

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет комп'ютерної інженерії та управління
(повна назва)

Кафедра електронних обчислювальних машин
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Методи інтерпретації рухів людини в системі
віддаленого керування елементами
розумного будинку
(тема)

Виконав:

студент II курсу, групи СПМ-22-5
Вінтонович М.С.
(прізвище, ініціали)

Спеціальність 123 «Комп'ютерна інженерія»
(код і повна назва спеціальності)

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

Освітня програма Системне програмування
(повна назва освітньої програми)

Керівник: доц. Барковська О.Ю.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ЕОМ

Коваленко А.А.
(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерної інженерії та управління
Кафедра _____ електронних обчислювальних машин
Рівень вищої освіти _____ другий (магістерський)
Спеціальність _____ 123 «Комп'ютерна інженерія»
(код і повна назва)
Тип програми _____ освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма _____ Системне програмування
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

“ _____ ” _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту _____ Вінтоновичу Микиті Сергійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Методи інтерпретації рухів людини в системі віддаленого керування
елементами розумного будинку

затверджена наказом по університету від “ 01 ” квітня 2024 р. № 257Ст

2. Термін подання студентом роботи до екзаменаційної комісії _____ 15 червня 2024 р.

3. Вхідні дані до роботи _____

операційна система iOS від 11.0 iPhone/iPad, macOS від 10.13.

бібліотеки та фреймворки – CoreML, Vision, ARKit

інструменти для розробки моделей - Create ML, TensorFlow, PyTorch

4. Перелік питань, що потрібно опрацювати у роботі _____

Аналіз методів детектування та розпізнавання динамічних об'єктів

Створення функціональної моделі системи

Розробка методології проведення досліджень

Проведення експериментів

Аналіз отриманих результатів

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) 15слайдів

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

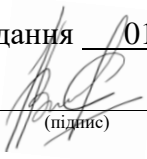
Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналіз моделей детектування та розпізнавання динамічних об'єктів	02.04.24-08.04.24	
2	Огляд систем аналізу жестів	09.04.24-16.04.24	
3	Створення функціональної моделі системи	17.04.24-22.04.24	
4	Розробка методології проведення досліджень	23.04.24-06.05.24	
5	Проведення експериментів	07.05.24-23.05.24	
6	Оформлення матеріалів кваліфікаційної роботи	24.05.24-03.06.24	
7	Подання кваліфікаційної роботи керівникові та її попередній захист	04.06.24-07.06.24	
8	Подання кваліфікаційної роботи на рецензування	08.06.24-12.06.24	

Дата видачі завдання 01 квітня 2024 р.

Студент



(підпис)

Керівник роботи



(підпис)

доц.Барковська О.Ю.

(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 62 с., 12 рис., 12 табл., 1 дод., 17 джерел.

МОДЕЛЬ, ДЕТЕКТУВАННЯ, РОЗПІЗНАВАННЯ, ЖЕСТИ, РУКИ, iOS.

Метою роботи є вдосконалення методів інтерпретації рухів людини для управління різними елементами розумного будинку через віддалене керування. Конкретно, мета полягає в створенні системи, яка забезпечує натуральну та ефективну взаємодію людини з розумним будинком шляхом інтерпретації рухів користувача.

В рамках роботи були вирішені наступні задачі – розглянуто сенсорні пристрої для реєстрації рухів людини, проаналізовано технології детектування та розпізнавання жестів руки, створено функціональну модель системи розпізнавання жестів, в якій головне дослідження зосереджено на модулі нейромережевого аналізу покращених вхідних даних. Результати проведених досліджень з вдосконалення базових нейромережевих моделей та використання гібридного підходу показали, що досягти найвищої точності розпізнавання жестів рук (розкрита долоня та стиснутий кулак) можливо на 6 шарах при 50 епохах навчання, досягаючи точності в 97% із тестовими стратами у 0,17. Дослідження щодо розташування гаджетів відносно досліджуваного об'єкту показало, що діагональне розташування гаджетів із попередньо визначеними технологіями має перевагу перед використанням монотехнології, а також перед використанням паралельного розташування гаджетів із однаковими технологіями на 0.62%.

Вирішення цих задач дозволить досягти поставленої мети і розробити систему, яка буде забезпечувати натуральну та ефективну взаємодію людини з розумним будинком через віддалене керування.

ABSTRACT

Master's thesis: 62 pages, 12 figures, 12 tables, 1 appendices, 17 sources.

MODEL, DETECTION, RECOGNITION, GESTURES, HANDS, iOS.

The aim of the work is to improve methods of interpreting human movements for controlling various elements of a smart home through remote management. Specifically, the goal is to create a system that provides natural and efficient interaction between a person and a smart home by interpreting the user's movements.

Within the scope of the work, the following tasks were accomplished: sensor devices for recording human movements were reviewed, hand gesture detection and recognition technologies were analyzed, and a functional model of a gesture recognition system was created, with the main focus on the neural network analysis module of enhanced input data. The results of the research on improving basic neural network models and using a hybrid approach showed that achieving the highest accuracy in hand gesture recognition (open palm and clenched fist) is possible with 6 layers and 50 training epochs, reaching an accuracy of 97% with a test loss of 0.17. Research on the placement of gadgets relative to the studied object showed that diagonal placement of gadgets with predefined technologies has an advantage over using a mono-technology and also over using parallel placement of gadgets with identical technologies by 0.62%.

Solving these tasks will allow achieving the set goal and developing a system that ensures natural and efficient interaction between a person and a smart home through remote management.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	8
ВСТУП	9
1 ОГЛЯД ПРОБЛЕМНОЇ ОБЛАСТІ	11
1.1 Розширене обґрунтування актуальності теми роботи.....	11
1.2 Огляд існуючих систем віддаленого керування елементами розумного будинку.....	14
1.2.1 Метод на основі отримання зображень	15
1.2.2. Методи розпізнавання жестів, які не використовують зображення в якості вхідної інформації	17
1.3 Мета та задачі дослідження	19
2 АПАРАТНО-МЕТОДОЛОГІЧНЕ ПІДГРУНТЯ ДЛЯ ПРОВЕДЕННЯ ДОСЛІДЖЕННЯ	20
2.1 Огляд моделей та методів розпізнавання жестів та рухів людини	20
2.1.1 Традиційна архітектура системи інтерпретації рухів	22
2.2 Алгоритми машинного навчання для аналізу рухів	24
2.3 Порівняльний аналіз технічних засобів для збору та аналізу даних про рухи людини	27
2.4 Можливості обраної технології детектування жестів	28
2.5 Моделі та методи для розпізнавання жестів людини для керування елементами розумного будинку.....	30
3 РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ	34
3.1 Обґрунтування вибору технології детектування та розпізнавання жестів людини	34
3.2 Запропонована модель системи розпізнавання жестів	36
3.3 Проведення експериментальних досліджень	40

3.3.1 Експеримент 1. Тестування інструменту Create ML від Apple на власному наборі жестів.....	40
3.3.2 Експеримент 2. Тестування сервісу Microsoft Azure на власному наборі жестів.....	42
3.3.3.Експеримент 3. Тестування сервісу Google Cloud AI Platform на наборі визначених жестів	45
3.3.4 Експеримент 4. Тестування запропонованої гібридної системи, заснованої на двох обраних вище моделях	48
ВИСНОВКИ.....	50
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	52
ДОДАТОК А Графічний матеріал кваліфікаційної роботи.....	54

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ
І ТЕРМІНІВ

AI – штучний інтелект (англ., Artificial Intelligence)

ANN – штучна нейронна мережа (англ., Artificial Neural Network)

CNN – згорткова нейронна мережа (англ., Convolutional Neural Network)

DL – глибоке навчання (англ., Deep Learning)

IP – обробка зображень (англ., Image Processing)

IR – розпізнавання зображень (англ., Image Recognition,)

ML – машинне навчання (англ., Machine Learning)

RNN – рекурентна нейронна мережа (англ., Recurrent Neural Network)

ВСТУП

Жести – це різновид відео, які здатний бачити комп'ютер завдяки моделям машинного навчання. Для зору комп'ютера руки складний об'єкт, вони активно жестикулюють, можуть перекривати один одного, змінювати форму від розкритої долоні до кулака, схрещувати пальці. На руках немає активних точок, як, наприклад, очі та рот на обличчі, що заважає природному сприйняттю мови рухів.

Велике зацікавлення викликає концепція взаємодії між людьми та роботами. Замість того, щоб роботи замінили людей на робочих місцях, співпраця людини з роботами стала важливим напрямком. Вона не тільки зберігає робочі місця, але й створює можливість спільної праці. Це відкриває нові перспективи у вирішенні важливих завдань, звільняючи людей від рутинних та тяжких обов'язків. Однак, на сьогоднішній день, канали зв'язку між людьми та роботами залишаються обмеженими. Щоправда, технологія розпізнавання жестів успішно використовується як засіб комунікації між людиною та комп'ютером. У цьому контексті важливим є не лише огляд сучасних технологій розпізнавання жестів, але і розгляд можливостей їх застосування у взаємодії людини та робота. Такий підхід відкриває нові можливості у створенні моделей взаємодії, які б враховували потреби обох сторін.

У сучасному світі велике значення набуває вивчення взаємодії між технологіями та людьми. Однією з ключових тем цього дослідження є розпізнавання жестів. Жести, як складна форма візуального вираження, мають великий потенціал для взаємодії з комп'ютерними системами. Використання методів машинного навчання в цій області відкриває нові перспективи та можливості для створення інноваційних технологій. Зокрема, розпізнавання жестів дозволяє комп'ютерам розуміти та інтерпретувати рухи людських рук, що створює можливість для реалізації низки застосувань у різних галузях, від відеозв'язку до керування різноманітними пристроями.

Проте, розробка систем розпізнавання жестів викликає певні виклики та труднощі. Наприклад, для успішної ідентифікації та аналізу рухів необхідно мати надійні алгоритми детекції об'єктів у відеопотоці, які були б ефективними в реальному часі. Також важливо враховувати фактори, які можуть впливати на точність розпізнавання, такі як відстань від камери, освітлення та інші зовнішні умови. Тому, подальше вдосконалення та дослідження у цій області має велике значення для подальшого розвитку технологій взаємодії між людьми та комп'ютерами.

Задача розпізнавання жесту на відеопослідовності включає в себе декілька основних підзадач: локалізацію жесту на елементі відеопослідовності (пошук цільового об'єкта), виділення ключових ознак локалізованого жесту та порівняння цих ознак з еталонними значеннями з бази даних. Залежно від області застосування жестового інтерфейсу ці підзадачі можуть розв'язуватися по-різному для забезпечення вимог щодо продуктивності та точності розпізнавання жестів.

1 ОГЛЯД ПРОБЛЕМНОЇ ОБЛАСТІ

1.1 Розширене обґрунтування актуальності теми роботи

У сучасному світі, де технології стрімко розвиваються, системи розпізнавання жестів рук виявляються дуже перспективними. Вони є складовою частиною широкого спектру застосувань, від інтерфейсів користувача та віртуальної реальності до медичних систем і робототехніки. Розробка ефективних систем розпізнавання жестів рук має велике значення, оскільки вони можуть робити взаємодію з технологією більш доступною та природньою для людини. Системи розпізнавання жестів рук зазвичай складаються з трьох етапів: виявлення, відстеження і розпізнавання. Проте, якщо мова йде про систему керування жестами, то при такій взаємодії людини з машиною, кількість кроків буде доповнена фазою відгуку, тобто: збір даних, аналіз даних, ухвалення рішення, відгук (рисунок 1.1).



