

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____

Кафедра _____ Програмної інженерії _____

АТЕСТАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти – другий (магістерський)

Дослідження методів для розробки програмної системи визначення вектору
направленості обличчя на зображенні

Виконав: студент 2 курсу, групи ІПЗм-18-2 _____

_____ Ракова А.О. _____

(прізвище, ініціали)

спеціальності 121- Інженерія програмного забезпечення
(код і повна назва спеціальності)

Освітньо-наукової програми
(тип програми)

Інженерія програмного забезпечення
(повна назва освітньої програми)

Керівник _____ проф. Білоус Н.В. _____

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри, проф. _____

З.В.Дудар

2020 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Програмної інженерії

Рівень вищої освіти - другий (магістерський)

Спеціальність 121-Інженерія програмного забезпечення

(код і повна назва)

Тип програми освітньо-наукова програма

Освітня програма Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

« ____ » _____ 20 ____ р.

ЗАВДАННЯ НА АТЕСТАЦІЙНУ РОБОТУ

студентові Раковій Аліні Олегівні

(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів для розробки програмної системи визначення вектору направленості обличчя на зображенні
затверджена наказом університету від "27" 03 2020 р № 473Ст
2. Термін подання студентом роботи до екзаменаційної комісії _____
3. Вихідні дані до роботи модуль відстеження положення голови людини, пояснювальна записка.
4. Перелік питань, що потрібно опрацювати в роботі мета роботи, аналіз проблемної галузі і постановка задачі, аналіз методів визначення вектору направленості обличчя на двомірному зображенні для розробки системи з відстеження рухів голови людини, аналіз методів порівняння рухів голови для визначення точності виконання вправ

Наступний аркуш завдання

5 Консультанти розділів роботи

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Спецчастина	проф. Білоус Н.В.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка *
1.	Аналіз предметної галузі	27.01.20 – 19.02.20	виконано
2.	Огляд методів визначення положення голови	20.02.20 – 10.03.20	виконано
3.	Аналіз обраних методів з точки зору використання у системі моніторингу рухів людини	11.03.20 – 25.03.20	виконано
4.	Програмна реалізація	26.03.20 – 15.04.20	виконано
5.	Підготовка пояснювальної записки	15.04.20 – 02.05.20	виконано
6.	Підготовка презентації та доповіді	03.05.20 – 06.05.20	виконано
7.	Попередній захист	07.05.20	виконано
8.	Нормоконтроль, рецензування	10.05.20 – 15.05.20	виконано
9.	Занесення диплома в електронний архів	16.05.20 – 20.05.20	виконано
10.	Допуск до захисту у зав. кафедри	20.05.20	виконано

Дата видачі завдання 27 лютого 2020 р.Студент _____
(підпис)Керівник роботи _____
(підпис)проф. Білоус Н.В.
(посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Атестаційна робота магістра містить: 39 с., 14 рис., 1 табл., 17 джерел.

НАПРАВЛЕНІСТЬ ОБЛИЧЧЯ, РОЗПІЗНАВАННЯ, COMPUTER VISION, DEEP LEARNING, PYTHON.

Об'єктом дослідження є вектор направленості обличчя. Предметом дослідження є використання методів визначення вектору направленості обличчя для моніторингу рухів. Метою роботи є розробка додатку відстеження рухів голови людини на потоковому відео.

У роботі досліджуються та порівнюються існуючі методи визначення вектору обличчя з точки зору їх експлуатації у системах відстеження рухів. Розглядалися такі характеристики методів, як точність, нечутливість до умов зйомки та швидкість роботи. У результаті дослідження був виявлений найбільш точний метод, з використанням якого був створений додаток для відстеження та перевірки рухів голови людини. У подальшому планується впровадження цього додатку у комплексні системи моніторингу руху людини.

COMPUTER VISION, DEEP LEARNING, FACE POSE, PYTHON, RECOGNITION.

The object of the study is the face direction vector. The subject of the study is use of methods for determining the face direction vector in movements monitoring. The purpose of the study is implementation of an application for tracking human head movements.

The study investigates and compares existing methods of determining the face vector in terms of their operation in motion tracking systems. Characteristics of methods such as accuracy, insensitivity to photo conditions and speed of work were considered. The study revealed the most accurate method, which was used to create an application for tracking and checking the movements of the human head. In the future, it is planned to implement this application in integrated human movement monitoring systems.

ЗМІСТ

Вступ.....	6
1 Аналіз предметної галузі.....	8
1.1 Існуючі підходи до визначення вектору направленості обличчя	8
1.2 Існуючі програмні продукти	15
1.3 Постановка задачі	16
2 Опис теоретичних досліджень	18
2.1 Вибір методів визначення вектору обличчя	18
2.2 Використані датасети	22
3 Опис практичних досліджень	24
4 Програмна реалізація	28
4.1 Архітектура додатку	28
4.2 Розробка програмної реалізації	31
5 Впровадження результатів дослідження у практичну діяльність	34
5.1. Опис варіантів використання отриманих результатів	34
5.2 Існуючі проблеми та напрямки подальшого дослідження	35
Висновки	37
Перелік джерел посилання	38
Додаток А Слайди презентації	40
Додаток Б Лістинг коду	48
Додаток В Апробація результатів атестаційної роботи	49

ВСТУП

Напрямок вектору людського обличчя, який є індикатором уваги людини, має безліч застосувань в нашому повсякденному житті, таких як взаємодія людини з комп'ютером, телеконференції, віртуальна реальність і 3D - рендеринг звуку. Також визначення положення голови може бути використано для порівняння виконуваних людиною вправ з певним еталоном. Це дозволить нам визначати точність виконання вправ, що в свою чергу корисно для реабілітаційних установ, фітнес центрів, розважальних приставок.

Протягом останніх років методи розпізнавання обличчя, його окремих частин та погляду розвивалися у деяких різних напрямках. З метою розпізнавання людських рухів у іграх, була розроблена оптична система Kinect, яка завоювала популярність завдяки високій точності розпізнавання. Але ця система має значні обмеження на використання, такі як максимальна кількість людей на зображенні та неможливість використання під прямими сонячними променями. Паралельно розвивався інший напрямок розпізнавання з двомірного зображення без використання датчиків глибини. Цей напрямок теж поділяється на різні методи, але останніми роками саме методи засновані на використанні CNN показали найкращі результати у точності та швидкості аналізу зображення.

Актуальність роботи полягає у тому, що запропоновані методи здатні подолати обмеження та незручності, які зв'язані з процесом визначення положення голови. Розробка та дослідження цих методів дозволить спростити інтеграцію модулю в комплексні системи відстеження руху, не вимагаючи додаткових засобів стеження, аніж двомірне зображення людини.

Метою даної роботи є розробка додатку, що дозволить відстежувати та аналізувати рухи голови людини на потоковому відео. Об'єктом дослідження є вектор направленості обличчя людини. Предметом дослідження є використання методів визначення вектору направленості обличчя у системах моніторингу рухів людини.

У рамках дослідження були розглянуті роботи пов'язані з пошуком напрямку обличчя, опубліковані за останні п'ять років. Були обрані методи для проведення експериментальної перевірки характеристик, важливих для відстеження рухів у реальному часі. Найбільш точний метод серед досліджених було обрано реалізації програми здатної відстежувати та перевіряти рухи голови людини. Таким чином була доведена можливість та доцільність використання методів пошуку вектору повороту обличчя у програмах моніторингу рухів людини.

Стаття, що описує це дослідження була подана до публікації у журнал «Радіоелектроніка, інформатика, управління», що видається Національним університетом «Запорізька політехніка».

У подальшому цей додаток може стати частиною системи з моніторингу рухів людського тіла, який може знайти використання у реабілітаційних та фітнес центрах, або у розважальній сфері. У порівнянні з апаратними системами відстеження руху людини, програмна реалізація є більш гнучкою для адаптації до різних умов використання та створює менше незручностей людині за якою ведеться спостереження.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Існуючі підходи до визначення вектору направленості обличчя

Різні підходи до визначення вектору повороту обличчя використовують різні показники. Виділяють два основні способи вираження положення голови: положення камери відносно голови, відхилення голови від положення анфас.

У комп'ютерному зорі поза предмета означає його відносну орієнтацію та положення щодо камери. Ви можете змінити позу, перемістивши об'єкт відносно камери, або камеру відносно об'єкта. Якщо ми бажаємо виразити положення голови, таким чином, то мета полягає в тому, щоб знайти позу об'єкту, коли у нас є калібрована камера, і ми знаємо розташування n 3D точок на об'єкті та відповідні 2D проекції на зображенні [1].

3D-жорсткий об'єкт має лише два види руху щодо камери:

- translation: Переміщення камери з її поточного 3D-місця (X, Y, Z) в нове 3D-місце (X', Y', Z') називається трансляцією. Як бачите, трансляція має 3 ступеня свободи - ви можете рухатись у напрямку X , Y або Z . Трансляція представлена вектором, який дорівнює $(X' - X, Y' - Y, Z' - Z)$;

- rotation: Ви також можете обертати камеру навколо осей X , Y і Z . Тому обертання також має три ступеня свободи. Існує багато способів подання обертання. Ви можете представити його за допомогою кутів Ейлера (поворот, крок і похил), матриці повороту 3×3 або напрямку обертання (тобто осі) та кута.

Для обчислення 3D-позиції об'єкта на зображенні вам потрібна наступна інформація:

- 2D координати кількох точок. Вам потрібно розташування кількох точок на зображенні. У випадку з обличчям ви можете вибрати куточки очей, кінчик носа, куточки рота тощо;

- 3D-розташування тих самих точок. Вам потрібні тривимірні місця розташування кількох точок у довільному кадрі (див. рис. 1.1). Оскільки отримати

точну тривимірну модель будь-якої голови з одного зображення не має можливості, використовується узагальнена модель голови людини.

