

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук
(повна назва)

Кафедра _____ Програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти _____ другий (магістерський)

**Дослідження способів застосування і реалізацій технологій
комп'ютерного зору для розробки мобільного
застосування оцінки положення людини у просторі**
(тема)

Виконав:

Випускник 2 курсу, групи ІПЗм-19-2
Свідін О.Є
(прізвище, ініціали)

Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми Освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Керівник проф. Білоус Н.В.
(посада, прізвище)

Допускається до захисту

Зав. кафедри _____ 3.В. Дудар
(підпис) (прізвище, ініціали)

2021 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Програмної інженерії
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Інженерія програмного забезпечення
(повна назва)

ЗАТВЕРДЖУЮ:

Зав.кафедри _____

(підпис)

« ____ » _____ 2021 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

студента Свідіна Олега Євгеновича
(прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження способів застосування і реалізацій технологій комп'ютерного зору для розробки мобільного застосування оцінки положення у просторі»

затверджена наказом університету від 26.03.2021 № 385Ст

2. Термін подання студентом роботи до екзаменаційної комісії «08» травня 2021р

3. Вихідні дані до роботи методи оцифровування та ідентифікації поз людини засобами мобільно пристрою

4. Перелік питань, що необхідно опрацювати в роботі мета роботи, аналіз предметної галузі і постановка задачі, методи оцифровування рухів людини, методи ідентифікації поз людини

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, ілюстрацій (слайдів) мета завдання, обґрунтування доцільності розроблення,

постановка задачі, методи і архітектурні підходи, структурно-логічна схема взаємодії даних, опис отриманих результатів, демонстраційні матеріали

6. 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	25.01.21 – 19.02.21	Виконано
2	Огляд існуючих методів до оцифровування рухів	20.02.21 – 10.03.21	Виконано
3	Огляд існуючих методів до ідентифікації пози та розробка нового	11.03.21 – 25.03.21	Виконано
4	Підготовка пояснювальної записки	26.03.21 – 15.04.21	Виконано
5	Спецчастина	15.04.21 – 17.04.21	Виконано
6	Підготовка презентації та доповіді	18.04.21 – 19.04.21	Виконано
9	Попередній захист	12.05.21	Виконано
7	Нормконтроль, рецензування	14.05.21 – 17.05.21	Виконано
8	Занесення диплома в електронний архів	20.05.21	Виконано
10	Допуск до захисту у зав. кафедри	23.05.21	Виконано

Дата видачі завдання 25 _____ січня _____ 2021 р.

Студент _____
(підпис)

Керівник роботи _____ проф. Білоус Н.В.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Кваліфікаційна робота магістра містить: 75с., 15 рис., 24 джерела.

ПОЗА, ВПРАВА, РУХ, ОЦИФРОВУВАННЯ РУХУ, МЕТОД КУТІВ, МЕТОД ВІДСТАНЕЙ, ARKIT, UNITY.

Об'єктом дослідження є ідентифікація вправ, виконаних людиною, у рухах записних на камеру.

Метою роботи є дослідження підходів для вирішення проблеми ідентифікації фізичних вправ, виконуваних людьми, за допомогою камери та обчислювальних потужностей мобільного пристрою.

Результатом роботи є розроблений метод, що дозволяє працювати з даними про рухи більш ефективно, рекомендації, щодо оцифровування рухів за допомогою камери смартфона, прототип застосунку, що демонструє переваги розробленого метода.

POSITION, EXERCISE, MOVEMENT, DIGIT DIGITALIZATION, ANGLE METHOD, DISTANCE METHOD, ARKIT, UNITY.

The object of the research is the identification of exercises performed by a person in the movements of the camera.

The aim of the work is to study approaches to solving the problem of identifying physical exercises performed by humans using a camera and computing power of a mobile device.

The result is a developed method that allows you to work with data about movements more efficiently, recommendations for digitizing movements using a smartphone camera, a prototype application that demonstrates the benefits of the developed method.

Я, *Свідін Олег Євгенович*, студент групи ІПЗм-19-2, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження способів застосування і реалізацій технологій комп'ютерного зору для розробки мобільного застосування оцінки положення людини у просторі», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ФОП	Формальний опис пози
СЗР	Система захоплення рухів
ВПЛ	Визначення пози людини
СЗР	Системи захоплення руху
FLIR	Forward-looking infrared
SVM	Support vector machine
PCA	Principal component analysis

ЗМІСТ

ВСТУП.....	9
1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ І ПОСТАНОВКА ЗАВДАННЯ	11
1.1 Загальний опис підходів до оцифровування пози людини	11
1.2 Існуючі системи для захвату рухів	12
1.3 Підходи до ідентифікації поз	15
1.4 Огляд літератури	17
2 ПОСТАНОВКА ЗАДАЧІ	20
2.1 Постановка дослідницької задачі.....	20
2.2 Формування вимог до експериментального додатку.....	21
3 МЕТОД ІДЕНТИФІКАЦІЇ	22
3.1 Опис вирішуваної проблеми	22
3.2 Корегування положень суглобів	23
3.3 Усереднення значень та фільтрація	26
3.4 Ідентифікація поз.....	28
4 ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА МЕТОДА ІДЕНТИФІКАЦІЇ ПОЗ	31
4.1 Методологія проведення експериментів	31
4.2 Результати тестів	32
4.3 Аналіз результатів	34
5 ПРОГРАМНА РЕАЛІЗАЦІЯ	36
5.1 UML-моделювання.....	36
5.2 Загальний опис.....	38
5.3 Зберігання опису пози.....	40
5.4 Приклади коду	41
ВИСНОВКИ	42
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ	44
ДОДАТОК А	47
ДОДАТОК Б.....	48

	8
ДОДАТОК В	49
ДОДАТОК Г	57
ДОДАТОК Д	67

ВСТУП

В медицині, спорті та деяких інших галузях існують практичні завдання для вирішення яких, необхідно виконати аналіз рухів людини. Пацієнт під час реабілітації не завжди може знаходитися під наглядом лікаря, щоб той перевіряв, наскільки якісно пацієнт виконує вправи. Людям, що займаються йогою, важко зрозуміти, чи правильно вони відтворюють послідовність дій, якщо вони не знаходяться під наглядом тренера. Для вирішення цієї проблеми необхідно оцифрувати рухи людини у тривимірному просторі, та проаналізувати, чи виконала вона необхідні вправи.

Дослідження з комп'ютерного зору є однією зі сфер наукової діяльності кафедри Програмної інженерії, детально проблематикою розпізнавання тіла, рухів та обличчя людини займається лабораторія «Інформаційних технологій в освіті та системах комп'ютерного зору» (проф. Білоус. Н.В).

Метою роботи є дослідження підходів для вирішення проблеми ідентифікації фізичних вправ, виконуваних людьми, за допомогою камери та обчислювальних потужностей мобільного пристрою. Для досягнення мети необхідно провести аналіз предметної галузі, обрати технологію захвату рухів людини, що задовольняє вимогам та запропонувати метод для ідентифікації рухів, що зможе віднайти задану множину вправ на відео.

Об'єкт дослідження – ідентифікація вправ, виконаних людиною, у рухах записаних на камеру.

Предмет дослідження – методи ідентифікації вправ людини на відео, що були записані на камеру мобільного пристрою.

Розроблено набір даних з відео, на яких записані різноманітні вправи з різних ракурсів, кожна вправа була виконана декілька разів з постійним пришвидшенням, окрім цього кожна вправа була виконана з різним ступенем правильності: правильна, з варіаціями помилок, неправильна.

Для вибору системи захоплення руху(СЗР) результати роботи різних систем СЗР були порівнянні, за допомогою накладання їх результатів на 3D модель людини і порівняння з відео яке було опрацьоване.

Для визначення кращого методу ідентифікації вправ, кожен з методів був використаний на наборі відео, після чого було порівняне число правильно ідентифікованих вправ, та число вправ, що були ідентифіковані, але на справді не були правильно виконані.

В результаті дослідження доведено можливість ідентифікації фізичних вправ за допомогою мобільного пристрою. Запропоновано оригінальний метод для їх ідентифікації, який за результати тестів має високу стійкість до помилок, що виникли у даних під час оцифровування рухів, що позитивно відрізняє його від інших методів. Має можливість до швидкого розширення списку прав, що можуть бути розпізнані та широкі налаштування, які конфігурують які помилки у вправі були припустимими.

Перевірка розробленого метода довела його придатність для використання у медичній та спортивній галузях, так за допомогою метода може бути розроблений мобільний застосунок, що дозволить тренеру чи лікарю обрати вправи, що мають виконати їх підопічні. Після чого лише необхідно встановити телефон з камерою та записати виконання вправ на відео. На виході метод вкаже які вправи були виконані вірно, та у чому були помилки у тих, які були виконані невірно.