Рисунок 1.1 – Чотириступінчаста модель обробки відеоінформації при взаємодії людини із машиною

Фаза виявлення відповідає за сегментацію області руки, використовуючи глибину та колір. Після цього створюється модель руху руки, а фаза розпізнавання інтерпретує жести як команди. Варіанти можливих жестів навкedListо на рисунку 1.2.

Існують різноманітні підходи до реалізації систем розпізнавання жестів. Деякі дослідження використовують алгоритми граничного аналізу, інші – технології машинного навчання. У деяких роботах використовується комбінація різних методів для покращення точності і швидкості розпізнавання жестів.

Існуючі принципи обробки та аналізу зображень є наступними:

- сегментація – комп'ютер бачить зображення як сітку пікселів, кожен із яких має різні відтінки. Під час аналізу алгоритм сегментує їх виділення об'єктів, кордонів, ліній. Кожен піксель отримує свою мітку, якою належить до певної категорії;

- контурний аналіз – нейромережа розпізнає графічні об'єкти за їхніми контурами. Метод підходить для пошуку об'єктів заданої форми, але не дозволяє аналізувати вміст – внутрішні точки, що дають уявлення про глибину, колір та інші параметри;

- Template matching – принцип пошуку на зображенні ділянок, які відповідають заданому шаблону;

- Feature Detection and Matching – концепція створення абстракції зображення та порівняння її з оригіналом за ключовими точками для виявлення спільних характеристик;

- пошук за шаблонами – метод, що найбільш застосовується в CV в цілому. Він ґрунтується на математичній інтерпретації попиксельного збігу зображень. Методи, що працюють на основі нейромереж, використовують мільйони ознак та семантичних смислів, які мережа запам'ятала під час навчання та безперервно доповнює. Навчання багат шарових нейромереж призведе до того, що комп'ютерний зір зможе вирішувати складні завдання, бачити і точно аналізувати навіть незнайомі об'єкти.

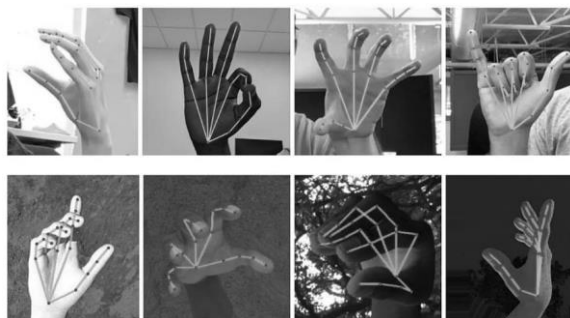


Рисунок 1.2 – Варіанти жестів рук, які можуть бути розпізнані системою віддаленого керування

Складність розпізнавання жестів для AI полягає в тому, що руки та пальці можуть загороджувати один одного, більше того, жестикуляція може бути занадто швидка і недостатньо чітка, через що розпізнавати їх точно та миттєво для алгоритмів дуже складно.

Для впровадження систем розпізнавання жестів у реальному часі важливо мати швидкі та ефективні алгоритми, оскільки вони зазвичай працюють в режимі реального часу. Такі системи можуть знайти широке застосування в сферах, де потрібна швидка та точна взаємодія з технологією.

Щоб прискорити процес, дослідники вирішили зменшити обсяг даних, які алгоритм має опрацювати. Наприклад, вони відмовилися від розпізнавання руки цілком: натомість від системи потрібно знайти долоню, після чого проаналізувати кожен із п'яти пальців. Окремий алгоритм фіксує 21 точку на отриманому зображенні, відстань між ними та їхнє взаємне розташування. Таким чином, створюється високоточна «карта» долоні та пальців. Водночас алгоритм Google може розпізнавати кілька рук.

Далі новий розпізнаний жест порівнюється з базою вже відомих жестів – від алфавіту та цифр мовою глухонімих до жестів, що символізують, наприклад, "рок" "світ", "ОК". Для цього дослідники створили датасет із 30 тисяч зображень з руками у різних жестах та під різним освітленням, на яких вручну позначили цю 21 координату.

Окрім того, апаратна реалізація алгоритмів обробки зображень може забезпечити більшу ефективність та надійність у порівнянні з програмними рішеннями. Застосування апаратних засобів також може дозволити значно знизити споживання енергії та збільшити швидкість обробки даних.

Отже, розробка та вдосконалення систем розпізнавання жестів рук має великий потенціал у різних галузях. Використання різноманітних методів та апаратних засобів може сприяти розвитку більш ефективних і точних систем, що відповідають потребам сучасного світу.

1.2 Огляд існуючих систем віддаленого керування елементами розумного будинку

Перед процесом розпізнавання жестів робочі дані мають бути зібрані датчиками. У цьому розділі датчики аналізуються з урахуванням різних технологій захоплення даних: засновані на зображеннях та не засновані на зображеннях (рисунок 1.3).

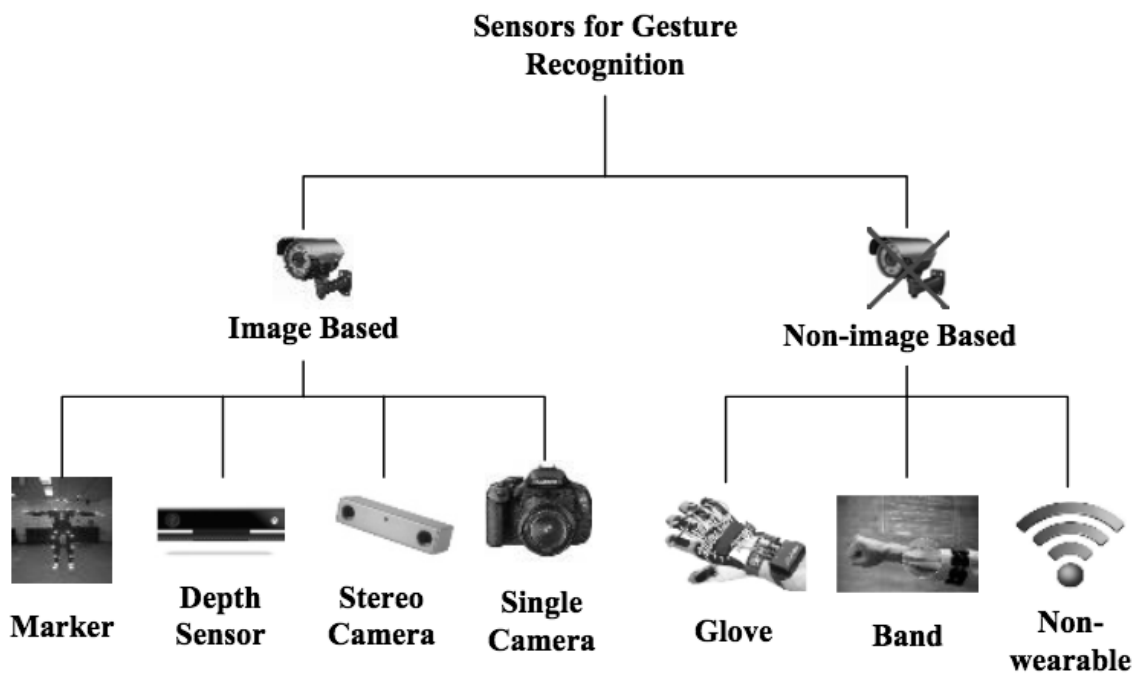


Рисунок 1.3 – Типи датчиків для розпізнавання жестів рук

Щоб розпізнати пальці та долоні, система повинна вивчити кілька точок у тривимірній проекції. Вона оцінює глибину, висоту і довжину кожної їх, зіставляє їх становище. По точках визначається позиція кожного пальця, формується їхнє векторне уявлення. На основі отриманих даних нейромережа класифікує виявлені закономірності, порівнює спостерігається жест зі своїми внутрішніми уявленнями різних жестів і робить висновок.

Якщо обробка та аналіз інформації відбуваються безпосередньо в камері, вона називається смарт-камерою. Звичайна камера передає сигнал комп'ютеру, де відбуваються обчислювальні дії.

1.2.1 Метод на основі отримання зображень

Технології часто надихаються природою. Ми використовуємо наші очі, щоб розпізнавати жести. Тому для роботів розумно використовувати камери, щоб "бачити" жести. Підходи, засновані на роботі із зображеннями, далі поділяються на чотири категорії.

1. У маркерному підході датчиком є оптична камера. У більшості рішень на основі маркерів користувачам необхідно мати видимі маркери. Сьогодні ми отримуємо набагато швидшу графічну обробку порівняно з двадцятьма роками раніше. В результаті на ринку є безліч датчиків розпізнавання жестів.

2. На початку 90-х дослідники почали аналізувати жести за допомогою одиночної камери. Недоліком однокамерного підходу є обмеження кута огляду, що впливає надійність системи. Однак у недавніх дослідженнях застосовувався однокамерний підхід до високошвидкісного розпізнавання жестів. Система використовує датчик швидкості та спеціально розроблений процесор візуальних обчислень для досягнення високошвидкісного розпізнавання жестів.

3. Щоб досягти надійного розпізнавання жестів, дослідники запропонували стереопідхід до створення 3D-зірки. Тут ми визначаємо стереокамерні підходи як програми, які використовують дві оптичні камери для створення інформації про глибину 3D. Багато стереопідходів до камер відповідали аналогічному процесу. Хоча системи стереокамер покращили стійкість у зовнішньому середовищі, вони, як і раніше, страждають від таких проблем, як складність обчислень та труднощі з калібруванням.

4. Останнім часом технології глибокого зондування швидко розвиваються. Ми визначаємо датчик глибини як монодатчик глибини. Моно-датчики глибини мають ряд переваг у порівнянні з традиційними стереокамерами. Наприклад, можна запобігти проблемам налаштування калібрування та умов освітлення. Крім того, вихідною інформацією датчика

глибини є інформація про глибину 3D. У порівнянні з інформацією про колір інформація про глибину 3D спрощує проблему ідентифікації жестів. Існує два типи загальних не стерео-датчиків глибини: камера з часом прольоту (ToF) і Microsoft Kinect (або аналогічні ІЧ-датчики) (рисунк 1.4).



Рисунк 1.4 – Приклади не стерео-датчиків глибини: а) камера з часом прольоту (ToF); б) Microsoft Kinect (або аналогічні ІЧ-датчики)

Основним принципом камер ToF є визначення часу проходження світла [19]. У різних публікаціях були запроваджені приклади розпізнавання жестів на основі ToF-камер [18, 20]. Перевагою камер ToF є вища частота кадрів. Обмеження камери ToF полягає в тому, що роздільна здатність камери сильно залежить від її світлочутливості та рефлексії [21].

Microsoft Kinect надає дешеве та просте рішення для розпізнавання жестів. Kinect – інфрачервоний датчик глибини. Аналогічними датчиками є ASUS Xtion Pro та Apple PrimeSense. Kinect має ІЧ-випромінювач, ІЧ-датчик та датчик кольору. Він широко використовується у сфері розваг, освіти та досліджень з великою спільнотою розробників [22-24]. Доступно безліч інструментів та проектів з відкритим вихідним кодом. Деякі дослідники реалізували системи розпізнавання жестів на основі Kinect на коротких відстанях [25-27]. Через обмежену роздільну здатність Kinect може

використовуватися для розпізнавання жестів тіла і розпізнавання жестів рук на невеликій відстані. Для розпізнавання жестів рук та кистей рук на відстані понад 2 метри краще використовувати інші підходи.

