Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)
Кафедра Інформатики
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

ДОСЛІДЖЕННЯ КЛАСИФІКАТОРА ЗОБРАЖЕНЬ ІЗ ВИКОРИСТАННЯМ АНСАМБЛЕВИХ ЗАСОБІВ АНАЛІЗУ СКЛАДУ СТРУКТУРНОГО ОПИСУ
(тема)

Виконав:
студент 2 курсу, групи ІНФм-21-1
Жадан О.В.
(прізвище, ініціали)

Спеціальності 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-професійна

Освітня програма Інформатика
(повна назва освітньої програми)

Керівник проф. Гороховатський В.О.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ Кобилін О.А.
(підпис) (прізвище, ініціали)

2022 р.

Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)

Кафедра Інформатики
(повна назва)

Рівень вищої освіти другий (магістерський)

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

Тип програми освітньо-професійна

Освітня програма Інформатика
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«____»_____ 2022 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Жадану Олексію Віталійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження класифікатора зображень із використанням ансамблевих засобів аналізу складу структурного опису

затверджена наказом по університету від 9 листопада 2022 року № 1469Ст

2. Термін подання студентом роботи до екзаменаційної комісії 21 листопада 2022 р.

3. Вихідні дані до роботи ознаки візуальних об'єктів, множина ключових точок, дескриптор OBR, поняття медоїду множини, інформативність дескриптору, відстань Хемінга, статистичний розподіл даних, багатокомпонентна модель даних, класифікація зображень.

4. Перелік питань, що потрібно опрацювати в роботі
1. Огляд теоретичного матеріалу щодо класифікації зображень зі структурним підходом.
2. Розробка способів побудови інформативних множин центрів класів.
3. Дослідження алгоритмів класифікації з підтримкою декількох центрів для класу.
4. Програмна реалізація впроваджених процедур.
5. Огляд отриманих результатів та формулювання висновків.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) постановка задачі, схеми моделей класифікації з підтримкою багатокомпонентної моделі даних, база еталонних зображень, виділені дескриптори ORB.

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

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Консультант з дотримання діючих стандартів та норм	Доцент Творошенко І.С.		

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	09.11.22	**виконано**
2	Аналіз завдання, підбір літератури	09.11.22	**виконано**
3	Аналіз літератури з досліджуваної проблеми	10.11.22-11.11.22	**виконано**
4	Аналіз технічних засобів	12.11.22	**виконано**
5	Розробка методів класифікації з використанням багатокомпонентних моделей даних	13.11.22-14.11.22	**виконано**
6	Програмна реалізація	14.11.22-15.11.22	**виконано**
7	Оформлення пояснювальної записки	16.11.22-17.11.22	**виконано**
8	Перевірка на плагіат	18.11.22	**виконано**
9	Рецензування	21.11.22-25.11.22	**виконано**
10	Підготовка презентації та доповіді	26.11.22-29.11.22	**виконано**
11	Занесення роботи в електронний архів	30.11.22	
12	Попередній захист кваліфікаційної роботи	30.11.22	

Дата видачі завдання 9 листопада 2022 р.

Студент _____
(підпис)

Керівник роботи _____ проф. Гороховатський В.О.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ/ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 64 с., 18 табл., 16 рис., 47 джерел.

КОМП'ЮТЕРНИЙ ЗІР, РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ, СТРУКТУРНІ МЕТОДИ КЛАСИФІКАЦІЇ, ДЕСКРИПТОР ORB, СИСТЕМА ЦЕНТРІВ ДАНИХ, БАГАТОКОМПОНЕНТНА МОДЕЛЬ.

Об'єктом дослідження є методи класифікації зображень з використанням множин дескрипторів ключових точок у системах комп'ютерного зору.

Метою дослідження є впровадження технологій класифікації зображень на підставі багатокомпонентної моделі даних на множині бінарних дескрипторів для опису еталонних зображень.

Використано дескриптори ORB, апарат теорії множин і векторного простору, метричні моделі для визначення релевантності множин багатовимірних векторів, елементи теорії ймовірностей та програмне моделювання. Дослідження направлені на застосування багатокомпонентної моделі даних у структурних підходах прийняття рішень про клас об'єкту задля покращення класифікаційних властивостей.

