

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М. Є. Жуковського  
«Харківський авіаційний інститут»

Факультет програмної інженерії та бізнесу

Кафедра інженерії програмного забезпечення

## **Пояснювальна записка до дипломного проєкту**

магістра

(освітній ступінь)

на тему «Аналіз ефективності використання методів комп'ютерного  
розпізнавання образів на зображеннях в ігровій індустрії»

XAI.603.6-95ПЗ1.121.168912.200

Виконав: студент 6 курсу групи № 6-95пз1  
Спеціальність 121 – Інженерія програмного  
забезпечення

(код та найменування)

Освітня програма Хмарні обчислення та  
Інтернет речей

(найменування)

Таран А.В.

(прізвище й ініціали студента)

Керівник Волобуєва Л.О.

(прізвище та ініціали)

Рецензент Філімончук Т.В.

(прізвище та ініціали)

Харків – 2020

**Міністерство світи і науки України**  
**Національний аерокосмічний університет ім. М. Є. Жуковського**  
**«Харківський авіаційний інститут»**

Факультет програмної інженерії та бізнесу

(повне найменування)

Кафедра інженерії програмного забезпечення

(повне найменування)

Рівень вищої освіти другий (магістерський)

Спеціальність 121 – інженерія програмного забезпечення

(код та найменування)

Освітня програма хмарні обчислення та Інтернет речей

(найменування)

**ЗАТВЕРДЖУЮ**

**Завідувач кафедри**

І. Б. Туркін

(підпис)

(ініціали та прізвище)

“ ”

2020 року

**З А В Д А Н Н Я**  
**НА ДИПЛОМНИЙ ПРОЄКТ (РОБОТУ) СТУДЕНТУ**

Тарану Андрію Валерійовичу

(прізвище, ім'я, по батькові)

1. Тема дипломного проекту Аналіз ефективності використання методів комп'ютерного розпізнавання образів на зображеннях в ігровій індустрії

керівник дипломного проекту Волобуєва Ліна Олексіївна, к.т.н., доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від “ ” 2020 року №

2. Термін подання студентом роботи

3. Вихідні дані до роботи: методи комп'ютерного розпізнавання образів

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити)  
огляд і аналіз технологій комп'ютерного зору; аналіз існуючих підходів до розробки аналогічних застосувань; критичний аналіз методів розпізнавання жестів; метод для розпізнавання образів на відеопослідовності; прототип ігрового додатку для демонстрації роботи запропонованого методу

5. Перелік графічного матеріалу

РПЗ – стор. 73, рисунків – 29 шт., таблиць – 2 шт., презентація – 16 слайдів.

## 6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Волобуєва Л.О., доц. каф. 603		
2	Волобуєва Л.О., доц. каф. 603		
3	Волобуєва Л.О., доц. каф. 603		

8. Нормоконтроль \_\_\_\_\_ В.А. Постернакова « \_\_\_\_ » \_\_\_\_\_ 2020 р.  
(підпис) (ініціали та прізвище)

7. Дата видачі завдання \_\_\_\_\_

## КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту	Строк виконання етапів проекту	Примітка
1	Отримання і затвердження теми диплому	03.09.2019	
2	Аналіз предметної області	04.09.2019	
3	Постановка задачі	20.11.2019	
4	Проведення теоретичних досліджень	22.11.2019	
5	Розробка прототипу ПЗ	02.09.2020	
6	Підготовка пояснювальної записки	22.10.2020	
7	Оформлення пояснювальної записки до дипломного проекту	10.11.2020	
8	Передзахист дипломного проекту	27.11.2020	
9	Захист дипломного проекту	24.12.2020	

Студент

\_\_\_\_\_ (підпис)

Гаран А.В.

(прізвище та ініціали)

Керівник роботи

\_\_\_\_\_ (підпис)

Волобуєва Л.О.

(прізвище та ініціали)

## РЕФЕРАТ

Пояснювальна записка до дипломного проєкту містить 73 стор., 29 рис., 20 джерел.

Об'єкт дослідження - процес комп'ютерного розпізнавання образів.

Предмет дослідження – методи комп'ютерного розпізнавання образів на зображеннях в ігровій індустрії.

Метою роботи є підвищення ефективності комп'ютерного розпізнавання образів за рахунок розробки ігрового застосунку, основна механіка якого будується через розпізнавання в реальному часі жестів гравця.

Для досягнення поставленої мети необхідно вирішити ряд завдань: огляд і аналіз технологій комп'ютерного зору; аналіз існуючих підходів до розробки аналогічних застосувань; критичний аналіз методів розпізнавання жестів; розроблено метод для розпізнавання образів на відеопослідовності; розроблено прототип ігрового додатку для демонстрації роботи пропонованого методу.

Наукова новизна. Удосконалено метод розпізнавання жестів на відеопослідовності, який на відміну від існуючих використовує вейвлет-перетворення і метод головних компонент для розпізнавання образів на відеопослідовностях, що надасть змогу зменшити ймовірність помилкових виявлень і підвищити швидкодію роботи системи.

Практична значимість отриманих результатів. В результаті досліджень був розроблений прототип мобільного ігрового додатку, що використовує основну механіку, засновану на комп'ютерному розпізнаванні жестів рук на зображеннях.

РОЗПІЗНАВАННЯ ЖЕСТІВ, КОМП'ЮТЕРНИЙ ЗІР, МЕТОДИ РОЗПІЗНАВАННЯ ОБРАЗІВ, МЕТОД ВІОЛІ-ДЖОНСА, ІГРОВА ІНДУСТРІЯ, МОБІЛЬНА ГРА

## **ABSTRACT**

Explanatory note to the master's thesis 73 pp., 29 fig., 20 sources.

The object of research is the process of computer pattern recognition.

The subject of research is methods of computer recognition of images on images in the gaming industry.

The aim of the work is to increase the efficiency of computer image recognition through the development of a game application, the basic mechanics of which is built through the real-time recognition of player gestures.

To achieve this goal it is necessary to solve a number of tasks: review and analysis of computer vision technologies; analysis of existing approaches to the development of similar applications; critical analysis of gesture recognition methods; developed a method for pattern recognition on video sequences; developed a prototype game application to demonstrate the operation of the proposed method.

Scientific novelty. The method of video sequence gesture recognition has been improved, which, unlike the existing ones, uses wavelet transform and the method of main components for image recognition on video sequences, which will reduce the probability of erroneous detection and increase system performance.

The practical significance of the obtained results. As a result of research, a prototype of a mobile game application was developed, which uses basic mechanics based on computer recognition of hand gestures in images.

**GESTURE RECOGNITION, COMPUTER VISION, PATTERN RECOGNITION METHODS, VIOLA-JONES METHOD, GAME INDUSTRY, MOBILE GAME**

## ЗМІСТ

Вступ .....	6
1 Аналіз предметної області .....	8
1.1 Теорія і технологія комп'ютерного зору .....	8
1.2 Теорія розпізнавання образів .....	13
1.3 Аналіз можливостей існуючих систем .....	17
2 Дослідження методів машинного зору .....	22
2.1 Фільтрація .....	22
2.2 Сегментація .....	24
2.3 Виявлення ознак .....	27
2.4 Модель Bag-of-words .....	32
2.5 Машинне навчання .....	34
3 Розробка ігрового додатку .....	37
3.1 Задача комп'ютерного розпізнавання жестів .....	37
3.2 Алгоритм для розпізнавання жестів .....	40
3.3 Дизайн ігрового додатку .....	49
3.4 UML проектування програмної системи .....	53
3.5 Створення UI\UX .....	65
Висновки .....	71
Перелік посилань .....	72

## ВСТУП

В історії розвитку персонального комп'ютера спостерігається еволюція користувальницького інтерфейсу в людино-машинній взаємодії від текстового командного інтерфейсу до графічного інтерфейсу, від простої клавіатури до миші, джойстика, електронної ручки, відеокамери. Пристрої ЛМВ стали більш зручними і природними для користувача. В даний час потрібна розробка більш потужних і зручних способів взаємодії людини з комп'ютерною системою.

В якості одного із способів забезпечення комфортної взаємодії з комп'ютером, людська рука може бути використана в якості інтерфейсу вводу. Жести рук є потужним каналом зв'язку, який формує основну частину передачі інформації в нашому повсякденному житті. У порівнянні з традиційними пристроями ЛМВ, жести є менш нав'язливим, більш простим та зручним і природним способом взаємодії для користувачів. Проте, виразність жестів все ще залишається недостатньо вивченою для вирішення проблеми людино-машинного взаємодії.

Останнім часом з'явилася і почала швидко розвиватися тенденція використання жестів, особливо жестів руки, як способу взаємодії з комп'ютерною системою. Розпізнавання жестів, таким чином, стало найважливішою частиною в людино-машинній інтелектуальній інтеракції і почало приваблювати багатьох дослідників. Крім того, розроблені в ЛМП технології також знаходять застосування в інших областях, таких як управління роботами, телеконференції, переклад мови жестів для глухонімих, управління комп'ютерними іграми і т.п.

Відеокамера являє собою недорогий і зручний пристрій вводу інформації, який може служити ефективним каналом зв'язку при реалізації людино-машинної взаємодії. Сучасні досягнення в технології комп'ютерного зору і висока продуктивність комп'ютерної техніки роблять відстеження та розпізнавання жестів руки в режимі реального часу перспективним напрямком дослідження з можливістю широкого застосування.

Серед різних підходів до вирішення задачі розпізнавання жестів, розпізнавання жестів на основі комп'ютерного зору виявляється домінантною тенденцією завдяки новим досягненням в області комп'ютерного зору, підвищеній продуктивності комп'ютерів, і також популярності і високій якості сучасних відеокамер. При цьому важливим є той факт, що системи розпізнавання жестів на основі комп'ютерного зору забезпечують більш інтуїтивний канал взаємодії людини з комп'ютером.

Об'єкт дослідження - процес комп'ютерного розпізнавання образів.

Предмет дослідження – методи комп'ютерного розпізнавання образів на зображеннях в ігровій індустрії.

Метою роботи є підвищення ефективності комп'ютерного розпізнавання образів за рахунок розробки ігрового застосунку, основна механіка якого будується через розпізнавання в реальному часі жестів гравця.

Для досягнення поставленої мети необхідно вирішити ряд завдань: огляд і аналіз технологій комп'ютерного зору; аналіз існуючих підходів до розробки аналогічних застосувань; критичний аналіз методів розпізнавання жестів; розроблено метод для розпізнавання образів на відеопослідовності; розроблено прототип ігрового додатку для демонстрації роботи пропонованого методу.

Наукова новизна. Удосконалено метод розпізнавання жестів на відеопослідовності, який на відміну від існуючих використовує вейвлет-перетворення і метод головних компонент для розпізнавання образів на відеопослідовностях, що надасть змогу зменшити ймовірність помилкових виявлень і підвищити швидкодію роботи системи.

Практична значимість отриманих результатів. В результаті досліджень був розроблений прототип мобільного ігрового додатку, що використовує основну механіку, засновану на комп'ютерному розпізнаванні жестів рук на зображеннях.



# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

## 1.1 Теорія і технологія комп'ютерного зору

Комп'ютерний зір покликаний вирішувати задачі, пов'язані зі збором і аналізом візуальної інформації в різних областях виробництва, при цьому частково або повністю замінюючи людину. Питання, що стосуються комп'ютерного зору, були популярні ще в 60-і роки ХХ століття, але зараз вони знаходяться в особливій стадії свого розвитку, оскільки нещодавно з'явилася можливість доступного використання комп'ютерів і програмного забезпечення великою кількістю користувачів для обробки зображень і створення програмних засобів на основі ідеї комп'ютерного зору.

Як наукова дисципліна, комп'ютерний зір відноситься до теорії і технології створення штучних комп'ютерних систем, які здійснюють виявлення, класифікацію і відстеження об'єктів [1]. Відео дані можуть бути представлені безліччю форм, таких як відеопослідовність, зображення з різних камер або тривимірними даними, наприклад з пристрою Kinect або медичного сканера.

Мета комп'ютерного зору полягає в формуванні корисних висновків щодо об'єктів і сцен реального світу на основі аналізу зображень, отриманих за допомогою датчиків.

Як технологічна дисципліна, комп'ютерний зір прагне застосувати теорії та моделі комп'ютерного зору до створення систем комп'ютерного зору.

Прикладами застосування таких систем можуть бути:

- Системи управління процесами (промислові роботи, автономні транспортні засоби).
- Системи відеонагляду.
- Системи організації інформації (наприклад, для індексації баз даних зображень).
- Системи моделювання об'єктів або навколишнього середовища (аналіз медичних зображень, топографічне моделювання).

- Системи взаємодії (наприклад, пристрої вводу для системи людино-машинного взаємодії).
- Системи доповненої реальності.
- Обчислювальна фотографія, наприклад для мобільних пристроїв з камерами.

Підрозділи комп'ютерного зору включають відтворення дій, виявлення подій, стеження, розпізнавання образів, відновлення зображень, тощо.

Область комп'ютерного зору може бути охарактеризована як молода, різноманітна і динамічна. І хоча існують більш ранні роботи, можна сказати, що тільки з кінця 1970-х почалося інтенсивне вивчення цієї проблеми, коли комп'ютери змогли управляти обробкою великих наборів даних, таких як зображення. Однак ці дослідження зазвичай починалися з інших областей, і, отже, немає стандартного формулювання проблеми комп'ютерного зору. Також, і це навіть більш важливо, немає стандартного формулювання того, як повинна вирішуватися проблема комп'ютерного зору. Замість цього, існує маса методів для вирішення різних строго певних завдань комп'ютерного зору, де методи часто залежать від завдань і рідко можуть бути узагальнені для широкого кола застосування.

Чимало методів і програм все ще знаходяться в стадії фундаментальних досліджень, але все більша кількість методів знаходить застосування в комерційних продуктах, де вони часто складають частину більшої системи, яка може вирішувати складні завдання (наприклад, в області медичних зображень або вимірювання і контролю якості в процесах виготовлення). У більшості практичних застосувань комп'ютерного зору комп'ютери попередньо запрограмовані для вирішення окремих завдань, але методи, засновані на знаннях, стають все більш загальними.

Важливу частину в області штучного інтелекту займає автоматичне планування або прийняття рішень в системах, які можуть виконувати механічні дії, такі як переміщення робота через деяке середовище [2]. Цей тип обробки зазвичай потребує вхідних даних, що надаються системами комп'ютерного зору,

що діють як відеосенсор і надають високорівневу інформацію про середовище і робота. Інші області, які іноді описуються як належні до штучного інтелекту, і які використовуються для комп'ютерного зору, – це розпізнавання образів і навчальні методи. В результаті, комп'ютерний зір іноді розглядається як частина області штучного інтелекту або області комп'ютерних наук взагалі.

Існує багато публікацій, присвячених комп'ютерному зору. Даними питаннями займаються як російські вчені, такі як С.Ю. Желтов [3], О.М. Писаревський та ін., так і закордонні, такі як Л. Шапіро, Дж. Стокман, Д. Форсайт та ін., розглядаючи сучасні підходи в цій області.

В своїх роботах Б.В. Анісімов, В.Д. Курганов, В.К. Злобин особливу увагу приділяють ідентифікації точкових зображень при розпізнаванні образів і розпізнавання просторових об'єктів за їх плоским зображенням. Вони розглядають особливості цифрової обробки точкових зображень, методи обробки плоских зображень просторових об'єктів.