За результатами атестаційної роботи є публікація тез на конференції, які описують розроблений метод ідентифікації поз та порівнюють його результати з іншими методами.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ І ПОСТАНОВКА ЗАВДАННЯ

1.1 Загальний опис підходів до оцифрування пози людини

Визначення пози людини (ВПЛ) активно вивчається у літературі з комп'ютерного зору, ВПЛ включає визначення конфігурацій людського тіла з вхідних даних зібраних сенсорами, зокрема з зображень та відео.

ВПЛ надає інформацію про позу тіла, як геометричні дані ключових точок та як інформацію про рух тіла, ці дані можуть бути використані у великому крузі задач таких як AR, VR, аналіз рухів та інших.

Внаслідок бурхливого розвитку рішень для глибокого навчання в останні роки було наведено приклади того, що такі рішення перевершують класичні методи комп'ютерного зору в різних завданнях, включаючи класифікацію зображень [1], семантичну сегментацію [2] та виявлення об'єктів [3]. Значний прогрес і надзвичайна ефективність вже досягнуті завдяки використанню методів глибокого навчання у завданнях ВПЛ.

Вирішення задачі зазвичай відбувається через ідентифікацію, пошук та відслідковування ключових точок. Для об'єктів це можуть бути кути чи інші важливі ознаки, а для людини ключовими точками є основні суглоби, такі як коліна чи лікті.

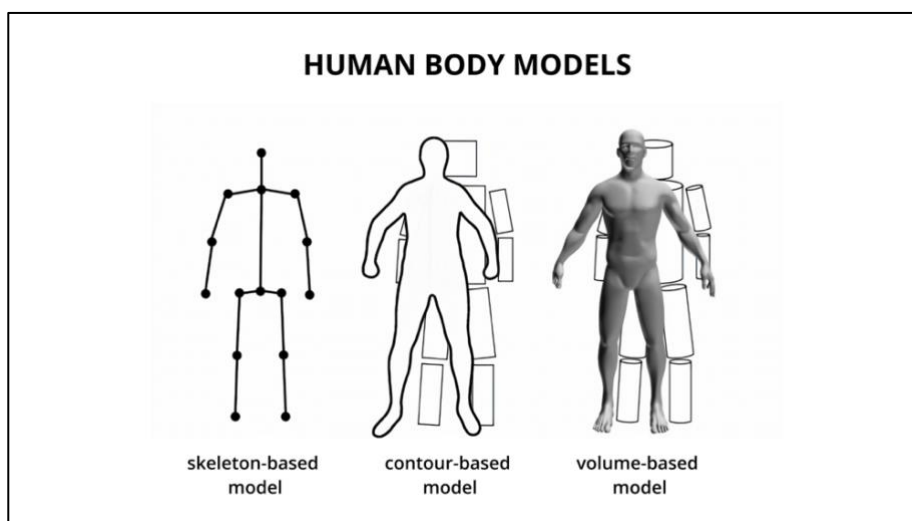


Рисунок 1 – Різні підходи до пошуку пози людини

Існують і інші підходи, так замість ключових точок можуть знаходитися контури людини у випадку двовимірного простору, чи форму у випадку тривимірного простору. У подальшому буде розглядатися метод ключових точок через його більшу простоту та універсальність та через те, що обидва інші варіанти дуже складні у випадку роботи з гнучким і динамічним об'єктом, яким є людина.

Впродовж довгого часу алгоритми комп'ютерного зору могли знаходити лише двовимірні координати людини, після чого на наступних етапах намагалися скорегувати дані, враховуючи проекцію та масштаб. Це дозволяло отримувати непогані результати, але тільки якщо відео було зняте з правильного ракурсу та не з усіма позами. Результати були гарні, коли людина стояла у одній площині та була повернута до камери і рухи також відбувались у тій же самій площині [4].

Визначення пози у тривимірному просторі дозволяє нам визначити реальне положення у просторі зображеної людини. Це є значно складнішою задачею для машинного навчання, так як результат залежить від значно більшої кількості факторів: фон зображення, освітлення та інше [5].

1.2 Існуючі системи для захвату рухів

Вже існують рішення для оцифровування рухів людини. Вони розподіляються на 2 види:

- маркерні, які використовують спеціалізоване обладнання, наприклад, костюми з датчиками;
- безмаркерні, які потребують лише зовнішнього спостереження крізь камеру[6].

Для вирішення практичних завдань, маркерні системи не підходять. Потенційні користувачі не будуть купувати та носити необхідне обладнання. Безмаркерні ж системи підходять, бо не потребують жодного обладнання, окрім камери, та в них є свої проблеми, такі як значні неточності у отриманих даних.

Гарними прикладами безмаркерних рішень є OpenPose, ARKit та VideoPose3D. OpenPose це програмне забезпечення з відкритим кодом, може бути скомпільований на 3 основних платформах Linux, Mac, Windows. Для Windows існує скомпільована версія для роботи з командної строки.

Перевагами цього рішення є можливість працювати з відео у реальному часі, вбудована підтримка різноманітних джерел відео сигналу, таких як ip камери та flir камери. Система розпізнає значну кількість точок та налаштовує які саме точки будуть шукатися, можна увімкнути розпізнання тільки точок тіла та ніг або лише рук чи обличчя.

Недоліками системи є те, що незважаючи на можливість виконувати обробку паралельно, продуктивність залишається відносно низькою. Так на процесорі intel i5 9 покоління обробка взагалі не почалася через недостатню потужність, а на відеокарті 1060 обробка відео с однією людиною у 2D відбувалася зі швидкістю 7 кадрів на секунду.

Для пошуку 3D координат необхідно використовувати Flir камеру певного типу, що має два об'єктиви синхронізовані між собою або використовувати декілька камер, які були відкалібровані та не змінюють ракурс та своє положення. Калібрування не є автоматичним та вимагає певного часу та зусиль. Також потрібно додати, що у реальному часі визначення пози у 3D працює тільки з Flir камерою, у інших випадках занадто багато часу витрачається на передачу даних.

ARKit це пропрієтарна бібліотека Apple, що підтримує визначення 3D пози людини у кадрі. Її основною перевагою є можливість роботи на мобільному пристрої та велика кількість контрольних точок, особливістю є те, що не усі точки, що надаються насправді відстежуються частина з них розраховується відносно іншої контрольної точки, так наприклад пальці рук не відстежуються, а тільки додаються на основі положення зап'ястя. Точність визначення точок є доволі середньою. Серед явних недоліків є велика кількість обмежень до пристроїв на яких бібліотека може бути запущена - лише iPhone та iPad на чипі A12 та вище.

VideoPose3D ця бібліотека може на відмінну від попередньо визначати координати ключових точок у тривимірному просторі за допомогою однієї камери.

Для цього вона використовує оцінку траєкторії рухів для більш точного позиціювання. Дає більшу продуктивність так на відеокарті Nvidia 1060 продуктивність складає 20 кадрів. Надає інформацію про 17 точок, що достатнім для більшої кількості задач. Недоліком є те, що вона не може бути використана на мобільних пристроях.

Після оцифровування рухів людини майже усі без маркерні технології повертають список суглобів з їх координатами, так на (див. рис.2) наведена схематична модель положення суглобів, вони позначенні синіми крапками, а для суглобів хребта додані їх найменування. Різні технології повертають різну кількість суглобів, так OpenPose повертає інформацію про 24 суглоби, а ARKit про 91.

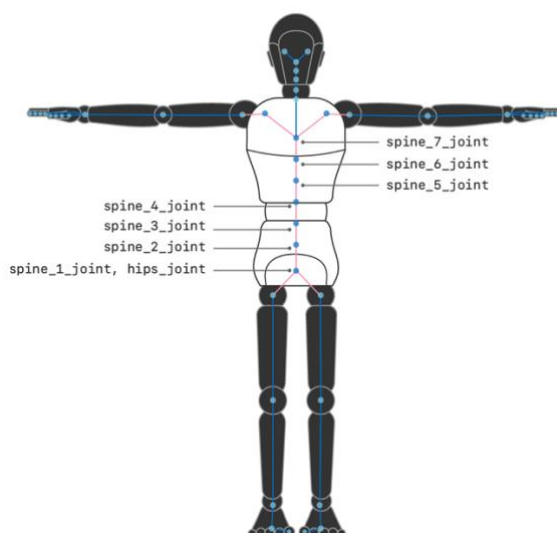


Рисунок 2 – Схематична модель розташування суглобів, що повертається технологією ARKit

Отримані дані про суглоби людини, не відповідають на питання, чи виконала людина певні вправи під час реабілітації, або чи виконала вона певні пози під час занять йогою. Задля цього ми повині ідентифікувати рухи людини.

1.3 Підходи до ідентифікації поз

В результаті роботи моделі ми отримуємо список суглобів з їх позиціюванням у тривимірному просторі. Існують різні способи ідентифікації поз людини і який з них необхідно обрати залежить від задачі та обставин.

Найпоширеніший спосіб ідентифікації поз це порівняння їх координат з такими самим позами які вже були виконанні та промарковані[7].

