

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
«КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Кафедра інформатики факультету інформатики

**Інформаційна система обробки відеозаписів для процесу
розпізнавання образів
Текстова частина до курсової роботи
за спеціальністю «Комп’ютерні науки» - 122**

Керівник курсової роботи

к.ф.-м.н., доцент

Нагірна А.М.

(Підпис)

“___” _____ 2022 року

Виконала студентка КН-МП1

Оксюта І.М.

“___” _____ 2022 року

Київ 2022

Зміст

Вступ	3
Розділ 1. Системи відеопошуку на основі контенту.	5
1.1 Основні характеристики та складові	5
1.2 Області застосувань	7
1.3 Особливості побудови систем та основні підходи до розробки	8
1.4 Висновки	10
Розділ 2. Методи обробки для вилучення облич.	11
2.1 Детекція	11
2.2 Дискретизація	14
2.3 Кластеризація	15
2.4 Висновки	17
Розділ 3. Опис створеної системи	18
3.1 Загальний опис системи	18
3.2 Оцінка якості кадрів	20
3.3 Опис програмного продукту та інтерфейсу	24
3.4 Аналіз результатів	26
3.5 Подальші дослідження	27
3.6 Висновки	28
Висновки	29
Джерела	30

Анотація

У даній роботі розглядається використання методів машинного навчання та алгоритмів розпізнавання образів для побудови відеопошукових систем. Описано основні складники такої системи направленої на пошук людських облич та запропоновано метрики оцінки якості вхідних даних, для відбору найвдаліших. Результатом роботи є система, яка направлена на обробку відеофайлів з метою вилучення найбільш інформативних кадрів на основі ранжування зображень за якістю. Отримані результати можуть бути використані в системах обробки облич або для формування пакету вхідних даних для навчання нових моделей машинного навчання в сфері розпізнавання образів.

Вступ

У наш час біометричні технології, такі як сканування райдужки ока, відбитків пальців та обличчя, стають все більш важливими та є ключовими у сучасних системах безпеки та розпізнавання людей. Так обробка облич здійснюється для ідентифікації, визначення віку, статі, етнічної приналежності чи емоційного стану. Тому зазвичай обличчя вважається найважливішим біометричним ідентифікатором. Подібні рішення широко імплементуються в системах відеоспостереження, як для державних органів безпеки так і для приватних цілей.

Основною перешкодою для досягнення високої точності в відеопошукових системах на відміну від статичних зображень є переповненість оточення. Рух, оклюзія зображення, зміна освітленості чи роздільної здатності, наявність додаткових перепон чи рух камери заважають коректно відпрацьовувати стандартним методам, що пристосовані до фото.

Таким чином розпізнавання облич на кожному кадрі відеофайлу буде надмірним та дуже ресурсозатратним. Крім того, результати детекції в цих кадрах можуть бути або некоректними або просто низької якості, непридатними для розпізнавання. З іншого боку, процес збору даних з кожного відеокадру займає багато часу. Тому, щоб вирішити проблему величезної кількості кадрів даних у відео, нам потрібно обрати кілька кадрів, облич на яких достатньо для завдань розпізнавання обличчя.

У цій роботі розглянуто загальну структуру та підхід до побудови системи для вилучення найважливіших кадрів, запропоновано методи оцінювання якості обличчя в кадрі. Також представлено інтерфейс для обробки відеофайлів і розбиття на ключові зображення. Дана система може стати одним із модулів відеопошукової системи, що направлена на роботу з

людськими обличчями. Або використовуватись як застосунок для препроцесингу даних у набір зображень, що можуть бути використані для навчання моделей машинного навчання, враховуючи специфіку предметної області.

У першому розділі наведено загальний огляд систем відеопошуку, області застосування таких систем, визначено основний напрям роботи та описано кроки необхідні при розробці системи вилучення ключових зображень з обличчями людей присутніх на відео.

У другому розділі детально описано методи які використовуються на кожному з кроків алгоритму вилучення кадрів. Наведено порівняння існуючих рішень.

У третьому розділі описано алгоритм дії розробленої системи, запропоновано метод оцінки якості зображень. Також реалізовано програмний продукт з користувацьким інтерфейсом. Наведено аналіз отриманих результатів та пропозиції подальшого дослідження.

Розділ 1. Системи відеопошуку на основі контенту.

1.1 Основні характеристики та складові

Система відеопошуку на основі вмісту (Content-based video retrieval) - це система, яка відноситься до пошукових засобів в архівах цифрового відеовмісту, де пошук базується на результатах аналізу відеоконтенту з вилученням даних, які можна індексувати для процесу пошуку. До таких даних відносяться: зображення, об'єкти на відео, текст і інші. Зокрема системний підхід використовує методи машинного навчання для вилучення ключових кадрів для швидкого перегляду та ефективного індексування відео.

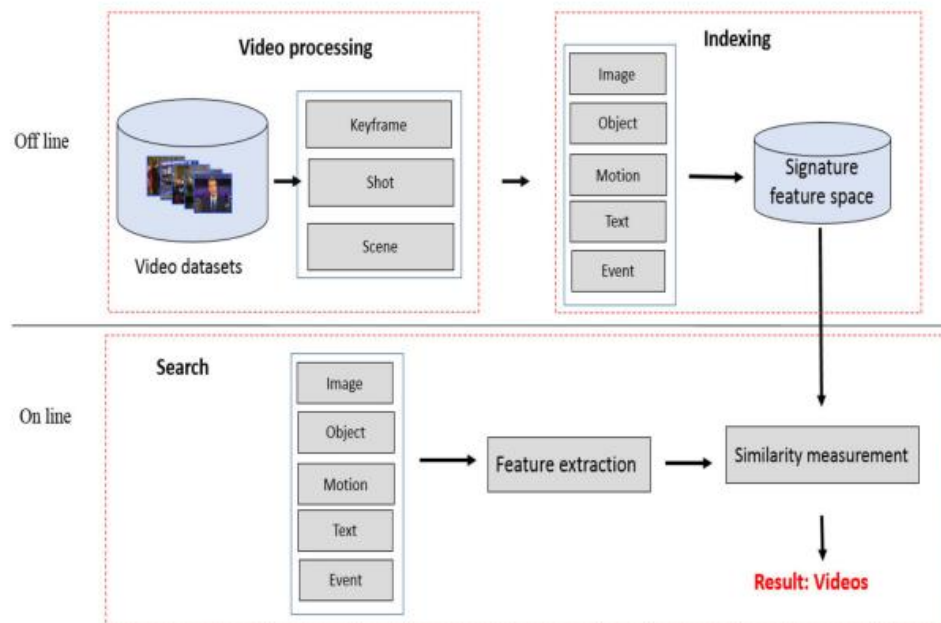


Рис.1.1 Схеми системи CBVR

Система пошуку відео на основі вмісту (CBVR) зазвичай має три етапи, показані на рисунку 1.1. Детально розглянемо кожен з них:

1. Етап Offline обробки відео зі сховища. Для цього з кожного файлу система вилучає відповідні елементи для опису відео. Зазвичай опис реалізують шляхом виділення набору кадрів і сцен, ключових кадрів. Набір кадрів визначається як неперервний набір зображень з однаковим вмістом. Сцена — це набір кадрів, згрупованих за чітко визначеними критеріями (подібний зміст, рух персонажів, тощо). Останній елемент, ключові кадри — це відеозображення, які описують вміст відео, усуваючи будь-яку надмірність. Виокремлення набору ключових кадрів називається узагальненням відео.
2. Етап індексування передбачає вилучення відповідних об'єктів, які будуть добре описувати відеофайл. Таким об'єктом може бути зображення, конкретний об'єкт, рух, текст, тощо. Вилучені елементи будуть проіндексовані в компактних і відповідних сигнатурах і збережені в просторі ознак пошуку.
3. Етап Online на основі витягнутих елементів на першому кроці, має обчислити подібність між об'єктами запиту та кожним вектором ознак у просторі ознак пошуку. Мета — повернути найбільш релевантні відео до запиту.

Відеоузагальнення є важливою технікою в програмах обробки відео. Метою цього процесу є надання короткого та репрезентативного резюме для оригінального відео. Існують два основних підходи: статичне та динамічне узагальнення. Статичний підсумок повертає набір найбільш репрезентативних кадрів для повністю завантаженого відео. Цей процес також відомий як вилучення ключового кадру. З іншого боку, динамічний підсумок — це процес відбору найбільш репрезентативних відеосегментів

самого відео. Але такий процес є досить завданням і вимагає різних модулів, а також ускладнює розуміння процесу пошуку за статичним запитом. Тому в даній роботі буде більш детально розглянуто саме статичний підхід до узагальнення відеофайлів. А саме запропоновано алгоритм для системи вилучення ключових кадрів відео як підсистеми CBVR.

1.2 Області застосувань

Системи відеопошуку направлені на формування стоку оброблених даних які описані деякими мітками. За допомогою запитів користувачі можуть взаємодіяти з системою з метою пошуку найбільш релевантного відеофайлу. Механізм схожий до стандартних пошукових систем, таких як Google, які індексують вміст веб-сторінок. Саме завдяки схожості та розуміння користувачами принципу пошуку, системи CBVR можуть бути імплементовані в багатьох сферах, що вимагають роботи з відеоматеріалом.

Так, наприклад, системи відеопошуку можуть бути частинами системи безпеки та використовуватись як інструмент для пошуку людей на відео. Або ж використовуватись у таких платформах як YouTube для пошуку релевантних відеоматеріалів.

Окрім цього, оскільки дослідницьке поле досить широке та є дуже багато варіантів для проектування подібних систем, частою практикою є використання окремих частин системи як самостійних алгоритмів. Модуль розпізнавання образів з навченою моделлю на специфічних даних або механізм вилучення ключових кадрів можуть бути використаними при розробці всіх систем, що базуються на машинному навчанні та направлені на обробку візуальної інформації.

1.3 Основні підходи до побудови системи статичного узагальнення

Оскільки для створення статичного підсумку використовується лише візуальна інформація, ця категорія відеозведення є більш простою, швидшою та придатною для індексування та пошуку відеоспостереження. Головним процесом є саме вилучення ключових кадрів, які найкращим чином будуть описувати відеофайл, та можуть бути використані в системі розпізнавання образів для отримання текстуального опису відеоряду всередині системи CBVR.

Вибір ключових кадрів може бути заснований на конкретних об'єктах у кадрах (рух, фон, обличчя, кольори). Більшість методів вибору ключових кадрів базуються на глобальних функціях. Ці методи враховують елементи, які представляють кадр у глобальній манері, як-от колір, текстура, гістограма зображення тощо. З іншого боку, методи, засновані на локальних особливостях, використовують характерні точки або регіони для створення підсумку, що дозволяє нам зосередитися на конкретних об'єктах у фреймах, що дає нам ключові кадри відповідно до конкретного об'єкта. У рамках цієї роботи було конкретизовано застосування подібних систем безпосередньо для розпізнавання облич. Тому розглянемо ключові особливості обробки відеоряду для такого завдання.

Незважаючи на свою важливість, методи вилучення ключових кадрів, засновані на ключових точках, не підходять для систем розпізнавання обличчя. По-перше, ці підходи не враховують труднощі, з якими потрібно зіткнутися для обробки зображень, такі як умови освітлення, зміна пози, вираз обличчя, оклюзія, відстань між обличчям і камерою тощо. По-друге, використовуючи ключові точки, ми не можемо гарантувати, що ми вилучимо найбільш нейтральне зображення обличчя, яке буде найкраще

розпізнаватись. З цих причин логічно зосередитись над виділенням ключового кадру на основі оцінки якості обличчя(face quality assessment). FQA формується на основі геометрії обличчя, включаючи позу і роздільну здатність зображення, фактор точності виявлення лиця, освітлення, вираз обличчя тощо.