У результаті програмно реалізована та досліджена модель класифікації, зроблений висновок щодо її ефективності.

COMPUTER VISION, IMAGE RECOGNITION, STRUCTURAL CLASSIFICATION METHODS, ORB DESCRIPTOR, DATA CENTER SYSTEM, MULTICOMPONENT MODEL.

The object of research is methods of image classification using sets of key point descriptors in computer vision systems.

The purpose of the research is to introduce image classification technologies based on a multicomponent data model on a set of binary descriptors to describe reference images.

The ORB descriptors, the apparatus of set theory and vector space, metric models for determining the relevance of sets of multidimensional vectors, elements of probability theory and software modeling were used. The research is aimed at applying a multicomponent data model in structural approaches to making decisions about the class of an object to improve classification properties.

As a result, the classification model was implemented and studied in software, and a conclusion was made about its effectiveness.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

КТ – ключова точка

SIFT – Scale-Invariant Feature Transform (масштабонезалежне перетворення ознак)

SURF – Speeded Up Robust Features (прискорені надійні ознаки)

MSER – Maximally Stable Extremal Regions (максимально стабільні екстремальні області)

BRISK – Binary Robust Invariant Scalable Keypoints (двійкові надійні інваріантні масштабовані ключові точки)

FREAK – Fast Retina Keypoints (швидкі ключові точки сітківки)

FAST – Features from Accelerated Segment Test (ознаки прискореного сегментного тесту)

BRIEF – Binary Robust Independent Elementary Features (двійкові надійні незалежні елементарні функції)

ORB – Oriented FAST and Rotated BRIEF (орієнтований FAST і обернутий BRIEF)

JPEG – Joint Photographic Experts Group (об’єднана експертна група з фотографій)

LINQ – Language-Integrated Query (інтегрований у мову запит)

ВСТУП

Розпізнавання зображень – це завдання ідентифікації об’єктів на зображенні та розпізнавання, до якої категорії належить візуальний образ. Коли людина бачить якийсь об’єкт або сцену, вона автоматично ідентифікує що перед нею та пов’язує це з окремими визначеннями. Однак візуальне розпізнавання є дуже складною задачею для машин, які оперують одними 0 та 1.

Розпізнавання зображень за допомогою штучного інтелекту є давньою проблемою досліджень у сфері комп’ютерного зору. Хоча різні методи імітації людського зору розвивалися з часом, спільною метою розпізнавання зображень є класифікація виявлених об’єктів у різні категорії (визначення категорії, до якої належить зображення). Це є досить складним питанням, що містить у собі набір комплексних перетворень та обчислень для визначення належності до тієї чи іншої групи подібних візуальних об’єктів. На сьогодні, незважаючи на велику популярність, сучасні системи комп’ютерного зору все рівно залишаються не дослідженні у повній мірі і потребують дієвих та дешевих за ресурсами алгоритмів розпізнавання образів. У цьому і полягає актуальність даного дослідження.

У структурних підходах, прийняття класифікаційних рішень про клас зображеного тіла відбувається на підставі використання множин дескрипторів ключових точок. При цьому сукупність векторів, яка являється описом візуального об’єкту, зіставляється з набором векторів, що відображають інформацію про центри класів. У ході алгоритму здійснюється обчислення оцінки релевантності об’ємних багатовимірних множин даних та бази еталонних зображень.

У роботах [1–4] дослідники запропонували дієві методи для обчислення значень релевантності за рахунок формування векторів статистичних розподілів між центрами (один центр на клас) для кожного елементу множини опису візуального об’єкту. Дані алгоритми є більш дієвими для

багатовимірних просторів де спостерігається певне скупчення елементів, але, попри те, немає ніяких обмежень у структурі досліджуваних даних. Варто зауважити, що ефективність класифікації з використанням цих методів значною мірою залежить від складу описів еталонів.

Через необмежене різноманіття задач та видів досліджуваних даних, визначення класу об’єкта на підставі єдиного центру може зменшити точність та необхідну ефективність прийнятого рішення. Таким чином, можна застосувати більш розширену модель опису класу, а саме з декількома класами (багатокомпонентне подання) для більше детального та поглибленого представлення еталонів. Перевірка працездатності даного підходу є головним завданням цього дослідження.

