

УДК 004.93

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Системотехніки
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

Рівень вищої освіти другий (магістерський)
ГЮИК. 502840.001 ПЗ

Дослідження методів розпізнавання об'єктів, що швидко рухаються
(тема)

Виконав:

Студент 2 курсу, групи СПРМ-19-2

Спеціальність 122 – Комп'ютерні науки
(код і повна назва напрямку)

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

Освітня програма Системне проектування
(повна назва освітньої програми)

Антонюк М.В.
(прізвище, ініціали)

Керівник Губаренко Є.В.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____
(підпис)

Гребеннік І. В.
(прізвище, ініціали)

2021 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Системотехніки
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 122 – Комп'ютерні науки
(код і повна назва)

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

Освітня програма Системне проектування
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

« _____ » _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Антонюку Максиму Васильовичу
(прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження методів розпізнавання об'єктів, що швидко рухаються» затверджена наказом по університету від «23» 03 2021 р. № 389Ст
2. Термін подання студентом роботи до екзаменаційної комісії 18 травня 2021 р.
3. Вихідні дані до роботи Стан системи: оператор налаштування системи, апаратне забезпечення системи, камери відеоспостереження, кадри зображень отримані з камер відеоспостереження.
4. Перелік питань, що потрібно опрацювати в роботі Вступ. Аналіз предметної області та постановка задачі. Попередня обробка зображень у відеопотоці. Навчання системи в відеопотоці та алгоритм розпізнавання зображень. Результати дослідження та експериментальні дані. Висновки.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслеників, плакатів, комп'ютерних ілюстрацій) Плакати на аркушах формату А4: причини розмиття образу об'єкта на зображенні, схема штучної нейронної мережі з трьома шарами, приклади головних компонент алгоритму Eigenfaces, приклад роботи алгоритму обмеженої машини Больцмана, співставлення інваріантного та еквіваріантного детектора локальних ознак, схема навчання моделі в потоці даних, схема алгоритму виділення локальних ознак, схема алгоритму розпізнавання.

6. Консультанти розділів роботи (проекту)

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

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання, аналіз завдання, уточнення плану роботи	23.03.2021	
2	Аналіз існуючих методів розпізнавання зображень в відеопотоці	26.03.2021	
3	Постановка задачі та вибір методу її вирішення	03.04.2021	
4	Проведення експериментальних досліджень	18.04.2021	
5	Оформлення пояснювальної записки	27.04.2021	
6	Підготовка презентації	11.05.2021	
7	Подання закінченої роботи науковому керівникові	12.05.2021	
8	Подання роботи на рецензування	12.05.2021	
9	Попередній захист	14.05.2021	
10	Подання роботи до комісії	17.05.2021	

Дата видачі завдання 23 березня 2021 р.

Студент

(підпис)

Антонюк М.В.

Керівник роботи

(підпис)

доц. Губаренко Є.В.

(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 85 с., 7 табл., 8 рис., 4 додатки, 47 джерел інформації.

ЕКВІВАРІАНТНИЙ ДЕТЕКТОР, ЗМЕНШЕННЯ РОЗМІРНОСТІ, ІНВАРІАНТНИЙ ДЕТЕКТОР, ЛОКАЛЬНІ ОЗНАКИ, ПОРІВНЯННЯ ЗОБРАЖЕНЬ, РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ, СТИСНЕННЯ, ШТУЧНА НЕЙРОННА МЕРЕЖА.

Об'єкт дослідження – системи моніторингу і трекінгу, які використовують відеокамери як засіб для отримання вхідних даних та алгоритми розпізнавання зображень для обробки вхідних даних.

Предмет дослідження – алгоритм розпізнавання зображень, що базується на нейронних мережах та використовує еквіваріантний детектор як елементарну одиницю мережі, а також алгоритм навчання такої мережі.

Мета кваліфікаційної роботи – проектування алгоритму розпізнавання зображень в відеопотоці та алгоритму для його навчання.

Методи досліджень – методи на базі нейронних мереж, інваріантні та еквіваріантні детектори, алгоритми розмиття, методи стиснення зниження розмірності зображень, методи навчання штучних нейронних мереж, методи розпізнавання зображень, які базуються на локальних ознаках та на каскадах примітивів.

Результат кваліфікаційної роботи – алгоритм розпізнавання зображень у відеопотоці, комплекс методів для попередньої обробки кадрів з потоку для ефективної роботи алгоритму та метод навчання нейронної мережі, який приймає відеопоток як вхідні дані для навчання.

Галузь застосування – системи моніторингу та трекінгу, які використовують відеокамери: відеоспостереження на дорожніх магістралях, пропускних пунктах, шлагбаумах, відеоспостереження в місцях трафіку осіб, аеропорти, торговельні центри, вокзали, метро.

ABSTRACT

Master's explanatory note: 85 pages, 7 tables, 8 figures, 4 applications, 47 sources.

ARTIFICIAL NEURAL NETWORK, COMPARISON OF IMAGES, COMPRESSION, DIMENSION REDUCTION, EQUIVARIANT DETECTOR, IMAGE RECOGNITION, INVARIANT DETECTOR, LOCAL SIGNS.

The object of research – monitoring and tracking systems that use video cameras as a base to obtain input data and image recognition algorithms to process input data.

The subject of research – image recognition algorithm based on neural networks and using an equivariant detector as an elementary unit of the network, as well as an algorithm for learning such type of network.

The purpose of the attestation work – designing an algorithm for image recognition in a video stream and an algorithm for its training.

Research methods – methods based on neural networks, invariant and equivariant detectors, blur algorithms, compression methods to reduce the dimensionality of images, methods of training artificial neural networks, image recognition methods based on local features and cascades of primitives.

The result of the attestation work – image recognition algorithm in a video stream, a set of methods for pre-processing frames from the stream for efficient operation of the algorithm and a method of learning a neural network that takes the video stream as input for learning.

Field of application – monitoring and tracking systems that use video cameras: video surveillance on highways, checkpoints, barriers, video surveillance in places of human trafficking, airports, shopping malls, railway stations, subways.

ЗМІСТ

Перелік умовних позначень	6
Вступ.....	7
1 Аналіз предметної області та постановка задачі	10
1.1 Інформативні параметри вимірювальних систем	10
1.2 Підходи формування нейронних мереж	19
1.3 Методи попередньої обробки зображення	26
1.4 Методи оцінки ефективності розпізнавання	33
1.5 Постановка задачі та обґрунтування вибору методу розв'язування	36
2 Порередня обробка зображень у відеопотоці.....	39
2.1 Інформативність зображення в відеопотоці.....	39
2.2 Оцінка контрастності.....	40
2.3 Оцінка різкості.....	42
2.4 Компенсація освітлення	47
3 Навчання нейронної мережі розпізнавання зображень у відеопотоці.....	50
3.1 Використання еквіваріантного детектора.....	50
3.2 Використання потоку візуальних даних для навчання	53
3.3 Алгоритм виділення локальних ознак	56
3.4 Розпізнавання зображення шляхом послідовної активації детекторів.....	62
4 Результати дослідження та експерименту	67
4.1 Планування експерименту	67
4.2 Оцінка ефективності окремого еквіваріантного детектору	69
4.3 Оцінка ефективності розпізнавання зображень	74
4.4 Локалізація об'єктів в композитних сценах.....	77
Висновки	79
Перелік джерел посилання	82
Додаток А	86
Додаток Б.....	97
Додаток В	100
Додаток Г	102

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

CBIR – Content-based image retrieval.

HPID – Head Pose Image Database.

JPEG – Joint Photographic Experts Group.

PCA – principal component analysis.

PSNR – peak signal-to-noise ratio.

RGB – red green blue.

SNR – signal-to-noise ratio.

SURF – surface.

ВСТУП

Розпізнавання зображень являє собою один з найважливіших компонентів систем управління та обробки інформації, автоматизованих систем та систем прийняття рішень. Завдання, пов'язані з класифікацією і ідентифікацією предметів, явищ і сигналів, що характеризуються кінцевим набором деяких властивостей і ознак, виникають в таких сферах як робототехніка, інформаційний пошук, моніторинг та аналіз візуальних даних, дослідження штучного інтелекту. Алгоритмічна обробка і класифікація зображень застосовуються в системах безпеки, контролю та управління доступом, в системах відеоспостереження, системах віртуальної реальності та інформаційних пошукових системах. На даний момент у виробництві широко використовуються системи розпізнавання рукописного тексту, автомобільних номерів, відбитків пальців або людських осіб, що знаходять застосування в інтерфейсах програмних продуктів, системах безпеки та ідентифікації особистості.

Сучасний розвиток транспортної та виробничої галузей народного господарства пов'язане зі збільшенням ролі і кількості вирішуваних завдань інформаційними і автоматизованими системами і комплексами, які впроваджуються безпосередньо в технологічні та суспільно-соціальні процеси, розширюється проблематика протидії тероризму, зростає значення своєчасного виявлення дефектів транспортної інфраструктури та дорожнього полотна, дистанційного і безперервного моніторингу та управління транспортними та пасажирськими потоками, віддаленого контролю і безпілотної експлуатації транспортно-технологічних комплексів. Сукупність транспортних засобів, інфраструктури і пасажиропотоку є основні об'єкти транспортної системи. Існуючі вимірювальні системи контролю та моніторингу транспортних засобів, об'єктів транспортної інфраструктури, пасажиропотоків, комплексної безпеки з самого початку орієнтовані на вирішення досить вузького кола завдань і для вирішення нового завдання, фактично потрібне використання іншої системи, а не налаштування і перепрофілювання існуючої. Також великі проблеми

виникають з можливою зміною інтерфейсних програм для створення розподіленої системи контролю або надання додаткових сервісів оператору. При цьому найбільш своєчасними і актуальними є завдання розробки інформаційного, математичного, алгоритмічного, програмного та апаратного забезпечення, що дозволяє визначати широкий спектр параметрів рухомих об'єктів, мінімізуючи використання активних пристроїв. Одним з перспективних напрямків розвитку таких методів є отримання характеристик реальних об'єктів через дослідження їх образів на зображеннях. Реалізація таких методів володіє такими незаперечними перевагами, як скритність для систем виявлення та протидії, а також малої помітності для об'єктів моніторингу і контролю в зв'язку з використанням сучасних фото- і відеодетектора малих розмірів, не випромінюючи електромагнітні імпульси.

За останній час істотний прогрес в розпізнаванні візуальних образів був досягнутий з появою методів зниження розмірності, згортальних нейронних мереж і констеляційних моделей. Однак, незважаючи на досягнуті успіхи, сучасні дослідження підтверджують той факт, що алгоритми розпізнавання зображень до сих пір не мають повноцінними здібностями біологічних зорових систем.

Так, актуальним проблемним питанням, визнаним науковим співтовариством, залишається розпізнавання зображень тривимірних об'єктів під різними кутами зору, підданих перетворенням обертання, масштабування і трансляції. Сучасні підходи до вирішення цього питання, такі як багат шарові згорткові нейронні мережі, а також використання інваріантних детекторів ознак SIFT і ORB, в даний момент пропонують часткові рішення, що не забезпечують достатньої точності розпізнавання і втрачають інформацію про структуру об'єкта. Також існують проблеми з обробкою потокового відео і виявлення об'єктів, що рухаються. Також проблемою є розпізнавання нечітких розмитих зображень або зображення перекриті іншими об'єктами в отриманих кадрах. У даній роботі детально розглядаються існуючі підходи до розпізнавання трансформованих об'єктів і пропонується система розпізнавання зображень, розроблена для вирішення цього проблемного питання.

Використання камер як уніфікованого пристрою для визначення множини параметрів рухомих об'єктів (відстані, швидкості, метричні параметри) дозволить знизити собівартість системи і спростити формалізацію одержуваної інформації за рахунок зменшення різновидів застосовуваних технічних пристроїв, а також без додаткових налаштувань робочого місця і додавання апаратних засобів підвищити багатофункціональність системи контролю. Крім того, уніфікація пристроїв і автоматизація системи отримання і обробки даних призведе до зменшення задіяння великої кількості людей в діагностичних роботах, що дозволить залучати їх до вирішення більш нетривіальних завдань і підвищить економічну ефективність всієї системи.

Мета роботи полягає в аналізі існуючих методів розпізнавання предметів які рухаються та впровадження покращень, для автоматизації систем моніторингу та аналізу предметів які рухаються за допомогою методів і алгоритмів аналізу зображень.

Об'єктом дослідження є геометричні, кінематичні та динамічні характеристики стану і поведінки рухомих і нерухомих об'єктів, які виявляються і детекційних інформаційно-вимірювальними системами моніторингу в системах з рухомими об'єктами.

Предмет дослідження – методи, програмне та алгоритмічне забезпечення вимірювальних систем віддаленого контролю, моделювання взаємодії різних елементів системи між собою і з оператором.

Теоретична і практична значущість роботи полягає в тому, що розроблене і реалізоване алгоритмічне, математичне і програмне забезпечення становить основу вимірювальної системи, яка може використовуватися для віддаленого контролю і моніторингу транспортних засобів, людей, технологічних процесів в різних сферах діяльності, а також входити до складу комплексу технічного зору, що забезпечує автономне функціонування транспортно-технологічних комплексів, безпілотних систем.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Інформативні параметри вимірювальних систем

Щоб описати параметри стану і поведінки об'єкта, які визначаються в інформаційно-вимірювальній системі моніторингу об'єктів транспорту, необхідно вибрати використовувані інформативні параметри, за допомогою яких потім можна описати:

а) Стан об'єкта (S), який представляє собою вектор значень його характеристик в даний момент часу t_i , причому деякі з них залежать від часу, а деякі – ні:

$$S(t_i) = \{r(t_i), v(t_i), dir(t_i), k(t_i), d(t_i), sh(t_i), h, l, w, class, type, add\}, \quad (1.1)$$

де $r(t_i)$ – віддаленість об'єкту, $v(t_i)$ – швидкість об'єкту, $dir(t_i)$ – напрямок руху об'єкту, $k(t_i)$ – контрастність об'єкту, $sh(t_i)$ – форма об'єкту, $d(t_i)$ – динамічні характеристики об'єкту, $class$ – клас об'єкту (транспортний засіб, частина оточення, тварина і т.д), add – додаткова множина признаков в залежності від класу, $type$ – тип в класі.

б) Поведінка об'єкта (B), що має на увазі комплекс дій, змін даного об'єкту і його реакцій на зовнішній вплив: зміна, розвиток, модернізацію:

$$B = f(t, Q), \quad (1.2)$$

де Q – комплекс параметрів об'єкта.

в) Зв'язок між окремими компонентами системи об'єктів, яка виражається в формі взаємних обмежень впливу елементів один на одного.

В результаті, з урахуванням прийнятих величин, система формально представляється як якась впорядкована послідовність виду:

$$\Sigma = \langle T, X, \Omega, Y, V, H, G, F, Z \rangle, \quad (1.3)$$

де $T = \{(t_i, t_{i+1})\}^N$ – ось часу; $X = \{x_j\}^N$ – множина вхідної інформації; $\Omega = \{\omega_i\}^N$, ω_i – оператор вводу, Ω – множина вхідних взаємодій; $Y = \{y_i\}^M$ – множина результатів; $V = \{\gamma_j\}^M$ – множина вихідних взаємодій.

На перетворення вхідної інформації X в вихідну інформацію Y на осі T вплив надають три функціональних фактори: G – алгоритм, функція виходу; H – функція поведінки об'єкта, функція переходу; F – функція управління, яка впливає на зміну як, G так і H ; Z – ряд внутрішніх станів об'єкта і $Z = \{Z_j\}$, також є безліч умов обмежень.

$$\left\{ \begin{array}{l} G: (X \times Z) \rightarrow Y, \\ H: (X \times Z) \rightarrow Z, \\ F: (X \times Z \times T) \rightarrow (G \times H), \\ \Omega: T \rightarrow^\infty X, \\ V: T^Y \rightarrow Y. \end{array} \right. \quad (1.4)$$

Можливість провести вимірювання тих чи інших параметрів об'єктів незалежно від характеру виконуваного розрахунку пов'язана з можливістю складання системи незалежних рівнянь:

$$\begin{aligned} X_1 &= f_1(C_1, C_2, \dots, C_i, \dots, C_k), \\ &\dots \\ X_i &= f_i(C_1, C_2, \dots, C_i, \dots, C_k), \\ X_{k-1} &= f_{k-1}(C_1, C_2, \dots, C_i, \dots, C_k), \\ C_1 + C_2 + \dots + C_i + \dots + C_k &= 1, \end{aligned} \quad (1.5)$$

де X_1, \dots, X_k – змінні параметри середовища яке аналізуємо, C_1, \dots, C_k – концентрації компонентів середовища яке аналізуємо, f_1, \dots, f_k – функції, які виражають характер залежності вимірюваних параметрів від складу середовища.

До основних технічних характеристик засобів вимірювання відносяться чутливість, діапазон вимірювань, час тривалості вимірювань.

Чутливість визначається відношенням зміни вихідного сигналу приладу до викликаючого його зміни вимірюваної величини, окремо розглядають абсолютну (1.6) і відносну (1.7) чутливості.

$$S_{absolute} = \Delta X_n / \Delta X, \quad (1.6)$$

$$S_{relative} = \Delta X_n / \left(\frac{\Delta X}{X} \right), \quad (1.7)$$

де ΔX – зміна величини на вході вимірювального приладу, ΔX_n – зміна величини на вході вимірювального пристрою, X – миттєве значення вхідної величини.

Також чутливість характеризується порогом – мінімальною зміною вхідної величини, яку можна зафіксувати без використання додаткових коштів, і стабільністю в часі при можливих впливах різних факторів.

Діапазон вимірювань являє собою область значень вимірюваної величини, для якої нормовані допустимі похибки приладу. При цьому визначаються верхній і нижній межі вимірювання обраних приладів. Діапазон $D = X_{max} / X_{min}$ визначається як відношення найбільшого верхньої межі вимірювань приладу X_{max} до мінімального значення X_{min} , який вимірюється приладом.

Проміжок часу від моменту стрибкоподібного вимірювання вхідної величини до моменту встановлення показань відлікового пристрою (із заданою точністю) характеризує час встановлення показань.

Для визначення параметрів стану і поведінки об'єкта в інформаційно-вимірювальної системи необхідно вибрати інформативні параметри, які будуть враховувати використовувані методи отримання первинної інформації про об'єкт.

Інформативним параметром називається параметр, функціонально пов'язаний з вимірюваним властивістю або який є самим вимірюваним властивістю об'єкта.