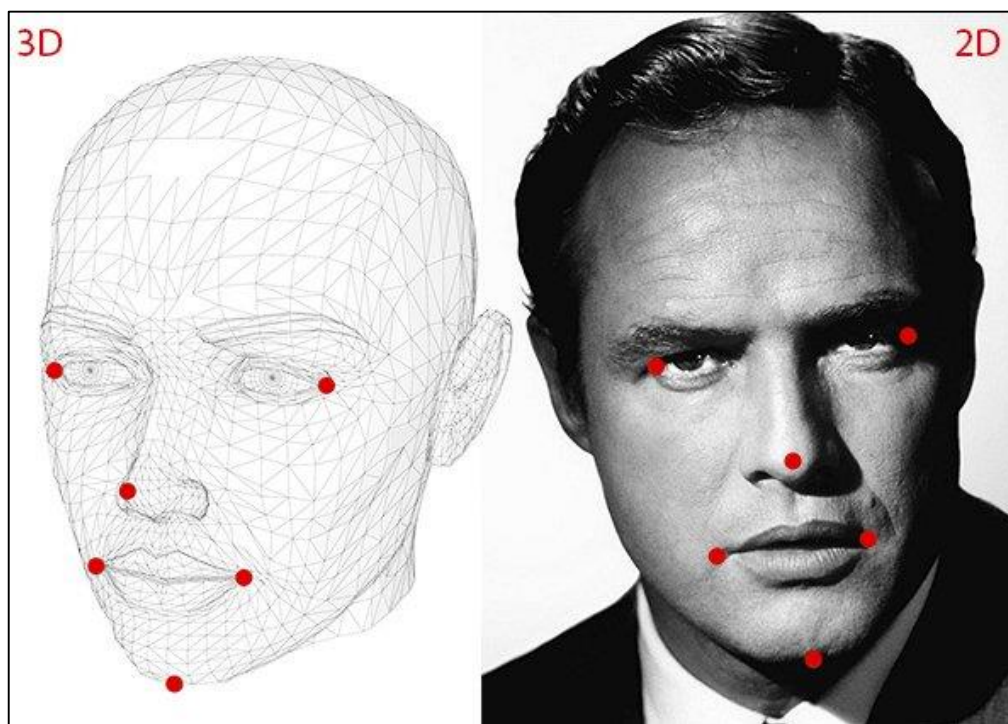


Рисунок 1.1 – Відповідність точок на двовимірному та тривимірному зображенні.

Показниками, за якими визначається положення голови при використанні другого способу вираження положення є кут на який голова відхиляється від положення анфас за трьома наступними напрямками: pitch (нахил вниз / вгору), roll (нахил до лівого / правого плеча), yaw (поворот) (див. рис. 1.2). Тому положення голови за цими показниками можна записати у вигляді вектору (r, p, y) .

У той час як перший спосіб підходить для визначення положення будь-якого предмету відносно камери, другий можна використовувати лише для визначення положення голови. Зазвичай саме другий спосіб використовується у методах які засновані на створенні CNN моделі.

Одна з відомих технології, які відіграють вирішальну роль у визначенні пози людини і пози голови - камера Kinect розроблена Microsoft. Камера Kinect має явну

перевагу перед іншими 3D камерами, тому що отримує більш точну інформацію про глибину об'єкта і працює досить швидко.

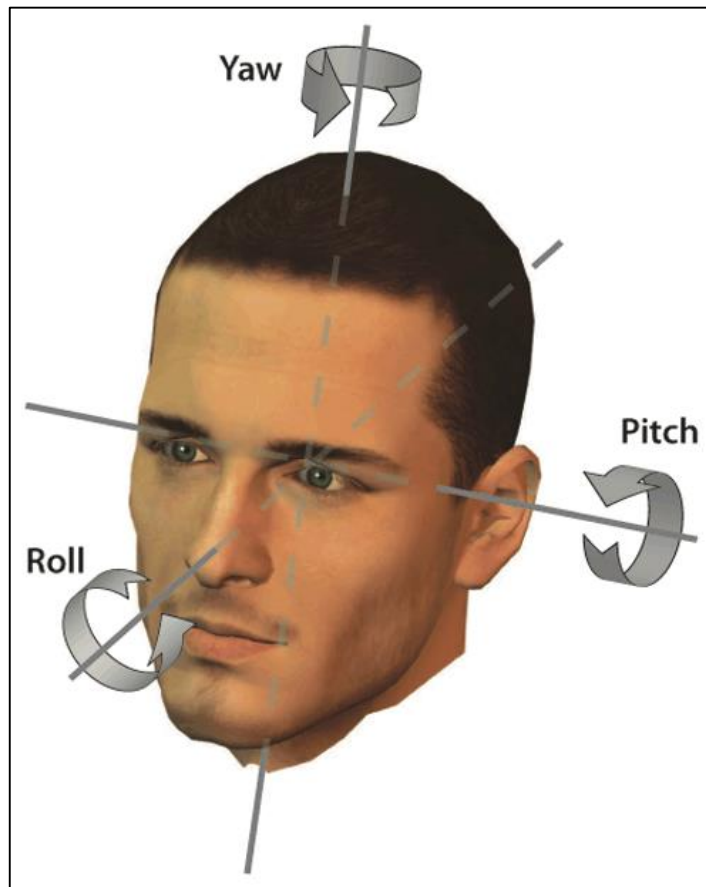


Рисунок 1.2 – Напрямки нахилу голови.

Використовуючи Kinect, можна відстежувати до шести чоловік одночасно а також отримати аналіз руху з функцією екстракції, що дозволяє визначати місцезнаходження точок зчленування людського скелета. Надзвичайно корисна в приміщенні, вона не може бути використана на вулиці, оскільки інфрачервоний датчик глибини надзвичайно чутливий до сонячного світла. Глибинний датчик в цій цифровій камері є інфрачервоним лазером, який об'єднаний з монохромним CMOS датчиком. Датчик глибини може легко записувати і захоплювати відео дані в 3D при будь-якому освітленні. Система (RGB камера, датчик глибини і мульти-мікрофон) може розпізнавати і інтерпретувати конкретні жести. Методи засновані на використанні інформації отриманої за допомогою глибинних датчиків

дозволяють добитися точності при якій відхилення від реальних даних становить не більше двох градусів [2].

Kinect є відмінним варіантом для розпізнавання напряду обличчя людини, однак на використання цієї технології накладаються наступні обмеження:

- камера може одночасно відстежувати не більше шести людей;
- камеру неможливо використовувати на відкритому повітрі. Kinect працює шляхом лазерного проектування структурованого точкового малюнка в ІК-діапазоні, захоплення його ІК-камерою і обробки спотворення точкового малюнка для визначення відстані між точками. Інфрачервоне структуроване освітлення повністю втрачається в навколишньому ІК в сонячну погоду.

У зв'язку з наявністю обмежень на використання Kinect ми розглянемо способи розпізнавання положення голови людини на основі двомірного зображення RGB камери. Історично склалося кілька основних підходів до моделювання обличчя: дискримінаційні, орієнтовані підходи і параметризовані моделі зовнішнього вигляду, або РАМ. Не зважаючи на обраний підхід, кожен з методів використовує детектор обличчя для отримання регіону зображення з обличчям, який у подальшому будуть аналізуватися.

У зв'язку з наявністю обмежень на використання Kinect ми розглянемо способи розпізнавання положення голови людини на основі двомірного зображення RGB камери. Історично склалося кілька основних підходів до моделювання обличчя: дискримінаційні, орієнтовані підходи і параметризовані моделі зовнішнього вигляду, або РАМ. Не зважаючи на обраний підхід, кожен з методів використовує детектор обличчя для отримання регіону зображення з обличчям, який у подальшому будуть аналізуватися.

Розпізнавання обличчя - також його називають виявленням обличчя - це комп'ютерна технологія на основі штучного інтелекту (AI), яка використовується для пошуку та ідентифікації обличч людини на цифрових зображеннях. Технологія виявлення обличч може застосовуватися в різних сферах - включаючи безпеку, біометрику, правоохоронні органи, розваги та особисту безпеку - для забезпечення спостереження та відстеження людей у режимі реального часу.

Існують декілька підходів до задачі розпізнавання обличч, деякі з них базуються реалізації алгоритму Віоли-Джонса[3], на вейвлет-трансформації [4] або методі головних компонент. Розпізнавання обличчя прогресувало від рудиментарних методик комп'ютерного зору до машинного навчання (ML), все більш досконалих штучних нейронних мереж (ANN) та пов'язаних з ними технологій. Результатом стали постійні поліпшення продуктивності. Зараз воно відіграє важливу роль як перший крок у багатьох ключових програмах, включаючи

відстеження обличчя, аналіз та розпізнавання. Розпізнавання обличчя суттєво впливає на ефективність послідовних операцій у програмі.

Під час аналізу розпізнавання обличчя допомагає визначити, на яких частинах зображення чи відео слід зосередити увагу, щоб визначити вік, стать та емоції за допомогою міміки. У системі розпізнавання, яка математично відображає риси обличчя людини і зберігає дані як відбиток, потрібні дані виявлення для алгоритмів, які визначають, які частини зображення або відео потрібні для створення відбитка обличчя. Після ідентифікації новий відбиток обличчя можна порівняти зі збереженими відбитками, щоб визначити, чи є відповідність.

Методи, що застосовуються при виявленні обличчя, можуть бути на основі знань, на основі особливостей, відповідності шаблонів або на зовнішньому вигляді. У кожного є свої переваги та недоліки:

- методи на основі знань описують обличчя на основі правил. Проблемою такого підходу є складність розробки чітко визначених правил;
- інваріантні методи особливостей, які використовують такі ознаки, як очі або ніс людини для виявлення обличчя - можуть відчувати негативний вплив вод шуму і світла;
- методи відповідності шаблонів ґрунтуються на порівнянні зображень із стандартними візерунками обличчя або особливостями, які були описані раніше, та співвідношенні їх для виявлення особи. На жаль, ці методи не враховують змін у позі, масштабі та формі;
- методи, засновані на зовнішньому вигляді, використовують статистичний аналіз та машинне навчання для пошуку відповідних характеристик зображень обличчя. Цей метод, який також використовується для вилучення особливостей для розпізнавання обличчя, поділяється на підметоди.

В останні роки з'явилися методи, які безпосередньо беруть 2D (face landmark) лицьові точки, використовуючи deep learning інструменти навчання [5]. Вони стали домінуючими підходами до аналізу повороту обличчя завдяки їх гнучкості і стійкості до екстремальних змін пози.

Пошук пози відбувається завдяки визначенню спотворення застосовуваного до 3D моделі. Щоб домогтися цього, детектор орієнтирів повинен знайти десятки точок на обличчі, такі як куточки рота, куточки очей, силует щелеп і багато іншого. Багато алгоритмів були розроблені і реалізовані в OpenCV. Виявлення орієнтирів особи починається з виявлення осіб на зображенні і їх екстентів (обмежувальних рамок). Виявлення особи довгий час вважалося вирішеною проблемою, і OpenCV містить один з перших надійних детекторів особи, вільно доступних для громадськості. Фактично, OpenCV в перші роки своєї діяльності був відомий перш за все і використовувався для швидкого розпізнавання осіб, реалізації канонічного алгоритму покращеного каскадного класифікатора Віоли-Джонса і надання попередньо навченої моделі. Хоча виявлення осіб значно розвинулося з тих ранніх днів, найшвидшим і простим способом виявлення осіб в OpenCV як і раніше є використання пов'язаних каскадних класифікаторів за допомогою класу `cv::CascadeClassifier`, передбаченого в модулі `core`. Детектор міток буде працювати навколо виявлених осіб, починаючи з обмежувальних рамок.