1.2.2. Методи розпізнавання жестів, які не використовують зображення в якості вхідної інформації

У розпізнаванні жестів протягом тривалого часу домінували датчики на основі зображень. Нещодавні розробки в MEMS та сенсорних технологіях значно покращили технології розпізнавання жестів, засновані не на зображеннях.

1. Жестові інтерфейси з урахуванням рукавичок також використовуються для розпізнавання жестів. Зазвичай методи на основі рукавичок вимагають провідного підключення акселерометрів та гіроскопів. Однак громіздка рукавичка з проводами може викликати проблеми у НРС. Підходи на основі рукавичок також мають складності у процедурах калібрування та налаштування.

2. Інша безконтактна технологія використовує детектори на браслетах. Сенсори встановлені на браслеті або аналогічних пристроях, що носяться. Сенсори на браслетах дозволяють використовувати бездротові технології та датчики електроміограми, що дозволяє уникнути підключення кабелів. Сенсори повинні контактувати із зап'ястям; руки та пальці користувача можуть бути вільними. Приклад браслетного датчика є пристрій Myo. Нещодавно було опубліковано кілька систем жестового керування на основі браслетів.

3. У третьому типі технологій, не пов'язаних із зображеннями, використовуються датчики, які не призначені для носіння. Безконтактні датчики можуть виявляти жести без контакту з тілом людини. Google представив Project Soli, систему радіолокаційного контролю та розпізнавання жестів на радіочастотному спектрі (Радар) [34]. Пристрій може розпізнавати

різні жести рук на невеликій відстані. Протягом багатьох років MIT є провідним новатором у галузі розпізнавання жестів. Технологія електричного поля була вперше розроблена у MIT [35]. Нещодавно Adib [36-38] з MIT представив систему WiTrack та RF-Capture, яка відстежує рух користувача за радіочастотними сигналами, відбитими від людського тіла. Система здатна захоплювати людські жести навіть з іншої кімнати через стіну з точністю до 20 см. Таким чином, технології, не придатні для носіння, є перспективними та швидкозрослими сенсорними технологіями для розпізнавання жестів.

Таблиця 1.1 містить порівняння різних сенсорних технологій. Показано переваги та недоліки різних підходів. Зрозуміло, що жоден датчик не підходить для всіх програм. На основі вищевказаних методів представлено два види сенсорних технологій.

Таблиця 1.1 – Переваги та недоліки різних сенсорних технологій

	Переваги	Недоліки
Маркери	Низьке обчислювальне навантаження	Маркери на тілі користувача
Камера	Простота встановлення	Низький рівень надійності
Стереокамера	Надійність	Складність обчислень, труднощі з калібруванням
ToF камера	Висока частота кадрів	Дозвіл залежить від потужності світла та відображення
Microsoft Kinect	Підтримка програмного забезпечення для розпізнавання жестів тіла	Не можна використовувати для розпізнавання жестів руки на відстані понад 2 метри
Рукавичка	Швидкість відгуку, точність трекінгу	Громіздкий пристрій із проводами
Браслет (сенсори на базі браслета, електроніка, що носитья)	Швидкість відгуку, область дії	Браслет повинен контактувати із людським тілом
Браслет (сенсори на базі браслета, електроніка, що носитья)	Швидкість відгуку, область дії	Браслет повинен контактувати із людським тілом
Безконтактні пристрої	Не вимагає контакту з тілом	Низька роздільна здатність, технологія недостатньо зріла

Сенсори для використання всередині приміщень: датчики глибини є найперспективнішими технологіями на основі зображень. Датчики глибини мають переваги простоти калібрування, встановлення та швидкості обробки даних. Існує велика спільнота розробників додатків, яка надає готові рішення.

Безконтактні сенсори є найперспективнішою технологією серед підходів, заснованих не на зображенні. Вони не потребують безпосереднього контакту з користувачами. Ця категорія сенсорів також є зростаючою галуззю на ринку технологій.

1.3 Мета та задачі дослідження

Метою дослідження є вдосконалення методів інтерпретації рухів людини для управління різними елементами розумного будинку через віддалене керування. Конкретно, мета полягає в створенні системи, яка забезпечує натуральну та ефективну взаємодію людини з розумним будинком шляхом інтерпретації рухів користувача.

Для досягнення поставленої мети необхідно розв'язати ряд задач:

- огляд сенсорних пристроїв для реєстрації рухів людини;
- аналіз технологій детектування та розпізнавання жестів руки;
- створення моделі системи розпізнавання жестів;
- розробка методології проведення досліджень з вдосконалення базових нейромережових моделей та використання гібридного підходу;
- аналіз отриманих результатів.

Вирішення цих задач дозволить досягти поставленої мети і розробити систему, яка буде забезпечувати натуральну та ефективну взаємодію людини з розумним будинком через віддалене керування.

2 АПАРАТНО-МЕТОДОЛОГІЧНЕ ПІДГРУНТЯ ДЛЯ ПРОВЕДЕННЯ ДОСЛІДЖЕННЯ

2.1 Огляд моделей та методів розпізнавання жестів та рухів людини

В сучасному світі інтелектуальних технологій існує ряд інноваційних рішень, спрямованих на поліпшення комфорту та ефективності використання побутових пристроїв. Зокрема, наявні системи управління, які реагують на жести та рухи людини з метою керування освітленням, температурою повітря та іншими аспектами домашнього комфорту (таблиця 2.1).

Деякі з цих систем використовують вбудовані в камери датчики руху та аналіз зображень для визначення певних жестів користувача. Наприклад, підняття руки може використовуватися як команда для включення освітлення, а зниження - для регулювання температури повітря. Інші системи можуть використовувати носимі пристрої з акселерометрами для відстеження рухів та жестів користувача. Цей підхід дозволяє більш точно визначати наміри користувача та швидко реагувати на них.

Застосування мікроконтролерів, таких як STM або Arduino, у таких системах дозволяє забезпечити швидку обробку даних та ефективне управління пристроями у реальному часі. Крім того, інтеграція з хмарними сервісами, такими як AWS, розширює можливості збереження та аналізу даних, що сприяє подальшому розвитку систем управління розумним будинком.

Розглянемо кілька прикладів існуючих моделей систем управління освітленням, температурою повітря та іншими побутовими пристроями з використанням жестів та рухів людини.

Однією з таких моделей є система, яка використовує камери з вбудованими датчиками руху. Ця система сприймає рухи та жести користувача перед камерою і аналізує їх, щоб визначити відповідні команди.

Наприклад, рух руки вгору може включити світло, а рух руки вниз - вимкнути його. Додатково, ця система може реагувати на інші жести, такі як покласти руки в бік або злегка покрутити їх, щоб змінити яскравість світла або температуру.

Ще однією моделлю є система, що використовує носимі пристрої з акселерометрами. Ці пристрої надають системі інформацію про рухи та жести користувача в реальному часі. Наприклад, якщо користувач підіймає руку, щоб змінити температуру повітря, акселерометр реєструє цей рух, і система автоматично змінює температуру відповідно до заданого сценарію.

Усі ці моделі використовують мікроконтролери для обробки даних та керування пристроями. Мікроконтролери, такі як STM або Arduino, забезпечують швидку та ефективну обробку даних у реальному часі, що робить їх ідеальним вибором для використання в таких системах. Крім того, інтеграція з хмарними сервісами, такими як AWS, дозволяє зберігати та аналізувати дані, що збираються системою, що розширює можливості та функціонал таких систем управління.

Таблиця 2.1 – Порівняльний аналіз підходів розпізнавання жестів рук при використанні різних сенсорів

Назва	Обладнання	Вхідні величини	Алгоритм	Переваги	Недоліки
1	2	3	4	5	6
Розпізнавання жестів за допомогою інфрачервоних сенсорів на основі Intel RealSense	Пристрій Intel RealSense, обчислювальна машина	Тривимірні дані про скелет руки	Використання інфрачервоних сенсорів, отримання зв'язків з подальшою класифікацією	Відсутність впливу світла на роботу алгоритму, висока точність розпізнавання	Необхідність наявності додаткового пристрою Intel RealSense, обмежена дальність роботи, складність алгоритму

Продовження таблиці 2.1

1	2	3	4	5	6
Динамічне розпізнавання жестів за допомогою направлених імпульсних нейронних мереж	Пристрій Microsoft Kinect, обчислювальна машина	Тривимірні дані про ключові точки тіла людини	Перетворення завдання розпізнавання на задачу найкоротшого шляху за допомогою PCNN	Відсутність впливу світла на роботу алгоритму, швидкість роботи алгоритму	Складність реалізації, великі обчислювальні витрати, необхідність наявності додаткового пристрою Microsoft Kinect
Розпізнавання жестів в реальному часі на основі мережі перекалібрування функцій з багатомасштабною інформацією	Камера, обчислювальна машина	Зображення, що подається на вхід згорткового шару	Обробка зображення згортковим шаром для виділення ознак різних рівнів	Висока точність класифікації жестів, хороша дальність роботи	

2.1.1 Традиційна архітектура системи інтерпретації рухів

Традиційна архітектура системи інтерпретації рухів людини забезпечує ефективне керування побутовими пристроями на основі жестів та рухів користувача. Вона включає в себе використання різноманітних технологій та компонентів, таких як датчики руху, мікроконтролери, алгоритми обробки сигналів та програмне забезпечення.

Для збору даних про рухи та жести користувача використовуються різні типи датчиків, включаючи акселерометри та гіроскопи. Ці датчики

можуть бути вбудовані в носимі пристрої, такі як смартфони або смарт-годинники, або розташовані в спеціальних пристроях, які розміщуються в приміщенні. Вони надають інформацію про рухи користувача, яка використовується для подальшого аналізу та інтерпретації.

Отримані від датчиків дані про рухи передаються на мікроконтролер, такий як STM або Arduino, для подальшої обробки. Мікроконтролер виконує різні алгоритми обробки сигналів та аналізу рухів, щоб визначити наміри користувача. Він взаємодіє з програмним забезпеченням і здійснює керування побутовими пристроями відповідно до інтерпретованих даних.

Для ефективної обробки сигналів, отриманих від датчиків руху, використовуються різні алгоритми, які дозволяють визначити тип руху чи жесту. Наприклад, алгоритми класифікації можуть відрізнити різні жести, а алгоритми прогнозування можуть передбачати наступні дії користувача.

Програмне забезпечення виконує роль посередника між мікроконтролером та побутовими пристроями. Воно отримує інтерпретовані дані про рухи користувача і надсилає команди для керування освітленням, опаленням, або іншими побутовими пристроями.

Один із прикладів традиційної архітектури – це система, яка виявляє рухи користувача за допомогою сенсорів, передає дані на мікроконтролер Arduino, де вони обробляються, і використовуються для керування освітленням або опаленням у приміщенні. Інший приклад – це система, яка використовує мікроконтролер STM для аналізу жестів, виявлених сенсорами, і автоматично вмикає або вимикає побутові пристрої залежно від рухів користувача.

Традиційна архітектура системи інтерпретації рухів забезпечує ефективно та зручне керування побутовими пристроями на основі рухів користувача. Її використання дозволяє оптимізувати взаємодію з пристроями та створює можливість для реалізації різноманітних сценаріїв автоматизації у побуті.

