

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет інформаційних технологій
Кафедра інтелектуальних технологій**

“На правах рукопису”
УДК _____

МАГІСТЕРСЬКА РОБОТА

**Випускна кваліфікаційна робота
на здобуття освітнього ступеня «магістр»
на тему:**

**«Розробка системи розпізнавання обличчя людини
у відеопотоці з доповненою реальністю»**

Галузь знань: 12 - «Інформаційні технології»

Спеціальність: 122 «Комп'ютерні науки»

Освітньо-наукова програма «Технології штучного інтелекту»

Виконавець:

Короткова Тетяна Сергіївна
студентка магістратури

Науковий керівник:

Іларіонов Олег Євгенович
к. т. н., доцент кафедри
інтелектуальних технологій

Магістерська робота допущена до захисту
рішенням кафедри інтелектуальних та інформаційних систем
Протокол № _____ від «__» _____ 20__ р.
Зав. кафедри _____ проф. Снитюк В.Є.

Зміст

Перелік умовних позначень.....	4
Вступ.....	6
РОЗДІЛ 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ. ПОСТАНОВКА ЗАДАЧІ.....	8
1.1 Аналіз сучасного стану питання та обґрунтування вибору теми дослідження.....	8
1.2 Огляд існуючих рішень.....	11
1.3 Постановка задачі.....	17
1.4 Висновки до першого розділу	17
РОЗДІЛ 2. ТЕОРЕТИЧНІ ВІДОМОСТІ.....	19
2.1 Основні поняття доповненої реальності	19
2.2 Основні поняття процесу розпізнавання обличчя	22
2.3 Порівняння існуючих підходів до розпізнавання обличчя людини у відеопотоці	24
2.4 Згорткові нейронні мережі.....	31
2.5 Висновки до другого розділу.....	39
РОЗДІЛ 3 ОПИС РЕАЛІЗАЦІЇ ПРОГРАМНОГО ЗАСТОСУВАННЯ.....	40
3.1 Побудова цільової функції та аналіз технічного завдання	40
3.2 Вибір архітектури згорткової мережі для розпізнавання обличчя	40
3.3 Розробка алгоритму програми.....	43
3.4 Обґрунтування вибору засобів розробки програми	46
3.5 Вибір методів оцінки програми.....	47
3.6 Розробка програмного застосування.....	48
3.7 Опис інтерфейсу програми	52
3.8 Аналіз результатів роботи програмного застосування	54

3.9 Висновки до третього розділу	64
Висновки.....	66
Список використаних джерел.....	67
Додаток А Лістинг програми.....	73

Перелік умовних позначень

3D (з англ. 3-dimensional) – тривимірна графіка

AI (з англ. Artificial Intelligence) - Штучний інтелект

API (з англ. Application Programming Interface) – Інтерфейс прикладного програмування

AR (з англ. Augmented Reality) – Доповнена реальність

ARBM (з англ. Adaptive Regional Blend Matching) – Адаптивне зіставлення змішаних областей

EER (з англ. Equal error rate) – Величина рівності помилок

FAR (з англ. False Acceptance Rate) – Коефіцієнт помилкового допуску, помилка другого роду

FRR (з англ. False Rejection Rate) – Коефіцієнт помилкової відмови, помилка першого роду

FTP (з англ. File Transfer Protocol) – Протокол передачі файлів

GLVQ (з англ. Generalized Learning Vector Quantization) – Узагальнений алгоритм векторного квантування, що навчається

GMFD (з англ. General Matching Face Detection) – Метод узагальненої відповідності

GUI (з англ. Graphical User Interface) – Графічний інтерфейс користувача

LBP (з англ. Local binary patterns) – Локальні бінарні шаблони

LFW (з англ. Labeled Faces in the Wild) – Дослівно “Маркеровані обличчя у дикій природі”. База даних для вивчення розпізнавання облич у необмежених середовищах

HOG (з англ. Histogram of oriented gradient) – Гістограма орієнтованих градієнтів

PSM (з англ. Perturbation Space Method) – Метод спотворення простору

ReLU (з англ. Rectified linear unit) – Випрямляч, одна з функцій активації

VIO (з англ. Visual Inertial Odometry) - Візуальна інерціальна одометрія

VR (з англ. Virtual Reality) – Віртуальна реальність

БД – База даних

БНМ – Багатошарова нейронна мережа

ЗНМ – Згорткова нейронна мережа

ЗНМХЛ – Згорткова нейронна мережа з хешуючим лісом

НМ – Нейронна мережа

ПЗ – Програмне забезпечення

Вступ

Темою даної дипломної роботи є “Розробка системи розпізнавання обличчя людини у відеопотоці з доповненою реальністю”.

Технологія розпізнавання обличчя використовується у багатьох галузях: безпека, медицина, розважальна індустрія і т.д. Багато з нас щодня користуються цією технологією (наприклад, щоб розблокувати свій смартфон). Проте розпізнавання обличчя, поєднане з технологіями доповненої реальності, може встановити нову планку. Наприклад, використовувати Augmented Reality (AR) (з англ. – “доповнена реальність”) і Artificial Intelligence (AI) (з англ. – “штучний інтелект”), щоб поєднати зображення людського обличчя з камери та інформацію про реальну людину. Можливості AR технологій дуже великі - вони можуть знайти застосування чи не в усіх сферах нашого життя. Доповнена реальність збагатить та покращить якість комунікації, отримання й обміну інформації, ведення підприємницької діяльності. Віртуальний екскурсовод проведе нас по руїнах замку, та ще й покаже сценку, як саме цей замок розвалили, і яким він був до того. Мобільний додаток з AR допоможе позбутись паперових інструкцій, провівши інтерактивне заняття з користування щойно придбаним предметом. Віртуальна примірочна одягу, аксесуарів та макіяжу дозволить легше створити новий образ. Щоб алгоритм візуального відстеження був корисним у реальних сценаріях, він повинен бути в змозі обробляти та долати випадки, коли зовнішній вигляд об’єкту змінюється від кадру до кадру. Значна і швидка зміна зовнішнього вигляду внаслідок шуму, оклюзії, фонові перешкоди, пози, масштабу та змін освітленості є основними проблемами, які потрібно подолати детектору [1]. Саме тому важливо розуміти, який з методів розпізнавання буде найбільш ефективним для тої чи іншої задачі.

Зважаючи на актуальність цієї теми, принципово важливими є подальші розробки та дослідження в даній сфері як на практичному, так і на теоретичному рівнях.

Метою роботи є дослідження ефективності розпізнавання обличчя людини у відеопотоці з доповненою реальністю методом нейронних мереж, розробка програмної системи та проведення експериментів.

Об'єктом дослідження даної роботи є процес розпізнавання обличчя людини у відеопотоці.

Предметом дослідження є інструментальні засоби вирішення задачі розпізнавання обличчя людини у відеопотоці з доповненою реальністю.

Для створення програмної системи буде використана високорівнева мова програмування загального призначення Python та інтегроване середовище розробки програмного забезпечення PyCharm Community Edition 2019.1.3. Відлагодження та тестування проводитиметься на комп'ютері типу IBM PC з процесором Intel(R) Core(TM) i7-3520U , 2.9 ГГц, ОЗУ 16 Гб.

Відповідно до поставленої мети дослідження визначено структуру роботи. Вона складається зі вступу, трьох розділів, висновків та списку використаних джерел.

РОЗДІЛ 1

АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ. ПОСТАНОВКА ЗАДАЧІ

1.1 Аналіз сучасного стану питання та обґрунтування вибору теми дослідження

Одним з найважливіших напрямків науково-технічного прогресу початку ХХІ століття є розвиток систем штучного інтелекту, здатних розширити коло розв'язуваних людством задач [2]. Прикладами систем, які можуть полегшити наше життя, є: системи керування процесами (промислові роботи, автономні транспортні засоби), системи відеоспостереження, системи організації інформації (наприклад, для індексації баз даних зображень), системи моделювання об'єктів або оточуючого середовища (аналіз медичних зображень, топографічне моделювання), системи взаємодії (наприклад, пристрої введення для систем людино-машинної взаємодії) [3]. Усі ці системи підпадають під сферу впливу комп'ютерного зору.

Так, зі зростанням потужності комп'ютерів і розповсюдженням високоякісних і недорогих відеокамер, такі галузі, як ігри, медицина, засоби масової інформації, автомобільна промисловість та освіта, почали приділяти більше уваги технології комп'ютерного зору і почали використовувати її переваги, збагачуючи свою продукцію інтелектом.

Багато додатків демонструють важливий вплив на повсякденне життя людини, наприклад, для зменшення кількості нещасних випадків на дорогах, автовиробники розробили розумну систему, яка виявляє втому людини та попереджає водіїв про це [4]. Зображення, що отримуються за допомогою камери у кабіні транспортного засобу, застосовуються для визначення положення голови, а також для перевірки, чи очі розплющені [5]. Увівши датчик Kinect до ігрової індустрії, Microsoft змогла позиціонувати новий ігровий пристрій без контролера як абсолютно новий спосіб розваг. З Kinect

гравцям більше не потрібно запам'ятовувати різні команди для ручного керування, вони самі є контролерами [6]. Доповнена реальність може використовуватися для швидкої примірки в магазинах: ідея зайти в меблевий магазин і тут же на тестовому стенді зібрати собі кімнату з меблями і побутовою технікою, користуючись підказками по поєднуваності, напрошується сама собою [7].

Щоб алгоритм візуального відстеження був корисним у реальних сценаріях, він повинен бути в змозі обробляти та долати випадки, коли зовнішній вигляд об'єкту змінюється від кадру до кадру. Значна і швидка зміна зовнішнього вигляду внаслідок шуму, оклюзії, фонові перешкоди, пози, масштабу та змін освітленості є основними проблемами, які потрібно подолати детектору [8].

На сьогоднішній день було запропоновано багато нових методів для вирішення кожної з цієї задачі. Точність підготовленого детектора обличчя в значній мірі пов'язана з даними та алгоритмом, які використовуються для навчання. Саме тому важливо розуміти, який з них буде найбільш ефективним для тої чи іншої задачі.

На сьогодні існує досить мало комп'ютерних систем, ціллю яких є саме розпізнавання облич із використанням доповненої реальності, а не тільки детекція. Найчастіше це або програми для комерційного використання у корпораціях, які відповідають за аутентифікацію та авторизацію, або програми, розроблені для держави чи силових структур, спрямовані на організацію належного рівня безпеки всередині країни - таке програмне забезпечення є пропрієтарним. У процесі дослідження не було знайдено реалізацій цільного продукту по розпізнаванню обличчя з доповненою реальністю у відкритому доступі, оскільки більшість розробок у цій сфері виробляються під індивідуальні вимоги великих замовників, і крім того, що

вимагають багато ресурсів на реалізацію (часові, фінансові, інтелектуальні), використовують чутливу інформацію, розголошення якої неприпустимо.

У роботі реалізовано демо-версію комп'ютерної системи по розпізнаванню обличчя з доповненою реальністю. Програма по зображенню у реальному часі ідентифікує людину та виводить короткі відомості про неї. Функціональність даної системи може бути скоригована під вимоги кінцевих користувачів – у залежності від сфери. Прикладами застосування є:

1) Використання у багатоквартирних будинках як засіб підвищення рівня безпеки. Встановлення домофону з відеокамерою дозволяє не тільки чути, але й бачити людину, яка хоче потрапити до під'їзду. Часто люди можуть впустити незнайомців, які приходять під виглядом працівників комунальних підприємств або доставки, хоча насправді є хуліганями чи крадіями. У випадку, якщо існуватиме єдина база облич з інформацією про працівників вище названих служб, запропоноване програмне забезпечення може бути інтегроване з відео-домофоном. Таким чином, жителі будинку зможуть бачити на екрані пульта, чи справді незнайомиць є працівником тої чи іншої організації – відеопотік буде доповнюватись відповідною інформацією.

2) Використання на місцях аварій та природних катастроф. Перед тим, як відправляти рятувальників на місце події, можна відправити квадрокоптери або роботів, оснащених 360-градусною камерою та технологією розпізнавання облич з доповненою реальністю. Ці машини зможуть передавати відео в реальному часі на гарнітуру AR у безпечну зону, щоб користувачі могли взаємодіяти з місцем інциденту так, ніби вони насправді там, при цьому залишаючись на відстані. Якщо під час інциденту на місці події знаходяться люди, програмне забезпечення для розпізнавання облич зможе ідентифікувати їх та надіслати сигнальний маячок про те, де вони знаходяться. Окрім того, така система в режимі реального часу дасть

можливість відрізнати постраждалих, які не занесені у базу, від рятувальників та оперативно надавати першим допомогу.

3) Використання у медичній сфері. Комп'ютерна система розпізнавання обличчя з доповненою реальністю може стати у нагоді лікареві для швидкого отримання інформації про наявність алергій, хронічних захворювань пацієнта або результати попередньої зустрічі без необхідності перегляду медичної картки. Більш корисним таке ПЗ буде за наявності централізованої карткової системи, оскільки дозволить агрегувати інформацію з різних медичних закладів. Таким чином, зникне потреба заводити нову реєстраційну картку, якщо до цього пацієнт обслуговувався у іншій лікарні. Особливо зручною така опція стане при введенні окулярів доповненої реальності у більш широкий вжиток.

1.2 Огляд існуючих рішень

Як зазначалось вище, цілісних аналогів систем розпізнавання облич із використанням доповненої реальності у відкритому доступі немає. Тому у роботі було розглянуто існуючі рішення по супутнім темам – по виявленню облич у системах доповненої реальності, а також розпізнаванню облич в цілому.

У ході дослідження ставало очевидним, що багато організацій намагаються знайти спрощений спосіб створення та плавної інтеграції розпізнавання об'єктів і облич у середовищах AR і VR, хоча і підхід кожної з них до реалізації дещо відрізнявся в залежності від ресурсів та сфери застосування. Універсальність розпізнавання об'єктів і облич можна побачити, спостерігаючи різноманітність галузей, у яких може використовуватись дана технологія. Нижче наведені деякі застосування, створені із використанням систем розпізнавання об'єктів і облич: розпізнавання тривимірних об'єктів для роботизованих систем захоплення у

заводському середовищі [8], система розпізнавання осіб для користувачів з вадами зору [9], система виявлення головок броколі на основі 3D-технології, яка допомагає підвищити ефективність методів збору врожаю [10].

Незважаючи на різноманітні програми та підходи до розробки для розпізнавання об'єктів і облич, у всій сфері застосування цієї технології можна побачити кілька основних математичних концепцій. Алгоритм Віоли-Джонса, який був запропонований в 2001 році, “об'єднує нові алгоритми та ідеї для створення основи для надійного і надзвичайно швидкого візуального виявлення” [11]. Цей алгоритм став основою програмного забезпечення для розпізнавання облич і використовується як база у різних дослідницьких проектах. Однією з основних функцій алгоритму Віоли-Джонса є виявлення ознак Хаара – областей, які є спільними майже для кожного людського обличчя, мають подібні візерунки світла і тіні. Наприклад, перенісся людського носа світліше, ніж верхні щоки. Область навколо очей темніша, ніж область навколо щік. Існує багато варіантів алгоритму Віоли-Джонса, але основна математика залишилася такою, як було вперше запропоновано. Іншими загальними математичними методами, які часто можна виявити під час дослідження розпізнавання облич, є Eigenfaces [12], Fisherfaces [13] і Laplacianfaces [14]. Ці три методи роботи з зображеннями спрямовані на те, щоб визначити, чи містить зображення обличчя і, якщо так, то в якій області зображення воно знаходиться. Але процес, за допомогою якого вони виявляють обличчя, зовсім різний, незважаючи на те, що концепції, на яких побудовані ці методи, в основному однакові.

Сьогодні існує багато методів розпізнавання облич. При їх розробці брались до уваги вимоги щодо розміру сформованих біометричних шаблонів облич, швидкості їх формування і порівняння та якості розпізнавання як основного критерію. Біометричний шаблон (або шаблон) – це цифровий еталон особливих характеристик, вилучених з біометричного зразка, за якими

одну людину можна відрізнити від іншої. Є кілька основних підходів. Наразі метрика Хеммінга для порівняння бінарних шаблонів є найшвидшою [15,16]. Але у свою чергу кращі показники якості в розпізнаванні облич демонструють глибокі згорткові нейронні мережі, що використовують небінарні біометричні шаблони [17,18].

Раніше був запропонований цілий ряд підходів до формування біометричних шаблонів зображень облич, які можуть використовуватися для біометричної верифікації або ідентифікації осіб [19, 20]. Проте практично всі останні результати в розпізнаванні осіб були отримані за допомогою використання глибинних згорткових нейронних мереж [21]. Існує багато підходів до навчання мереж для розпізнавання людей - мережі можуть бути навчені як на класифікацію, коли кожній особі відповідає свій клас, так і за принципом сіамських мереж на основі попарного порівняння для найкращої верифікації. Також є реалізації, що об'єднують обидва підходи. Кращі сучасні результати на базі облич Labeled Faces in the Wild (LFW) отримані композицією глибоких мереж, навчених на різних фрагментах (так званих патчах) облич [22]. Проте використання об'єднання кількох архітектур нейронних мереж вимагає значно більших обчислювальних ресурсів, в той час як деякі однопатчеві мережі можуть бути досить ефективними. Як біометричний шаблон зазвичай використовується вихідний сигнал одного з шарів, найчастіше це останній прихований шар в мережі. Для зниження розмірності отриманих шаблонів зазвичай використовується метод головних компонент, а порівняння шаблонів здійснюється за допомогою косинусної або L2-відстані [22].

Наприклад, у роботі Ю.В. Візільтера, В.С. Горбацевича, А.В. Воротнікова, Н.А. Костромова [21] на основі згорткової мережі з хешуючим шаром було отримано згорткову нейронну мережу з хешуючим лісом (ЗНМХЛ). Ця модифікація дозволяє змінювати розмір шаблону, що

використовується від надкомпактних (десятки байтів) до великих шаблонів (сотні байтів). Відмінність в них буде у якості розпізнавання. Це дозволило будувати компактні біометричні шаблони (250 байт) з одночасно покращеними (у порівнянні з базовою мережею) показниками верифікації (98,14%) і ідентифікації (91% rank-1) на базі осіб LFW, а також ультракомпактні (25 байт) шаблони. З іншого боку, описаний підхід дозволяє будувати великі біометричні шаблони, що дозволяють досягати високої якості розпізнавання - ЗНМХЛ з 2000 кодуєчими 7-бітними деревами досягає 98,59% точності верифікації 93% rank-1 на LFW (надбавка більш ніж в 3% в rank-1 щодо вихідної базової архітектури мережі).

Подібні роботи показують, що дослідження у сфері комп'ютерного зору тривають і є простір для вдосконалення методів та технологій.

Оскільки всі методи розпізнавання обличчя мають свої переваги і недоліки, проведення порівняння для пошуку найкращого без прив'язки до конкретної задачі є неефективним. Тому було вирішено оцінити існуючі методи розпізнавання об'єктів і облич, які потенційно можуть бути сумісними із середовищами AR та VR. Більш детальний опис цих методів наведено у пункті 2.3.