Отримавши орієнтири на обличчі, ми можемо спробувати визначити напрямок обличчя. 2D орієнтири обличчя по суті відповідають формі голови. Отже, з огляду на тривимірну модель людської голови, ми можемо знайти приблизні відповідні тривимірні точки для ряду лицьових орієнтирів (див. рис. 1.1).

З цих 2D-3D-відповідностей ми можемо розрахувати тривимірну позу (обертання і переміщення) голови щодо камери за допомогою алгоритму Point-n-Perspective (PnP) [6]. Камера, яка зробила зображення, має фіксовану трансформацію, тобто вона змістилася на деяку відстань від об'єкта, а також кілька повернута відносно нього. У більш широкому сенсі ми можемо записати відносини між точками на зображенні (поруч з камерою) і об'єктом наступним чином:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = s \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 & r_2 & r_3 & t_1 \\ r_4 & r_5 & r_6 & t_2 \\ r_7 & r_8 & r_9 & t_3 \end{pmatrix} \begin{pmatrix} U \\ V \\ W \\ 1 \end{pmatrix} \quad (1.1)$$

У наведеній вище формулі U , V , W позиція об'єкта в тривимірному просторі, x , y координати точок на зображенні, f_x та f_y – це фокусні відстані у напрямку x та y , c_x та c_y – це фокальний центр, r_i – елементи вектору повороту, t_i – елементи вектору переміщення камери. OpenCV надає реалізацію для пошуку повороту і переміщення за допомогою своїх функцій `cv::solvePnP` модуля `calib3d`.

Оскільки модель зазвичай використовується усереднена, то це може призводити до помилок розпізнавання положення в разі значних відмінності голови випробуваного від моделі. Ще однією проблемою в застосуванні цього методу є залежність від точності визначення опорних точок особи.

Більш новими і просунутими методами визначення пози голови є методи, засновані на тренуванні нейронних мереж. Детектори особи навчаються за зображеннями особи з різними дискретними позами і, об'єднують виходи ряду класифікаторів. Ці існуючі методи оцінки пози голови зазвичай навчаються на основі класифікації бін-поз або регресії в одній позі. Методи, засновані на класифікації, виконуються шляхом зіставлення зображення з мітками дискретних поз, в той час як методи, засновані на регресії, безпосередньо виводять значення безперервної пози. Однак ці методи навряд чи використовують класифікацію і втрату регресії одночасно. Методи, засновані на класифікації, є більш надійними, щодо змін в неідеальних умовах. Їх мітки істинності, приймають кутові інтервали (зазвичай більше 10°), тому відповідні мітки є вузьким місцем для подальшого підвищення точності. Методи, засновані на регресії, можуть більш точно передбачити позу голови, але їх ефективність залежить від початкової позиції голови і від варіацій пози голови, і погано працюють в неідеальних умовах.

Останнім часом дослідні роботи з використанням глибоких нейронних мереж стали лідирувати в точності оцінки пози голови. Дж. Парк і С. Квон використовували глибокі нейронні мережі, такі як `Lenet`, щоб оцінити безперервну позу голови [7]. Массіміліано і Анджело досліджували роль адаптивних градієнтних методів для поліпшення характеристик CNN в оцінці пози голови [8]. Вищезазначені роботи будують свій основний блок, використовуючи менше п'яти шарів згортки, і витягають більш виразні функції з навчальних наборів даних.

Кумар і співавтори модифікували архітектуру GoogleNet для спільного прогнозування орієнтирів обличчя і пози голови [9]. Сюй і співавтори адаптували глобальні та локальні особливості особи, отримані з допомогою CNN для coarse-to-fine оцінки пози голови[10]. Вони використовували глобальні мережі для прогнозування вихідної пози голови і локальні мережі для поновлення пози відповідно до поточної формою. Б. Хуанг і співавтори створили метод оцінки пози голови з використанням двоступеневих груп з усередненою top-k регресією [11].

Попередні роботи показали, що оцінка пози голови вразлива до змін у фоновій сцені навколо цільового обличчя, хоча це може бути не так дивно; методи вешті-решт засновані на зовнішності. Однак було виявлено, що правильне регулювання меж, які обрамляють обличчя на зображенні, зменшує цю вразливість. Крім того, були досліджені функції втрат, які використовуються у CNN для покращення оцінки повороту обличчя, що сприяє подальшому вдосконаленню.

Причина, через яку ці важливі фактори не помічаються в ранніх дослідженнях, може полягати в тому, що завдання з оцінки голови розглядалося як допоміжне завдання (або побічний продукт) інших основних завдань, таких як вирівнювання обличчя, розпізнавання обличчя та виявлення.

1.2 Існуючі програмні продукти

Серед програмних продуктів які надають можливість відстежувати та аналізувати обличчя людини розглядалися:

- Face++;
- Visage Technologies.

Face++ аналізує ряд атрибутів, пов'язаних з обличчям, включаючи вік, стать, інтенсивність посмішки, позу голови, стан очей, емоції, красу, погляд очей, стан рота, стан шкіри, етнічну приналежність, якість зображення обличчя та розмитість. Приклад демо можна побачити на рисунку 1.3.

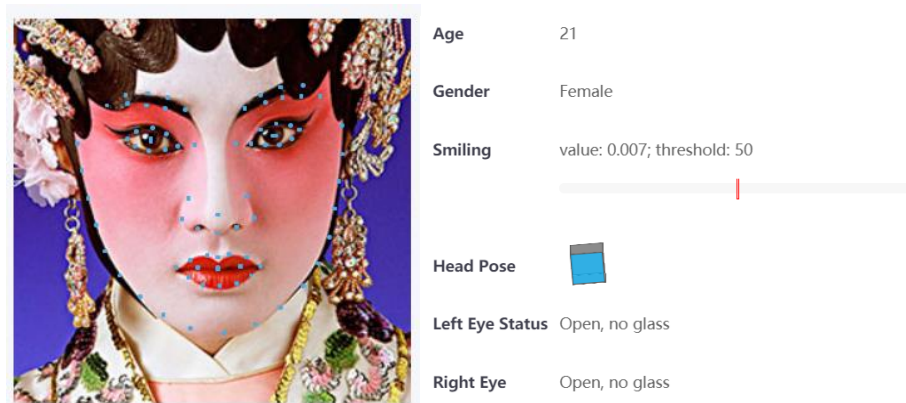


Рисунок 1.3 – Демо Face++

Visage Technologies виявляє обличчя на зображеннях чи відео, а потім використовує підпрограми для відстеження обличчя та дій, для розпізнавання статі, емоцій та віку для обличчя у фронтальному положенні. Ця система теж надає можливість відстежувати позу обличчя.

Основна причина з якої наявне SDK цих систем погано підходить для вирішення поставленої задачі – це різносторонність функціоналу цих програм. Оскільки основною функцією є пошук ключових точок обличчя, усі інші функції залежать від точності визначення цих точок. У попередньому підрозділі були описані недоліки систем, які засновані на попередньому виявленні ключових точок. Це погіршення точності у разі відхилення рис обличчя від еталонної моделі та значна затрата ресурсу для визначення атрибутів обличчя, які є лише ланкою в процесі отримання даних про положення голови.

Наявні недоліки змушують розглядати інші методи, які заточені під пошук вектору направленості обличчя та не витрачають ресурсу на побічні обчислення.

1.3 Постановка задачі

Базуючись на дослідженні існуючих систем та методів, які виконують визначення вектору обличчя на зображеннях, можна описати певні вимоги до

побудови системи яка дозволить розпізнавати та порівнювати позу голови людини на зображенні та потоковому відео.

Основним критерієм за яким слід розглядати існуючі методи є точність розрахунку вектору та стабільність роботи у різних умовах освітлення та якості відео. Також відстань на якій знаходиться людина не має впливати на якість визначення вектору так само як і стать, вік, наявність окуляр або незвичної зачіски. Слід також розглянути результати роботи різних методів у разі часткової закритості обличчя іншою частиною тіла або волоссям.

Важливим показником за яким слід досліджувати методи є максимальний кут, який вони здатні розпізнавати. Значення максимального куту розпізнавання може відрізнятись для різних осей повороту голови.

Система, яка проводить відстеження багатьох людей одночасно у режимі реального часу має певні вимоги до швидкості обробки зображення. Так світовий стандарт кадрової частоти є 24 кадри на секунду, тобто швидкість обробки одного фрейму має складати приблизно 40 мс. Це дозволить максимально точно порівнювати швидкі рухи, які частіше всього доводиться спостерігати в спорті.

Виходячи з вищенаведених умов були сформовані наступні задачі дослідження:

- дослідити різні підходи та методи пошуку вектору напрямку обличчя, у першу чергу беручи до уваги методи запропоновані за останні 5 років;
- експериментально перевірити можливості цих методів на відповідність вищенаведеним критеріям, та виявити найбільш підходящий метод для використання в умовах відстеження рухів голови;
- створити додаток для розпізнавання та перевірки рухів голови людини з використанням обраного серед досліджених варіантів методу.

2 ОПИС ТЕОРЕТИЧНИХ ДОСЛІДЖЕНЬ

2.1 Вибір методів визначення вектору обличчя

Як було зазначено у попередньому розділі, існують 2 основні напрямки методів знаходження вектору повороту обличчя. Пошук напрямку голови через ключові точки базується загалом на пошуку опорних точок та знаходженні відповідності між двовимірним зображенням та тривимірною моделлю цей метод ми назвемо метод моделі [12]. У порівнянні з методами, що базуються на моделях, методи, що базуються на зовнішньому вигляді, в повній мірі використовують інформацію зображення для оцінки позиції голови. Методи, що базуються на зовнішньому вигляді, можна розділити на класифікаційні методи та нелінійні методи регресії. Оцінка пози голови розглядається як багатокласова класифікаційна проблема в методах класифікації. Хоча методи класифікації легко застосувати та мають високу продуктивність у режимі реального часу, вони можуть лише передбачити приблизний діапазон позиції голови і страждають від грануляції оцінених кутів, тому важко навчати класи, кути яких дуже схожі. Тож їх сфера застосування обмежена.