2.2 Алгоритми машинного навчання для аналізу рухів

Для покращення швидкодії запропоновано алгоритм, оснований на згортковій нейронній мережі з невеликою архітектурою. Основою згорткової нейронної мережі є шари згортки. Кожен шар включає фільтри для кожного каналу. Вони обробляють попередній шар частинами (шляхом сумування матричних фрагментів). Усі ваги ядра згортки наперед не відомі і змінюються в процесі навчання залежно від вхідних даних. На кінці шару згортки завжди стоїть функція активації.

Для передачі інформації з одного шару на інший потрібна функція активації. Вона перетворює інформацію (числові значення) з усіх нейронів попереднього шару в певне значення для нейрона поточного шару. Вихід залежить від функції активації і може бути як дійсним, так і цілим. Значення виходу є показником активізації нейрона поточного шару.

Оскільки кожному фільтру згортки відповідає одна карта ознак, це дозволяє нейронній мережі навчитися виділяти ознаки незалежно від їх розташування в вхідному зображенні.

Пулінг можна тлумачити так: якщо на попередній операції згортки були виявлені деякі ознаки, то для подальшої обробки так детальне зображення вже не потрібно, і воно зменшується в розмірності, тобто стискається до менш детального. Крім того, фільтрація вже непотрібних деталей зменшує перенавчання. Шар пулінгу, як правило, вставляється після шару згортки перед наступним шаром згортки.

Завдяки шару пулінгу мережа стає найбільш стійкою до змін вхідного зображення, наприклад, до його зміщень. Також зменшується розмірність наступних шарів.

Повнозв'язаний шар (багат шаровий перцептрон) - прихований шар, пов'язаний з усіма нейронами попереднього шару. Останнім шаром багат шарового перцептронів є один або декілька нейронів, кількість яких дорівнює кількості класів. Простіше кажучи, на вхід усій згортковій

нейронній мережі надходить зображення, а на виході мережа видає клас, до якого відноситься це зображення.

Згорткові нейронні мережі забезпечують часткову стійкість до змін масштабу, зсувів, обертань, зміни ракурсу та інших спотворень. Загальна топологія зображена на рисунку 2.1.

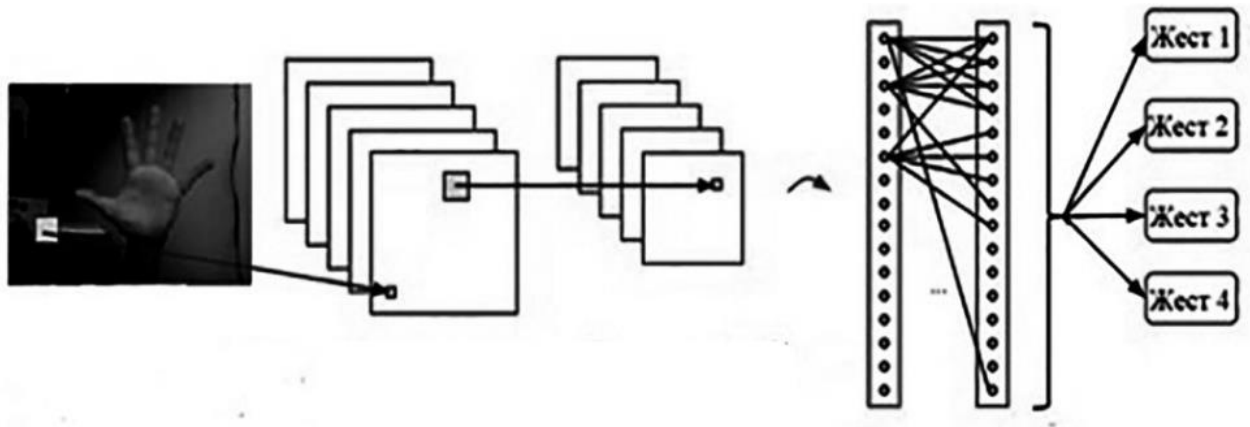


Рисунок 2.1 – Обробка жесту руки на нейронній мережі згорткового типу

Класифікація жестів – це останній і найважливіший крок у розпізнаванні жестів. Більшість людських жестів – це динамічні жести. Один динамічний жест завжди складається із кількох кадрів. Щоб класифікувати динамічні жести, класифікація жестів повинна виконуватись після або разом із відстеженням жестів.

Однією з тенденцій є підхід глибокого навчання. Основними обмеженнями глибокого навчання обмежені обчислювальні потужності. Однак обчислювальна потужність, що експоненційно зростає, може легко вирішити цю проблему. Кількість програм класифікації жестів, заснованих на глибокому навчанні, швидко зростає. Інша тенденція полягає у поєднанні різних алгоритмів класифікації. Кожен алгоритм класифікації має свої переваги та недоліки. Щоб використовувати це, різні класифікатори можуть бути поєднані для досягнення кращої продуктивності. Ми також зауважили,

що важливо координувати алгоритми класифікації жестів за допомогою алгоритмів ідентифікації та трекінгу жестів.

Переваги та недоліки підходів до класифікації жестів наведені у таблиці 2.2.

Таблиця 2.2 – Переваги та недоліки підходів до класифікації жестів

Підхід	Переваги	Недоліки
Метод К-найближчих сусідів	Простота	К-параметр слід обирати обережно
Прихована Марківська модель	Гнучкість навчання та перевірки, прозорість моделі	Необхідно відрегулювати безліч вільних параметрів
Метод опорних векторів	Можуть застосовуватися різні функції ядра	Число опорних векторів зростає лінійно з розміром навчального набору
Метод ансамблю	Не потрібно великої кількості даних навчання	Легко перенавчити, чутливість до шумів та викидів (outliers)
Динамічне деформування часових рядів	Надійне нелінійне вирівнювання між зразками	Складність часу та за обсягом даних
Штучні нейронні мережі	Може виявляти складні нелінійні залежності між змінними	Принцип «чорної скриньки» не може використовуватися за наявності невеликого набору даних для навчання
Глибоке навчання	Не потребують хорошої підготовки ознак, перевершує інші методи машинного навчання	Потрібна велика кількість навчальних даних та значні обчислювальні потужності

2.3 Порівняльний аналіз технічних засобів для збору та аналізу даних про рухи людини

У рамках даного розділу буде проведено докладний аналіз технічних засобів, які використовуються для збору та аналізу даних про рухи людини. Дослідження цих засобів є важливим аспектом для розвитку систем інтерпретації рухів та керування різними технічними пристроями. Розглядатимуться різні параметри технічних засобів, включаючи їх характеристики, можливості та обмеження.

Intel RealSense: Система RealSense використовує інфрачервоні сенсори для збору тривимірних даних про рухи людини. Один з його основних переваг - висока точність розпізнавання жестів, незалежність від освітлення та широкий спектр застосування. Проте, для ефективної роботи необхідна додаткова підтримка обладнання та обмеження у дальності роботи.

Microsoft Kinect: Кінект використовує нариси тіла для аналізу рухів. Це надає високу швидкість роботи та нечутливість до впливу світла. Проте, його складність реалізації та великі обчислювальні витрати можуть бути обмеженнями.

AWS: Платформа AWS може бути використана для збору та аналізу даних про рухи людини у хмарних середовищах. Її переваги включають високу масштабованість та доступність великого обсягу ресурсів, але можливі затримки у зв'язку з мережею та питання конфіденційності даних.

Arduino: Платформа Arduino використовується для збору даних про рухи за допомогою різноманітних сенсорів. Її переваги полягають у низькій вартості та простоті використання, але обмежені можливості обробки даних та необхідність у додатковому обладнанні.

Саморобні технічні засоби: На відміну від комерційних пристроїв, існують також саморобні технічні засоби для збору та аналізу даних про рухи людини. Їх перевагою є можливість налаштування під конкретні потреби, але

такі системи можуть бути складнішими у реалізації та потребувати додаткового часу та зусиль для створення.

Порівняльний аналіз: Для порівняльного аналізу цих технічних засобів будуть враховуватися їх характеристики, ефективність, вартість та відповідність конкретним потребам дослідження. Такий аналіз допоможе визначити найбільш оптимальний інструмент для конкретного застосування.

Цей аналіз технічних засобів для збору та аналізу даних про рухи людини є важливим етапом для подальшого розвитку систем інтерпретації рухів та керування.

2.4 Можливості обраної технології детектування жестів

Вимоги до ПЗ та АЗ для використання CoreML наведені нижче.

Програмне забезпечення (ПЗ):

- операційна система iOS від 11.0 iPhone/iPad;
- операційна система macOS від 10.13.

Інтеграційні середовища: Xcode від 9.0.

Мова програмування: Swift 4+.

Бібліотеки та фреймворки:

- CoreML для машинного навчання;
- Vision для обробки зображень та відео;
- ARKit для доповненої реальності.

Інструменти для розробки моделей:

- Create ML (для створення та тренування моделей на macOS);
- TensorFlow, PyTorch (для створення моделей, які потім можна конвертувати у формат CoreML).

Апаратне забезпечення (АЗ):

- пристрої від iPhone 6s;
- пристрої від iPad Air 2;

- MacBook з підтримкою macOS від 10.13.

Процесор:

- A9 або новіший для iOS пристроїв;
- Intel або Apple Silicon (M1, M1 Pro, M1 Max) для Mac.

Оперативна пам'ять: від 2 ГБ ОЗП для базової роботи.

Камера: вбудована камера iPhone, iPad або Mac для зйомки відео та фотографій.

У майбутньому також можна викорисати датчики акселерометр, гіроскоп, та інші, для покращення роботи з доповненою реальністю (ARKit).

Жести, які вміє розпізнавати технологія (CoreML з Vision):

- розпізнавання обличчя – виявлення облич, розпізнавання характеристик обличчя (наприклад, очі, рот, ніс);
- розпізнавання рук - виявлення положення рук, визначення положення пальців;
- розпізнавання поз – виявлення пози тіла, визначення розташування основних точок на тілі (наприклад, лікті, коліна).

Приклади жестів, які можна розпізнати за допомогою CoreML з Vision (таблиця 2.3):

- жести руками – відкрита долоня, стиснутий кулак, вказівний палець;
- жести головою – кивок, поворот голови;
- пози тіла - положення стоячи, положення сидячи, підняття рук вгору.

CoreML дозволяє додавати нові жести шляхом навчання моделей з новими даними. Використовуючи Create ML або інші інструменти, такі як TensorFlow або PyTorch, можна створити кастомні моделі для специфічних жестів, які потім конвертуються у формат CoreML для використання на пристроях Apple.

Процес розширення бази жестів включає в себе: збір даних (відео або зображення з новими жестами), анотація даних (позначення ключових точок на зображеннях або відео), навчання моделі (використання інструментів для машинного навчання для тренування моделі), конвертація моделі (у формат CoreML), інтеграція (додавання нової моделі у застосунок за допомогою Xcode та Swift). Цей процес дозволяє постійно оновлювати і покращувати

базу розпізнаваних жестів, роблячи застосунок більш функціональним та адаптивним до нових вимог.