Процес обробки і аналіз зображень в задачах комп'ютерного зору детально представлені в дослідженнях Ю.В. Візільтера, С.Ю. Желтова. Автори в аналізі зображень використовують математичні моделі, описують завдання лінійної і нелінійної фільтрації, застосовуючи прийоми виділення контурних точок, гістограмного обробку зображень, бінаризація напівтонових зображень, оператори їх обчислення, проблеми постобробки и контурного зображення.

С.Ю. Желтов і С.І. Протасов розглядають алгоритми і методи аналізу, візуалізації та передачі даних в системах комп'ютерного зору.

Таким чином, можна зробити висновок, що поняття «машинний зір» на сьогоднішній день включає в себе: комп'ютерний зір, розпізнавання зорових образів, аналіз і обробку зображень та багато іншого.

Серед основних задач машинного зору можна виділити:

розпізнавання об'єктів (див. рис.

1.1); ідентифікація об'єктів;

виявлення об'єктів; розпізнавання

тексту;

відновлення 3D форми по 2D

зображенню; оцінка руху; відновлення

зображень;

виділення на зображеннях структур певного виду, сегментація

зображень;

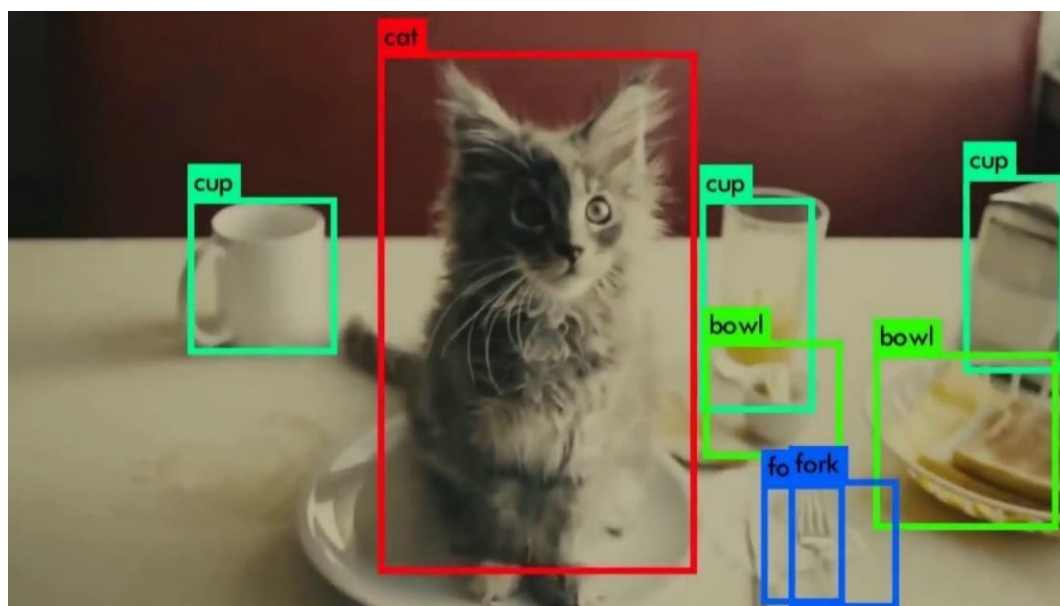


Рисунок 1.1 – Результат роботи алгоритму розпізнавання

Результатом роботи систем комп'ютерного зору може бути видозмінене зображення або список значень певних параметрів зображення, наприклад, розмір об'єкта, колір, орієнтація по відношенню до камери, швидкість.

В даний час машинний зір знаходить найбільше застосування в медицині, робототехніці, автомобільної промисловості, військової галузі та біотехнологіях [4]. Це пов'язано з тим, що в цих галузях вже є чітко сформульовані завдання для комп'ютерного зору, вирішенням яких займаються провідні дослідницькі центри і компанії. При цьому можливе поле застосування машинного зору, посправжньому, величезне – ці технології можуть бути використані практично у всіх сферах життєдіяльності. За рахунок зростання складності розв'язуваних науково-технічних завдань, автоматична обробка і аналіз візуальної інформації стають все більш актуальними питаннями.

Разом з зростанням кількості та складності потенційних місць застосування технологій машинного зору також зростає загальний рівень технологій у світі, які намагаються вирішити якнайбільше з цих задач. Переламний момент, що викликав появу величезної кількості нових задач для автоматизації за допомогою технологій машинного зору, наступив при появі смартфонів з камерами та потужними процесорами, здатними проводити необхідні обчислення.

Велика кількість цих задач були вирішені и на сьогоднішній день: існують додатки для зчитування та перекладу тексту за допомогою камери смартфона у реальному часі (див. рис. 1.2).



Рисунок 1.2 – Розпізнавання тексту у Google Translate

Провідні виробники техніки використовують техніки розпізнавання обличчя для забезпечення безпечної авторизації; існують додатки які дозволяють примірити на себе одяг та косметику не виходячи з дому; автоматизовані няні, що повідомляють батьків про дії дитини та багато іншого.

## 1.2 Теорія розпізнавання образів

Розпізнавання образів – науковий напрямок, пов'язаний з розробкою принципів і побудовою систем, призначених для визначення приналежності даного об'єкту до одного із задалегідь виділених класів об'єктів.

Розпізнавання образів (об'єктів, сигналів, ситуацій, явищ або процесів) – завдання ідентифікації об'єкта або визначення будь-яких його властивостей по його зображенню (оптичне розпізнавання) або аудіозаписи (акустичне розпізнавання) і іншим характеристикам.

З розвитком обчислювальної техніки стало можливим вирішити ряд завдань, що виникають в процесі життєдіяльності, полегшити, прискорити, підвищити якість результату. Наприклад, робота різних систем життєзабезпечення, взаємодія людини з комп'ютером, поява роботизованих систем і ін. Проте, слід відзначити, що забезпечити задовільний результат в деяких завданнях (розпізнавання об'єктів, що швидко рухаються, розпізнавання рукописного тексту) в даний час не вдається.

Задача пошуку зображення за зразком є частиною (підзадачею) більш загальної задачі розпізнавання образів. При несистематизованому і ненаправленому пошуку «схожих» об'єктів з безлічі об'єктів, їх можна перераховувати безкінечно довго і не прийти до завершення із заданою вірогідністю. В окремих випадках об'єкти характеризуються такими ідентифікаційними параметрами (ознаками), як форма, колір, положення, рухливість, по відмінних рис, їх комбінації і т.п. Залежно від цих факторів об'єкти піддаються класифікації. Часто стоїть не глобальне завдання класифікації всіх оточуючих об'єктів, а необхідність виділити в відео-потоці, що надходить, об'єкти певного роду.

Класична постановка задачі розпізнавання образів: Дана множина об'єктів. Щодо них необхідно провести класифікацію. Множина представлено підмножинами, які називаються класами. Задані: інформація про класи, опис

усієї множини і опис інформації про об'єкт, належність якого до певного класу невідома. Потрібно за наявною інформацією про класи і описі об'єкта встановити, до якого класу належить цей об'єкт [5].

Найбільш часто в задачах розпізнавання образів розглядаються монохромні зображення, що дає можливість розглядати зображення як функцію на площині. Якщо розглянути точкову множину на площині, де функція  $(x, y)$  виражає в кожній точці зображення його характеристику — яскравість, прозорість, оптичну щільність, то така функція є формальним записом зображення.

Множина всіх можливих функцій  $(x, y)$  на площині, є моделлю множини всіх зображень. Уводячи поняття подібності між образами можна поставити задачу розпізнавання. Конкретний вид такої постановки сильно залежить від наступних етапів при розпізнаванні відповідно до тих чи інших підходом.

Одним з базових понять є поняття, що не має конкретного формулювання – поняття множини. У комп'ютері множина представляється набором неповторюваних однотипних елементів. Слово «неповторюваних» означає, що якийсь елемент в множині або є, або його там немає. Універсальна множина включає всі можливі для розв'язуваної задачі елементи, порожня не містить жодного.

Образ – класифікаційне угруповання в системі класифікації, яка об'єднує (виділяє) певну групу об'єктів за певною ознакою. Образи мають характерну властивість, що виявляється в тому, що ознайомлення з кінцевим числом явищ з однієї і тієї ж множини дає можливість дізнаватися як завгодно велике число її представників. Образи мають характерні об'єктивні властивості в тому сенсі, що різні люди, які навчаються на різному матеріалі спостережень, здебільшого однаково і незалежно один від одного класифікують одні і ті ж об'єкти. У класичній постановці задачі розпізнавання універсальна множина розбивається на частини-образи. Кожне відображення будь-якого об'єкта на сприймаючі органи системи, що розпізнає, незалежно від його положення щодо цих органів,

прийнято називати зображенням об'єкта, а множини таких зображень, об'єднані певними загальними властивостями, являють собою образи.

Методика віднесення елемента до якого-небудь образу називається вирішальним правилом. Ще одне важливе поняття – метрика, спосіб визначення відстані між елементами універсальної множини. Чим менший цей період, тим більше схожими є об'єкти (символи, звуки та ін.) – те, що ми розпізнаємо. Зазвичай елементи задаються у вигляді набору чисел, а метрика - у вигляді функції. Від вибору уявлення образів і реалізації метрики залежить ефективність програми, один алгоритм розпізнавання з різними метриками буде помилятися з різною частотою.

Навчанням зазвичай називають процес вироблення в деякій системі тієї чи іншої реакції на групи зовнішніх ідентичних сигналів шляхом багаторазового впливу на систему зовнішньої коректування. Таку зовнішню коректування в навчанні прийнято називати «заохоченнями» і «покараннями». Механізм генерації цього коректування практично повністю визначає алгоритм навчання. Самонавчання відрізняється від навчання тим, що тут додаткова інформація про вірність реакції системі не повідомляється.

Адаптація – це процес зміни параметрів і структури системи, а можливо, і керуючих впливів, на основі поточної інформації з метою досягнення певного стану системи при початковій невизначеності і мінливих умовах роботи.

Навчання – це процес, в результаті якого система поступово набуває здатність відповідати потрібними реакціями на певні сукупності зовнішніх впливів, а адаптація – це підстроювання параметрів і структури системи з метою досягнення необхідної якості управління в умовах безперервних змін зовнішніх умов.

Загалом, можна виділити чотири методи розпізнавання образів.

Перший – метод перебору. У цьому випадку проводиться порівняння з базою даних, де для кожного виду об'єктів представлені різноманітні модифікації відображення. Наприклад, для оптичного розпізнавання образів можна застосувати метод перебору виду об'єкта під різними кутами,



масштабами, зміщеннями, деформаціями, тощо. Для букв потрібно перебирати шрифт, властивості шрифту, тощо. У разі розпізнавання звукових образів, відповідно, відбувається порівняння з деякими відомими шаблонами (наприклад, слово, вимовлене кількома людьми).

Другий підхід – проводиться глибший аналіз характеристик образу. У разі оптичного розпізнавання це може бути визначення різних геометричних характеристик. Звуковий зразок в цьому випадку піддається частотному, амплітудному аналізу, тощо.

Наступний метод – використання штучних нейронних мереж (ШНМ). Цей метод вимагає або великої кількості прикладів завдання розпізнавання при навчанні, або спеціальної структури нейронної мережі, яка враховує специфіку даного завдання. Проте, його відрізняє більш висока ефективність і продуктивність.

Останній метод – експертний метод, заснований на безперервному навчанні експертної системи в процесі експлуатації.

Загальна структура системи розпізнавання і етапи в процесі її розробки показані на рис. 1.3.

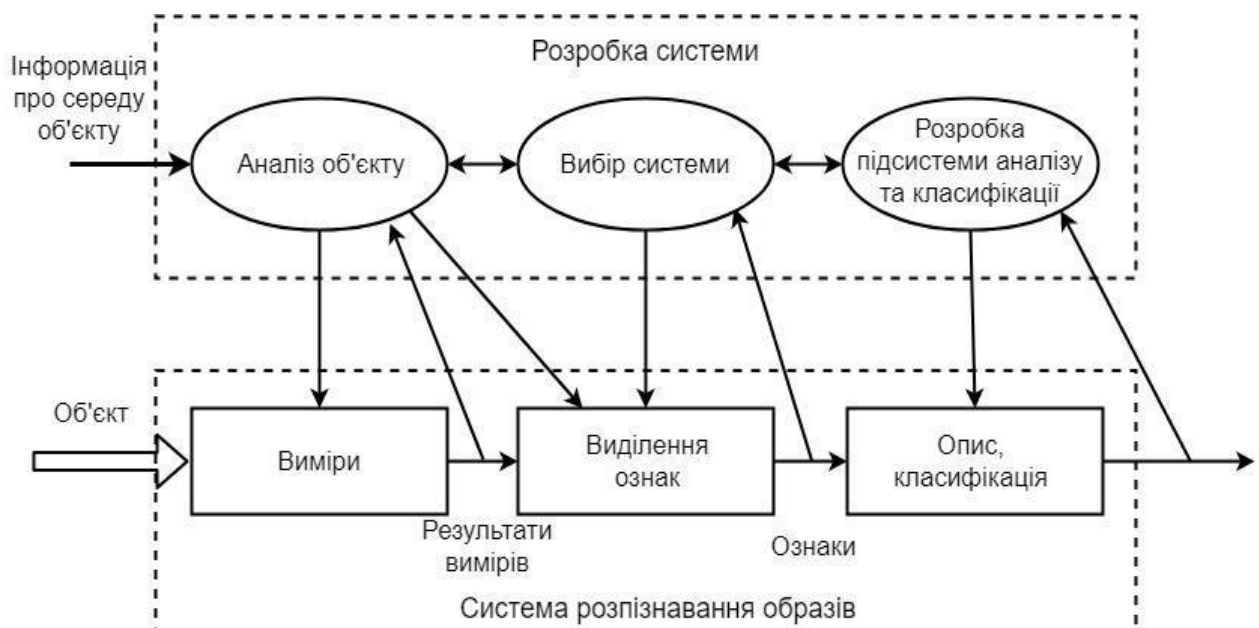


Рисунок 1.3 – Приклад структури системи розпізнавання

При постановці задач розпізнавання намагаються користуватися математичною мовою, прагнучи – на відміну від теорії штучних нейронних мереж, де основою є отримання результату шляхом експерименту, – замінити експеримент логічними міркуваннями і математичними доказами.

### 1.3 Аналіз можливостей існуючих систем

На даний момент на ринку вже існують приклади розважальних додатків та ігрових проектів, що в тій чи іншій мірі використовують засоби та методи машинного зору та розпізнавання образів. Наприклад, широко відомі різноманітні додатки, що використовують технології розпізнавання обличчя для редагування фотографій та відео, додання фільтрів, тощо.

Гарним прикладом є мобільний додаток Snapchat – мультимедійний додаток для обміну повідомленнями, що використовується по всьому світу, розроблений Snap Inc.

Snapchat став відомий тим, що представляє новий мобільний напрямок для соціальних мереж, і приділяє велику увагу користувачам, які взаємодіють з віртуальними наклейками і об'єктами доповненої реальності. Функція «Lens», представлена у вересні 2015 року, дозволяє користувачам додавати ефекти в реальному часі в свої знімки за допомогою технології розпізнавання обличчя (див. рис. 1.4).



Рис. 1.4 – Результат роботи gender-swapping фільтру Snapchat

Фільтр Snapchat являє собою продукт технології комп'ютерного зору, який працює з програмним забезпеченням для розпізнавання осіб. Додаток, створений українським стартапом, дозволяє користувачеві змінювати свої риси обличчя для зображень або відеочату. Після придбання стартапа Snapchat почав працювати над тим, що ми всі зараз використовуємо в нашому повсякденному житті. Фільтри Snapchat використовують обробку зображень в якості певної області для комп'ютерного відео. Він перетворює зображення, виконуючи алгоритм для кожного пікселя зображення. Додаток слід і працює на наступних елементах:

- виявлення обличчя;
- виявлення рис обличчя;
- обробка зображення;
- застосування фільтру для зміни обличчя.