Методи нелінійної регресії використовують навчальний набір для створення нелінійного відображення від зображень до поз, а CNN є частиною цих методів. Оскільки CNN мають можливість зменшити розміри та автоматично витягувати характеристики, вони досягали хороших результатів у різних сферах. Використання CNN значно покращило точність оцінки голови, однак відмінні показники демонструються лише в тому ж типі зображень та умов, що є у навчальному наборі через сильне перенапруження набору тренувань.

FSA-Net [13] є представником методу регресії та використовує плавну ступінчасту регресійну схему. Існуючі методи агрегування функцій обробляють вхідні дані як набір особливостей i , таким чином, ігнорують їх просторове відношення на мапі особливостей. Метод пропонує дрібнозернисте (fine-grained) відображення структури для просторового групування ознак перед агрегацією.

FSA-Net побудоване на SSR-Net, яке забезпечує компактну модель для оцінки віку з одного зображення. Натхненний DEX, SSR-Net ставить проблему регресії оцінки віку як проблему класифікації шляхом поділу на вікову область на кілька вікових класів (бін). Мережа виконує завдання з класифікації та виводить розподіл ймовірностей для вікових класів. Враховуючи розподіл ймовірностей, вік оцінюють як очікуване значення. Для того, щоб мати компактну модель, SSR-Net приймає стратегію coarse-to-fine точності для класифікації. Кожен етап проводить лише проміжну класифікацію з невеликою кількістю класів, наприклад, "відносно молодший", "відносно старший" в межах поточної вікової групи. Наступний етап уточнює рішення у віковій групі, призначеній попереднім етапом. Підсумовуючи, SSR-Net здійснює ієрархічну класифікацію і використовує наступну стадію регресії для оцінки віку \tilde{y} :

$$\tilde{y} = \sum_{k=1}^K \vec{p}^{(k)} \cdot \vec{\mu}^{(k)} \quad (2.1)$$

У наведеному рівнянні K – це кількість етапів, $\vec{p}^{(k)}$ – розподіл вірогідностей для k -того етапу, $\vec{\mu}^{(k)}$ - вектор, який складається з показникових значень вікових груп на k -тому етапі. Поступова ступінчаста регресія може застосовуватися до будь-якої проблеми регресії. У цій роботі застосована ступінчаста регресія до проблеми оцінки пози з одного зображення. На відмінну від проблеми оцінювання віку, проблема оцінки пози оцінює вектор, а не скаляр.

На рисунку 2.1 зображена архітектура FSA-Net. Вхідне зображення проходить через два потоки. Існує K етапів ($K = 3$ на рисунку 2.1). Кожен потік отримує мапу особливостей. Для k -ї стадії отримані карти особливостей зливаються разом. Модуль поєднання спочатку поєднує в собі дві мапи особливостей. Потім він застосовує $n \times 1$ згортки для перетворення об'єднаної карти об'єктів у n каналів. Нарешті, для зменшення розміру використовується усереднююче об'єднання мап особливостей до 3×3 год.

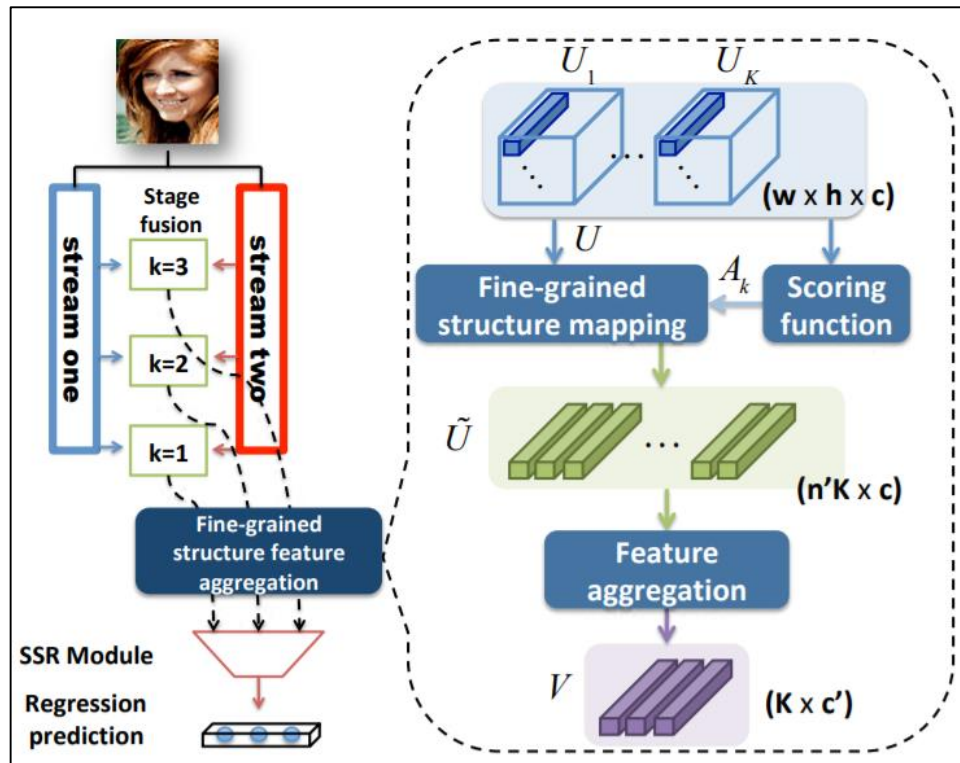


Рисунок 2.1 – Архітектура FSA-Net.

FSA-Net виконує просторове групування ознак перед подачею їх у процес агрегації. Розробники цього методу стверджують, що завдяки його використанню вдається добитися точності, що перевищує можливості систем заснованих на обробці RGBD зображень, тобто зображень з камер з глибинними датчиками.

ResNet50 [14] поєднує класифікаційний та регресійний підхід для створення CNN яка визначає положення голови без попереднього знаходження опорних точок. ResNet50 використовує три окремі втрати (loss), по одній для кожного кута. Кожна втрата - це поєднання двох компонентів: класифікована поза та регресійний компонент. Будь-який мережу можна використовувати та доповнювати трьома повністю з'єднаними шарами, які прогнозують кути. Ці три повністю пов'язані шари поділяють попередні згорткові шари мережі.

Ідея цього підходу полягає в тому, що, здійснюючи класифікацію, використовується шар softmax і крос-ентропія, таким чином мережа вчиться чітко прогнозувати сусідню позу. Маючи три перехресні ентропійні втрати, по одному для кожного кута Ейлера, у нас є три сигнали, які передаються в мережу що покращує навчання. Для того, щоб отримати більш чіткі прогнози, обчислюється

очікування кожного кута виходу для отриманого класу виведення. Архітектура мережі представлена на рисунку 2.2.

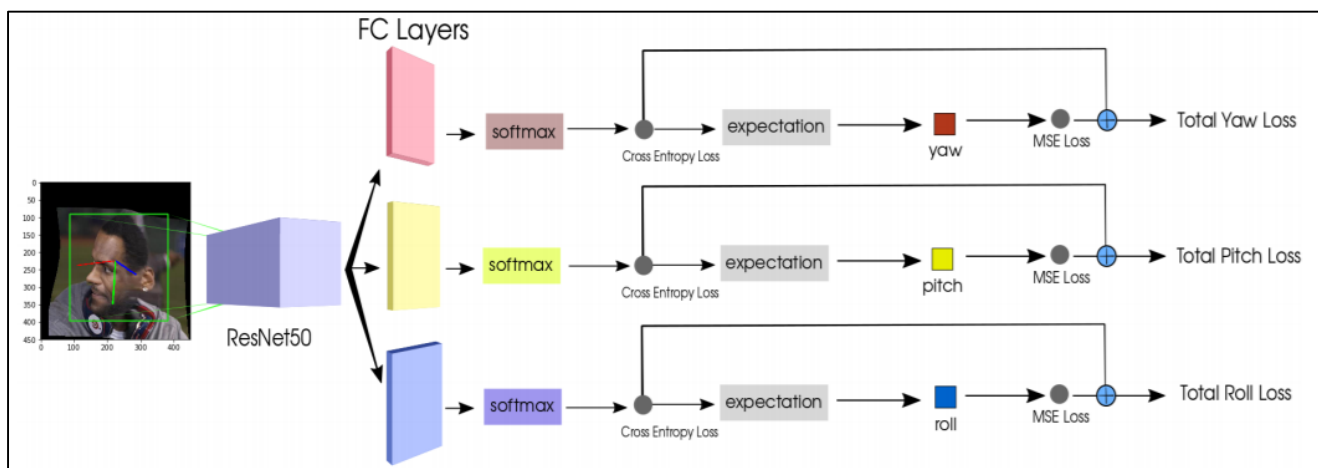


Рисунок 2.2 – Архітектура ResNet50.

Потім до мережі додаються втрати регресії, а саме середньоквадратична втрата помилок, щоб покращити прогнози. Є три кінцеві втрати, по одному на кожен кут, і кожен є лінійною комбінацією як відповідної класифікації, так і регресійних втрат.

Сучасні методи оцінки пози голови повинні розроблятися, щоб покращити оцінку вектору голови на зображеннях поганої якості або низької роздільності. Глибокі мережі, які прогнозують позицію безпосередньо від інтенсивності зображення, є хорошим методом для вирішення цієї проблеми, оскільки надійність може бути вбудована в них шляхом зміни мережі або розширення її навчальних даних розумними способами. Навчаючи мережу на даних з різною роздільною здатністю, можна спромогтись створити ефективне представлення для зображень різної якості. Це має особливе значення для розпізнавання людей на великій відстані. Методи засновані на використанні камер з глибинними датчиками значно програють методам розпізнавання на двомірному зображенні.

2.2 Використані датасети

Для того, щоб досягти прогресу в проблемі прогнозування пози по інтенсивності зображення, ми повинні знайти реальні набори даних, які містять точні анотації пози, численні зображення різних людей, різні умови освітлення, і значне різноманіття поз. Ми виділяємо два дуже різних набори даних, які відповідають цим вимогам.