Таблиця 2.3 – Приклади жестів, які можна розпізнати за допомогою CoreML з Vision, включаючи жести руками, жести головою та пози тіла

Категорія	Приклади жестів
Жести руками	Піднята рука (salute)
	Махання рукою (wave)
	Стиснутий кулак (fist)
	Показаний великий палець (thumbs up)
	Вказівний палець (pointing)
	Жест "ОК" (OK sign)
	Показаний знак "V" (peace sign)
	Аплодування (clapping)
	Відкриті долоні (open palm)
	Перехрещені пальці (crossed fingers)
	Жести головою
Похитування головою (shake)	
Нахил голови вбік (tilt)	
Поворот голови вліво/вправо (turn)	
Нахил голови вперед/назад (tilt forward/backward)	
Підняття брів (raising eyebrows)	
Опускання голови (bowing)	
Пози тіла	Стійка прямо (standing)
	Сидіння (sitting)
	Нахил вперед (leaning forward)
	Руки на стегнах (hands on hips)
	Схрещені руки (arms crossed)
	Нахил убік (leaning sideways)
	Розставлені ноги (legs apart)
	Сидіння з ногою на ногу (crossed legs sitting)
	Сидіння на колінах (kneeling)
Присідання (squatting)	

2.5 Моделі та методи для розпізнавання жестів людини для керування елементами розумного будинку

Для реалізації розпізнавання жестів людини за допомогою CoreML від Apple використовуються різні моделі та методи машинного навчання,

орієнтовані на обробку зображень і відео. Ці моделі дозволяють ефективно розпізнавати жести для інтерактивного керування елементами розумного будинку.

Традиційна послідовність етапів роботи системи розпізнавання динамічних об'єктів наведена на рисунку 2.2.



Рисунок 2.2 – Традиційна послідовність етапів роботи системи розпізнавання динамічних об'єктів

Нейромережева модель Convolutional Neural Networks (CNN) має архітектуру, яка складаються з згорткових шарів (convolutional layers), шарів об'єднання (pooling layers) і повнозв'язних шарів (fully connected layers). Ці шари працюють разом для автоматичного виділення характеристик з зображень. Принцип роботи полягає у ефективному розпізнаванні патернів в зображеннях, такі як контури, текстури і форми. Для розпізнавання жестів часто використовуються моделі, попередньо натреновані на великих наборах даних (наприклад, ResNet, MobileNet). Для використання на пристроях iOS необхідна конвертація попередньо навчених CNN моделей у формат CoreML.

Нейромережева модель Recurrent Neural Networks (RNN) та Long Short-Term Memory (LSTM) використовуються для обробки послідовностей даних, таких як відео. LSTM має здатність запам'ятовувати довгострокові залежності в даних, що робить його ефективним для розпізнавання

послідовних жестів. Принцип роботи полягає у тому, що RNN та LSTM обробляють кадри відео послідовно, враховуючи попередні кадри для кращого розуміння контексту жестів. LSTM моделі необхідні для аналізу відео потоку з метою розпізнавання складних послідовностей жестів.

Pose Estimation Models - ці моделі використовують CNN для виявлення ключових точок на тілі людини, таких як суглоби та кінцівки. OpenPose є прикладом такої моделі, а оцінки пози визначають положення ключових точок на тілі, що дозволяє визначити жести на основі положення рук, ніг, голови тощо у реальному часі.

Описані моделі, які використовують алгоритми штучного інтелекту, зазвичай використовуються на етапі аналізу підготовлених та покращених вхідних зображень.

Вимогами до нейромережових моделей є:

- модель повинна мати високу точність розпізнавання жестів, щоб мінімізувати помилкові спрацьовування;
- швидкість обробки має бути достатньою для роботи в реальному часі, щоб забезпечити миттєвий відгук на жести користувача;
- модель повинна бути оптимізованою для роботи на мобільних пристроях, щоб знизити споживання енергії;
- модель повинна підтримувати додавання нових жестів без значного погіршення продуктивності.

Попередня обробка зображень включає в себе нормалізацію (масштабування значень пікселів до єдиного діапазону для покращення точності моделі) та аугментацію даних (розширення набору даних шляхом створення модифікованих версій існуючих зображень, наприклад, повороти, зсуви, зміни яскравості).

Класифікація жестів включає в себе класифікація зображень (використання CNN для класифікації окремих кадрів або зображень з відео) та аналіз послідовностей (використання RNN або LSTM для обробки відео послідовностей і визначення жестів на основі контексту).

Інтеграція з CoreML включає в себе конвертацію моделей (тренування моделей у TensorFlow, PyTorch або інших фреймворках і конвертація їх у формат CoreML за допомогою CoreML Tools) та використання CoreML API (Завантаження та інференс моделей на пристроях iOS за допомогою CoreML API).

Методи підвищення точності включають в себе Transfer Learning (використання попередньо навчених моделей і їх адаптація до специфічного завдання шляхом додаткового навчання на новому наборі даних) та Hyperparameter Tuning (налаштування гіперпараметрів моделей для покращення їх продуктивності).

Приклади жестів для керування елементами розумного будинку, які були розглянуті в роботі:

- відкрита долоня – використовується для вмикання світла;
- стиснутий кулак – використовується для вимикання світла.

Використання CoreML для розпізнавання жестів дозволяє створювати інтерактивні інтерфейси для керування елементами розумного будинку. Комбінація CNN, RNN/LSTM та моделей оцінки поз дозволяє забезпечити високу точність, швидкість та гнучкість у розпізнаванні жестів. Інтеграція цих моделей у додатки на iOS забезпечує зручність використання та високий рівень продуктивності, що робить їх ідеальними для таких завдань.

3 РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

3.1 Обґрунтування вибору технології детектування та розпізнавання жестів людини

Для розпізнавання тексту було обрано технології від Apple, так як останні роки ця компанія активно розвиває цей напрямок та надає власні фреймворки для використання на своїх платформах, що в свою чергу дозволяє легко інтегрувати CoreML (фреймворк для роботи з машинним навчанням) з іншими фреймворками компанії, такими як Vision, ARKit, Create ML.

Так як CoreML нативна розробка Apple, вона оптимізована для роботи на усіх пристроях компанії та забезпечує високу продуктивність та енергоефективність. Також сюди входить використання сучасної та продуктивної мови програмування Swift, яка має безпечний синтаксис, а документація Apple забезпечує приклади коду для полегшення та прискорення процесу розробки.

До переваг CoreML ще можна віднести безпеку та конфіденційність даних користувача, так як цей фреймворк обробляє дані безпосередньо на пристрої, так як Apple має строгу політику конфіденціальності. Також CoreML легко інтегрувати у застосунок, який використовує ARKit (фреймворк для створення доповненої реальності) та Vision (фреймворк для попередньої обробки зображень та розпізнавань об'єктів).

Окрім продуктів Apple існують й інші технології від інших компаній, такі як TensorFlow від Google, яка має підтримку моделей ML, включаючи конвертацію з TensorFlow. Добре підходить для різних платформ, включаючи Android і iOS, але є складнішою у налаштуванні для оптимальної роботи на iOS порівняно з CoreML (не так глибоко інтегрований в екосистему Apple); OpenCV + Custom ML Models, яка має потужний

інструмент для комп'ютерного зору з відкритим кодом, має багато готових рішень, проте потребує більше часу на налаштування та інтеграцію і не оптимізовано продуктивність на iOS; MediaPipe від Google, який є оптимізованим для реального часу розпізнавання жестів і поз, підтримує кросплатформені рішення, проте вимагає більшої обчислювальної потужності та не настільки сильно інтегрований в екосистему iOS.

Таблиця 3.1 – Системи на основі штучного інтелекту, які базуються на поєднанні апаратних та програмних засобів для створення інтерактивних інтерфейсів

	Amazon Recognition	Google Cloud Vision:	Microsoft Azure Cognitive Services	MediaPipe	RNN та CNN
Необхідність додаткового апаратного забезпечення	Ні	Ні	Ні	Ні	Так/Ні
Легкість масштабування системи	Висока	Висока	Висока	Середня	Середня
Перелік жестів	Обмежений	Обмежений	Обмежений	Широкий	Необмежений
Можливість розширення бази жестів	Обмежений	Обмежений	Обмежений	Висока	Висока
Легкість інтеграції	Висока	Висока	Висока	Середня	Залежить
Швидкість обробки запиту	Висока	Висока	Висока	Висока	Висока
Вартість	Залежить від обсягу використання	Залежить від обсягу використання	Залежить від обсягу використання	Безкоштовно	Залежить

Вибір технологій від Apple для розпізнавання жестів має декілька важливих переваг, включаючи інтеграцію з екосистемою Apple, високу продуктивність, безпеку та конфіденційність даних. Альтернативні варіанти мають свої переваги, але потребують більше часу та ресурсів для інтеграції.

Amazon Recognition - інструмент для аналізу зображень, що підтримує розпізнавання жестів. Google Cloud Vision - набір API для обробки зображень та розпізнавання жестів. Microsoft Azure Cognitive Services – бібліотека від Google, яка пропонує готові рішення для розпізнавання рук та жестів у реальному часі. RNN та CNN – алгоритми машинного та глибокого навчання, які підтримують можливість виділення ознак та класифікації жестів безпосередньо з зображень, а також застосовуються для аналізу послідовностей жестів у відео.

3.2 Запропонована модель системи розпізнавання жестів

Модель системи, яка запропонована, реалізована та протестована в роботі, наведена на рисунку 3.2.

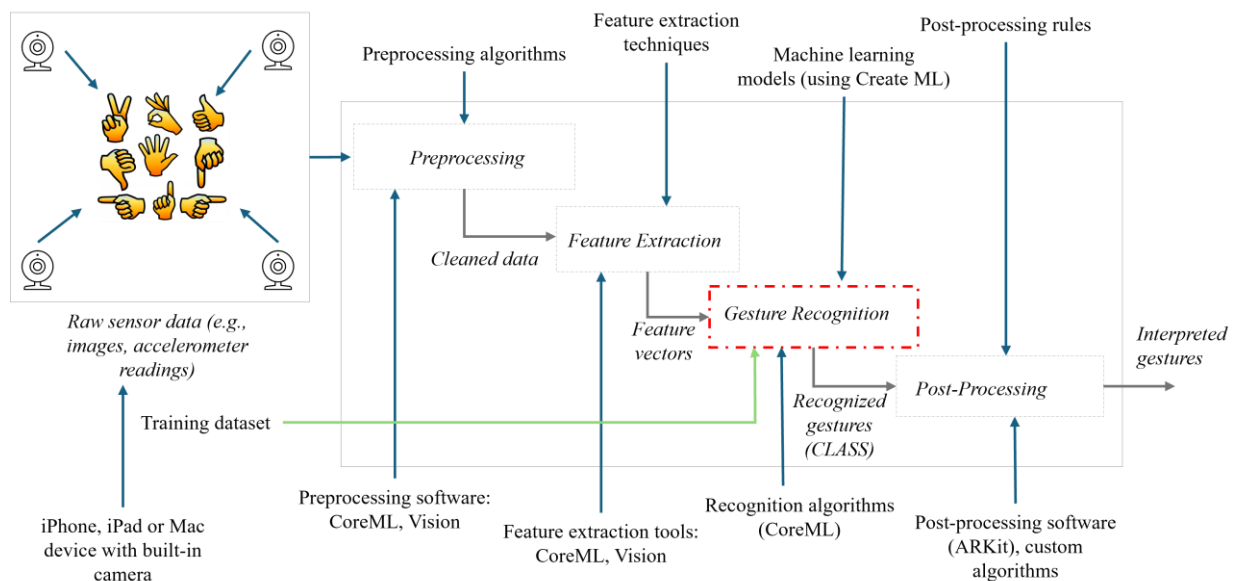


Рисунок 3.2 – Запропонована функціональна модель системи розпізнавання жестів користувача

Модуль попередньої обробки зображень відповідає за підготовку вхідних зображень перед їх подальшою обробкою і аналізом. Основні завдання модуля включають зменшення розмірності зображень (зможуть бути зменшені до стандартного розміру або до розміру, який підходить для подальшого аналізу моделлю машинного навчання) та нормалізацію, яка підготовлює зображення, вирівнюючи яскравість, контраст та кольорові властивості. Це може включати стандартизацію діапазону значень пікселів або зменшення впливу освітлення.