1 ТЕХНОЛОГІЇ СТРУКТУРНОГО РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ

1.1 Дескриптори ключових точок як ознаки візуальних об’єктів

Зіставлення зображень є фундаментальним аспектом багатьох проблем комп’ютерного зору, включаючи розпізнавання об’єктів, створення тривимірної структури з кількох зображень, стереозбіг, а також відстеження маршруту та сегментації. Даний процес складається з виявлення локальних областей ознак, їх опису, зіставлення, оцінки релевантності та вирівнювання двох зображень. Етап виявлення та опису є критичним, оскільки він визначає, чи можна виконати подальшу обробку чи ні, та наскільки коректним буде результат.

У рамках даного дослідження для опису зображення чи візуального об’єкту використовується механізм *ключових точок (КТ)* та їх *дескрипторів*. Це найпопулярніший спосіб характеристики зображень, оскільки їх легко знайти та описати порівняно з іншими ознаками.

Ключові точки (також відомі як особливі точки або точки інтересу) – це множина точок з деякими характеристиками, що відрізняють їх від інших точок на зображенні, містять вагому для аналізу інформацію та можуть бути стабільно виявлені [5]. Як правило, це кутові, граничні точки, з різким перепадом кольорів або яскравості, тощо. На рисунку 1.1 наведений приклад знайдених КТ на зображенні.

Дескриптор ключової точки – це числовий опис ділянки навколо КТ [5]. Сама по собі ключова точка не несе ніякої інформації, окрім їх позиціонування на зображенні, тобто координат X та Y. Через це були створені дескриптори, побудова яких полягає у аналізі околу КТ з врахуванням яскравості, інтенсивності кольорів та інших властивостей, які дозволять ідентифікувати цю точку.

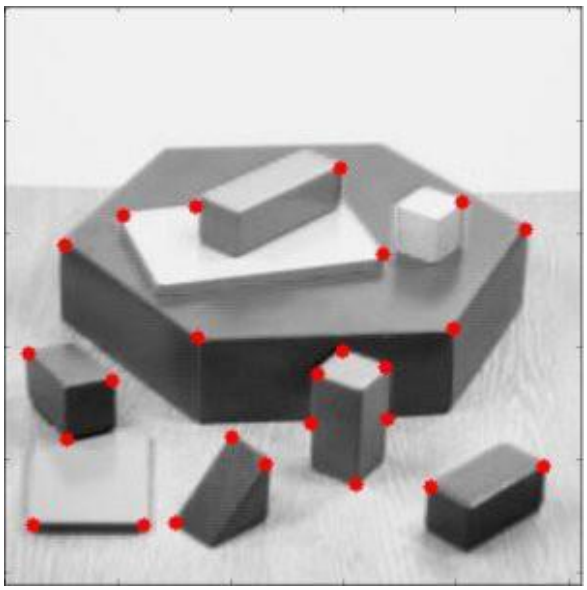

Рисунок 1.1 – Зображення з виділеними КТ [6]

Дескриптори мають багато властивостей, які роблять їх придатними для зіставлення різних візуальних об’єктів чи сцен. Вони є інваріантними щодо масштабування та обертання зображення, а також частково інваріантними щодо змін освітлення та точки огляду 3D-камери. Добре локалізовані як у просторовій, так і в частотних областях, зменшуючи ймовірність перешкод через оклюзію або шум. З типових зображень, за допомогою ефективних алгоритмів, можна отримати велику кількість дескрипторів. Найважливіше те, що вони мають вагому відмінність, що дозволяє з високою ймовірністю правильно зіставити одну точку з великою множиною ознак, забезпечуючи основу для розпізнавання об’єктів.

Як правило, алгоритми побудови дескрипторів містять у собі дві складові – роботу детектора КТ та обчислення значень дескрипторів. Витрати на вилучення множини дескрипторів мінімізовано завдяки підходу послідовного фільтрування, у якому більш дорогі операції застосовуються лише в місцях, які пройшли початковий тест. Нижче наведено основні етапи обчислень для створення набору характеристик зображення [7].