Якщо ми хочемо ідентифікувати пози однієї людини то ми можемо порівнювати їх з позами, що вона попередньо виконала та записала напрямку використовуючи відстані між точками. Для цього необхідно вирахувати евклідову відстань (див. рис. 3).

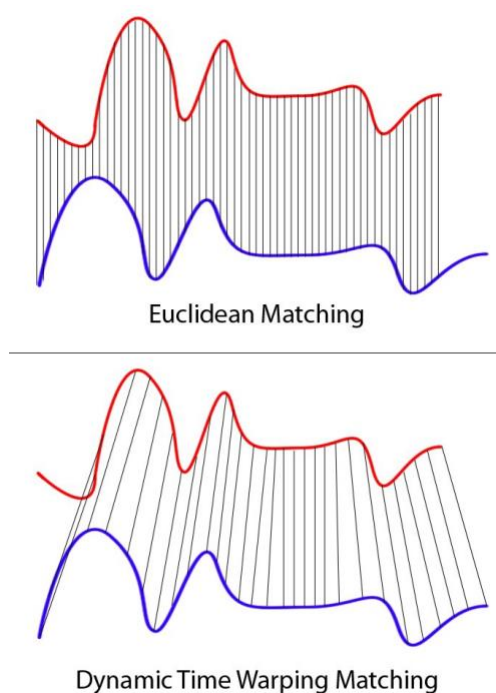


Рисунок 3 – Порівняння двох рухів через евклідові відстані

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

де d – евклідова відстань, n – розмірність простору

Проблемою цього метода є те, що порівнюються саме рухи, а це значить що у відео має відшукуватися необхідний рух, після чого необхідно нормалізувати розміри людини, а потім за допомогою алгоритму DTW вони мають бути синхронізовані. І при усьому цьому навіть ідентичні рухи будуть трохи відрізнятися, у результаті чого алгоритм ніколи не зможе дати сто відсоткову гарантію того, що це саме той рух та пози, що були потрібні.

Якщо ж люди різні то пропорції їх тіла різні і необхідно використовувати підходи, що не залежать від пропорцій тіла, до таких відноситься метод порівняння кутів у суглобах [8]. Для використання цього методу необхідно обрати суглоб кут якого ми будемо вимірювати, та провести від нього лінії до двох інших суглобів, щоб утворився кут який ми будемо вимірювати (див. рис. 4).

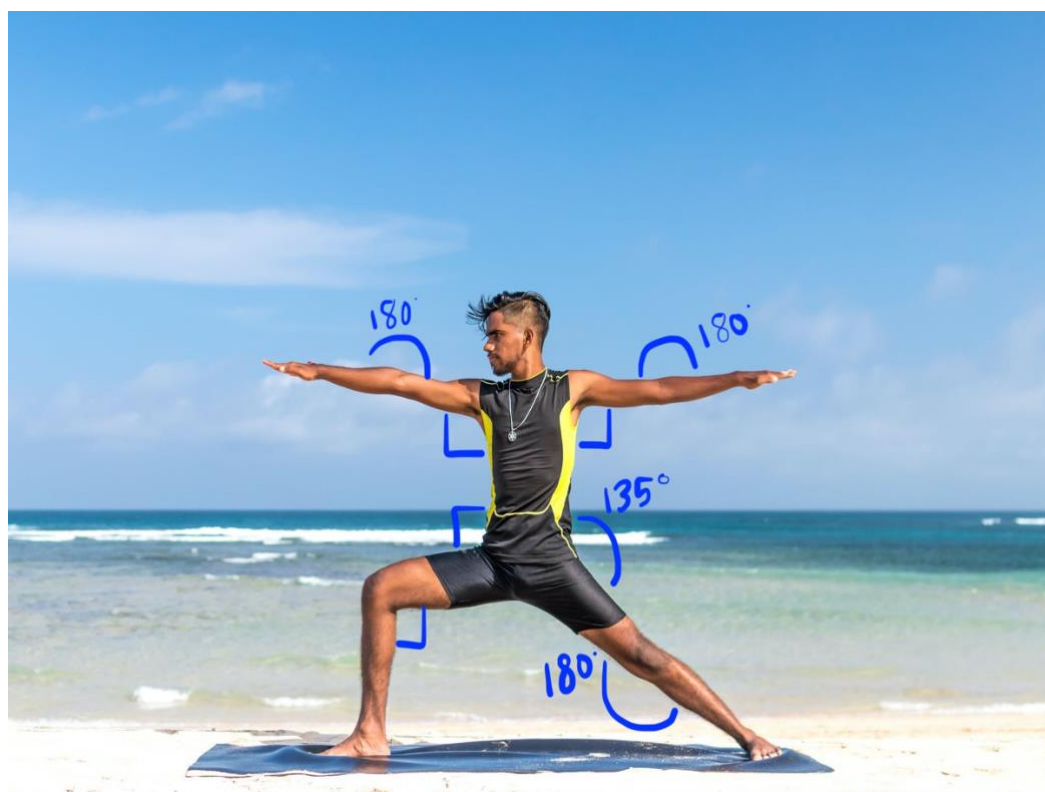


Рис 4 – Пози представлена у вигляді кутів

$$\varphi_{\vec{r}_1 \vec{r}_2} = \arccos \frac{(\vec{r}_1 * \vec{r}_2)}{|\vec{r}_1| * |\vec{r}_2|}$$

$\vec{r}_1 = (x_1 - x_2, y_1 - y_2, z_1 - z_2)$ – перший вектор

$\vec{r}_2 = (x_3 - x_2, y_3 - y_2, z_3 - z_2)$ – другий вектор

Цей метод кращий за попередній, бо він дозволяє ідентифікувати окремі пози не тільки через порівняння з попередньо відзнятими та промаркованими, а й через порівняння з описом. Так позу можна описати у вигляді множини кутів між суглобами. Головною проблемою цього методу є точність, він гарно працює коли людина стоїть так як наведено на рисунку вище, але коли вона починає рухатися у інших площинах точність результатів падає. Це відбувається через те, що для вирахування кута необхідно використовувати дев'ять значень по три координати трьох суглобів, кожна з цих координат матиме певні відхилення від реальних значень, а координата глибини і зовсім матиме приблизне значення. У результаті значення отримані цим методом використовувати доволі важко.

Наступним методом є метод комп'ютерного навчання, який дозволяє натренувати класифікатор поз. Однією з основних переваг цього методу є висока точність, так натренований на даних, що мали неточності він зможе адекватно трактувати подібні неточності у майбутньому. Головною ж проблемами є:

- його доволі важко натренувати бо для цього необхідні значні набори даних різноманітних поз;
- незрозуміло як впізнає ту чи іншу позу, що робить складним реалізацію підказок про те, що необхідно змінити людині, щоб поза стала правильною;
- важко додавати нові пози, для цього потрібно знову проводити тренування.

1.4 Огляд літератури

Для вирішення задачі ідентифікації поз необхідно порівнювати координати суглобів. З врахуванням того, що не усі суглоби приймають участь у русі та є важливими для його ідентифікації.

В існуючих роботах уже розглядаються різні підходи до пошуку найважливіших суглобів, рух яких є основним для опису дії. У роботі F.Ofli та ін.

[1] були визначені способи пошуку найбільш інформативних суглобів на основі кута суглоба, кутової швидкості суглоба, тощо.

T. Batabyal та ін. [2] спостерігали, що суглоби, які не беруть безпосередньої участі в послідовності дій (наприклад, гомілковостопний суглоб у випадку плескання долоні), мають стабільно низьку частоту зубчастих шумів. Було помічено, що міжсуглобові варіації з точки зору окремих координат (наприклад, зміна координати x одного суглоба відносно зміни координати z іншого суглоба під час руху), є більш важливими, ніж варіації положення суглобів загалом (наприклад, варіації положення лівого коліна щодо положення правого коліна).

Цей конкретний факт ігнорується в попередніх роботах, таких як [9] або [11].

Для вирішення проблеми шуму в роботах [12,13] пропонується використовувати вбудовування розрідженої матриці коваріації в лінійний колектор низької розмірності, а для класифікації дій використовувати SVM.

Більшість попередніх підходів використовують положення 3D хмари точок, представлення спільних кутів або їхні варіанти, зменшені PCA [14, 15, 16, 17].

Використання евклідових відстаней було детально розглянуте у роботі Forbes K та ін.[18, 19] Де вони проводили експерименти, у яких порівнювали однакові рухи людини зняті з різних ракурсів. Їх експерименти показали, що можливо досягти гарних результатів лише за умови, що обидва відео були зняті зі схожого кута, та рухи на них відрізнялися лише швидкістю.

Для свого метода вони використовували підходи для обробки інформації про рух, які до цього використовувалися лише для обробки звуку [20, 21].

Однією з проблем таких підходів є їх чутливість до деформацій пози, що може відбуватися в логічно пов'язаних рухах, а також чутливість до неточностей у даних.