Існуючі інформаційно-вимірювальні системи можна класифікувати по інформативному параметру, який визначається в першу чергу і на основі якого потім відбувається оцінка необхідних характеристик об'єкта контролю. Первинний інформативний параметр залежить від обраного методу отримання даних про об'єкт і використовуваних для цього пристроїв і виду контролю об'єкта. При використанні методів неруйнівного контролю можливі такі інформативні параметри, як: реєстрація коерцитивної сили об'єкта, його магнітної проникності, намагніченості, напруженості магнітного поля, залишкової індукції, щільності потоку енергії іонізуючого випромінювання, параметрів магнітного шуму, теплового потоку, температури, ємності ділянки об'єкта або розподілу потенціалів на ньому, тощо – в залежності від пропонованого виду контролю: магнітного, електричного, радіохвильового, теплового, радіаційного, акустичного, оптичного. В останньому з них в якості первинного інформативного параметра може виступати: час проходження хвилі, геометрія інтенсивності хвильового пучка, поляризації хвиль, спектра випромінювання, амплітуда хвиль реєстрованого сигналу яку визначають як:

$$U_0(t) = [U_{m0} + kx(t)]\sin(\omega_0 t + \varphi_0). \quad (1.8)$$

При цьому коефіцієнт k вибирається таким чином щоб при всіх можливих від'ємних значеннях x зберігалась умова $U_{m0} \geq 0$.

Так, наприклад, при використанні інфрачервоних і ультразвукових методів в інформаційно-вимірювальних системах визначення місця розташування об'єктів можливо на відстані до 10 метрів за допомогою часопролітного (time-of-flight) вимірювання поширення сигналу від джерела до приймача. При цьому висока точність позиціонування (від 3 до 30 сантиметрів) заміщається чутливістю до перешкод від зовнішніх, оточуючих аналізований об'єкт елементам.

Значення інформативних параметрів служать вихідними даними для подальшої обробки сигналів повідомлень [1, 2]. Математична обробка інформативних параметрів може полягати в порівнянні з певних вирішальним правилам отриманих значень зі значеннями, прийнятими за норму чи за гранично допустимі, для прийняття відповідного рішення за результатами випробувань або для прийняття діагностичних рішень.

Інформативні параметри можуть вибиратися в залежності від здійснення того чи іншого процесу в системі. Наприклад, при передачі інформації до таких параметрів можна віднести: швидкість передачі інформації, достовірність передачі інформації, вид і параметри модуляції сигналу, смуга частот системи, потужність, що витрачається на передачу інформації. Вони пов'язані між собою таким чином, що з них можна вибирати той чи інший параметр як першорядних, виходячи з вимог, що пред'являються до подібних процесів в системі.

При роботі з зображеннями інформативність того чи іншого способу зазвичай визначається шляхом порівняння поточного зображення з еталонним, отримуючи при цьому коефіцієнт кореляції. Важливою підзадачею при цьому є вибір ділянки еталонного зображення для порівняння з точки зору інформативності. Недолік показника кореляції зазвичай полягає в тому, що він обчислюється в процесі порівняння поточного способу з еталоном, в той час як показник інформативності повинен обчислюватися апріорно, вказуючи на ті ділянки еталонного зображення, які будуть мати найбільш надійну метрику подібності. Невдалий вибір еталона може істотно погіршити результат міри схожості.

Завдання вибору інформативних ділянок для надійного і точного зіставлення зображень не отримала до теперішнього часу остаточного рішення, хоча і мало вивчена. Основна проблема тут криється у виборі оптимального поєднання суперечливих вимог гнучкого і адаптивного аналізу інформативності і швидких алгоритмів обчислення оцінок.

Найбільш поширені інформативні параметри фрагмента зображення: дисперсія яскравості, відношення сигнал/шум, радіус кореляції сигналу.

Точність прив'язки до еталонного зображення тим вище, чим більше неоднорідним є відповідні фрагменти зображення. Одним з обчислювально

простих і ефективних показників неоднорідності фрагмента є дисперсія яскравості всередині фрагмента, яка визначається виразом:

$$\sigma^2(x_0, y_0, N) = \frac{1}{(2N + 1)^2} \sum_{x=-N}^N \sum_{y=-N}^N (f(x + x_0, y + y_0))^2 - \left(\frac{1}{(2N + 1)^2} \sum_{x=-N}^N \sum_{y=-N}^N (f(x + x_0, y + y_0)) \right)^2, \quad (1.9)$$

де x_0, y_0 – точка зображення, для якої визначається інформативність, N – обраний розмір околиці точки.

Очевидно, чим вище дисперсія, тим більше неоднорідним є фрагмент. Таким чином, найбільш проста функція інформативності фрагмента має вигляд:

$$Inf(x_0, y_0, N) = \sigma(x_0, y_0, N). \quad (1.10)$$

Показник «ставлення сигнал-шум» (signal-to-noise ratio, SNR) визначається наступним чином:

$$SNR(x_0, y_0, N) = \frac{\sigma_s(x_0, y_0, N)}{\sigma_n(x_0, y_0, N)}, \quad (1.11)$$

де σ_s – середньоквадратичне відхилення сигналу, σ_n – середньоквадратичне відхилення шуму на фрагменті зображення.

Сенс розгляду SNR в якості показника інформативності фрагмента полягає в наступному. Якщо фрагмент однорідний, тобто варіації сигналу в ньому малі, то значення SNR теж мало. Якщо на фрагменті присутні сильні коливання яскравості, то дисперсія сигналу, як видно з попереднього пункту, велика, отже, і значення SNR велике. Даний показник є мірою неоднорідності: на більш

однорідних фрагментах шум превалує, тому ставлення сигнал/шум менше. Таким чином, функція інформативності фрагмента тут має вигляд:

$$Inf(x_0, y_0, N) = SNR(x_0, y_0, N). \quad (1.12)$$

Радіус кореляції сигналу показує, на якій відстані заміри сигналу можна вважати статистично незалежними. Чим більш однорідною є область зображення, тим більше для неї радіус кореляції. Радіус кореляції в точці зображення може бути обчислений за допомогою автокореляційної функції, наприклад, за формулою:

$$R_d(x_0, y_0, N) = \frac{\sum_{k=-d, k \neq 0}^d C_v(k, 0) + C(0, 0)}{C(0, 0)}, \quad (1.13)$$

де d – розмір апертури для обчислення радіуса кореляції, $C_v(k, l)$ – автокореляційна функція фрагмента зображення, яка обчислюється на області $\{(k, l): -d \leq k \leq d, l = 0\}$; (x_0, y_0) – координати точки зображення, N – розмір околиці заданої точки, C – інтегроване (середнє) значення C_v поблизу точки $(0, 0)$.

Таким чином, функція (1.13) також є мірою інформативності ділянок зображення.

Крім розглянутих параметрів в якості подібних заходів інформативності можна виділити також контрастність і чіткість фрагментів.

Співвідношення світлин на зображенні в цілому або на його окремих ділянках характеризують також контрастом – відношенням коефіцієнтів відбиття. Світлість окремої деталі може бути більшою або меншою в залежності від середньої яскравості зображення, його граничної яскравості або яскравості локального фону, тобто на ділянці, що безпосередньо прилягає до даної деталі. Тому існує як загальний контрасти збережений ($K = I_{max}/I_{min}$), так і локальний $K_{\text{локальний}} = I(x, y)/I(x + \Delta x, y + \Delta y)$ і детальний ($K = I_{\text{детальний}}/I_{min}$) контрасти, де I_{max} , I_{min} ,

$I_{\text{детальний}}$ – максимальна, мінімальна і детальна інтенсивності зображення, $I(x,y)$ – інтенсивність точки з координатами (x,y) .

Очевидно, що інтервал щільності всього зображення може бути визначений і через логарифм загального його контрасту:

$$D = \lg K = D_{\max} - D_{\min}. \quad (1.14)$$

Важливим параметром зображення є чіткість, що характеризує якість відтворення на ньому дрібних деталей і контурів. Цей параметр можна однозначно визначити через число темних і світлих деталей мінімального розміру, які можуть бути розміщені в кадрі, будучи встановленими в шаховому порядку, тобто визначається як:

$$N = kz^2, \quad (1.15)$$

де k – коефіцієнт формату кадру, а z – число рядків в даному стандарті розкладання. При цьому передбачається, що довжина пікселя (одичного елемента) зображення дорівнює його ширині.

У фотографії різкість s визначається головним чином роздільною здатністю об'єктивів і емульсій, розсіюванням світла в фотослої та іншими подібними факторами і оцінюється зворотною величиною зони розмиття σ контуру об'єкта: $s=1/\sigma$.

У системах з використанням зображень, наприклад в телебаченні, різкість можна оцінити як безрозмірне відношення апертури D до граничного шару σ (розмиття образу) $s=D/\sigma$. Подібну оцінку можливо поширити і на растрові зображення, прийнявши в якості елемента розкладання растровий осередок.

Зона розмиття є прикордонний шар між фоном і об'єктом. Параметри прикордонного шару можуть не тільки свідчити про якість зображення, але і бути повноцінним інформативним критерієм, що визначає сукупність параметрів всієї системи моніторингу і контролю, а також параметрів стану і поведінки

досліджуваного об'єкта, який спробуємо порівняти з дисперсією оптичного сигналу, що передається по лініях волоконно-оптичних систем передачі даних.

Основні фактори, що визначають необхідність кардинальних перетворень в електроенергетиці під впливом складних умов, на основі попереднього дослідження можна згрупувати наступним чином як зображено на рис 1.1.



Рисунок 1.1 – Причини розмиття образу об'єкта на зображенні

Так як розмиття образу залежить як від різних зовнішніх умов, так і від параметрів об'єкта спостереження, то воно може свідчити не тільки про якість одержуваного зображення, але і бути повноцінним інформативним критерієм, що визначає сукупність параметрів всієї системи моніторингу і контролю, а також параметрів стану і поведінки досліджуваного об'єкта, зіставляється з дисперсією оптичного сигналу, що передається по лініях волоконно-оптичних систем передачі даних. Деталізація факторів, що впливають на розмиття, а

значить і описують його (рис. 1.1). Подібна інтерпретація функції розмиття прикордонного шару об'єкта дозволяє використовувати її як важливий інформативний параметр зображення, щоб потім за образом об'єктів визначати їх реальні параметри в інформаційно-вимірювальній системі моніторингу.

Вибір величини розмиття як інформативного параметра поєднує в собі розглянуті раніше параметри дисперсії, відносини сигнал/шум, радіуса кореляції сигналу для опису зображення і дозволяє вирішувати задачу вибору інформативних ділянок через порівняння значень розмитість окремих елементів зображення.

1.2 Підходи формування нейронних мереж

Один з основних підходів, який найбільш широко використовувався в області розпізнавання зображень, являє собою застосування класичних моделей-класифікаторів, що навчаються з учителем. Для навчання таких моделей використовуються маркована вибірка даних, що складається з масиву зображень і відповідного їм масиву міток, що визначають категорію, до якої належить зображення. В процесі навчання масив даних розділяється на дві нерівні частини – навчальну вибірку і тестову вибірку, потім за допомогою специфічного для конкретного алгоритму правила навчання параметри моделі налаштовуються з використанням навчальної вибірки таким чином, щоб отримавши в якості вхідних даних зображення, модель на виході виробляла б мітку відповідного класу. Цей підхід представлений безліччю моделей, серед яких найбільш широко використовуваними є регресійна модель, штучна нейронна мережа (багатошаровий перцептрон), метод опорних векторів, а також дерева прийняття рішень і моделі ансамблі, що представляють собою поєднання деяких перерахованих моделей.

Багатошарові перцептрони, які навчаються методом зворотного поширення помилки, широко використовуються для розпізнавання різних категорій зображень, таких як рукописні цифри [3], почерк, людські обличчя [4]

і дані зорових сенсорів робототехнічних систем [5]. Модель багатошарового перцептрона являє собою сукупність штучних нейронів – обчислювальної одиниці моделі – об'єднаних в рівні (шари), задані в ієрархічному порядку.

Штучний нейрон являє собою модель біологічного нейрона (нервової клітини), представлену одним або декількома входами, одним виходом і функцією активації [6]. Крім цього, кожен вхід штучного нейрона має асоційований коефіцієнт або вага. Поведінка нейрона будується наступним чином: нехай є $m + 1$ входів, значення яких дорівнюють, $x_0, x_1 \dots x_m$, а значення їх ваг рівні, $\omega_0, \omega_1 \dots \omega_m$, при цьому перший вхідний елемент, як правило, являє собою фіксоване значення зміщення $x_0 = 1$. Тоді вихідне значення нейрона є значення функції активації від зваженої суми його вхідних значень [7]:

$$y = \varphi \left(\sum_{i=0}^m \omega_i x_i \right). \quad (1.16)$$

Як функції активації може виступати функція, що володіє властивостями нелінійності, нормалізації вхідних даних, і деякими іншими. Поширеними кандидатами є функція сигмоїд (1.17), гіперболічний тангенс (1.18) або випрямляюча лінійна функція (1.19).

$$\sigma(z) = \frac{1}{1 + e^{-z}}. \quad (1.17)$$

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (1.18)$$

$$y(z) = \begin{cases} 0, & z < 0; \\ 1, & z \geq 0. \end{cases} \quad (1.19)$$

При об'єднанні штучних нейронів в мережу вхідні значення нейрона шару l є вихідні значення нейронів попереднього шару $l - 1$. При цьому нейрони

першого (вхідного) шару отримують в якості вхідних значення безпосередньо дані, що підлягають розпізнаванню, які в разі розпізнавання зображення представляють собою значення інтенсивності складових його пікселів (точкових елементів). Вихідний шар мережі може варіюватися в залежності від завдання, але класична архітектура має на увазі формування його числом нейронів, рівній кількості класів розпізнавання, при цьому вихідне значення кожного нейрона нормується по інтервалу $\{0,1\}$, і являє собою ймовірність приналежності вхідного зображення до відповідного класу [7]. Згідно з теоремою про універсальну апроксимації, такі багатошарові нейронні мережі здатні інкапсулювати будь-яку безперервну функцію за допомогою довільного набору нейронів [7].

Оскільки сформулювати аналітично правило класифікації зображень за категоріями розпізнавання часто представляється нетривіальним, здатність навчатися на базі вибірки робить нейронні мережі та споріднені з ними моделі придатними для розпізнавання природних зображень навколишнього світу, що відрізняються нечіткою структурою і безліччю варіацій в межах класу. Приклад схеми штучної нейронної мережі з трьома шарами зображено на рис 1.2.

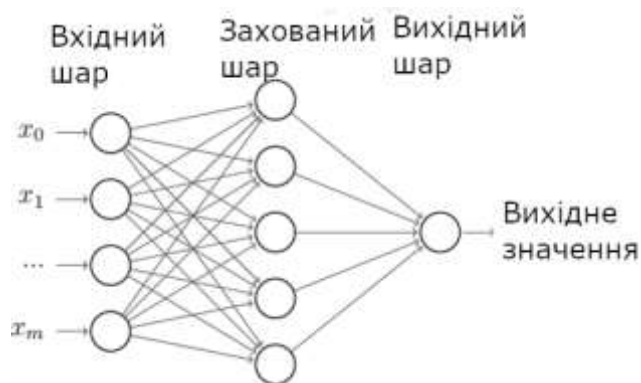


Рисунок 1.2 – Схема штучної нейронної мережі з трьома шарами

Навчання мережі методом зворотного поширення полягає в наступному: нехай є деяка невідома функція розпізнавання $g: X \rightarrow Y$, аргументом якої є

зображення $n \in X$, представлені у вигляді вектору довжини n , а значеннями функцій – безліч класів (категорій) $y \in Y$. Навчальна вибірка представляє собою підмножину значень цієї функції $D = \{(, y), (, y), \dots, (, Y)\}$.

Завдання навчання моделі розпізнавання полягає в знаходженні такої функції $h: X \rightarrow Y$, яка б апроксимувати функцію g на всій її області визначення, в тому числі значеннях, які не включені в D , і являє собою додаток теорії оптимізації.

Нехай $h(x)$ – вихідне значення мережі, отримане шляхом послідовної активації нейронів кожного шару, а $g(x)$ – значення функції, що апроксимується для одного і того ж зображення. Далі виконується крок зворотного поширення, що полягає в обчисленні приватної похідною від виходу мережі для кожного нейрона мережі по відношенню до його ваг:

$$\frac{\partial E}{\partial \omega_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial \omega_{ij}}. \quad (1.20)$$

$$E = \frac{1}{2} (h(x) - g(x))^2. \quad (1.21)$$

У формулі (1.21) E – середньоквадратичне відхилення мережі, ω_{ij} – вага між i -м та j -м нейроном, net_j – вхідне, а o_j – вихідне значення j -го нейрону. Далі на кожному кроці навчання ваги нейронів інкрементуються значеннями приватних похідних відповідно до методу градієнтного спуску [7]. Варіації навчального алгоритму включають в себе виключення додаткових параметрів регуляризації з метою захисту від перенавчання, і використання різних оптимізаторів – методу Ньютона, методу імітації відпалу, L-BFGS та інших [8].

Багатошарові перцептрони демонструють успішні результати при використанні їх для розпізнавання зображень деяких окремих обмежених категорій, таких як символи природної мови, рукописні цифри і почерк [9]. В даний час в більшості додатків, що використовують пряме навчання з учителем для розпізнавання зображень, нейронні мережі витіснені методом опорних

векторів, що пропонують більш ефективно з точки зору обсягу обчислювальних ресурсів рішення [10].

Метод опорних векторів розглядає кожен екземпляр даних (зображення) як точку в n -вимірному просторі, де n відповідає розмірності даних або загальної кількості пікселів зображення [11]. Кожна з точок належить до певного класу (категорії). При цьому завдання розпізнавання представляється у вигляді завдання по знаходженню такої гіперплощини в n -вимірному просторі, який би відділяв всі крапки, відповідні зображенням даного класу, від інших, які не належать йому. Припускаючи, що таких гіперплощин може існувати багато, метод опорних векторів ставить за мету відшукання площині, відстань до якої від найближчої точки максимально в межах множини можливих варіантів – в тому числі оптимальну розділяючу гіперплощину і відповідний їй оптимальний класифікатор. Вхідні дані, таким чином, мають вигляд:

$$\{(x_0, y_0), (x_1, y_1), \dots, (x_m, y_m)\}, \quad (1.22)$$

де x_i – i -те зображення, а y_i – i -тий клас, який представлений цілим числом. Значення x_i , що представляють собою m -мірний вектор, нормалізуються в межах інтервалу $\{0,1\}$. Розділяє площину задається параметром w – перпендикуляром (нормальним вектором) від точки до площини, і описується рівнянням $wx - b = 0$. Таким чином, завдання зводиться до мінімізації $\|w\|$. По теоремі Куна-Таккера [12] гіперплощина може бути представлена як лінійна комбінація векторів навчальної вибірки:

$$w = \sum_{i=1}^n a_i y_i x_i, \quad (1.23)$$

де a_i – деякі множники Лагранжа. Знаходження значення w таким чином дозволяє отримати лінійні гіперплощини, тому такий метод відносять до розділу так званих лінійних опорних векторів. Функція класифікації при цьому має вигляд:

$$F(x) = \text{sign}((w, x) + b), \quad (1.24)$$

де b – допоміжний параметр зміщення [8]. На практиці випадки, де дані в задачі розпізнавання можуть бути розділені лінійно, досить рідкісні. У таких випадках застосовується метод використання ядер, запропонований Б. Босером, І. Гийон і В. Вапніком, і полягає в тому, що елементи, для навчання вибірки вкладаються в простір x' більш високої розмірності з допомогою спеціального відображення $\varphi: \mathbb{R}^n \rightarrow x'$ [12]. При цьому відображення φ вибирається так, щоб в просторі x' вибірка була здатна розділятися лінійно. Ядром класифікатора називається вираз:

$$\hat{x}(x, x') = (\varphi(x), \varphi(x')). \quad (1.25)$$

Вираз (1.25) задає відображення вибірки в новий простір, і його роль може виконувати будь-яка позитивна визначена симетрична функція двох змінних. На практиці зустрічаються наступні ядра: поліноміальне, радіально базисна функція, гаусова базисна функція, сигмоїда [12].