Етап 1. *Вибір піку в масштабному просторі*. На першому етапі обчислення має здійснюватися пошук у всіх масштабах і розташуваннях зображення. Ідентифікація потенційних точок інтересу, незмінних відносно

масштабу та орієнтації, може бути ефективно реалізована за допомогою різниці Гауса.

Етап 2. *Локалізація ключової точки*. Для кожного потенційного місця розташування КТ застосовується детальна модель для встановлення розташування, масштабу та контрасту. Ключові точки вибираються на основі показників їх стабільності.

Етап 3. *Призначення орієнтації*. Для кожної КТ на основі локальних властивостей зображення призначається одна або кілька орієнтацій. Усі майбутні операції виконуються відносно орієнтації, масштабу та розташування, що забезпечує інваріантність цих перетворень.

Етап 4. *Дескриптор ключової точки*. У вибраному масштабі в області навколо кожної ключової точки вимірюються локальні градієнти зображення та перетворюються на представлення, яке враховує спотворення форми та зміну освітлення.

Важливим аспектом цього підходу є те, що він створює велику кількість дескрипторів, які щільно охоплюють зображення в повному діапазоні масштабів і місць розташування. Типове зображення розміром 500 на 500 пікселів дасть приблизно 2000 стабільних характеристик (хоча ця кількість залежить як від вмісту зображення, так і від вибору різних параметрів детекторів).

Для розпізнавання зображень ознаки спочатку витягуються з набору еталонних зображень, а потім зберігаються в базі даних. Нове зображення зіставляється шляхом окремого порівняння кожного його дескриптора з побудованою базою, з метою пошуку відповідності їхніх ознак на основі певних метрик. Дескриптори КТ мають високу відмінність, що дозволяє з високою ймовірністю знайти правильний відповідник, однак на безладному зображенні багато дескрипторів не матимуть правильних збігів, що породжує значну кількість помилкових зіставлень. Правильні збіги можна відфільтрувати з повного набору, визначивши підмножини КТ, які збігаються з об’єктом і його розташуванням, масштабом і орієнтацією на новому

зображенні. Імовірність того, що кілька дескрипторів випадково збігаються за цими параметрами, набагато нижча, ніж ймовірність того, що збіг є помилковим.

Існує велика кількість алгоритмів для побудови дескрипторів. На рисунку 1.2 зображена графічна інтерпретація деяких двійкових дескрипторів.