Серед програмних пакетів по розпізнаванню облич було розглянуто продукти корпорації NEC (Японія), Cognitec Systems GmbH (Німеччина), «VisionLabs» (Росія) «FACE ++» і Neurotechnology (Литва).

«NEC's Face Recognition» - одна з передових систем розпізнавання облич, розроблена японською компанією «NEC», що дозволяє ідентифікувати людей по кадрам багаторічної давності навіть, якщо людина знаходиться в окулярах або змінює вираз обличчя. Усі розпізнані обличчя зберігаються в базі даних, тому в разі необхідності можна підняти всю історію відеореєстрації і переглянути дату і час будь-якого збереженого зображення. Технологія NEC перевершує безліч інших систем розпізнавання своєю точністю і швидкістю.

Вона має хороші показники продуктивності в різних ситуаціях, в тому числі при роботі з відео низької якості і сильно стиснутими зображеннями. NEC аналізує індивідуальні ознаки обличчя (розмір, форму зіниць, лінії носа і рота), їх взаємне розташування, і знаходить потім по цій інформації відповідну людину в базі даних. Система включає в себе кілька модулів, що реалізують такі алгоритми:

1. Використовується метод узагальненої відповідності (GMFD), який забезпечує високу швидкість детектування і високу точність розпізнавання особи. Основна логіка розпізнавання обличчя в рамках GMFD – це модифікований алгоритм GLVQ. Метод GLVQ заснований на нейронних мережах і здійснює попередній пошук пар очей [23].

2. Алгоритм PSM (Perturbation Space Method), який дозволяє ефективно справлятися зі складнощами, пов'язаними з розташуванням обличчя у кадрі (обличчя під нахилом або певним кутом).

3. Метод ARBM (Adaptive Regional Blend Matching), який зменшує вплив невеликих змін на обличчі (наприклад, зміни виразу обличчя, наявність окулярів, головного убору) на точність розпізнавання [24].

«FaceVACS-VideoScan» - просте у використанні, налаштоване програмне забезпечення для розпізнавання облич і відеопотоці в режимі реального часу, яку пропонує компанія «Cognitec Systems». Система «FaceVACS-VideoScan» складається з декількох компонентів системи: відеомагазину, що керує відеопотоками; сервера відеосканування, який координує всі компоненти системи і виконує основні біометричні операції; обчислювального вузла, використовуваного для розподілу обчислювального навантаження; інтерфейсу користувача; диспетчера сигналів, що отримує повідомлення про події та обслуговуючого мобільні пристрої; операційної бази даних; комплекту інтеграторів. На сьогоднішній день технологія FaceVACS використовує алгоритм розпізнавання облич B10T9. Цей алгоритм,

стійкий до змін міміки, поворотів обличчя (на $\pm 15^\circ$), часткового його закриття, використання окулярів та зміни освітлення [25]. Порівняння облич відбувається в режимі реального часу, застосовується C++ API та Web Services API.

«LUNA SDK» - спеціалізоване ПЗ відслідковування і розпізнавання облич на цифрових фотографіях або у відеопотоці від компанії «VisionLabs». Має одні з кращих в світі показників повноти і точності розпізнавання в реальних умовах. Розпізнавання осіб в LUNA SDK працює на основі глибинних нейронних мереж і складається з декількох ключових етапів: визначення положення і розмірів всіх облич (детекції); визначення розташування характерних рис обличчя і трансформації його в стандартизовану форму (вирівнювання); витяг дескрипторів (числових векторів, які підсумовують характерні ознаки обличчя); порівняння облич з базою зображень і запобігання підміни облич. LUNA SDK повністю розроблена на C ++, підтримує платформи Windows і Linux та надає можливість працювати в режимі багатопоточності.

«VeriLook SDK» - технологія ідентифікації облич, розроблена компанією «Neurotechnology». Це система виявлення облич з можливістю одночасного множинного розпізнавання і швидкої ідентифікації. VeriLook SDK доступна у вигляді пакету для розробки ПЗ і підтримує пристрої на Windows, Linux, Mac OS X, iOS і Android [26]. Алгоритм VeriLook реалізує локалізацію особи з використанням алгоритмів обробки цифрових зображень заснованих на глибоких нейронних мережах. Основними перевагами системи VeriLook є відсутність необхідності контакту із засобами сканування і швидке впровадження функцій біометричної ідентифікації в прикладні системи замовника [27].

1.3 Постановка задачі

Дослідити ефективність одного з методів розпізнавання обличчя людини у відеопотоці з доповненою реальністю. Досягнення поставленої мети передбачає вирішення наступних завдань:

1) здійснення інформаційного пошуку та проведення аналізу існуючих підходів до розпізнавання обличчя людини у відеопотоці з доповненою реальністю, виявлення сильних та слабких сторін використовуваних методів;

2) визначення методу реалізації розпізнавання обличчя людини у відеопотоці;

3) розробка застосування, що здійснює розпізнавання обличчя людини у відеопотоці з доповненою реальністю;

4) проведення експериментальних досліджень та визначення оцінки ефективності обраного методу розпізнавання обличчя людини у відеопотоці з доповненою реальністю.

1.4 Висновки до першого розділу

У розділі було проведено аналіз сучасного стану дослідження задачі розпізнавання обличчя з використанням технології доповненої реальності. Було виявлено, що на сьогодні у відкритому доступі немає комп'ютерних систем, ціллю яких є саме розпізнавання обличчя із використанням доповненої реальності, а не тільки виявлення обличчя. З огляду на це було прийнято рішення створити систему, яка б забезпечувала виконання цієї задачі, та дослідити її якість та ефективність. Оскільки аналогів систем розпізнавання обличчя із використанням доповненої реальності не було знайдено, було розглянуто існуючі рішення по супутнім темам – по виявленню обличчя у системах доповненої реальності, а також розпізнаванню обличчя в цілому. Оцінку та більш детальний опис існуючих методів розпізнавання об'єктів і обличчя, які потенційно можуть бути сумісними із середовищами AR та VR наведено у

пункті 2.3. У результаті аналізу предметної області були сформульовані завдання, виконання яких забезпечить вирішення поставленої задачі.

РОЗДІЛ 2. ТЕОРЕТИЧНІ ВІДОМОСТІ

2.1 Основні поняття доповненої реальності

Доповнена реальність (AR) - це технологія, що розміщує в навколишньому світі графічні або текстові об'єкти у режимі реального часу за допомогою комп'ютерних пристроїв [28].

Необхідно розрізняти доповнену, віртуальну та змішану реальності. У доповненій реальності віртуальні об'єкти переносяться на реальний світ, доповнюють дійсне середовище. У віртуальній реальності світ не є справжнім, хоч і може бути схожим на нього. Він створюється за допомогою комп'ютерних систем, сприйняття якого забезпечує сенсорна система людини (зорові, тактильні та інші відчуття). Змішана або гібридна реальність поєднує ці поняття. Так, світ, у який занурюється людина, створюється за допомогою VR (з англ. "Virtual Reality" – віртуальна реальність), а додання віртуальних елементів забезпечується AR. Отже, віртуальна реальність взаємодіє лише з користувачами, а доповнена реальність - з усім зовнішнім світом [29].

Створення доповненої реальності у загальному випадку відбувається наступним чином: пристрій, оснащений доповненою реальністю (спеціальні окуляри, смартфон тощо), використовуючи камеру, захоплює об'єкт фізичного світу; за допомогою програмного забезпечення пристрою у режимі реального часу проводиться розпізнавання об'єкта, накладання відповідної цифрової інформації та виведення кінцевого зображення на пристрій візуалізації.

Є декілька різних технологій, які використовуються для роботи AR. Розглянемо їх.

1. Доповнена реальність, що базується на маркерах. Цей тип технології використовує камеру та спеціальний пасивний візуальний маркер, наприклад QR-код (з англ. "quick response code" – код швидкого відгуку) або зображення

якогось об'єкту. Запрограмований результат буде показано лише тоді, коли сенсор читає маркер. Таким чином вдається вирізнити віртуальні об'єкти з реального світу [30].

2. Безмаркерна доповнена реальність. Безмаркерна технологія працює за особливими алгоритмам розпізнавання, де на навколишній ландшафт, знятий камерою, накладається віртуальна сітка. На цій сітці програмні алгоритми знаходять якісь опорні точки, за якими визначають точне місце, до якого буде “прив’язана” віртуальна модель. Перевага такої технології в тому, що об'єкти реального світу служать маркерами самі по собі і для них не потрібно створювати спеціальних візуальних ідентифікаторів, які б запускали відображення цифрового контенту. Віртуальний об'єкт можна розмістити у будь-якому місці, незалежно від навколишнього середовища, масштабувати та повертати його [31].

3. Просторова доповнена реальність. Є підвидом безмаркерної доповненої реальності. Інколи її ще називають координатно-, або GPS-орієнтованою (з англ. “GPS – Global Positioning System” - система глобального позиціювання). Щоб надати дані про ваше місцеперебування, вона може використовувати GPS, цифровий компас, датчик швидкості або акселерометр, якими оснащено ваш пристрій [32]. Завдяки масовому розповсюдженню смартфонів та планшетів ця технологія використовується найчастіше на даний момент. Найпоширеніші випадки використання – це позначення напрямків, пошук потрібних місць, таких як кафе чи офіс, або ж у додатках, що орієнтовані на конкретне місцеположення. Відомим застосування є гра Pokemon Go.

4. Доповнена реальність, що базується на проекції. Є підвидом безмаркерної доповненої реальності. Вона працює шляхом проектування світлових проекцій на фізичні поверхні. Спеціальні додатки допомагають здійснювати взаємодію між людиною та проекцією, визначаючи моменти

дотику людини до світла, яке проектується. Це досягається за допомогою порівняння очікуваної проекції та зміненої певними перешкодами, наприклад через дотик рукою [33]. Ще один цікавий спосіб – застосування плазмової технології, завдяки якій можна створювати тривимірні проекції (голограми) у просторі.

5. Доповнена реальність, що базується на візуальній інерційній одометрії (від англ. “Visual Inertial Odometry, VIO”). Є підвидом безмаркерної доповненої реальності. VIO – це технологія, яка допомагає відслідковувати позицію та орієнтуватися в просторі за допомогою сенсорів та камери. Завдяки цьому можливо створити точну 3D-модель простору навколо пристрою, оновлювати її в реальному часі, визначати в ній положення, передавати ці дані всім додаткам та накладати поверх неї додаткові шари. Можливості цієї технології насправді унікальні: можна вимірювати відстані, вставляти різноманітні об’єкти в інтер’єр та взаємодіяти з ними. VIO обіцяє стати найперспективнішою технологією в AR, на даний момент її використовують такі корпоративні гіганти, як компанія Google в своєму Project Tango та компанія Apple в ARKit [34].

Запропонована система розпізнавання облич з AR використовує безмаркерну доповнену реальність. Процес відслідковування обличчя складається з трьох основних кроків: розпізнавання, ініціалізація та відслідковування. Розпізнавання дозволяє визначити, який об’єкт знаходиться на поточному зображенні. У даній системі об’єктом є виявлене обличчя людини. Під час ініціалізації у відеопотоці необхідний об’єкт виявляється на основі одного або кількох ключових кадрів. Ключові кадри — це зображення, які визначають початкову та кінцеву точку для будь-якого плавного переходу [35]. Об’єкт відслідковується у відеопотоці на основі ключового кадру та інформації про попередню позицію об’єкта. Схема процесу розпізнавання та відслідковування об’єкта зображена на рисунку 2.1.

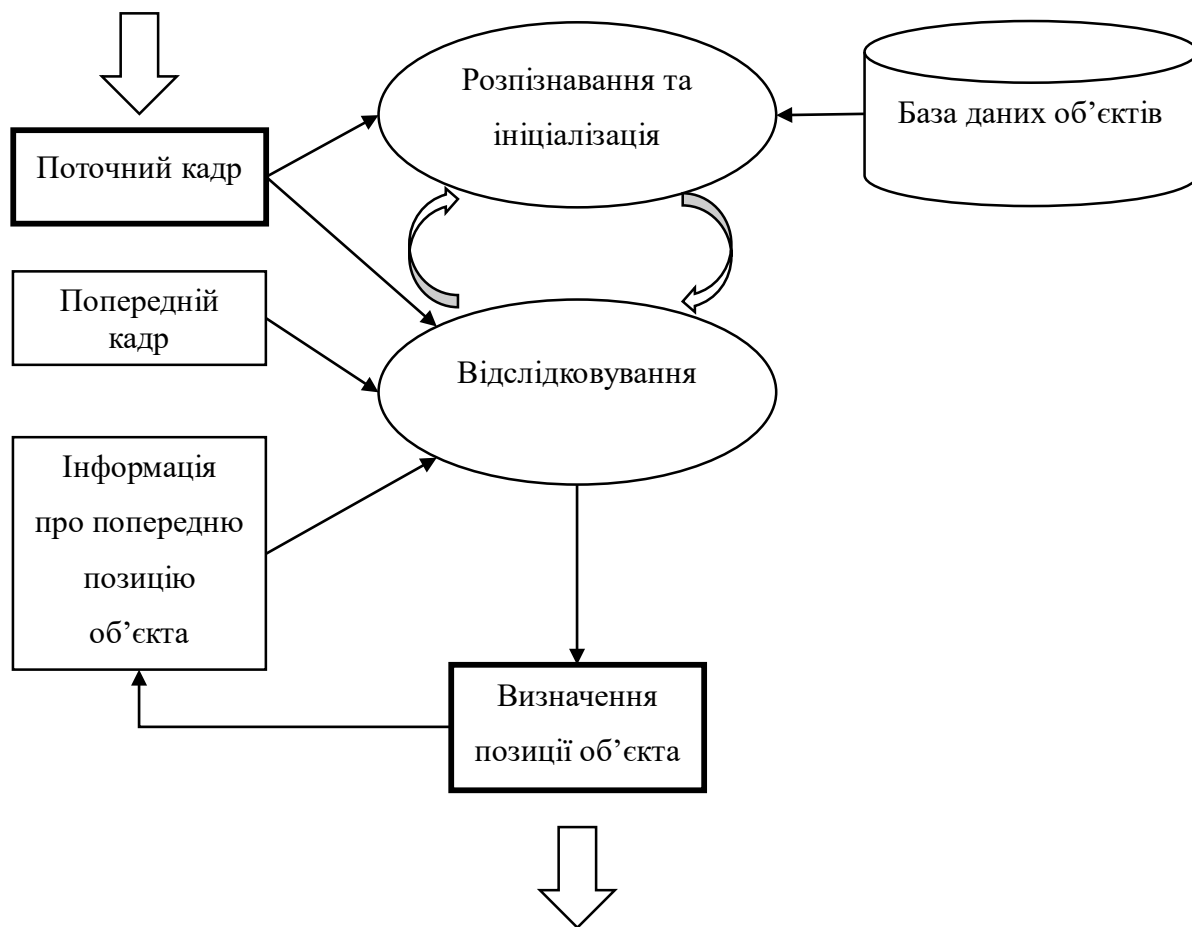


Рисунок 2.1 - Процес розпізнавання та відслідковування об'єкта у доповненій реальності

2.2 Основні поняття процесу розпізнавання обличчя

Загальний процес розпізнавання облич складається з виявлення облич, відстеження, вилучення ознак, зменшення розмірів та відповідно розпізнавання.

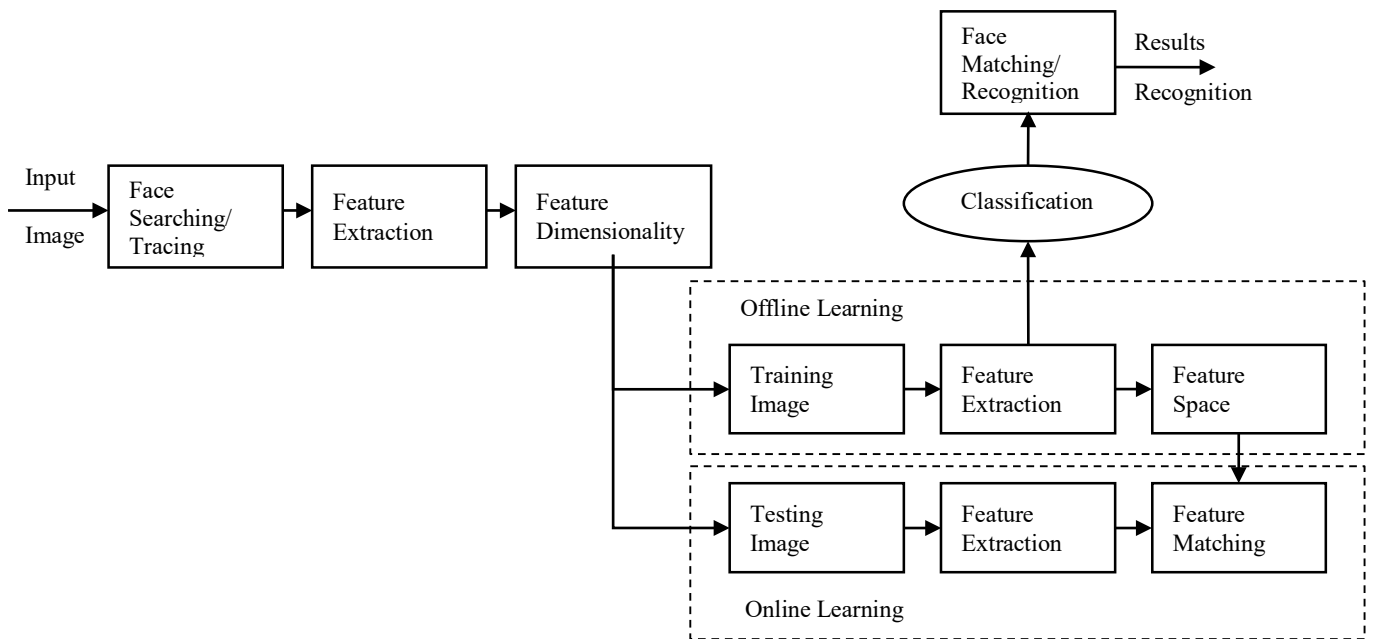


Рисунок 2.2 - Загальний процес розпізнавання обличчя людини

На рисунку 2.2 виявлення обличчя є першим кроком для реалізації розпізнавання облич. Його мета полягає в автоматичному пошуку області обличчя у вхідному зображенні. Якщо є ділянка обличчя, потрібно визначити конкретне місце та діапазон обличчя. Розпізнавання ділиться на виявлення обличчя на зображенні та на відео [36]. Якщо вхід є нерухомим зображенням, то кожне зображення виявляється; якщо на вхід подається відео, то виявлення обличчя виконується протягом всієї відеопослідовності.