Навчання моделі, що використовує метод опорних векторів проводиться методами квадратичного програмування, такими як послідовна мінімальна оптимізація.

Метод опорних векторів має деякі переваги і недоліки по відношенню до використання багатосарових перцептронів:

а) багатосаровий перцептрон є модель з безліччю прихованих параметрів, що залежать від числа нейронів мережі. Параметризована модель потенційно здатна до інкапсуляції більш складних, високорівневих функцій, але при цьому

вимагає більше часу і обчислювальних ресурсів для навчання та налаштування параметрів. Метод опорних векторів використовує вектори, відібрані з навчальної вибірки, при цьому кількість параметрів обмежена зверху розміром вибірки, а на практиці може бути проріджені за рахунок використання інженерії ознак [8].

б) на відміну від навчання нейронної мережі, яке здійснюється за допомогою методу градієнтного спуску (і його варіацій) і оцінки помилки мережі, навчання моделі опорних векторів включає в себе не тільки оцінку помилки, але і метрику складності отриманої гіперплощини. Пошук оптимального значення нейронної мережі вразливий до наявності локального мінімуму, здатного зупинити процес градієнтного спуску, при цьому метод опорних векторів при коректному виборі метапараметрів гарантує знаходження глобального рішення.

в) навчена нейронна мережа вимагає мінімальних обчислювальних ресурсів для роботи в режимі розпізнавання (передбачення категорій). Метод опорних векторів в деяких випадках, коли число векторів велике в порівнянні з розміром вибірки, будує передбачення істотно повільніше [8].

д) у порівнянні з нелінійним (використовувачим ядра) методом опорних векторів, нейронна мережа демонструє розширені можливості до онлайн-навчання, коли розмір вибірки не фіксований і поповнюється за рахунок надходження нових даних.

У більшості сучасних додатків алгоритмів розпізнавання і машинного навчання зараз віддано перевагу методу опорних векторів [9] за рахунок скорочення часу навчання та стійкості до локального мінімуму. Метод опорних векторів також широко використовується для розпізнавання таких зображень, таких як людські особи, демонструючи високу точність розпізнавання (80-85% успішно розпізнаних зображень) для вирівнюється вибірки.

Особливість завдання розпізнавання зображень полягає в тому, що дані, що представляють собою візуальні сигнали, демонструють вкрай низьку інформаційну ємність – тобто, більша частина точок растрового зображення

(наприклад, відповідні ділянкам однотонного або рівномірно розподіленого фону) не містить інформації, що впливає на розпізнавання. При цьому розмірність зображень, що використовуються в системах обробки інформації, як правило, досить велика – сучасні засоби мультимедіа, графічні дисплеї і сенсори забезпечують масове поширення зображень (фотографій, кадрів відео, комп'ютерної графіки) з високою роздільною здатністю, розмірність яких вимірюється мільйонами точок. Для класичних методів розпізнавання образів характерна пряма залежність між розмірністю (числом параметрів) даних навчальної вибірки і часом навчання, а також показниками збіжності при оптимізації моделі. Наявність великої кількості параметрів, основна частина яких не містить істотної для розпізнавання інформації, негативно впливає на продуктивність моделі, і крім вимоги значно більш високих обчислювальних ресурсів веде до появи так званої проблеми перенавчання [7], коли функція розпізнавання, апроксимована моделлю, задовільно класифікує навчальну вибірку, але є при цьому не генералізованою і демонструє низьку точність в тестовій вибірці. Для вирішення цієї проблеми використовується підхід пошуку компактного представлення зображення – виділення обмеженого числа генералізованих ознак, що містять основну інформацію, необхідну для розпізнавання.

1.3 Методи попередньої обробки зображення

Одна з особливостей розпізнавання зображень в порівнянні з іншими додатками теорії розпізнавання образів полягає в тому, що зображення в растровому вигляді у вигляді двовимірної матриці пікселів, кожен з яких має деякий значення яскравості або кольору), мають високу розмірність – середньостатистична фотографія може бути представлена вектором довжини $\sim 10^6$. Обробка даних, представлених розмірністю таких порядків, вимагає виняткових обчислювальних ресурсів, і для неспеціалізованих обчислювальних машин представляється нездійсненним або непрактичною завданням. При

цьому, однак, лише невелика частина цих параметрів критична для задачі розпізнавання, що дозволяє зображень демонструвати низьку чутливість до випадкового шуму і глобальним спотворень. Ця особливість успішно використовується в алгоритмах стиснення з втратами – так, за допомогою алгоритму JPEG зображення може бути стисло аж до 10%, при цьому зміни залишаються непомітними для людського ока. З огляду на цю особливість, стає можливим застосування до природних зображень статистичних методів зниження розмірності, таких як метод головних компонент. Суть методу полягає в тому, щоб представити вхідні дані у вигляді лінійної суми компонент з деякими коефіцієнтами.

Нехай x_i – i -те зображення. Знайдемо матрицю коваріацій:

$$\Sigma = \frac{1}{m} \sum_{i=1}^m (x_i x_i^T). \quad (1.26)$$

Власні вектори і власні значення матриці Σ будуть відповідати головним компонентам перетвореного зображення. Виберемо p головних компонент в якості базису для представлення даних. Побудуємо матрицю

$$U = (u_1 | u_2 | \dots | u_p), \quad (1.27)$$

де u_1, u_2, \dots, u_p – p перших власних векторів матриці коваріації (вибраних по порядку спадання відповідних їм значенням). Тоді ми зможемо отримати компактне представлення зображення як $x^{PCA} = U^T x$, де розмірність x^{PCA} буде знижена до значення p .

Класичний метод головних компонент, однак, непридатний для більшості зображень через обчислювальну складність побудови матриці коваріації. М. Турк і А. Пентланд в 1991 р запропонували алгоритм розпізнавання Eigenfaces, де використовували альтернативний, прийнятний для сучасних обчислювальних

машин метод розрахунку власних векторів. В їхньому прикладі метод використовувався на фронтальних фотографіях людських обличч. Підтверджуючи припущення про те, що розмірність зображення може бути значно знижена, зберігаючи при цьому досить інформації для успішного розпізнавання людиною, вони показали, що кожна з осіб вибірки можна уявити за допомогою обмеженого (<10) набору головних компонент.



Рисунок 1.3 – Приклади головних компонент алгоритму Eigenfaces

Для розпізнавання тестові зображення проектувалися на базис обраних головних компонент, тобто представлялися у вигляді лінійної суми p доданків. Потім на представлених таким чином даних тренували модель, яка використовує навчання з учителем (багатошаровий перцептрон або SVM), і таким чином, завдання зводилася до класичної. Використання Eigenfaces дозволяло ефективно розпізнавати обличчя при різному освітленні і давало деяку стійкості до орієнтації; проте, алгоритм демонстрував незадовільні результати для зображень осіб різного розміру (варіації масштабу). Крім того, алгоритм був розрахований на те, що вхідні дані будуть являти собою особи, зорієнтовані відповідним чином, не пропонуючи методу відшукування цікавить фрагмента особи серед зображення композитної сцени.

Крім перерахованих, метод головних компонент мав і інші обмеження, які сприяли появі нових методів представлення зображень. Б. Ольшозен в своїй роботі [9] показав, що алгоритм, названим їм розрідженим кодуванням здатний ефективніше представляти внутрішню структуру зображення і об'єктів в ньому, при цьому демонструючи деякі властивості, вражаюче схожі з властивостями

клітин зорової кори головного мозку (так званих «простих клітин» зони V1). Цей алгоритм, проте, на противагу PCA, представляв дані у вигляді переповненого базису векторів, кожен з яких, таким чином, не був лінійно незалежним від інших. Інакше, якщо зображення $x \in \mathbb{R}^n$, то розріджений уявлення складається з зваженої суми таких векторів φ , що

$$x = \sum_{i=1}^k a_i \varphi_i, \quad (1.28)$$

де $k > n$ (умова переповнення базису).

Розрідженість отриманого уявлення забезпечується тим, що для окремо взятого зображення x , більшість компонентів φ матимуть коефіцієнт, що дорівнює нулю. Ця умова мотивовано тим фактом, що природні зображення, як правило, можуть бути представлені за допомогою комбінації невеликого числа ненульових компонентів примітивів, таких як краю або кордону (в області алгоритмів розрідженого кодування відповідні компоненти зветься «атомів» або «кодових слів»). Таким чином, розріджений кодування забезпечує великий набір компонентів, які можуть значно відрізнитися один від одного, при цьому гарантуючи, що окремо взяте зображення буде представлено за допомогою суми всього лише деяких з них.

Оскільки розріджений уявлення все ще має апроксимувати вихідні дані досить ефективним для розпізнавання чином, на відповідний базис накладається додаткова умова максимальної відповідності. Таким чином, для вибірки з m зображень компоненти можуть бути отримані за допомогою мінімізації наступної функції

$$J(x, a, \varphi) = \sum_{j=1}^m \left(x^j - \sum_{i=1}^k a_i^j \varphi_i \right)^2 + \lambda \sum_{i=1}^k S(a_i^j), \quad (1.29)$$

де $S(a_i^j)$ – функція умови розрідженості, що обмежує a_i^j таким чином, щоб більшість значення аргументів дорівнювали нулю. Існують кілька кандидатів на роль $S(a_i^j)$; на практиці застосовуються норми L_0 , L_1 і функція $S(a_i^j) = \log(1 + (a_i^j)^2)$.

Існують різні алгоритми пошуку розрідженого коду для вибірки зображень, таких як ортогональне узгоджене переслідування, регресія найменшого кута і використання специфічних нейронних мереж – розріджених автоенкодерів. Перевага цього методу в порівнянні з методом головних компонент виражається в тому, що компоненти, отримані за допомогою другого способу, завжди представляють собою лінійні перетворення вхідних даних, тоді як у разі розрідженого коду (і деяких інших уявлення) компоненти можуть бути нелінійними, інкапсулюю, таким чином, більш складні функції представлення даних.

Інший широко використовується клас алгоритмів, здатний формувати цілісні уявлення об'єктів – обмежені машини Больцмана.

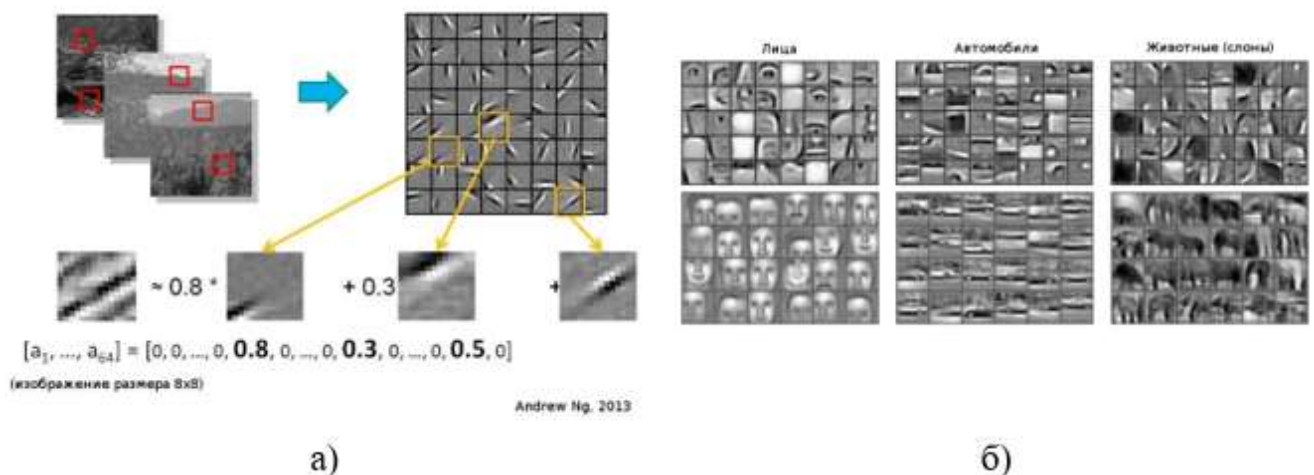


Рисунок 1.4 – Приклад роботи алгоритму обмеженої машини Больцмана

На рис. 1.4 – а) демонстрація представлення ділянок зображення за допомогою розрідженого коду. б) приклади компонент для трьох категорій об'єктів (на різних рівнях).

Обмежена машина Больцмана являє собою генеративну стохастичну нейронну мережу, яка навчається завдяки формуванню деякого імовірнісного розподілу своїх входів. Вони являють собою модифікацію класичних машин Больцмана, які, в свою чергу, є варіаціями мереж Хопфілда [13].

Зв'язки між нейронами такої мережі являють собою двочастковий граф (біграф), де одна частина відповідає вхідному шару мережі, а друга – прихованому шару. Кожен вхідний нейрон з'єднаний з усіма прихованими нейронами за допомогою симетричних зв'язків, при цьому нейрони в межах кожної частини біграфа не мають зв'язків один з одним (на відміну від класичних, «необмежених» машин Больцмана, де такі зв'язки можливі). Це обмеження дозволяє ефективно навчати мережу, використовуючи алгоритм зіставлення розбіжність (contrastive divergence).

Істотною перевагою уявлень, які формуються за допомогою розрідженого кодування і обмеженою машини Больцмана в порівнянні з методом головних компонент є їх нелінійність, що дозволяє розглядати методи нарощування таких репрезентативних моделей. Цей підхід відомий під назвою глибокого навчання, і відзначений різким підвищенням точності розпізнавання в безлічі сфер машинного навчання, в тому числі в розпізнаванні зображень. В його основі лежить припущення про те, що уявлення, яким навчаються репрезентативні моделі, мають ієрархічну природу, і таким чином, існує можливість навчання каскаду моделей, кожен з яких приймає в якості вхідних даних уявлення, що виробляються моделлю яка стоїть вище. Метод головних компонент, таким чином, не здатний формувати глибокі ієрархії уявлень, оскільки будь-яка, необмежено велика комбінація лінійних перетворень тотожна одному лінійному перетворенню [13].

Для розпізнавання зображень успішно застосовувалися глибокі моделі, що складаються з обмежених машин Больцмана – так звані глибокі мережі довіри. Використання ієрархічних уявлень дозволяє такої моделі навчатися складним, масштабним об'єктам, забезпечуючи додаткові рівні стійкості до інваріантним перетворенням на кожному шарі уявлення. Так, глибока модель, навчена на базі

людських обличч, здатна розпізнавати значно більш суттєві викривлення, ніж модель Eigenfaces, що включають в себе обертання об'єкта в межах обмежених кутів. Глибокі моделі також можуть будуватися і на базі методів розрідженого кодування – одним з найбільш відомих є глибокий автоенкодер, яких навчають пошарово, жадібним чином. В цілому глибокі моделі забезпечують більш гнучкі і багаті уявлення, які підходять для об'єктів зі складною структурою. Зворотною стороною цієї переваги є ускладнений процес навчання, в окремих випадках (для глибоких мереж довіри) вимагає розробки окремих алгоритмів, і в загальному випадку – споживає більше обчислювальних ресурсів.

Компактні цілісні уявлення дозволяють позбутися від «прокляття розмірності», перетворюючи складні в обробці, об'ємні зображення в компактний вид, забезпечуючи при цьому деяку стійкість до варіативності. Методи, які здійснюють нелінійні перетворення, такі як розріджений кодування, можуть використовуватися для отримання багаторівневих уявлень (див. рисунок 2.3б), використовуючи глибоке навчання і властивість стаціонарності природних зображень (той факт, що статистичні характеристики локальних ділянок зображень, як правило, розподілені рівномірно). При цьому підхід відшукування компактних цілісних уявлень демонструє високі результати для об'єктів, що мають в цілому схожу форму (як людські особи, вирівняні однаковим чином), але не здатний справлятися з об'єктами, що характеризуються високою варіативністю в рамках категорії.

Більш того, оскільки розпізнаються об'єкти зазвичай мають тривимірну природу, вони здатні істотно змінювати форму під впливом геометричних трансформацій (так, зображення особи в профіль не може бути представлено сумою компонентів, отриманих декомпозицією зображення особи анфас). В силу умови цілісності отримані уявлення уразливі до проблеми неповних даних – ситуацій, коли частина об'єкта загорожена або невиразні через шум. Для отримання компактних цілісних уявлень, таким чином, необхідна суворо підібрана вибірка об'єктів, вирівняних по загальній орієнтації (див. рисунок 2.2).

Складання подібної вибірки має на увазі участь експериментатора і обробки вихідних зображень людиною.

Ці особливості і обмеження методу зниження розмірності привели до розвитку альтернативного підходу до розпізнавання, специфічного для сфери розпізнавання зображень і використовує виявлення локальних ознак, що є стійкими компоненти (частини) зображеного об'єкта.

1.4 Методи оцінки ефективності розпізнавання

Оцінка ефективності роботи системи розпізнавання є ключовим етапом дослідження, що дозволяє чисельно виміряти різні аспекти якості розробленої системи. Для отримання чисельного значення оцінки широко використовуються як загальні методи математичної статистики, так і специфічні показники, що застосовуються для оцінки алгоритмів машинного навчання.