Технологія, над якою працює Snapchat, все ще відносно нова, але за роки була проведена велика робота по її вдосконаленню. Саме досягнення полягає в тому, що додаток доступний кожній людині, і тепер ми розуміємо, чому Snapchat заплатив більше мільйона доларів за його придбання.

Одним із відомих прикладів використання технологій розпізнавання візуальних образів та для створення гри є невелика гра від Google під назвою Quick, Draw! (див. рис. 1.5).

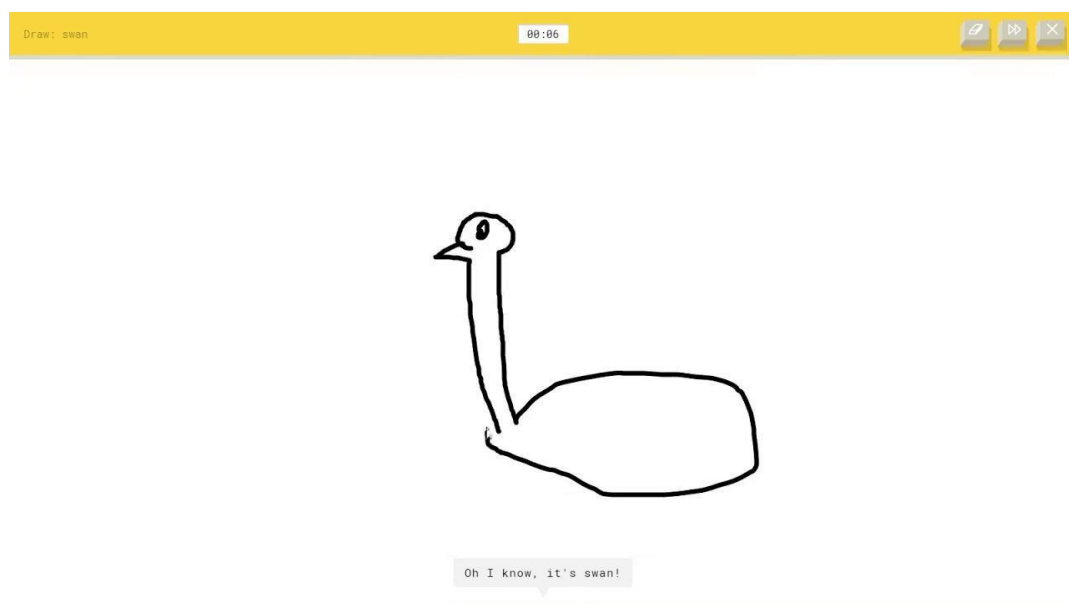


Рис. 1.5 – Процес гри Quick, Draw!

Quick, Draw! – це онлайн-гра, яка змушує гравців малювати зображення об'єкта або ідеї, а потім використовує нейронний штучний інтелект, щоб вгадати, що зображують малюнки.

Штучний інтелект навчається на кожному малюнку, збільшуючи його здатність правильно вгадувати в майбутньому. Гравець має тільки обмежений час для малювання (40 секунд).

Поняття, що вгадує штучний інтелект, можуть бути простими, наприклад, «ноги», або більш складними, наприклад, «міграція тварин». Ця гра є однією з багатьох простих ігор, створених Google, які засновані на штучному інтелекті як частина проекту, відомого як 'A.I. Experiments'.

Один із цікавих проектів, що використовував технології розпізнавання образів для ігрової індустрії – технологія Kinect від Microsoft.

Камера Kinect була випущена корпорацією Microsoft для взаємодії з ігровими приставками на початку 2010 року, і основна її ідея полягала в тому, щоб розробити інтерфейсний пристрій, що дозволяє уникати пристроїв з ручним контролером, використовуючи дані глибини для перетворення рухів гравця в елементи управління, а також дані голосу для голосових команд. Перша версія була випущена з консолі Xbox 360. Друга версія була випущена в 2014 році разом з консоллю Xbox One і мала більшу навіть продуктивність в порівнянні з першою.

Датчик другої версії Kinect – це популярне сенсорний пристрій введення, яке служить природним інтерфейсом користувача для комп'ютерів та ігрових приставок (Xbox one). Датчик Kinect v2 використовується в багатьох областях техніки, так як він може вимірювати глибину, захоплювати кольорові зображення, випромінювати інфрачервоні промені і сприймає аудіо інформацію.

Кольорова камера Kinect v2 (див. рис. 1.6) відповідає за захоплення і потокову передачу кольорового відео для виявлення червоного, синього і зеленого кольорів з Kinect з роздільною здатністю 1920x1080 пікселів, тоді як датчик глибини генерує інформацію про глибину об'єкта перед Kinect з дозволом 512x424 пікселів. Тим часом інфрачервоний випромінювач (ІК-камера

і ІК-проектор) дозволяє датчику бачити світло незалежно. Новий датчик глибини для Kinect v2 має більш високу глибину, точність і значно покращений рівень шуму, а завдяки 3D-візуалізації ми можемо бачити більш дрібні об'єкти і всі об'єкти чіткіше. Оскільки Kinect v2 використовує систему Time-of-Flight, вона модулює джерело світла камери прямокутної хвилею, використовуючи фазовий детектування, щоб виміряти час, який потрібно світлу, щоб пройти від джерела світла до об'єкта і назад до датчика. Глибина потім розраховується на основі швидкості світла в повітрі [6].



Рис. 1.6 – Архітектура Kinect v2

Заснований на додатковому периферійному пристрої в стилі веб-камери, Kinect дозволяв користувачам контролювати і взаємодіяти зі своєю консоллю / комп'ютером без використання ігрового контролера через природний користувальницький інтерфейс за допомогою жестів і голосових команд. Незважаючи на те, що ігрова лінійка не набула великого поширення і в кінцевому підсумку була припинена, сторонні розробники і дослідники виявили декілька застосувань після виходу на ринок для просунутих функцій недорогих сенсорів Kinect, що призвело Microsoft до просування лінійки продуктів в сторону більш нейтральних додатків, включаючи інтеграцію пристрій з платформою хмарних обчислень Microsoft Azure.

Незважаючи на те, що Kinect v2 був в основному розроблено Microsoft як інтерфейс для ігор з консоллю Xbox One, його датчик глибини, заснований на принципі часу прольоту, зробив його популярним в науковому співтоваристві, і дослідники протестували і використовували його в багатьох областях.

Наприклад, в рамках заходу Faculty Summit 2013 азіатське дослідницький підрозділ Microsoft представило нову систему, яка здатна розпізнавати і переводити мову жестів з використанням сенсора рухів Kinect і спеціального програмного забезпечення.

Система була розроблена пекінською командою корпорації у співпраці з Інститутом комп'ютерних технологій Китайської Академії Наук (CAS) за підтримки Microsoft Research Connections в рамках дослідження можливостей використання сенсорних технологій в розпізнаванні жестових мов. Призначена розробка зовсім не для геймерів - в першу чергу вона покликана полегшити взаємодію з комп'ютером людям з обмеженими можливостями здоров'я, які не здатні використовувати традиційне голосове управління.

У першому режимі система сприймає візуальну інформацію і перетворює її в текст англійською або китайською мовою, у другому виробляє зворотний інтерпретацію в жести, які відображаються на екрані за допомогою тривимірного аватару. Творці відзначають, що здатність Kinect одночасно розпізнавати декілька видів даних, в тому числі колір і глибину зображення, сприяє більш точному і швидкому відстеження рухів. Поки розробники реалізували лише підтримку американської мови жестів, але в майбутньому цей список може розширитися.

## 2 ДОСЛІДЖЕННЯ МЕТОДІВ МАШИННОГО ЗОРУ

### 2.1 Фільтрація

У реальному світі сигнали не існують без шуму, який виникає під час придбання (оцифрування) та / або передачі зображення (див. рис. 2.1). Коли є зображення придбані за допомогою камери, рівні освітленості та температури датчика є основними факторами, що впливають на кількість шуму. Під час передачі зображення пошкоджені головним чином через перешкоди в каналі, що використовується для передачі. Видалення шуму зображень – це важлива проблема при обробці зображень [7].



Рисунок 2.1 – Приклад зображення з шумом

Шум - це небажана інформація, яка руйнує якість зображення. У процесі видалення шуму, інформація про тип шуму, присутній у вихідному зображенні відіграє значну роль. Типові зображення пошкоджені шумом, який моделюється як з гауссовським, рівномірним, так і з розповсюдженням солі та перцю.

Гауссовий шум, який має функцію щільності імовірності нормальної розподілу, також відомий як гауссовський розподіл. Інакше кажучи, значення, які може спричинити шум, є гауссово розподіленими. Гауссовський шум належним чином визначається як шум з розподілом амплітуди Гауса.

Шум солі та перцю обумовлений різкими, раптовими порушеннями сигналу зображення, а його поява випадковим чином розподіляється білим або чорним (або обома) пікселями над зображенням. Зображення, що містить шум солі та перцю, матиме темні пікселі у яскравих областях та яскраві пікселі у темних областях (див. рис. 2.2). Цей тип шуму може бути викликаний мертвими пікселями, помилками аналогових та цифрових перетворювачів та бітових помилок у передачі.



Рисунок 2.2 – Приклад зображення з шумом солі та перцю

Mean фільтрація – це простий, інтуїтивно зрозумілий та простий у застосуванні спосіб зменшення шуму на зображеннях, зменшуючи різницю інтенсивності інтенсивності між одним пікселем і наступним. Він часто використовується для згладжування. Ідея mean фільтрації полягає в тому, щоб просто замінити кожне значення пікселя у образі середнім або середнім значенням сусідів, включаючи саме його. Це призводить до знищення значень пікселів, які не відображають їх оточення. Середнє значення фільтрації зазвичай розглядається як фільтр згортки. Як і інші згортки, воно ґрунтується навколо ядра, яке відображає форму та розмір сусіднього регіону для відбору під час обчислення середнього значення. Часто використовуються ядра з квадратів  $3 \times 3$  пікселі, хоча для більш жорсткого згладжування можна використовувати більш великі ядра, наприклад квадрати  $5 \times 5$ . Основна проблема mean фільтрації полягає



в тому, що один піксель з дуже непредставленим значенням може суттєво вплинути на середнє значення всіх пікселів у його сусідстві. Приклад роботи різних фільтрів представлено на рис. 2.3.

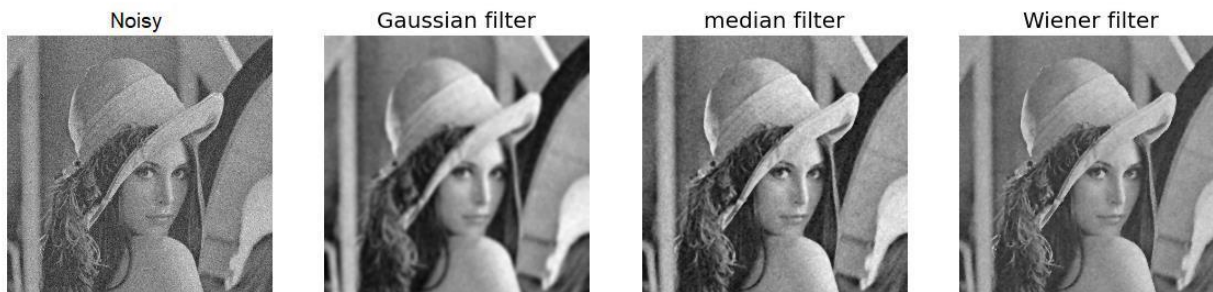


Рисунок 2.3 – Результати роботи різних фільтрів шуму

Медіанний фільтр часто робить кращу роботу, ніж mean фільтр, збереження корисної деталі на зображенні. Як mean фільтр, середній фільтр обмінюється кожним пікселем у зображенні по черзі і розглядає сусідів, що оточують їх, щоб вирішити, чи є вона репрезентативною оточення. Замість того, щоб просто замінити значення пікселя середнім значенням сусідніх значень пікселів, він замінює його медіаною цих значень.

## 2.2 Сегментація

Сегментація зображення - це процес розбиття цифрового зображення на кілька сегментів (наборів пікселів, також відомих як суперпікселі). Мета сегментації полягає у спрощенні та / або зміні представлення зображення на те, що є більш значущим і простішим для аналізу. Сегментація зображення зазвичай використовується для визначення об'єктів та меж (ліній, кривих тощо) у зображеннях. Точніше, сегментація зображення - це процес присвоєння мітки кожному пікселю у зображенні таким чином, щоб пікселі з однаковими мітками мали певні характеристики.

Результатом сегментації зображення є набір сегментів, які разом покривають все зображення або набір контурів, витягнутих із зображення. Кожен з пікселів у регіоні є подібним щодо певного характеристичного чи обчисленого властивості, наприклад, кольору, інтенсивності або текстури (див. рис. 2.4).

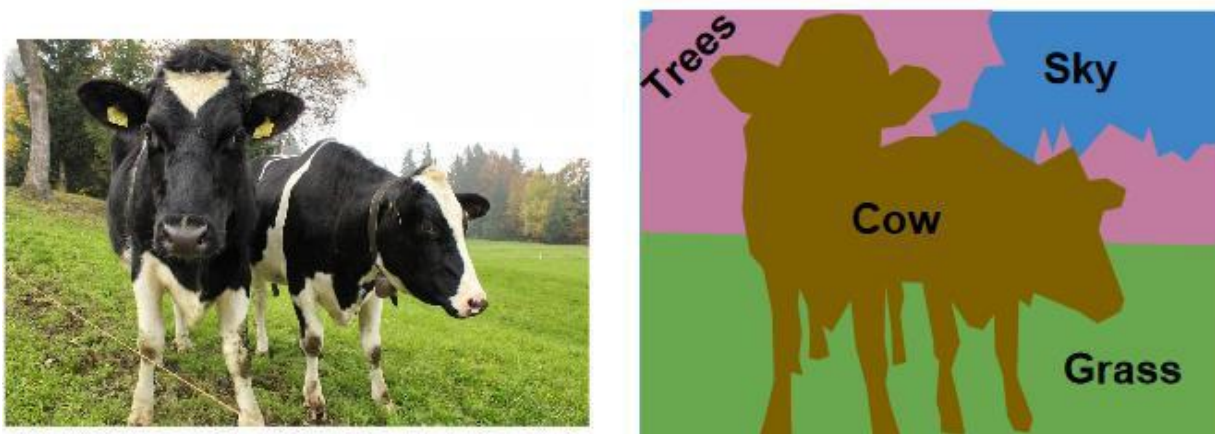


Рисунок 2.4 – Принцип сегментації зображень

Сусідні регіони значно відрізняються по відношенню до однакової характеристики. При застосуванні до стеків зображень, характерних для медичного зображення, результуючі контури після сегментації зображень можуть бути використані для створення 3D-реконструкції за допомогою алгоритмів інтерполяції, таких як *marching cubes*.

В промисловості використовуються кілька популярних методів, включаючи метод максимального ентропії, метод Оцу (максимальна дисперсія) та кластеризація *k-means*.

*K-means* кластеризація – це метод векторного квантування, що був створений для обробки сигналів, а тепер є популярним для кластерного аналізу при пошуку даних. Метод кластеризації *k-means* спрямований на розбиття *n* спостережень на *k* кластерів, в яких кожне спостереження належить до кластера з найближчим середнім значенням, служачи прототипом кластера. Це призводить до розбиття простору даних на клітини Вороного (див. рис. 2.5).