В загальному вигляді систему статичного узагальнення для пошуку облич можна описати наступними кроками:

1. **Face Detection.** На початку роботи з кожного кадру відеофайлу необхідно виокремити області з обличчями. Найбільш поширеними алгоритмами детекції є моделі з використанням згорткових нейронних мереж. Такі алгоритми досить якісно оброблюють зображення та дозволяють оцінити точність виявлення обличчя на знімку. В результаті цього кроку, з відеофайлу отримуються набір зображень виявлених облич.
2. **Clustering.** Для якісного процесингу та підрахунку якості для кожного обличчя наявного у відеоряді(може бути багато людей на відео) потрібно відділити їх один від одного. Для цього застосовують алгоритми кластеризації, тобто зображення на основі їх вмісту розділяються на окремі групи з присвоєнням міток за допомогою алгоритмів машинного навчання. В результаті цього кроку, набір всіх облич розділяється на окремі класи(які відносяться до різних людей).
3. **FQA.** Для оцінювання кожного зображення можуть бути застосовані різні метрики на розсуд розробника. Головною метою є різностороння оцінка зображення, та komponування загальної оцінки.

4. Ранжування. Відбір N найкращих кадрів для кожного обличчя в залежності від результатів FQA.

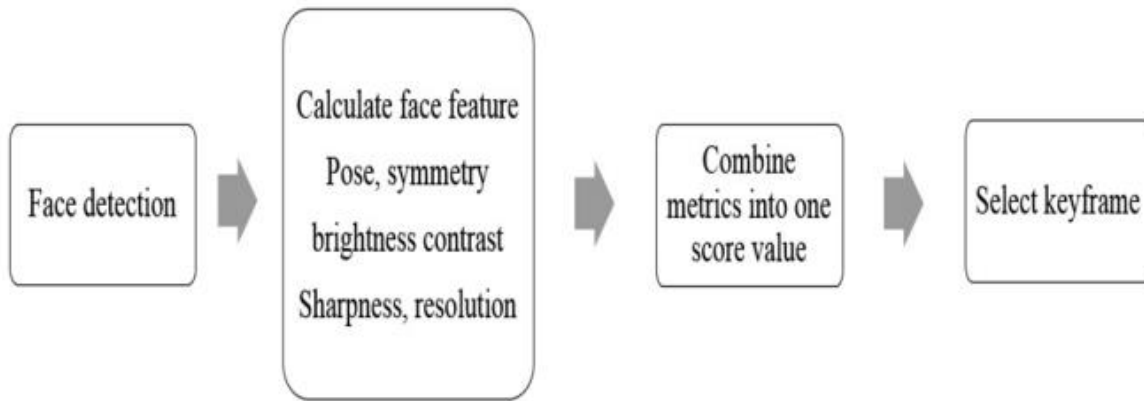


Рис. 1.2 Схема системи статичного узагальнення з використанням FQA

До основних проблем які можуть заважати коректній роботі такої системи є похибки алгоритмів на кожному з кроків. І детектори і кластеризатори оброблюють вхідні дані з деякою похибкою, тому важливо підібрати оптимальні параметри для цих алгоритмів. Також важливим є час обробки відеофайлів в такій системі. Важливо максимально оптимізувати процес, використовуючи швидкодійні мови програмування та принципи паралелізму.

1.4 Висновки

У даному розділі було розглянуто основні характеристики та особливості систем відеопошуку на основі вмісту, області використання. Визначено напрям роботи, а саме розробка модуля статичного відеоузагальнення для пошуку облич. Описано основні складові алгоритму вилучення ключових кадрів з використанням методу оцінки якості вмісту.

Розділ 2. Методи обробки для вилучення облич.

2.1 Детекція

На першому етапі обробки вхідних відеофайлів необхідно з послідовності кадрів виокремити обличчя людей присутніх на відео. Для цього використовують моделі детекції. Зазвичай такі алгоритми проходять навчання на величезних тренувальних датасетах та мають високі показники точності. Для задач детекції облич найчастіше використовують один з наступних алгоритмів.

1. Viola-Jones

Детектор обличчя Viola-Jones широко використовується для застосінків, які працюють в режимі реального часу. Використання функцій Haar і AdaBoost робить навчання повільнішим, але тестування дуже швидким. До переваг даного алгоритму можна віднести роботу з зображеннями різної роздільної здатності та локалізацію очей та області рота. Це дозволяє працювати з цими компонентами обличчя та порівнювати їх положення відносно ідеального зразка. Тим не менш, детектор виявляє лише фронтальні обличчя.

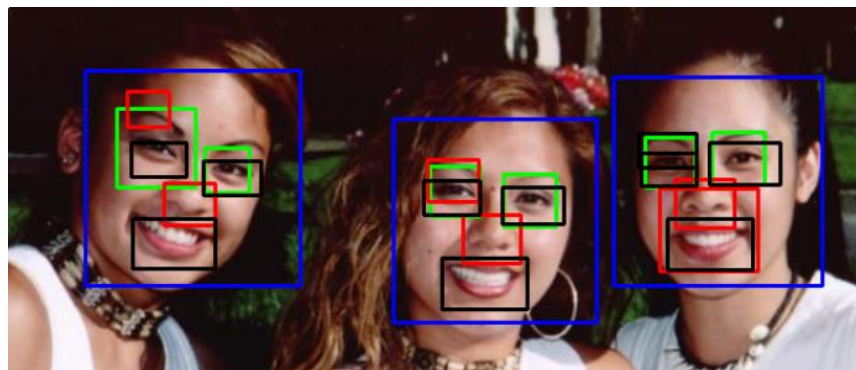


Рис.2.1 Результат роботи детектора Viola-Jones

2. Dlib

Dlib – це кросплатформенна бібліотека для задач комп'ютерного зору та машинного навчання. Алгоритм детекції обличчя в цій бібліотеці демонструє більш високу продуктивність на фронтальних або трохи майже фронтальних зображеннях. Але в той же час, він доволі чутливий до широких варіацій поз і зображень з низькою роздільною здатністю (мінімальний необхідний розмір – 80x80). До переваг можна віднести здатність виявити невелику оклюзію зображення та знаходження 68 міток обличчя, які дуже допомагають, наприклад, при аналізі емоційного стану.

