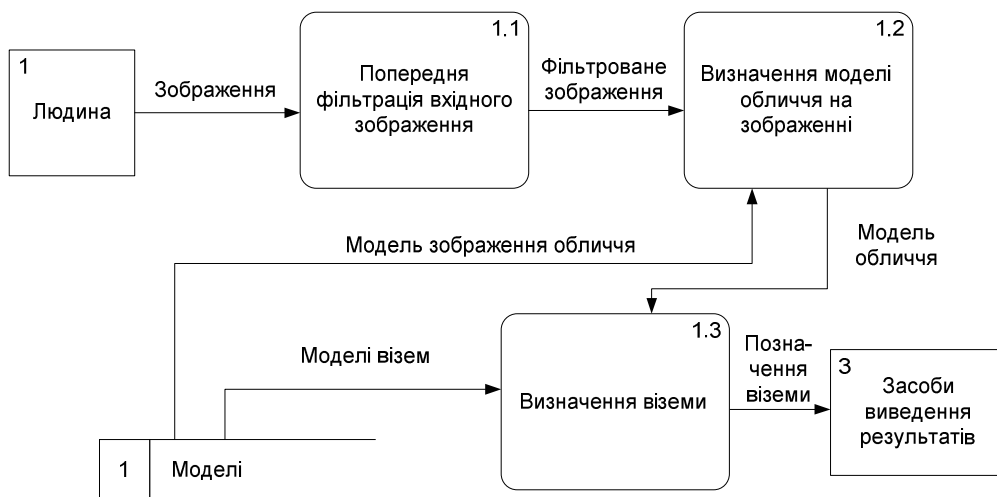


Рис. 3. Діаграми потоків даних для підсистеми розпізнавання артикуляції

Метод розпізнавання артикуляції та міміки обличчя

Розпізнавання параметрів артикуляції та міміки обличчя виконують або із допомогою статистичних методів [4], або із використанням 3D-моделювання [1], або моделювання з застосуванням ASM (Active Shape Model) [2] та AAM (Active Appearance Model) [2].

Серед цих методів як базовий для реалізації використано метод ASM, переваги якого полягають у ефективному розпізнаванні форми об'єкта на зображенні, навіть у разі сильної деформації об'єкта [2].

Розглянемо, як розв'язується задача розпізнавання артикуляції з використанням модифікованого методу ASM. Вхідними даними алгоритму є зображення, яке отримано з відеокамери, а вихідними даними є позначення ідентифікованої віземи. Алгоритм розв'язання цієї задачі складається з таких етапів.

1. Приведення розміру зображення до стандартних розмірів масштабуванням.
2. Попередня фільтрація зображення. Виконується для зменшення впливу змінних параметрів середовища (освітлення, колір шкіри тощо) на результат розпізнавання застосуванням

фільтрів, які дають змогу виконати нормалізацію гістограми зображення, знайти контури, виділити кольорову компоненту яскравості та медіанний фільтр.

3. Знаходження моделі обличчя. За допомогою методу ASM здійснюється пошук ключових точок на зображенні.

4. Ідентифікація віземи. Використано метод найменших квадратів для обчислення попарних відстаней усіх ключових точок, за якими знаходять модель віземи, найбільш схожу на модель обличчя зображення з відеокамери.

Аналіз реалізації системи розпізнавання артикуляції

Систему розпізнавання артикуляції реалізовано мовою програмування C++ з використанням бібліотеки OpenCV, з якої використано функції отримання зображення з відеокамери та пошуку обличчя на зображенні. Тестування системи здійснено на тестових даних з відкритої бази облич "BioID" [6] та бази відеозаписів слів української мови, створеної розробниками системи.

Відкриту базу облич "BioID" використано для тестування якості розпізнавання форми губ. Ця база містить 1521 зображення. Кожне зображення є фронтальним виглядом однієї з 23 осіб.

Тестування якості розпізнавання форми губ здійснювали у такій послідовності:

1. Вибрано одну з людину з 23, зображення яких є у базі, та всі зображення цієї людини поділено на дві групи: групу навчальних прикладів та контрольних прикладів. Розподіл виконано у випадковий спосіб у пропорції 1:4.

2. Навчено систему так, щоб вона успішно розпізнавала навчальні приклади.

3. Протестовано якість розпізнавання контрольних прикладів. Для кожного контрольного прикладу візуально оцінено правильність визначення ключових точок. Якщо ключові точки розставлені правильно, то результат тестування цього контрольного прикладу вважається успішним, інакше – неуспішним.

Протестовано зображення десяти людей. Успішний результат розпізнавання отримано для 81% зображень.

Приклад успішного та неуспішного результату тестування показано на рис. 4.