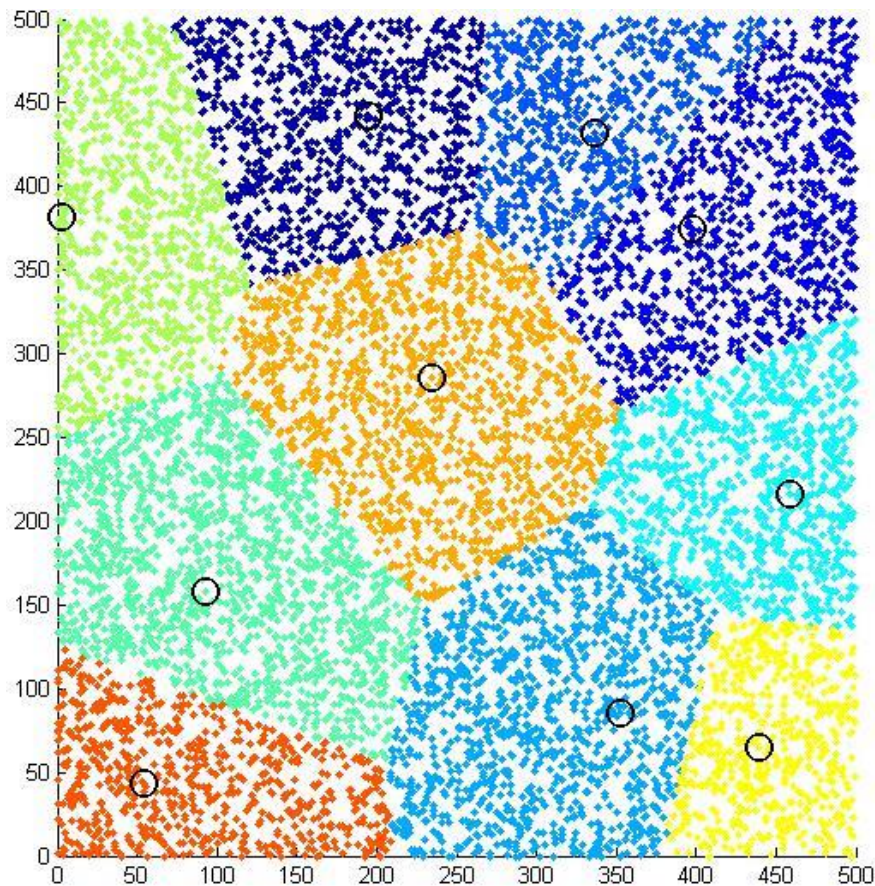


Рисунок 2.5 – Розбиття даних на кластери k-means

Проблема обчислювальної складності (NP-складна); однак існують ефективні евристичні алгоритми, які зазвичай використовуються і швидко сходяться до локального оптимуму. Вони, як правило, аналогічні алгоритму максимізації очікувань для сумішей гауссових розподілів за допомогою ітеративного методу удосконалення, що використовується як моделлю k-means, так і гауссової суміші. Крім того, вони обидва використовують центри кластерів для моделювання даних; однак, k-means кластеризація прагне знайти кластери порівняної просторової межі, тоді як механізм максимізації очікувань дозволяє кластерам мати різні форми.

Алгоритм має вільне відношення до класичного класифікатора k-nearest neighbors, популярної методики машинного навчання для класифікації, яка часто плутається з k-means через назву. Можна застосувати 1-найближчий класифікатор сусідів до кластерних центрів, отриманих за допомогою k-засобу для класифікації нових даних у існуючі кластери.

Найбільш поширеним евристичним алгоритмом є  $k$ -means++ - це алгоритм вибору початкових значень (або "насіння") для алгоритму кластеризації  $k$ -means. Це було запропоновано в 2007 році Девідом Артуром та Сергієм Васильовицьким, як алгоритм наближення для проблеми NP-hard- $k$ -means - спосіб уникнути іноді поганих кластерингів, виявлених за допомогою стандартного алгоритму  $k$ -means [8].

Проблема  $k$ -means полягає в тому, щоб знайти кластерні центри, які мінімізують дисперсію всередині класу, тобто сума квадратних відстаней від кожної точки даних кластеризована до його кластерного центру (найближчого до нього центру). Хоча пошук точного рішення задачі  $k$ -засобу для довільного введення NP-складний, широко використовується стандартний підхід до пошуку наближеного рішення (часто його називають алгоритмом Ллойда або алгоритмом  $k$ -means), який часто знаходить розумні рішення швидко.

Однак алгоритм  $k$ -means має принаймні два основні теоретичні недоліки: по-перше, було показано, що найгірший час роботи алгоритму є супер-поліноміальним у розмірі входу; по-друге, знайдене наближення може бути як завгодно поганим стосовно об'єктивної функції порівняно з оптимальним кластеризацією.

Алгоритм  $k$ -means++ вирішує другу з цих перешкод, дозволяючи вказати процедуру ініціалізації кластерних центрів, перш ніж продовжувати стандартні ітерації оптимізації  $k$ -means. За допомогою ініціалізації  $k$ -means++ алгоритм гарантує пошук рішення, яке є  $O(\log k)$ , конкурентоспроможним для оптимального рішення  $k$ -means.

### 2.3 Виявлення ознак

Виявлення ознак включає в себе методи для обчислення абстракцій візуальних даних та прийняття локальних рішень на кожному зображенні, вказує на те, чи присутня на ньому ознака зображення певного типу чи ні.

Отримані ознаки будуть підмножинами домену зображення, часто у вигляді ізольованих точок, безперервних кривих або з'єднаних областей (див. рис. 2.6).



Рисунок 2.6 – Виявлення ознак різних типів

Немає універсального чи точного визначення того, що являє собою ознаку, і точне визначення часто залежить від проблеми або типу програми. З огляду на це, ознака визначається як "цікава" частина зображення, і ознаки використовуються як відправна точка для багатьох алгоритмів машинного зору. Оскільки ознаки використовуються як відправна точка та основні примітиви для наступних алгоритмів, загальний алгоритм буде часто настільки ефективний, як його детектор ознак. Отже, бажаною властивістю детектора ознак є повторюваність: чи буде виявлена однакова ознака у двох або більше різних зображеннях однієї і тієї ж сцени.

Виявлення ознак - це обробка зображення на низькому рівні. Тобто, це зазвичай виконується як перша операція на зображенні, і розглядає кожен піксель, щоб побачити, чи присутня ознака на цьому пікселі. Якщо це частина більшого алгоритму, то алгоритм, як правило, розглядає лише зображення в регіоні ознак. В якості вбудованого попереднього визначення ознаки, вхідне зображення зазвичай згладжується у представленому масштабному просторі та обчислюється одне чи кілька зображень ознак, часто виражених у термінах операцій місцевих похідних.

Іноді, коли виявлення ознак має велику обчислювальну складність та існують часові обмеження, алгоритм вищого рівня може бути використаний для керування етапом виявлення ознак, тому для пошуку потрібні лише певні частини зображення. Багато алгоритмів машинного зору використовують виявлення функцій як початковий етап, тому в результаті розробляється дуже велика кількість детекторів. Вони широко розрізняються за видами виявленої ознаки, обчислювальною складністю та повторюваністю. Ознаки можна умовно поділити на кілька типів: ребра, кути або точки інтересу, регіони інтересу та хребти (див. рис. 2.7).

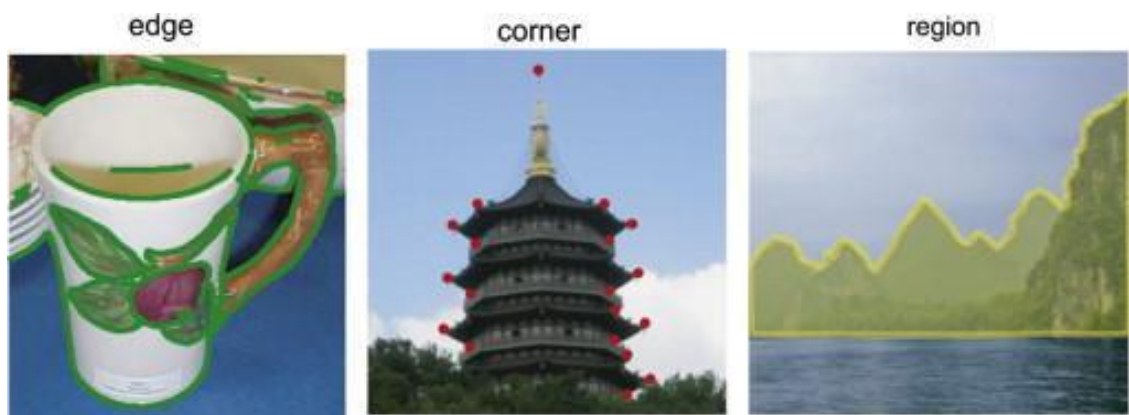


Рисунок 2.7 – Види ознак зображень

Ребра – це точки, де є межа (або край) між двома областями зображення. Загалом, ребро може мати майже довільну форму і може включати в себе з'єднання. На практиці, ребра, як правило, визначаються як набори точок на зображенні, які мають сильну градієнтну величину. Крім того, деякі загальні алгоритми потім об'єднують високі градієнтні точки разом, щоб сформувати більш повний опис ребра. Ці алгоритми зазвичай ставлять певні обмеження на властивості ребра, такі як форма, гладкість та градієнтне значення. На місцевості ребра мають одномірну структуру.

Терміни "кути" та "точки інтересу" використовуються дещо взаємозамінно та стосуються точкових функцій зображення, які мають локальну двомірну структуру. Назва "кут" виникла з тих пір, коли ранні алгоритми вперше

виконували виявлення краю, а потім проаналізували краї, щоб знайти швидкі зміни в напрямку. Потім ці алгоритми були розроблені таким чином, що для явного виявлення країв більше не потрібно, наприклад, шукати високий рівень кривизни градієнта зображення. Потім було помічено, що так звані кути також були виявлені на частинах зображення, які не були кутами в традиційному значенні (наприклад, невелика яскрава пляма на темному тлі). Ці моменти часто відомі як точки інтересу, але термін "кут" використовується за традицією.

Регіони інтересу надають додатковий опис структури зображення з точки зору великих фрагментів зображення, на відміну від кутів, які є більш точними (див. рис. 2.8).



Рисунок 2.8 – Приклад виявлених регіонів інтересу

Тим не менше, дескриптори регіонів інтересу часто можуть містити переважну точку (місцевий максимум реакції оператора або центр тяжкості), що означає, що багато детекторів регіонів також можуть розглядатися як детектори точок інтересу. Детектори регіонів інтересу можуть виявляти ділянки на зображенні, які занадто гладкі, щоб бути виявлені детектором кутів.

Якщо зменшити зображення, а потім виконувати виявлення кутів, детектор буде реагувати на точки, які різкі у зменшеному зображенні, але можуть бути гладкими у вихідному зображенні. Саме в цьому випадку різниця між детектором кута та детектором регіонів стає дещо розпливчастою. значною мірою ці відмінності можна виправити шляхом включення відповідного поняття масштабу.

Для подовжених об'єктів поняття хребтів є природним інструментом. Дескриптор хребта, обчислений із зображення сірого рівня, можна розглядати як узагальнення медіальної осі. З практичної точки зору, хребет можна розглядати як одномірну криву, яка являє собою вісь симетрії, і, крім того, має атрибут локальної ширини хребта, пов'язаної з кожною точкою хребта. Однак, на жаль, алгоритмічно складніше витягувати властивості хребта з загальних класів зображень, крім краєвих, кутових чи чорно-білих. Тим не менше, дескриптори хребта часто використовуються для виявлення дороги на повітряних зображеннях та кровоносних судин у медичних зображеннях (див. рис. 2.9).

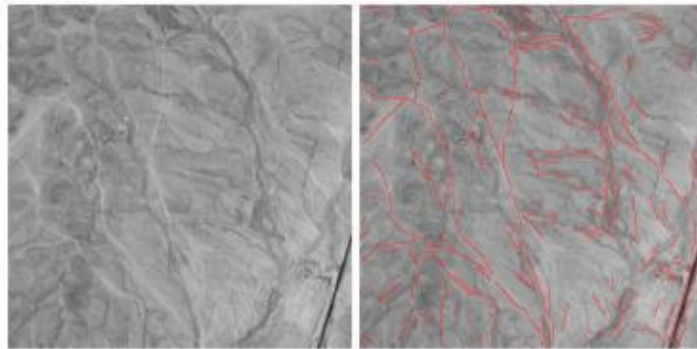


Рисунок 2.9 – виявлення хребтів за допомогою SIFT

Speeded Up Robust Features (SURF) – це запатентований детектор і дескриптор локальних ознак. Він може використовуватися для таких завдань, як розпізнавання об'єктів, реєстрація зображень, класифікація або 3D-реконструкція. Частково він надиханий дескриптором Scale-Invariant Feature Transform (SIFT). Стандартна версія SURF в кілька разів швидше, ніж SIFT, і її автори заявляють про більшу надійність проти різних перетворень зображення, ніж SIFT [9].

Для виявлення точок інтересу SURF використовує цілочисельне наближення детермінанта Гессенського блочного детектора, який можна обчислити за допомогою 3 цілочисельних операцій з використанням попередньо інтегрованого зображення. Його дескриптор опирається на суму значень вейвлета Хаара в точці інтересу. Вони також можуть бути обчислені за допомогою інтегрального зображення.



Алгоритм SURF базується на тих самих принципах та етапах, що і SIFT, але деталі на кожному кроці різні. Алгоритм має три основні частини: виявлення точок інтересу, опис локального середовища та підбір. SURF використовує детектор регіонів інтересу на основі матриці Гессе для пошуку цікавих об'єктів. Визначник матриці Гессе використовується як міра локальних змін навколо точки, і точки вибираються там, де цей детермінант має максимальне значення.

## 2.4 Модель Bag-of-words

Модель "Bag-of-words" (BoW) є спрощувальним представленням, що використовується при обробці природних мов та пошуку інформації. Також відомий як векторна просторова модель. У цій моделі текст (наприклад, речення або документ) представляється як сумка (мультисет) її слів, ігноруючи граматику і навіть порядок слів, але зберігаючи множинність. Модель BoW зазвичай використовується в методах класифікації документів, де частота появи кожного слова використовується як функція для навчання класифікатора.

У машинному зорі модель BoW може бути застосована до класифікації зображень, обробляючи ознаки зображення як слова (див. рис. 2.10). Тобто, гістограма над словником.

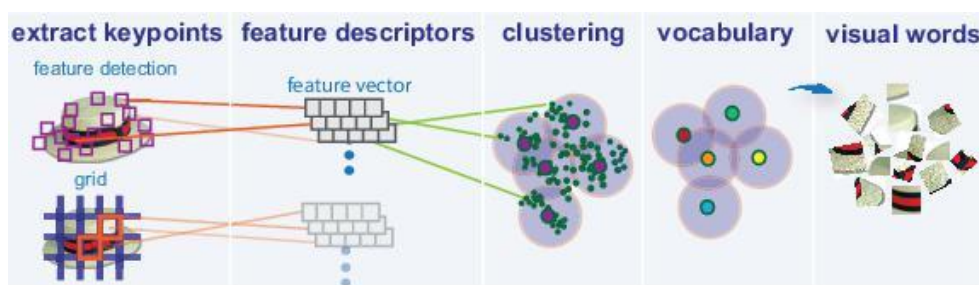


Рисунок 2.10 – Схема роботи BoW для візуальних даних

У машинному зорі мішок візуальних слів є вектором розбіжностей у елементів словника локальних ознак зображення. Для представлення зображення за допомогою моделі BoW зображення можна розглядати як документ. Подібним

чином, також слід визначити "слова" у зображеннях. Це, як правило, включає в себе три етапи: виявлення ознак, опис ознак та генерація словника. Визначення моделі VoW може бути "представлення у виді гистограми на основі незалежних ознак".