Однією з найбільш простих метрик оцінки ефективності є процентна частка коректно розпізнаних зображень [8, 11]. Коректним розпізнаванням вважається отримання на виході алгоритму класу, відповідного попередньо заданому класу. Для оцінки використовується вибірка, спроектована аналогічно навчальній вибірці, але містить зображення, до яких алгоритм не міг мати доступ в процесі навчання. Для цього, як правило, вихідна загальна вибірка поділяється на дві нерівні частини (розмір тестової вибірки при цьому може відрізнятись, і складати 20-30% розміру загальної вибірки [8]). Тоді якщо функція $c(x_i)$ представляє собою функцію-індикатор коректності розпізнавання i -го зображення:

$$c(x_i) = \begin{cases} 1, & h(x_i) = y_i, \\ 0, & \text{в інших випадках.} \end{cases} \quad (1.30)$$

Тоді частка коректно розпізнаних зображень оцінюється наступним чином [22]:

$$Q = \frac{\sum_{i=1}^n c(x_i)}{n}, \quad (1.31)$$

де n – кількість зображень в тестовій вибірці. Таким чином, оцінка приймає значення на відрізку $[0,1]$, де 1 відповідає коректному розпізнаванню всіх зображень вибірки.

Розглянутий показник є найбільш узагальненим і підходить для безлічі завдань розпізнавання з обмеженою кількістю класів. У завданнях, де кількість класів не фіксоване, і завдання розпізнавання являє собою задачу ідентифікації об'єкта певної категорії серед безлічі інших, потенційно необмежених категорій, замість неї застосовуються такі показники як точність і повнота оцінки. Їх використання дозволяє розрізнити помилково-позитивні (класифікатор прийняв позитивне рішення по зображенню, що не містить шуканого об'єкта) і помилково-негативні (класифікатор не пізнав об'єкт на зображенні, де він був присутній) помилки розпізнавання, або помилки першого і другого роду. Таким чином, точність оцінки в межах класу представляє собою частку зображень, дійсно належать даному класу щодо всіх зображень які система віднесла до цього класу. Повнота системи – це частка знайдених класифікатором зображень, що належать класу щодо всіх зображень цього класу в тестовій вибірці [14]. Тобто:

$$P = \frac{TP}{TP + FP}. \quad (1.32)$$

$$R = \frac{TP}{TP + FN}. \quad (1.33)$$

Де точність виражається рівністю (1.32), повнота виражається рівністю (1.33), TP – кількість істино-позитивних рішень, TN – кількість істино-негативних рішень, FP – кількість помилково-позитивних рішень, FN – кількість помилково-негативних рішень.

Показники точності і повноти широко використовуються в області обробки інформації, і як правило, розраховуються спільно. При цьому існує кілька методик зіставлення двох показників:

- кожен показник враховується окремо;
- для випадків, коли між показниками спостерігається залежність, проводиться оцінка одного з показників при фіксації іншого;
- двоє показників можуть бути скомбіновані в один.

Серед прикладів агрегованих метрик для третьої методики зустрічаються такі показники як коефіцієнт кореляції Метьюза, регресивні коефіцієнти ΔP та $\Delta P'$ [14]. Також для деяких цілей розглядають зважене арифметичне середнє показника точності і значення, зворотного показником повноти, і навпаки [8]. З точки зору теорії ймовірності ці показники можуть інтерпретуватися наступним чином: точність відповідає ймовірності того, що випадково обране з безлічі позитивно упізнаних зображень дійсно розпізнано коректно, при цьому повнота є ймовірність того, що випадково обране із загальної вибірки зображення буде коректно класифіковано алгоритмом (позитивно або негативно). Так, в залежності від програми завдання, до продуктивності методу розпізнавання можуть бути пред'явлені вимоги, що стосуються максимізації повноти (для випадків, коли певна кількість хибно-позитивних рішень допускається), так і збалансованого значення двох показників або міри, вираженої в такий спосіб:

$$F = 2 \frac{RP}{R + P}. \quad (1.34)$$

Таким чином ця міра є гармонічним середнім значенням показників точності та повноти [7].

Крім перерахованих показників, важливу роль в оцінці ефективності методів розпізнавання грають такі показники, як розмір довірчого інтервалу для вибірки зображень, оцінка числа ітерацій і часу роботи алгоритму при навчанні, необхідних для досягнення збіжності, темп збіжності та інше.

Для оцінки розміру вибірки, що забезпечує статистично достовірні результати, використовується метод розрахунків довірчих інтервалів. Для параметру θ – кількість зображень вибірки, яка представляє випадкову величину X з рівнем довіри p , таку, що $X=(x_1, x_2, \dots, x_n)$ інтервал довіри представляють за допомогою таких границь $l(x_1, x_2, \dots, x_n)$ та $u(x_1, x_2, \dots, x_n)$, які є реалізаціями випадкових величин $L(X_1, X_2, \dots, X_N)$, $U(X_1, X_2, \dots, X_N)$, таких, що:

$$P(L \leq \theta \leq U) = p. \quad (1.34)$$

При цьому граничні точки інтервалу довіри l та u представляють собою границі довіри. Рівень довіри p вибирається виходячи зі специфіки поставленої задачі і вимогам до системи обробки інформації, а також зв'язаних з нею ризиків.

Досягнення збіжності при навчанні це етап, коли $\Delta J = J_i - J_{i-1} < \alpha$, тобто зміні функції ціни навчається моделі на черговому кроці оптимізації не перевищує деякого фіксованого мінімального значення, вибір якого може проводитися з урахуванням поставленого завдання. Існує набір методик оптимізації, що дозволяють варіювати розміри цього порогового значення виходячи з темпів збіжності на конкретній ділянці області визначення функції, заданої навчається моделлю [11].

1.5 Постановка задачі та обґрунтування вибору методу розв'язування

Алгоритми та методології розпізнавання об'єктів що швидко рухаються базуються на предметній області та апаратному забезпеченні. На даний момент складно створити універсальну систему для розпізнавання різних класів об'єктів, для подальшого їх аналізу. Вузькими містами зараз є обчислювальні можливості комп'ютерів, недосконалість алгоритмів. Для вирішення поставленої задачі у рамках кваліфікаційної роботи буде розглядатися переважно розпізнавання рухомих об'єктів транспортного типу. Зараз це є одна з найзатребуваніших прикладних областей для даного типу програмного забезпечення. Таке

забезпечення може використовуватися на постах ДАІ, шлагбаумах, пропускних пунктах, залізничних шляхах, магістралях, метрополітену та багато інших, для моніторингу трафіку, знаходження вузьких місць в трафіку на дорогах, для перепустки в якісь місця та багато іншого.

Порівняння різних алгоритмів, методологій, їх синтез, адаптування та оптимізація їх частин в єдине рішення і є метою даної кваліфікаційної роботи. Для систематизації знань а також для структуризації рішення проблеми виділені наступні етапи роботи алгоритму:

- оптимізація обробки відеопотоку: вибір зображень з серії кадрів які відрізняються, знайти кращий кадр за якістю серед схожих для більш детального аналізу зображення. Даний крок потрібен для оптимізації швидкості роботи алгоритму, не має сенсу повністю аналізувати кожен кадр з відеопотоку, адже велика ймовірність того що зображення ідентичні або майже незмінні. Також важливо вибрати більш чітке і контрастне зображення для подальшого аналізу серед вибірки з декількох схожих задля покращення процесу розпізнавання;

- стиснення, зменшення розмірності зображення з мінімальною втратою інформативності. Даний крок необхідний для оптимізації швидкодії алгоритму, адже як правило необроблені зображення мають дуже високі розмірності, що ускладнює обробку і потребує більше ресурсів і часу для аналізу;

- знаходження границь об'єктів на зображенні. В даному кроці знаходяться контури об'єктів, встановлення їх кількості і розташування на зображенні. Таким чином надалі ми будемо аналізувати лише частини зображення які нас цікавлять;

- аналіз інформативних параметрів об'єкта. В даному кроці знаходить колір, розміри, віддаленість від камери, швидкість руху, позиція відносно камери, напрям руху. Всі ці характеристики далі використовуються для вирішення бізнес-задач конкретної системи;

- виділення класу об'єкта та його вид. На даному кроці ми аналізуємо що за об'єкт був знайдений, його клас та тип, вирішується чи потрібен він для роботи системи чи є аномалією, далі може класифікуватися а також використовуватися для навчання адаптивних нейронних мереж для самовдосконалення системи.

Для оцінки результатів дослідження слід провести експерименти в яких порівняти результати роботи алгоритму розпізнавання з відомими аналогами алгоритмів та порівняти долю успішно розпізнаних зображень. Тестовими вибірками можуть бути згенеровані зображення спеціальними програмами або ж відомі тестові набори які знаходяться в відкритому доступі. Також слід провести порівняння з відомими алгоритмами на тестових вибірках, які зазнали впливу шумів та оклюзій. Окремим експериментом слід перевірити роботу алгоритму при розпізнаванні об'єктів на відео матеріалах.

Кожен з даних етапів може бути декомпозованим на більш дрібні задачі, для кожного з етапів використовуються різні види алгоритмів та методологій. В даній кваліфікаційній роботі будуть розглядатися алгоритми для кожного з етапів, вибрані та оптимізовані для складання цілісної системи для розпізнавання об'єктів що швидко рухаються з використанням відеокамер, без інших вимірювальних приладів та датчиків.

2 ПОПЕРЕДНЯ ОБРОБКА ЗОБРАЖЕНЬ У ВІДЕОПОТОЦІ

2.1 Інформативність зображення в відеопотоці

Для систем розпізнавання об'єктів часто існують неформалізовані вимоги, що умови освітлення повинні дозволяти реєструвати зображення об'єктів, на яких можуть бути помітний основний рельєф та контури (достатній контраст). Важливою умовою є наявність прямого чи близького до нього кута падіння сонячних променів на фронтальну площину об'єктів, а також наявність розсіяного джерела світла. Негативний вплив на якість розпізнавання надає бокове засвічення об'єктів.

В більшості систем відеоспостереження використовується частота кадрів не менш ніж 25 кадрів в секунду. Зображення об'єкту який аналізується на декількох послідовних кадрах майже не відрізняється, і кадри можуть повністю співпадати, або відрізнятися так що це майже непомітно. Аналіз таких майже ідентичних кадрів не має сенсу, адже це нераціональне використання обчислювальних ресурсів. Повний аналіз зображень, а тим більше розпізнавання на ньому об'єктів це складні процеси, які витрачають багато ресурсів обчислювальної техніки. Таким чином, слід проводити фільтрацію кадрів які майже не відрізняються. Було знайдено 3 основних видів підходів, які дозволяють порівнювати зображення.

а) використання хеш-функцій на зменшеній копії зображення і порівняння отриманих хешів. Даний метод має високу швидкість обчислень, але як результат дає грубу оцінку, яка напряму залежить від зменшення зображення. Через це він може використовуватися лише для пошуку віддаленої візуальної схожості об'єктів, наприклад для пошуку схожих зображень в колекції. Метод відноситься до категорії пошуку зображень за їх змістом (англ. Content-based image retrieval (CBIR)) [15] – розділ комп'ютерного зору, який вирішує задачу пошуку зображень, які мають необхідний контент, в великих колекціях цифрових зображень.

б) обчислення коефіцієнту кореляції по одній з існуючих формул. Такий підхід чутливий до незначних змін зображення, але потребує значних обчислювальних ресурсів. Його використання досить обмежене специфічними областями роботи з зображеннями [16].

в) порівняння зображень через побудову і аналіз SURF-дескрипторів, які описують деякі невеликі зони зображення. Дослідження показують, що даний метод показує найкращі результати при порівнянні зображень об'єктів, які знімають під різними кутами і в різних умовах освітленості, але він потребує в декілька разів більше обчислювальних ресурсів навіть ніж метод обчислення коефіцієнту кореляції [17]. Існує модифікація даного методу (BRISK), яка має високу швидкість обробки [18], але даний метод складно використовувати для пошуку незначних змін в кадрах, і попри це дана модифікація потребує набагато більше обчислювальних ресурсів ніж використання хеш-функцій.

Таким чином, задача порівняння кадрів не має простого і ефективного рішення, яке б можна було використовувати для відео потоку високої роздільної здатності. Однак можна створити гібрид з існуючих методів, а також створити можливість вибору і переходу від одного методу до іншого, в залежності від ситуації, об'єктів які розпізнаються, віддаленості камери від об'єктів спостереження, типу об'єктів.

2.2 Оцінка контрастності

Перед тим, як покращувати якість зображення, необхідно дати оцінку його якості. Людина одним поглядом може сказати про зображення, яскраве воно чи темне, чітке чи розмите, контрастне чи ні. Алгоритми працюють детально, аналізуючи кожен піксель, або оброблюють невелику групу пікселів. Тому, на основі роботи алгоритму, часто дуже складно дати загальну оцінку якості зображення.

Контраст – в оптиці, сенситометрії та фотометрії різниця в характеристиках різних ділянок зображення, здатність фотографічного матеріалу чи оптичної системи відтворювати цю різницю, а також характеристика чутливості ока (зорової системи) відносно яскравості і кольору.

Контрастність (також, іноді в різних контекстах використовується саме слово контраст або коефіцієнт контрасту) – ступінь контрасту, частіше за все виражається безрозмірною величиною, відношенням або логарифмом відношень. Мірою контрастності є величина k .

$$k = \frac{B_b - B_s}{B_s}, \quad (2.1)$$

де B_b – яскравість фону, B_s – яскравість предмету чи його зображення.

Контрастність зображення є показником його інформативності, так як цей показник бере до уваги освітленість і можливість розрізнити об'єкти на зображенні. Цього параметру недостатньо, щоб можна було стверджувати що зображення якісне, так як бокове засвічення об'єктів надає високу оцінку контрасту, але ускладнює виділення деталей зображення.

У випадку, коли зображення вдається нормалізувати, контрастність стає найбільш корисною оцінкою.

Розрізняють два типи контрастності: відносно яскравості та тонова.

Контрастність відносно яскравості є різницею між фізичною або видимою яскравістю окремих ділянок зображення. Розрахунок фізичної або видимої яскравості можна розглядати як конвертацію кольорового зображення в ахроматичні кольори. Тому контрастність відносно яскравості – це порівняння двох ділянок зображення, які привели до ахроматичних кольорів.

Якщо проаналізувати RGB-гістограми, то можна зробити висновок, що у контрастного зображення кількість темних та світлих пікселів повинна бути приблизно однаковою, а різниця в їх яскравості великою, а основним місцем концентрації пікселів – біля границь діапазону.

Тонова контрастність бере до уваги не тільки переходи яскравості, але і переходи кольору. Її слід використовувати при аналізі кольорових зображень, так як, незважаючи на зростання складності обчислень, результуюча оцінка буде біль точною.

Доцільним критерієм оцінки контрастності відносно яскравості може бути дисперсія яскравості пікселів зображення[19]:

$$\sigma^2 = \frac{1}{N} \sum_{p=1}^N (Y_p - Y)^2, \quad (2.2)$$

де Y – середнє значення контрастної відносно яскравості всього зображення, Y_p – значення контрастності відносно яскравості в точці p , N – загальна кількість точок зображення.

Оцінка контрастності (2.2) може бути нормована шляхом розрахунку відношення середньоквадратичного відхилення до максимально можливого значенню яскравості:

$$C = \frac{2\sigma}{Y_{max}}. \quad (2.3)$$

Діапазон значення C – $[0;1]$. Значення 0 відповідає однотонному зображенню, значення 1 – максимально контрастному. Оптимальне значення контрастності (2.3) залежить від типу об'єкта, який представлений на зображенні.

Контрастність зображення не може бути точним показником якості зображення: розмиті зображення менш інформативні, проте мають більшу високую оцінку контрастності, ніж чіткі зображення з меншою кількістю точок максимальної і мінімальної яскравості.

2.3 Оцінка різкості

Різкість частіше за все трактують як якість відтворення на фото достатньо великих фрагментів. В цьому, умовно, її відмінність від чіткості зображення, яка характеризує відмінність маленьких деталей.

Підвищення різкості – один з найбільш поширених способів покращення зображень. Зазвичай завжди потрібно в тій чи іншій мірі підсилювати різкість початкового зображення.

Об'єкти на зображенні відрізняються один від одного за яскравістю чи оптичній щільності, що є причиною їх відрізняємості. Границя між об'єктами завжди має ту чи іншу ступінь розмитості.

Ширина розмитості теоретично різкої границі між світлими (яскравими) та темними ділянками (об'єктами) зображення характеризує різкість зображення. Це характерно для будь-яких зображень, як кольорових, так і чорно-білих.

Чіткість зображення – це параметр, який характеризує якість отриманого зображення і показує, наскільки чітко помітні дрібні об'єкти на зображенні і визначається мінімальною відстанню між ними.

Інакше кажучи, чіткість показує повноту відтворення дрібних об'єктів на зображенні і наскільки у них різкі контури.

Зображення високої чіткості має різкі, чітко видимі границі.

Різкість може бути ще одним критерієм інформативності зображення, в загальному випадку вона може бути визначена наявністю контурів контрастних переходів. Однозначного методу для визначення контрастності не існує, але є декілька алгоритмів, які дозволяють обчислити оцінку різкості із різноманітних характеристик зображення.

Головна відмінність всіх методів які пропонують в формулах, які визначають кількісну оцінку контрастності зображення. В цілому всі методи можна розділити на три групи:

- методи, які розраховують середнє значення диференціальних операторів зображення;
- методи, які розраховують ексцес амплітудного спектру зображення;
- методи, які розраховують відношення значимих і незначимих точок амплітудного спектру зображення.

Перша група методів розраховує характеристику на основі різності, напряму з вхідного зображення, наприклад за допомогою оператора Лапласа чи оператора Робертса. Друга і третя групи розраховують частотні характеристики зображення (перетворення Фур'є, чи Wavelet-перетворення), і розраховує кількість контрастних переходів з спектру зображення.

Основна ідея методу, який використовує диференціальні оператори в тому, що він розраховує середнє значення яскравості від перетвореного оператором зображення. Такий метод чутливий до якості зображення, а також до характеру зображених об'єктів, тобто він може дати високу оцінку різкості для поганого зображення зі складним об'єктом, і меншу – для якісного зображення з малою кількістю контрастних переходів.

Оператор Лапласа – диференційний оператор, сума других часткових похідних по координатам зображення, еквівалентний операціям градієнту та дивергенції:

$$\text{Laplace}(f) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \quad (2.4)$$

Оператор Робертса – це згортка зображення з наступними ядрами:

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \text{ та } \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (2.5)$$

Оператор Робертса в порівнянні з оператором Лапласа потребує менше обчислень, більш точно виділяє тонкі контури, але він дуже чутливий до шумів.

В такому випадку слід розглянути варіант використання оператора Робертса, попередньо обробив зображення, щоб видалити шуми. Алгоритми шумозаглушення зазвичай спеціалізуються на заглушуванні якогось конкретного шуму.