Для ідентифікації вправ у відео важливо розуміти зі скількох поз вони складаються і де ці пози починаються та закінчуються. Для цього запропоновані алгоритми описані в [22, 23], вони дозволяють шукати локальні екстремуми графіку руху.

Методи розглянуті вище гарно працюють для двовимірних даних. У існуючих роботах можна побачити велику кількість досліджень на тему того, як з обмеженнями двовимірного простору досягти максимального результату. Наприклад, робота де пропонується метод для пошуку напрямку зору, який дозволяв краще зрозуміти у якій проекції до камери була знята людина [24].

Це не актуально при використанні провайдера даних, який надає результати у тривимірному просторі. У такому випадку на перший крок виходить не пошук способу скорегувати пропорції спотворені проекцію камери, а пошук підходів до коректного трактування інформації про z координату.

2 ПОСТАНОВКА ЗАДАЧІ

2.1 Постановка дослідницької задачі

Для створення мобільного застосунку, що може зможе оцінювати пози людей, необхідно розглянути технології та методи, що при цьому використовуються. Оцінка пози людини складається з двох частин:

- збір даних про положення суглобів людини у просторі;
- ідентифікація поз, у яких перебуває людина.

Виходячи з цього необхідно дослідити технології, що можуть збирати дані про рухи людини через камеру. У процесі дослідження кожен технологію необхідно перевірити за наступним критерієм:

- необхідні потужності;
- необхідна кількість камер;
- точність отриманих даних по кожній з осей;
- чутливість результатів до ракурсу зйомки;
- обмеження по апаратному забезпеченню, де технологія може бути запущена;

Наступним кроком буде дослідження методів з ідентифікації поз, вони будуть перевірятися за наступними критеріями:

- чутливість до неточностей у даних отриманих від технології захвату рухів.
- відсоток правильно розпізнаних поз;
- легкість додання поз, що будуть розшукуватися у записі;
- можливість ідентифікувати пози, що схожі на розшукувані навіть, якщо вони виконані неправильно;
- можливість надання інформації про те, чим саме ідентифікована поза не відповідає її правильному виконаному аналогу.

2.2 Формування вимог до експериментального додатку

У результаті виконання роботи має бути створений мобільний застосунок, що буде мати наступний функціонал:

- працювати на мобільному пристрої з камерою без застосування додаткового обладнання;
- підтримувати додання нових поз, що підлягають розпізнанню;
- створювати записи рухів, що були виконані людиною на камеру;
- відтворювати записи рухів за допомогою 3D моделі;
- ідентифікувати пози, що були присутні у русі, та повертати їх інформацію про них.

3 МЕТОД ІДЕНТИФІКАЦІЇ

3.1 Опис вирішуваної проблеми

Проблема полягає у тому, щоб знайти на відео, оцифрованих за допомогою ARKit, вправи та оцінити правильність їх виконання. Для оцінки правильності потрібен еталон, відносно якого відбуватиметься порівняння вправ – цим еталоном будуть вправи виконані тренером.

Рухи людини можна описати як вектор поз $\overline{M} = \{m_1, m_2, \dots, m_n\}$, де n це кількість кадрів у відео. Кожна поза m_i описується матрицею де колонки це координати по одній з осей, а строки це координати певного суглоба. Усього вона має 91 строку, що відповідає кількості суглобів, які може відстежувати ARKit.

$$m_i = \begin{pmatrix} x_{hips} & y_{hips} & z_{hips} \\ x_{shoulder} & y_{shoulder} & z_{shoulder} \\ \vdots & \vdots & \vdots \\ x_{wrist} & y_{wrist} & z_{wrist} \end{pmatrix}$$

Рух тренера також є вектором поз ідентичним за структурою до інших рухів, але кожна поза у ньому вважається правильною, назовемо його $\overline{E}$.

Для людини, що не є тренером, має бути знайдена множина ключових поз L така, що є достатнім мінімум для опису рухів.

$$f_1 : \overline{M} \rightarrow L, L \subset \overline{M}$$

Пози людини, що повторює вправи, на відміну від поз тренера, не є безумовно правильними. Тому серед усіх ключових поз необхідно знайти правильні пози CL . У випадку коли усі пози виконані правильно, множина CL може бути ідентичною множині L .

$$f_2 : L \rightarrow CL, CL \subseteq L$$

Для тренера у множині $\bar{E}$ також обирається множина ключових поз, так як усі пози тренер виконує правильно, позначимо її як L_{ref} .

Необхідно знайти відсоток вірних поз Q , що виконує людина відносно до поз тренера L_{ref} .

$$Q = \frac{|L|}{|L_{ref}|}$$

Основними вимогами до методу є:

- нечутливість до неточностей у даних;
- можливість компенсувати різницю у швидкості між порівнюваними відео.

Якщо вищезазначені вимоги не виконуються, система не може розглядатися як якісна і потребує вдосконалення або пошуку нових підходів до вирішення проблеми.

3.2 Корегування положень суглобів

Інформація про позу тіла людини повертається у вигляді безлічі суглобів. Ця інформація може містити суттєві помилки і повинна бути оброблена для усунення неточностей для подальшого аналізу.

Помилки, що присутні у результатах систем захвату рухів частіше за все виглядають як невірна відстань між суглобами людини. Такі проблема викликані неточністю у даних про z координату. Незважаючи на те, що відстань неправильна напрямком вектору, що сполучає суглоби вірний. Так система може повернути

координати кисті руки на відстані від ліктя 40см у людини з променевою кісткою довжиною 30см.

Для того, щоб замінити завідомо невірні координати суглобів тіло людини має бути виміряне у позах де кількість помилок при вимірюванні відстаней буде мінімальною. Вимір кісток відбувається через пошук довжини вектору між парами суглобів за формулою:

$$|\vec{a}| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Враховуючи, що помилки як вже наводилося раніше виникають через невірну оцінку глибини, то поза яка буде використовуватися для вимірювання довжин кісток має бути двовимірною. Прикладом такої пози є рівна стійка з руками вздовж тіла.

Для корегування позицій суглобів ми повинні змінити їх положення так, щоб відстань між ними збільшилася, а напрямок зберігся (див. рис. 5). Неможливо це зробити для кожної пари окремо бо корегування однієї пари суглобів змінить положення одного з них у іншій парі (див. рис. 6). Як результат положення кожного наступного суглобу необхідно корегувати відносно положення попереднього.

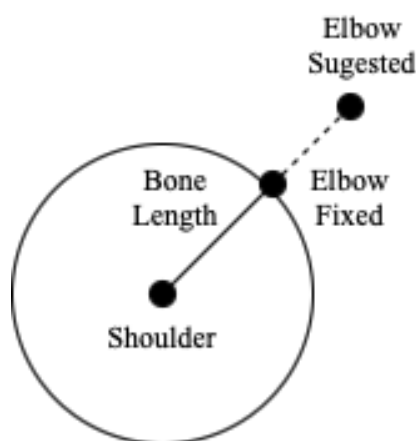


Рисунок 5 – Корекція позиції рухливого суглоба

Для вирішення цієї проблеми тіло повинно бути представлене у вигляді ієрархічної структури (див. рис. 7), яка з'єднує всі суглоби між собою. Дерево

суглобів повинно мати корінь, яким виступатиме певний суглоб. Коренем були обрані стегна, основними причинами стало їх позиціювання по середині тіла, а також те, що відносно них надаються усі координати у бібліотеці ARKit.

Не для всіх з'єднань суглобів це слід робити. Наприклад, хребет складається з великої кількості дрібних суглобів, відстань між якими важко визначити, враховуючи той факт, що людина може сутулитися.

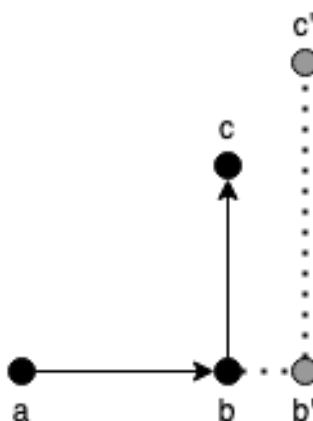


Рисунок 6 – Каскадний ефект після зміни позиції одного суглоба

Це неможливо побачити у багатьох проекціях, крім того, хребці можуть розтягуватися і стискатися, що робить відстань між ними нестабільною величиною.