Після виявлення ознак кожне зображення абстрагується кількома фрагментами. Методи представлення ознак стосуються як представляти фрагменти як числові вектори. Ці вектори називаються дескрипторами ознак. Хороший дескриптор повинен мати здатність певною мірою керувати інтенсивністю, обертанням, масштабом і афінними варіаціями. Одним із дескрипторів, що добре підходять для такої задачі є SURF. Після цього кроку кожне зображення являє собою набір векторів одного і того ж розміру, де порядок різних векторів не має значення.

Останній крок для моделі VoW полягає у перетворенні векторних фрагментів на "кодові слова" (аналогічні словам у текстових документах), що також створює словник. Приклади фрагментів, що можуть вважатися за слова показано на рис. 2.11.

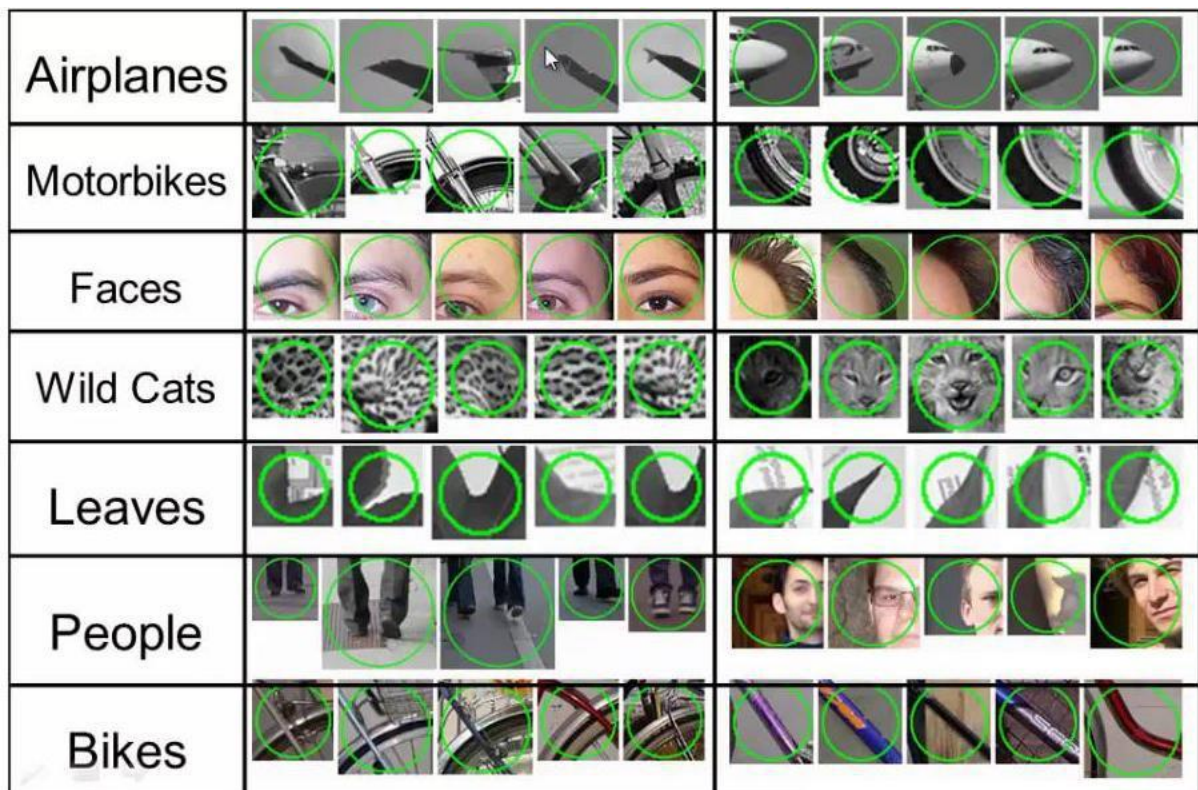


Рисунок 2.11 – Приклади візуальних слів

Кодове слово можна розглядати як представник кількох аналогічних фрагментів. Одним із простих методів є виконання k-means кластеризації над усіма векторами. Кодові слова потім визначаються як центри вивчених кластерів. Кількість кластерів - це розмір словника. Таким чином, кожен фрагмент у зображенні відображається до певного кодового слова через процес групування, і зображення може бути представлено гістограмою кодових слів.

## 2.5 Машинне навчання

Машинне навчання – це підмножина штучного інтелекту в галузі інформатики, яка часто використовує статистичні методи, щоб дати комп'ютерам можливість "вчитися" (тобто поступово покращувати продуктивність для конкретного завдання) з даними, без явного програмування.

Створене на базі дослідження розпізнавання образів та теорії обчислювального навчання в штучному інтелекті, машинне навчання досліджує вивчення та побудову алгоритмів, які можуть вивчати і робити прогнози щодо даних - такі алгоритми подолають строго статичні інструкції програми, створюючи прогнозовані дані чи рішення, шляхом створення моделі з вхідних даних вибірки. Машинне навчання використовується в ряді обчислювальних завдань, де проектування та програмування явних алгоритмів з високою продуктивністю важко або нездійсненно; Приклади застосування включають фільтрацію електронної пошти, виявлення шкідливих інсайдерів, що працюють над порушеннями даних, оптичне розпізнавання символів, навчання в ранзі та комп'ютерне бачення.

Є чіткі компроміси між традиційним комп'ютерним баченням та підходами з машинним навчанням [10]. Класичні алгоритми комп'ютерного зору є зрілими, перевіреними та оптимізованими для продуктивності та енергоефективності, а машинне навчання забезпечує більшу точність та універсальність - але вимагає

великої кількості обчислювальних ресурсів. Тим, хто прагне впровадити високопродуктивні системи, швидко стає зрозуміло, що гібридні підходи, які поєднують традиційне комп'ютерне бачення та машинне навчання, можуть запропонувати найкраще з обох світів. Наприклад, в камері безпеки, алгоритм машинного зору дозволяє ефективно виявляти обличчя або переміщення об'єктів на сцені. Потім менший сегмент зображення, де виявлено обличчя або об'єкт, обробляється шляхом машинного навчання для підтвердження тотожності або класифікації об'єктів - заощаджуючи значні обчислювальні ресурси.

Штучні нейронні мережі (ANNs) є обчислювальними системами, у деякому сенсі натхненними біологічними нейронними мережами, які утворюють тваринний мозок. Такі системи "вчаться" виконувати завдання, розглядаючи приклади, як правило, без програмування з будь-якими правилами конкретних завдань [11]. ANN базується на сукупності пов'язаних одиниць або вузлів, які називаються штучними нейронами (див. рис. 2.12), які вільно моделюють нейрони у біологічному мозку. Кожне з'єднання, як синапси в біологічному мозку, може передавати сигнал від одного штучного нейрона до іншого. Штучний нейрон, який отримує сигнал, може обробити його, а потім сигналізувати додаткові штучні нейрони, пов'язані з ним.

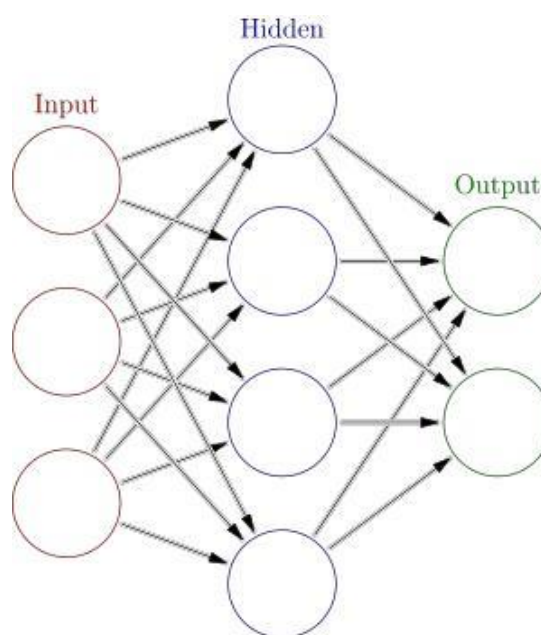


Рисунок 2.12 – Загальна схема нейронної мережі

У поширених реалізаціях ANN, сигнал на зв'язці між штучними нейронами є дійсним числом, а вихід кожного штучного нейрона обчислюється деякою нелінійною функцією суми його входів. З'єднання між штучними нейронами називають "краями". Штучні нейрони та ребра звичайно мають вагу, яка налаштовується у процесі навчання. Вага збільшує або зменшує сили сигналу при з'єднанні. Штучні нейрони можуть мати такий поріг, що сигнал надсилається лише тоді, коли сумарний сигнал перетинає цей поріг. Як правило, штучні нейрони агрегуються в шари. Різні шари можуть виконувати різні види перетворень на своїх вкладах. Сигнали проходять від першого шару (вхідного шару) до останнього шару (вихідний шар), можливо, після проходження по декількох шарах. Оригінальна мета методу ANN полягала в тому, щоб вирішити проблеми так само, як і людський мозок, проте з часом увага перейшла до виконання конкретних завдань, що призвело до відхилень від біології.

Багатошаровий перцептрон (MLP) – це клас нейронних мереж. MLP складається щонайменше з трьох шарів вузлів. Окрім вхідних вузлів, кожен вузол є нейроном, який використовує нелінійну функцію активації. MLP використовує контрольовану методику навчання, яка називається зворотне поширення. Його кілька шарів і нелінійна активація відрізняють MLP від лінійного перцептрона. Він може розрізнити дані, які не є лінійно відокремлюваними [12].

Навчання під контролем - завдання машинного навчання, яке дозволяє вивчати функцію, яка відображає вхідний сигнал на виході на прикладі вхідних та вихідних пар. Він виконує функцію від помічених навчальних даних, що складаються з набору навчальних прикладів. У керованому навчанні кожен приклад - це пара, що складається з вхідного об'єкта (як правило, вектора) та бажаного вихідного значення (також називається контрольний сигнал). Наглядний алгоритм навчання аналізує тренувальні дані та видає передбачувану функцію, яку можна використовувати для відображення нових прикладів. Оптимальний сценарій дозволить алгоритму правильно визначити мітки класів для нових даних без міток [13].

## 3 РОЗРОБКА ІГРОВОГО ДОДАТКУ

### 3.1 Задача комп'ютерного розпізнавання жестів

Жест – це рух чи дія для вираження емоції чи інформації замість розмови, або під час розмови. Дія або рух людського тіла або його частини, що має певне значення або сенс, тобто є знаком або символом.

Мова жестів багата на способи вираження людьми найрізноманітніших емоцій і значень, наприклад, образи, ворожості, дружелюбності або схвалення по відношенню до інших. Більшість людей використовують при розмові жести і мову тіла на додаток до слів. Багато жестів використовуються людьми підсвідомо. Вважається, що деякі етнічні групи використовують жести частіше інших, і культурно прийнятний обсяг жестикуляції різниться від одного місця до іншого.

Постановка завдання розпізнавання жестів комплексна і враховує неоднозначну природу статичних і динамічних жестів, проблеми виділення руки на навколишньому фоні, умови освітлення і «шуми» на відео- або фото-матеріалі. Вирішення задачі в більшості випадків передбачає вибір алгоритмів з використанням комп'ютерних ресурсів.

Залежно від типу вхідних даних підхід для інтерпретації жестів може бути виконаний різними шляхами. Однак більшість методів засновані на ключових показниках, представлених в тривимірній системі координат. Ґрунтуючись на їх відносному русі, жест може бути виявлений з високою точністю, в залежності від якості вхідних даних і підходу алгоритму.

Щоб інтерпретувати рух тіла, потрібно класифікувати його за загальними властивостями і повідомленнями, які можуть висловлювати рухи. Наприклад, в мові жестів кожен жест являє собою слово або фразу.

На рисунку 3.1 наведені деякі з статичних жестів, які можливо діагностувати з аналізу відео. Видно, що для діагностики жестів необхідно фіксувати положення руки і пальців, при цьому виникає безліч однотипних

ситуацій, які можуть бути представлені як жест, однак не будучи таким. При аналізі жестів необхідно зрозуміти, які може приймати положення і форми кисть, щодо стоїть / сидячої людини.



Рисунок 3.1 – Положення долоні і пальців під час відображення жестів

Кисть має п'ять пальців (це визначає набір жестів і відповідних алгоритмів розпізнавання). Розглянемо можливі комбінації з пальцями (див. рис. 3.1). Якщо долоню стиснути в кулак і окремо витягувати кожен палець, то вийде шість варіантів: кулак, відкритий великий, вказівний, середній, безіменний з мізинцем, мізинець. Далі можна збирати різні комбінації, при деяких комбінаціях розпрямляється третя фаланга стислих пальців (якщо це дозволяють кистьові суглоби). Звичайно, розтягнення зв'язок і обертання в суглобах індивідуально, так що тут наведені усереднені дані. Для пальців існує 2 позиції: стиснутий, розігнутий. Також кути повороту вибираються кратними  $90^\circ$  (це робиться для полегшення обробки даних).

У деяких літературних джерелах відрізняють два різних підходи розпізнавання жестів: на основі 3D-моделі і на основі зовнішнього вигляду [14]. Передові методи для розпізнавання жестів використовують тривимірну інформацію про ключові елементи частин тіла, щоб отримати кілька важливих параметрів, таких як положення долоні або кути суглоба. З іншого боку, системи на основі зовнішнього вигляду використовують зображення або відео для прямої інтерпретації.

Підхід 3D-моделі може використовувати об'ємні або скелетні моделі або навіть їх комбінацію. Об'ємні підходи широко використовувалися в індустрії комп'ютерної анімації та для цілей комп'ютерного зору. Моделі зазвичай

створюються з складних тривимірних поверхонь, таких як NURBS або полігональні сітки.

Замість інтенсивної обробки тривимірних моделей і роботи з великою кількістю параметрів можна просто використовувати спрощену версію параметрів кутів з'єднання разом з довжинами сегментів. Цей метод відомий як метод скелетного уявлення тіла, де обчислюється віртуальний скелет людини, і частини тіла відображаються на певні сегменти. Аналіз тут виконується з використанням положення і орієнтації цих сегментів і відносин між кожними з них.

Моделі на основі 2D відображення об'єкта не використовують просторове уявлення тіла, тому що вони отримують параметри безпосередньо із зображень або відео, використовуючи базу даних шаблонів [15]. Деякі з них засновані на деформуємих двомірних шаблонах людських частин тіла, особливо рук. Деформуємі шаблони – це набори точок на контурі об'єкта, які використовуються в якості вузлів інтерполяції для апроксимації контуру об'єкта. Одна з найпростіших інтерполяційних функцій – лінійна, яка виконує усереднення форми за наборами точок, параметрами мінливості точок і зовнішніми деформаторами. Ці засновані на шаблонах моделі в основному використовуються для відстеження рук, але можуть також використовуватися для простої класифікації жестів.