Знаходження ключових ознак базується на виявленні обличчя, тому входом для нього є виявлене зображення обличчя. Найпоширенішими ознаками є LBP (локальні двійкові структури), HOG (гістограма орієнтованого градієнта), фільтри Габора та ін [37]. HOG описує особливості меж. Через нечутливість до змін освітленості та малих переміщень, він описує загальну та місцеву інформацію людського обличчя. LBP показує локальні зміни текстури зображення з інваріантністю яскравості. Функція Габо фіксує локальний структурний зміст просторового положення, селективності напряму і просторової частоти. Він підходить для опису людських облич.

Зменшення розмірності ознак відбувається наступним чином. Особливість обличчя, як правило, представляється багатомірним вектором. Це збільшує часову і просторову складність алгоритму. Крім того, важко судити про здатність опису характеристик облич з великою розмірністю. Однак багатомірний вектор ознак обличчя може проектуватися на маломірний . Інформація маломірного підпростору може завершувати ідентифікацію ознак обличчя. Після видобування особливостей оригінальні особливості повторно поєднуються, щоб зменшити розмірність вектора обличчя.

Після цього ми порівнюємо існуючі об'єкти в базі даних облич і обличчя, які повинні бути визначені, на основі певної стратегії пошуку співпадінь, приймаючи, таким чином, остаточне рішення. Відповідне співпадіння може бути представлено за допомогою автономного навчання або моделей співпадіння онлайн-алгоритмів [38].

2.3 Порівняння існуючих підходів до розпізнавання обличчя людини у відеопотоці

Вибір алгоритму, який використовується для ідентифікації людини по зображенню обличчя, залежить від конкретних умов його застосування. Наприклад, із завданням розпізнавання в строго обмеженому колективі легко справляється багат шарова нейронна мережа. У той же час завдання виявлення конкретної людини в натовпі (з невизначеним складом) вимагає застосування витончених методів для зниження рівня помилкових спрацьовувань. Швидше за все, в цьому випадку буде потрібно багаторівнева система, що містить велику множину аналізаторів, що працюють в різних ознакових просторах, з прийняттям рішення методом голосування. На початкових етапах роботи система ідентифікації повинна відсікати свідомо невідповідних кандидатів і використовувати множину кандидатів, що залишилися, для прийняття остаточного рішення про ідентифікацію.

Нижче представлений огляд по найбільш широко використовуваних методах виявлення та ідентифікації людини по зображенню її обличчя.

1. Метод головних компонент

Вхідні вектори - відцентровані і приведені до єдиного масштабу зображення облич. Власні вектори, обчислені для всього набору зображень осіб, називаються власними обличчями (eigenfaces). За допомогою обчислених раніше матриць вхідне зображення розкладається на набір лінійних коефіцієнтів, які називаються головними компонентами. Сума N перших головних компонент, помножених на відповідні власні вектори, є апроксимацією зображення порядку N .

Для кожного зображення обличчя обчислюються його головні компоненти. Зазвичай береться від 5 до 200 головних компонент. Інші компоненти кодують дрібні відмінності між обличчями та шум. Процес розпізнавання полягає в порівнянні головних компонент невідомого зображення з компонентами всіх інших зображень. Для цього зазвичай застосовують будь-яку метрику (найпростіший випадок - Евклідова відстань). Додаткове підвищення надійності досягається за рахунок додаткового застосування аналізу головних компонент до окремих ділянок обличчя таких, як очі, ніс, рот [39].

Основні недоліки МГК такі. Метод власних осіб вимагає для свого застосування ідеалізованих умов таких, як єдині параметри освітленості, нейтральний вираз обличчя, відсутність перешкод на зразок окулярів чи бороди. При недотриманні цих умов головні компоненти не будуть відображати міжкласові варіації. Також обчислення набору власних векторів відрізняється високою трудомісткістю.

2. Лінійний дискримінантний аналіз

Fisherface - нащадок Eigenface, цей алгоритм забезпечує більш високу точність розпізнавання при змінах освітлення або виразу обличчя. В основі

Fisherfaces лежить лінійний дискримінантний аналіз, а саме лінійний дискримінант Фішера. Дія алгоритму Fisherface заснована на пошуку проекції даних, при якій класи зображень обличч максимумально розділювані. Ця відмінність з Eigenface дозволяє вирішити проблему високої чутливості до змін освітлення. Алгоритм Fisherface використовує таку проекцію простору зображень на простір ознак, яка мінімізує внутрішньокласову і максимізує міжкласову відстань у просторі ознак. У цих методах передбачається, що класи лінійно нероздільні (лінійна роздільність – можливість розділити множину точок, що відповідають вхідним значенням, геометрично).

Зазвичай тренувальний набір містить зображення обличч при декількох базових умовах освітленості, на основі яких за допомогою лінійних комбінацій можна отримати будь-які інші умови освітленості. Цей метод дає високу точність розпізнавання (близько 96%) для широкого діапазону умов освітленості, різних виразів обличчя і наявності або відсутності окулярів. Однак залишаються нез'ясованими питання, чи цей метод може бути застосований для пошуку у великих базах даних, чи може метод працювати, коли в тренувальній вибірці для деяких осіб є зображення тільки в одних умовах освітленості [40].

3. Локальні бінарні шаблони

LBP - ефективний і простий в реалізації метод розпізнавання обличч. Отримане зображення розділяється на області. Оператор LBP застосовується до кожного пікселя зображення. Він встановлює яскравість центрального пікселя як поріг та робить перевірку його околиці - восьми пікселів, розташованих навколо. Пікселі, які мають значення більше, ніж центральний піксель (або рівну йому), приймають значення "1", ті, які менше центрального, приймають значення "0". Таким чином виходить восьмирозрядний бінарний код, який описує околицю пікселя. Далі для кожної області обчислюється гістограма частоти кожного числа, що отримується (тобто кожної комбінації

пікселів яких менша або більша за центр). Ця гістограма представлена 256-мірним вектором ознак. Далі гістограми конкатенуються і отримується вектор ознак для всього зображення. Наступним кроком є порівняння за допомогою методів машинного навчання. У класичному варіанті використовується метод найближчого сусіда. На відміну від EigenFaces, алгоритм стійкий до монотонних змін освітлення, що робить його придатним для розпізнавання облич у системах обробки в реальному часі.

4. Гнучкі контурні моделі обличчя

У даних методах розпізнавання проводиться на основі порівняння контурів особи. Контури зазвичай беруться для ліній голови, вух, губ, носа, брів і очей. Контури представлені ключовими позиціями, між якими положення точок, що належать контуру, обчислюються інтерполяцією. Зазвичай ключові точки розміщуються вручну на наборі тренувальних зображень. При пошуку контурів нового обличчя використовується метод імітації відпалу з цільовою функцією з двох складових. Для порівняння зображень використовуються значення головних компонент, обчислені на наборі векторів, що представляють собою координати ключових точок [41]. Головним завданням при розпізнаванні по контурах є правильне виділення цих контурів. У загальному вигляді це завдання за складністю можна порівняти безпосередньо з розпізнаванням зображень.

5. Порівняння еластичних графів

У цьому методі обличчя представляється у вигляді графа, вершини якого розташовані на ключових точках обличчя таких, як контури голови, губ, носа та їх крайніх точках. Кожна грань позначена відстанями між її вершинами. У кожній такій точці обчислюються коефіцієнти розкладання по функціях Габор для п'яти різних частот і восьми орієнтацій. Набір таких коефіцієнтів $J = \{J_j\}$ називається джетом (jet). Джети характеризують локальні області зображень і служать для двох цілей: по-перше, для знаходження точок

відповідності в заданій області на двох різних зображеннях; по-друге - для порівняння двох відповідних областей різних зображень [42]. Даний метод здатний досить надійно розпізнавати при змінах ракурсу до 20° ; при великих кутах точність розпізнавання різко зменшується, функція подібності виявляється більш чутливою до ракурсу, ніж до міжкласових відмінностей.

6. Методи, засновані на геометричних характеристиках обличчя

Суть полягає у виділенні набору ключових точок (або областей) особи і наступному виділенні набору ознак. Кожна ознака є або відстанню між ключовими точками, або відношенням таких відстаней. На відміну від методу порівняння еластичних графів, тут відстані вибираються не як дуги графів. Набори найбільш інформативних ознак виділяються експериментально.

У процесі розпізнавання порівнюються ознаки невідомого обличчя з ознаками, що зберігаються в базі. Завдання знаходження ключових точок наближається до трудомісткості безпосередньо розпізнавання, і правильне знаходження ключових точок на зображенні багато в чому визначає успіх розпізнавання. Тому зображення обличчя людини повинно бути без перешкод, що заважають процесу пошуку ключових точок. До таких перешкод відносять окуляри, бороди, прикраси, елементи зачіски і макіяжу. Освітлення бажано рівномірне і однакове для всіх зображень. Крім того, зображення обличчя повинно мати фронтальний ракурс, можливо, з невеликими відхиленнями. Вираз обличчя має бути нейтральним [43].

Таким чином, даний метод пред'являє досить суворі вимоги до умов зйомки і потребує надійного механізми знаходження ключових точок для загального випадку. Крім того, потрібне застосування більш досконалих методів класифікації або побудови моделі змін. У загальному випадку цей метод не є найоптимальнішим, однак для деяких специфічних завдань все ж перспективний. До таких задач можна віднести документний контроль, коли

потрібно порівняти зображення обличчя, отриманого в поточний момент, з фотографією в документі.

7. Порівняння шаблонів

Порівняння шаблонів полягає у виділенні областей обличчя на зображенні і подальшому порівнянні цих областей для двох різних зображень. Кожна область, що збігається, збільшує міру схожості зображень. Для порівняння областей використовуються найпростіші алгоритми на кшталт порівняння по кожному пікселю [44].

Недолік цього методу полягає в тому, що він вимагає багато ресурсів як для зберігання ділянок, так і для їх порівняння. З огляду на те, що використовується найпростіший алгоритм порівняння, зображення повинні бути зняті в строго встановлених умовах: не допускається помітних змін ракурсу, освітлення, емоційного вираження і ін.

8. Багатошарові нейронні мережі (БНМ)

БНМ може використовуватися для класифікації зображень обличчя на основі таких характеристик, як відстані між деякими специфічними частинами особи (ніс, рот, очі). У класичній БНМ нейронні з'єднання між шарами повнозв'язні, і зображення представлено у вигляді одновимірної вектора, хоча воно двовимірне. Архітектура згорткової НМ спрямована на подолання цих недоліків. У ній використовуються локальні рецепторні поля (забезпечують локальну двовимірну зв'язність нейронів), загальні ваги (забезпечують детектування деяких рис в будь-якому місці зображення) і ієрархічна організація з просторовими підвибірками (з англ. "spatial subsampling"). Згорткова НМ (ЗНМ) забезпечує часткову стійкість до змін масштабу, зсувів, поворотів, спотворень. Для ЗНМ характерні висока швидкість навчання і роботи, а також точність [45].

На основі зібраної інформації було проведено порівняльний аналіз та виявлено, наскільки сильно деякі фактори впливають на якість розпізнавання кожного методу. Для аналізу було взято наступні параметри:

- рівномірність та постійність освітлення;
- кут нахилу голови;
- наявність бороди, вусів, окулярів тощо;
- зміна емоцій на обличчі;
- обчислювальна складність методу.

Для відносного порівняння методів для кожного методу було присвоєно оцінку від 1 до 5, яке показує наскільки погіршується якість розпізнавання з динамічністю того чи іншого параметру (окрім обчислювальної складності), де 1 – не розпізнає, а 5 – розпізнає, як при еталонних умовах. Для обчислювальної складності значення 1 означає, що метод потребує багато обчислювальних ресурсів та займає багато часу, а 5 – не потребує особливих обчислювальних ресурсів. Результати аналізу зображені у таблиці 2.1

Умовні позначення: Освітленість – О, Нахил, поворот – НП, Аксесуари – А, Емоції – Е, Обчислювальна складність - ОС

Таблиця 2.1 – Порівняння методів розпізнавання обличчя

Метод	Параметри порівняння				
	О	НП	А	У	ОС
EigenFaces	1	1	1	1	4
FisherFaces	2	1	3	2	4
LBP	4	2	2	3	5
Гнучкі контурні моделі обличчя	3	4	2	2	5
Порівняння еластичних графів	3	5	2	5	3

Методи, засновані на геометричних характеристиках обличчя	1	2	1	2	4
Порівняння шаблонів	2	3	1	2	2
Багатошарові нейронні мережі	4	4	4	4	4

Таким чином, багатошарові нейронні мережі є більш стійкими та нечутливими до сторонніх впливів.

2.4 Згорткові нейронні мережі

Згорткові нейронні мережі (ЗНМ) мають спільні властивості з мережами прямого поширення типу перцептрон: вони складаються з нейронів, нейрони поєднуються між собою зв'язками, котрі характеризуються вагами, що коригуються під час навчання, а також використовують параметр зсуву (рисунок 2.3). Однак ЗНМ враховують, що вхідні дані є зображеннями. Таким чином, певні властивості можна закодувати в архітектурі мережі, що дозволить зменшити кількість параметрів нейронної мережі.

Нейронні мережі отримують вхідні дані (єдиний вектор), проводять перетворення над ним та передають через послідовність прихованих шарів. Кожен прихований шар складається з певної кількості нейронів, кожен з яких пов'язаний з усіма нейронами попереднього шару і де нейрони на кожному шарі повністю незалежні від інших нейронів на цьому ж рівні. Останній повнозв'язний шар називається вихідним шаром і в задачах класифікації відповідає за розподіл вхідних даних по класах.

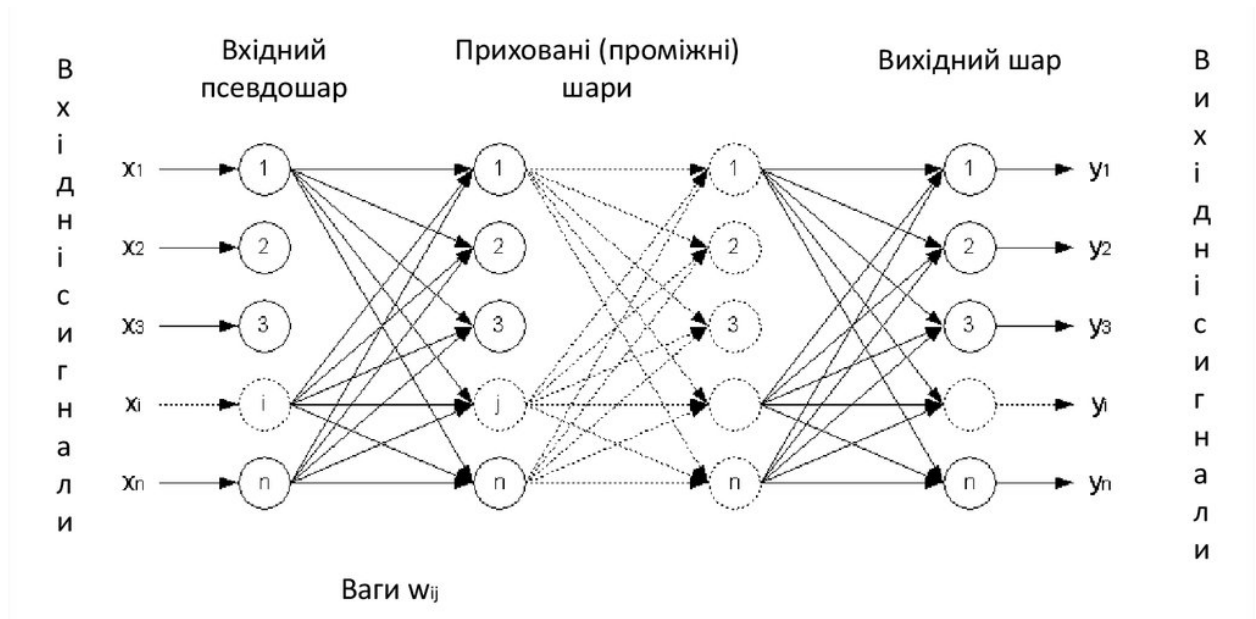


Рисунок 2.3 – Багатошарова нейронна мережа з прямим зв'язками. Ресурс: [\[https://ppt-online.org/134069\]](https://ppt-online.org/134069)

Припустимо, що вхідні зображення мають розмір $W = 20, H = 20, D = 3$, де W – ширина зображення, H – висота зображення, D – кількість каналів або глибина зображення, тоді один повнозв'язний нейрон у першому прихованому шарі звичайної нейронної мережі буде мати $20 \times 20 \times 3 = 1200$ ваг. Однак на практиці зображення зазвичай мають більш високу роздільну здатність, а кількість нейронів обчислюється кількома сотнями. Це призводить до зростання обчислювальної складності і, отже, до збільшення часу навчання мережі.

Згорткові мережі мають перевагу перед іншими типами архітектури: вони використовують локальні рецептивні поля, що дозволяє зменшити кількість параметрів навчання, а також вміють стискати дані, виділяючи з них тільки найважливішу інформацію [46].

Оскільки вхідними даними є зображення, нейрони розташовуються у трьох вимірах: ширина, висота, глибина (рисунок 2.4). ЗНМ фігурує 3 поняттями, які дозволяють досягти інваріантності до зміщення,

масштабування і спотворення зображення: локальні рецептивні поля, загальні ваги, субдискретизація.

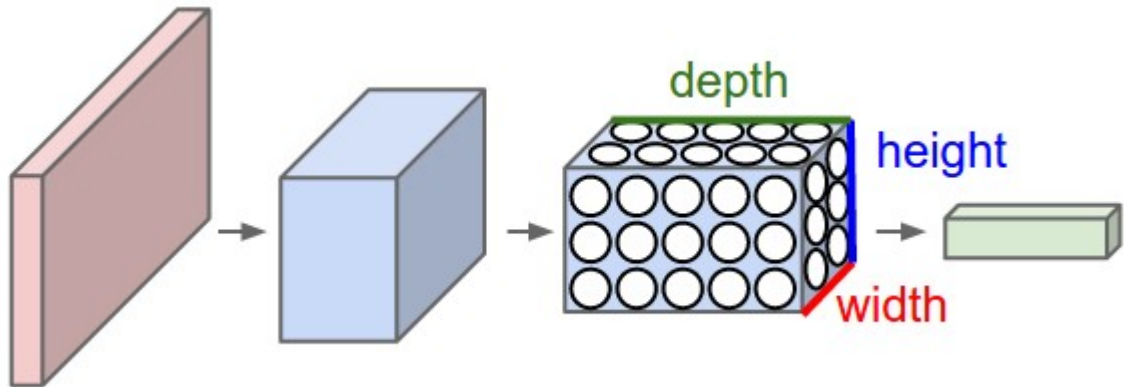


Рисунок 2.4 – Розташування нейронів згорткової нейронної мережі. Ресурс: [\[https://habr.com/ru/post/456186/\]](https://habr.com/ru/post/456186/)

Локальні рецептивні поля. Ідея полягає в тому, щоб з'єднати нейрони прихованого шару лише з тими нейронами попереднього шару, які входять в деяку маленьку область $n \times n$. Причому для кожного нейрона вибирається своя окрема область. Дана область і називається локальним рецептивним полем. Рецептивні поля виявляють елементарні візуальні ознаки такі, як кут, край, а їх комбінації дають можливість отримати більш складні ознаки. На рисунку 2.5 представлений приклад з'єднання нейронів для поля розміру 3×3 .