Набір даних BIWI [15] збирається в лабораторних умовах шляхом запису відео RGB-D різних людей з різними позами голови за допомогою пристрою Kinect v2. Він містить приблизно 15 000 кадрів, а кути обертю $\pm 75^\circ$ для повороту (yaw), $\pm 60^\circ$ для нахилу (pitch) і $\pm 50^\circ$ для обертання (roll). 3D-модель була створена для кожного відмінного обличчя окремо та обертання голови відстежувалися, щоб створювати анотації поз. Цей набір даних зазвичай використовується як орієнтир для оцінки пози за допомогою методів глибини, що підтверджує точність їх міток. Приклади зображень з датасету BIWI представлені на рисунку 2.3.

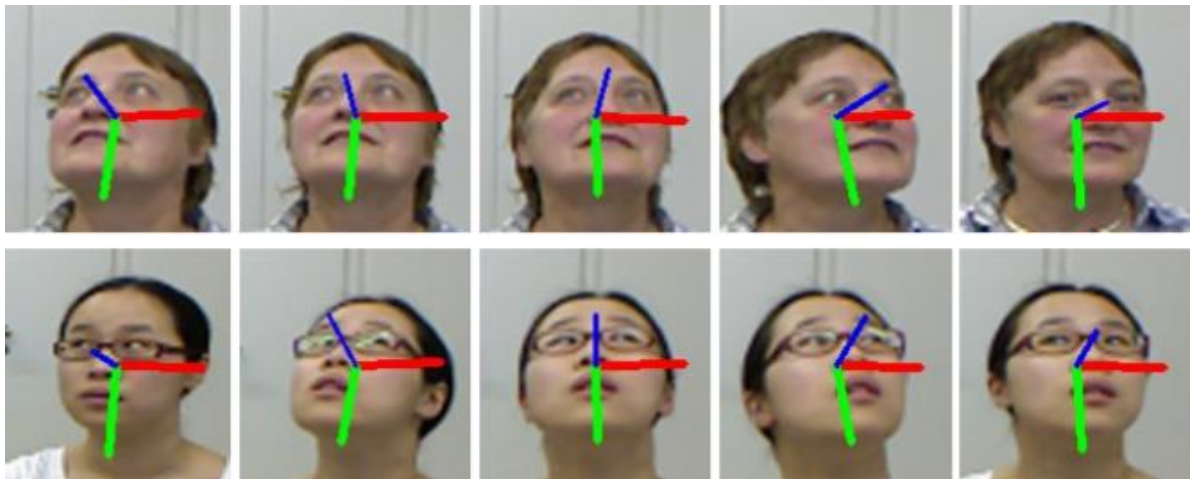


Рисунок 2.3 – Приклад зображень з датасета BIWI.

Інший набір даних 300W-LP[16] є розширенням з 300W, що стандартизує безліч баз даних вирівнювання з 68 орієнтирами, включаючи AFW, LFPW, HELEN,

IBUG та XM2VTS. Цей датасет пропонує синтетично розширені дані для підготовки моделей з виявлення орієнтирів. Це колекція популярних двовимірних наборів даних, які були згруповані та повторно анотовані(див. рис. 2.4).



Рисунок 2.4 – Приклад синтетичних зображень з набору 300W-LP

Модель обличчя формується до кожного зображення, і зображення спотворюється, щоб змінювати поворот обличчя, що дає нам змогу з фронтального обличчя створити безліч поз. Синтетичність набору дає змогу створювати зображення з різними значеннями відхилення по всіх трьох напрямках і при цьому отримувати максимально точні мітки, що значно впливає на якість тренованих CNN моделей.

Для тестування був використаний датасет AFLW2000. Це набір даних, побудований шляхом повторного анотування перших 2000 зображень від AFLW, використовуючи 68 3D-орієнтирів у відповідності з зображеннями з 300W-LP-3D. Обличчя цього набору даних містять великі варіації поз (відхилення від -90° до 90°) з різними виразами та умовами освітлення.

3 ОПИС ПРАКТИЧНИХ ДОСЛІДЖЕНЬ

Розглянуті раніше методи були досліджені за наступними показниками:

- точність роботи методу;
- швидкість роботи;
- максимальний кут розпізнавання.

Точність роботи методу з визначення пози голови зазвичай обчислюється у MAE (Mean absolute error). У статистиці середня абсолютна помилка (MAE) - це міра різниці між двома безперервними змінними. Припустимо, X і Y є змінними парних спостережень, які виражають одне і те ж явище. Приклади Y порівняно з X включають порівняння прогнозованого порівняно із спостережуваним, подальший час проти початкового часу та одну техніку вимірювання проти альтернативної методики вимірювання. Розглянемо графік розсіяння n точок, де точка i має координати (x_i, y_i) . Середня абсолютна помилка (MAE) - середня вертикальна відстань між кожною точкою та ідентичною лінією. MAE - це також середня горизонтальна відстань між кожною точкою та ідентичною лінією. MAE задається наступною формулою:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n} \quad (3.1)$$

Середня абсолютна помилка - це середнє значення абсолютних помилок $|e_i| = |y_i - x_i|$, де y_i - це передбачене значення, а x_i - це реальне значення.

Експеримент проводився з використанням двох різних датасетів: BIWI, AFLW2000. Порівняння значень за основними напрямками yaw, pitch, roll та середнє значення помилки при використанні BIWI датасету представлені на графіку 3.1.

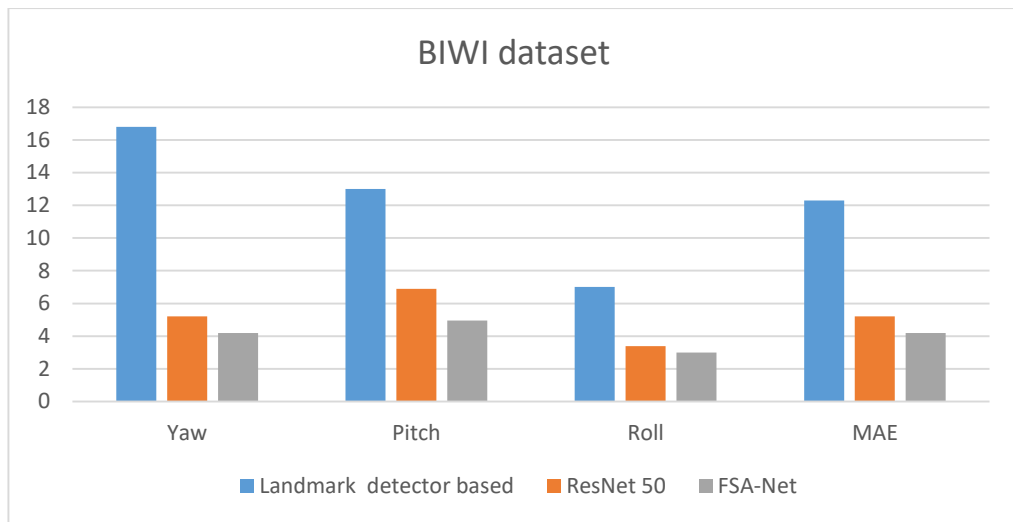


Рисунок 3.1 – Графік результатів роботи методів з використанням BIWI датасету

Той самий експеримент було проведено з датасетом AFLW2000 (див. рис. 3.2). Основна відмінність цих датасетів полягає у тому, що BIWI – це датасет, отриманий в лабораторних умовах, тому зображення мають однакове освітлення та досить хорошу якість. AFLW2000 [17] – це датасет отриманий у реальних умовах, де освітлення, розмитість та якість зображення дуже різні.

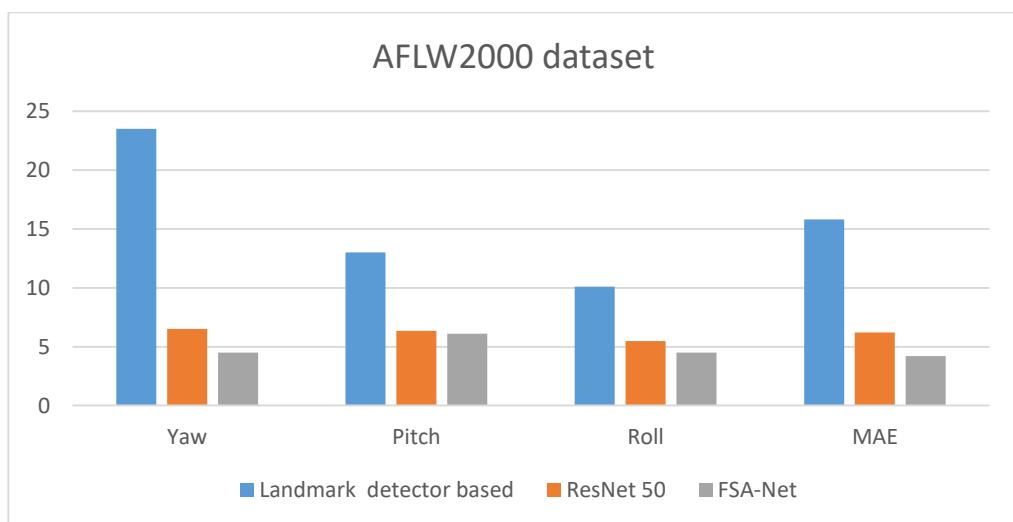


Рисунок 3.2 – Графік результатів роботи методів з використанням AFLW2000 датасету

З даних отриманих під час експерименту видно, що методи, які не базуються на пошуку ключових точок показують досить хороші результати навіть на датасеті

з AFLW2000. Середнє значення помилки для landmark-based методу складає 15.8, що означає помилку в 15.8°. У той час, як ResNet 50 та FSA-Net дозволяють досягти точності у 6.4° та 4.2° відповідно. Це є приводом вважати, що потенціал методів, які базуються на CNN вищий за методи ключових точок.

Швидкість роботи досліджувалася на фреймах отриманих зі звичайної камери з FPS 24. Розрахунки проводилися з використанням процесору intel core i7, RAM 16Gb. Швидкість розрахунку пози для одного фрейма та одного обличчя у фреймі наведена у таблиці 3.1. Виходячи з результатів, що ми отримали, найбільш точний метод FSA-Net показує найгірший результат швидкості 133 мс, що у 3 рази більше, ніж бажана швидкість обробки у 40 мс. Якщо використовувати цей метод, то ми зможемо порівнювати кожен четвертий фрейм у відео. Така низька швидкість виконання обумовлена великою кількістю кроків обробки зображення та значно залежить від швидкості роботи детектора облич, без якого жоден метод не може обійтись.

