

А. М. Глибовець

Метод розпізнавання слів з відео потоку на основі аналізу просторового розміщення характеристичних точок губ

Національний університет «Кієво-Могилянська академія», Київ, Україна.

У цій роботі розглянуті сучасні методи та підходи до вирішення задачі розпізнавання природної мови з відеопотоку. Запропоновано новий підхід та програмну реалізацію алгоритму, який бере за атомарну частину розпізнавання саме слово. Новий метод розпізнавання слів з відеопотоку базується на аналізі просторового розміщення ключових точок губ. Результати тестування програмного застосунку реалізації модуля дозволяють рекомендувати його використання в системах, де можливе безпосереднє навчання системи користувачем.

Опишемо основні концепції розробленої програмної системи розпізнавання слова із відео потоку. Для тестування розпізнавання слів була створена своя база даних, її особливості описані далі. Правильність розпізнавання слів перевірялась за допомогою контрольного запису.

Програма складається з таких основних модулів: розпізнавання обличчя людини, локалізацію губ, виокремлення характеристик з кожного кадру, класифікації вхідних параметрів та отримання результату. Архітектура системи приведена на рисунку 1. Для перших двох етапів програми був використаний метод який був реалізований за допомогою інструментів OpenCV[1] та бібліотеки dlib[2].

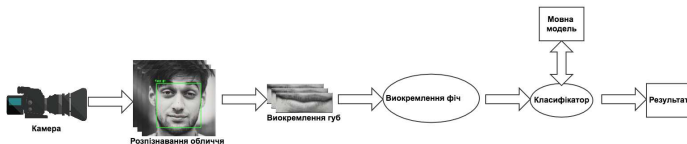


Рис. 1. Архітектура системи

Знаходження ключових точок обличчя є підзадачою розпізнавання образів. Отримавши вхідне зображення, розпізнавач образів (shape predictor) знаходить ключові точки, які визначають

об'єкт. В контексті розпізнавання обличчя, алгоритм розпізнає основні ключові елементи обличчя: очі, рот, ніс, брови, підборіддя та багато інших точок, які допомагають класифікувати основні характеристики обличчя.

Для вирішення задачі знаходження обличчя та губ було вирішено використати `deep learning`[3] та бібліотеку `dlib`.

Бібліотека `dlib` використовує навчальну вибірку зображень облич, що були попередньо розмічені вручну, із визначеними конкретними (x, y) -координатами областей, що оточують кожну характеристику обличчя. Точки шукаються на основі ймовірності між двома вхідними парами пікселів.

На основі заданої навчальної вибірки будуються дерева лінійних регресій, які навчаються визначати основні точки на обличчі безпосередньо з інтенсивності вхідних пікселів. Як результат, ми отримуємо зображення (рис. 3), на якому знайдені усі ключові точки обличчя з високою точністю.

В початковій навчальній вибірці детектор який закладений в основу `dlink`, який використовує 68 (x, y) координат, які задають структуру обличчя.

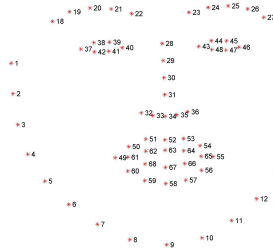


Рис. 2. Ключові точки обличчя

Після знаходження детектором усіх ключових точок обличчя ми можемо виокремити губи з загального зображення просто звернувшись до них за відомими індексами.

Виокремлення характеристик – задача непроста, кожні губи унікальні, кожні мають свої особливості, тому для узагальнення губ були використані підходи, які базуються на усередненні відношення висоти до ширини вуст та усереднення просторових координат ключових точок губ.

Для розпізнавання слова не потрібно знати кожну ознаку губ користувача, а лише деяку узагальнену інформацію. Так в першому випадку було використано співвідношення ширини до дов-

жини, як характеристику, яка узагальнено характеризує положення губ для великої групи людей.

Першим способом, який був запропонований став пошук мінімальної помилки. На перший погляд, спосіб здається досить легким для реалізації, але не потрібно забувати, що одна і та сама людина вимовляє слова з різною швидкістю, в залежності від емоційного стану чи інших факторів. Тому перша задача, яку потрібно було вирішити – стала задача зведення векторів до однакової довжини. Запропонований алгоритм працює наступним чином.

Обираємо вектор з мінімальною довжиною. Для вектора з більшою довжиною виокремлюємо підвектори з довжиною меншого вектора, зберігаючи послідовність елементів. Рахуємо сумарну помилку для кожного підвектора. Запам'ятовуємо кандидата з найменшою помилкою. Ітеруємо по всіх словах нашого словника, слово з найменшою сумарною помилкою і буде шукане слово. Алгоритм пошуку мінімальної помилки з відкиданням максимального та мінімального значень відрізняється від попереднього лише видаленням відповідного найбільшого та найменшого значення з обох векторів на першому кроці.

Інший підхід, який також використовує співвідношення сторін, як основну характеристику, полягає в тому, що замість пошуку підмножини з більшого вектора – обчислюють середнє значення співвідношення сторін для кожного слова і також шукають в словнику за мінімальною помилкою.

Другий підхід до визначення слів, який було досліджено, базувався на основі положення 20 точок, які визначали контури губ. Ці точки можна побачити на рисунку 3.

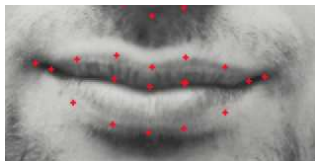


Рис. 3. Характеристичні точки

Першим способом який, базувався на 20 контрольних точках був спосіб усереднення усіх 20 координат в кадрі в одну точку - центроїд. Такий процес виконувався для кожного кадру.

Алгоритм усереднення центроїдів кадрів усього слова відрізняється від попереднього лише тим, що замість порівняння центро-

їдів кожного кадру слова, шукаються центроїди кожного кадру і далі усереднюються по всьому слову. Тобто розглядаються середні значення центроїдів кадрів.

У підході усереднення по положені точки розглядаються усі 20 точок рота, але на відмінно від попереднього методу замість того, щоб рахувати центроїд рахується середнє положення контрольних точок під час вимови слова. Після того, пораховано середню координату точки під час вимови, шукається слово в нашому словнику з найменшою помилкою.

Виходячи з результатів дослідження, було показано, що система показує найкращі результати при використанні просторового підходу з методом визначення слова, який базується на усередненні значень контрольних точок по кадру, тобто знаходження центроїда кадру. При розмірі словника в 10 слів, наш алгоритм показав точність розпізнавання 40% для людини, яка навчала систему та 20% для людини, яка не навчала систему.

[1] OpenCV - <http://opencv.org>

[2] DLib - <http://dlib.net/>

[3] Modern Face Recognition with Deep Learning -
<https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cfc121d78>

E-mail: ✉ andriy@glybovets.com.ua.