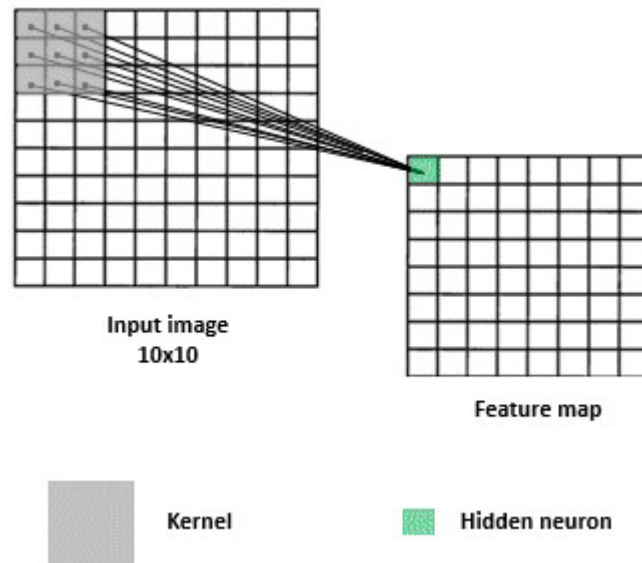


Рисунок 2.5 - Приклад з'єднання нейронів згорткової нейронної мережі.

Ресурс: [<http://www.ashokrahulgade.com/skills/DL/CNN/Introduction.html>]

Загальні ваги W і значення зсуву b нейронів забезпечують знаходження певної ознаки в різних частинах зображення. Нейрони, що мають загальні параметри, групуються і створюють підшар. Один прихований шар мережі може мати кілька підшарів, кожен з яких відповідає за певну ознаку, відмінних від інших. Таким чином, в одній і тій же області зображення можуть бути виявлені кілька різних ознак. Підшари формують карту ознак даного прихованого шару (рисунок 2.6).

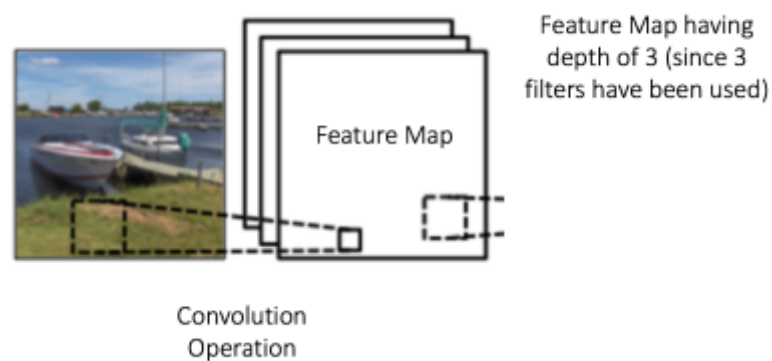


Рисунок 2.6 – Карта ознак. Ресурс:

[https://medium.com/@chriskevin_80184/feature-maps-ee8e11a71f9e]

Субдискретизація або пулінг. Приблизне розташування ознак об'єкта відносно один одного є більш важливою інформацією, ніж їх точні місця розташування в зображенні, оскільки в різних зразках один і той же об'єкт може бути зображений по-різному. Простим і ефективним способом зниження точності даних про становище ознак є ущільнення карти ознак (рисунк 2.7) [47].

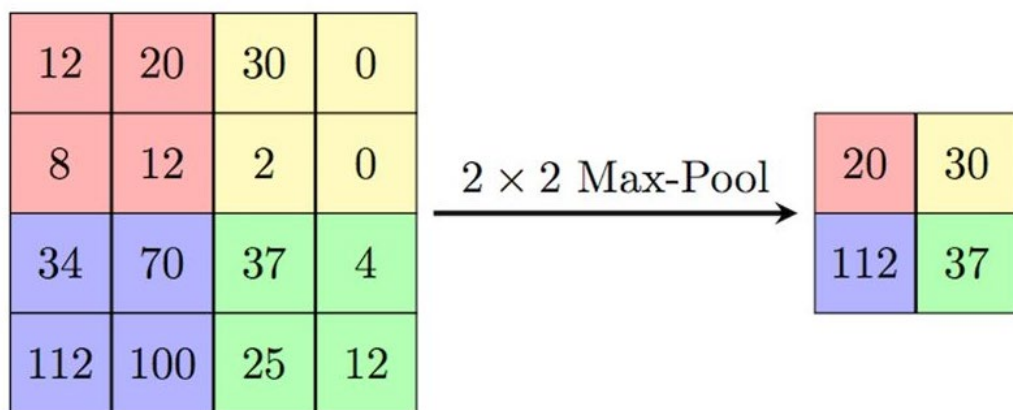


Рисунок 2.7 – Приклад субдискретизації. Ресурс:

[\[https://slideplayer.fr/slide/13975436/\]](https://slideplayer.fr/slide/13975436/)

Найбільш поширена форма архітектури згорткової мережі складається з кількох шарів згортки і ReLU, потім йде шар пулінгу, така послідовність повторюється кілька разів, поки зображення не вийде досить маленького розміру. В кінці додаються кілька повнозв'язних шарів, які забезпечують виведення потрібної інформації, наприклад, номер класу, що відповідає вхідному тестовому зображенню. Шаблон архітектури можна представити наступним чином (2.1):

$$Input \rightarrow [[Conv \rightarrow ReLU] * N \rightarrow Pool?] * M \rightarrow [Fc \rightarrow ReLU] * K \rightarrow Fc, \quad (2.1)$$

де *Input* - вхідний шар, *Conv* - шар згортки, *ReLU* - шар функції активації, *Pool* - шар пулінгу, *Fc* – повнозв'язний шар, * *X* – повтор блоку *X* разів, ? – включення 1 або 0 разів.

Нейрон згорткового шару працює за таким самим принципом, як і нейрон повнозв'язної мережі типу перцептрон, за винятком того, що з'єднання з нейронами попереднього шару обмежена деякою областю. По суті, виконується операція згортки: значення пікселів зображення поелементно множаться на ядро згортки, тобто на значення ваг нейрона, а потім отримані добутки підсумовуються. Результат є значенням даного нейрона. Ілюстрація цих тверджень зображена на рисунку 2.8.

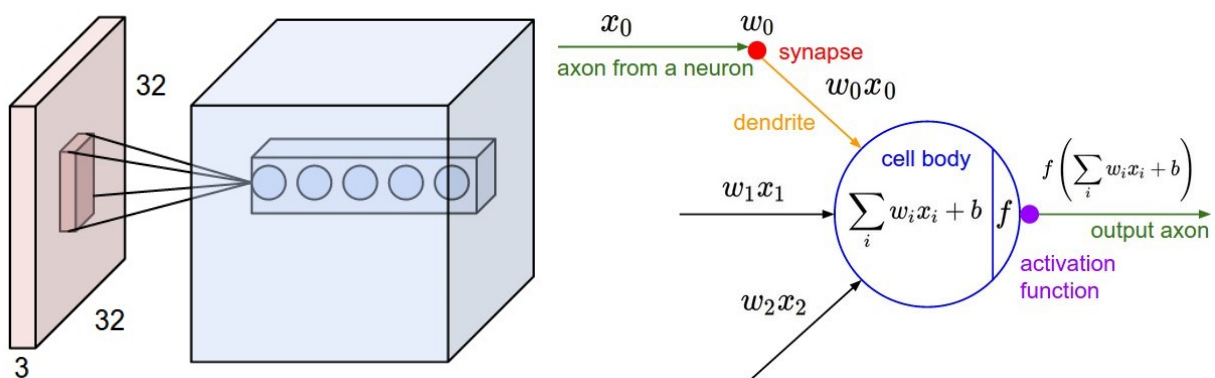


Рисунок 2.8 - Нейрон згорткового шару. Ресурс:

[\[https://habr.com/ru/post/456186/\]](https://habr.com/ru/post/456186/)

Параметри згорткового шару:

- **K** - кількість фільтрів або глибина;
- **F** - розмір фільтра;
- **S** – крок, з яким фільтр виконує прохід; розмір кроку визначає величину зсуву фільтра на кожному кроці. Чим більше крок, тим менше фільтр застосовується і тим менше розмір вихідної матриці;
- **P** - кількість заповнень нулями. Найчастіше ми використовуємо дану властивість, щоб зберегти просторовий розмір вхідної інформації з метою збереження вхідних та вихідних значень висоти і ширини однаковими. Без використання доповнення нулями отримуємо вузьку згортку [47].

Розмір вихідних даних $W_2 \times H_2 \times D_2$ обчислюється таким чином (2.2 – 2.4):

$$W_2 = \frac{W_1 - F + 2P}{S} + 1 \quad (2.2)$$

$$H_2 = \frac{H_1 - F + 2P}{S} + 1 \quad (2.3)$$

$$D_2 = K \quad (2.4)$$

де W_1 - ширина, H_1 - висота, D_1 - глибина вхідних даних.

На шарі ReLU (Rectified linear unit) застосовується функція активації до кожного елемента вхідного шару. Ця функція має наступну формулу (2.5):

$$f(x) = \max(0, x) \quad (2.5)$$

та реалізує пороговий перехід в нулі. На відміну від сигмоїдної і гіперболічної функцій ReLU не вимагає операцій з великими накладними витратами, крім того було встановлено, що швидкість збіжності градієнтного спуску набагато вище швидкості самих функцій [48].

Як згадувалося раніше, шар субдискретизації або пулінг необхідний для ущільнення карти ознак, що в свою чергу веде до суттєвого зменшення кількості параметрів і операцій, а також дозволяє уникнути перенавчання. Кожен підшар карти проходить через нелінійне перетворення, і в результаті зменшується в розмірі. Найчастіше в якості опції перетворення використовується функція максимуму або усередненого значення (рисунок 2.9).

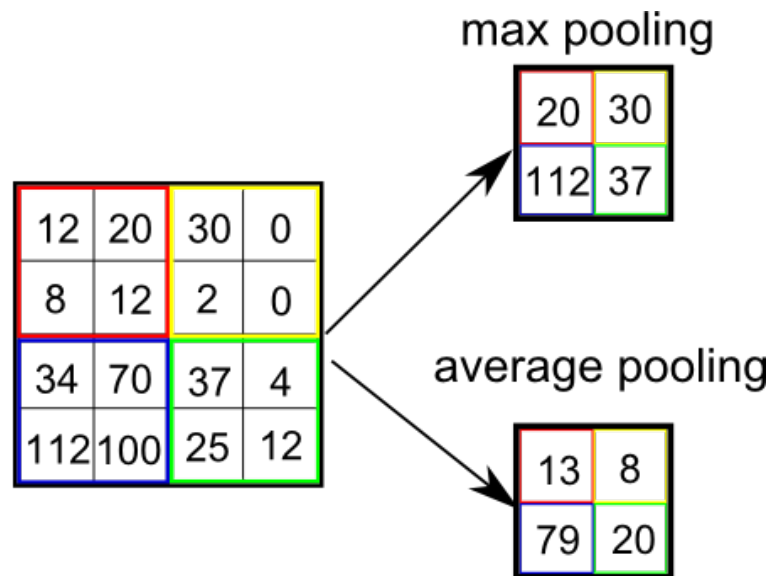


Рисунок 2.9 - Приклад пулінгу при розмірі фільтра $F = 2 \times 2$ і кроці $S = 2$

Ресурс: [<http://www.ashokrahulgade.com/skills/DL/CNN/Introduction.html>]

Даний шар має 2 параметри:

- F - розмір фільтра, найчастіше $F = 2 \times 2$;
- S - крок.

Розмір вихідного шару $W_2 \times H_2 \times D_2$ розраховується за такими формулами (2.6 – 2.8):

$$W_2 = \frac{W_1 - F}{S} + 1 \quad (2.6)$$

$$H_2 = \frac{H_1 - F}{S} + 1 \quad (2.7)$$

$$D_2 = D_1 \quad (2.8)$$

де W_1 - ширина, H_1 - висота, D_1 - глибина даних з вхідного шару.

2.5 Висновки до другого розділу

У розділі було опрацьовано основні поняття доповненої реальності, вивчено процес розпізнавання та відслідковування об'єкта у доповненій реальності, розкрито основні поняття процесу розпізнавання обличчя. Оскільки підходів до виявлення та ідентифікації людини по зображенню її обличчя багато, було досліджено найбільш розповсюджені методи розв'язання цих задач, виявлено їх сильні та слабкі сторони, а також проведено порівняльний аналіз. Було встановлено, що багатосарові нейронні мережі є більш стійкими та нечутливими до сторонніх впливів, таких як нерівномірність освітлення, нахил та поворот голови, наявність бороди, вусів, окулярів тощо, зміни емоцій на обличчі. Окрім того, БНМ виявились прийнятними і по обчислювальній складності методу.

Оскільки видів нейронних мереж багато, необхідно було визначитись із тим, який саме використовувати для поставленої задачі. Найбільш підходящою виявилась згорткова нейронна мережа, оскільки вона направлена на роботу з зображеннями та має властивості, які дозволяють досягти інваріантності до зміщення, масштабування і спотворення зображення: локальні рецептивні поля, загальні ваги, субдискретизація.

РОЗДІЛ 3

ОПИС РЕАЛІЗАЦІЇ ПРОГРАМНОГО ЗАСТОСУВАННЯ

3.1 Побудова цільової функції та аналіз технічного завдання

Для поставленої задачі була розроблена наступна цільова функція (3.1):

$$\sum_{i=0}^k BestMatch_i \left(Descr \left(Detec \left(T_{features} \right) \right), Descr \left(Detec \left(E_{features} \right) \right) \right) \rightarrow max,$$

$$BestMatch_i = \{ \overrightarrow{Match}_{t,e} \mid t.dist < 0,7 \cdot e.dist, t = \overline{0, n-1}, e = \overline{0, n-1} \} \quad (3.1)$$

де T - тестове зображення, E - еталонне зображення, $T_{features}$ і $E_{features}$ – особливості T і E відповідно, $Detec()$ - детектор, $Descr()$ - дескриптор, $\overrightarrow{Match}_{t,e}$ – вектор відповідностей між дескрипторами зображень T та E
 $BestMatch_i$ – кількість відповідностей між дескрипторами, що задовольняють обмеженню, $dist$ - відстань між дескрипторами.

Вхідними даними розробленої програми є база даних обличч людей та інформації про них. У результаті виконання програми знаходиться найкраще співпадіння обличчя, відображуваного у відеопотоці, з обличчям з бази даних та відтворення інформації про цю людину на відео шляхом використання доповненої реальності.

Програма виконана у вигляді застосування з графічним інтерфейсом, де взаємодія з користувачем відбувається через GUI.

3.2 Вибір архітектури згорткової мережі для розпізнавання обличчя

Після перемоги AlexNet [48] на конкурсі класифікації LSVRC2012, глибинна залишкова мережа (ResNet), розроблена у Microsoft Research [49], стала, можливо, найбільш новаторською роботою в області комп'ютерного зору та спільноті глибинного навчання за останні кілька років. ResNet дає

можливість навчати до сотень або навіть тисяч шарів і досягає переконливих показників.

Скориставшись своєю потужною репрезентативною здатністю, продуктивність багатьох програм комп'ютерного зору, відмінних від класифікації зображень, була підвищена. Наприклад, виявлення об'єктів і розпізнавання облич.

Згідно з загальною теоремою апроксимації, ми знаємо, що мережа прямого поширення з одним шаром може відтворити будь-яку функцію. Але шар має бути великим, через що нейрона мережа може бути схильною до перенавчання, що дасть погані результати. Для розв'язання цієї проблеми мережева архітектура має бути більш глибокою.

Починаючи з AlexNet, сучасна архітектура згорткових нейронних мереж (CNN) стає все глибшою. Якщо AlexNet мала лише 5 згорткових шарів, то архітектура мереж наступного покоління VGG та GoogleNet мали вже 19 і 22 шари відповідно. Однак збільшення глибини мережі не полягає просто в об'єднанні все більшої кількості шарів. Глибинні мережі важко тренувати через проблему зникнення градієнта - оскільки градієнт повертається назад до більш ранніх шарів, повторне множення може зробити градієнт нескінченно малим [50]. У результаті, оскільки мережа стає глибше, її продуктивність стає насиченою або швидко починає знижуватись (рисунок 3.1).

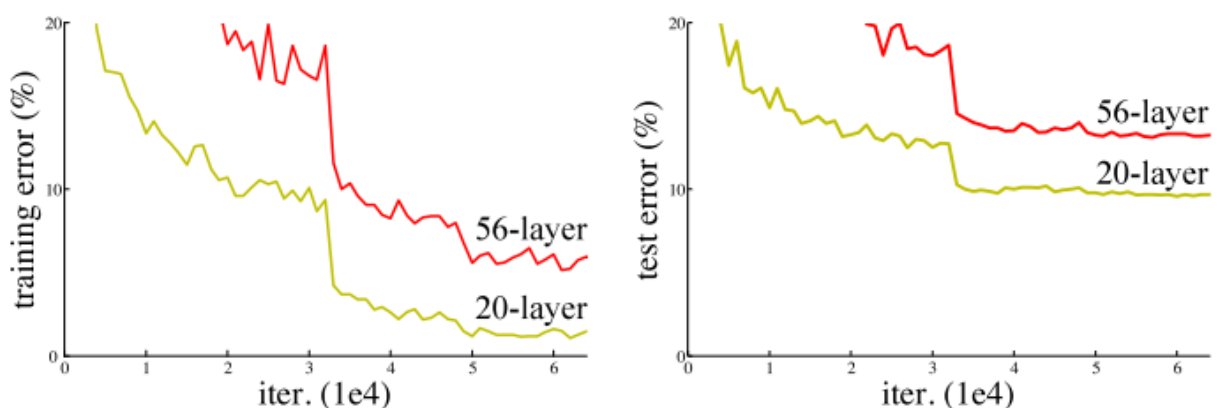


Рисунок 3.1 - Збільшення кількості шарів призводить до погіршення продуктивності

Основною ідеєю ResNet є введення блоку ідентичності (від англ. “identity shortcut connection”) або залишкового шару, який пропускає один або кілька шарів, як показано на рис. 3.2.