Таблиця 3.1 – Тест швидкості

№	Назва методу	Середня швидкість виконання
1	Landmark-based detector	75 мс
2	ResNet 50	105 мс
3	FSA-Net	133 мс

Максимальний кут розпізнавання досліджувався за допомогою датасету BIWI, який містить приблизно 15 000 кадрів, а кути оберту $\pm 75^\circ$ для повороту (yaw), $\pm 60^\circ$ для нахилу (pitch) і $\pm 50^\circ$ для обертання (roll). Результати точності розпізнавання куту в залежності від його значення наведено на рисунку 3.3.

Експеримент показав, що усі методи майже однаково добре розпізнають усі доступні кути датасету. Однак landmark-based метод показав найменш точне розпізнавання різниці між граничними кутами, що обумовлено скритістю частини опорних точок при повороті голови. Для landmark-based методів недостатність

значної частини точок через перекривання обличчя іншими частинами тіла або головним убором призводить до неможливості визначити кут.

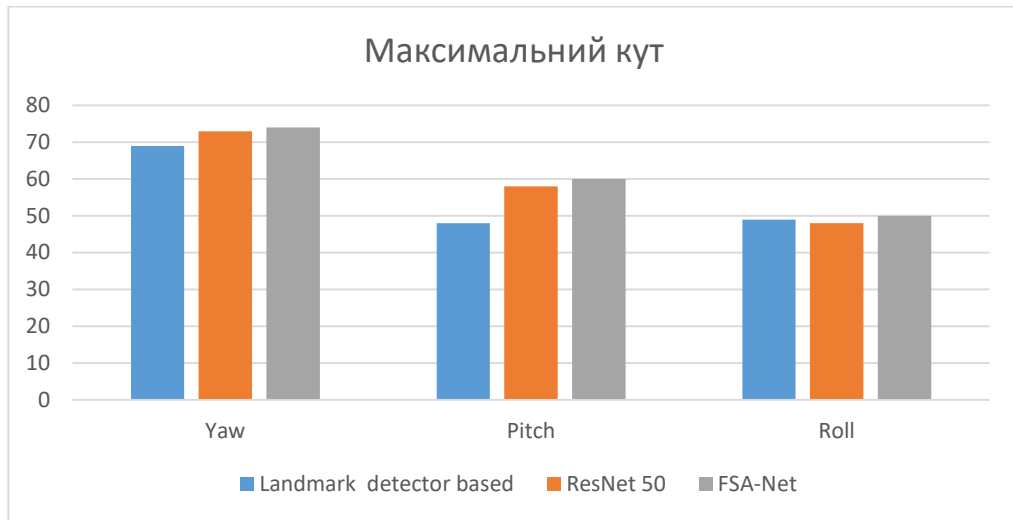


Рисунок 3.7 – Максимальний кут повороту голови

За результатами досліджень алгоритми, що не базуються на опорних точках показали значно кращі результати точності особливо на датасеті в реальних умовах та при поганому освітленні. Різниця в 2° для класифікаційно-регресійних моделей є досить важливим кроком у покращенні точності розпізнавання вектору направленості обличчя. FSA-Net, який дозволяє отримати 98,8% точності (з розрахунку помилки у 4.2° з 360°) є кращим на сьогодні кандидатом для використання у системах оцінки точності рухів голови людини.

4 ПРОГРАМНА РЕАЛІЗАЦІЯ

4.1 Архітектура додатку

Оскільки практична цінність методів розпізнавання вектору повороту обличчя найбільш значима у збірних системах моніторингу рухів тіла, то доцільно створити модуль, який стане складовою частиною більш складної системи. Такий модуль надаватиме змогу зчитувати потік кадрів з відео файлу або з камери та порівнюватиме рухи голови людей на кадрі з еталонною моделлю.

Виходячи з результатів попереднього дослідження, найліпшим методом з точки зору точності знаходження вектору обличчя є FSA-Net. Мережа, яка описана у цьому методі, була натренована за допомогою TensorFlow. TensorFlow – це відкрита програмна бібліотека для машинного навчання, розроблена компанією Google для вирішення завдань побудови і тренування нейронної мережі з метою автоматичного знаходження та класифікації образів, досягаючи якості людського сприйняття. Застосовується як для досліджень, так і для розробки власних продуктів Google. Основний API для роботи з бібліотекою реалізований для Python, також існують реалізації для C Sharp, C ++, Haskell, Java, Go і Swift. TensorFlow є системою машинного навчання Google Brain другого покоління. У той час як еталонна реалізація працює на одиничних пристроях, TensorFlow може працювати на багатьох паралельних процесорах, як CPU, так і GPU, спираючись на архітектуру CUDA для підтримки обчислень загального призначення на графічних процесорах. TensorFlow доступна для 64-розрядних Linux, macOS, Windows, і для мобільних обчислювальних платформ, включаючи Android і iOS.

Для спрощення взаємодії з TensorFlow була використана бібліотека Keras. Keras - це API нейронних мереж високого рівня, написаний на Python і здатний працювати над TensorFlow, CNTK або Theano. Він був розроблений з акцентом на швидке експериментування. Keras дозволяє легко та швидко розробляти прототипи, підтримує як конволюційні мережі, так і періодичні мережі, а також їх комбінації, працює безперебійно на CPU та GPU.

Модуль може працювати у двох режимах: запису рухів еталону, порівняння записаної послідовності рухів з рухами людини на відео. Модуль має підтримувати роботу з потоковим відео з камери, тому співставлення має проходити у режимі реального часу.

Розглядаючи можливі методи співставлення поз голови були виділені два методи: покадрове співставлення, порівняння опорних точок. Перший метод є значно простішим з точки зору реалізації, але у нього є значні недоліки. По-перше, точність цього методу залежить від швидкості роботи методу пошуку вектору обличчя, якщо ця швидкість нижча за 40 мс (проміжок між кадрами при записі відео FPS 24), то деякі кадри будуть пропущені, що призведе до помилково негативного сценарію. Також людина, що виконує вправи, може змінювати темп руху. Тому факт виконання вправи не слід пов'язувати з часом.

Метод опорних точок спрощує процес запису еталонної вправи, бо відтепер ми не маємо записувати кожне положення голови, а лише ті положення у яких рух почався у іншому напрямку. Цей метод полягає у відстеженні вектору, за яким обличчя рухається у часі. Оскільки вектор направленості голови можна представити як $\vec{h} = (r, p, y)$, де r – це кут нахилу голови по осі z , p – кут нахилу голови по осі x , y – кут нахилу голови по осі y , то опорною точкою є такий вектор направленості, при якому вектор зміни положення голови змінює свій напрямок. Опорну точку можна знайти за наступною формулою

$$|x_i| - |x_{i+1}| \leq |x_i| - |x_{i-1}|, \quad (4.1)$$

де x_i – це значення куту на i -тому фреймі. Якщо нерівність виконується хоча б для одного з показників, то положення запам'ятовується як опорна точка.

Загальний алгоритм роботи наведено на рисунку 4.1. Потік фреймів з камери або відео файлу поступає до детектору облич. Кожне знайдене обличчя виділяється з фрейму, нормалізується та передається на вхід до моделі, що повертає передбачення пози. Кожен вектор пози перевіряється на відповідність нерівності

4.1 та у разі співпадіння записується до носія. Кожен фрейм виводиться на екран з векторами, що вказують напрямок голови за кожним з критеріїв (roll, pitch, yaw).

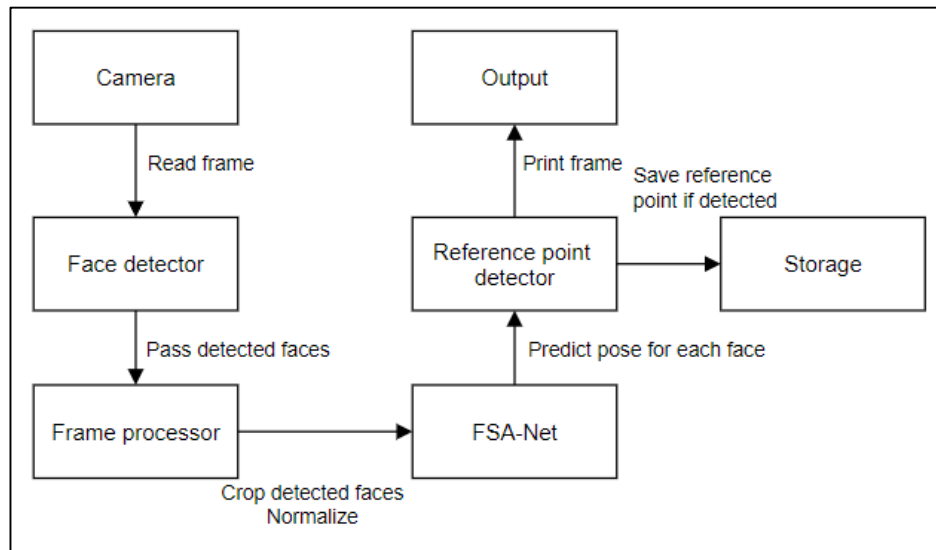


Рисунок 4.1 – Схема роботи алгоритму запису опорних точок.

Майже таким самим чином проходить перевірка на відповідність еталону вправи. На вхід подається файл у який ми попередньо записали еталонні рухи у вигляді опорних точок. Кроки з моменту зчитування з камери і до отримання передбаченої пози незмінні. Так само кожна поза перевіряється на можливість опорної точки та порівнюється з наступною опорною точною еталону з допуском у 10° . Якщо пози співпали, то точка вважається пройденою та для порівняння обирається наступна.

Людина що виконує вправи має знати у якому напрямку їй рухатись, щоб виконати вказану вправу. Тому до режиму перевірки були додані направляючі вектори, що вказують на бажаний напрям руху відносно положення у якому знаходиться обличчя людини зараз. Один вектор для pitch та yaw та інший для roll. Одного вектору недостатньо, щоб вказати три напрями одночасно.

Одне з важливих питань полягає у тому, який допуск є найліпшим для порівняння. З одного боку допуск має бути більше, ніж максимальна похибка алгоритму, з іншого достатньо мала, щоб це не спричинило похибково позитивні результати виконання вправи.

4.2 Розробка програмної реалізації

У результаті реалізації методу був отриманий модуль на мові Python. Модуль використовує бібліотеку OpenCV для роботи з зображенням, TensorFlow та Keras для створення моделі пошуку вектору обличчя. Детектор обличчя використовує претреновану модель SSD.

У режимі запису еталонної послідовності рухів слід використовувати відео на якому є лише одна персона. На рисунку 4.2 зображений фрейм з камери у режимі запису. Вектор синього кольору вказує напрям голови з урахуванням осі x та y. Червоний та зелений вектори враховують поворот голови за вектором z.