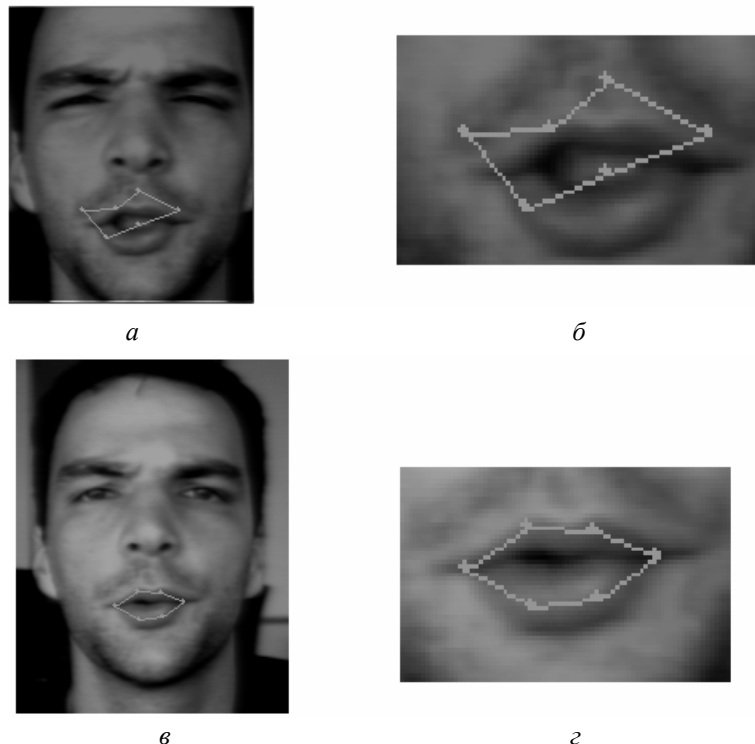


Рис. 4. Результату тестування під час розпізнавання форми губ на обличчі з бази облич BioID: а – неуспішний результат розпізнавання; б – неуспішний результат розпізнавання (збільшене зображення губ); в – успішний результат розпізнавання; з – успішний результат розпізнавання (збільшене зображення губ)

Автори статті створили базу відеозаписів артикуляції слів української мови, яка складається з 50 відеозаписів. За допомогою цієї бази здійснено тестування якості розпізнавання візем.

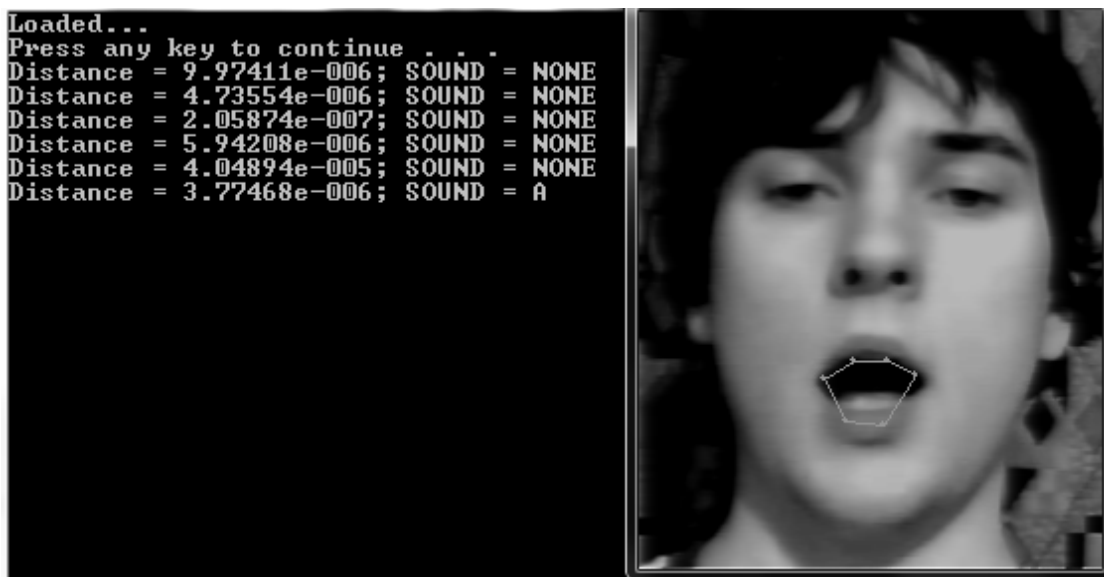
Тестування якості розпізнавання візем відбувалось у такій послідовності.

1. Для кожної віземи, яку система повинна розпізнавати, створено її зображення. Для тестування вибрано шість візем.