Шум в відео потоці можна умовно розділити на просторовий і часовий. Часовим шумом називають неприємний візуальний ефект, який з'являється через випадкову чи корельовану зміну просторового шуму від кадру до кадру. Тому фільтри шумозаглушення часто поділяють також на часові та просторові: перші використовують між декількома кадрами, а другі для роботи з одним зображенням. На практиці часто використовують комбінацію просторових і часових фільтрів, першим використовують просторовий фільтр, оскільки після його застосування процедура знаходження руху в часовому фільтрі стає

простіше і точніше. Такі комбіновані фільтри називають 3D-фільтри. Основною проблемою при використанні просторових фільтрів є ймовірність погубити чіткість країв об'єктів на зображенні, а також дрібні деталі, які мають розміри по амплітуді як і шум. Також складною проблемою є оцінка якості заглушування шуму. Часто використовують підхід при якому на зображення штучно наносять шум, а потім порівнюють оригінальне зображення з зашумленим за допомогою метрик. Найпопулярнішою метрикою є PSNR (peak signal-to-noise ration), яке вимірюють за формулою:

$$PSNR(x, y) = 20 * \log_{10} \frac{255}{\sqrt{\frac{1}{N} \sum_{i=1..N} d(x_i, y_i)^2}}, \quad (2.6)$$

де x_i, y_i – i -ті пікселі двох зображень, N – сумарна кількість пікселів на кожному зображенні, а $d(x_i, y_i)$ – різниця між кольорами відповідних пікселів (для ахроматичного зображення це різниця значень пікселів, для кольорових зображень – евклідова відстань між пікселями в тримірному просторі). Чим ближче відфільтроване зображення до оригінального, тим більше значення PSNR, і тим вище вважається якість роботи алгоритму.

Основними базовими підходами просторового шумозаглушення:

- лінійне усереднювання пікселів по сусідам;
- медіанна фільтрація;
- математична морфологія;
- розмиття по Гаусу;
- методи на основі вейвлет-перетворення;
- метод головних компонент;
- анізотропна дифузія.

Деякі з методів використовують і для часової модифікації. Алгоритми на основі вейвлет-перетворення і методу головних компонент не можуть використовуватися в відео обробці в реальному часі, адже працюють дуже довго, хоча і показують найкращі показники.

Для поставленої задачі слід підбирати алгоритми на основі тестових вибірок даних в індивідуальних умовах. В деяких випадках шумом можна вважати навіть частини зображення, так наприклад це може бути дощ під час зйомки, тому слід враховувати дані ситуації при побудові системи.

Шумозаглушення можна розглядати також як окремий етап попередньої обробки кадрів, перед безпосереднім розпізнаванням об'єктів.

В роботі [20] розглянули метод оцінки різкості на основі перетворень Фур'є, головна мета якого в обчисленні ексцесу зображення в частотній області:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}, \quad (2.7)$$

де μ_2 – другий центральний момент зображення, μ_4 – четвертий центральний момент, σ – стандарте відхилення.

В роботі [21] також використовується перетворення зображення в частотну область, але в якості критерію оцінки різкості використовують наступне відношення:

$$FM = \frac{T_H}{M * N}, \quad (2.8)$$

де T_H – число значимих точок, $M*N$ – розмір зображення амплітудного спектру. Значимі точки визначають як точки, значення в яких перевищує деяку долю максимуму амплітуди спектру.

Дослідження методів показало, що оцінка різкості з використанням оператора Лапласа дає суттєвий розкид результатів при схожих умовах освітлення і ракурсу, що пов'язане з відсутністю нормування результату і способом виділення об'єкта. Оцінка (2.7) показує більш стабільні результати, але майже не реагує на фактичне зниження різкості, після застосування розмиття по Гаусу. Оцінка (8) найбільш точно відповідає початковій задачі пошуку різких зображень, вона коректно реагує на зміни різкості при русі і відсутності фокусу,

до цього ж має нормування, яке дозволяє по заданому рівню відсіювати зображення, які незадовільні за встановленою якістю.

Попри переваги оцінки (2.4), вона схильна до тих же проблем, що й оцінка з використанням Лапласіана. Головна причина цього в тому, що різкість зображення оцінюється кількісно, і тому пов'язана з загальною кількістю контрастних переходів, а ця кількість залежить від кількості об'єктів на зображенні. Роблячи висновок, можна вимірювати лише відносну різкість зображення, яка оцінюється для одного об'єкту на серії кадрів.

2.4 Компенсація освітлення

Різкість частіше за все трактують Підвищити шанси для ефективного використання методів (2.2), (2.3), (2.7), (2.9) може компенсація зображення об'єкту по умовам освітлення.

Для нормування вхідного зображення по умовам освітлення існує декілька методів, наприклад DoG-фільтрація [22] і частотна фільтрація смуговим фільтром [23], які видаляють високі і низькі частоти. Ці методи дозволяють ефективно вирівнювати зображення при дисбалансі яскравості при освітленні.

DoG-фільтрація знаходить краї шляхом використання розмиття зображення по Гаусу з вказаним значенням стандартного відхилення. Результуюче зображення є розмитою версією вхідного. Далі використовується інше розмиття з меншим значенням, що надає більш різке зображення, ніж в попередньому випадку. Фінальне зображення розраховується заміною кожного пікселя різницею від пікселів двох розмитих зображень.

Для нормалізації яскравості зображення з цим методом часто використовується гама-корекція.

Частотний смуговий фільтр при нормалізації освітлення заглушує високочастотні шуми і згладжує нерівномірне освітлення зображення, закодоване в зображенні з низькими частотами.

Два цих методи підвищують чіткість зображення, але не усувають повністю вплив нерівномірного освітлення.

Також для підвищення контрасту зображення використовується адаптивне вирівнювання гістограми з обмеженням (CLAHE) [24]. Спочатку зображення розбивається на прямокутні ділянки (блоки). В кожному блоці обчислюється гістограма розподілу яскравості. Щоб уникнути перенасичення монотонних ділянок зображення, обмежується нахил функції розподілу яскравості шляхом введення обмеження на максимальне значення частоти яскравості. Нехай M – кількість рядків в блоці, N – кількість стовбців в блоці, α – параметр, який задає обмеження. Граничне значення гістограми розраховується відповідно від формули:

$$hist_{lim} = hist_{min} + \alpha(MN - hist_{min}), \quad (2.9)$$

Де $hist_{min} = MN / (L + 1)$, L – максимальне значення яскравості.

Після обмеження для кожної області формується своя функція перетворення яскравості на основі вирівнювання гістограми розподілу яскравості. Для вхідного зображення розраховується значення яскравості результуючого зображення відповідно до формули:

$$LUT(k) = \sum_{j=0}^k p_r(r_j) = \sum_{j=0}^k \frac{n_j}{n}, \quad (2.10)$$

де $LUT(k)$ – значення яскравості результуючого зображення, відповідно до яскравості r_k вхідного зображення, $k \in [0, L]$; r_j – яскравість елемента вхідного зображення, n_j – кількість елементів вхідного зображення зі значеннями r_j , n – загальна кількість елементів зображення, $p_r(r_j)$ – ймовірність появи елемента вхідного зображення з яскравістю r_j .

Таке перетворення дозволяє більш рівномірно розподілити рівні яскравості по частоті, в основі перетворення базується припущення, що в контрастних чорно-білих зображеннях значення яскравості розподілені рівномірно.

Щоб зруйнувати блокову структуру отриманого зображення, результуюче зображення формується шляхом білінійної інтерполяції по чотирьом найближчим *LUT*.

3 НАВЧАННЯ НЕЙРОННОЇ МЕРЕЖІ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ У ВІДЕОПОТОЦІ

3.1 Використання еквіваріантного детектора

В рамках дослідження була розроблена ієрархічна модель репрезентації зображення, що дозволяє представляти зображені об'єкти з урахуванням їх просторових характеристик. Сформульовано принципи побудови моделі репрезентації зображення, а також аналітична формулювання моделі та її компонентів.

Цифрове растрове зображення являє собою матрицю чисел, де кожне значення матриці відповідає деякому характеристичному параметру найменшого логічного елемента зображення (пікселя). Як параметр можуть виступати значення інтенсивності (яскравості) пікселя для чорно-білих зображень, числове значення кольору для деякої колірної моделі та ін. Це подання цифрового растрового зображення зручно для використання в інформаційних системах і системах мультимедіа, а також для зберігання даних, проте не містить психофізіологічних характеристик, які могли б використовуватися для розпізнавання. Таким чином, для побудови системи розпізнавання зображень, що відповідає поставленій меті, необхідна альтернативна репрезентація об'єкта у вигляді певної кількості параметрів або ознак, які є характеристичними по відношенню до класу розпізнавання. Перш ніж формулювати вимоги до моделі репрезентації, розглянемо деякі ситуації, які можуть виникнути при отриманні зображення об'єкта:

- об'єкт може бути присутнім на зображенні не повністю, відчувати на собі вплив різного роду візуальних шумів, перекриття, або розмиття;
- разом з шуканим об'єктом на зображенні можуть бути присутніми деталі навколишнього оточення, фону, або інші об'єкти;

– в силу застосування до об'єкту просторових трансформацій, таких як обертання, масштабування або трансляції, форма об'єкта на зображенні, що представляє собою проекцію тривимірного об'єкту на двовимірну поверхню сенсора камери, може змінюватися відповідно правилам проєктивної геометрії.

З урахуванням цих спостережень, а також беручи до уваги проведеній аналітичний аналіз літератури з розпізнавання зображень, сформулюємо основні принципи побудови моделі репрезентації зображення, дотримання яких сприятиме досягненню поставленої мети:

Крок 1. Модель повинна являти собою структуру даних, елементи якої відповідають локальним ознаками зображення і їх комбінацій. Підставою для цього служить можливість часткової присутності об'єкта або декількох об'єктів на зображенні – ситуація, в якій неможливо використовувати підхід компактних цілісних уявлень.

Крок 2. Елементи такої структури даних повинні бути організовані ієрархічно, тобто таким чином, що елементи першого рівня моделі представляють собою ознаки, витягнуті безпосередньо з локальних підмножин растрової матриці, а елементи наступних рівнів – результати комбінацій елементів попереднього рівня. Підставою служить багаторазово зазначене дослідниками [25] властивість представляти і описувати спостережувані об'єкти у вигляді композиції їх фрагментів або частин. Ця особливість проєктованої моделі відрізняє її від констеляційної моделі, що не використовує ієрархічну угруповання.

Крок 3. З урахуванням необхідності розпізнавати об'єкт з різних кутів зору, модель репрезентації об'єкта повинна включати в свою структуру елементи (ознаки), отримані шляхом спостереження за об'єктом в процесі трансформацій в реальному часі в достатньому обсязі для того, щоб оцінити тривимірну форму об'єкта. Таким чином, система розпізнавання, яка використовує розробляється модель, повинна навчатися в потоці візуальних даних, що представляє собою фрагменти відеозйомки об'єкта в русі. Дана вимога є одним з найбільш помітних

характерних відмінностей розроблювальної моделі від альтернатив, розглянутих раніше, і накладає додаткові умови на процес навчання системи розпізнавання;

Крок 4. Локальні ознаки зображення і їх комбінації повинні виражатися з використанням принципу еківаріантності. Даний принцип, який застосовується в статистичному аналізі даних [26], протиставляється більш поширеній поняттю інваріантності. Різниця полягає в тому, що при інваріантному перетворенні даних відповідно деякої функції значення функції залишається незмінним, в той час як при еківаріантному перетворенні воно змінюється певним чином, відповідним характером перетворення. Дотримання принципу еківаріантності особливо важливо при побудові ієрархії локальних ознак, так як елементи більш високих рівнів ієрархії отримують необхідну інформацію про параметризації просторового перетворення на попередньому рівні, в разі ж інваріантних ознак відповідна інформація втрачається. Відмінності між результатами еківаріантного і інваріантного перетворення приведені на рис 3.1.



Рисунок 3.1 – Співставлення інваріантного та еківаріантного детектору локальних ознак.

Приклад на рис 3.1 демонструє, що в результаті інваріантного розпізнавання окремих ознак ми втрачаємо інформацію, необхідну для

розпізнавання їх композиції – так, зображення обличчя в профіль признаються інваріантним детектором однаковими з точки зору набору знайдених ознак. Еквіваріантний детектор дозволяє проводити відмінності між ознаками зображення, отриманого з різних кутів огляду.

Еквіваріантний детектор являє собою відокремлену математичну конструкцію в межах одного рівня ієрархії моделі. Окремий детектор є частиною ієрархічної ланцюжка, отримуючи в якості вхідних даних результати детекторів нижнього рівня, і відправляючи вихідні дані на наступний рівень моделі. Детектор виконує наступні функції:

- отримавши на вході зображення трансформованого об'єкта, повертає зображення об'єкта в первісному вигляді (за замовчуванням), тобто не поставляє під трансформації (функція ідентифікації чи класифікаційний компонент);
- отримавши на вході зображення трансформованого об'єкта, повертає матрицю афінної трансформації (функція відновлення трансформації або регресійний компонент).

Остання функція детектора є реалізацією принципу еквіваріантності і цим відрізняється від детекторів локальних ознак, які використовують в методах детекції локальних ознак.

3.2 Використання потоку візуальних даних для навчання

Система розпізнавання зображень що рухаються яка проектується включає в себе безліч детекторів локальних ознак, що спеціалізуються в межах рівня ієрархії на певні унікальні ознаки об'єкта і здатних відрізнити області зображення, які містять такі ознаки, від інших. Для коректного функціонування детекторів, кожен з них повинен реалізовувати відповідну класифікує правило, для чого запропоновано метод навчання детекторів на вибірці зображень. Таким чином, що розробляється система розпізнавання зображень є навчається систему [8].

Оскільки окремих детектор виконує дві функції, здійснюючи як класифікацію ознак, так і оцінку параметрів ініціалізації, то навчальна вибірка, відповідно, повинна складатися із зображень, анотування вектором параметрів ініціалізації, а також зображень, які не належать шуканого класу. Остання частина вибірки може бути виключена для розглянутого випадку використання детектора на базі алгоритму трансформує автоенкодера, однак при цьому вибірка повинна містити згруповані пари зображень трансформованого об'єкта. Навчальну вибірку з m елементів можна виразити як:

$$S_{train} = \{(I_i, I_i^t, T_i); i \in 1..m\}, \quad (3.1)$$

де I_i – вхідне зображення, яке представляє собою матрицю пікселів, I_i^t – зображення трансформованого об'єкта, також є матрицею пікселів, T_i – числове значення трансформації (матриця чи інша репрезентація), в результаті якого об'єкт перейшов з вхідного стану в трансформоване.

Необхідність отримання такої вибірки може істотно ускладнити процес навчання, оскільки відомі існуючі відкриті бази даних, як правило [27], не містять необхідної інформації. У деяких джерелах, що описують навчання трансформує автоенкодера [28], пропонується використовувати штучно згенеровані бази даних, наприклад, за допомогою комп'ютерної графіки. Однак суттєвим недоліком такого підходу для розроблюваної системи розпізнавання зображень є відсутність можливості виділення локальних областей зображень, що містять відокремлені ознаки.

Розроблено метод навчання детекторів в потоці даних, що представляє собою відеофрагмент або послідовність кадрів відеозйомки. Представлений метод навчання реалізований за допомогою двох розроблених взаємопов'язаних алгоритмів.

Перший алгоритм – виділення локальних ознак, що здійснює пошук на окремому кадрі I_j локальних областей де S_j^i – i -та область j -го кадру, таких що

використання знайдених областей S_j^i в якості вхідних даних детекторів першого рівня ієрархії максимізує точність розпізнавання зображення I_j .

Другий алгоритм – оптичного трекінгу і оцінки трансформації, що здійснює відстеження виявлених локальних областей S_j^i зображення в потоці даних і який проводить оцінку трансформації об'єкта для двох сусідніх кадрів I_{j-1}, I_j .

Використовуючи методи проективної геометрії і ефект паралакса руху, пропонується аналізувати послідовність кадрів I_0, I_1, \dots, I_m які змінюються в часі, обчислюючи значення трансформації T_j , що впливає на об'єкт в інтервалі, позначеному кадрами I_{j-1}, I_j . Локальні ділянки послідовних кадрів I_{j-1}, I_j використовуються в якості примірників вибірки вихідних зображень (x це координата) і результатів трансформації t .

Перед тим як розглянути детально обидва алгоритму, наведемо етапи методу навчання в потоці даних:

Етап 1. Для кожного кадру $I_j, j > 0$ в потоці даних проводиться обчислення трансформації $T_j = f(I_{j-1}, I_j)$, форма яка в загальному випадку залежить від постановки задачі і обмежень в русі об'єктів і камери. Обчислення значення трансформації проводиться за допомогою ефекту монокулярного паралакса руху і методів проективної геометрії.

Етап 2. Для кадру I_{j-1} проводиться пошук локальних ознак об'єкта $S_{j-1}^0, S_{j-1}^1, \dots, S_{j-1}^k$.

Етап 3. Для кадру I_j проводиться відстеження кожної з таких ознак з метою пошуку відповідного відображення в кадрі I_j . Для цього використовуємо алгоритм оптичного трекінгу.

Етап 4. Отримані пари ознак $\{S_{j-1}^0, S_j^0\}, \{S_{j-1}^1, S_j^1\}, \dots, \{S_{j-1}^k, S_j^k\}$ та трансформація T утворюють вибірку для навчання еквіваріантного детектора.

Етап 5. Вибірка поповнюється за рахунок наступної пари кадрів $\{I_j, I_{j+1}\}$. Додатково зберігається можливість використовувати в вибірці пари зображень, які не є сусідніми – наприклад, $\{S_{j-1}^k, S_{j+1}^k\}$, і $2T_j$ в якості значення трансформації.

Етап 6. Еквіваріантний детектор навчається на базі отриманої вибірки методом зворотного поширення помилки [7].

Схема навчання моделі в потоці даних приведена на рис 3.2.