Головний недолік методів на основі просторового уявлення тіла полягає в тому, що вони вимагають значних обчислювальних ресурсів, а також для роботи з ними необхідна камера з можливістю визначення глибини зображення. Тому для невеликих прикладних програм раціональніше використовувати методи, засновані на 2D поданні об'єкта, так як в них знижується обчислювальна складність, а також відпадає необхідність в спеціальному обладнанні: для отримання зображення може бути використана звичайна камера мобільного пристрою. Оскільки 2D методи можуть бути менш точними, перспективним напрямком досліджень буде їх удосконалення для отримання не ресурсоємного і ефективного алгоритму розпізнавання жестів рук.



Для локалізації жестів на елементах відеопослідовності часто використовується метод Віоли-Джонса. Цей підхід заснований на принципах інтегрального представлення зображень, побудови класифікаторів на основі адаптивного посилення і каскадного комбінування класифікаторів. Даний метод вимагає наявності навчальної вибірки зображень, володіє низькою швидкістю навчання класифікаторів, проте відрізняється високою швидкістю роботи, що дозволяє його використовувати в системах реального часу. Важливо відзначити, що застосовність методу Віоли-Джонса обмежується кінцевим безліччю можливих жестів для розпізнавання (відповідно до навчальної вибірки). Алгоритм добре працює при невеликому куті відхилення, однак при відхиленні кута більше 30 градусів якість результатів різко падає. В якості альтернативи може використовуватися метод сегментації зображення на основі колірного кластера шкіри, який має низьку обчислювальну вартість, однак не забезпечує достатньої робастності жестів при наявності схожих за кольором об'єктів на вхідній відеопослідовності. Якість роботи даного методу також сильно залежить від умов навколишнього середовища.

Виділення ключових ознак може здійснюватися за допомогою вейвлет-перетворення. Цей підхід показав совою ефективність у багатьох задачах обробки зображень. Важливо відзначити, що подібний підхід також вимагає чітко визначеного домену можливих до розпізнавання жестів і попереднього навчання.

### **3.2 Алгоритм для розпізнавання жестів**

Одним з перспективних методів розпізнавання образів вважається метод Віоли-Джонса – алгоритм, що дозволяє виявляти об'єкти на зображеннях в реальному часі. Цей метод запропонований в 2001 році Полом Віолою і Майклом Джонсом. Даний метод є основоположним для пошуку об'єктів на зображенні в реальному часі в більшості існуючих алгоритмів розпізнавання та

ідентифікації. Також він є одним з кращих по співвідношенню ефективності розпізнавання і швидкості роботи. Алгоритм показує відмінні результати і розпізнає об'єкти під невеликим кутом, приблизно до 30 градусів, і при різних умовах освітленості [16].

Основні принципи, на яких базується метод:

- можливість подання зображення в інтегральному вигляді, що дозволяє швидко обчислювати необхідні об'єкти;
- використання ознак Хаара, тобто ознак цифрового зображення, які використовуються в розпізнаванні образів, за допомогою яких відбувається пошук потрібного об'єкта;
- застосування алгоритму бустінга (від англ. boost- поліпшення, посилення) для вибору найбільш підходящих ознак для шуканого об'єкта на даній частині зображення, процедури послідовного побудови композиції алгоритмів машинного навчання, коли кожен наступний алгоритм прагне компенсувати недоліки композиції всіх попередніх алгоритмів;
- використання каскадів ознак для швидкого відкидання вікон, де не знайдений об'єкт.

Метод Віоли-Джонса з використанням ознак Хаара є одним з кращих алгоритмів для вирішення задач виявлення об'єктів на зображеннях в реальному часі, використовуючи двомірні зображення. Особливістю ознак Хаара, є найбільша, в порівнянні з іншими ознаками, швидкість, тому, для вирішення завдання побудови алгоритму виявлення руки в реальному часі, візьмемо за основу метод Віоли-Джонса.

Першочерговим завданням у багатьох методах розпізнавання об'єктів в реальному часі є інтегральне представлення зображення. Воно використовується, і в вейвлет-перетвореннях – інтегральних перетвореннях, які є згортку функції з сигналом, і в фільтрах Хаара – дозволяють аналізувати різні частотні компоненти даних, в SURF – методі пошуку особливих точок зображення і створення їх дескрипторів, інваріантних до масштабу і обертання, а також багатьох інших алгоритмах.

Інтегральне уявлення дозволяє швидко розраховувати сумарну яскравість довільного прямокутника, причому час розрахунку не залежить від площі прямокутника. Інтегральне представлення зображення являє собою матрицю, розмірність якої збігається з розмірністю вихідного зображення. У кожному елементі матриці зберігається сума інтенсивностей всіх пікселів, що знаходяться лівише і вище даного елемента. Елементи матриці розраховуються за такою формулою:

$$I(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (3.1)$$

де  $I(i, j)$  — яскравість пікселя вхідного зображення.

Кожен елемент матриці  $[I, J]$  являє собою суму пікселів в прямокутнику від  $(0,0)$  до  $(I, J)$ , тобто значення кожного пікселя  $(i, j)$  дорівнює сумі значень усіх пікселів лівише і вище даного пікселя  $(i, j)$ . Розрахунок матриці займає лінійний час, пропорційне числу пікселів в зображенні, тому інтегральне зображення прораховується за один прохід.

Розрахунок матриці можна виробляти за рекурентною формулою:

$$I(x, y) = I(x, y-1) + I(x-1, y) - I(x-1, y-1) \quad (3.2)$$

За такою інтегральною матрицею можна дуже швидко вирахувати суму пікселів довільного прямокутника, довільної площі.

Нехай в прямокутнику ABCD є цікавий для нас об'єкт D (див. рис. 3.2).

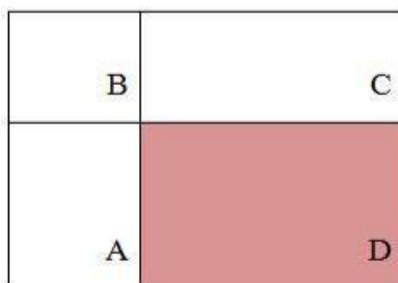


Рисунок 3.2 – Прямокутник ABCD

З рисунка зрозуміло, що суму усередині прямокутника можна виразити через суми і різниці суміжних прямокутників за такою формулою:

$$I(x, y) = I(x, y+1) + I(x, y-1) - I(x, y+1) - I(x, y-1). \quad (3.3)$$

З точки зору необхідності використання досить простих алгоритмів отримання ознак, перспективним є застосування хаар-подібних характеристик (Haar wavelet-like features), що представляють собою результат порівняння яскравості в двох прямокутних областях зображення. Зокрема, як уже зазначалося вище, Віола і Джонс запропонували використовувати три види характеристик. Значенням характеристики з двох прямокутників є різниця між сумою пікселів в цих прямокутних областях. Области мають однаковий розмір і форму і по горизонталі і по вертикалі.

Каскад Хаара – це набір ознак, для яких вважається їх згортка з зображенням. Ознака Хаара – відображення  $f: \mathcal{X} \rightarrow \mathcal{Y}$ , де  $\mathcal{Y}$  – множина допустимих значень ознаки. Якщо задані ознаки  $f_1, \dots, f_n$ , то вектор ознак  $\mathbf{f} = (f_1(x), \dots, f_n(x))$  називається ознаковим описом об'єкта  $x \in \mathcal{X}$ . Ознакові описи допустимо ототожнювати з самими об'єктами. При цьому множина  $\mathcal{F} = \mathcal{X} \times \mathcal{Y}^n$  називають ознаковим простором.

Обчислюваним значенням такої ознаки буде:

$$f(x) = \frac{1}{|A|} \sum_{(x,y) \in A} I(x,y) - \frac{1}{|B|} \sum_{(x,y) \in B} I(x,y), \quad (3.4)$$

де  $|A|$  – сума значень яскравості точок закриваються світлою частиною ознаки, а  $|B|$  – сума значень яскравості точок закриваються темної частиною ознаки.

Для їх обчислення використовується поняття інтегрального зображення, розгляньте вище.

Ознаки поділяються на такі типи в залежності від множини  $\mathcal{Y}$ :

– бінарна ознака,  $\mathcal{Y} = \{0,1\}$ ;

– номінальна ознака:  $\mathcal{Y}$  – скінченна множина;

- порядкова ознака:— кінцева впорядкована множина;
- кількісна ознака:— множина дійсних чисел.

Звісно, бувають прикладні завдання з різнотипними ознаками, для їх вирішення підходять далеко не всі методи.

У стандартному методі Віоли - Джонса використовуються прямокутні ознаки, зображені на малюнку нижче, вони називаються примітивами Хаара (див. рис. 3.3).

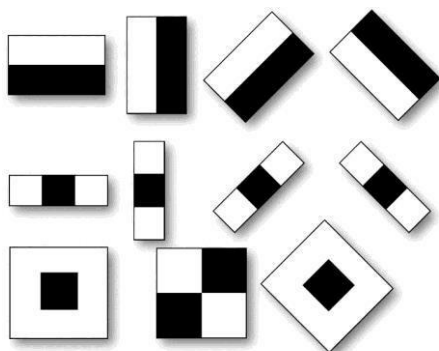


Рисунок 3.3 – Примітиви Хаара

У розширеному методі Віоли-Джонса, що використовується в бібліотеці OpenCV використовуються додаткові ознаки [17] (див. рис. 3.4).

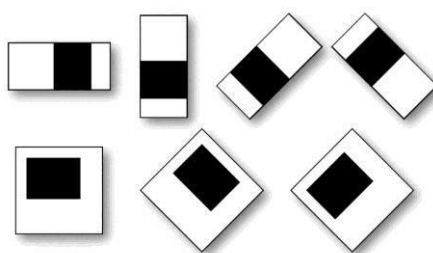


Рисунок 3.4 – Додаткові ознаки

Ознаки Хаара дають точкове значення перепаду яскравості по осі и відповідно.

Для вибору найбільш придатних ознак для шуканого об'єкта на частині зображення використовують алгоритм бустінга. Даний метод дозволяє посилити

прості класифікатори шляхом комбінування примітивних «слабких» класифікаторів в один «сильний». Під «силою» класифікатора в даному випадку мається на увазі ефективність (якість) рішення задачі класифікації.

Розвитком даного підходу є розробка більш досконалого сімейства алгоритмів бустінга - AdaBoost (adaptive boosting - адаптоване поліпшення), запропонована Йоавом Фройндом (Freund) і Робертом Шапіро (Schapire) в 1999 році, який може використовувати довільне число класифікаторів і виробляти навчання на одному наборі прикладів, по черзі застосовуючи їх на різних етапах.

В результаті роботи алгоритму бустінга на кожній ітерації формується простий класифікатор виду:

$$h(z) = \begin{cases} 1, & \text{якщо } (z) < \tau, \\ 0, & \text{інакше.} \end{cases}$$

В де  $z$  - вікно зображення розміром  $20 \times 20$  пікселів;  $\tau$  - показує напрямок знака нерівності;  $(z)$  - обчислене значення ознаки;  $\tau$  - значення порога.

Отриманий класифікатор має мінімальну помилку по відношенню до поточних значень ваг, задіяним в процедурі навчання для визначення помилки.

Для підвищення швидкості виявлення використовується каскадна структура, фокусуються свою роботу на найбільш інформативних областях зображення. Каскад складається з шарів, які представляють собою класифікатори, навчені за допомогою процедури бустінга.

Алгоритм CAMShift був створений Гаррі Брадскі в 1998 р і здатний відстежувати особи. Він комбінує алгоритм відстеження об'єкта Mean Shift, заснований на мапі ймовірності кольору шкіри, з адаптивним кроком зміни розміру області відстеження. Ймовірність кольору шкіри кожного пікселя зображення визначається методом Histogram Backprojection, заснованим на кольорі, представленим у вигляді колірної тону (Hue) моделі HSV. Так як алгоритм CAMShift здатний відстежувати особи на основі ймовірності кольору

шкіри, то він може застосовуватися для відстеження руки.

Перевагами даного алгоритму є: низькі вимоги до обчислювальних ресурсів, гнучкі настройки точності позиціювання, можливість роботи в різних умовах освітленості. Також додатковою перевагою алгоритму є можливість роботи в умовах часткового перекриття відслідковується об'єкта. Зазначені вище властивості алгоритму обумовлені використанням моделі об'єкта, побудованої на основі гістограми яскравості і кольору, а також використанням процедури Mean Shift для точного позиціонування положення об'єкта.

Вейвлет-перетворення широко використовується для аналізу нестационарних процесів. Воно показало свою ефективність для вирішення широкого класу задач, пов'язаних з обробкою зображення. Коефіцієнти вейвлет-перетворення містять інформацію про аналізованому процесі і використовуюваному вейвлет. Тому вибір аналізує вейвлета визначається тим, яку інформацію необхідно витягти з процесу. Кожен вейвлет має характерні особливості в тимчасовій і частотній областях, тому іноді за допомогою різних вейвлетів можна повніше виявити і підкреслити ті чи інші властивості аналізованого процесу.

Метод головних компонент (Principal Component Analysis, PCA) - один з найбільш поширених методів для зменшення розмірності даних, втрати найменшої кількості інформації. Він полягає в лінійному ортогональному перетворенні вхідного вектора  $P$  розмірності  $N$  у вихідний вектор  $Q$  розмірності  $M$ ,  $M < N$ . Компоненти вектора  $Q$  є некорельованими, і загальна дисперсія після перетворення залишається незмінною.

Обчислення головних компонент зводиться до обчислення власних векторів і власних значень коваріаційної матриці, яка розраховується для зображення. Сума головних компонент, помножених на відповідні власні вектори, є реконструкцією зображення. Для кожного зображення об'єкта обчислюються його головні компоненти. Зазвичай береться від 5 до 200 головних компонент. Інші компоненти кодують дрібні відмінності між об'єктами і шум. Процес розпізнавання полягає в порівнянні головних компонент

невідомого зображення з компонентами всіх відомих зображень. З бази даних вибираються зображення-кандидати, що мають найменшу відстань від вхідного (невідомого) зображення.

Метою даної роботи є розпізнавання жестів на цифрових зображеннях. Для вирішення цієї задачі запропоновано новий алгоритм, заснований на застосуванні вейвлет-перетворення і методу головних компонент. Запропонований алгоритм складається з двох процесів: добування і збереження ознак відомих жестів в базі даних і розпізнавання жестів. Процес вилучення і збереження ознак відомих жестів відбувається наступним чином:

- перетворення зображення області жесту в напівтонове;
- зміна розміру області жесту до  $64 \times 64$  пікселів;
- застосування до отриманого зображення вейвлет-перетворення для вилучення ознак жесту (вейвлет-коефіцієнтів);
- збереження витягнутих ознак в базі даних.

У процесі розпізнавання невідомого жесту здійснюються кроки 1-3, потім отримані ознаки порівнюються з ознаками, що зберігаються в базі даних, на основі застосування методу головних компонент. Функціональна схема запропонованого алгоритму представлена на рис 3.5.