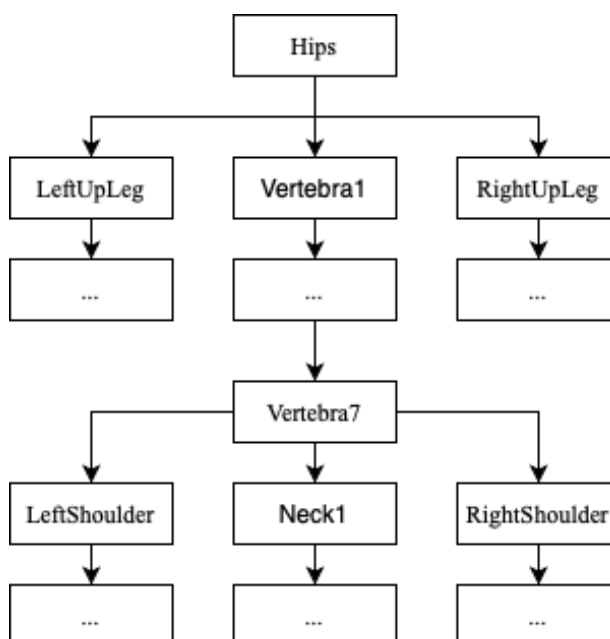


Рисунок 7 – Дерево суглобів

Натомість є суглоби, відстань між якими досить велика і стабільна, а також видима з більшості ракурсів. Наприклад, відстань між плечем і ліктем, а також між ліктем і зап'ястям.

Хоча може здатися, що після корегування позиціювання суглобів стане більш відповідним до позиції людини у реальності це не зовсім так, насправді бувають випадки коли отримані позиції, відрізняються навіть більше ніж ті, що були спочатку. Головною метою корегування є отримання теоретично можливої пози людини, без цього ідентифікація пози значно ускладнюється бо ми не можемо сподіватися на логічні висновки, що до тіла людини(наприклад, коліно не може бути зігнуте уперед).

3.3 Усереднення значень та фільтрація

Кожен варіант координат на наступному кадрі відео є новою позою m_i , але більшість із цих поз не представляють інтересу, оскільки вони не є визначальними для руху $\overline{M}$. Ключовими позами руху L є локальні екстремуми руху, але знайти їх важко, оскільки, хоча викиди відфільтровані, діаграма все ще має зубчатий вигляд (див. рис. 8).

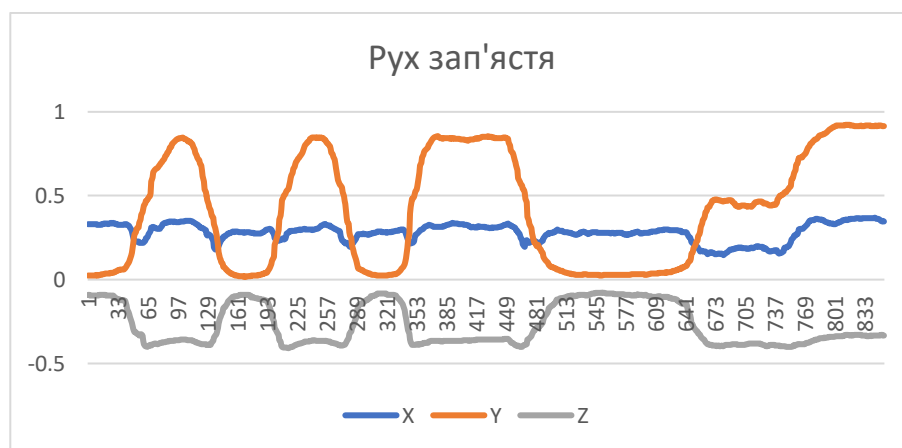


Рисунок 8 - Рух зап'ястя в трьох напрямках. Вісь x - відстань у метрах. вісь y – номер кадру.

Поява цих зубців зумовлена неточністю розпізнавання скелета. У різних постачальників даних вона може бути різною, для ARKit вона не перевищує 0,05 м.

Для того, що позбутися від зубців було вирішено використовувати алгоритм усереднення значень із заданим кроком, який дорівнює помилці постачальника даних s . Алгоритм усереднення застосовується окремо для руху вздовж кожної осі. Його сенс полягає в тому, щоб пройти всі точки по кожній з осей і послідовно об'єднати точки в групи (див. рис. 9). В групі різниця між максимальним і мінімальним значенням не має перевищувати значення похибки.

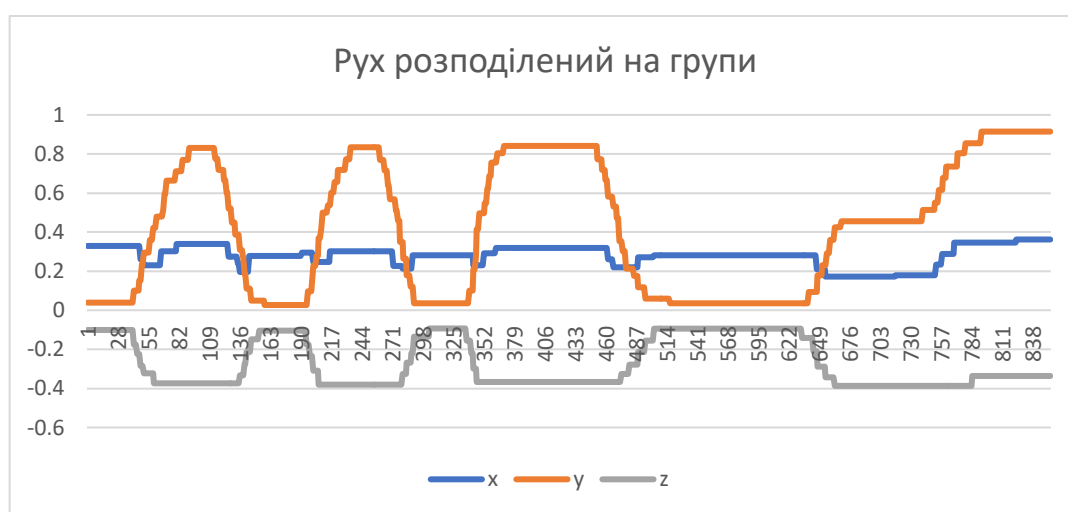


Рисунок 9 – координати по кожній осі об'єднані у групи

Після поділу графіка на групи можна побачити, що він тепер складається з великої кількості невеликих груп. Не всі групи мають значення, відбираються лише ті групи де:

- обидва сусіди нижче чи обидва сусіди вище;
- їх довжина перевищує задане значення l .

Друга умова необхідна, якщо ми хочемо зберегти інформацію про тривалі паузи, які були впродовж руху. Якщо значення занадто мале, то ми залишимо багато проміжних значень, і порівняння пройдуть тільки однакові записи. Значення решти груп зводиться до їх середнього значення (див. рис. 10), на цьому графіку групи це рівні лінії паралельні осі x .



Рисунок 10 – Результати підготовки даних про рух кисті для подальшого аналізу

Отримані групи є ключовими моментами відео. В теорії з кожної групи можна узяти випадковий кадр та назвати його ключовою позою. На практиці краще брати той кадр який уходить до максимальної кількості груп, тобто у момент коли максимальна кількість суглобів змінювала напрям руху.

3.4 Ідентифікація поз

Для пошуку поз запропонований метод ідентифікації, який працює наступним чином: кожна поза може бути формально описана як взаємне розташування суглобів. Наприклад, права рука вище правого плеча, ліва рука нижче лівого плеча, тощо.

Кожне твердження має складатися з трьох складових:

- назва одного суглоба;
- операція порівняння;
- назва іншого суглоба.

Використовуються наступні операції порівняння:

- вище;
- нижче;

- лівіше;
- правіше;
- позаду;
- попереду.

Вони описують відносне положення по кожній з осей. Важливо зауважити, що серед операцій немає операцій рівності. Це викликано тим, що в них ніколи не можна бути впевненим через неточність даних. Якщо ж ввести проміжок, який буде вважатися можливим відхиленням, при якому ми продовжуємо вважати, що значення рівні то він буде занадто великим. У разі використання ARKit можливе відхилення складає 5см. Тобто потрібно б було вважати, що два значення рівні коли їх відхилення складає 10 см, а це вже більше схоже на усвідомлений рух.

Таких тверджень може бути будь яка кількість, але необхідно пам'ятати, що бажано не робити їх надто багато. Може статися, що більшість людей не зможуть пройти цей тест через специфіку їх виконання, яка при цьому не порушає ідею вправи.

Для більш наглядної демонстрації правил їх можна візуалізувати у декількох проекціях. Для вправи присідання достатньо представлення у двох з трьох можливих проекцій. На візуалізації видно, що присідання має 4 правила для осі аплікату, та 3 правила для осі ординат (див. рис 10) та 1 правило по осі абсцисс (див рис 11).

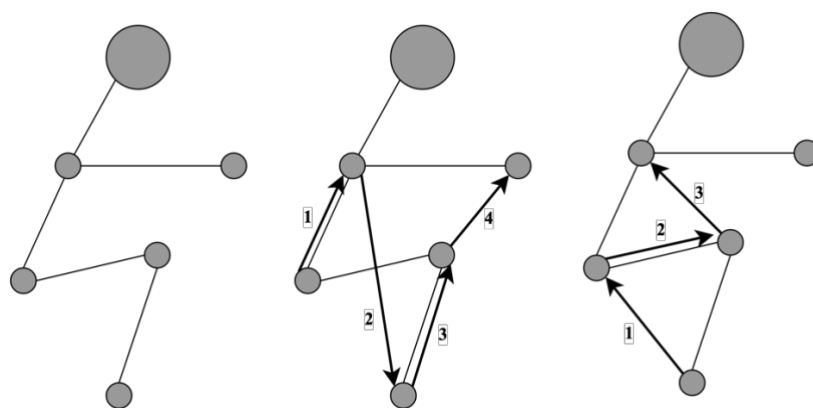


Рисунок 11 – Зображення правил для присяду по осях ординат та аплікату

На рисунках можна побачити, що правил могло б бути і більше. Не всі правила додані усвідомлено. Доданих правил достатньо, щоб у реальному оточенні тільки присідання пройшло цей тест.