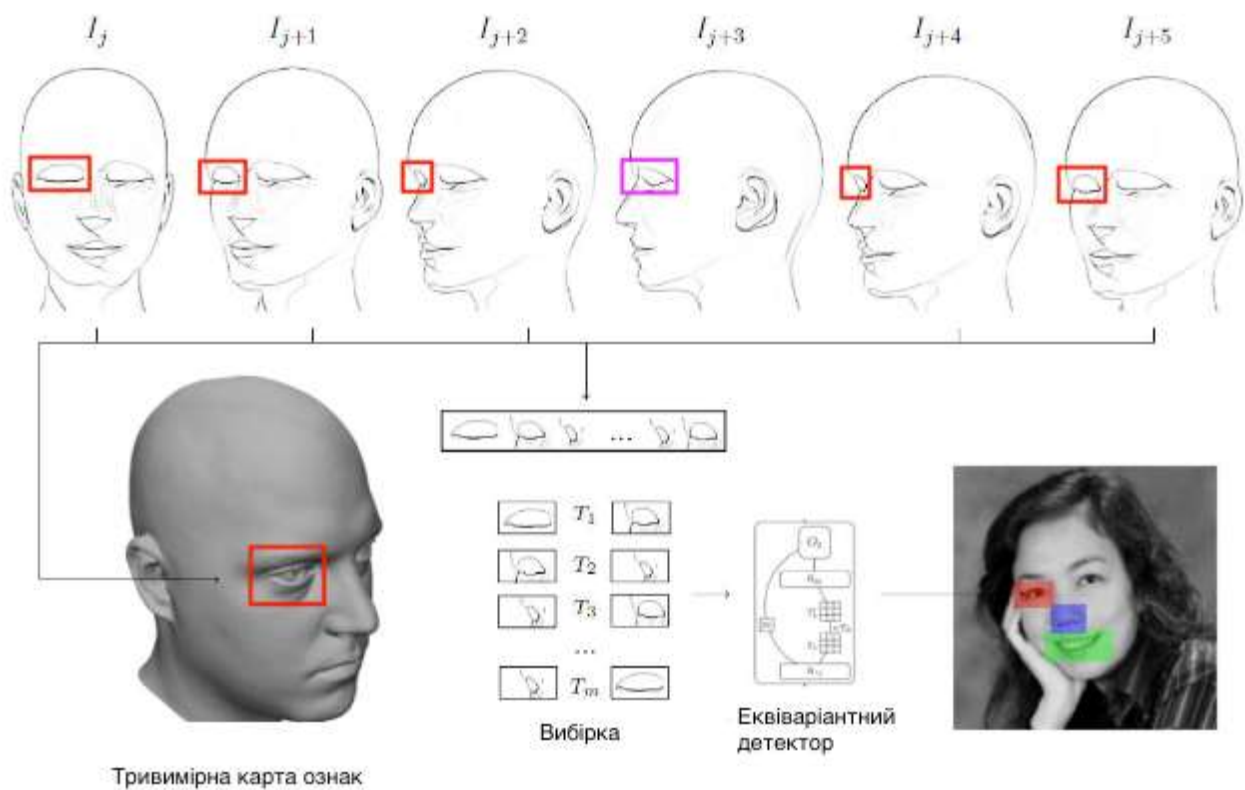


Рисунок 3.2 – Схема навчання моделі в потоці даних

3.3 Алгоритм виділення локальних ознак

Алгоритм виділення локальних ознак F_j^i виконує завдання пошуку локальних областей S_j^i j -того кадру для навчання еквіваріантних детекторів. Ключовою вимогою до таких локальних областей зображення є їх інформативність – тобто, отримані ознаки повинні максимізувати точність

розпізнавання в просторі можливих альтернативних локальних ознак зображення.

Для оцінки областей зображення з використанням даного критерію інформаційної ємності використовуємо поняття візуальної помітності (салієнтності) – дискримінаційною характеристикою сприйняття, що відображає ступінь відмінності фрагмента зображення від сусідніх ділянок (фону) та інших об'єктів в полі зору [29]. Розглянутий алгоритм виробляє оцінку показника помітності за допомогою сукупності двох взаємно доповнюють один одного показників:

- показник локальної інформаційної ентропії ділянки зображення по Шеннону $H'(x, y, R)$;

- показник відстані до найближчого екстремуму представлення яке масштабується $(\hat{x}, \hat{y}, \hat{t})$, де \hat{x}, \hat{y} – координати екстремуму, \hat{t} – значення масштабу.

Обидва показники розраховуються незалежно. Результат алгоритму являє собою перетин областей, знайдених кожним з методів. Далі розглянемо по черзі розрахунок обох показників і функцію їх об'єднання.

Оцінка локальної інформаційної ентропії. Локальна ентропія по Шеннону для області зображення являє собою суму добутків нормалізованих значень пікселів зображення в області на логарифм цих значень за основою два, відповідно до загальновідомої формули інформаційної ентропії. Виразим ентропію і фрагмент зображення функціями від трьох змінних: координат x, y , і радіусу області R :

$$H(x, y, R) = - \sum p(x, y, R) \log_2 p(x, y, R). \quad (3.2)$$

Локальні максимуми отриманої функції ентропії є шукані області зображення. Розглянемо також доповнення до розрахунку даного показника, представлене Т. Кадір і М. Бреді [30]: розрахуємо для кожної точки зображення, похідну від значень пікселів зображення по радіусу області:

$$W(x, y, R) = \frac{R^2}{2R - 1} \sum \frac{\partial}{\partial R} p(x, y, R). \quad (3.3)$$

Потім помножимо розрахований показник ентропії на це значення [52]:

$$H'(x, y, R) = H(x, y, R)W(x, y, R). \quad (3.4)$$

Для кожної точки зображення x, y потім розраховується показник локальної ентропії і вибирається такий R , для якого значення показника максимально. Зіставляючи ці показники для різних x, y , алгоритм потім ранжує ділянки зображення по спадаючій помітності і вибирає відповідні найбільш високих показників. Використання різних значень радіуса області R дозволяє здійснювати оцінку помітності методом, інваріантним до розміру зображення, і використовувати метод для об'єктів, зображених з різним масштабом. На практиці доцільно обмежувати можливі розміри радіуса як зверху, так і знизу, вибираючи R так, що $R_{min} < R < R_{max}$.

Для пошуку локальних ознак, що представляють собою обмежені замкнуті області зображення, пропонується виконати кластеризацію простору зображення, що може бути зроблено за допомогою наступних кроків:

Крок 1. Застосування глобального порогу τ до зображення (вибір пікселів зображення, інтенсивністю не нижче τ).

Крок 2. Вибір глобальної точки максимуму в області визначення помітності $H'(x, y, R)$.

Крок 3. Для знайденої точки максимуму проводиться пошук k найближчих сусідів на зображенні, де k – задана константа.

Крок 4. Для кожної з точок k визначається коефіцієнт варіації V_k і відстань до центральної точки D_k .

Крок 5. Кожна точка, для якої виконуються умови $D_k > \hat{R}$ та $V_k > V_t$, додається в масив виявлених фрагментів. Тут R – середнє значення радіуса області, V_k – задане граничне значення коефіцієнта варіації.

Крок 6. З k точок вибирається наступна за значенням помітності, і процес повторюється заново з кроку 2.

Отримана множина фрагментів високої помітності буде являти собою «слабкий» результат алгоритму, що включає в себе множину фрагментів, що характеризуються високою помітністю в порівнянні з навколишнім фоном, але не є структурно унікальними в межах зображення. Для знаходження структурно унікальних ділянок зображення використовується друга частина алгоритму, яка полягає в знаходженні екстремумів уявлення яке масштабується.

Пошук екстремумів масштабується уявлення проводиться за допомогою відомого методу оцінки різниці гауссіан зображення і обчислення визначника гауссіана. Для знаходження ділянок структурної помітності (що представляють собою локальну концентрацію високодеталізованих елементів зображення) використовуємо обчислення різниці Гауссіан [31]. Представимо зображення I у вигляді функції від двох змінних $f(x,y)$. Розглянемо уявлення що масштабується для зображення $L(x,y,t)$ – результат згортки $f(x,y)$ функцією Гаусса:

$$g(x, y, t) = \frac{1}{2\pi t^2} e^{-\frac{x^2+y^2}{2t^2}}, \quad (3.5)$$

$$L(x, y, t) = L(x, y, t) * f(x, y).$$

Застосування оператора Лапласа до даного уявлення що масштабується дозволяє отримати детектор, який приймає позитивні значення в областях яскравих ділянок неоднорідності радіуса $2t$. Для визначення ділянок різного радіусу використовується мультимасштабний підхід і різні значення t . Шуканий показник, таким чином, являє собою оператор Лапласа від $L(x,y,t)$, нормалізований за масштабом [31]:

$$\nabla_{norm}^2 L(x, y, t) = t(L_{xx} + L_{yy}). \quad (3.6)$$

Оскільки отримане уявлення що масштабується задовольняє рівняння дифузії $\partial_2 L = \frac{1}{2} \nabla^2 L$, стає можливою апроксимація нормалізованого лапласіана за допомогою різниці двох уявлень, згладжених гаусовими функціями з параметрами $t - \Delta t$ та $t + \Delta t$:

$$\nabla_{norm}^2 L(x, y, t) \approx \frac{t}{\Delta t} (L(x, y, t + \Delta t) + L(x, y, t - \Delta t)). \quad (3.7)$$

Отримана апроксимація є різницею Гауссіан функції $f(x, y)$. Використовуємо цю різницю для пошуку ділянок високої деталізації на зображенні. Для цього застосуємо до зображення фільтр Канни [32], що використовується для виділення меж, конвертуємо отримане зображення в бінарне, де $f(x, y) = 0$ в разі від'ємного значення детектора Канни, і 1 в іншому випадку. Тоді ділянки зображення, відповідні високої концентрації кордонів Канни, будуть оброблятися методом різниці Гауссіан як області помітності. Таким чином різниця Гауссіан можна використовувати для знаходження морфологічних деталей зображеного об'єкта, що не відрізняються забарвленням (інтенсивністю) і не дають помітних ефектів освітлення, але характеризуються відокремленою концентрацією деталізованих елементів текстури.

Для виявлення точок неоднорідною інтенсивності використовуємо відомий метод обчислення оператора гессіан. Для точки зображення з координатами (x, y) і обраного масштабу t визначник гессіан задається наступним чином [33]:

$$\det HL(x, y, t) = t^2 (L_{xx} L_{yy} - L_{xy}^2), \quad (3.8)$$

де HL є матрицею Гессе від уявлення що масштабується L . Точки максимуму для такого визначника є координатами і масштабними розмірами ділянок неоднорідності, дозволяючи визначнику виступати в якості диференціального детектора [34]:

$$(\hat{x}, \hat{y}, \hat{t}) = \underset{x,y,t}{\operatorname{argmax}} \det HL(x, y, t). \quad (3.9)$$

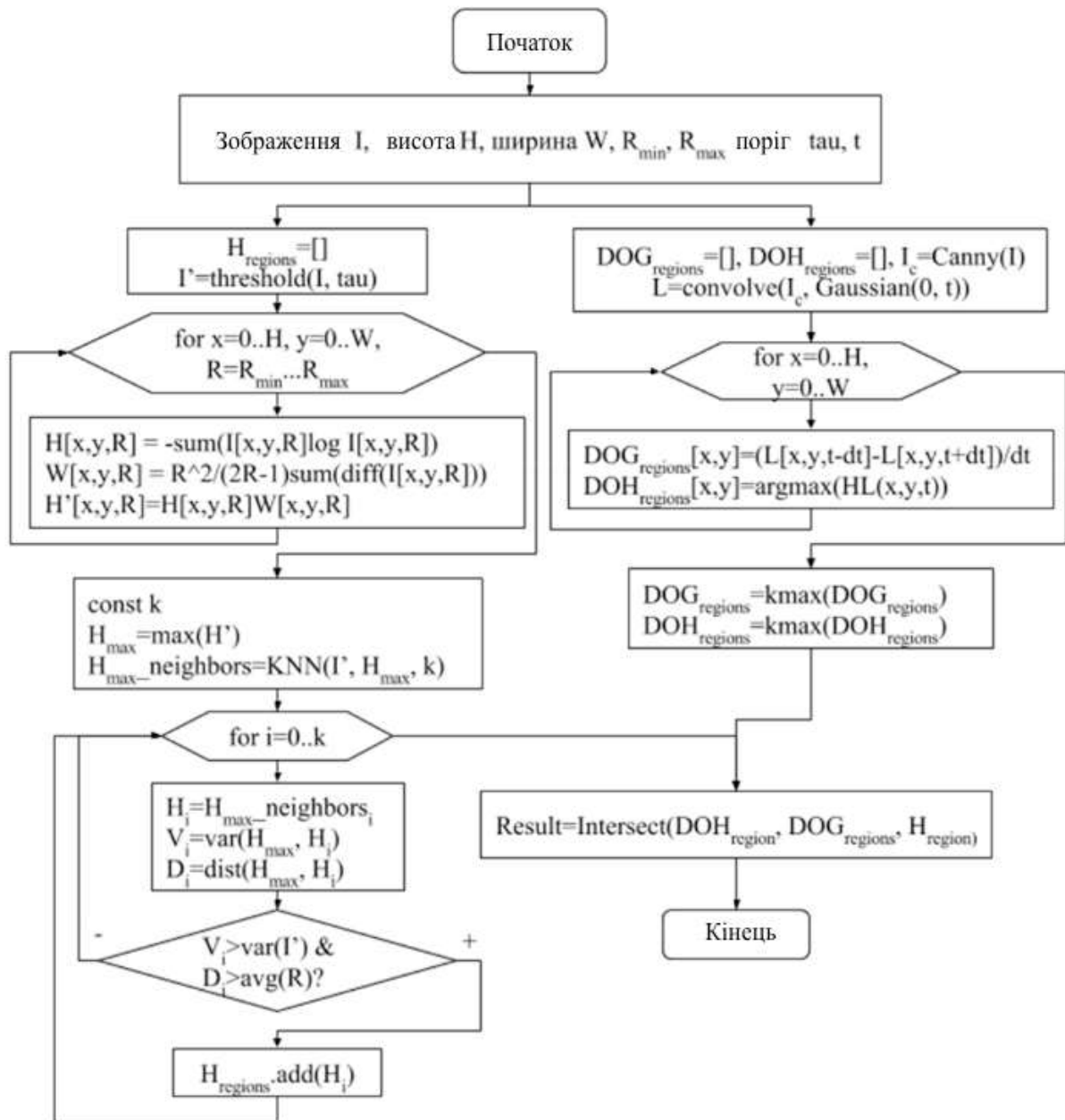


Рисунок 3.3 – Схема алгоритму виділення локальних ознак

Координати отриманих ділянок неоднорідності (\hat{x}, \hat{y}) і їх радіус \hat{t} отримані за допомогою визначника Гессе, коваріантного по відношенню до трансляції, обертання і масштабування зображення. Для ділянок зображення, отриманих таким чином, крім того, характерно дещо більша стійкість до афінних перетворень в порівнянні з різницею гаусіан. Крім того, метод визначника Гессе дозволяє виявити як яскраві, так і темні ділянки неоднорідності в зображенні, і має значно нижчу обчислювальну складність, що робить його підходящим кандидатом для запропонованого комбінованого методу.

Об'єднання показників для алгоритму виділення локальних ознак проводиться шляхом знаходження перетинають областей зображення, виявлених обома методами. Використання двох показників дозволяє відшукувати області зображення, які характеризуються властивостями помітності (максимальної локальної ентропією) і унікальності, як по структурному змістом, так і з порівняльної інтенсивності. Схема алгоритму виділення локальних ознак наведена на рис 3.3:

3.4 Розпізнавання зображення шляхом послідовної активації детекторів

Алгоритм розпізнавання складається з послідовної активації детекторів моделі, починаючи з першого рівня, на заданому зображенні. Нехай ϵ шукана функція розпізнавання $g: X \rightarrow Y$, аргументом якої є зображення $x_n \in X$, які представлені в вигляді вектору довжини n , а значеннями функції – множина класів (категорій) $y \in Y$, варійоване в залежності від поставленого завдання. Є підмножина пар аргументів і значень функції $D = \{(x_0, y_0), \dots, (x_m, y_m)\}$. Таким чином, ієрархія детекторів що послідовно активуються реалізує функцію $h: X \rightarrow Y$, яка апроксимує функцію g на всій її області визначення, в тому числі в точках, які не включені в D . Для розрахунку значення $h(x)$ зображення надходить на вхід першого шару навченої системи, потім виконується послідовна активація локальних детекторів на кожному із рівнів. Вихідна значення ієрархії являє

собою бінарне число, що визначає приналежність зображення до класу, при цьому вихідний рівень також проводить оцінку параметрів локалізації зображеного об'єкта, якщо значення активації дорівнює 1 (зображення успішно розпізнано і належить класу).

У такому вигляді розроблена система розпізнавання використовується для вирішення завдання унарною класифікації, коли множина класів $u \in Y$ представлена двома елементами і дорівнює $\{0,1\}$. Функція, таким чином, дорівнює 1 у випадках, коли зображення, що служить її аргументом, містить об'єкт, що належить класу, і 0 в іншому випадку. Для випадків, коли потрібно розпізнати зображення серед кількох можливих класів (завдання мультикласової класифікації), проводиться навчання окремої ієрархії детекторів для кожного конкретного класу, і потім проводиться послідовна перевірка зображення на позитивне відповідність кожної з них. В цьому випадку розглядається функція g' , визначена на множині X' , значеннями якої є множина класів, така що для обраного j -того класу $Y' = \{y_i, U_{j \neq i} y_j\}$.

Нехай дана навчена ієрархія детекторів M_c для деякого класу зображень c (наприклад, людських обличч), або кілька ієрархій для завдання мультикласового розпізнавання, і зображення I , яке необхідно розпізнати. Алгоритм розпізнавання складається з наступних кроків:

Крок 1. Вибираємо ієрархію детекторів M_c .

Крок 2. Для l -того рівня ієрархії, починаючи з першого, і для кожного детектору відповідного рівня $D_j^{(l)}$ складемо вектор вихідних даних $Z^{(l)}$.

Крок 3. Якщо $l = 0$, то заповнимо вектор вихідних даних наступним чином: розраховуємо значення функції ідентифікації детектору $d_{I_j}^{(l)}$ і значення впевненості $d_{p_j}^{(l)}$ для кожної ділянки зображення $I(x..x+w, y..y+h)$, де w та h відповідають розмірам локальної ділянки зображення, і проведемо конкатенацію вектору $Z^{(l)}$ з вихідними значеннями детектору для тих ділянок, де $d_{I_j}^{(l)} = 1$, та $d_{p_j}^{(l)} > t$, де t – обрана ступінь впевненості.

Крок 4. Якщо $l \neq 0$, вектор вхідних даних отримаємо як $Z^{(l)} = (D_0^{(l)}(Z^{(l-1)})) \parallel \dots \parallel (D_j^{(l)}(Z^{(l-1)})) \parallel (0_0 \parallel \dots \parallel 0_{N_l})$, де N_l – максимальна кількість можливих детекторів на рівні l .

Крок 5. Якщо для рівня l активації всіх детекторів цього рівня негативні (дорівнюють нулю), то зображення не належить класу c .

Крок 6. В іншому разі інкрементуємо значення l і повторюємо алгоритм з другого кроку.

Крок 7. Якщо активація останнього рівня позитивна, то зображення належить класу c . Інакше повторюємо алгоритм з першого кроку.