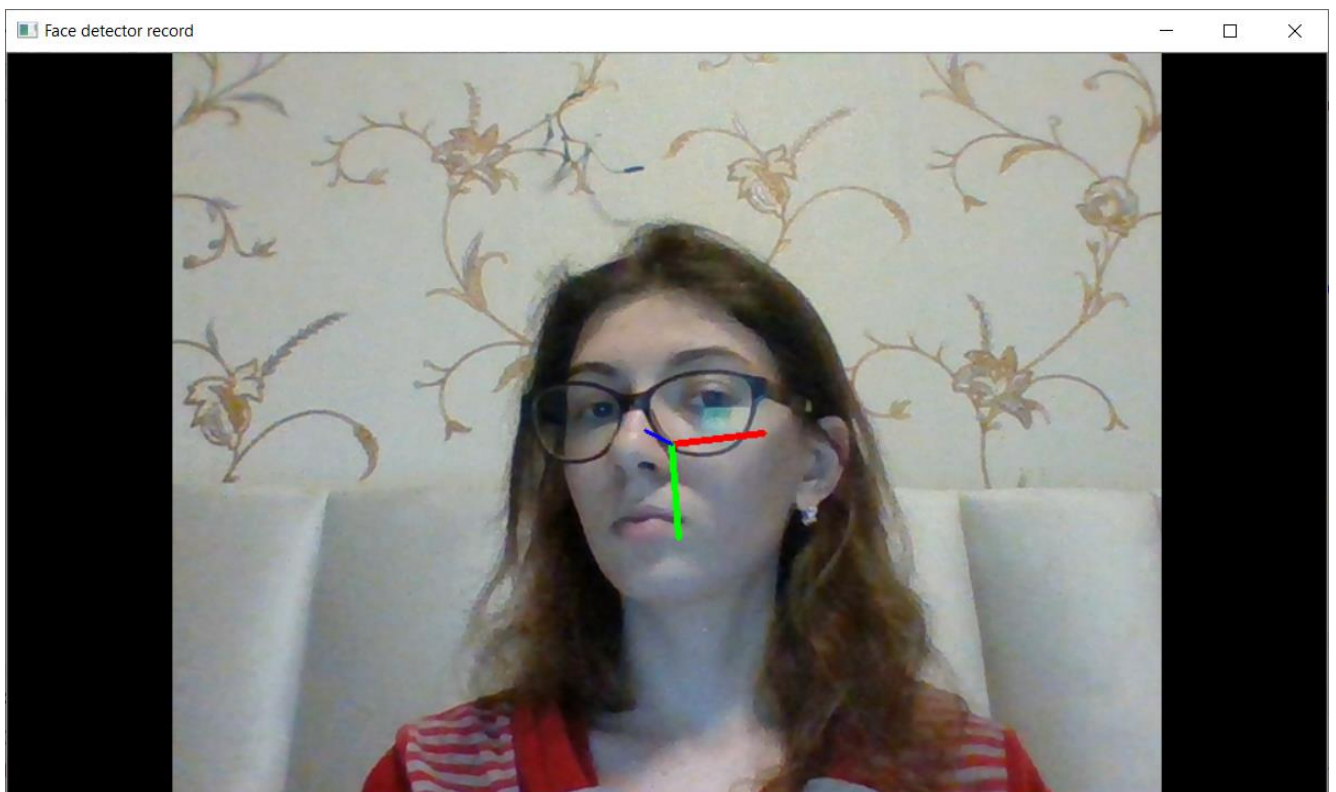


Рисунок 4.2 – Кадр з вектором напрямку обличчя у режимі запису.

Кути направленості голови у опорних точках записуються у текстовий файл в послідовності “yaw pitch roll”. У подальшому ці дані можуть бути записані у базу

даних разом з даними про загальне положення тіла, які будуть отримані з інших модулів.

У режимі перевірки вектори, що представляють положення голови не відображаються. Натомість зображується вектор підказка, що вказує бажаний напрямок руху (див. рис. 4.3).

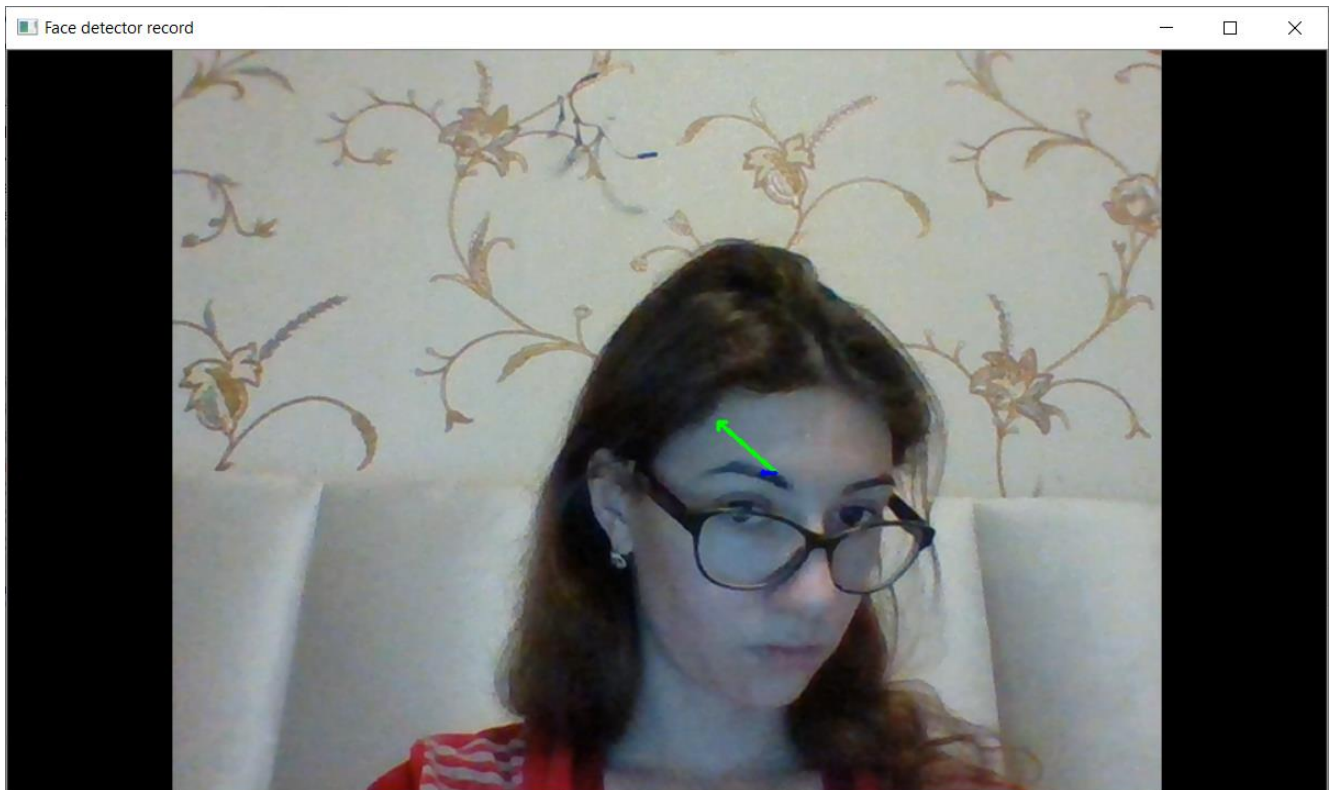


Рисунок 4.3 – Кадр з вектором, що вказує напрям руху.

Якщо опорна точка співпала з поточним положенням голови, то програма дає користувачеві сигнал зеленим прямокутником (див. рис. 4.4) та зображується наступний вектор підказка доки вправа не буде виконана. Оскільки на зображенні може бути кілька людей, підказка відображається посередині обличчя кожної окремої людини.

Оскільки у даній роботі насамперед розглядалися методи визначення направленості обличчя питання пошуку найбільш підходящого детектора облич залишається відкритим. Найбільшою проблемою детекторів з точки зору пошуку

вектору є ігнорування облич на значній відстані від камери. Через це робота алгоритму з малою кількістю людей набагато ефективніша за роботу згуртом.

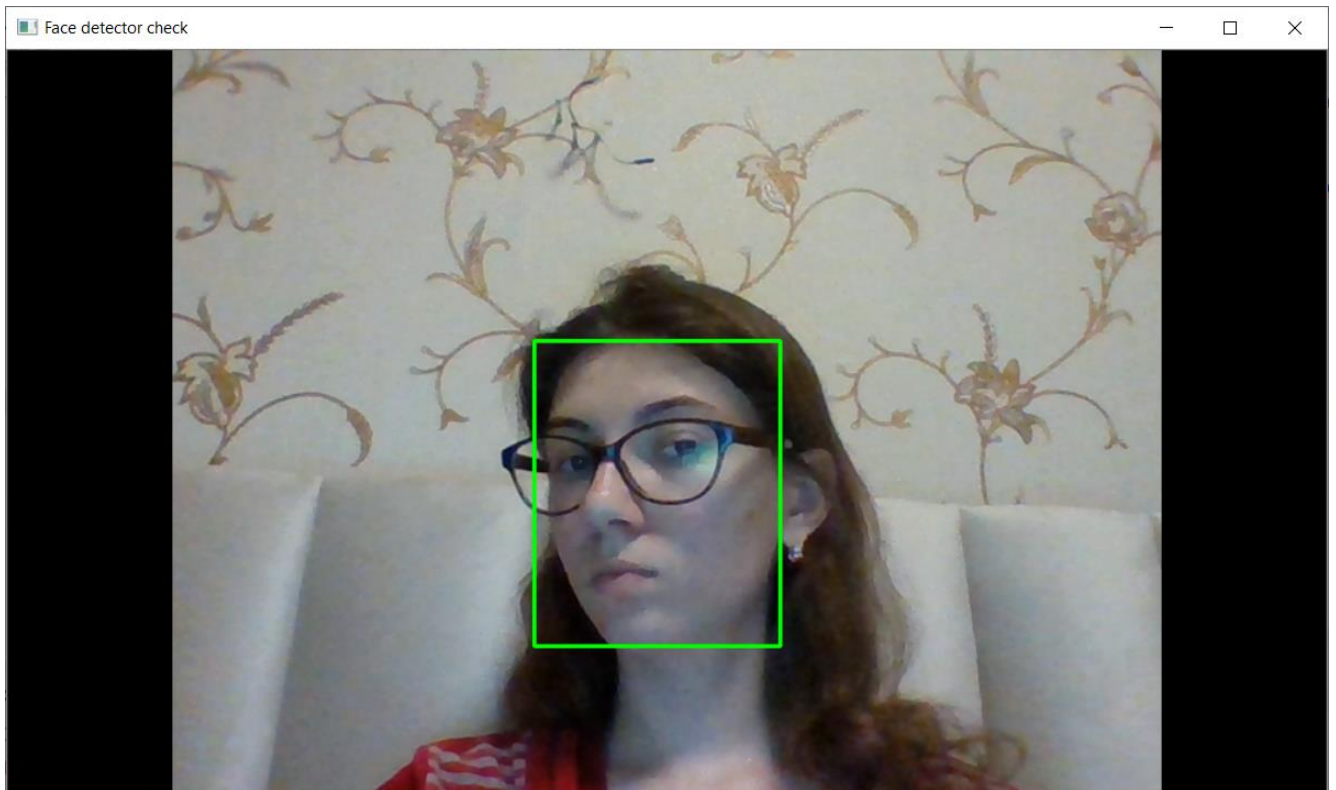


Рисунок 4.4 – Кадр з виконаною позою.