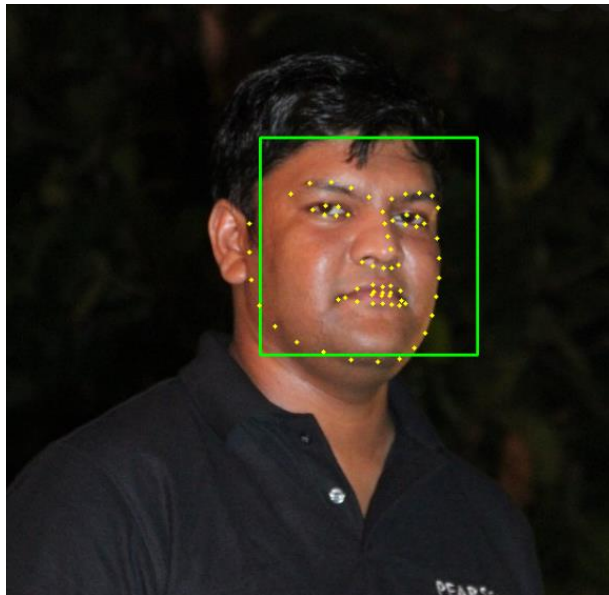


Рис.2.2 Результат роботи детектора Dlib

3. MTCNN

Multi-task Cascaded Convolutional Networks (MTCNN) - фреймворк, розроблений як для виявлення обличчя, так і для задачі групування обличчя. Алгоритм складається з трьох етапів згорткових мереж, які здатні розпізнавати обличчя та орієнтири, такі як очі, ніс і рот. На першому етапі використовується невелика CNN для швидкого

створення обличч-кандидатів. На другому етапі відбувається уточнення запропонованих зразків за допомогою більш складної CNN. І нарешті, на третьому етапі використовується третя CNN, складніша за інші, для фінального уточнення результату та визначення орієнтирів обличчя. Крім того, алгоритм розпізнає обличчя, що мають невелику оклюзію та малу роздільну здатність.



Рис.2.3 Результат роботи детектора MTCNN

Цей детектор в результаті обчислює п'ять орієнтирів, показаних на рисунку 2.3 (два ока, два кути роту і ніс), оцінку впевненості, яка відображає ймовірність того, що виявлений об'єкт є обличчям, і обмежує обличчя в рамку.

Для подальшої роботи вирішено використовувати саме цей детектор обличчя. По-перше, MTCNN дозволяє усунути непотрібні кадри. Відхилені кадри – це кадри, в яких детектор не може виявити орієнтири облич. Наприклад, обличчя з великою різницею поз або незбалансованим розподілом яскравості, що може приховувати більшість деталей обличчя. По-друге, використовуючи показник впевненості, можна обрахувати додаткову

метрику якості зображення і таким чином проранжувати кадри-кандидати. Нарешті, на основі п'яти орієнтирів, наданих MTCNN, та їх координат, можна досить просто обрахувати оцінку пози голови.

2.2 Дискретизація

Після процесу вилучення облич з кадрів відео необхідно логічно розділити їх на окремі класи(показники того, що обличчя належать різним людям) шляхом кластеризації набору отриманих зображень. Для того щоб алгоритми кластеризації працювали, необхідно для початку перевести кожне зображення в деякий вектор замість значень пікселів у RGB. Для цього існують різні методи дискретизації та спеціалізовані дескриптори зображень, які вилучають ключові показники зображень і формують вектор характеристик.

Для конкретної задачі дискретизації обличчя замість використання дескрипторів зображень, які працюють з будь-якою вхідної інформацією, доцільно застосувати алгоритм який призначений для задач обробки обличчя. FaceNet — це система розпізнавання облич, розроблена в 2015 році дослідниками з Google, яка досягла найсучасніших результатів у ряді контрольних наборів даних розпізнавання обличчя. Систему FaceNet можна використовувати в багатьох задачах завдяки кільком реалізаціям моделі з відкритим вихідним кодом від сторонніх розробників і наявності попередньо навчених моделей. Систему FaceNet можна використовувати для вилучення високоякісних характеристик облич, які називаються embeddings облич. Модель являє собою глибоку згорткову нейронну мережу, навчену через triplet loss-function, яка спонукає вектори для одного обличчя ставати більш схожими (менша відстань), тоді як очікується, що вектори для різних облич

стануть менш схожими (більша відстань). Вихідним результатом роботи такого алгоритму є 128D вектор, який містить інформацію про характеристики зображення обличчя.

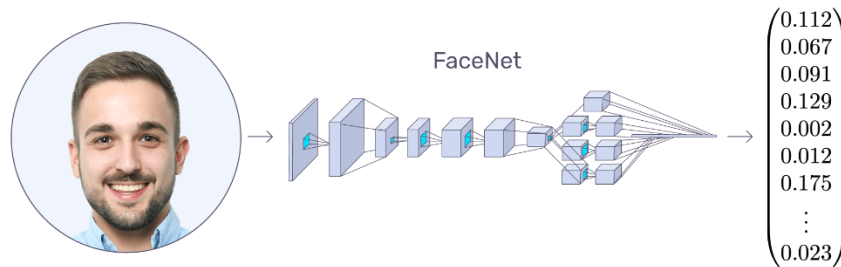


Рис.2.4 Схема роботи моделі FaceNet

2.3 Кластеризація

Після переведення набору зображень у векторну форму необхідно провести кластеризацію цих векторів та виділити угруповування які містять вектори, що описують одну й ту саму людину. Для цього можна застосувати різні алгоритми машинного навчання. Для даної задачі було використано алгоритм ієрархічної кластеризації. Ієрархічна кластеризація — це впорядкована декомпозиція даних на основі групової подібності.

Існують два методи для пошуку цих ієрархічних кластерів: агломеративний та роздільний. Агломеративна кластеризація використовує підхід знизу догори, при якому кожна точка даних виділяється в окремий кластер. Потім ці кластери об'єднуються, на основі обрахованих відстаней між ними, знаходячи найбільш подібні. Відстань може вимірюватись по різному: евклідова, косинусна, манхеттенська, тощо. Роздільна кластеризація використовує підхід зверху вниз, при якому всі точки даних починаються в

одному кластері. Потім використовуючи параметричний алгоритм кластеризації, наприклад K-Means, проводиться поділ на два кластери. І так для кожного кластера проводиться поділ поки не буде досягнуто потрібної кількості кластерів.