Використовуючи той факт, що низькорівневі локальні ділянки зображення (взяті з достатнім масштабом), як правило, представляють собою краю, межі і кути, слід зазначити має місце тенденцію до зростання різноманітності серед еквіваріантних детекторів на більш високих рівнях моделі. Експерименти показують, що кількість детекторів першого рівня при навчанні не перевищує ≈ 10 , при цьому ієрархії, навчені на об'єктах різних категорій, здатні розділяти між собою частину детекторів першого рівня, демонструючи ефект, що нагадує трансферне навчання або попереднє навчання без вчителя.

Для прикладів зображень об'єктів, розглянутих в даному дослідженні, таких як людські особи, число рівнів ієрархії, достатню для ефективного розпізнавання, не перевищує двох. У загальному випадку евристичне правило вибору кількості рівнів для моделі залежить від кількості ступенів свободи об'єкта, що цікавить, і не може бути однозначно встановлено по зображенню в силу впливу шуму і обмеженого простору спостереження.

Розроблений алгоритм розпізнавання відповідає вимогам поставленого завдання і дозволяє здійснювати розпізнавання зображених об'єктів з обчисленням параметрів інстанціювання в тривимірному просторі.

Схема алгоритму розпізнавання приведена на рис 3.4.

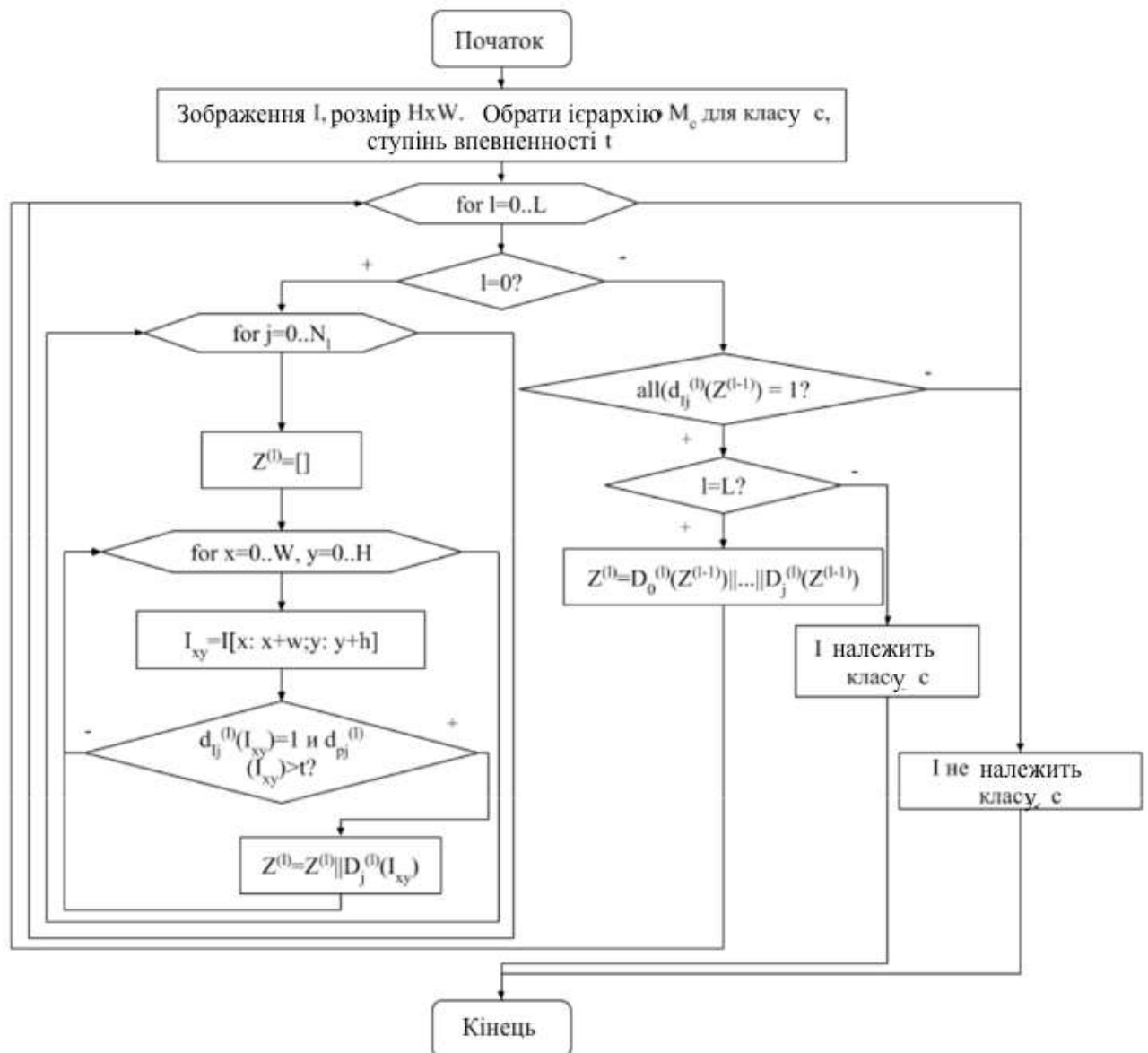


Рисунок 3.4 – Схема алгоритму розпізнавання

Застосування ієрархічного підходу в реалізації алгоритму дозволяє здійснювати розпізнавання шляхом активацій обмеженого числа локальних ділянок зображення. Використання алгоритмом розпізнавання еківаріантних детекторів дозволяє алгоритму об'єднувати в одній категорії зображення, різні з інформатико-теоретичний точки зору (такі, як зображення людського обличчя в профіль і в фас), але відповідні при цьому однієї категорії приналежності.

Використання локальних детекторів дозволяє досягти стійкості до оклюзії (перекритті частини зображення стороннім фоном). Оскільки детектори моделі

являють собою специфічні нейронні мережі, які навчаються компактним репрезентаціям даних, окремий локальний детектор забезпечує стійкість до ефектів шуму і розмиття в своїй області, дозволяючи алгоритму розпізнавання успішно справлятися з актуальними проблемами при розпізнаванні образів – нечіткістю вхідних даних і ефектами класової варіативності.

4 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ ТА ЕКСПЕРИМЕНТУ

4.1 Планування експерименту

Для оцінки ефективності розробленого алгоритму розпізнавання зображень проведемо експерименти трьох варіантів. алгоритму об'єднувати в одній категорії зображення, різні з інформатико-теоретичної точки зору (такі, як зображення людського обличчя в профіль і в фас), але відповідні при цьому одній категорії приналежності.

Варіант 1. Навчання локального еквіваріантного детектора в якості окремого класифікатора. Метою експериментів є визначення точності відновлення фрагментів зображення по компактним репрезентації, адекватність сформованих репрезентацій, їх відповідність значень трансформацій. Крім цього, тестується ефективність роботи детектора в якості опції-ідентифікатора. У завершальній частині проводиться зіставлення з іншими алгоритмами унарної класифікації та відновлення трансформації.

Варіант 2. Розпізнавання зображень за допомогою повноцінної ієрархії локальних еквіваріантних ознак. Об'єктами експерименту є показники помилки розпізнавання (в порівнянні з відомими алгоритмами), а також формуються моделлю просторові структури еквіваріантних детекторів.

Варіант 3. Робота системи на узагальнених відеофрагментах, що містять безліч об'єктів різних класів. Досліджується здатність моделі кластеризувати сцену, виявляючи окремі об'єкти і формуючи відповідні їм репрезентації.

Оскільки представлена модель вимагає для навчання обмежену вибірку з повно зв'язних відеофрагментів – тип даних, який не отримав поширення серед сучасних методів розпізнавання зображень – то навчання моделей для експерименту проводилося на даних, представлених автором. Навчальні вибірки включають в себе як відеофрагменти, отримані в результаті обробки

згенерованих з використанням комп'ютерної графіки тривимірних моделей, так і фрагменти живої зйомки навколишнього світу.

Оцінка ефективності розпізнавання людських обличч проводилася на відкритих базах даних LFW [35], HPID (Head Pose Image Database) [36]. Деякі моделі, що використовують альтернативні методи розпізнавання, надані бібліотекою Caffe [37]. Також для експериментів використовувалися дані, отримані за допомогою програми тривимірного моделювання FaceGen [38], що представляють собою марковані зображення людських обличч в різних орієнтаціях щодо камери.

Експериментальна перевірка розробленого методу розпізнавання проведена відповідно до рекомендацій по навчанню і порівнянні моделей розпізнавання образів в рамках апарату теорії розпізнавання образів і машина навчання [7, 8]. Для оцінки ефективності методів, які беруть участь в експерименті, використовувалися показники точності і повноти (що відображають помилки першого і другого роду).

Для оцінки результативних показників застосовувався метод розрахунку довірчих інтервалів. Відповідно до прийнятих практиками постановки експерименту в області розпізнавання і рекомендаціями по вибору рівня довіри, значення рівня довіри p вибрано рівним 0.95 [11]. Експериментальні дані продемонстрували, що для досягнення відповідного рівня досить вибірки близько 500 примірників, що для авторського методу відповідає короткому відеофрагменту, за часом не перевершує 15 секунд, Для якого частота зміни кадрів дорівнює 24 кадру в секунду, за умови використання методу пермутації.

Оскільки специфіка розробленого методу полягає у визначенні просторової локалізації об'єктів під різними кутами зору, то експериментальні вибірки для навчання і тестування містять набори відповідних зображень. Для оцінки здатності методу розпізнавання стійким чином обробляти об'єкти під впливом відповідних інваріантних перетворень, для кожного етапу експерименту результати обчислювалися за допомогою агрегації показників серії експериментальних перевірок. При цьому кожна серія складається з

наступних елементів: навчання і тестування на загальній випадковою вибіркою і навчання і тестування на кластеризованій вибірці. Кластеризація включає в себе розбивку вибірки по відеофрагментів, що містить зображення окремих об'єктів (конкретних людських осіб), при цьому дані відеофрагментів навчання не перетинаються з тестовими.

Таким чином, результатом експерименту є агрегований показник по серії з двох стадій навчання і тестування з урахуванням розрахованої на базі складових показників стандартної помилки.

В рамках експерименту проводиться зіставлення результатів розробленого методу і альтернативних, що використовуються в даний час як методи розпізнавання зображень. Вибрані альтернативи широко застосовуються як у виробництві (метод Віоли-Джонса [39], згорткові нейронні мережі [40]), так і в академічних дослідженнях. При виборі альтернатив використовувалися дані відкритих змагань по машинному навчання в області розпізнавання зображень, таких як ILSVRC [41]. Реалізації алгоритмів, що використовувалися в ході експерименту, представлені авторами відповідних методів, і отримані з відкритих джерел [7,42].

4.2 Оцінка ефективності окремого еквіваріантного детектору

В даному експерименті окремий еквіваріантний детектор, реалізований на базі нейронної мережі, навчається на послідовності кадрів обертання людського обличчя (модель згенерована за допомогою комп'ютерної графіки). Виявлені на етапі оптичного трекінгу ділянки помітності відповідають елементам, які виділяються в складі людського обличчя і можуть бути згруповані за типами «око», «ніс», «рот», «деталі зовнішнього контуру». На базі тестових відеофрагментів, отриманих за допомогою програми моделювання обличчя FaceGen [38] складена навчальна вибірка, що включає в себе фрагменти зображень кожного з локальних ознак.

На базі отриманої вибірки, згрупованої попарно, проводиться навчання розрідженого трансформуєчого автоенкодера. В експерименті брали участь чотири моделі автоенкодерів, зазначені в нижчезазначених результатах як А, В, С і D, і що відрізняються кількістю нейронів першого і останнього прихованих шарів.

- А – 36x36 нейронів;
- В – 64x64 нейрони;
- С – 256x256 нейронів;
- D – 768x768 нейронів.

Вибір кількості нейронів зроблено відповідно до відомих експериментів з навчання автоасоціативних нейронних мереж [43]. Трансформації представляють собою поворот камери по осях X і Z .

В рамках експерименту була виконана оцінка метопараметрів моделі з використанням методики крос-валідації. На рис 4.1 і 4.2 наведені графіки оптимізації функції ціни автоенкодера в залежності від змінюваних метопараметрів розрідженості s і β . По осі y зображена безрозмірна величина середньоквадратичного відхилення (помилки розпізнавання).

У таблиці 5.1 наведені результати розпізнавання для різних категорій фрагментів зображень, витягнутих з відеофрагментів (складання вибірки проводилося в автоматичному режимі, методом трекінгу; категорії промарковані назвами для спрощення і зручності аналізу).

З отриманих даних можна зробити висновок про оптимальність моделі – збільшення числа нейронів веде до незначного підвищення точності реконструкції, при цьому збільшуючи обчислювальну навантаження. Також більше число нейронів вимагає більшого числа ітерацій навчання для досягнення збіжності.

Таблиця 4.1 – Показники точності розпізнавання окремого детектору, %

	A (36x36)	B (64x64)	C (256x256)	D (768x768)
Обличчя: очі	79±4	95±3	96±3	96±3
Обличчя: ніс	82±3	92±3	93±3	94±2
Обличчя: рот	84±3	97±4	97±3	97±3
Обличчя: вуха	72±3	91±3	93±4	93±3
Обличчя: деталі контуру	92±4	96±3	97±4	98±3
Автомобіль: колеса	72±4	92±4	94±3	95±4
Автомобіль: фари	74±3	91±3	92±2	92±2
Автомобіль: лобове скло	80±3	94±4	95±2	95±3
Автомобіль: заднє скло	82±4	92±2	93±3	93±3
Автомобіль: бампер	79±4	90±3	92±4	93±4
Автомобіль: деталі контуру	78±2	89±4	90±3	91±3

Таблиця 4.2 – Показники точності і повноти розпізнавання окремого детектору на вибірці HPID, %

	Точність	Повнота
Обличчя: очі	96±3	95±3
Обличчя: ніс	98±4	96±4
Обличчя: рот	97±4	96±4
Обличчя: вуха	98±2	94±3
Обличчя: деталі контуру	99±2	95±3

Отримані результати узгоджуються з результатами, отриманими іншими дослідниками при навчанні автоасоціативних нейронних мереж [43]. Слід зазначити відмінності в показниках точності для окремих категорій зображень –

так, статистично більш прості елементи (кути, деталі контурів) розпізнаються успішно, ніж складні (окремі риси обличчя, наприклад, очі).

Друга частина експерименту проводилася на базі вибірки HPID з використанням обраної архітектури нейронної мережі. Відповідно до обраної методикою оцінки ефективності було проведено розрахунок показників точності і повноти. Результати відображені в таблиці 4.2.

Найбільш проблемними випадками для роботи еківаріантного детектора є аномальні ситуації, які не зустрілися в відеофрагменті – наприклад, оклюзія шуканої ділянки зображення сторонніми предметами (так, детектор демонструє зростання помилки до 10% при обробці зображень обличч, на яких присутні окуляри). Для зображень, семантично відповідних відеофрагменту, що використовувалися для навчання, детектор демонструє гнучкість в умовах обмежене вибірки – так, 2-3 відеофрагментів, отриманих з використанням комп'ютерної графіки, виявилось досить для розпізнавання вибірки HPID з присутністю 15 обличч, що відрізняються статтю, расою та особливостями міміки.

Отримані ваги демонструють тенденцію до формування розрідженого коду. Серед ваг практично не зустрічаються дублікати, при цьому характерна наявність декількох (3-5) «змішаних нейронів», візуально відповідних комбінації (накладення) різних ознак. За рахунок відповідних нейронів детектор здатний реконструювати (з використанням функції реконструкції) результати впливу трансформацій, відсутніх в його навчальній вибірці. Так, якщо вибірка містить два результату повороту об'єкта на кут φ_1 і φ_2 відповідно, то навчений детектор здатний передбачити проміжний результат трансформації $\varphi_1 < \varphi < \varphi_2$.

Крім ідентифікації ділянок зображення, вторинної функцією еківаріантного детектора є оцінка параметрів інстанціювання або оцінка позиції об'єкта. Для експериментальної перевірки точності цієї оцінки використовувалися змішані дані вибірки HPID і згенерованої вибірки зображень осіб FaceGen з подальшою пост-обробкою візуальними ефектами розмиття і

випадкової оклюзії. Для зіставлення результатів використовувалися такі методи як класичний алгоритм POSIT [9], і навчається з учителем регресор – випадковий ліс [44]. Оскільки прогноз трансформації є являє собою регресію, а не класифікацію, як показника ефективності обрана відносна похибка, виражена у відсотках. Результати представлені в таблиці 4.3.

Таблиця 4.3 – Показники точності розпізнавання окремого детектору, %

	POSIT	Випадковий ліс	Авторський метод
Обличчя: очі	25±3	17±5	11±3
Обличчя: ніс	27±3	14±2	12±3
Обличчя: рот	18±4	16±4	10±4
Обличчя: вуха	24±4	12±4	11±2
Обличчя: деталі контуру	20±2	14±3	14±4
Розмиття: очі	27±3	19±3	12±3
Розмиття: ніс	32±3	18±3	14±2
Розмиття: рот	25±3	15±3	12±4
Розмиття: вуха	21±2	17±2	14±3
Розмиття: деталі контуру	31±4	16±4	12±3
Оклюдія: очі	32±4	20±3	10±5
Оклюдія: ніс	28±2	16±5	14±2
Оклюдія: рот	31±5	16±2	13±3
Оклюдія: вуха	31±3	18±4	11±4
Оклюдія: деталі контуру	30±4	17±4	13±4

В результаті експерименту виявлено, що еквіваріантний детектор показує більш точну оцінку просторових параметрів об'єкта в порівнянні з альтернативними методами. Істотну перевагу над класичними методами

комп'ютерного зору пояснюється тим, що алгоритм POSIT вимагає для оцінки позиції наявності маркерів, які можуть бути спроектовані на об'єкт за допомогою методів епіполярної геометрії. Так, оцінка позиції людського обличчя, як правило, проводиться тільки для обмеженого простору трансформацій, в умовах, коли очі і рот знаходяться в полі зору камери. Для зображень, незначно відхиляються від фронтальних, ефективність еквіваріантного детектора можна порівняти з альтернативними методами.

4.3 Оцінка ефективності розпізнавання зображень

Було проведено дослідження ефективності простий дворівневої моделі на трьох категоріях зображень людських обличч: звичайні зображення осіб, зображення з штучним зашумлення за допомогою оклюзії і розмиття. Ефективність розпізнавати розраховувалася за допомогою показників точності і повноти і порівнювалася з аналогічними показниками альтернативних методів розпізнавання.