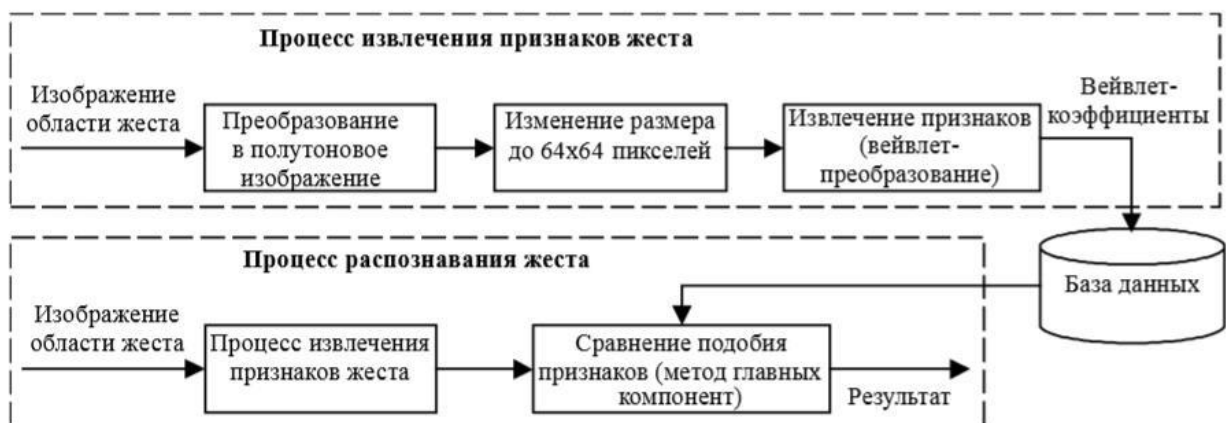


Рисунок 3.5 – Схема алгоритму розпізнавання жестів

У даній роботі також розглядається задача розпізнавання жестів на



відеопослідовності в режимі реального часу. Для вирішення цієї задачі запропоновано оригінальний комплексний алгоритм, заснований на застосуванні методу Віюлі - Джонса, алгоритму CAMShift, вейвлет-перетворення і методу головних компонент. Процес розпізнавання жестів на відеопослідовності відбувається наступним чином.

Крок 1 – запит чергового відеофрейма. Перетворення відеофрейма в півтонове зображення. Застосування до напівтонового зображення методу Віюлі-Джонса для пошуку області руки.

Крок 2 – якщо область руки виявлена, то виконується крок 3. У зворотному випадку здійснюється повернення на крок 1.

Крок 3 – запит чергового відеофрейма. Відстеження області руки на основі застосування алгоритму CAMShift.

Крок 4 – якщо відстеження здійснено, то виконується крок 5. У зворотному випадку відбувається повернення на крок 1.

Крок 5 – виконання процесу розпізнавання жесту (рис. 3.5).

Крок 6 – повернення на крок 3.

В результаті даний алгоритм дозволяє зменшити ймовірність помилкових виявлень і підвищити швидкодію роботи системи, допомагає позбутися від обмежень у вигляді перешкод на зображенні за допомогою попередньої обробки зображення, а також позбутися від проблем, пов'язаних з кутом нахилу руки шляхом тренування нових каскадів.

Даний підхід є універсальним, розширюваним і якщо використовувати достатню кількість каскадів класифікаторів, то метод знаходження працює дуже швидко і практично безпомилково.

### 3.3 Дизайн ігрового додатку

На сьогоднішній день ринок розваг є одним з найприбутковіших. З моменту початку інформаційно-технічної революції світ стрімко рухається в майбутнє, створюючи все більш досконалі комп'ютерні системи, щоб полегшити життя людини, а так само зайняти його дозвілля.

З появою персональних комп'ютерів, з кожним роком, їх роль в житті людей постійно зростає. Комп'ютер став незамінним помічником не тільки в сфері економічних розрахунків, а й є потужним центром розваг. Сфери впливу кіно і літератури відчувають потужний тиск з боку інтерактивних розваг, пристроїв доповненої реальності і інших, впроваджуваних завдяки розвитку комп'ютерних технологій. Сучасна людина іноді навіть не замислюється, що і його смартфон є аналогічним мобільним персональним комп'ютером. З ростом ролі комп'ютерів в житті людини, комп'ютерна техніка суттєво впливає і на модель поведінки людини. За останніми дослідження середній вік гравця комп'ютерних ігор становить 30 років і вище.

Комп'ютерні розваги роблять життя людини багатше, більш насиченим і як наслідок – це потужна економічна сфера приносить величезні доходи. Тому не випадково, що особлива роль у житті сучасної людини відводиться комп'ютерним іграм, перші з яких існували на самій зорі комп'ютерної техніки.

Пристрої вводу, що використовуються для ігор, залежать від платформи. Звичайні контролери включають геймпади, джойстики, комп'ютерні миші, клавіатури, сенсорні екрани мобільних пристроїв або навіть тіло людини з використанням датчика Kinect. Гравці можуть переглядати гру на пристрої відображення, наприклад на телевізорі або моніторі комп'ютера, а іноді і на віртуальних шоломах віртуальної реальності. Деякі ігри 2000-х років включають тактильні пристрої, що створюють вібрацію ефекти, периферію зі зворотним зв'язком і гарнітури віртуальної реальності.

У 2010-х комерційна важливість індустрії відеоігор зростає. Азіатські ринки, зокрема мобільні ігри для смартфонів, стимулюють зростання галузі. Станом на 2018 рік продажі відеоігор в усьому світі становили 134,9 млрд. дол. США в рік і були третім за величиною сегментом на ринку розваг в США після радіо і кабельного телебачення [18].

Задля демонстрації роботи алгоритму розпізнавання жестів завданням даної роботи є створення ігрового додатку для мобільних платформ, що міг би використовувати жести людської руки в як основну ігрову механіку. Вибір саме мобільної платформи обґрунтований, по-перше, популярністю мобільних пристроїв серед сучасних гравців, а по-друге, тим фактом, що зараз кожен смартфон має потужну камеру, що може бути використана для захвату відеопотоку.

Ідея мобільного додатку полягає в створенні мобільної гри на основі відомої усім дитячої гри «камінь-папір-ножиці». Камінь-папір-ножиці – це ручна гра, в яку зазвичай грають дві людини, в якій кожен гравець одночасно формує одну з трьох фігур з витягнутою рукою (див. рис. 4.1).

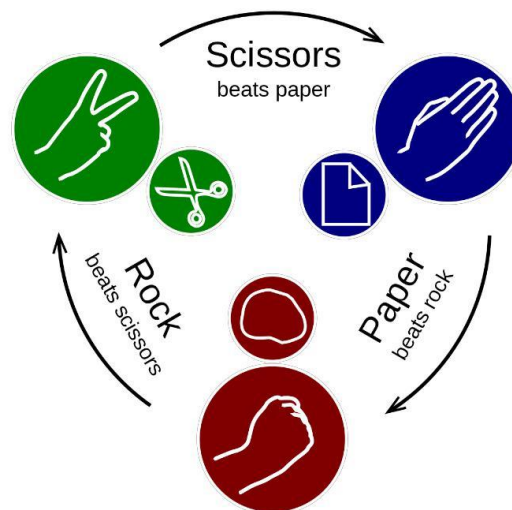


Рисунок 3.8 – Схема гри «камінь-папір-ножиці»

Цими формами є «камінь» (закритий кулак), «папір» (плоска рука) і «ножиці» (кулак з витягнутим вказівним і середнім пальцями, утворюючи V). «Ножиці» ідентичні двухпалому знаку V (також позначає «перемогу» або «світ»)

за винятком того, що він спрямований горизонтально, а не тримається вертикально в повітрі. Одночасна гра з нульовою сумою має тільки два можливих результату: нічия або виграш для одного гравця і програш для іншого.

Камінь-ножиці-папір часто використовується в якості методу справедливого вибору між двома людьми, схожого на підкидання монет, малювання соломинок або кидання костей для врегулювання спору або прийняття неупередженого групового рішення. Однак, на відміну від дійсно випадкових методів відбору, рок-папір-ножиці можна грати з певним ступенем майстерності, розпізнаючи і використовуючи не випадкове поведінку супротивників.

У багатьох іграх для групи можливих варіантів характерно взаємодія в стилі «камінь-папір-ножиці», де кожен вибір сильний проти певного вибору, але слабкий проти іншого. Така механіка може зробити гру в деякій мірі збалансованою і не допустити перевантаженості ігрового процесу однією домінуючою стратегією.

Проектуємий ігровий додаток буде використовувати дану ігрову механіку, дещо доповнивши її задля цікавості ігрового процесу.

За сюжетом гри користувач бере на себе роль чаклуна, що може використовувати природні стихії задля участі у турнірах з іншими чаклунами. Усього доступно п'ять стихій: вогонь, дерево, вода, метал та земля.

Ігровий процес гри досить простий. Гравець може розпочати магічний турнір з іншим гравцем або комп'ютером. Він використовує камеру свого мобільного пристрою, щоб увімкнути відеотрансляцію. Після цього гравцю необхідно рукою виконати один із жестів стихій, так, щоб камера мобільного пристрою захватила зображення долоні гравця.

Жестом руки гравець показує символ обраної стихії. Жести долоні для стихій приведені на рисунку 4.2. Після цього додаток починає аналіз жесту на відеопослідовності. Якщо жест не був розпізнаний коректно, або на відеопослідовності не було долоні, раунд вважається зіграним внічию і починається спочатку.



Рис. 4.2 – Жести стихій. Зліва направо: вогонь, дерево, вода, метал, земля.

Алгоритм розпізнавання приймає на вхід відеопослідовність з жестом гравця, після чого відносить жест до одного із п'яти класів – стихій. Після цього за схемою ієрархії стихій (рис. 4.3) визначається переможець раунду. За наведеною схемою, по аналогії з грою «камінь-папір-ножиці», магічні елементи мають переваги один перед іншим. Вода перемагає вогонь, вогонь перемагає дерево, дерево перемагає землю і т.п.

Після трьох раундів (або більше, якщо три були зіграні в нічию) визначається переможець турніру. Після цього гравець виходить із турніру та потрапляє у головне меню гри.

Гравець може отримувати і переглядати різноманітні досягнення у спеціальному меню, а також заробляти бали успіху. В меню рейтингу гравців можна відстежувати власний рейтинг та змагатися з іншими гравцями за звання найсильнішого чаклуна.

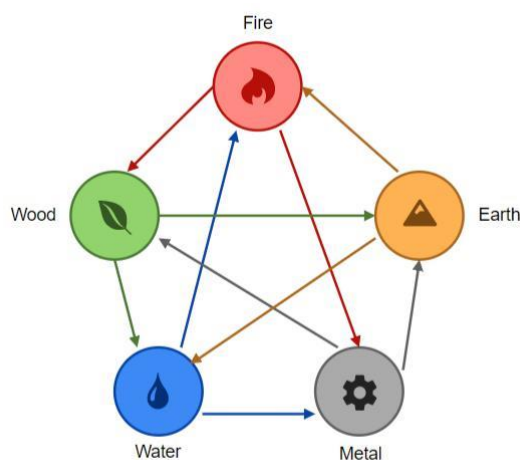


Рис. 4.3 – Ієрархія стихій

Таким чином, моделюємий ігровий додаток являє собою дещо розширену версію «камінь-папір-ножиці» у східному магічному сеттінгу, що використовує камеру мобільного пристрою для розпізнавання жестів на відеопослідовності. Гра дуже проста для розуміння, демонструє можливості використання машинного зору та допоможе майбутнім користувачам весело скоротати час.

### 3.4 UML проектування програмної системи

Моделювання розроблюваної системи проводиться з використанням мови UML для побудови діаграм, що допоможуть відобразити функціональність та внутрішню структуру системи. UML – мова графічного опису для об'єктного моделювання в області розробки програмного забезпечення. UML є мовою широкого профілю, це відкритий стандарт, який використовує графічні позначення для створення абстрактної моделі системи, що називається UML-моделлю.

Можна виділити наступні переваги UML:

– UML об'єктно-орієнтований, в результаті чого методи опису результатів аналізу і проектування семантично близькі до методів програмування на сучасних об'єктно-орієнтованих мовах;

– діаграми UML порівняно прості для читання після досить швидкого ознайомлення з його синтаксисом;

– UML розширює і дозволяє вводити власні текстові та графічні стереотипи.

При проектуванні системи були розроблені діаграми наступних типів: діаграма діяльності та діаграма станів.

Діаграма діяльності (англ. activity diagram) — в UML, візуальне представлення графу діяльностей. Граф діяльностей є різновидом графу станів скінченного автомату, вершинами якого є певні дії, а переходи відбуваються по завершенню дій [19].

Дія є фундаментальною одиницею визначення поведінки в специфікації. Дія отримує множину вхідних сигналів, та перетворює їх на множину вихідних сигналів. Одна із цих множин, або обидві водночас, можуть бути порожніми. Виконання дії відповідає виконанню окремої дії. Подібно до цього, виконання діяльності є виконанням окремої діяльності, буквально, включно із виконанням тих дій, що містяться в діяльності. Кожна дія в діяльності може виконуватись один, два, або більше разів під час одного виконання діяльності. Щонайменше, дії мають отримувати дані, перетворювати їх та тестувати, деякі дії можуть вимагати певної послідовності. Специфікація діяльності (на вищих рівнях сумісності) може дозволяти виконання декількох (логічних) потоків, та існування механізмів синхронізації для гарантування виконання дій у правильному порядку.

На діаграмі діяльності (див. рис. 4.4) зображений процес, коли користувач заходить у гру. Для цього користувачу необхідно спочатку авторизуватись у системі. Щоб авторизуватись, необхідно зареєструватись. Потім користувач опиняється у головному меню, де може вибирати – переглянути власні досягнення, подивитись на рейтинг гравців або почати ігрову сесію з іншим гравцем або комп'ютером. Після вибору «почати гру» автоматично запускається меню ігрового турніру.

Діаграма станів — орієнтований граф для кінцевого автомата, в якому: вершини позначають стани, а дуги показують переходи між двома станами [20].

Діаграма станів (див. рис. 4.5) у UML 2.0 значно розширює поняття кінцевого автомата, що його розробляють у класичних комп'ютерних науках. Вона вводить таке поняття, як ієрархічні стани, що є близькими до поняття наслідування в об'єктно-орієнтовних мовах програмування.

Використовуються такі умовні позначення:

- коло, що позначає початковий стан;
- коло з маленьким колом усередині, що позначає кінцевий стан (якщо існує);
- округлений прямокутник, що позначає стан; верхівка прямокутника

- містить назву стану. В середині може бути горизонтальна лінія, під якою записуються активності, що відбуваються в даному стані;
- стрілка, що позначає перехід; назва події (якщо є), що викликає перехід, відзначається поруч зі стрілкою.