Для алгоритму кластеризації необхідним параметром є кількість кластерів. Для визначення цього параметра необхідно проаналізувати дендрограму. Дендрограми відповідають графічному представленню ієрархічного дерева, тож перед початком обробки необхідно побудувати ієрархічне дерево і на основі отриманого результату визначити кількість кластерів.



Рис. 2.5 Принцип дії ієрархічної кластеризації.

2.4 Висновки

У цьому розділі було розглянуто основні алгоритми для побудови системи вилучення ключових кадрів облич з вхідного відеоряду. Описано вже реалізовані рішення та виконано їх порівняння для визначення найбільш оптимального методу для подальшої імплементації в алгоритмі системи.

Розділ 3. Опис створеної системи

3.1 Загальний опис системи

Для реалізації системи вилучення ключових кадрів запропоновано наступний алгоритм.

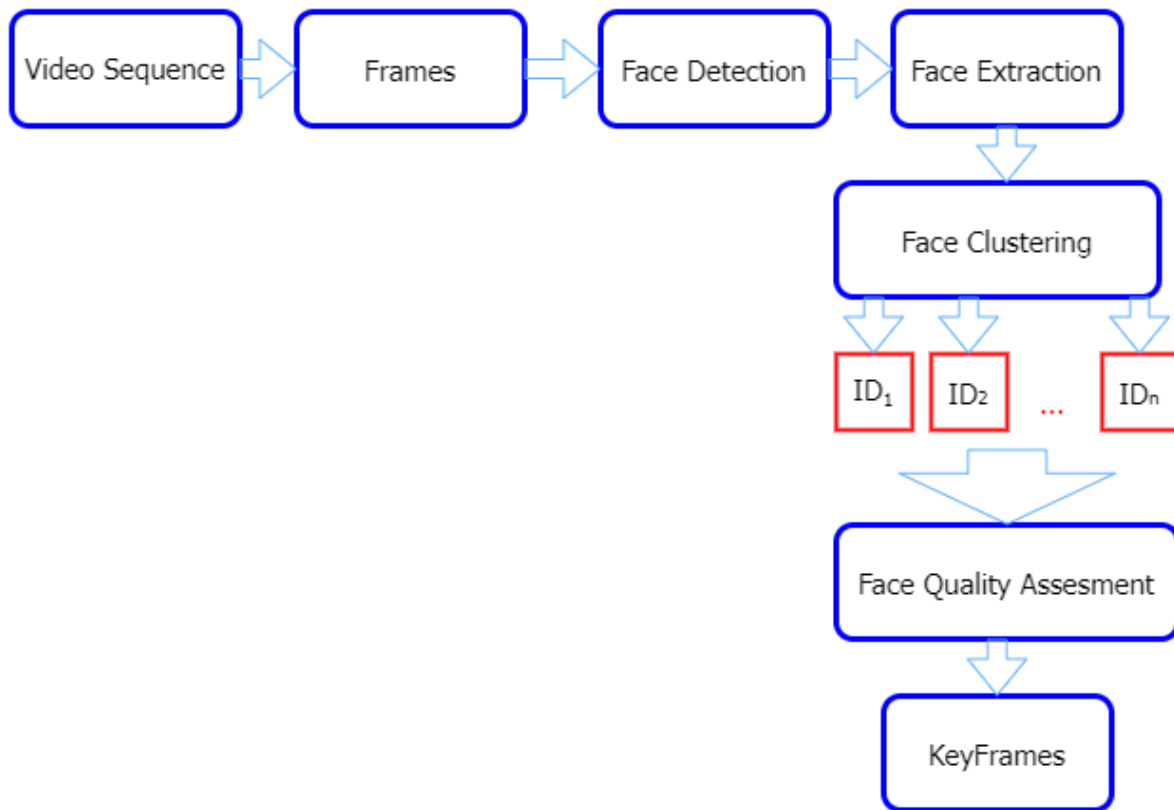


Рис. 3.1 Схема алгоритму

Опис алгоритму:

1. Вхідний відеофайл ділиться на фрейми(кадри) за допомогою обробника.
2. Для кожного кадру застосовується детектор облич. Як вже було описано в розділі 2, найкращим з точки зору реалізації буде

використання детектора MTCNN. Метод окрім виділення зон обличчя також надає додаткову інформацію про ключові точки обличчя, яка буде використовуватись для оцінки якості.

3. Вилучення областей з обличчями в окремі зображення. Результат роботи детекції виглядає як підмасив вхідного зображення, тому досить легко відділити обличчя від загального фрейму.
4. Тепер для зображень в отриманому наборі необхідно провести кластеризацію, щоб відділити обличчя, які належать різним людям. Для початку застосуємо дискретизацію зображень, тобто переведення фрейму у вектор який описує ключові характеристики зображення. Для цього використано попередньо натреновану модель FaceNet, яка призначена для створення embedding фото та переводить значення пікселів у 128-вимірний вектор ознак. Далі для кластеризації застосовуємо ієрархічну кластеризацію. Кількість кластерів, як необхідний параметр алгоритму визначаємо шляхом побудови дендрограми. В результаті отримуємо вхідний набір ображень з присвоєними мітками класів.
5. Для зображень кожного класу проводимо оцінку якості облич. В результаті отримуємо відсортований набір для кожного класу.
6. Обираємо K ключових кадрів кожного класу. Значення параметру K задає користувач.

Функціонування алгоритму в такому вигляді є достатньо оптимальним з точки зору швидкодії та отриманої якості обробки. Про метод формування оцінки для кожного зображення обличчя більш детально описано у наступному розділі.

3.2 Оцінка якості кадрів

Оцінка якості обличчя (FQA) проводиться для кожного кандидата. У цій роботі використано п'ять метрик для оцінки якості обличчя на зображенні: точність детектора, поза голови, різкість, яскравість і роздільна здатність. Найкращі за якістю кадри для кожного обличчя будуть відібрані як ключові.