Модуль виділення ознак підготовленого зображення відповідає за вилучення різних ознак з попередньо оброблених зображень. Основні завдання модуля включають витягування характеристик (використання конволюційних нейронних мереж (CNN) або інших алгоритмів для автоматичного виявлення важливих ознак, таких як рухи рук) та підсумкові ознаки (збір та агрегація витягнутих ознак в компактні представлення, які використовуються для подальшого аналізу).

Модуль аналізу виділених ознак здійснює аналіз вилучених ознак для класифікації або інтерпретації жестів. Основні завдання модуля включають класифікацію (використання класифікаційних моделей, таких як мережі глибокого навчання або інші алгоритми машинного навчання, для визначення, який жест був виконаний на основі виділених ознак) та повторне використання інформації (про вибірку для визначення значущості та інтерпретації даних).

Робота даного модулю можлива лише після навчання нейромережевої моделі, яка лежить в основі модулю. Навчання відбувається на створеному датасеті для рішення поставленої задачі. Створений датасет має наступні характеристики:

- різні умови освітлення;
- обрані руки для тренування на базі класів: ліва рука, права рука;
- фото власних рук (близько 200 шт для кожного класу) по ~2.8 Мб (Size: 3024x4032);

- наявність на фото аксесуарів (каблучка, годинник);
- середовище - кімната, для проведення експериментів;
- враховується поворот руки близько 50'.

Модуль інтерпретації класифікованого жесту відповідає за інтерпретацію результатів класифікації жесту. Основні завдання модуля включають оцінку достовірності (визначення достовірності класифікації і відмови від знань) та результати та рекомендації (відображення результатів для кінцевого користувача та надання рекомендацій для подальших дій).

На даному етапі передбачена комунікація (фото)датчиків із обчислювальним сервером для прийняття рішення на основі отриманих результатів розпізнаних жестів окремими гаджетами із попередньо навченою моделлю для локального отримання результату. Організаційна схема системи наведена на рисунку 3.3. Це є графічним представленням структури системи, відображає різні елементи системи та їх взаємозв'язки. Організаційна схема запропонованої системи допомагає зрозуміти, як система організована і як компоненти взаємодіють один з одним.

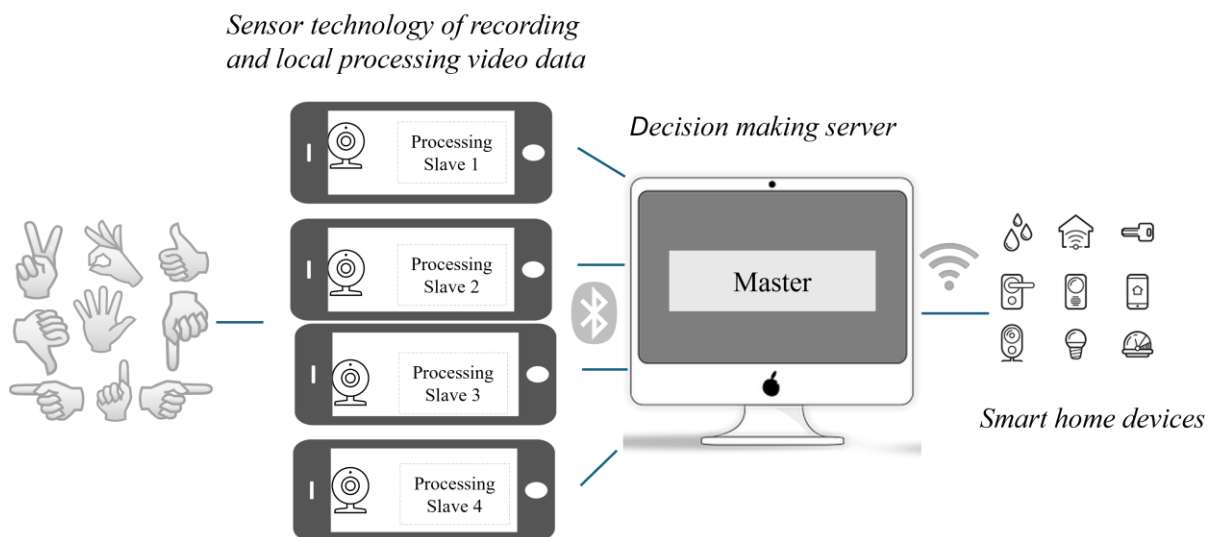
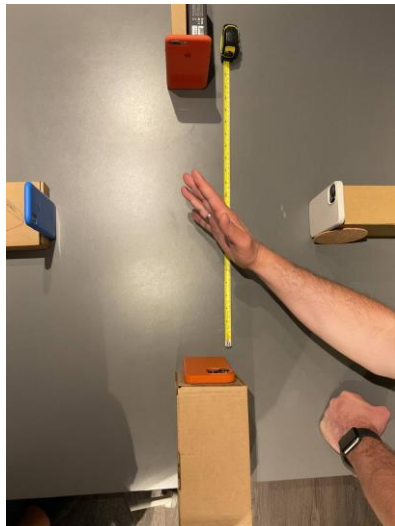


Рисунок 3.3 – Організаційна схема запропонованої системи

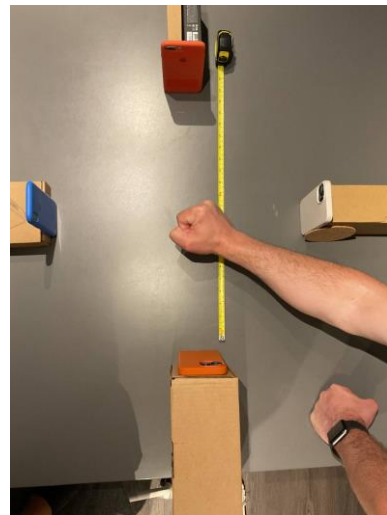
Сенсорні пристрої (Processing Slave 1-4) записують відеодані та здійснюють локальну обробку для визначення жестів користувача. Оброблені

дані з сенсорних пристроїв передаються на сервер прийняття рішень за допомогою бездротового з'єднання. Сервер прийняття рішень "Master" отримує оброблені дані та на їх основі приймає рішення щодо керування пристроями розумного будинку. Сервер прийняття рішень передає команди до пристроїв розумного будинку через бездротове з'єднання, що дозволяє здійснювати контроль та управління різними елементами будинку на основі розпізнаних жестів.

Запропонована система є гнучкою та портативною, що дозволяє використовувати її при зміні місця перебування особи, на особливих жестах якої навчається система. Надійність та відмовостійкість системи полягає у тому, що забезпечити її працездатність можливо як на 2, так і на 4 гаджетах.



а)



б)

Рисунок 3.4 – Тестовий стенд для проведення експериментальних досліджень: а) для розпізнавання розкритої долоні, б) для розпізнавання зібраної у кулак долоні

Для тестування працездатності запропонованої системи на основі зазначених технологій було організовано тестовий стенд (рисунок 3.4), який дозволив виконати ряд експериментів, наведених далі.

3.3 Проведення експериментальних досліджень

3.3.1 Експеримент 1. Тестування інструменту Create ML від Apple на власному наборі жестів

В даному експерименті розглядається інструменту Create ML від Apple, яка застосовується на всіх чотирьох обчислювачах, розміщених по кутах досліджуваної області. Тобто всі девайси, які входять до складу системи детектування та розпізнавання жестів людини, навчені на одній і тій самій неймережевій моделі, яка лежить в основі інструменту Create ML від Apple. Create ML (Apple) - інструмент, доступний через середовище розробки Xcode, який дозволяє виконати навчання моделі за заданою колекцією зображень локально на MacBook/iMac/...

Таблиця 3.2 – Загальне дослідження продуктивності технології розпізнавання жестів людини. Показники для інструменту Create ML від Apple

Кількість шарів моделі	Кількість епох тренування	Точність тестова	Втрати тестові
1	2	3	4
Навчальна: тренувальна : валідаційна = 50:25:25			
2	10	85%	0.35
	20	88%	0.28
	30	90%	0.25
	50	92%	0.20
4	10	87%	0.32
	20	90%	0.25
	30	92%	0.20
	50	94%	0.18
6	10	88%	0.30
	20	91%	0.24
	30	93%	0.19
	50	95%	0.16
Навчальна: тренувальна : валідаційна = 70:15:15			
2	10	87%	0.33
	20	89%	0.27
	30	91%	0.23
	50	93%	0.19
4	10	89%	0.30
	20	92%	0.23

Продовження таблиці 3.2

1	2	3	4
	30	94%	0.18
	50	96%	0.15
6	10	90%	0.28
	20	93%	0.22
	30	95%	0.17
	50	97%	0.14

Результати експериментальних досліджень, проведених із метою визначення архітектури та особливостей навчання моделі для досягнення найвищої точності наведені в таблиці 3.2.

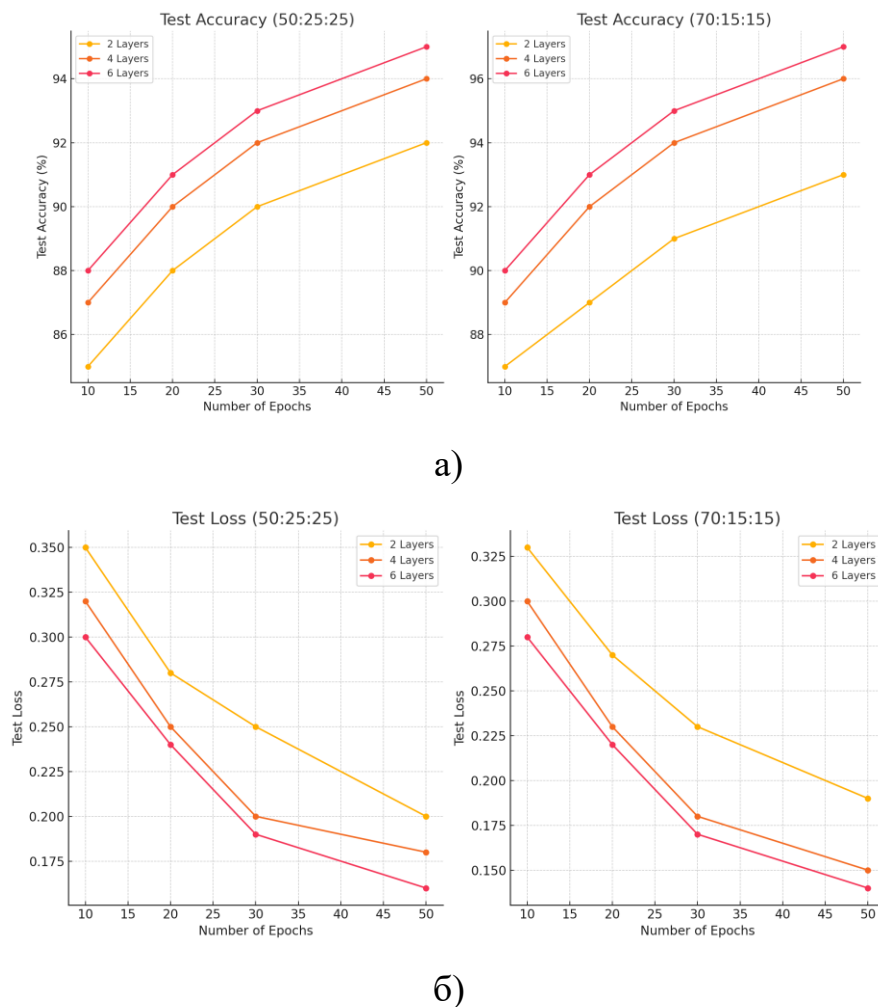


Рисунок 3.5 – Залежність тестової точності та тестових втрат від кількості епох для різних кількостей шарів в CNN моделі від Create ML

Аналіз отриманих результатів показав, що розподіл датасету у співвідношенні 70:15:15 для навчальної: тренувальної : валідаційної вибірок відповідно, дає вищий показник тестової точності та тестових втрат, досягаючи 97%. При цьому, середня точність розпізнавання двох досліджуваних жестів становить 96.66% та 95.94% відповідно для відкритої та стиснутої долоні (таблиця 3.3).