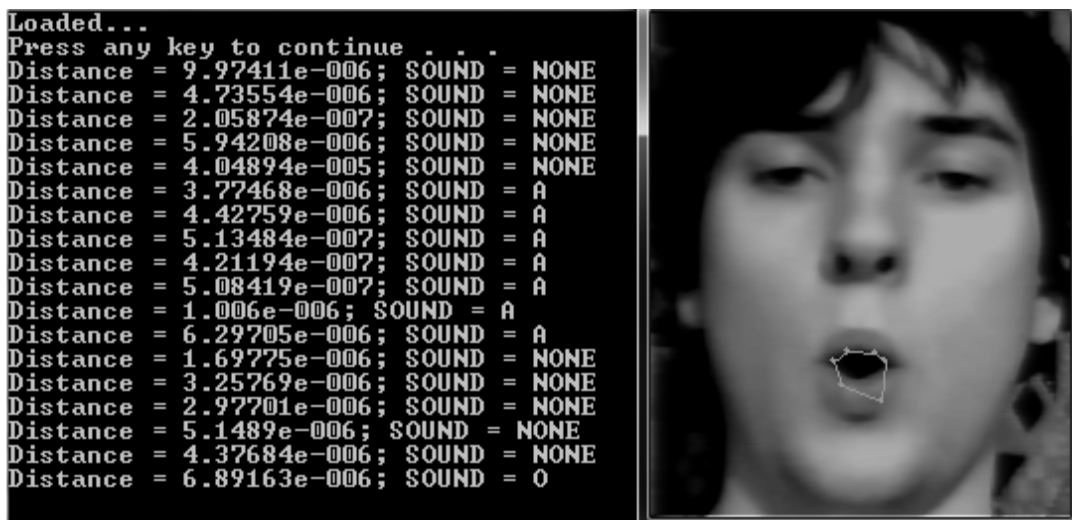
2. Формуються навчальні приклади із зображень візем, отриманих на попередньому кроці, та вручну розставлених контрольних точок на цих зображеннях.

3. Для кожного відеозапису визначається, чи успішно розпізнано всі віземи слова. Віземи вважаються успішно розпізнаними, якщо для кожної віземи в слові правильно розпізнано більше від половини кадрів із зображенням цієї віземи.

Протестовано усі відеозаписи з бази відеозаписів. Успішний результат тестування отримано для 74 % відеозаписів. Приклад розпізнавання візем слова «Тато» показано на рис. 5.



а



б

Рис. 5. Розпізнавання візем слова «Тато».
Ліворуч – записи про визначені віземи (окремий кадр).
Праворуч – кадр з відеозапису: а – візема «А»; б – візема «О»

Висновки

Запропоновані в статті алгоритми програмно реалізовано у комп'ютерній системі розпізнавання артикуляції. Часові параметри програмної реалізації уможливають її використання у системах реального часу. Для тестування роботи алгоритмів використано тестові набори даних із зображеннями облич різних людей. Розроблена система показала високий відсоток розпізнавання артикуляції.

Планується інтеграція системи у навчальний тренажер української жестової мови для контролю правильності положення губ людини під час виконання жесту. Розроблені алгоритми можна використовувати в системах для розпізнавання слів за артикуляцією губ. Це – системи-тренажери артикуляції, системи автоматичного перекладу з жестової мови на словесну, інші системи аудіовізуального розпізнавання мови.

1. Інформаційна технологія для моделювання української жестової мови / Ю.Г. Кривонос, Ю.В. Крак, О.В. Бармак, А.С. Тернов, Б.А. Троценко // *Искусственный интеллект*. – 2009. – № 3. – С.186–197. 2. Stan Z. Li. *Handbook of face recognition* / Stan Z. Li, Anil K. Jain. – Springer-Verlag London Limited. – 2005. – 395 с. 3. Suat Akyol *Evaluation of ASM head tracker for robustness against occlusion* / Suat Akyol, Jörg Zieren. – *Proceedings of the International Conference on Imaging Science, Systems, and Technology (CISST 02), June 24-27, Las Vegas, Nevada. Volume I, CSREA Press, ISBN 1-892512-95-5*. 4. Давидов М.В. Вибір ефективного методу опрацювання зображень на основі еталону для ідентифікації елементів жестової мови / М.В. Давидов, В.В. Пасічник, Ю.В. Нікольський // *Автоматизированные системы управления и приборы автоматики. Всеукраинский межведомственный научно-технический сборник*. – 2008. – № 139. — С. 59–68. 5. Давидов М.В. Методи та засоби опрацювання зображень реального часу для ідентифікації елементів жестової мови / М.В. Давидов, Ю.В. Нікольський // *Штучний інтелект. Державний університет інформатики і штучного інтелекту*. – Донецьк, 2008. – № 1. – С. 131–138. 6. *BioID face database* [Електронний ресурс]: <http://www.bioid.com/support/downloads/software/bioid-face-database.html>.