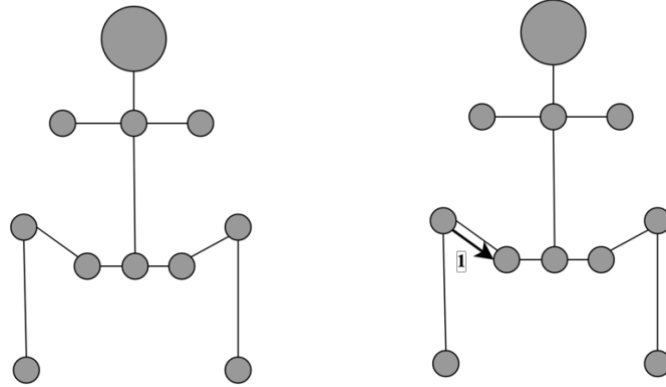


Рисунок 12 – Зображення правил для присяду по осі абсцис

Описати позу, наприклад, присідання T_{squat} можна як кон'юнкцію тверджень:

Метод ідентифікації не дає точної вказівки на близькість двох поз одна до одної. Хоча ми можемо порівняти кількість тверджень, що задовольняє одна поза і не задовольняє інша. Головне, що метод дозволяє стверджувати, що пози однакові, а також він дуже стійкий до помилок у даних.

$$T_{squat} = f(p_i) = (x_{wrist} > x_{shoulder}) \wedge (y_{knee} > y_{hips}) \wedge (y_{hips} > y_{foot}) \wedge \dots \wedge (z_{hips} > z_{knee})$$

Обробивши ключові пози цією функцією, ми можемо отримати множину правильно виконаних поз CL , які присутні у русі. Потім можна обробити відео с тренером та порівняти набір поз, отриманих з нього L_{ref} , з тим, який був знайдений до цього. Порівнявши ці значення ми можемо порахувати який відсоток Q поз відповідав позам тренера.

4 ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА МЕТОДА ІДЕНТИФІКАЦІЇ ПОЗ

4.1 Методологія проведення експериментів

Метою проведення експериментів є тестування працездатності запропонованого методу ідентифікації поз та порівняння його результатів з результатами інших методів.

Дослідження проводилось на пристроях iPhone XR, iPhone 12, iPhone 12 Pro з операційною системою iOS 14.2. Дані скелету були записані з використанням технології ARKit Motion Tracker. Інформація про скелет зберігалася як положення суглобів відносно кореня тіла.

Усі записи починаються з пози, коли людина стоїть з руками вздовж тіла, і закінчуються нею. Це необов'язкова умова для роботи методу, але вона дозволяє проводити більш точні тести.

Граничне значення помилки, що використовується в алгоритмі, встановлюється як 0,05м. Це значення було отримано експериментально, як максимальне значення помилки для ARKit. Паузи довжиною 24 кадри або більше вважаються значними паузами в русі, за умови, що швидкість запису становить 24 кадри в секунду.

У якості даних для тесту будуть використані записи 3 базових вправ: присідання, віджимання, згинання. Кожна вправа буде записана у декількох варіантах: правильний, прискорений, частково неправильний, повністю неправильний. Правильне відео також буде записано під різними кутами до обличчя людини з кроком 45 градусів. Вправи матимуть десять повторів.

4.2 Результати тестів

Кожна вправа складалася з двох поз, загалом на кожному записі вправа повторювалася 10 разів, ідеальним результатом роботи алгоритму буде розпізнання 20 поз. Після перевірки методу на відео, де всі вправи були виконані правильно під різними кутами, були отримані наступні значення для кожного кута (Табл. 1).

Таблиця 1 – Відсоток поз, що були розпізнанні на записі при різни кутах зйомки

Кут\Вправа	Присідання	Віджимання	Нахил
0	90%	85%	95%
45	100%	95%	100%
90	100%	100%	100%
135	95%	100%	100%
180	90%	85%	90%

Далі йде перевірка на помилково позитивні та помилково негативні результати, для цього записується 4 відео:

- вправи виконані правильно;
- вправи виконані дуже швидко;
- вправи виконані частково неправильно (не усі вимоги пози були виконані);
- вправи не виконувалися, але були присутні рухи, що є схожі на необхідні вправи.

У результаті були вираховані відсотки розпізнаних поз, для кожної вправи з різноманітними її виконаннями.

Таблиця 2 – Відсоток поз, що були розпізнані при виконанні вправ з різним рівнем правильності

	Присідання	Віджимання	Нахили
правильне	100%	100%	100%
пришвидшене	95%	90%	100%
частково правильне	0%	5%	0%
неправильне	0%	0%	0%

У наступному експерименті відбулося вимірювання показників схожості між чотирьома відео, де виконувалися одні й ті самі вправи. Перше відео взялося як еталон який зняв тренер, а інші були з ним порівнянні. Результат у таблиці 3

Таблиця 3 – Порівняння схожості виконання вправ

	Присідання, %	Віджимання, %	Нахили, %
1	85	90	95
2	85	95	90
3	90	85	85

У наступних випробуваннях відбулась ідентифікація тих самих вправ, але за допомогою інших існуючих методів ідентифікації.

Таблиця 4 – Порівняння схожості виконання вправ, за допомогою методу порівняння кутів

	Присідання, %	Віджимання, %	Нахили, %
1	32	90	86
2	45	78	81
3	48	86	78

Таблиця 5 – Порівняння схожості виконання вправ, за допомогою методу інтервалів

	Присідання, %	Віджимання, %	Нахили, %
1	40	82	84
2	55	72	82
3	44	73	79

У всіх експериментах, крім першого, порівнювалися відео записані під кутом 90 градусів до обличчя людини.

4.3 Аналіз результатів

В результаті випробувань було доведено, що для отримання найбільш точного результату потрібно знімати у площинах, де відбуваються найбільш суттєві зміни. Це пов'язано з тим, що постачальнику даних важко оцінити глибину зображення, і іноді він повертає абсолютно неправильні дані, які неможливо виправити, навіть використовуючи підходи, вбудовані в метод. Однак метод в цілому демонструє високу стійкість до неточностей.

Цим викликані показники з першого тесту у яких видно, що найкращі результати були досягнені при зйомці під кутом 90 градусів. Якщо дивитися з цього ракурсу, то майже усі зміни відбуваються без залучання осі z.

Перш ніж застосовувати метод ідентифікації, необхідно обрати правильний поріг неточності наданих даних, а також вибрати значення довжини для інтервалів, які не є екстремумами і не повинні ігноруватися. У разі неправильного вибору цих параметрів можна отримати два крайні випадки: знайдено забагато проміжних поз або пропущено забагато ключових поз.

Крім того, цей метод досить добре справляється з визначенням подібності рухів у відеоролику в цілому та у кожному русі окремо. Так, при значній зміні швидкості рухів, метод розпізнає більше 90% поз.

Недоліком високого відсотка розпізнавання є наявність помилково позитивних результатів. Вони виникають внаслідок того, що людина може виконувати абсолютно різні рухи, попадаючи при цьому у надзвичайно схожі пози. Вірогідність такого збігу дуже мала, бо для цього має зійтися надто багато факторів.

Порівняно з іншими методами ідентифікації, розроблений метод має майже однакові результати у випадку коли у даних мінімальна кількість помилок, і впізнає приблизно на 30% більше поз у випадках коли дані насичені помилками.

5 ПРОГРАМНА РЕАЛІЗАЦІЯ

5.1 UML-моделювання

Use Case діаграма показує основні можливості, що мають користувачі мобільного застосунку (див. рис. 13).

Користувачеві доступний наступний функціонал:

- зняти відео;
- отримати інформацію про присутні у ньому пози;
- перевірити їх правильність;
- програти відео.