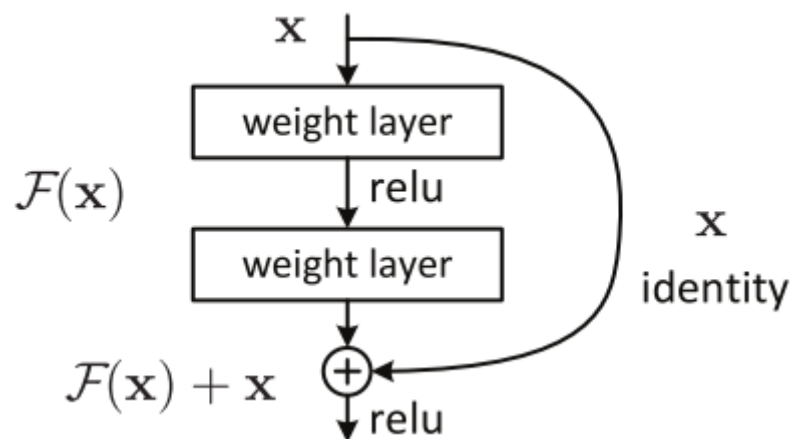


Рисунок 3.2 - Блок нейромережі ResNet

Автори цієї архітектури стверджують, що об’єднання шарів не повинно погіршувати продуктивність мережі, тому що можна просто складати ідентифікаційні відображення (шар, який нічого не робить) на поточній мережі, і продуктивність нейромережі з такою архітектурою не зміниться [51]. Це вказує на те, що більш глибока модель не повинна виробляти тренувальної похибки вище, ніж її дрібніші аналоги.

Отже, може бути корисним явно змусити мережу вивчити ідентифікаційне відображення, дізнавшись залишковий вхід і вихід деяких шарів (або підмереж). Припустимо, що вхід підмережі є x , true вихід це $H(x) = x$. Залишковість, це різниця між ними (3.2):

$$F(x) = H(x) - x \quad (3.2)$$

Оскільки ми зацікавлені в знаходженні true, що лежить в основі виходу підмережі, ми потім переписуємо це рівняння на наступне (3.3):

$$H(x) = F(x) + x \quad (3.3)$$

Потім це значення проходить через Rectified Linear Unit (ReLU). Це дозволяє переносити важливу інформацію з попереднього шару до наступного. При цьому ми можемо запобігти зникненню проблеми градієнта. Навіть якщо це з'єднання виглядає як доповнення до стандартного підходу CNN, це пришвидшує тренування мережі.

Ось у цьому полягає різниця між ResNet та традиційними нейронними мережами. Там де звичайні нейромережі будуть вчити $H(x)$ напряму, мережі ResNet будуть вивчати різницю між входом і виходом підмережі. Це надасть мережі можливість просто пропустити підмережі шляхом створення $F(x) = 0$, так що $H(x) = x$. Іншими словами, вихід певної підмережі - це тільки вихід останньої підмережі.

Під час зворотнього поширення похибки, вивчення саме різниці між входом та виходом дає нам приємну властивість. Завдяки формулюванню мережа може ігнорувати градієнт деяких підмереж і пересувати градієнт від вищих шарів до нижніх шарів без будь-якої модифікації. Це означає, що ResNet може просто передати градієнт з останнього шару, наприклад, шару 151, безпосередньо до першого шару.

3.3 Розробка алгоритму програми

Для розпізнавання облич у програмі використовується наступний алгоритм:

1. Знайти всі обличчя на зображенні з відеопотоку.
2. Розпізнати кожне обличчя, навіть якщо воно знаходиться під кутом або під поганим освітленням.
3. Визначити унікальні риси обличчя, що відрізняють одну людину від іншої, наприклад, форма обличчя, розмір очей тощо.

4. Порівняти знайдені особливості цього обличчя з усіма людьми, обличчя яких система вже знає, щоб зрозуміти, хто зараз зображений на відео.

З'єднавши цей ланцюг дій та додавши накладення елементів доповненої реальності, ми отримаємо розв'язок поставленої задачі (рис. 3.3).

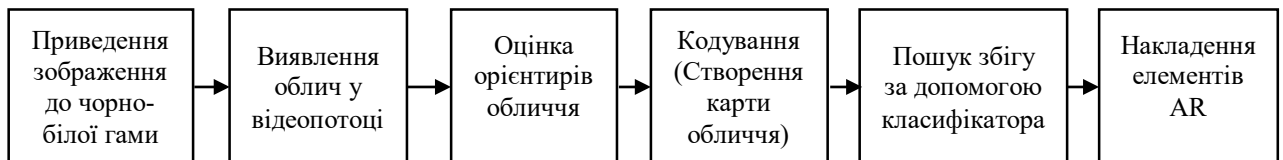


Рисунок 3.3 - Загальний алгоритм розробленої програми

Розглянемо кожен крок більш детально.

Крок 1. Пошук всіх облич.

На першому етапі алгоритму потрібно знайти всі обличчя. Для цього було використано гістограму направлених градієнтів (від англ. Histogram of Oriented Gradients, HOG).

Спочатку зображення приводиться до чорно-білої гами, тому що для пошуку обличчя дані про колір є надлишковими. Потім кожен піксель розглядається окремо і порівнюється з сусідніми. Мета цих дій – виявити, наскільки темним є поточний піксель у порівнянні з сусідніми. Потім малюється стрілка, що вказує, в якому напрямку зображення стає темнішим (градієнт). Таким чином алгоритм отримує градієнтне зображення обличчя. Для кожної людини напрям зміни кольору на обличчі від темного до світлого буде однаковим при будь-якому освітленні.

Крок 2. Оцінка орієнтирів обличчя

Після знаходження обличчя необхідно розібратися з проблемою, коли обличчя, повернуті у різні сторони, здаються комп'ютеру різними людьми.

Для вирішення цієї проблеми кожне зображення перетворюється так, щоб очі і губи завжди знаходились в якомусь визначеному місці. Це спрощує задачу порівняння облич на наступних кроках. Для цього використовується алгоритм оцінки орієнтирів обличчя (від англ. – “face landmarks estimation”).

Основа ідея полягає в тому, що ми відмічаємо 68 особливих точок (які називаються орієнтирами), що існують на кожному обличчі – верхня частина підборіддя, зовнішня точка кожного ока, внутрішня точка кожної брови тощо. Після цього алгоритм машинного навчання знаходить ці 68 точок на кожному обличчі.

Коли відомо, де очі, а де рот, ми можемо повертати, масштабувати і зсувати зображення так, щоб очі і рот були якомога краще відцентровані. Для цього було використано афінні перетворення.

Крок 3. Кодування обличчя

Потрібно знайти спосіб зняти кілька основних вимірів з кожного обличчя, які можна було б порівняти із найближчими відомими вимірами та знайти найбільш схоже обличчя. Наприклад, можна змінити розмір кожного вуха, відстань між очима, довжину носа тощо.

Проблема в тому що виміри, що здаються людям очевидними (наприклад, колір очей), насправді не мають сенсу для комп'ютера, що роздивляється окремі пікселі зображення. Дослідники показали, що найбільш точний підхід – дозволити комп'ютеру самому зробити виміри того, що йому потрібно. Алгоритми глибинного навчання самостійно визначають, які частини обличчя потрібно вимірювати, а які – ні.

Розв'язок полягає у створенні згорткової нейронної мережі, яка може створювати 128 вимірів для кожного обличчя. 128 вимірів кожного обличчя називають картою. Ідея перетворення масива необроблених даних, наприклад, зображення, в список генерованих комп'ютером чисел вкрай важлива для машинного навчання.

Кодування зображення обличчя

Цей процес навчання згорткової нейронної мережі для отримання мапи обличчя потребує великої кількості даних і великих обчислювальних потужностей. Саме тому було використано модель, навчену на кількох

наборах даних упродовж кількох тижнів. Таким чином залишається тільки пропустити зображення обличчя через готову навчену нейронну мережу, щоб отримати 128 вимірювань для кожного обличчя.

Крок 4: Пошук інформації людини за кодуванням

Останній етап є найпростішим у всьому процесі. Все, що потрібно, це знайти людину в базі відомих для розробленої програми людей, вимірювання якої найближчі до отриманих даних в реальному часі. Це робиться за допомогою класифікатора, що вимірює відстань між значеннями, і визначає, наскільки вони схожі.

3.4 Обґрунтування вибору засобів розробки програми

Для реалізації алгоритму обрано високорівневу мову програмування загального призначення Python версії 3.8. Синтаксис ядра Python мінімалістичний, однак у той же час для Python створено багато пакетів, що включають великий обсяг корисних функцій. Також на Python можна писати кросплатформні додатки, що у наш час є великим плюсом [52]. Для розпізнавання обличчя людей було використано бібліотеку `face_recognition`. Вона надає зручний API, який в свою чергу використовує бібліотеки `OpenFace`, `dlib`, `OpenCV`. Також ця бібліотека представляє методи роботи з якісно навченими моделями для визначення `face landmarks`, і згорткової нейронної мережі архітектури `ResNet34`.

Як середовище програмування обрано `PyCharm` - інтегроване середовище розробки для мови програмування Python. Воно надає засоби для аналізу коду, графічний відладчик, інструмент для запуску юніт-тестів і підтримує веб-розробку на `Django`. `PyCharm` розроблено компанією `JetBrains` на основі `IntelliJ IDEA`. Для створення застосування розпізнавання обличчя людини у відеопотоці з доповненою реальністю використано `PyCharm Community Edition`, оскільки ця версія є безкоштовною для освітніх установ і

проектів з відкритим вихідним кодом і нічим не відрізняється від платної версії продукту.

Інформаційна база проекту, а саме зображення облич людей, реалізована через збереження даних у локальному сховищі, у даному випадку – файловій системі, оскільки для даного дослідження можливість масштабування не є критично важливою, а даний спосіб збереження забезпечує ряд таких переваг, як швидкість, ціна та простота реалізації. Інформація про людей, місцеположення відповідного зображення їх обличчя, а також карта ознак зображення зберігається у базі даних MySQL.

3.5 Вибір методів оцінки програми

При виборі системи важливо правильно визначити не тільки мету, але і критерії для оцінки ефективності цієї системи. А це значить визначити пристосованість системи до роботи у певних умовах і різні пороги допустимих помилок. В даний час для оцінки ефективності систем розпізнавання виділяють 2 основні параметри. У біометрії їх називають Коефіцієнтом помилкового допуску (FAR) і Коефіцієнтом помилкової відмови (FRR). Коефіцієнт помилкової відмови показує ймовірність того, як часто обличчя бувають помилково відкинуті, або іншими словами - система не знайшла в базі “знайому” людину. Чим цей коефіцієнт менше, тим вище точність розпізнавання. Коефіцієнт помилкового допуску - це навпаки, ймовірність того, як часто система помилково класифікує обличчя, якого немає в базі, як знайоме. Існують ще один коефіцієнт, що дозволяє порівнювати біометричні системи - коефіцієнт EER (Equal error rate). Це коефіцієнт, при якому обидві помилки (помилка допуску і помилка відхилення) еквівалентні. Чим нижче коефіцієнт EER, тим вище точність біометричної системи [53]. На рисунку 3.4 показано взаємозв'язок характеристик FAR, FRR і EER.

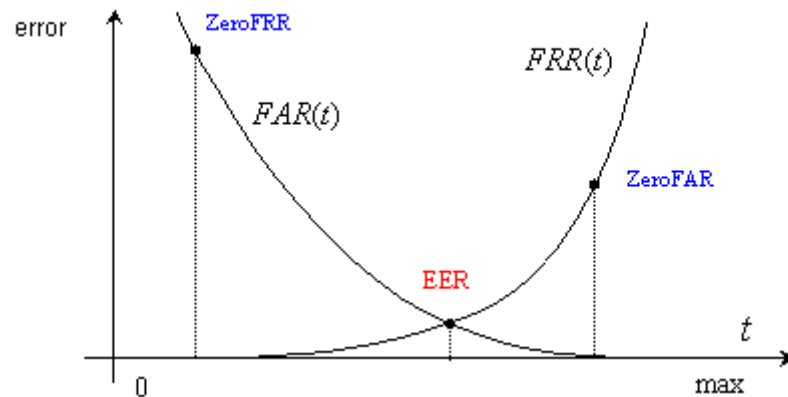


Рисунок 3.4 – Характеристики біометричних систем

Для порівняння систем і візуалізації залежності між помилками використовують графік ROC (Receiver Operating Characteristic). Крива ROC визначає, наскільки зразок повинен бути близький до шаблону, щоб система могла вважати це збігом. При регулюванні порогів помилок, варто не забувати, що величини FAR і FRR взаємодоповнюючі і якщо поріг зменшити, то буде менше помилкових відмов, але більше помилкових допусків.

3.6 Розробка програмного застосування

Спроектована архітектура системи розпізнавання обличчя у відеопотоці з доповненою реальністю зображена на рисунку 3.5.

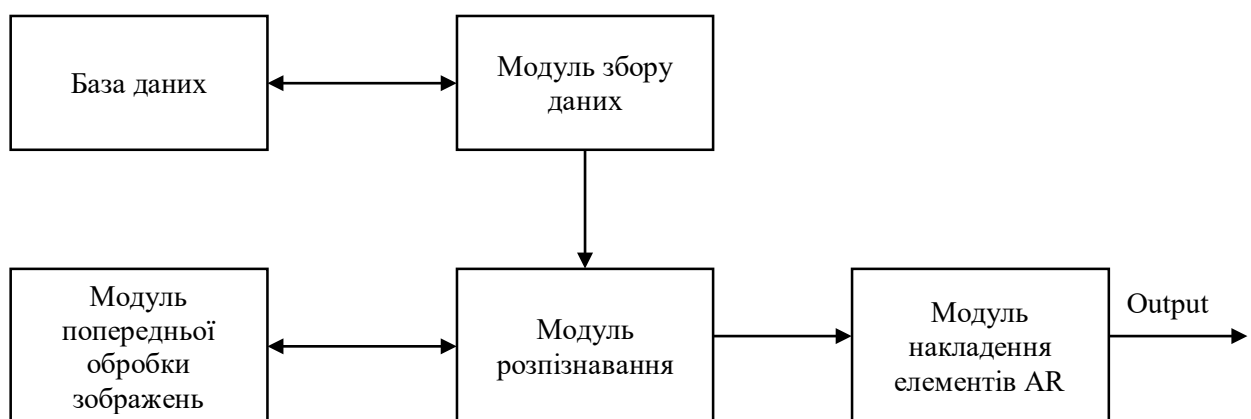


Рисунок 3.5 – Архітектура системи розпізнавання обличчя у відеопотоці з доповненою реальністю

Базою даних (БД) для збереження зображень типу .jpg даної системи слугує файлова система, оскільки у збереження файлів, як файлів є ряд переваг:

- швидкість: при зберіганні в БД нам необхідно зробити наступні дії для виведення зображення: зчитати стрім з бази, створити тимчасовий файл, записати в нього стрім; З файлової системи достатньо тільки зчитати.
- універсальність. Звичайні зображення можна використовувати для багатьох додатків: переслати через FTP, додати до листа, показати у браузері; бінарні дані з бази необхідно спочатку конвертувати.
- простота реалізації: завантажити файл на сервер і зберегти посилання на нього в таблицю легше, ніж реалізувати логіку збереження бінарних даних у ту ж саму таблицю, а простота - це швидкість розробки і внесення змін.
- надійність [54].

У базі даних MySQL міститься інформація про користувачів, місцеположення зображень у файловій системі, а також карти ознак облич. Збереження карти ознак відомих допомагає пришвидшити роботу програми, оскільки зникає необхідність щоразу перераховувати її для всіх наявних облич. Схему таблиць зображено на рисунку 3.6.



Рисунок 3.6 – Схема бази даних розробленої системи

Модуль збору даних включає:

- зчитування користувацьких даних;
- запис користувацьких даних у БД.

Модуль попередньої обробки зображень включає:

- приведення зображення до палітри сірих кольорів.

Модуль розпізнавання включає:

- виявлення обличчя;
- оцінка орієнтирів обличчя;
- виявлення ключових точок та обчислення дескрипторів;
- застосування матчера для зіставлення дескрипторів початкового та еталонного зображення.

Модуль накладення елементів AR відповідає за відображення графічних та текстових об'єктів у режимі реального часу на обличчя.

Розглянемо мережу ResNet34 (нейромережа на архітектурі ResNet з 34 шарами) на рис. 3.7.

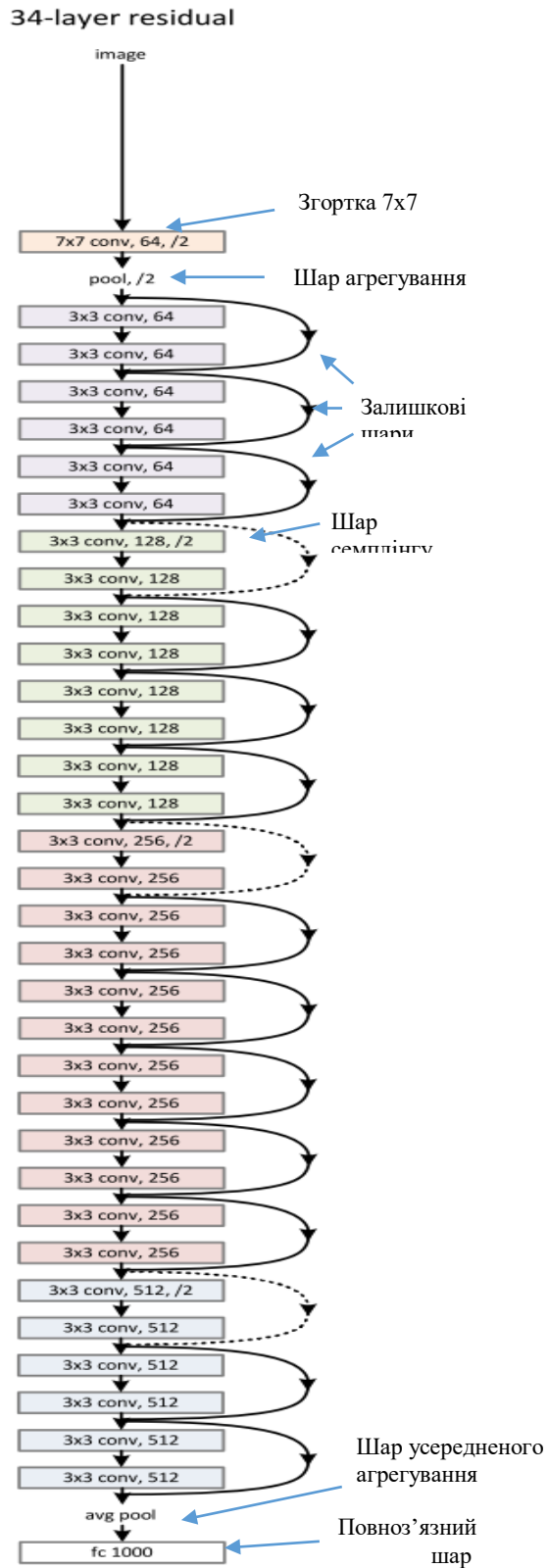


Рисунок 3.7 - Архітектура ResNet з 34 шарами. Ресурс:
[\[https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8\]](https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8)

У першому шарі ResNet використовує згортку 7×7 з кроком 2 для зменшення вхідного сигналу у 2 рази (від англ. “downsampling”), подібно шару агрегування (від англ. “pooling”). Потім йдуть три блоки ідентичності (або залишкові шари), після чого кількість пікселів вхідного зображення знову зменшується у 2 рази. Шар семплінгу (зменшення зображення) також є згортковим шаром, але не з’єднується з блоком ідентичності. Ці блоки архітектури повторюються аналогічним чином кілька разів. Останній шар - це шар усередненого агрегування (від англ. “average pooling”), який створює 1000 карт властивостей (для даних ImageNet) і середнє значення для кожної карти об’єктів. Результатом буде 1000-мірний вектор, який потім подається безпосередньо в шар Softmax, тому він повнозв’язний (від англ. “fully connected”) [55]. Основне призначення функції softmax полягає в перетворенні (ненормалізованого) виводу одиниць K (який, наприклад, представлений у вигляді вектору K елементів) повнозв’язного шару, у розподіл ймовірностей (нормалізований вихід), який часто представлений як вектор K елементів, кожен з яких знаходиться між 0 та 1 (ймовірність), а сума усіх цих елементів дорівнює 1 (розподіл ймовірності). У випадку задачі класифікацією i -й елемент вектору, що створюється функцією softmax, відповідає ймовірності входу мережі, що належить до i -го класу.