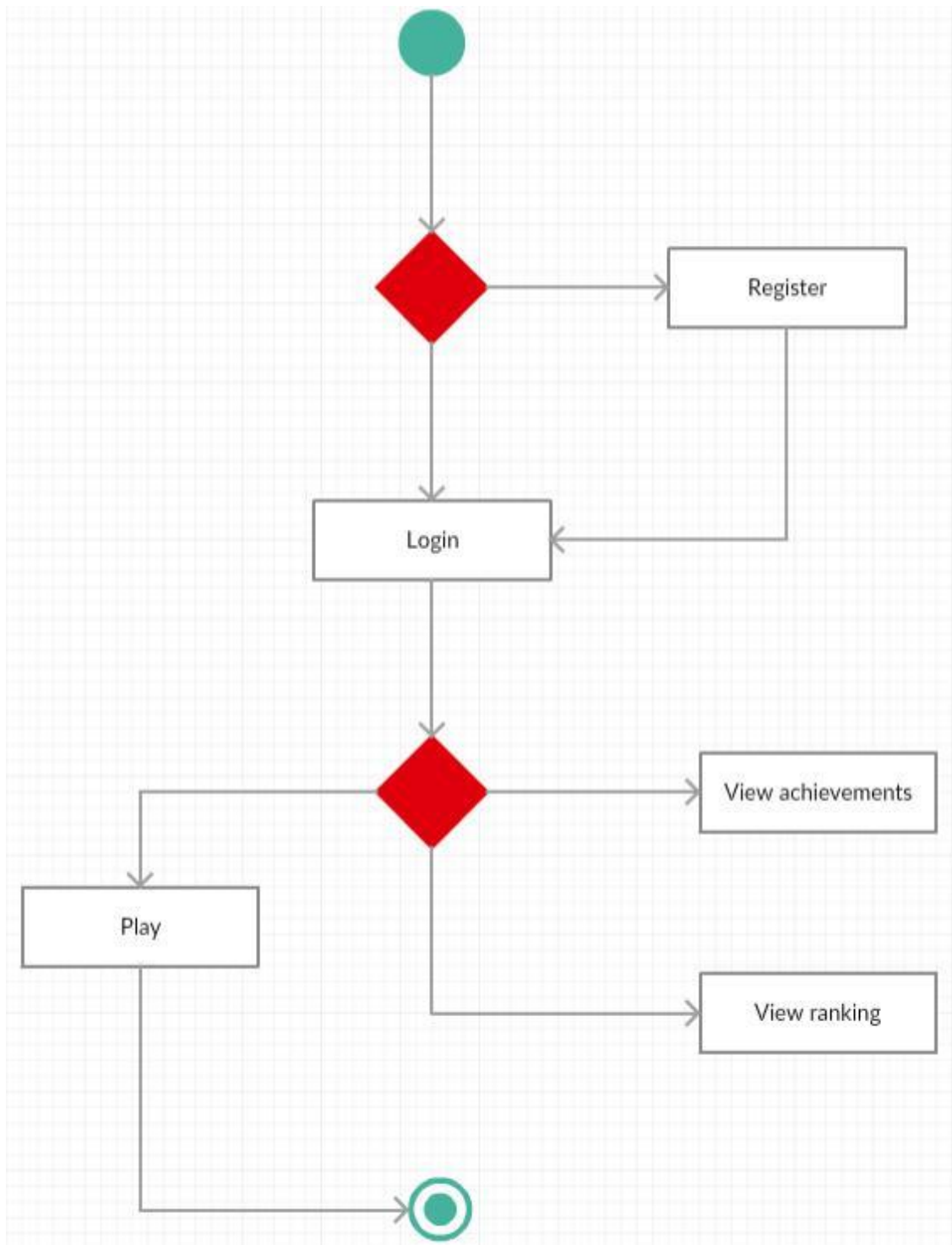


Рисунок 3.4 – Діаграма діяльності



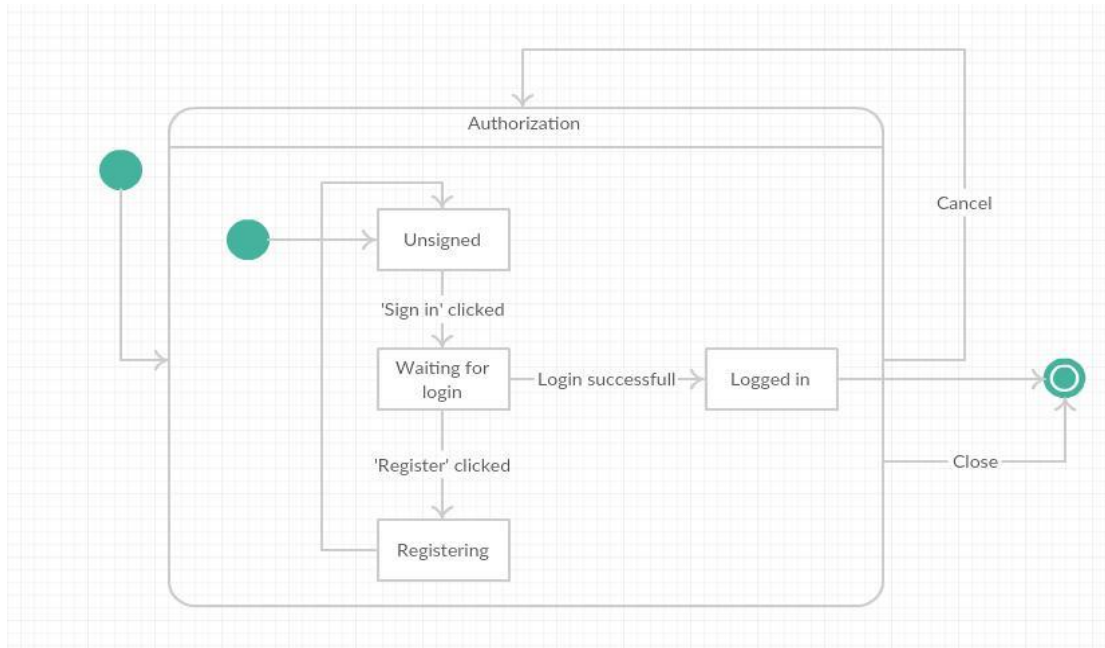


Рисунок 3.5 – Діаграма станів

Діаграма станів на рисунку 3.5 моделює процес реєстрації.

### 3.5 Створення UI\UX

UX /UI дизайн – це проектування будь-яких призначених для користувача інтерфейсів в яких зручність використання так само важлива як і зовнішній вигляд.

Інтерфейс користувача (англ. user interface, UI) — засіб зручної взаємодії користувача з інформаційною системою. Сукупність засобів для обробки та відображення інформації, максимально пристосованих для зручності користувача; у графічних системах інтерфейс користувача реалізовується багатівіконним режимом, змінами кольору, розміру, видимості (прозорість, напівпрозорість, невидимість) вікон, їхнім розташуванням, сортуванням елементів вікон, гнучкими налаштуваннями як самих вікон, так і окремих їхніх елементів (файли, папки, ярлики, шрифти тощо), доступністю багатокористувацьких налаштувань.

UX – це User Experience (дослівно: «досвід користувача»). Тобто це те, який досвід / враження отримує користувач від роботи з інтерфейсом. Чи вдається йому досягти мети і на скільки просто або складно це зробити.

Мобільний ігровий додаток має досить простий інтерфейс (див. рис. 4.6).

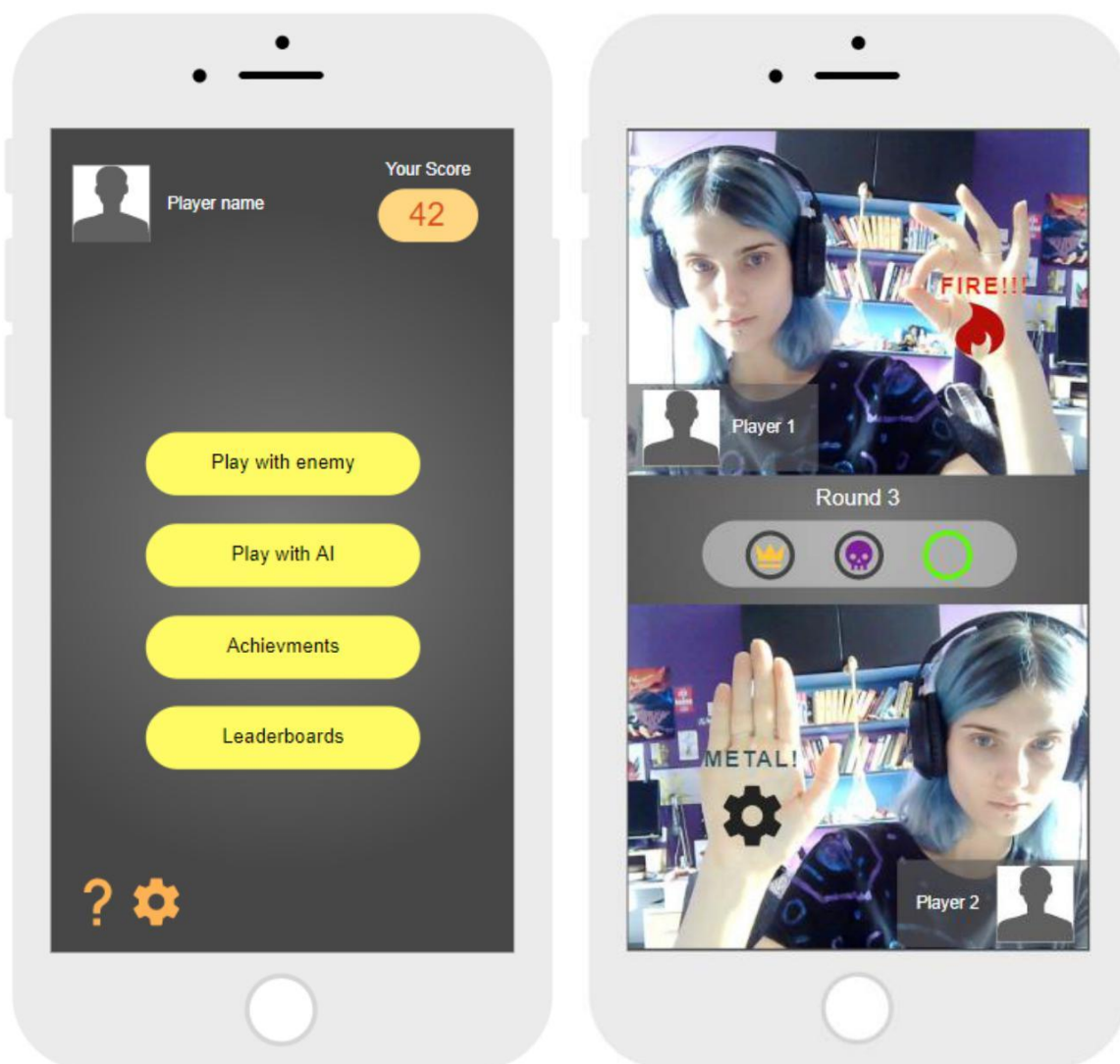


Рисунок 4.6 – Мокапи інтерфейсу мобільного додатку

Під час першого входу у гру користувач має змогу авторизуватися чи створити новий аккаунт, а також йому доступні налаштування гри та довідка. Після авторизації чи реєстрації користувач може почати використовувати програмний додаток.

У головному меню гри відображаються відомості про гравця, тобто його аватар, ім'я та набрані ігрові бали. З головного меню доступно два варіанта гри:

з іншим гравцем або з комп'ютером. Також є можливість перейти у меню ігрових досягнень або переглянути рейтинги найуспішніших гравців.

На другому зображенні показано вигляд основного геймплейного вікна. На екрані мобільного пристрою відображаються два відео: гравця та його супротивника. Обоє використовують жести для відображення символу одного із магічних елементів. Програмний додаток розпізнає жести та визначає переможця раунду. Після трьох раундів визначається переможець турніру, йому нараховуються ігрові бали, після чого гравець знову повертається у головне меню гри.

Приклади інтерфейсу були виконані у безкоштовному веб-додатку Moqups. Надалі інтерфейс ігрового мобільного додатку буде розроблятися, опираючись на дані зразки.

## ВИСНОВКИ

У ході атестаційної роботи магістра було проведено дослідження сучасних підходів та методів машинного зору у застосуванні до завдань розпізнавання образів; проведено аналіз предметної області; запропоновано алгоритм для розпізнавання жестів руки на відеопослідовностях, здатний працювати у режимі реального часу; спроектовано дизайн ігрового мобільного додатку для демонстрації можливості роботи алгоритму.

На початку роботи було проведено аналіз предметної області, в процесі якого були визначені основні проблеми, що існують в розглянутій галузі. Було проведено аналіз існуючих систем на основі машинного зору. На підставі аналізу предметної області була проведена постановка завдання.

У ході виконання поставленої задачі було:

- проведено аналіз предметної галузі;
- проведено аналіз існуючих підходів до розробки подібних застосувань та аналіз аналогів;
- проаналізовано методи та алгоритми машинного зору;
- запропоновано алгоритм для розпізнавання жестів на відеопослідовностях;
- виконано моделювання ігрового додатку для демонстрації роботи алгоритму.

На основі проведеного аналізу оптимальних алгоритмів і методів виявлення і розпізнавання жестів руки на відео в режимі реального часу було запропоновано метод, який дозволяє зменшити ймовірність помилкових виявлень і підвищити швидкодію роботи системи.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Szeliski, R. Computer Vision: Algorithms and Applications [Текст] / R.
2. Bishop, C. Pattern Recognition and Machine Learning [Текст] / C. Bishop –
3. Обработка и анализ изображений в задачах машинного зрения: Курс лекций и практических занятий / Ю. Визильтер, С. Желтов, А. Бондаренко, М. Ососков, А. Моржин. – М.: Физматкнига, 2010. – 672 с.
4. Шапиро Л., Стокман Д. Компьютерное зрение / Пер. с англ.-М.:БИНОМ. Лаборатория знаний, 2006. – 752 с.
5. Потапов А.С. Распознавание образов и машинное восприятие. - С-Пб.: Политехника, 2007. – 548 с.
6. Wang, C., Liu, Z., & Chan, S. (2015). Superpixel-based hand gesture recognition with Kinect depth camera. IEEE Transactions on Multimedia, Vol. 17, pp. 29-39.
7. Gonzalez R. C., Woods R.E., Digital Image Processing (4th Edition) – New York: Pearson, 2018. – 1148 с.
8. k-means++: The Advantages of Careful Seeding [Электронный ресурс] / Stanford Library. - Режим доступа: [www/ URL: http://ilpubs.stanford.edu:8090/778/1/2006-13.pdf](http://ilpubs.stanford.edu:8090/778/1/2006-13.pdf) - Загол. з екрану.
9. SURF: Speeded Up Robust Features [Электронный ресурс] / Computer Vision Laboratory, ETH Zurich. - Режим доступа: [www/ URL: http://www.vision.ee.ethz.ch/~surf/eccv06.pdf](http://www.vision.ee.ethz.ch/~surf/eccv06.pdf) - Загол. з екрану.
10. Bishop, C. Neural Networks for Pattern Recognition – London: Oxford University Press, 1995. – 576 с.
11. Fausett, L. V. Fundamental of Neural Networks: Architectures, Algorithms, and Applications – Paris: Prentice-Hall, 1993. – 463 с.
12. Girshick R., Donahue J., Darrell T., Malik J., Rich feature hierarchies for

accurate object detection and semantic segmentation [Текст] / R. Girshick, J. Donahue, T. Darrell, J. Malik, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference, 2014.

13. Krizhevsky A., Sutskever I., Hinton G. E., ImageNet Classification with Deep Convolutional Neural Networks [Текст] / A. Krizhevsky, I. Sutskever, G. E. Hinton, NIPS, 2012.

14. Vladimir I. Pavlovic, Rajeev Sharma, Thomas S. Huang. Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. – IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997.

15. Hand tracking and gesture recognition for human-computer interaction / [C. Manresa et al.] // Electronic Letters on Computer Vision and Image Analysis. – 2005.

16. Метод Виолы-Джонса (Viola-Jones) как основа для распознавания лиц [Электронный ресурс] – Режим доступа: [www/URL: //http://habrahabr.ru/post/133826/](http://habrahabr.ru/post/133826/) - Загол. з экрану.

17. OpenCV [Электронный ресурс] / OpenCV. – Режим доступа: [www/URL: https://opencv.org/](https://opencv.org/) - Загол. з экрану.

18. New Gaming Boom: Newzoo Ups Its 2017 Global Games Market Estimate to \$116.0Bn Growing to \$143.5Bn in 2020 | Newzoo. Newzoo. Retrieved December 15, 2017.

19. Фаулер, М. UML. Основы [Текст] : пер. с англ. А.: Петухов / М. Фаулер, К. Скотт. — СПб.: Символ, 2006. — 184 с.

20. Вікіпедія [Электронный ресурс] – Режим доступа: [www/URL: https://en.wikipedia.org/wiki/UML\\_state\\_machine](https://en.wikipedia.org/wiki/UML_state_machine) - Загол. з экрану.