Обрано саме ці показники через їх важливість для оцінки якості обличчя. По-перше, саме обличчя в анфас є найбільш адекватним для представлення особистості. По-друге, через переміщення людей перед камерою на зображення можуть бути шуми, це призводить до низької якості зображення обличчя. Третя обрана метрика – це вимірювання яскравості, оскільки незбалансована яскравість зображення може приховати деталі обличчя в будь-якому положенні(засвітити деякі важливі області). Нарешті, у зображенні з низькою роздільною здатністю неможливо візуалізувати окремих компонент обличчя. Також потрібно врахувати похибку з якою спрацьовує детектор обличчя. Модель MTCNN надає показник confidence, який відображає впевненість алгоритму що результатом детекції було обличчя. З цих причин для оцінки було обрано саме дані п'ять показників. Детально розглянемо кожен з них, та спосіб обчислення.

1. **Оцінка впевненості.** Визначається як один з параметрів, що повертається після роботи детектора MTCNN.
2. **Оцінка пози.** Для підрахунку необхідно визначити в якій мірі ключова точка обличчя зміщена від центру зображення. Для визначення ключової точки обличчя використана точка яка відповідає за ніс та визначається детектором в процесі обробки. Для визначення центру зображення використаємо формулу:

$$x_c = \frac{x_2 - x_1}{2} \quad y_c = \frac{y_2 - y_1}{2}$$

, де (x_1, y_1) – координати верхнього лівого кута, (x_2, y_2) – координати нижнього правого кута. Для оцінки вимірюємо евклідову відстань між цими точками.

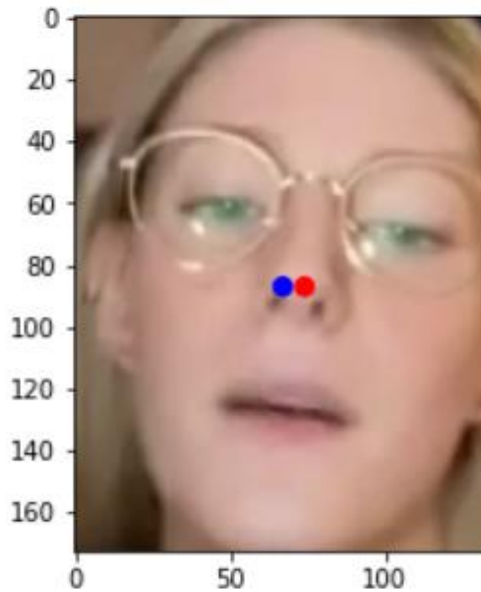


Рис.3.2 Оцінка пози. Синя крапка – центр зображення, червона – ніс.

$$Pose\ Score = \frac{1}{1 + \sqrt{(x_c - x_n)^2 + (y_c - y_n)^2}}$$

, де (x_n, y_n) – координати точки «ніс» відповідно до результатів роботи детектора.

3. **Оцінка яскравості.** Деякі області обличчя можуть бути погано помітні на темних зображеннях. Тому нам потрібно знайти зображення обличчя з найкращим розподілом освітлення. Показник

яскравості дорівнює сумі значень усіх пікселів, поділених на площу зображення.

$$\text{Brightness Score} = \frac{\sum_{i=1}^H \sum_{j=1}^W B_{ij}}{H * W}$$

, де H , W — висота та ширина зображення, а B_{ij} — значення пікселя в координатах i, j . Діапазон значень кожного пікселя $B \in [0, 255]$.

4. **Оцінка різкості.** Оскільки набір зображень буде відображати рух людей на відео, то деякі з кадрів матимуть ефект розмиття. Ранжування за різкістю націлене на знаходження найбільш нейтральних зображень, та відсіювання розмитих. Щоб обчислити показник різкості, необхідно виконати розмиття зображення обличчя за допомогою оператора Гауса. Тоді оцінка визначається як абсолютна різниця між вхідним і розмитим зображенням, поділена на розмір зображення.

$$\text{Sharpness Score} = \frac{\sum_{i=1}^H \sum_{j=1}^W |I - G(I)|}{H * W}$$

, де I — вихідне зображення, $G(I)$ — зображення після розмиття за Гаусом. Ядро $G(I)$ має розмір (5,5) зі стандартним відхиленням 1,0 у напрямках X і Y . W , H — ширина та висота зображення відповідно.

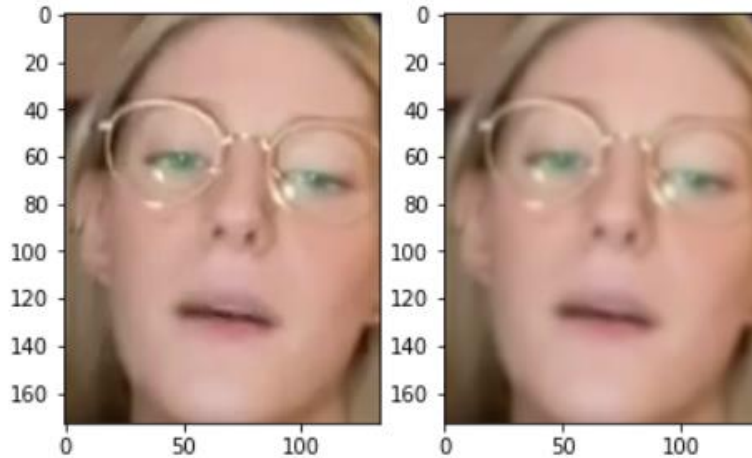


Рис.3.3 Розмиття за допомогою оператора Гауса

5. **Оцінка роздільної здатності.** Відображає наскільки придатними до подальшої обробки є зображення. Розраховується як добуток висоти на ширину зображення.

$$Resolution\ Score = H * W$$

Після обчислення п'яти показників для кожного зображення обличчя вилучені ознаки об'єднуються в єдине значення якості. Спочатку кожна оцінка нормалізується відповідно до максимального значення цього показника для даної послідовності. Потім всі оцінки характеристик об'єднуються в загальне значення номінальної якості зображення.

$$Face\ Quality = \sum_{i=1}^N \frac{S_i}{S_{i_{max}}}$$

, де S_i – значення оцінки i , $S_{i_{max}}$ – максимальне значення оцінки i для заданої послідовності зображень, $N=5$ – кількість оцінок.

Таким чином, для кожного набору зображень, які будуть отримані в результаті кластеризації(розділення на окремих людей) застосовуємо FQA. В результаті отримуємо набір відсортованих за спадання якості зображень для кожної людини, що присутня на відео та яку вдалося розпізнати детектором облич. В залежності від параметра K який задає користувач, та який вказує на бажану кількість ключових кадрів для кожного обличчя – формуємо результат з відсортованого списку.