3.7 Опис інтерфейсу програми

Інтерфейс додатку складається з головного вікна, на якому розміщено чотири кнопки та двома текстовими полями, призначеними для отримання інформації від користувача та додання її до бази даних. Вигляд інтерфейсу можна побачити на рисунку 3.8. На першому вікні розташовані дві кнопки: “Start recognition” та “Add to Database”.

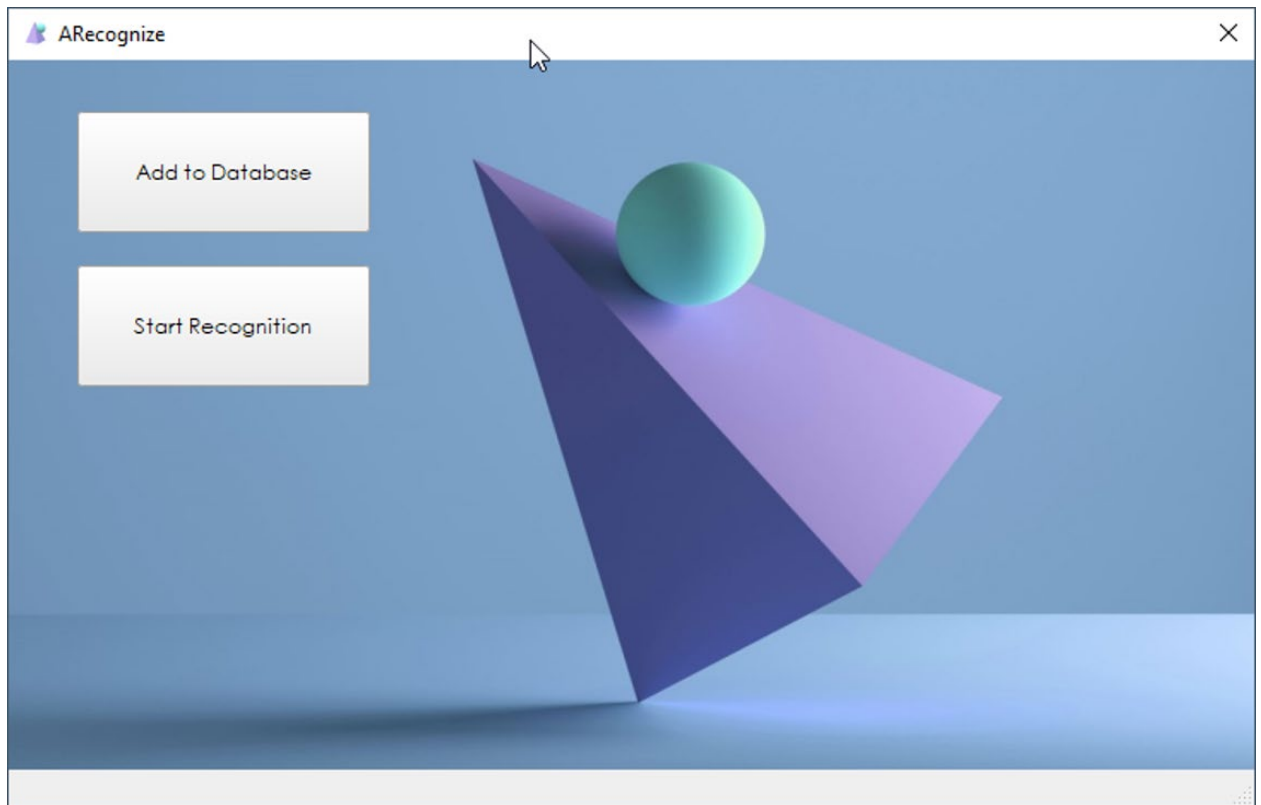


Рисунок 3.8 – Головне вікно програми

Кнопка “Start Recognition” запускає процес розпізнавання обличчя людини. На екрані з’являється вікно захоплення відеопотоку. Як тільки знайдено обличчя людини з бази даних, що найбільш збігається з зображуваним у відеопотоці, додаткова інформація про користувача витягується з бази та з’являється на відео у вигляді квадрата з текстом.

Кнопка “Add to Database” призначена для додавання нового обличчя людини та інформації про неї до бази даних (рисунок 3.9). Відкриється вікно, у якому користувачу необхідно вказати інформацію про себе (ім’я, вік, стать та національність). Фото можна зробити з програми за допомогою кнопки “Take Photo” - відкриється додаткове вікно, у якому користувачу необхідно зробити знімок – для цього потрібно натиснути клавішу “Q”. Іншою опцією додавання фотографії є завантаження зображення з файлової системи. Для цього необхідно скористатись кнопкою “Upload photo”. Завантажений або щойно зроблений знімок з’явиться справа від заповненої форми. Як тільки усі

дані заповнено, необхідно натиснути кнопку “Insert Info to Database” - алгоритм проведе обробку зображення, збудує карту ознак та разом з персональною інформацією занесеться до бази даних.

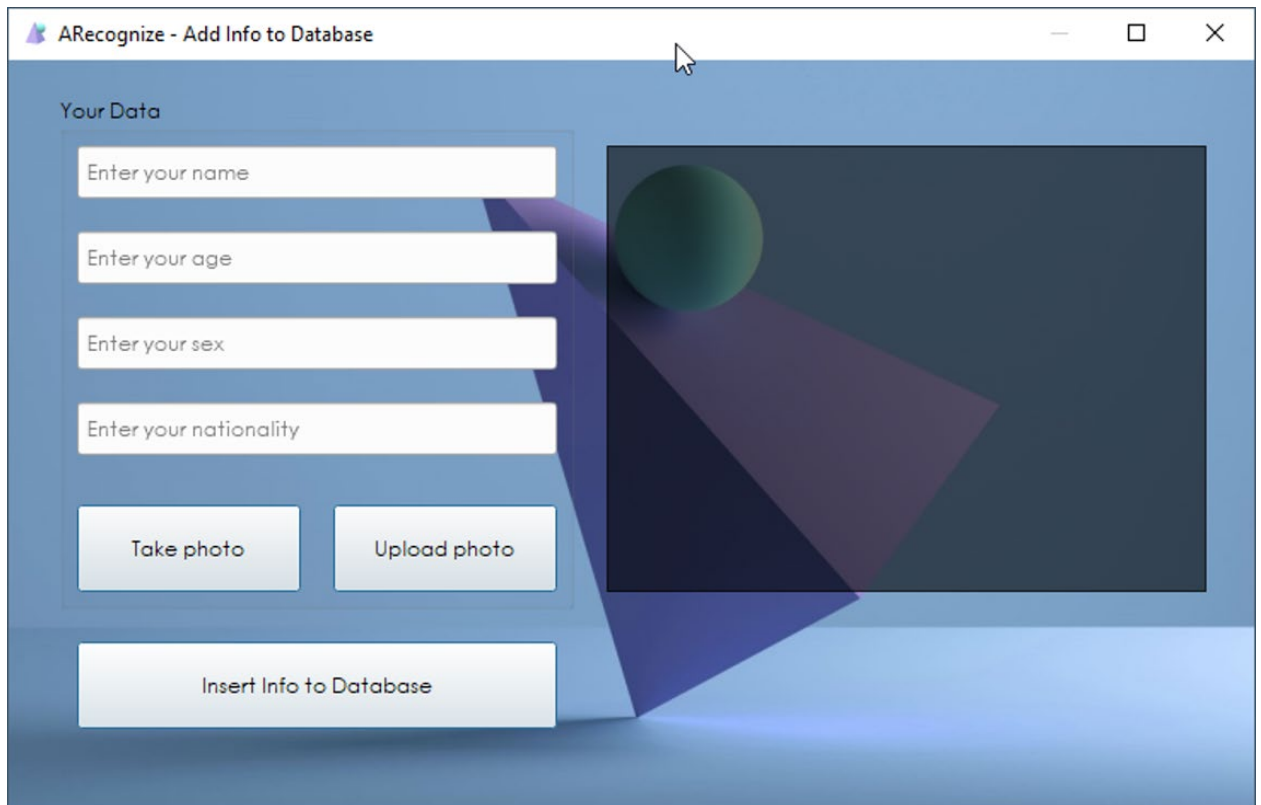


Рисунок 3.9 – Вікно програми, що дозволяє додавати нову інформацію про людину до бази даних

Для виходу з програми або закриття вікна необхідно скористатись стандартною кнопкою “Close”.

3.8 Аналіз результатів роботи програмного застосування

Розроблена програмна система, призначена для розпізнавання обличчя людини у відеопотоці з доповненою реальністю, показала прийнятні результати.

Додаток правильно розпізнає обличчя, наявне у базі даних, в умовах різного освітлення, повороту голови, за наявності аксесуарів, закриття деяких частин обличчя, зміни емоцій тощо. Аналогічно для розпізнання кількох облич

одночасно. Систему було протестовано на 840000 кадрах: 4 людей окремо та разом у різних комбінаціях (усього 15 комбінацій) за різних умов освітлення, шумів, перешкод, аксесуарів, емоцій. Умови та кількість кадрів для кожної комбінації наведено у таблиці 3.1.

Таблиця 3.1 - Умови та кількість кадрів для кожної комбінації, які використовувались для тестування

Умова	Кількість кадрів для кожної комбінації
Масштабування та нахил/поворот голови	7000
Рівномірна освітленість	4000
Нерівномірна освітленість	4000
Темрява	4000
Наявність шумів	7000
Використання аксесуарів	10000
Зміна емоцій	10000
Наявність фізичних перешкод	10000

Приклади результатів роботи застосування зображено на рисунках 3.10 – 3.21.