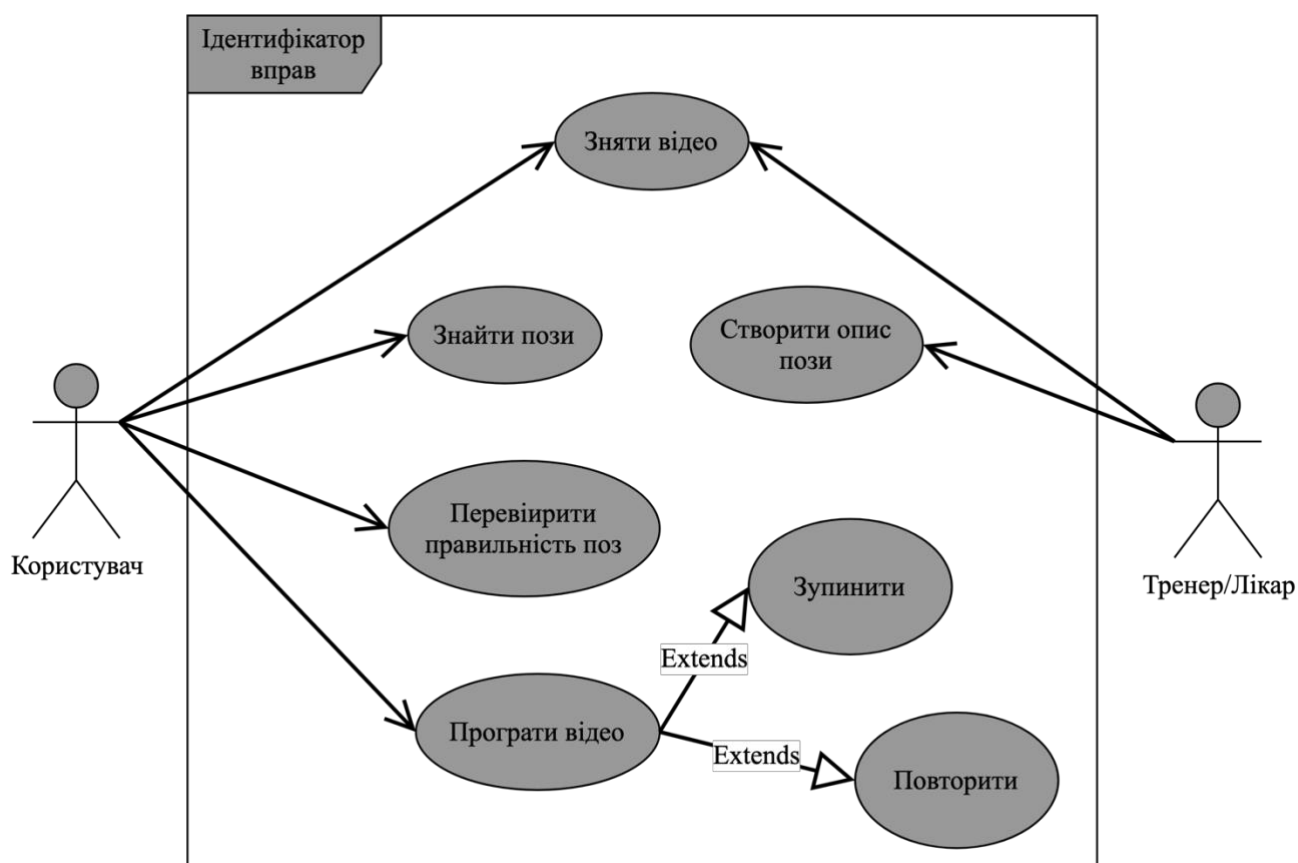


Рисунок 13 – Use Case діаграма мобільного застосунку

Тренеру/Лікару доступний наступний функціонал:

- зняти відео;

– додати опис пози.

Для того, щоб мати змогу не тільки ідентифікувати пози, що є повністю правильними, а також мати змогу знайти ті, що виконані з помилками процес ідентифікації є ітеративним. На кожному наступному кроці вимоги до пози зменшуються (див. рис. 14).

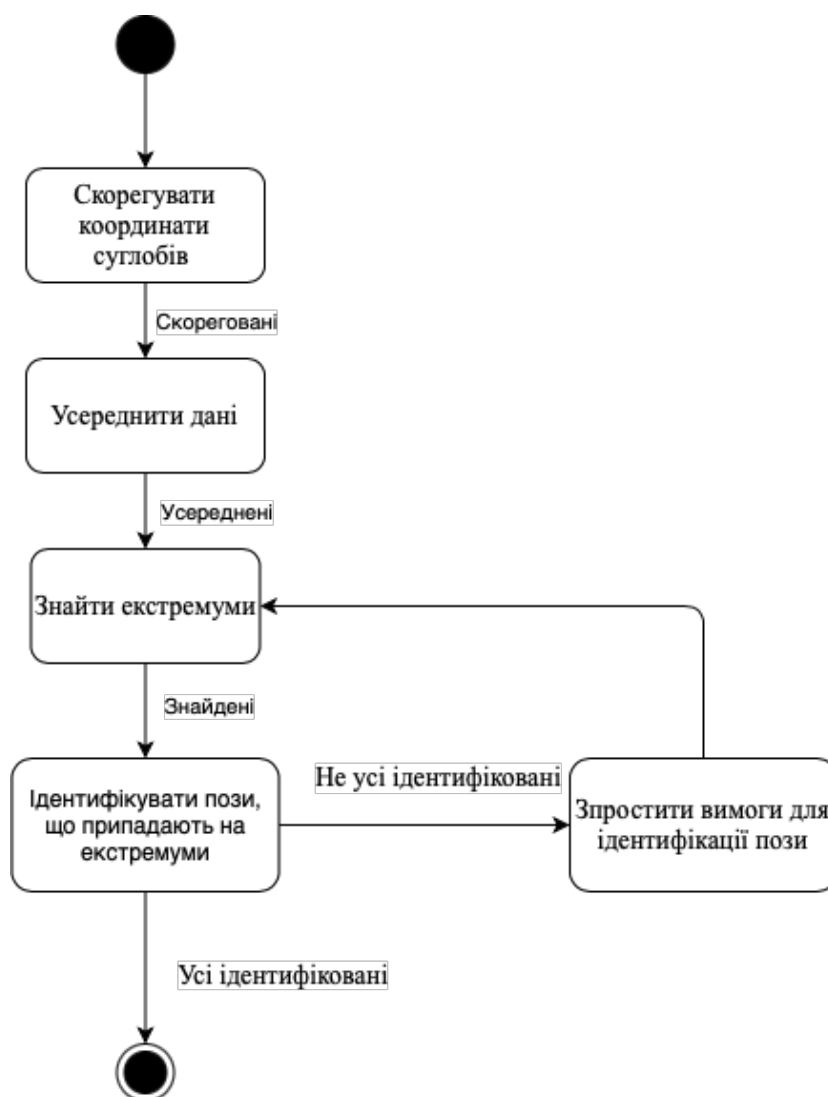


Рисунок 14 – Діаграма активності

Такий підхід дозволяє з певною вірогідністю стверджувати, що користувач намагався відтворити позу, але зробив це з помилками.

5.2 Загальний опис

Програмний застосунок створено з двох частин:

- рекордер;
- аналізатор.

Рекордер необхідний для того щоб отримати сирі дані про рухи людини з камери смартфона. Для написання рекордера розглядалися різні шляхи:

- використання смартфонів як ір камери задля отримання зображення з декількох ракурсів та подальша обробка зображення на сервері;
- об'єднання смартфонів у мережу де є головний та допоміжні, усі передають дані головному, який з обробляє отримані дані;
- використання лише одного смартфона та відразу обробляти на ньому.

Перший та другий метод мали теоретичні переваги так як надавали можливість надавати більш точні координати за рахунок того, що відео зняте з декількох ракурсів. Однак, на практиці їх важко реалізувати. Основною проблемою для обох є затримка при передачі даних, так проблеми синхронізації.

Тому для написання рекордеру був обраний 3 варіант, він має гірші показники у точності, але не потребує передачі даних. Існує не так багато технологій, що здатні розпізнавати тривимірні координати за допомогою лише однієї камери. Серед протестованих технологій прийнятна точність розпізнавання досягалася лише за допомогою ARKit Motion Capture. Оцифровування рухів людини не є основною задачею ARKit, але він напрочуд гарно з цим пореється.

ARKit можна використовувати у різних оточеннях:

- iOS SDK;
- Unity.

ARKit окрім стандартного підходу до використання через iOS SDK може використовуватися через ігровий движок Unity. У Таблиці 6 наведені можливості доступні при використанні кожного з підходів.

Таблиця 6 – Переваги та недоліки оточень для ARKit

Функції	ARKit + iOS SDK	ARKit + Unity
Додання 3D моделі без написання коду	-	+
Запуска програми на заздалегідь записаному відео	+	-
Перезапуск коду без перебудови програми	-	+
Можливість проводити налаштування через інтерфейс без використання коду	-	+
Стрімінг відео з телефона відразу на комп'ютер	-	+
Наявність детальної документації	-	+

Зважаючи наведені вище переваги та недоліки стає очевидним, що розробляти застосування з використанням ARKit швидше та простіше з використанням Unity.

Написаний рекордер після запису відео зберігає його до сховища пристрою. Перед цим серіалізує дані про позу у бінарному вигляді, цей підхід обраний через значний розмір отриманих даних, який зростає на пів мегабайта за секунду навіть при використанні бінарного формату.