3.3 Опис програмного продукту та інтерфейсу

Для реалізації цього проекту була обрана мова програмування Python 3.7. Вибір аргументований тим, що бібліотеки Python містять всі необхідні інструменти для реалізації поставленої задачі, час розробки значно менший та найчастіше для задач пов'язаних з розпізнавання образів, обробкою зображень, машинним навчанням використовують саме Python.

Опис використаних бібліотек для побудови алгоритму:

1. **Decord**. Бібліотека для швидкої обробки відеоряду. Працює швидше ніж стандартні інструменти зчитування.
2. **OpenCV**. Класична бібліотека для широкого спектру задач комп'ютерного зору. Має великий набір інструментів та вбудованих алгоритмів оптимізованих на найкоротший час виконання.
3. **Mtcnn**. Бібліотека безпосередньо для реалізації моделі MTCNN. Також має оптимізовані алгоритми.
4. **TensorFlow, Keras**. Фреймворк від Google для машинного навчання, містить зручні інструменти створення моделей нейронних мереж. Використовується безпосередньо для завантаження моделі FaceNet для побудови embeddings.

5. **Scikit-Learn.** Бібліотека з методами для машинного навчання. Використовується для реалізації алгоритму ієрархічної кластеризації.
6. **Numpy, Scipy, Matplotlib.** Стандартні бібліотеки для обчислень та побудови графіків.

Для створення користувацького інтерфейсу було обрано платформу **Streamlit.io**. Вона має пряму інтеграцію з Python та дозволяє досить швидко зробити обгортку для задач машинного навчання, містить інструменти для швидкого деплою веб додатків.

В результаті реалізований інтерфейс має наступний вигляд:

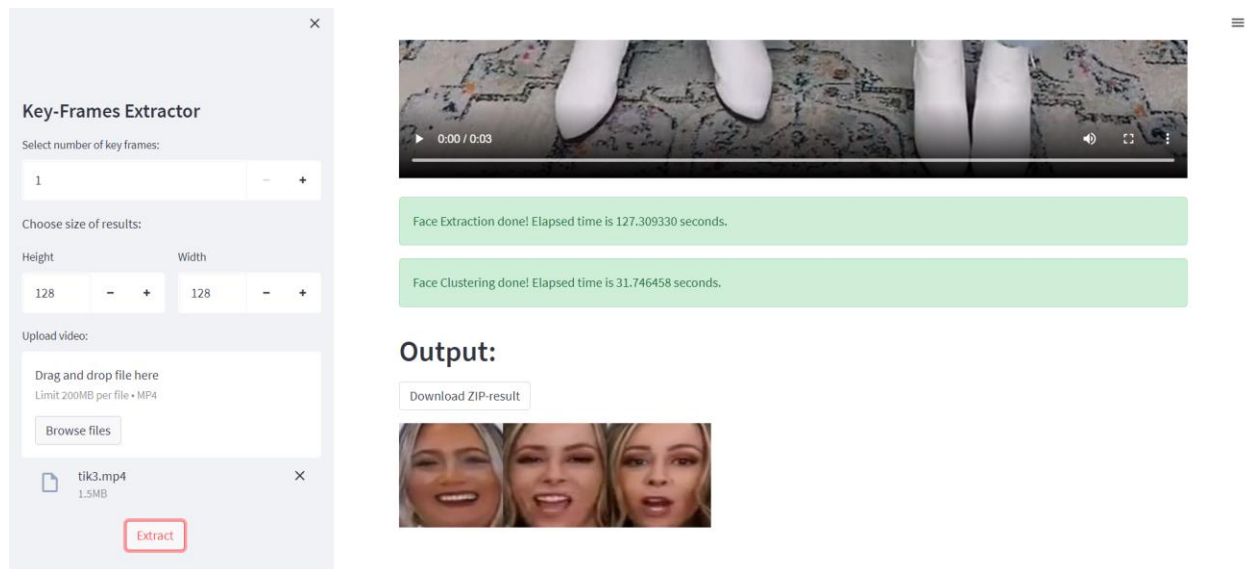


Рис. 3.4 Графічний інтерфейс програми

Користувач може завантажити відео в форматі MP4. Натискаючи Extract – запускається алгоритм. При цьому у вікні з'являється програвач з відео, яке оброблюється. Процес обробки займає деякий час, для цього по завершенню кроків детекції та кластеризації користувачу надходять сповіщення на екран. По завершенню роботи алгоритму, у вікні виводу

відображаються зображення отримані в результаті роботи. Є можливість завантажити ZIP-архів з результатами.

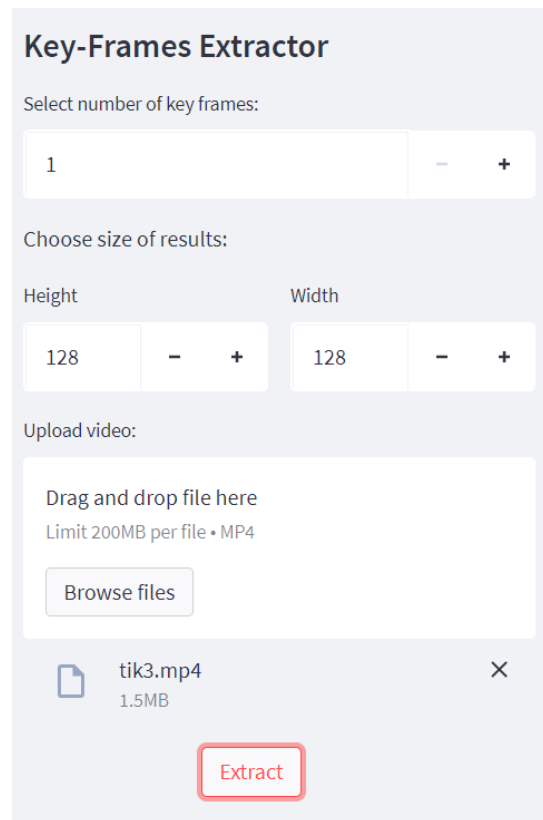


Рис. 3.5 Вікно редагування параметрів

У вікні редагування параметрів алгоритму користувач може задати бажану кількість ключових кадрів для кожного обличчя, розміри результуючих зображень та обрати відеофайл для обробки.