Як частина більшої системи, цей модуль може бути використаний для перевірки рухів голови, що разом з іншими даними про положення тіла дасть змогу оцінити точність фізичних вправ без використання глибинних камер та датчиків.

5 ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ У ПРАКТИЧНУ ДІЯЛЬНІСТЬ

5.1. Опис варіантів використання отриманих результатів

У роботі були розглянуті два основні напрями пошуку вектору направленості обличчя. Методи засновані на співставленні двомірного зображення з тривимірним використовують попередньо виявлені опорні точки обличчя. Це робить їх вдалим вибором для систем, які заточені під пошук орієнтирів обличчя, оскільки за результатами дослідження вони показують можливість найшвидшого виконання у порівнянні з іншими методами, та значно залежать від точності знайдених орієнтирів. Основним підходом до покращення точності цих алгоритмів є покращення точності детекторів орієнтирів.

Як було зазначено раніше детектори орієнтирів наразі втрачають точність в умовах недостатнього освітлення або низької якості зображення, тому методи засновані на зовнішності слід використовувати, якщо пошук орієнтирів не є основним завданням системи. Згідно з результатами дослідження, методи засновані на зовнішності менш вразливі до незвичних умов та здатні передбачити положення навіть у разі, якщо частина обличчя перекрита іншими предметами або зачіскою. Системи, що слідкують за поглядом водія або правильністю виконання вправ вірогідно будуть використовувати саме ці методи, у той час, як ігрові системи отримають більше можливостей з методами заснованими на орієнтирах. Подальший розвиток цієї групи методів полягає у вдосконаленні функцій втрат та попередній структуризації особливостей, що передаються до агрегатора. Найновіші дослідження у цій сфері намагаються поєднати два підходи та передавати на вхід до мережі положення орієнтирів обличчя. Створення синтетичних датасетів високої точності маркування розширило можливості методів, що базуються на CNN.

Окремо увагу слід приділити комбінації методів відстеження погляду та повороту голови. Відстеження погляду очей щільно пов'язано з загальною позою

обличчя. Оскільки наразі методи відстеження погляду не використовують дані про положення голови, значне відхилення голови від фронтального положення робить неможливим зчитування напрямку очей. Створення тримірної моделі голови за допомогою орієнтирів обличчя дозволить передбачити погляд з будь якого положення. На поєднанні інформації про напрям погляду та повороту обличчя можна створити систему безконтактного управління пристроями для людей з обмеженими можливостями.

5.2 Існуючі проблеми та напрямки подальшого дослідження

Задача розпізнавання напрямку голови досить добре вивчена на даний момент. Досягнута точність у 98%, є досить значним показником та може вважатися достатньою для певних сфер використання, але все ж залишається питання щодо розширення можливостей існуючих алгоритмів.

До недоліків, що характерні для всіх методів визначення пози обличчя належить обмежений діапазон кутів на якому вони здатні працювати. По-перше, це пов'язано з можливостями детекторів обличчя, які схильні до значних похибок або повного ігнорування обличчя, якщо кут повороту по осі у перевищує 90° . Незначні зміни у положенні рис обличчя при повороті призводять до того, що похибка у виявленні збільшується зі збільшенням значення кута.

По-друге, датасети, які використовуються у процесі навчання мережі зазвичай не містять зображень з діапазоном кутів повороту більшими за $95^\circ - 98^\circ$, тому створення нових датасетів, заточених під задачу виявлення вектору повороту обличчя з повним діапазоном можливих положень голови, дасть змогу покращити характеристики існуючих систем.

Швидкість роботи алгоритму є істотним показником при роботі з потоковим відео. За результатами експерименту більш точні алгоритми програють у

швидкості виконання через наявність додаткових кроків попередньої обробки, але за можливості використання потужних процесорів ця різниця не буде істотною.

Наразі алгоритм не бере до уваги швидкість виконання вправ, бо кожна людина може працювати у своєму темпі. У деяких випадках, на кшталт фітнесу або танців, можливість оцінювати синхронність рухів є досить важливою. Для вирішення цієї задачі має бути достатньо долучити до запису значення вектору напрямку голови інтервал між двома послідовними рухами, які у даній реалізації представлені опорними точками. Також можливе порівняння значень напрямів голови різних людей на одному зображенні, що спростить тренеру процес спостереження за рухами учнів.

ВИСНОВКИ

В рамках роботи було проведено дослідження методів визначення вектору направленості обличчя людини та їх порівняння. Ми ознайомилися з різними підходами до вирішення цієї проблеми. З них підходи засновані на попередньому знаходженні опорних точок та підходи засновані на тренуванні CNN моделі.

Для проведення аналізу були обрані 3 методи визначення пози голови: метод заснований на опорних точках, ResNet 50 та FSA-Net. За результатами проведеного дослідження методи засновані на регресії показали значно кращу точність та незалежність точності від освітлення та часткової закритості обличчя. Всі методи, а в особливості метод заснований на опорних точках показали погіршення в розпізнаванні граничних кутів через незначну відмінність характеристик обличчя. Це є вікном для можливого покращення роботи методу в умовах відстежування людини повернутої до камери під кутом більшим за 70° .

Значної оптимізації потребує швидкість роботи методу FSA-Net оскільки за швидкістю виконання в 3 рази перевищуючу менш точні аналоги, кілька фреймів може бути пропущено. Оскільки процес визначення пози голови є багатокроковим, оптимізація можна провести як на кроці знаходження облич на зображенні, використавши більш швидкі детектори, так і на кроці знаходження пози.

Дослідження показує, що методи знаходження пози голови людини можуть конкурувати за точністю з методами, які засновані на використанні глибинних датчиків. При цьому методи роботи з простим RGB зображенням мають значно менше обмежень за місцем використання.

З використанням найбільш точного методу з проаналізованих був розроблений програмний продукт, що здатен розпізнавати, записувати та перевіряти рухи голови людини з відео потоку. У подальшому цей продукт може стати частиною системи моніторингу людських рухів.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Satya Mallick, Head Pose Estimation using OpenCV and Dlib. 2016. URL: <https://www.learnopencv.com/head-pose-estimation-using-opencv-and-dlib/>. (дата звернення: 15.03.2020)
2. Guido Borghi, Matteo Fabbri, Roberto Vezzani, Simone Calderara, Rita Cucchiara, Face-from-Depth for Head Pose Estimation on Depth Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018. № 42, С. 596 – 609.
3. Paul Viola , Michael Jones, Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2001.
4. G. Shcherbakova, V. Krylov, N. Bilous, Methods of automated classification based on wavelet-transform for automated medical diagnostics. *Information Technologies in Innovation Business Conference (ITIB)*. 2015. С. 7 – 10.
5. Frank Wallhoff and Markus Ablaßmeier and Gerhard Rigoll. Multimodal Face Detection. *Head Orientation and Eye Gaze Tracking, IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*. 2006. С.13 – 18.
6. Yinqiang Z., Yubin K., Shigeki S., Kalle Å., Masatoshi O. Revisiting the PnP Problem: A Fast, General and Optimal Solution. *IEEE International Conference on Computer Vision*. 2013.
7. B. Ahn, J. Park, I.S. Kweon, Real-time head orientation from a monocular camera using deep neural network. *Asian Conference on Computer Vision*. 2014. С. 82 – 96.
8. M. Patacchiola, A. Cangelosi, Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. *Pattern Recognition*. 2017. № 71, С. 132 – 143.
9. Amit Kumar, Azadeh Alavi, and Rama Chellappa, KEPLER: Keypoint and pose estimation of unconstrained faces by learning efficient H-CNN regressors. *12th IEEE International Conference on Automatic Face & Gesture Recognition*. 2017. С. 258 – 265.

10. Xu X., Kakadiaris I.A. Joint head pose estimation and face alignment framework using global and local CNN features. *12th IEEE International Conference on Automatic Face & Gesture Recognition*. 2017. C. 642 – 649.
11. Bin Huang, Renwen Chen, Wang Xu, Qinbang Zhou, Improving head pose estimation using two-stage ensembles with top-k regression. *Image and Vision Computing*. 2020. № 93.
12. Lam Thanh Hien, Do Nang Toan, Tran Van Lang, Detection of Human Head Direction Based on Facial Normal Algorithm. *International Journal of Electronics Communication and Computer Engineering*. 2015. № 6, C. 110 – 114.
13. Tsun-Yi Y., Yi-Ting C, Yen-Yu L, Yung-Yu C., FSA-Net: Learning Fine-Grained Structure Aggregation for Head Pose Estimation from a Single Image. *Conference on Computer Vision and Pattern Recognition*. 2019. C. 1087 – 1096.
14. Nataniel Ruiz, Eunji Chong, and James M. Rehg. Finegrained head pose estimation without keypoints. *Conference on Computer Vision and Pattern Recognition*. 2018. C. 1821 – 1829.
15. Gabriele Fanelli, Matthias Dantone, Juergen Gall, Andrea Fossati, and Luc Van Gool., Random forests for real time 3D face analysis. *International Journal of Computer Vision*. 2013. № 101, C. 437 – 458.
16. Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, Stan Z. Li, Face alignment across large poses: A 3D solution. *Conference on Computer Vision and Pattern Recognition*. 2016. C. 146 – 155.
17. Martin Koestinger, Paul Wohlhart, Peter M. Roth, Horst Bischof. Annotated facial landmarks in the wild: A largescale, real-world database for facial landmark localization. *International Conference on Computer Vision Workshops*. 2011. C. 2144 – 2151.