Обрана для першого експерименту технологія показує кращі результати при використанні більшої кількості шарів та епох, особливо при співвідношенні 70:15:15, де точність тестування досягає найвищих значень – 97% точності при 6 шарах та 50 епохах навчання.

Таблиця 3.3 – Дослідження точності системи на різних жестах для інструменту Create ML від Apple (для визначеної архітектури)

	Точність розпізнавання
Точність розпізнавання жесту 1 (відкрита долонь)	96.15%
Точність розпізнавання жесту 2 (стиснутий кулак)	95.47%
Середня точність розпізнавання	95.81%

Втрати тестування знижуються зі збільшенням кількості шарів та епох, що свідчить про покращення загальної продуктивності моделі та здатність узагальнювати отримані знання на нових даних.

3.3.2 Експеримент 2. Тестування сервісу Microsoft Azure на власному наборі жестів

В даному експерименті розглядається сервіс Microsoft Azure, який застосовується на всіх чотирьох обчислювачах, розміщених по кутах досліджуваної області. Тобто всі девайси, які входять до складу системи детектування та розпізнавання жестів людини, навчені на одній і тій самій нейромережевій моделі, а саме - Microsoft Azure.

Таблиця 3.4 – Загальне дослідження продуктивності технології розпізнавання жестів людини. Показники для сервісу Microsoft Azure

Кількість моделі	шарів	Кількість тренування	епох	Точність тестова	Втрати тестові
Навчальна: тренувальна : валідаційна = 50:25:25					
2		10		82%	0.40
		20		85%	0.34
		30		87%	0.30
		50		90%	0.25
4		10		84%	0.38
		20		88%	0.31
		30		90%	0.27
		50		92%	0.23
6		10		85%	0.35
		20		89%	0.28
		30		91%	0.24
		50		94%	0.20
Навчальна: тренувальна : валідаційна = 70:15:15					
2		10		84%	0.37
		20		87%	0.32
		30		89%	0.28
		50		91%	0.24
4		10		86%	0.35
		20		89%	0.30
		30		92%	0.26
		50		94%	0.22
6		10		87%	0.33
		20		90%	0.27
		30		93%	0.23
		50		95%	0.19

Результати експериментальних досліджень, проведених із метою визначення архітектури та особливостей навчання моделі для досягнення найвищої точності наведені в таблиці 3.4. Навчена модель через сервіс від Microsoft конвертується у формат `.mlmodel`, з яким працює CoreML Framework.

Аналіз отриманих результатів показав, що розподіл датасету у співвідношенні 70:15:15 для навчальної: тренувальної : валідаційної вибірок відповідно, дає вищий показник тестової точності та тестових втрат, досягаючи 95% при навчанні мережі на 50 епохах навчання. При цьому,

середня точність розпізнавання двох досліджуваних жестів становить 94.84% та 93.15% відповідно для відкритої та стиснутої долоні (таблиця 3.5).

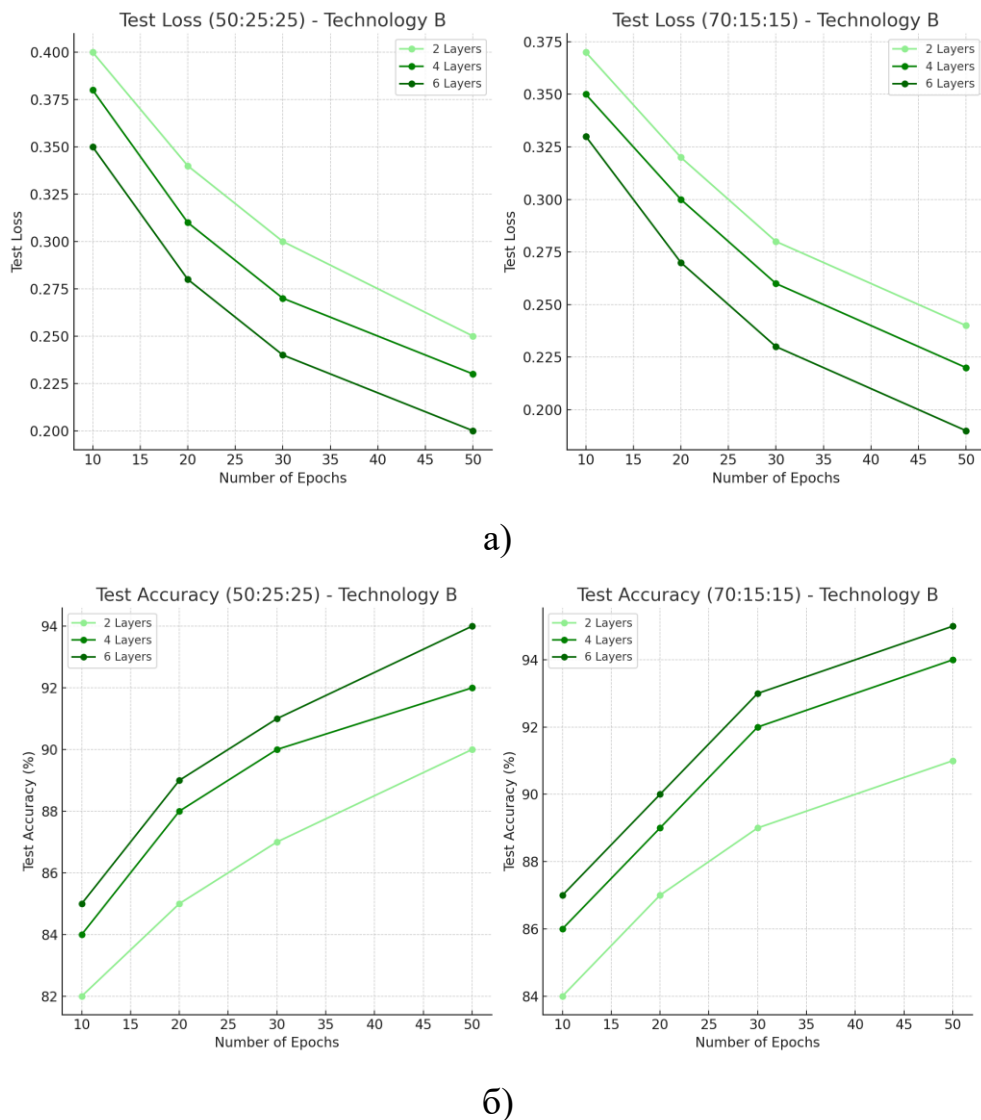


Рисунок 3.6 – Залежність тестової точності та тестових втрат від кількості епох для різних кількостей шарів моделі, наданої сервісом Microsoft Azure

Зі збільшенням кількості епох та шарів моделі точність тестування також покращується. Найвища точність (95%) досягнута при 6 шарах та 50 епохах. Точність зростає швидше у порівнянні з 50:25:25, що свідчить про кращу узагальнювальну здатність моделі при більшій кількості навчальних даних. Втрати тестування зменшуються зі збільшенням кількості епох та шарів моделі. Найменші втрати (0.19) при 6 шарах та 50 епохах.

Таблиця 3.5 – Дослідження точності системи на різних жестах для Microsoft Azure (для визначеної архітектури)

	Точність розпізнавання
Точність розпізнавання жесту 1 (відкрита долонь)	94.84%
Точність розпізнавання жесту 2 (стиснутий кулак)	93.15%
Середня точність розпізнавання	93.99%

3.3.3. Експеримент 3. Тестування сервісу Google Cloud AI Platform на наборі визначених жестів

В даному експерименті розглядається сервіс Google Cloud AI Platform, який застосовується на всіх чотирьох обчислювачах, розміщених по кутах досліджуваної області. Тобто всі девайси, які входять до складу системи детектування та розпізнавання жестів людини, навчені на одній і тій самій нейромережевій моделі, а саме - Google Cloud AI Platform.

Результати експериментальних досліджень, проведених із метою визначення архітектури та особливостей навчання моделі для досягнення найвищої точності наведені в таблиці 3.6. Навчена модель через сервіс від Google конвертується у формат .mlmodel, з яким працює CoreML Framework.

Аналіз отриманих результатів показав, що розподіл датасету у співвідношенні 70:15:15 для навчальної: тренувальної : валідаційної вибірок відповідно, дає вищий показник тестової точності та тестових втрат, досягаючи 97% при навчанні мережі на 50 епохах навчання із тестовими втратами 0.17. При цьому, середня точність розпізнавання двох досліджуваних жестів становить 90.56% та 88.3% відповідно для відкритої та стиснутої долоні (таблиця 3.7).

Розподіл датасету у співвідношення 70:15:15 показує, що зі збільшенням кількості епох та шарів моделі точність тестування також покращується. Точність зростає швидше у порівнянні з 50:25:25, що свідчить про кращу узагальнювальну здатність моделі при більшій кількості

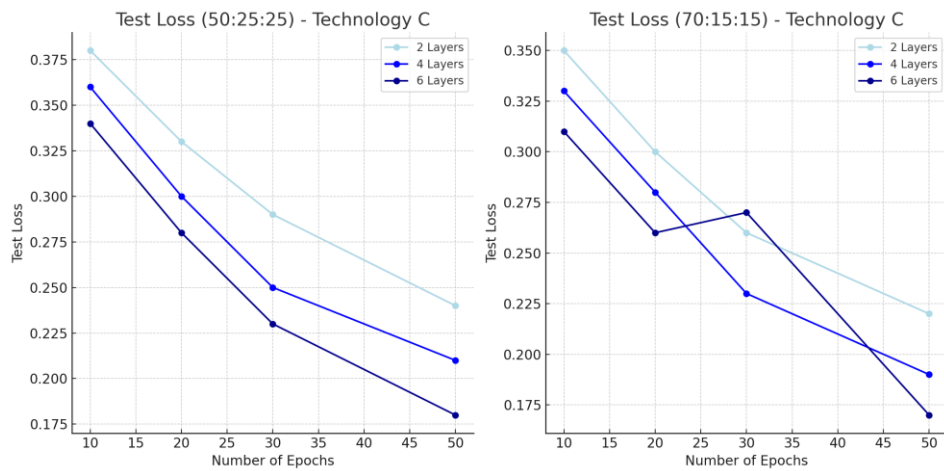
навчальних даних. Втрати тестування зменшуються зі збільшенням кількості епох та шарів моделі. Найменші втрати (0.17) при 6 шарах та 50 епохах.