3.4 Аналіз результатів

Аналізуючи роботу системи на різних вхідних даних вдалося виокремити два основних фактора які впливають на якість обробки.

1. Час обробки. Оскільки використані готові рішення(методи детекції, дискретизації, кластеризації) для цього алгоритму достатньо

оптимізовані на швидкодійній мові C++, основним чинником є потужність процесора комп'ютера на якому ведуться обчислення. Для відпрацювання використовувався стандартний процесор CPU. Тому для дуже довгих відео або відео з великою кількістю людей час обробки був досить великим. Цю проблему можна вирішити шляхом запуску алгоритму на процесорах GPU, використовуючи принципи паралелізму.

2. Параметр кількості кластерів. Оскільки в запропонованому алгоритмі для визначення цього параметра використовується метод побудови дендрограми, автоматичне обрання кластерів може бути не дуже точним. Але ручний метод оцінки не є зручним та не може виконуватись автоматично, тому в процесі кластеризації можуть виникати неточності. Цю проблему можна спробувати вирішити або іншим методом підбору кількості кластерів або винесенням цієї задачі на розсуд користувача як додатковий параметр алгоритму. Таким чином при у вікні редагування параметрів, користувач буде вказувати скільки облич наявно в даному відеофайлі.

3.5 Подальші дослідження

Отримані в ході роботи результати можуть бути використані при подальших дослідженнях систем відеопошуку на основі вмісту. Запропонований алгоритм може стати базовим для розробки інших систем або модифікацій існуючих. Відкритим питання залишають задача кластеризації зображень та метод оцінки якості. Для розділення облич на окремі класи можна спробувати реалізувати інші методи кластеризації. Для оцінки якості в свою чергу можуть бути досліджені показники, що базуються

на даних пікселів або використовують для оцінки інші методи комп'ютерного зору. Також можлива адаптація алгоритму до систем, які працюють в реально часі. Загалом реалізований алгоритм планується використовувати в дипломній роботі, що пов'язана з розпізнавання обличч на відео. А саме в якості обробника відеофайлів для формування вхідного датасету зображень, який виступить навчальним для моделі машинного навчання.

3.6 Висновки

У даному заключному розділі було описано алгоритм, створений з використанням методів комп'ютерного зору, для поставленої задачі вилучення ключових кадрів з відеофайлу. Також описано запропоновані метрики оцінки якості зображення обличчя. Представлено графічний інтерфейс, який дає змогу оброблювати вхідні дані від користувача запропонованим алгоритмом. Проведено аналіз результатів роботи та визначено шляхи подальших досліджень.

Висновки

У цій роботі було проведено дослідження функціонування систем відеопошуку на основі вмісту. Такі системи є потужними інструментами для задач комп'ютерного зору та мають широкий спектр використання. У якості завдання для курсової роботи було обрано створення одного з модулів такої системи – алгоритму вилучення ключових кадрів обличчя з відеофайлів. Для реалізації цієї задачі було визначено основні кроки та описано необхідні методи обробки облич. Сформовано алгоритм повної обробки вхідного відеозапису, запропоновано метрики оцінки якості зображень обличчя.

Також був повністю реалізований програмний продукт, включно з графічним інтерфейсом для користувача. Отримані результати роботи алгоритму можна аналізувати з двох сторін: час обробки та точність. На час обробки дуже впливає потужність комп'ютера на якому виконується програма. На точність впливає оцінка кількості людей на відео. Таким чином визначено основні засади для подальшого дослідження створеної системи.

Створений продукт може бути використаний як самостійний застосунок для обробки відео при вирішенні задач комп'ютерного зору та машинного навчання. Або стати повноцінним модулем системи відеопошуку на основі контенту.

Джерела

1. Y. Shi, C. Otto, i A. K. Jain, «Face Clustering: Representation and Pairwise Constraints», IEEE Trans.Inform.Forensic Secur., вип. 13, вип. 7, с. 1626–1640, Лип 2018.
2. E. M. Saoudi i S. Jai-Andalousi, «A distributed Content-Based Video Retrieval system for large datasets», Journal of Big Data, вип. 8, вип. 1, с. 87, Чер 2021.
3. L. Chang, A. Pérez-Suárez, i M. González-Mendoza, «Effective and Generalizable Graph-Based Clustering for Faces in the Wild», Computational Intelligence and Neuroscience, вип. 2019, с. e6065056, Груд 2019.
4. G. Ciaparrone, L. Chiariglione, i R. Tagliaferri, «A comparison of deep learning models for end-to-end face-based video retrieval in unconstrained videos», Neural Comput & Applic, вип. 34, вип. 10, с. 7489–7506, Трав 2022.
5. Y. Yang, B. C. Lovell, i F. Dadgostar, «Content-Based Video Retrieval (CBVR) System for CCTV Surveillance Videos», в 2009 Digital Image Computing: Techniques and Applications, Груд 2009, с. 183–187.
6. F. Schroff, D. Kalenichenko, i J. Philbin, «FaceNet: A Unified Embedding for Face Recognition and Clustering», в 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Чер 2015, с. 815–823.
7. M. I. Nouyed, «Evaluation and Understandability of Face Image Quality Assessment», MS, West Virginia University Libraries, 2019.

8. T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, i C. Busch, «Face Image Quality Assessment: A Literature Survey», ACM Comput. Surv., с. 3507901, Січ 2022.
9. S. Bhattacharya, C. Kyal, i A. Routray, «Simplified Face Quality Assessment (SFQA)», Pattern Recognition Letters, вип. 147, с. 108–114, Лип 2021.
10. Z. Gao, G. Lu, i P. Yan, «Key-frame selection for video summarization: an approach of multidimensional time series analysis», Multidim Syst Sign Process, вип. 29, вип. 4, с. 1485–1505, Жов 2018.
11. M. Tapaswi, M. Law, i S. Fidler, «Video Face Clustering With Unknown Number of Clusters», в 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), Жов 2019, с. 5026–5035.