Оскільки в якості однієї з основних переваг даної моделі є еквіваріантність – здатність до ідентифікації об'єктів на зображенні в різних орієнтаціях – то експериментальні вибірки згруповані таким чином: розглядається деяка вихідна позиція об'єкта з координатами обертання (кутів Ейлера) $(0,0,0)$, при цьому в групу, що характеризується значеннями $(\varphi_{min}, \varphi_{max})$, входять зображення об'єктів, які зазнали трансформації обертанням $(\varphi_i, \varphi_j, \varphi_k)$, такий, що для будь-якого $\varphi \in (\varphi_i, \varphi_j, \varphi_k)$ є вірною нерівність $\varphi_{min} < \varphi < \varphi_{max}$.

Оскільки для цього експерименту потрібна значна кількість зображень об'єктів з різних кутів огляду, для його проведення була використана вибірка, отримана з використанням комп'ютерної графіки і генерації осіб програмою FaceGen.

Тестування розпізнавання осіб проводилося шляхом зіставлення методом Віоли-Джонса, класифікатора SVM в поєднанні з обчисленням гістограм

орієнтованих градієнтів і згорткових мереж, навченої на вибірці ImageNet. Навчені моделі були надані бібліотеками Caffe з OpenCV.

Таблиця 4.4: Відсоток коректно розпізнаних зображень обличч для групи різних орієнтацій, %

	Метод Віоли-Джонса	SVM+HOG	CaffeNet	Авторський метод
(0°,15°)	85±3	84±5	87±4	92±3
(15°,30°)	72±3	75±2	85±4	91±3
(30°,45°)	67±4	72±4	86±2	89±4
(45°,60°)	66±4	78±4	83±3	86±2
(60°,90°)	86±2	77±3	82±3	90±4
(90°,120°)	67±3	80±3	83±3	86±3

Метод Віоли-Джонса потребує наявність окремої стадії навчання для кожної орієнтації. В рамках експерименту використовувалася попередньо навчена модель, яка продемонструвала зіставні результати для фронтальної і профільної орієнтації осіб, але вкрай низькі результати в проміжних станах. При використанні SVM в поєднанні з методами зниження розмірності і підвищення інваріантності (гістограма орієнтованих градієнтів) для класифікатора характерно зниження точності в міру включення до вибірки зображень різних орієнтацій. Така поведінка пов'язана з тим, що модель в процесі навчання намагається виробити ознаки, які однаково підходять для всіх зображень у вибірці, в результаті отримуємо рівномірно розподілені невисокі значення точності. Серед порівнянних методів авторська модель поступається тільки глибинної згорткової мережі, здатної до навчання різних (що не змішуються) локальних ознак для різних орієнтацій об'єкта.

Слід зазначити, що на відміну від методу Віоли-Джонса, розроблена система розпізнавання дозволяє виділити комплексну структуру голови людини,

не обмежуючись ділянкою, що містить очі і рот. Такий ефект є наслідком використання ознак деталей контуру обличчя.

Порівняння результатів розпізнавання зображень обличч, підданих впливу розмиття проводилося за допомогою методів SVM + HOG, глибинної згорткової мережі і алгоритму випадкового лісу. Вибірка проводилася на базі CVLAB Dataset і включала в себе об'єкти, подані з різних кутів. Шумом є розмиття по Гаусу зі значенням $\sigma = 0.5 \dots 2.5$.

Таблиця 4.5 – Відсоток коректно розпізнаних розмитих зображень обличч для групи різних орієнтацій, %

	Випадковий ліс	SVM+HOG	CaffeNet	Авторський метод
(0°,15°)	72±4	72±4	85±2	91±4
(15°,30°)	73±3	74±3	81±3	88±3
(30°,45°)	75±2	76±4	80±4	88±4
(45°,60°)	72±3	74±4	82±4	89±4
(60°,90°)	71±3	73±4	83±3	91±3
(90°,120°)	73±3	72±4	83±3	87±4

При впливі ефекту розмиття дисперсія результатів по відношенню до орієнтації об'єкта знижується.

Як завершальній групи експериментальної вибірки використовувалися зображення людських обличч під впливом шуму оклюзії (часткового перекриття). Шум оклюзії згенерований за допомогою випадкового розміщення на зображеннях осіб непрозорих геометричних фігур. При генерації шуму оклюзії параметри розміщення фігур підібрані таким чином, щоб залишати відкритою мінімум дві третини площі зображення.

Таблиця 4.6 – Відсоток вірно розпізнаних частково перекритих зображень обличчя для групи різних орієнтацій, %

	Випадковий ліс	SVM+HOG	CaffeNet	Авторський метод
(0°,15°)	72±4	72±3	84±3	90±5
(15°,30°)	69±3	66±4	82±3	89±4
(30°,45°)	70±3	68±4	83±2	88±3
(45°,60°)	71±4	70±4	81±4	89±4
(60°,90°)	70±2	67±3	84±5	88±3
(90°,120°)	72±3	67±3	85±4	87±3

Як демонструють результати, оклюзія є суттєвою перешкодою для алгоритмів розпізнавання, що використовують компактні цілісні уявлення, таких як випадковий ліс і SVM+HOG – для деяких груп орієнтацій спостерігається падіння точності розпізнавання до 10%. Методи, які використовують локальні ознаки, таким як розроблена система і мережу CaffeNet, менш чутливі до часткового перекриття локальних областей зображення. У таких ситуаціях продуктивність представленого методу наближається до показників основного конкурента – згортковій мережі.

4.4 Локалізація об'єктів в композитних сценах

В даному експерименті перевірялася здатність розробленої системи за допомогою багаторівневої моделі аналізувати сцени, що містять безліч об'єктів, і автономно (без наявності вчителя) класифікувати їх між собою. Як вибірки використовувалися відкриті дані, що містять відеозаписи камер спостереження за дорожнім трафіком. Оцінка продуктивності алгоритму проводилася в такий спосіб: підраховувалася кількість об'єктів в кадрі, аналогічним чином вручну

вироблялося з розбиття по групах, потім ці значення зіставлялися зі значеннями знайденими алгоритмом.

Таблиця 4.7 – Оцінка аналізу композитних сцен, % від фактичного значення

	Об'єктів знайдено	Груп об'єктів знайдено
Запис трафіку 1	70	96
Запис трафіку 2	72	100
Запис трафіку 3	75	75
Запис трафіку 4	74	75

Слід зазначити, що на відміну від ситуації з контрольованим трекінгом і рухом камери навколо об'єкта, алгоритм в режимі аналізу композитних сцен зі сторонніх відеофрагментів позбавлений інформації про фактичний рух об'єкта навколо камери. Так, система не здатна без додаткової інформації визначити, що рух від камери автомобіля що віддаляється являє собою переміщення по площині, не паралельній площині сенсору камери. У загальному випадку в такій ситуації алгоритм знаходиться в умовах невизначеності і здатний сформувати тільки обмежене уявлення об'єкта за допомогою моделі еквіваріантних детекторів. Існує можливість, використовуючи методи проективної геометрії, ввести деякі припущення в розрахунок оптичного трекінгу – так, для розглянутих експериментальних випадків стадія оптичного трекінгу була доповнена умовою жорсткості (фіксованою форми) об'єктів в кадрі. При цьому трекер інтерпретував зменшення розмірів об'єкта на зображенні як видалення від камери, і на підставі відносної зміни площі об'єкта і показників оптичного потоку обчислював напрямок його руху.

ВИСНОВКИ

Під час виконання кваліфікаційної роботи виконано аналіз існуючих методів розпізнавання зображень, класифіковано існуючі підходи до розпізнавання. Визначено проблему розпізнавання тривимірних об'єктів під впливом інваріативних просторових перетворень і запропоновано метод вирішення проблеми на основі практики існуючих досліджень і проведених дослідів.

Проаналізовано етапи розпізнавання зображень у відеопотоці, запропоновано етапи попередньої обробки кадрів для ефективного аналізу зображень та збільшення продуктивності системи.

Розроблено теоретичні положення нейронно-мережевої системи розпізнавання з використанням локально-еквіваріантної моделі репрезентації зображень, що дозволяє розпізнавати об'єкти і здійснювати їх просторову локалізацію. Сформовано принципи організації ієрархічної структури моделі і компонентів окремих рівнів ієрархії, описано математичну модель локального еквіваріантного детектора як елементарної одиниці ієрархічної структури репрезентації

Розроблено метод навчання системи розпізнавання в потоці даних, що дозволяє отримувати і обробляти ключову для розпізнавання інформацію про тривимірну структуру об'єкта, оцінювати його положення в просторі і локалізувати об'єкт на зображенні. Використовуючи запропонований метод, система здатна формувати ієрархію локальних ознак об'єкта, навчаючись на візуальному досвіді автономного інтелектуального агента з відеокамерою, що робить її придатною для обробки інформації в різних додатках штучних системах комп'ютерного зору.

Розроблено алгоритм розпізнавання зображень об'єктів під довільними кутами огляду на основі ієрархічної моделі представлення об'єктів. Розроблений алгоритм використовує послідовну активацію локальних еквіваріантних детекторів ієрархії для прийняття рішення про приналежність зображень до категорії розпізнавання, що робить його стійким для деяких категорій

візуального шуму, як-от розмиття і оклюзія. Алгоритм розпізнавання використовує зворотний зв'язок для корекції помилок першого і другого роду в нижніх рівнях ієрархії завдяки реконструкції високорівневими детекторами взаємного розташування локальних ознак попередніх рівнів.

Проведено експеримент, у процесі якого ефективність розробленої системи порівнювалася з показниками альтернативних відомих методів розпізнавання. Під час експерименту показана ефективність розробленої системи в порівнянні з альтернативами, при розпізнаванні зображень об'єктів, орієнтованих під довільними кутами по відношенню до камери. Також продемонстрована стійкість розробленої системи до візуальних шумів оклюзії і розмиття. Перевірка показала перевагу розробленої системи. Точність розпізнавання у разі використання запропонованого методу зростає на 3-5%. Розроблена система розпізнавання більш стійка до локального шуму: для зображень, підданих розмиттю і оклюзії, падіння точності розпізнавання розробленої системи становить 3-6% проти 5-10% у альтернативних відомих методах.

Результати досліджень, наведені в роботі, становлять практичний інтерес при проектуванні систем управління та обробки інформації в області комп'ютерного зору і розпізнаванні зображень, для тих завдань, де існує необхідність визначення просторових параметрів зображених об'єктів. Тобто метод доцільно використовувати в системах відеоспостереження для автоматичної ідентифікації класів об'єктів, наприклад на дорожніх магістралях чи інших моніторингових пунктах. Метод навчання системи базується на використанні відеозаписів, тобто система може навчатися на встановленому пункті моніторингу для адаптації до локальних образів. Також результати дослідження можуть використовуватись для засобів аутентифікації на базі їх реакції [45].

Для використання системи для моніторингу в режимі реального часу слід проаналізувати і спроектувати систему на базі розподілених обчислень для паралельного аналізу кадрів, адже результати локальних експериментів показують труднощі при багатьох кадрах в секунду, активація детекторів багатьох рівнів потребує значних обчислювальних ресурсів.

Тема дослідження методів розпізнавання об'єктів, що швидко рухаються пройшла апробацію на Міжнародній науково-практичній конференції «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В СОЦІОКУЛЬТУРНІЙ СФЕРІ, ОСВІТІ ТА ЕКОНОМІЦІ» (м. Київ) [46] та на Міжнародному молодіжному форумі «Радіоелектроніка і молодь в ХХІ віці» [47]. Частина наведених у роботі результатів була використана під час оформлення конкурсної роботи, яка приймала участь у Всеукраїнському конкурсі студентських наукових робіт 2019/2020 навчального року в галузі «Комп'ютерні науки», робота зайняла 2 місце у II турі, ксерокопію диплома про нагородження наведено у додатку В.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Локтєв Д.А., Кочнев В.А., Локтєв А.А. Вивчення функцій розмиття зображення у вигляді інформативного параметра стану і поведінки аналізованого об'єкта // Динаміка складних систем – ХХІ століття, No 2, 2020. – С. 16-27.
2. Loktev D. Image blurring function as an informative criterion / Loktev A., Loktev D. // Advances in Intelligent Systems and Computing, vol. 1258 AISC, 2021. – P. 173-183.
3. Bellman R.E. Perturbation techniques in mathematics, engineering and physics. Courier Corporation, 2003.
4. Aly M., Welinder P., Munich M., Perona P. Towards automated large scale discovery of image families // Computer Vision and Pattern Recognition. 2009. pp. 9-16.
5. Bengio Y., Erhan D. Why does unsupervised pre-training help deep learning? // The Journal of Machine Learning Research. 2010. No. 11. pp. 625-660.
6. Bengio Y. Learning deep architectures for AI // Foundations and trends in Machine Learning. 2009. No. 1. pp. 1-127.
7. Bishop C.M. Neural networks for pattern recognition. Oxford: Oxford university press, 1995.
8. Bishop C.M. Pattern recognition and machine learning. New York: Springer, 2006.
9. Duin R.P., Pekalska E. Open issues in pattern recognition // Computer Recognition Systems. 2005. pp. 27-42.
10. Grauman K., Leibe B. Visual object recognition. Morgan & Claypool Publishers, 2010.
11. Duda R.O., Hart P.E., Stork D.G. Pattern classification. John Wiley & Sons, 2012.
12. Hearst M.A., S.T. D., Osman E. Support vector machines // Intelligent Systems and their Applications, No. 4, 1998. pp. 18-28.
13. Bengio Y. Learning deep architectures for AI // Foundations and trends in Machine Learning. 2009. No. 1. pp. 1-127.

14. Sebe N. Machine learning in computer vision. New York: Springer Science & Business Media, 2005.
15. Charles E. Jacobs A. Finkelstein D.H. Sales in Fast Multiresolution Image Querying // Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98195.
16. АЛЬТМАН Е.А., Захаренко Є.І. Швидкий алгоритм обчислення двовимірної кореляції для відеонаблюдения // Доповіді ТУСУР. 2015. №2 (36) С.119-124.
17. Bay H., Ess A., Tuytelaars T., Van Gool L. SURF: Speeded up robust features // Computer Vision and Image Understanding. – 2008. – V. 110. – P. 346–359.
18. Leutenegger S. BRISK: Binary Robust invariant scalable keypoints / S. Leutenegger, M. Chli, R.Y. Siegwart // Proceedings of the 2011 International Conference on Computer Vision (ICCV '11), 6 November 2011, P. 2548-2555.
19. Фісенко В.Т. Комп'ютерна обробка і розпізнавання зображень: навчальний посібник / В.Т. Фісенко, Т.Ю. Фісенко – СПб: СПбГУ ІТМО, 2008. – 192 с.
20. Kanjar D. Image Sharpness Measure for Blurred Images in Frequency Domain / D. Kanjar, V. Masilamani // International Conference on Design and Manufacturing. – Procedia Engineering, 2013. – P. 149 – 158.
21. Hubel D.H., Wiesel T.N. Receptive fields and functional architecture of monkey striate cortex // The Journal of physiology. 1968. No. 195 (1). pp. 215–243.
22. Tan X. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions / X. Tan, B. Triggs // IEEE Transactions on image processing, Vol. 19, No 6, June 2010. – P. 1635-1650.
23. Гонсалес Р. Цифрова обробка зображень / Р. Гонсалес, Р. Вудс // Москва: Техносфера, 2012. – 1104 с.
24. T. Sund & A. Moystad: Sliding window adaptive histogram equalization of intra-oral radiographs: effect on diagnostic quality // Dentomaxillofac Radiol. 2006. P 133.
25. Changizi M.A. Perceiving the present and a systematization of illusions // Cognitive science. 2008. No. 32 (3). pp. 459-503.

26. Serfling R. Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization // *Journal of Nonparametric Statistics*, No. 22.7, 2010. pp. 915-936.
27. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg AC, Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge. arXiv, 2010.
28. Hinton G.E., Krizhevsky A., Wang S.D. Transforming auto-encoders // *Artificial Neural Networks and Machine Learning* ICANN 2011.2014.pp. 44-51.
29. Purves D., Lotto R.B. Why we see what we do: An empirical theory of vision. Sinauer Associates, 2003.
30. Kadir T., Zisserman A., Brady M. An affine invariant salient region detector // *Computer Vision-ECCV*. 2004. pp. 228-241.
31. Bundy A., Wallen L. Difference of Gaussians // *Catalog of Artificial Intelligence Tools*. 1984. pp. 30-35.
32. Canny J. A computational approach to edge detection // *Pattern Analysis and Machine Intelligence*, IEEE Transactions on. 1986. No. 6. pp. 679-698.
33. Lindeberg T. Scale-space theory: A basic tool for analyzing structures at different scales // *Journal of applied statistics*. 1994. No. 21.1-2. pp. 225-270.
34. Lindeberg T. Scale invariant feature transform // *Scholarpedia*. 2012.
35. Huang G.B., Ramesh M., Berg T., Learned-Miller E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments // *Technical Report 07-49*, University of Massachusetts. 2007. No. 1 (2). pp. 3-37.
36. Gourier N., Hall D., Crowley J.L. Estimating face orientation from robust detection of salient facial structures // *FG Net Workshop on Visual Observation of Deictic Gestures*, 2004.
37. Gourier N., Hall D., Crowley J.L. Caffe: Convolutional architecture for fast feature embedding // *FG Net Workshop on Visual Observation of Deictic Gestures*. 2004. pp. 1-9.
38. Singular Inversions. FaceGen modeller (Version 3.3) [computer software] // *Singular Inversions*, 2008.
39. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features, No. 1, 2001. pp. 502-511.

40. Szegedy C. Going deeper with convolutions // arXiv. 2014. No. 1409.4842.
41. Hinton G.E. A practical guide to training restricted Boltzmann machines // Momentum. 2010. No. 9 (1). P. 926.
42. Hubel D.H., Wiesel T.N. Brain and visual perception. ISBN13, 2005.
43. Hinton G.E., Salakhutdinov R.R. Reducing the dimensionality of data with neural networks // Science. 2006. No.313 (5786). pp. 504-507.
44. Leo B. Random forests // Machine learning. 2001. No. 45 (1). pp. 5-32.
45. Nechiporenko A.S., Gubarenko E.V., Gubarenko M.S. (2019) Authentication of users of mobile devices by their motor reactions. Telecommunications and Radio Engineering, 78 (11). pp. 987-1003. DOI: 10.1615 / TelecomRadEng.v78.i11.60.
46. Антонюк М.В., Губаренко Є.В. Інформаційні технології в соціокультурній сфері, освіті та економіці : матеріали Міжнар. наук.-практ. конф. Київ : КНУКіМ, 2021
47. Антонюк М.В. Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті». Конференція «Інформаційні інтелектуальні системи» : Міжнар. молодіжний форум Харків, 2019: Том 6 С. 204-205.