Таблиця 3.6 – Загальне дослідження продуктивності технології розпізнавання жестів людини. Показники для сервісу Google Cloud AI Platform

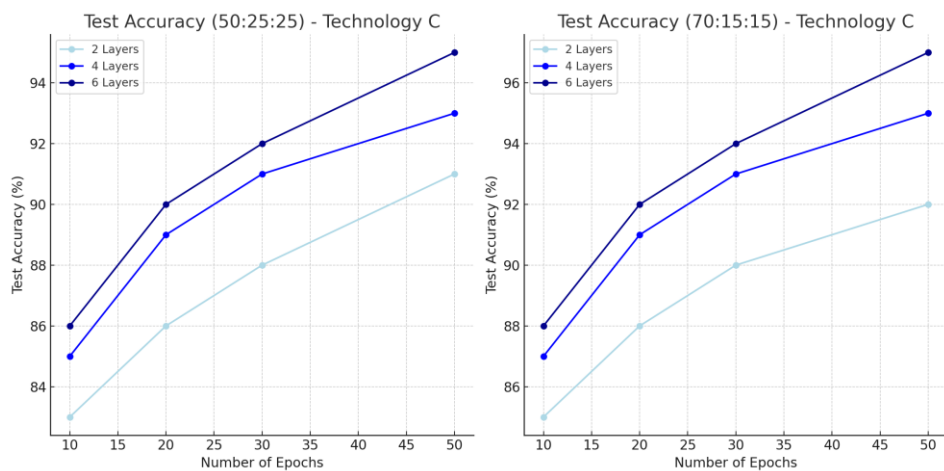
Кількість моделі	шарів	Кількість тренування	епох	Точність тестова	Втрати тестові
Навчальна: тренувальна : валідаційна = 50:25:25					
2		10		83%	0.38
		20		86%	0.33
		30		88%	0.29
		50		91%	0.24
4		10		85%	0.36
		20		89%	0.30
		30		91%	0.25
		50		93%	0.21
6		10		86%	0.34
		20		90%	0.28
		30		92%	0.23
		50		95%	0.18
Навчальна: тренувальна : валідаційна = 70:15:15					
2		10		85%	0.35
		20		88%	0.30
		30		90%	0.26
		50		92%	0.22
4		10		87%	0.33
		20		91%	0.28
		30		93%	0.23
		50		95%	0.19
6		10		88%	0.31
		20		92%	0.26
		30		94%	0.27
		50		97%	0.17

Таблиця 3.7 – Дослідження точності системи на різних жестах для сервісу Google Cloud AI Platform

	Точність розпізнавання
Точність розпізнавання жесту 1 (відкрита долонь)	90.56%
Точність розпізнавання жесту 2 (стиснутий кулак)	88.3%
Середня точність розпізнавання	89.43%



а)



б)



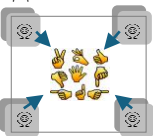
Рисунок 3.7 – Залежність тестової точності та тестових втрат від кількості епох для різних кількостей шарів моделі для сервісу Google Cloud AI Platform

Не залежно від обраної технології детектування та розпізнавання жестів людини, всі експерименти показують вищу середню точність розпізнавання відкритої долоні у порівнянні із середньою точністю розпізнавання стиснутого кулака. Причиною може бути те, що відкрита долонь має велику площу з чіткими, виразними рисами, такими як пальці, лінії на долоні, і форма руки. Це надає більше ключових точок та інформації для алгоритмів комп'ютерного зору для аналізу та розпізнавання. Кулак має меншу площу і менш виразні риси. Згорнуті пальці та зменшення видимості характерних ліній на руці ускладнюють завдання розпізнавання.

3.3.4 Експеримент 4. Тестування запропонованої гібридної системи, заснованої на двох обраних вище моделях

Останній експеримент присвячено використанню гібридних технологій детектування та розпізнавання жестів людини, а саме використанню двох моделей із трьох вище досліджених, які показали більш високі показники точності – моделі у складі інструменту від Apple та сервісу від Microsoft. Тобто далі робота зосереджена саме на цих 2 моделях. Спочатку два діагональні гаджети використовують єдину технологію, потім два паралельні гаджети використовують єдину технологію. Кожна камера має свій номер. Діагональне розміщення показало перевагу на 0,62% у порівнянні із паралельним розміщенням.

Таблиця 3.8 – Дослідження топології розміщення гаджетів із встановленими навченими моделями

			Точність розпізнавання жесту 1 (відкрита долонь)	Точність розпізнавання жесту 2 (стиснутий кулак)	Середня точність розпізнавання
Паралельне розміщення 	Камера 1 Камера 2	Технологія Create ML від Apple	97.24%	95.9%	95.86%
	Камера 3 Камера 4	Технологія Microsoft Azure	95.78%	94.52%	
Діагональне розміщення 	Камера 1 Камера 3	Технологія Create ML від Apple	98.05%	96.78%	96.43%
	Камера 2 Камера 4	Технологія Microsoft Azure	96.5%	94.39%	
Використання єдиної техноогії 	Камера 1 - 4	Технологія Create ML від Apple	96.15%	95.47%	95.81%
	Камера 1 - 4	Технологія Microsoft Azure	94.84%	93.15%	93.99%

Загальними висновками по експерименту 4 є те, що завдяки поєднанню двох різних технологій розпізнавання жестів людини на локальних

обчислювачах, загальна точність детектування жестів для діагонального розміщення склала 96.43%, що на 0.62% вище, ніж для паралельного розміщення (95.81%). Це свідчить про перевагу комбінованого підходу та оптимального розміщення сенсорних пристроїв для підвищення точності системи розпізнавання жестів.

Технологія Create ML від Apple показала вищу точність розпізнавання жестів у порівнянні з технологією Microsoft Azure у всіх розглянутих сценаріях. Це свідчить про більш високу ефективність технології від Apple для завдання розпізнавання жестів.

Використання єдиної технології для всіх камер також показало хороші результати, але менші у порівнянні з гібридним підходом. Середня точність розпізнавання для технології Create ML від Apple склала 95.81%, тоді як для Microsoft Azure – 93.99%.

ВИСНОВКИ

Широкий спектр застосувань систем інтерпретації рухів людини у сфері розумного будинку підкреслює її важливість. Дослідження показало, що розвиток таких систем може відкрити нові можливості для покращення зручності та енергоефективності житла.

В роботі розглянуті різноманітні технічні засоби, які можуть бути використані для збору та аналізу даних про рухи людини. Розглянуті системи, такі як Intel RealSense, Microsoft Kinect, AWS та Arduino, мають свої переваги та обмеження. Важливо враховувати конкретні потреби дослідження при виборі найбільш відповідного засобу.

Порівняльний аналіз різних технічних засобів показав, що кожен з них має свої переваги та обмеження. Вибір конкретного засобу повинен бути здійснений на основі потреб дослідження, вартості та ефективності. Правильний вибір технічного засобу є ключовим для успішного розвитку систем інтерпретації рухів та керування в розумному будинку.

Метою роботи є вдосконалення методів інтерпретації рухів людини для управління різними елементами розумного будинку через віддалене керування. Конкретно, мета полягає в створенні системи, яка забезпечує натуральну та ефективну взаємодію людини з розумним будинком шляхом інтерпретації рухів користувача.

В рамках роботи були вирішені наступні задачі – розглянуто сенсорні пристрої для реєстрації рухів людини, проаналізовано технології детектування та розпізнавання жестів руки, створено функціональну модель системи розпізнавання жестів, в якій головне дослідження зосереджено на модулі нейромережевого аналізу покращених вхідних даних. Результати проведених досліджень з вдосконалення базових нейромережевих моделей та використання гібридного підходу показали, що досягти найвищої точності розпізнавання жестів рук (розкрита долоня та стиснутий кулак) можливо на 6

шарах при 50 епохах навчання, досягаючи точності в 97% із тестовими стратами у 0,17. Дослідження щодо розташування гаджетів відносно досліджуваного об'єкту показало, що діагональне розташування гаджетів із попередньо визначеними технологіями має перевагу перед використанням монотехнології, а також перед використанням паралельного розташування гаджетів із однаковими технологіями на 0.62%.

Гібридний підхід, що використовує технології Create ML від Apple та Microsoft Azure, показує кращі результати у порівнянні з використанням лише однієї технології. Особливо ефективним виявилось діагональне розміщення камер, що забезпечує більш високу точність розпізнавання жестів. Це свідчить про перевагу комбінованого підходу та оптимального розміщення сенсорних пристроїв для підвищення точності системи розпізнавання жестів.

Вирішення цих задач дозволить досягти поставленої мети і розробити систему, яка буде забезпечувати натуральну та ефективну взаємодію людини з розумним будинком через віддалене керування.

На основі цього дослідження можна зробити висновок, що подальший розвиток систем інтерпретації рухів та керування буде сприяти покращенню якості життя в розумних будинках, забезпечуючи зручність, енергоефективність та безпеку.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. O. Barkovska, D. Rosinskiy, I. Mikhailov, M. Vintonovych ENSURING SAFE RESOURCE UTILIZATION OF LIVING SPACE THROUGH CONTROL OF THE MICROCLIMATE OF A SMART HOME //SCIENTIFIC BULLETIN OF KHERSON STATE UNIVERSITY. SERIES «ECONOMIC SCIENCES». – 2023. – №. 49. – С. 5-13. DOI: <https://doi.org/10.32999/ksu2307-8030/2023-49-1>.
2. F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, F. Durand, Capturing the human figure through a wall, ACM Transactions on Graphics (TOG), 34 (2015) 219.
3. F. Adib, D. Katabi, See through walls with WiFi, ACM, 2013.
4. F. Adib, Z. Kabelac, D. Katabi, R. C. Miller, 3d tracking via body radio reflections, in: Usenix NSDI, 2014.
5. S. Celebi, A. S. Aydin, T. T. Temiz, T. Arici, Gesture Recognition using Skeleton Data with Weighted Dynamic Time Warping, in: VISAPP (1), 2013, pp. 620–625.
6. T. Arici, S. Celebi, A. S. Aydin, T. T. Temiz, Robust gesture recognition using feature pre-processing and weighted dynamic time warping, Multimedia Tools and Applications, 72 (2014) 3045–3062.
7. S. S. Rautaray, A. Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, Artificial Intelligence Review, 43 (2015) 1–54.
8. H. Hasan, S. Abdul-Kareem, Static hand gesture recognition using neural networks, Artificial Intelligence Review, 41 (2014) 147–181.
9. T. D'Orazio, G. Attolico, G. Cicirelli, C. Guaragnella, A Neural Network Approach for Human Gesture Recognition with a Kinect Sensor, in: ICPRAM, 2014, pp. 741–746.
10. A. El-Baz, A. Tolba, An efficient algorithm for 3D hand gesture recognition using combined neural classifiers, Neural Computing and Applications, 22 (2013) 1477–1484.

11. K. Subramanian, S. Suresh, Human action recognition using meta-cognitive neuro-fuzzy inference system, *International journal of neural systems*, 22 (2012) 1250028.
12. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*, 521 (2015) 436–444.
13. J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks*, 61 (2015) 85–117.
14. J. Tompson, M. Stein, Y. Lecun, K. Perlin, Real-time continuous pose recovery of human hands using convolutional networks, *ACM Transactions on Graphics (TOG)*, 33 (2014) 169.
15. K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos, in: *Advances in Neural Information Processing Systems*, 2014, pp. 568–576.
16. A. Jain, J. Tompson, Y. LeCun, C. Bregler, Modeep: A deep learning framework using motion features for human pose estimation, in: *Computer Vision-ACCV 2014*, Springer, 2014, pp. 302–315.
17. S. Bilal, R. Akmeliawati, A. A. Shafie, M.J. E. Salami, Hidden Markov model for human to computer interaction: a study on human hand gesture recognition, *Artificial Intelligence Review*, 40 (2013) 495–516.