Аналізатор розпаковує збереження рекордером дані, після чого може:

- відтворити;
- знайти ключові пози;
- протестувати ключові пози, порівнявши з описами розшукуваних поз.

5.3 Зберігання опису пози

Для роботи методу ідентифікації пози повинні бути створені описи поз. Для цього був обраний формат Json через його гнучкість та можливість описати будь які дані.

Опис пози виглядає як масив правил для кожного суглобу, кожне правило складається з назви суглобу для якого воно приміряється, та масиву тверджень для кожної координати (див. рис. 13). Твердження складається з операції та назви суглобу з яким буде виконуватися порівняння.

```
{
  "joins": [
    {
      "name": "wrist",
      "X": {
        "rules": [
          {
            "type": "more",
            "target": "shoulder.X"
          }
        ]
      },
      "Y": {
        "rules": [
          {
            "type": "less",
            "target": "hips.Y"
          },
          {
            "type": "less",
            "target": "shoulder.Y"
          }
        ]
      },
      "Z": {
        "rules": []
      }
    },
    {
      "name": "shoulder"
```

Рисунок 15 – Опис правила у форматі json

5.4 Приклади коду

Код для відтворення пози, що була попередньо збережена

```
public void ApplyBodyPose(Record record, int frame){
    var rotations = record.rotations[frame];
    for (int i = 0; i < boneMapping.Length && i < rotations.Length;
++i){
        var bone = boneMapping[i];
        if (bone != null){
            bone.transform.localRotation = new
Quaternion(rotations[i].Item1, rotations[i].Item2,
rotations[i].Item3, rotations[i].Item4);
        }
    }
}
```

Код для пошуку груп під час усереднення значень графіку

```
List<Range> getRanges(List<double> values, double threshold) {
    final ranges = <Range>[];
    late int start;
    late double min;
    late double max;

    void initRange(int i) {
        start = i;
        min = values[i];
        max = values[i];
    }

    initRange(0);

    for (var i = 1; i < values.length; i++) {
        if (values[i] - min > threshold || values[i] - max < -threshold)
        {
            ranges.add(Range(values.sublist(start, i)));
            initRange(i);
        } else {
            if (values[i] > max) max = values[i];
            if (values[i] < min) min = values[i];
        }
    }

    ranges.add(Range(values.sublist(start, values.length)));
    return ranges;
}
```

ВИСНОВКИ

У роботі були проаналізовані існуючі підходи до вирішення задачі оцінки пози людини за допомогою мобільного пристрою. Для цього проводилося дослідження існуючих технологій аналізу руху та методів оцінки поз. У результаті для вирішення проблеми серед існуючих технологій була обрана ARKit MT, а серед існуючих методів для ідентифікації не було знайдено тих, щоб задовольняли усім умовам. Через це був запропонований новий метод до ідентифікації поз, що опирається на відносне позиціонування суглобів людини.

Запропонований метод ідентифікації показує достатні результати для практичних завдань, таких як допомога пацієнтам у процесі відновлення або людям, які займаються фітнесом. Враховуючи, що цей метод був запущений на основі даних, отриманих з телефонної камери, це відкриває широкі перспективи для його використання.

Перевагами методу є стійкість до неточностей даних, а також той факт, що він задає позу у зрозумілій для людини формі, це дозволяє проводити додатковий аналіз. Наприклад, зменшувати загальну кількість вимог до пози доки вона не буде ідентифікована, це дозволить визначити які саме умови не були виконані користувачем і, що він має скорегувати.

Поділ на пози який використовуються у методі, дозволяє шукати вправи за їх ключовими моментами. Для цього необхідно шукати схеми рухів, щоб, наприклад, знайти повторення того самого руху на відео.

З іншого боку, метод має деякі недоліки. Поза не буде виявлена, якщо її виконати трохи неправильно, для вирішення цієї проблеми є сенс спробувати поступове зменшення вимог до пози, доки вона не буде ідентифікована. Іншим недоліком методу є те, що навмисно його обдурити досить просто, наприклад, він не зможе розрізнити кругові та ромбоподібні рухи руками, оскільки він зосереджений на зміні векторів руху. Однак це не є важливим недоліком,

враховуючи зону використання вірогідність того, що користувач буде спеціально виконувати вправи неправильно незначна.

Однією з основних проблем, які були помічені під час експериментів, є неправильні результати, спричинені неправильними значеннями z координати. Хоча метод впорався на 30% краще за інші це досі важлива причина помилок. Невірну оцінку глибини досить складно виправити, але є ймовірність того, що в майбутньому, завдяки ширшому використанню лідарів у телефонах, можна буде уникнути неточностей.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Zanfir M. et al. Human synthesis and scene compositing //Proceedings of the AAAI Conference on Artificial Intelligence. – 2020. – Т. 34. – №. 07. – С. 12749-12756.
2. Bogu F. et al. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image //European conference on computer vision. – Springer, Cham, 2016. – С. 561-578.
3. Brau E., Jiang H. 3d human pose estimation via deep learning from 2d annotations //2016 Fourth International Conference on 3D Vision (3DV). – IEEE, 2016. – С. 582-591.
4. Chen C. H., Ramanan D. 3d human pose estimation= 2d pose estimation+ matching //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2017. – С. 7035-7043.
5. Collobert R., Puhersch C., Synnaeve G. Wav2letter: an end-to-end convnet-based speech recognition system //arXiv preprint arXiv:1609.03193. – 2016.
6. Maidi M., Lehiani Y., Preda M. Open Augmented Reality System For Mobile Markerless Tracking //2020 IEEE International Conference on Image Processing (ICIP). – IEEE, 2020. – С. 2591-2595.
7. Pourdamghani N., Rabiee H. R., Zolfaghari M. Metric learning for graph based semi-supervised human pose estimation //Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). – IEEE, 2012. – С. 3386-3389.
8. Thar M. C., Winn K. Z. N., Funabiki N. A Proposal of Yoga Pose Assessment Method Using Pose Detection for Self-Learning //2019 International Conference on Advanced Information Technologies (ICAIT). – IEEE, 2019. – С. 137-142.
9. Ofli F. et al. Sequence of the most informative joints (smij): A new representation for human skeletal action recognition //Journal of Visual Communication and Image Representation. – 2014. – Т. 25. – №. 1. – С. 24-38.

10. Batabyal T., Chattopadhyay T., Mukherjee D. P. Action recognition using joint coordinates of 3d skeleton data //2015 IEEE International Conference on Image Processing (ICIP). – IEEE, 2015. – C. 4107-4111.
11. Yang X., Tian Y. L. Eigenjoints-based action recognition using naive-bayes-nearest-neighbor //2012 IEEE computer society conference on computer vision and pattern recognition workshops. – IEEE, 2012. – C. 14-19.
12. Chiu C. Y. et al. Content-based retrieval for human motion data //Journal of visual communication and image representation. – 2004. – T. 15. – №. 3. – C. 446-466.
13. Kovar L., Gleicher M. Automated extraction and parameterization of motions in large data sets //ACM Transactions on Graphics (ToG). – 2004. – T. 23. – №. 3. – C. 559-568.
14. Keogh E. et al. Indexing large human-motion databases //Proceedings of the Thirtieth international conference on Very large data bases-Volume 30. – 2004. – C. 780-791.
15. Sakamoto Y., Kuriyama S., Kaneko T. Motion map: image-based retrieval and segmentation of motion data //Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation. – 2004. – C. 259-266.
16. Forbes K., Fiume E. An efficient search algorithm for motion data using weighted PCA //Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation. – 2005. – C. 67-76.
17. Hsu E., Pulli K., Popović J. Style translation for human motion //ACM SIGGRAPH 2005 Papers. – 2005. – C. 1082-1089.
18. Liu G. et al. A system for analyzing and indexing human-motion databases //Proceedings of the 2005 ACM SIGMOD international conference on Management of data. – 2005. – C. 924-926.
19. Müller M., Röder T., Clausen M. Efficient content-based retrieval of motion capture data //ACM SIGGRAPH 2005 Papers. – 2005. – C. 677-685.
20. Picone J. Fundamentals of speech recognition: A short course //Institute for Signal and Information Processing, Mississippi State University. – 1996.

21. Kovar L., Gleicher M. Flexible automatic motion blending with registration curves //Symposium on Computer Animation. – 2003. – T. 2.
22. Shcherbakova G. et al. Adjustement of wavelet function parameters for analysis of non-stationary periodic signals with multistart optimization //2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). – IEEE, 2017. – C. 110-112.
23. Hramm O., Bilous N., Ahebian I. Configurable Cell Segmentation Solution Using Hough Circles Transform and Watershed Algorithm //2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL). – IEEE, 2019. – C. 602-605.
24. Rakova A. O., Bilous N. V. REFERENCE POINTS METHOD FOR HUMAN HEAD MOVEMENTS TRACKING //Radio Electronics, Computer Science, Control. – 2020. – №. 3. – C. 121-128.