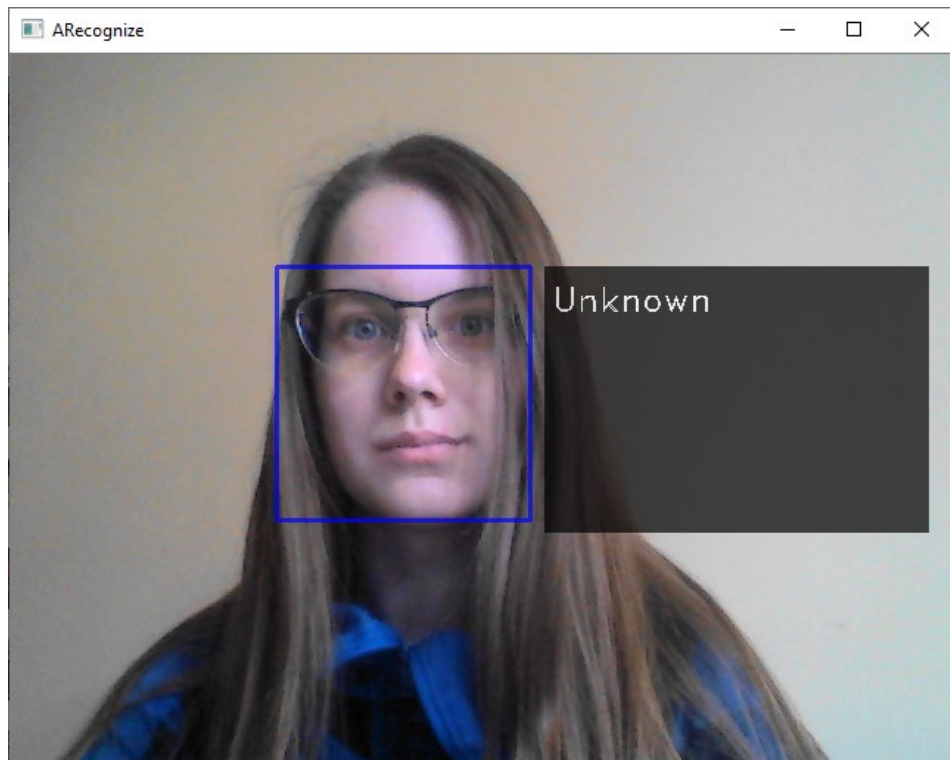


Рисунок 3.10 – Розпізнавання обличчя людини, не занесеної до бази

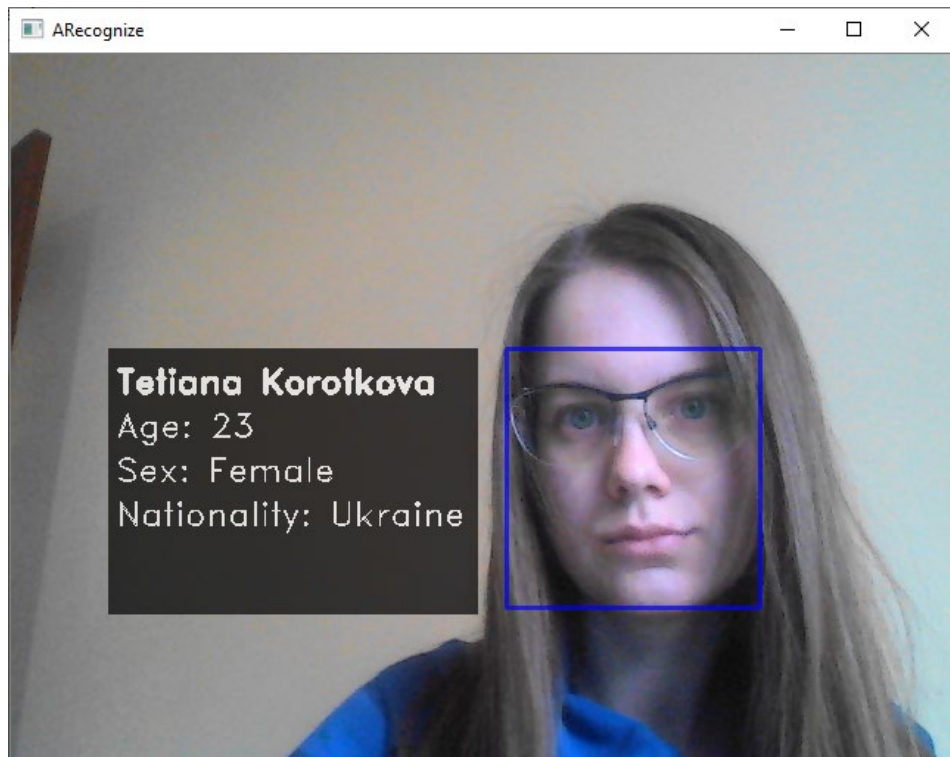


Рисунок 3.11 – Розпізнавання обличчя людини, занесеної до бази

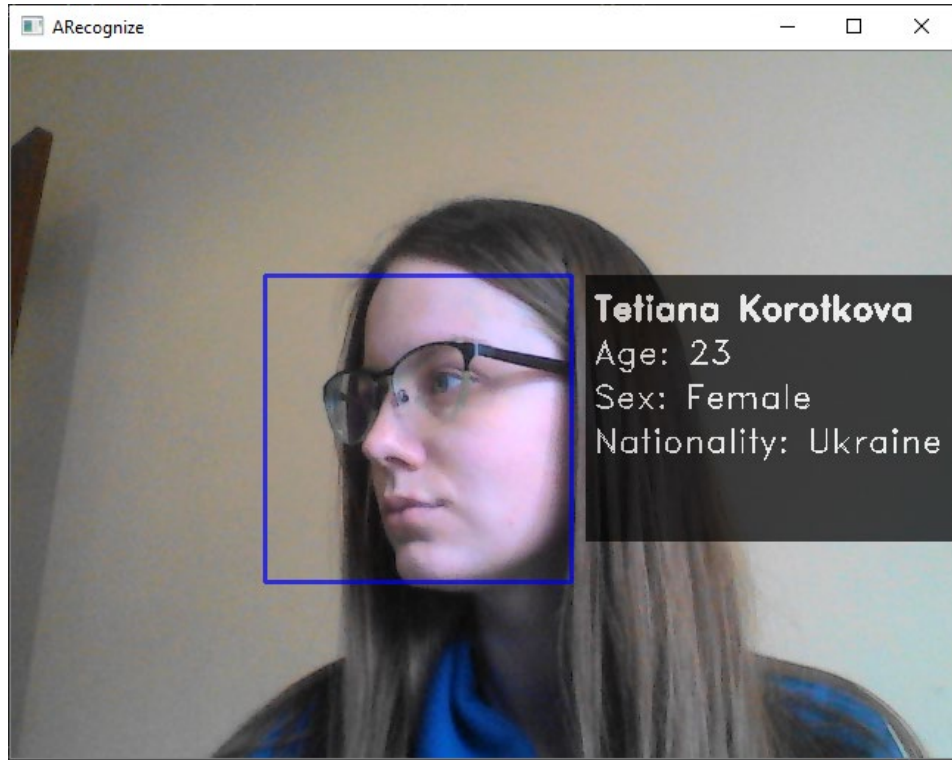


Рисунок 3.12 – Розпізнавання обличчя людини при повороті голови

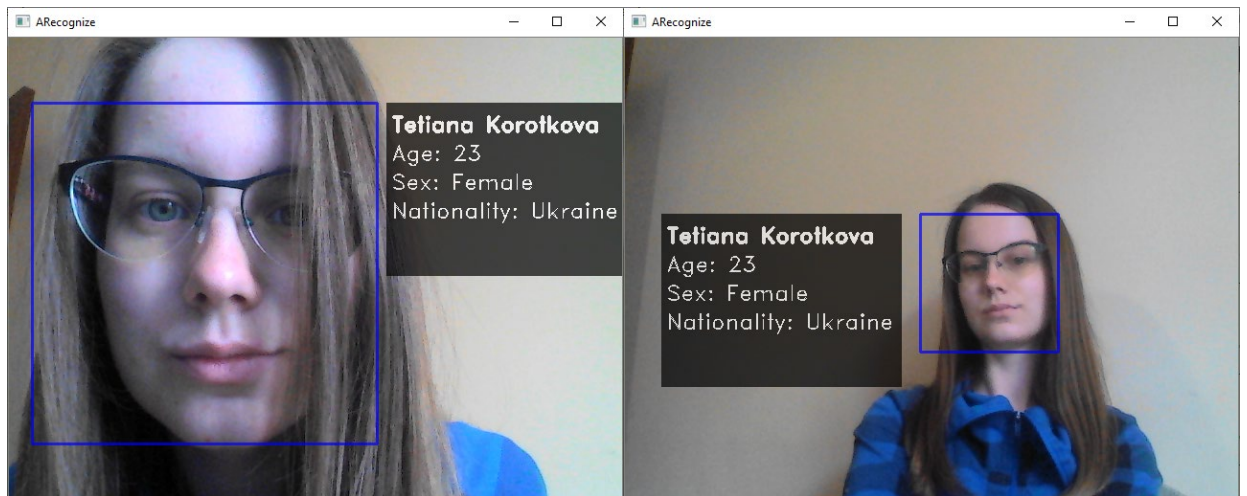


Рисунок 3.13 – Розпізнавання обличчя людини при масштабуванні

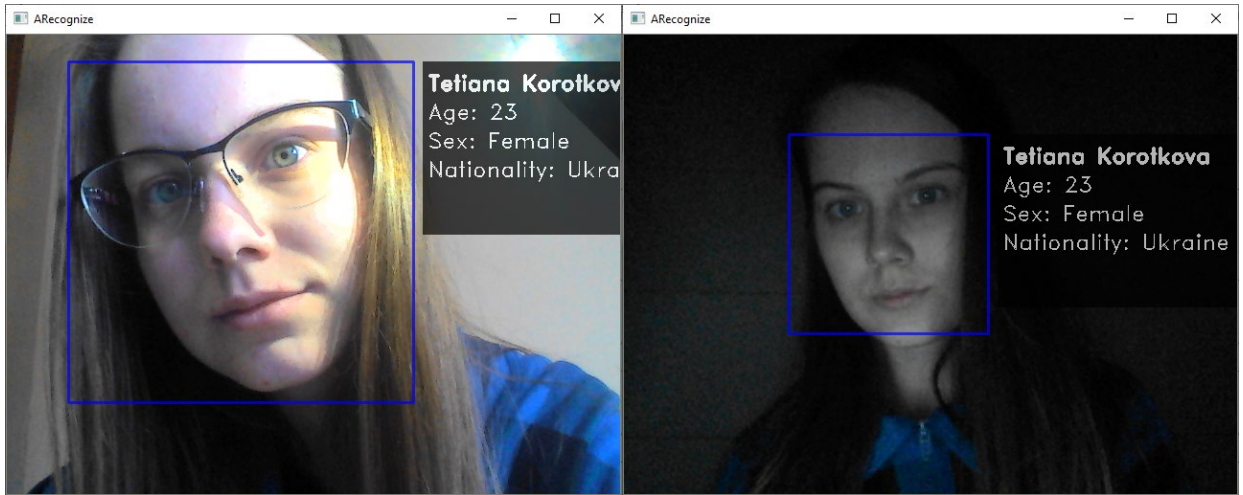


Рисунок 3.14 – Розпізнавання обличчя людини при різному освітленні

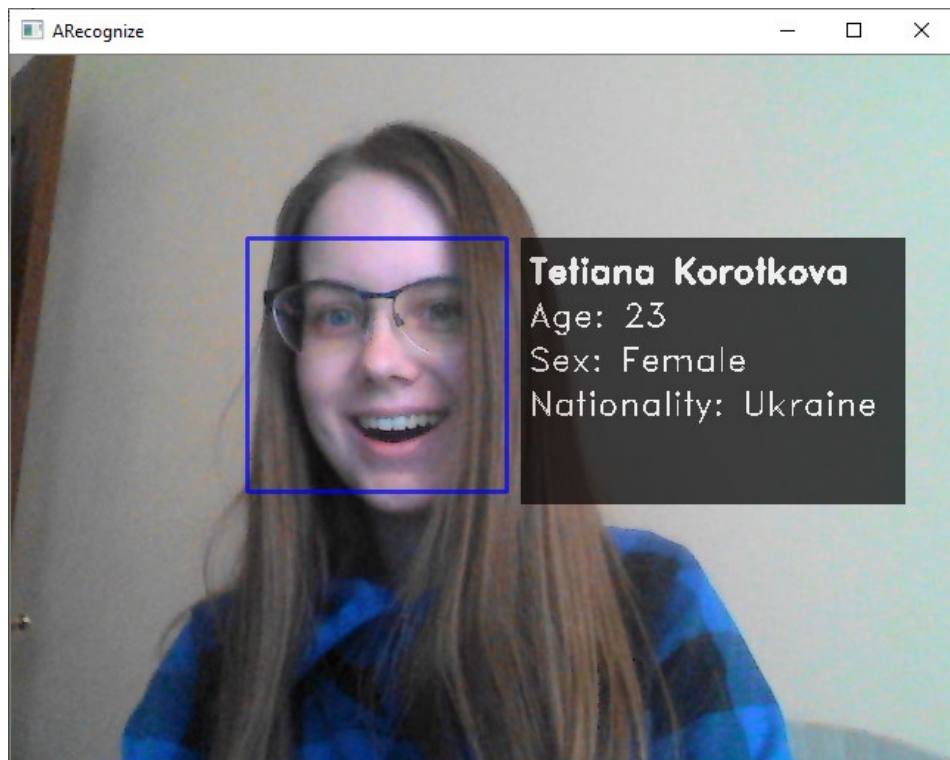


Рисунок 3.15 – Розпізнавання обличчя людини з радісним обличчям

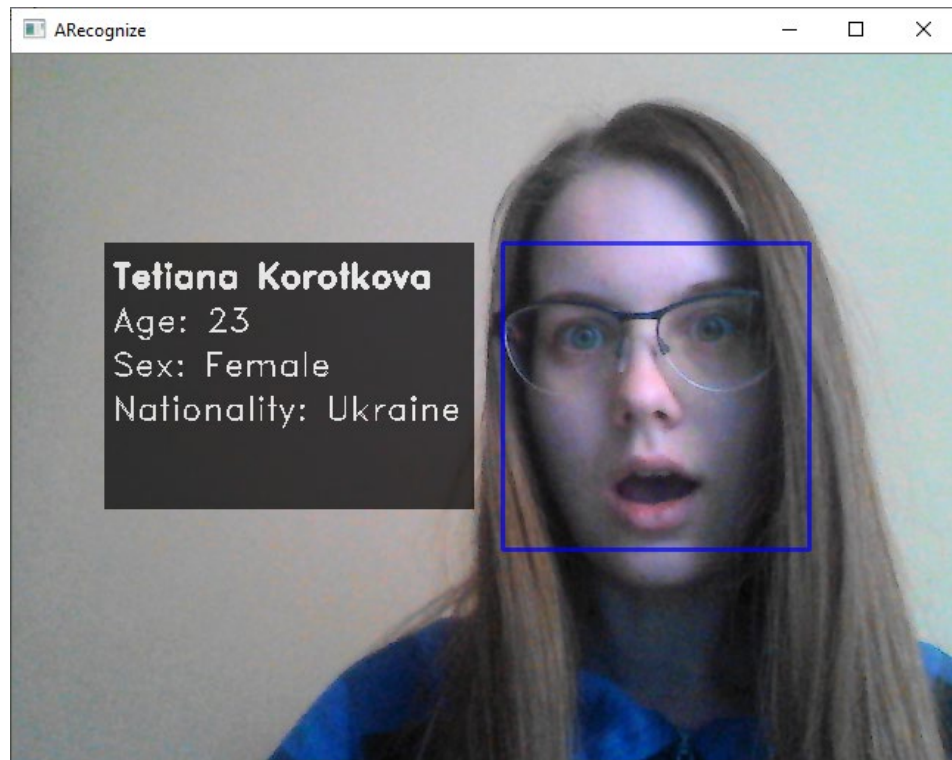


Рисунок 3.16 – Розпізнавання обличчя людини зі здивованим обличчям

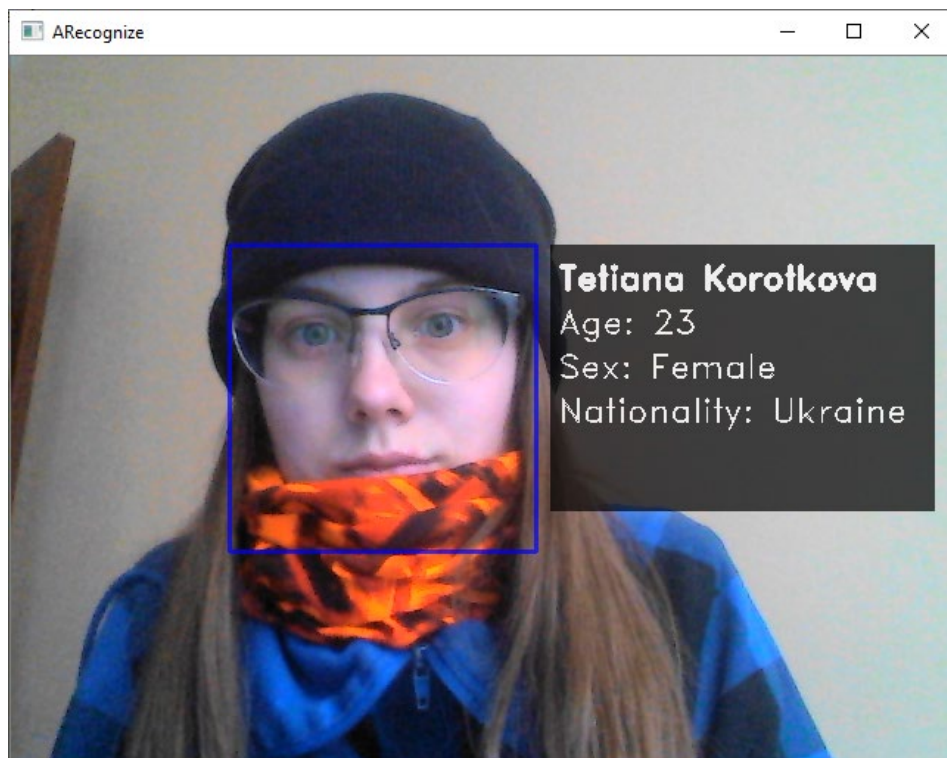


Рисунок 3.17 – Розпізнавання обличчя людини з аксесуарами

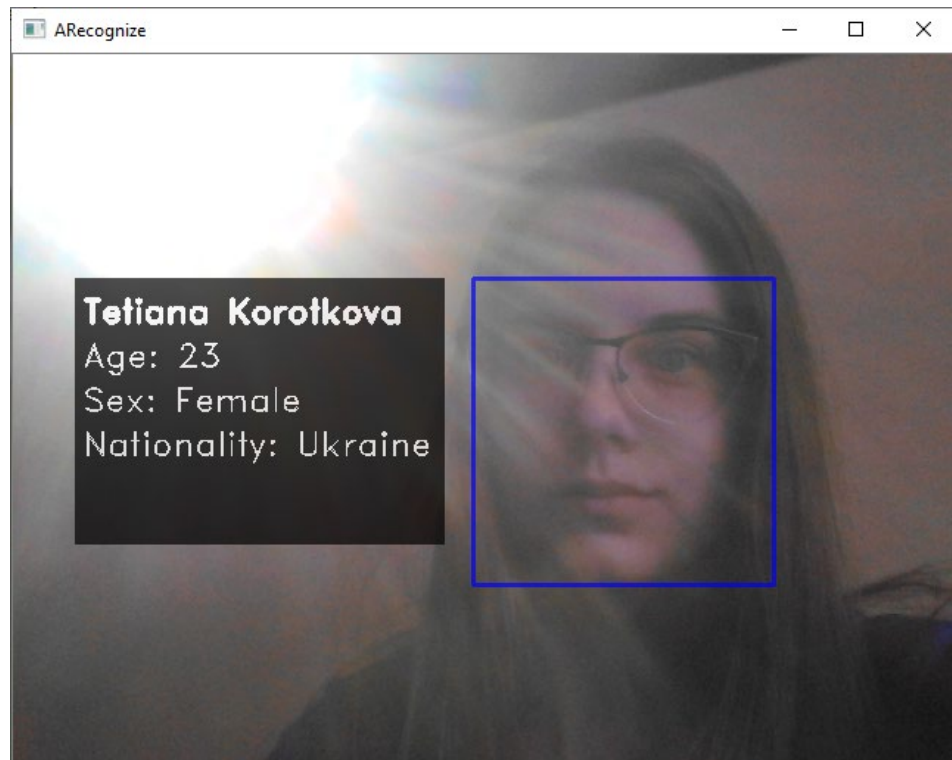
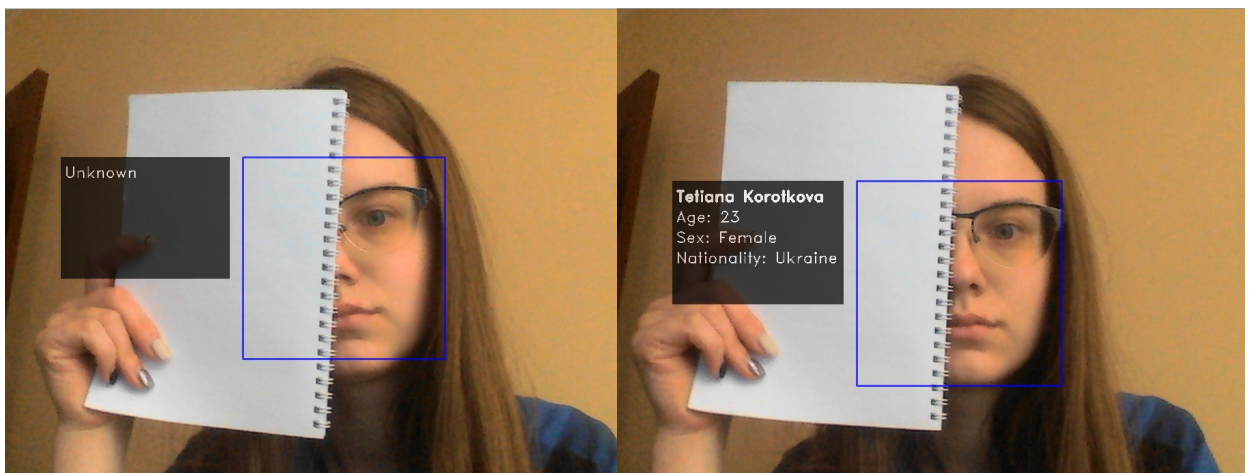


Рисунок 3.18 – Розпізнавання обличчя людини за наявності шумів



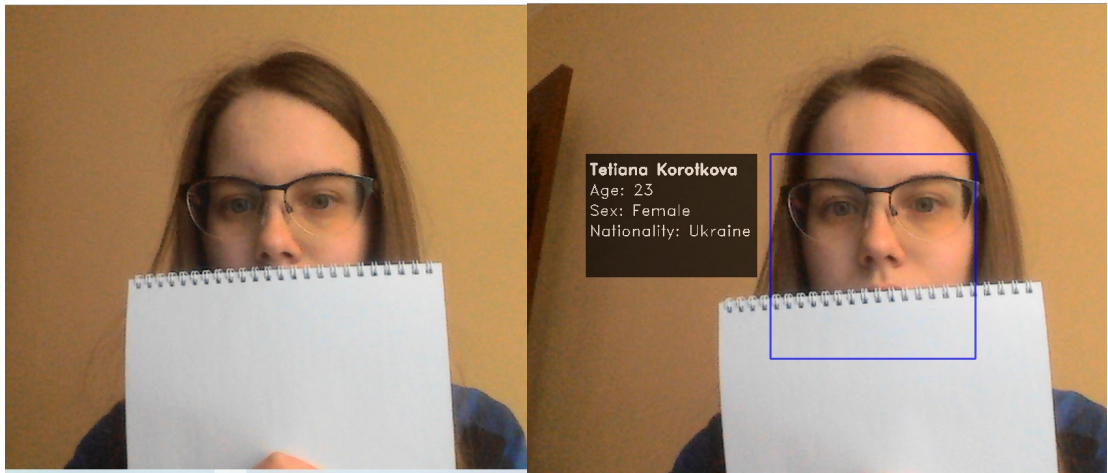


Рисунок 3.19 – Розпізнавання обличчя людини за наявності перешкод

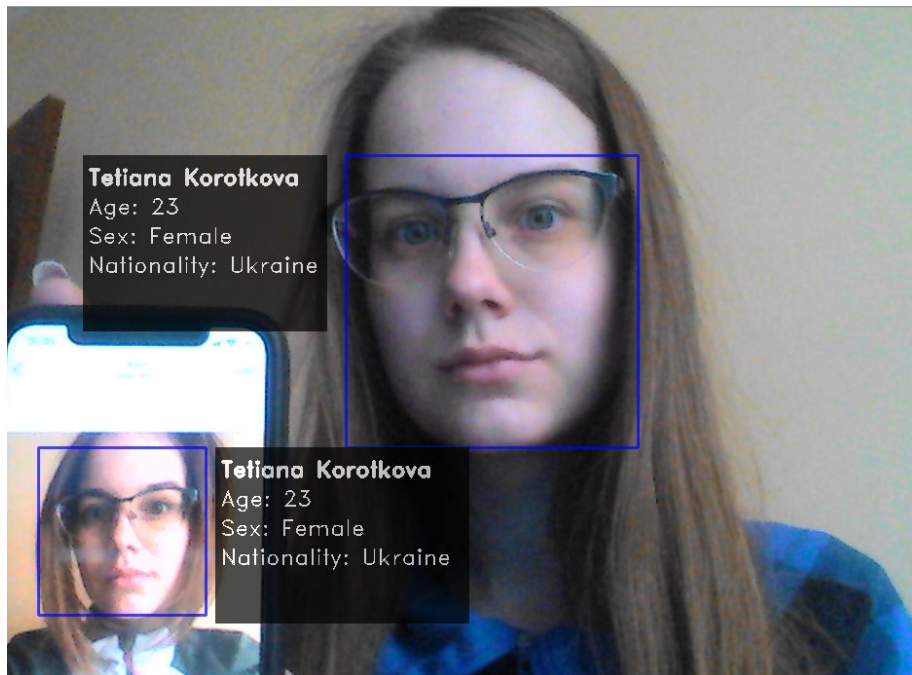


Рисунок 3.20 – Розпізнавання кількох облич одночасно - 1

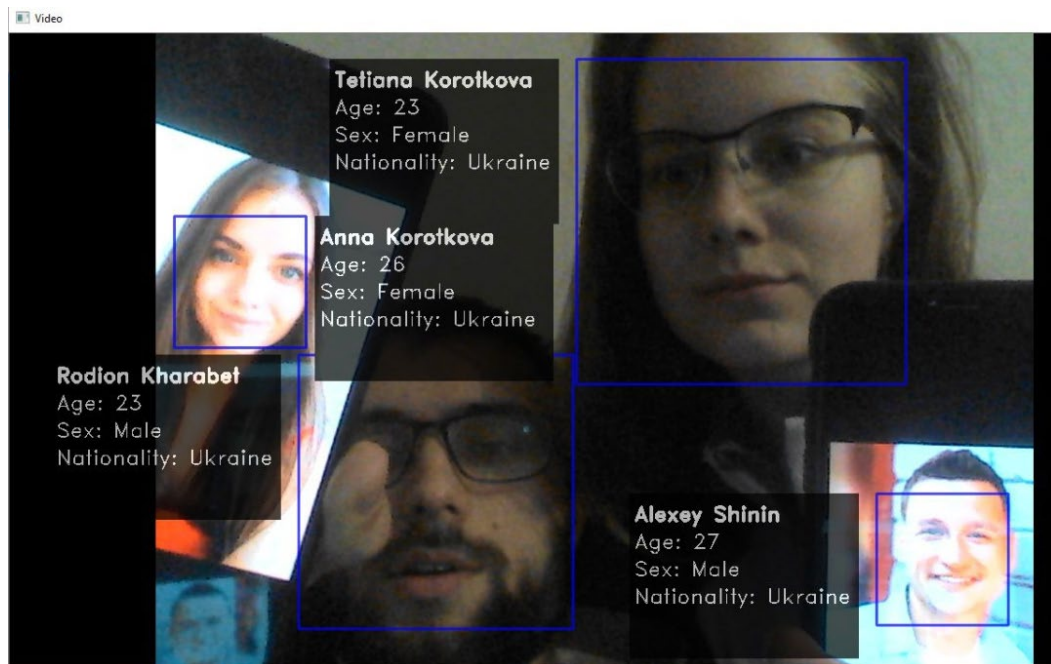


Рисунок 3.21– Розпізнавання кількох облич одночасно - 2

Як бачимо, експерименти показують стабільні результати. Однак за перешкод, а саме коли закривається частина обличчя, яка є визначальною для розпізнавання, програма не може визначити, кого зображено. Це пов'язано з тим, що алгоритм не може порівняти карти ознаки наявного зображення обличчя та обличчя з бази даних.

З 840000 кадрів програма не змогла виявити обличчя взагалі на 59715 кадрах. У більшості випадків це спричинено наявністю перешкод та сильним нахилом або поворотом голови.

Кількість пар фактичне значення – прогнозоване значення наведено у таблиці 3.1 – матриці неточностей (від англ. confusion matrix), де Точність класу (від англ. Class Precision) – це частина значень, які дійсно належать класу відносно всіх значень, які система віднесла до цього класу, Повнота класу (від англ. Class Recall) – це частка знайдених класифікатором значень, що належать класу відносно всіх значень цього класу в тестовій вибірці.

Таблиця 3.2 – Матриця неточностей результатів системи розпізнавання обличчя

Predicted → Real ↓	Клас 1 (Tetiana)	Клас 2 (Rodion)	Клас 3 (Ann)	Клас 4 (Alex)	Точність класу
Клас 1 (Tetiana)	165793	4276	14515	3361	187945 88,21%
Клас 2 (Rodion)	1563	187719	1367	7534	198183 94,72%
Клас 3 (Ann)	11170	398	180584	1421	193573 93,29%
Клас 4 (Alex)	1113	11819	4002	183471	200405 91,55%
Повнота класу	92,29%	91,92%	90,08%	93,71%	

Для оцінки якості детектування за допомогою розробленої програмної реалізації використовується показник, що дорівнює відношенню площі перетину прямокутника, отриманого в результаті детектування, і прямокутника з розмітки до площі їх об'єднання (Intersection over Union, IoU). Таким чином, вважається, що обличчя виявлено правильно, якщо даний показник перевищує певний поріг. В іншому випадку приймається, що обличчя не виявлено. Спрацьовування алгоритму детектування на області, де обличчя відсутнє, вважається помилковим спрацьовуванням.

На підставі наведеного показника здійснюється побудова ROC-кривої (рисунки 3.21 - 3.22), яка відображає залежність кількості помилкових спрацьовувань алгоритму детектування (false positive) від точності детектування (true positive rate).

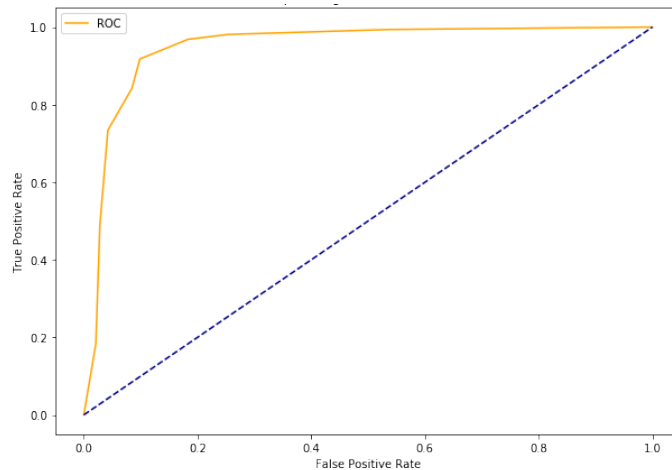


Рисунок 3.21 - ROC-крива при розпізнаванні одного обличчя, $AUC=0,95$

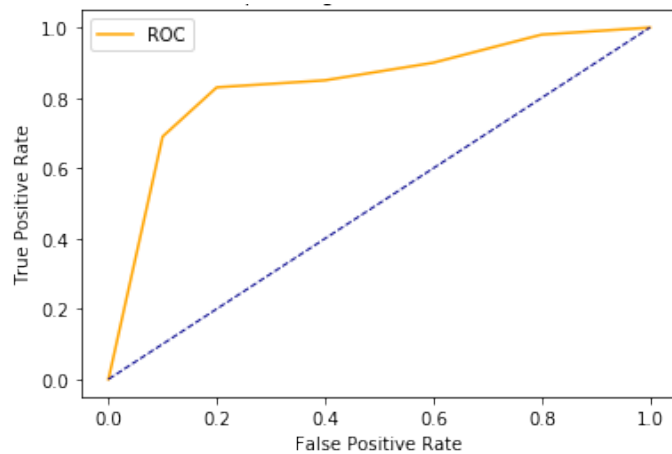


Рисунок 3.22 - ROC-крива при розпізнаванні кількох облич (максимум 4), $AUC=0,83$

3.9 Висновки до третього розділу

У розділі було побудовано цільову функцію для вирішення поставленої задачі, розроблено алгоритм та архітектуру програми. Для розпізнавання обличчя було обрано згорткову нейронну мережу архітектури ResNet34, оскільки вона є глибинною нейронною мережею, що сьогодні є необхідною умовою для роботи із зображеннями, а також долає проблему зникнення градієнта завдяки наявності у своїй архітектурі блоку ідентичності - залишкового шару.

Програма розроблялась на високорівневій мові програмування Python 3.8 у середовщі розробки PyCharm. Для розпізнавання облич людей було використано бібліотеку `face_recognition`, яка надає зручний API та представляє методи роботи з якісно навченими моделями для визначення `face landmarks`, і згорткової нейронної мережі архітектури ResNet34. Збереження зображень облич людей реалізовано через збереження даних у файловій системі, тоді як інформація про людей, місцеположення відповідного зображення їх обличчя, а також карта ознак зображення зберігається у базі даних MySQL.

Систему було протестовано на 840000 кадрах: 4 людей окремо та разом у різних комбінаціях (усього 15 комбінацій) за різних умов освітлення, шумів, перешкод, аксесуарів, емоцій. З 840000 кадрів програма не змогла виявити обличчя взагалі на 59715 кадрах. У більшості випадків це спричинено наявністю перешкод та сильним нахилом або поворотом голови. Середня точність класів – 91,94%, середня повнота класів – 92%.

Висновки

У даній роботі здійснено інформаційний пошук та визначено метод виявлення і розпізнавання обличчя у відеопотоці, інваріантний до масштабування, зміни сцени, поворотів голови, зміни освітленості, аксесуарів, зміни емоцій, а також розроблено додаток, що здійснює детектування та розпізнавання обличчя людини у відеопотоці з доповненою реальністю.

Для розпізнавання облич у відеопотоці використано нейромережу на архітектурі ResNet з 34 шарами, включаючи шар об'єднання, шари ідентичності, шари згортки та шар average pooling, який створює 1000 карт властивостей (для даних ImageNet) і середнє для кожної карти об'єктів.

Для детекції облич у відеопотоці було використано гістограму направлених градієнтів (HOG), для подолання проблеми повороту, масштабування і зсуву голови було застосовано алгоритм оцінки орієнтирів обличчя. Систему була протестована на 840000 кадрах: 4 людей окремо та разом у різних комбінаціях (усього 15 комбінацій) за різних умов освітлення, шумів, перешкод, аксесуарів, емоцій, і пройшла перевірку. Експерименти показують стабільні результати. За перешкод, а саме коли закривається частина обличчя, яка є визначальною для розпізнавання, програма не може визначити, кого зображено, оскільки не може порівняти особливі ознаки наявного зображення обличчя та обличчя з бази даних. Система є платформою для подальших розробок. Зокрема, планується проведення експериментальних досліджень з використанням інших методів розпізнавання обличчя у відеопотоці та виконати порівняльний аналіз отриманих результатів, а також створення більш зручного графічного інтерфейсу програми та адаптація для мобільної версії.

Список використаних джерел

1. P. Sharma, R. N. Yadav, and K. V. Arya, “Pose-invariant face recognition using curvelet neural network,” 2013.
2. Simmel, G. (1971). The metropolis and mental life. In Levine, D.N. (Ed.), On individuality and Social Forms - Selected Writings. Chicago, IL: University of Chicago Press
3. Комп’ютерний зір [Електронний ресурс]. // Режим доступу: https://uk.wikipedia.org/wiki/%D0%9A%D0%BE%D0%BC%D0%BF%27%D1%8E%D1%82%D0%B5%D1%80%D0%BD%D0%B8%D0%B9_%D0%B7%D1%96%D1%80
4. J. Jo, S. J. Lee, K. R. Park, I.-J. Kim, and J. Kim, “Detecting driver drowsiness using feature-level fusion and user-specific classification,” Expert Systems with Applications, vol. 41, no. 4, pp. 1139–1152, 2014.
5. C. Sun, J. H. Li, Y. Song, and L. Jin, “Real-time driver fatigue detection based on eye state recognition,” Applied Mechanics and Materials, vol. 457, pp. 944–952, 2014
6. T. C. Davies, T. Vinumon, L. Taylor, and J. Parsons, “Let’s kinect to increase balance and coordination of older people: Pilot testing of a balloon catching game,” 2014.
7. K. Lee, “Augmented Reality in Education and Training,” Techtrends: Linking Research & Practice To Improve Learning, 2012.
8. Chen, Chin-Sheng, et al. “Three-Dimensional Object Recognition and Registration for Robotic Grasping Systems Using a Modified Viewpoint Feature Histogram.”Sensors, vol. 16, no. 11, 2016, p. 1969., doi:10.3390/s16111969.
9. Neto, Laurindo Britto, et al. “A Kinect-Based Wearable Face Recognition System to Aid Visually Impaired Users.” IEEE Transactions on Human-Machine Systems, 2016, pp. 1–13., doi:10.1109/thms.2016.2604367.]

10. Kusumam, Keerthy, et al. "Can You Pick a Broccoli? 3D-Vision Based Detection and Localisation of Broccoli Heads in the Field." 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, doi:10.1109/iros.2016.7759121.
11. Viola, P., and M. Jones. "Robust Real-Time Face Detection." Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, doi:10.1109/iccv.2001.937709.
12. Zhujie, and Y.I. Yu. "Face Recognition with Eigenfaces." Proceedings of 1994 IEEE International Conference on Industrial Technology - ICIT '94, doi:10.1109/icit.1994.467155
13. Martinez, Aleix. "Fisherfaces." Scholarpedia, vol. 6, no. 2, 2011, p. 4282., doi:10.4249/scholarpedia.4282
14. He, Xiaofei. "Laplacianfaces." Scholarpedia, vol. 5, no. 8, 2010, p. 9324., doi:10.4249/scholarpedia.9324
15. Gong, Y. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval / Y. Gong, S. Lazebnik, A. Gordo, F. Perronnin // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2012. – Vol. 35, Issue 12. – P 2916-2929. – DOI: 10.1109/TPAMI.2012.193.
16. Grauman, K. Learning binary hash codes for large-scale image search / K. Grauman, R. Fergus. – In Book: Machine Learning for Computer Vision / ed. by R. Cipolla, S. Battiato, G.M. Farinella. – Berlin, Heidelberg: Springer, 2013. – P. 49-87. – ISBN: 978-3-642-28660-5. – DOI: 10.1007/978-3-642-28661-2_3.
17. Wang W, Yang J, Xiao J, Li S, Zhou D. Face recognition based on deep learning. In Book: Zu Q, Hu B, Gu N, Seng S, eds. Human Centered Computing. Springer; 2015: 812820. ISBN: 978-3-319-15553-1. DOI: 10.1007/978-3319-15554-8_73.

18. Wu X. Learning robust deep face representation. arXiv preprint arXiv:1507.04844 2015
19. Chen D, Cao X, Wen F, Sun J. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. Proc CVPR 2013: 3025-3032. DOI: 10.1109/CVPR.2013.389.
20. Nguyen H-V, Bai L. Cosine similarity metric learning for face verification. Proc ACCV 2010: 709-720. DOI: 10.1007/978-3-642-19309-5_55.
21. Ю.В. Визильтер, В.С. Горбацевич, А.В. Воротников, Н.А. Костромов “Идентификация лиц в реальном времени с использованием свёрточной нейронной сети и хэширующего леса”
22. Sun, Y. Deep learning face representation by joint identification-verification / Y. Sun, X. Wang, X. Tang // Proceedings of the 27th International Conference on Neural Information Processing Systems. – 2014. – P. 1988-1996.
23. David Manners “NEC installs face recognition at Rio games” [Электронный ресурс]. // Режим доступа: <https://www.electronicweekly.com/news/business/nec-face-recognition-rio-games-2016-08/>
24. Atsushi SATO, Hitoshi IMAOKA, Tetsuaki SUZUKI and Toshinori HOSOI “Advances in Face Detection and Recognition Technologies”
25. Cognitec “FaceVACS-VideoScan” [Электронный ресурс]. // Режим доступа: <https://www.fardata.pl/images/PDF/FaceVACS-VideoScan-PL.pdf>
26. VeriLook SDK - Technical specifications and usage recommendations [Электронный ресурс]. // Режим доступа: <http://www.neurotechnology.com/verilook-technical-specifications.html>
27. VeriLook SDK - Features and capabilities [Электронный ресурс]. // Режим доступа: <http://www.neurotechnology.com/verilook.html>

- 28.6 интересных идей применения дополненной реальности (AR) [Электронный ресурс]. // Режим доступа: <https://habr.com/ru/company/funcorp/blog/440698/>
- 29.Доповнена реальність, або AR-технології. Як це працює? [Электронный ресурс]. // Режим доступа: <http://thefuture.news/page1837780.html>
30. Технология дополненной реальности AR [Электронный ресурс]. // Режим доступа: https://funreality.ru/technology/augmented_reality/
31. H.-K. Jee, S. Lim, J. Youn, and J. Lee, “An augmented reality- based authoring tool for e-learning applications,” *Multimedia Tools and Applications*,
32. Normand, J.-M., Servières, M. & Moreau, G., 2012. A new typology of augmented reality applications. Megève, France, ACM, pp. 445-453.
33. Gee, A. P. et al., 2013. Augmented crime scenes: virtual annotation of physical environments for forensic investigation. Firenze, Italy, ACM, pp. 105-110.
34. Ключові кадри [Электронный ресурс]. // Режим доступа: https://uk.wikipedia.org/wiki/%D0%9A%D0%BB%D1%8E%D1%87%D0%BE%D0%B2%D1%96_%D0%BA%D0%B0%D0%B4%D1%80%D0%B8
35. Caudell, T. P. & Mizell, D. W., 1992. Augmented reality: an application of heads-up display technology to manual manufacturing processes. Kauai, HI, USA, IEEE, pp. 659-669.
36. Augmented reality virtual glasses try-on technology based on iOS platform [Электронный ресурс]. // Режим доступа: <https://jivp-urasipjournals.springeropen.com/articles/10.1186/s13640-018-0373-8>
37. Maatta J, Hadid A, Pietikainen M. Face spoofing detection from single images using texture and local shape analysis[J]. *IET Biometrics*, 2012, 1(1):3–10

38. Zhang V. Cluster Comput. 2017. [Електронний ресурс]. // Режим доступу: <https://doi.org/10.1007/s10586-017-1330-5>
39. І. Голуб'як “Методи Розпізнавання Облич: Короткий Огляд”. . [Електронний ресурс]. // Режим доступу: <http://itcm.comp-sc.if.ua/2017/Holubiak.pdf>
40. Выделение и распознавание лиц [Електронний ресурс]. // Режим доступу: http://wiki.technicalvision.ru/index.php/%D0%92%D1%8B%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B8_%D1%80%D0%B0%D1%81%D0%BF%D0%BE%D0%B7%D0%BD%D0%B0%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%BB%D0%B8%D1%86
41. Анализ существующих подходов к распознаванию лиц [Електронний ресурс]. // Режим доступу: <https://habr.com/ru/company/synesis/blog/238129/>
42. Овчарук В. В. “Порівняльний аналіз алгоритмів розпізнавання емоцій людини за зображенням обличчя” [Електронний ресурс]. // Режим доступу: [Посилання](#)
43. Біометрична ідентифікація і автентифікація особи за геометрією обличчя [Електронний ресурс]. // Режим доступу: https://www.researchgate.net/publication/321423112_Biometricna_identifikacia_i_avtentifikacia_osobi_za_geometriu_oblicca
44. Template matching [Електронний ресурс]. // Режим доступу: https://en.wikipedia.org/wiki/Template_matching
45. Convolutional neural network [Електронний ресурс]. // Режим доступу: https://en.wikipedia.org/wiki/Convolutional_neural_network
46. Іван Гудфелов, Йоша Бенгіо, Арон Коурвілле. «Машинне навчання» MIT Press, 2016. Режим доступу: <http://www.deeplearningbook.org>.

47. [Электронный ресурс]. // Режим доступа:
https://rk6.bmstu.ru/pub/diplom_labors/2017/2017_Chirkova_A_rpz.pdf
48. Deep Residual Learning for Image Recognition [Электронный ресурс]. // Режим доступа: <https://arxiv.org/abs/1512.03385>
49. ICML 2016 Tutorial on Deep Residual Networks [Электронный ресурс]. // Режим доступа: <http://kaiminghe.com/icml16tutorial/index.html>
50. Study of Residual Networks for Image Recognition [Электронный ресурс].
// Режим доступа: http://cs231n.stanford.edu/reports2016/264_Report.pdf
51. How does deep residual learning work? [Электронный ресурс]. // Режим доступа: <https://www.quora.com/How-does-deep-residual-learning-work>
52. Python [Электронный ресурс]. // Режим доступа:
<https://ru.wikipedia.org/wiki/Python>
53. Е.А. Никифорова “Разработка системы распознавания лиц для пропускной системы института”
54. Где хранить изображения [Электронный ресурс]. // Режим доступа:
<https://habr.com/ru/post/46466/>
55. Augustinus Kristiadi “Residual Net” [Электронный ресурс]. //Режим доступа: <https://wiseodd.github.io/techblog/2016/10/13/residual-net/>

Додаток А
Лістинг програми