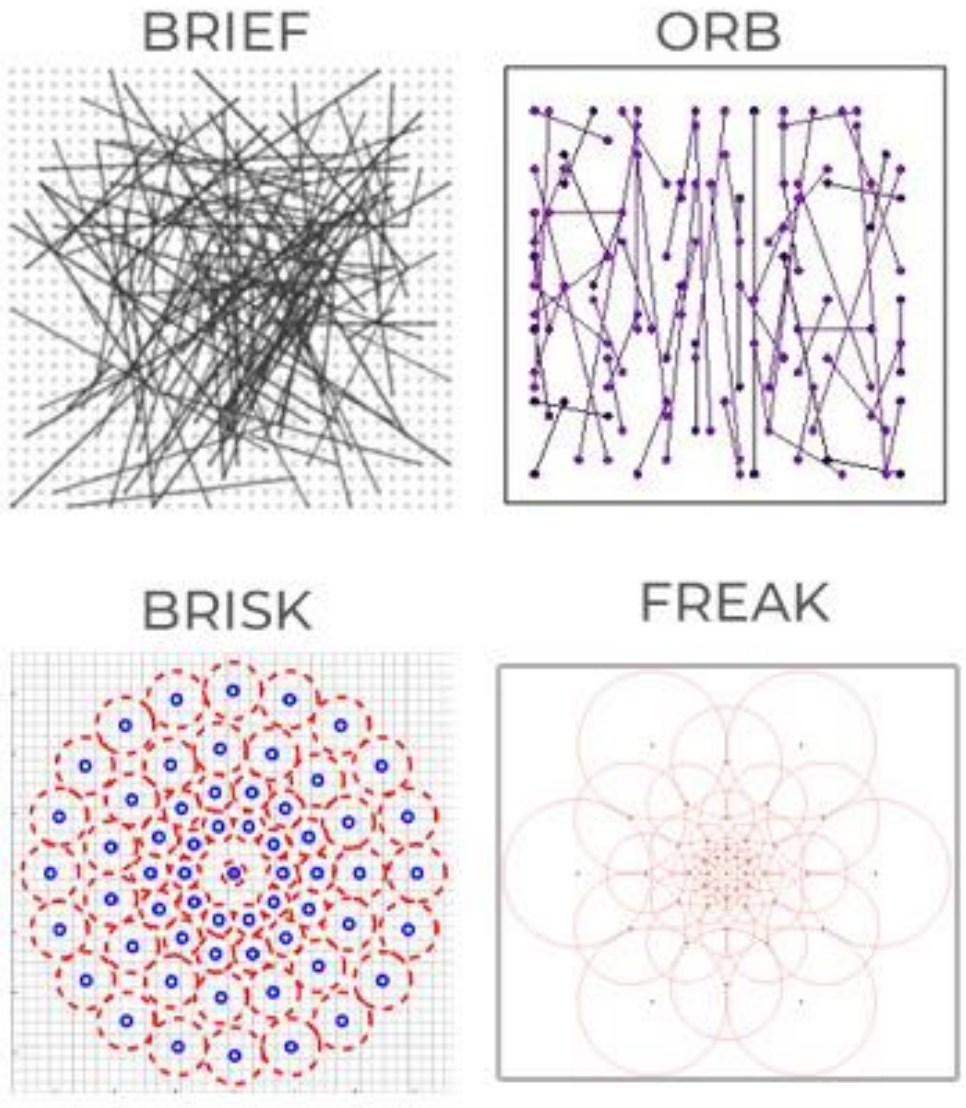

Рисунок 1.2 – Ілюстрація бінарних дескрипторів BRIEF, ORB, BRISK та FREAK [8, 9]

У роботі [10] дослідник порівняв різні комбінації популярних детекторів КТ та дескрипторів: SIFT з SIFT, SURF з SURF, MSER з SIFT,

BRISK з FREAK, BRISK з BRISK, ORB з ORB і FAST з BRIEF. Деякі назви повторюються, так як деякі методи містять у собі процедури виявлення множини КТ та обчислення значень дескрипторів. Дослідження проводились на предмет стійкості у випадку, коли зображення підлягає різноманітним перетворенням та трансформаціям: ефект від зміни масштабу, повороту об’єктів, розмиття зображення, підвищення рівня освітленості, стиснення JPEG та порівняння фото чистої стіни та з нанесеним графіті. Також, окремо проведені експерименти по розрахунку часу виявлення на кожну ключову точку.

Після аналізу отриманих результатів було виявлено, що комбінація ORB з ORB і MSER з SIFT може бути кращою майже в усіх можливих ситуаціях, коли розглядаються результати точності та відкликання. Крім того, швидкість FAST з BRIEF (обидва являються складовими ORB, детальніше у наступному підрозділі) перевершує інші. Тож, це є вагомим доводом для вибору дескрипторів ORB у якості опису зображень для подальших досліджень у рамках даної роботи.

1.2 Дескриптор ORB

ORB (Oriented FAST and Rotated BRIEF) розглядається як заміна SIFT і досягає кращої ефективності за допомогою двійкового опису [11]. Являється інваріантним, стійким до шумів та дешевим за обчислювальними ресурсами. Складається з двох кроків: вилучення множини КТ та побудови дескрипторів.

Для виявлення множини КТ алгоритм ORB використовує вдосконалений алгоритм *FAST* (*Features from accelerated segment test)*. Детектор FAST, запропонований Е. Ростеном і Т. Драммондом, є вдосконаленим детектором кутів Харріса шляхом бінарного обчислення та машинного навчання [12]. Ідея базується на тому, що якщо піксель значно

відрізняється від сусідніх, то, швидше за все, він буде кутовою точкою. Особливістю є його обчислювальна ефективність, що дає змогу використовувати його для обробки зображень та відео в режимі реального часу. Процес виявлення складається з 4 кроків [13].

Крок 1. *Визначення потенційних точок інтересу*. Спочатку обирається піксель p на зображенні та припускається, що його яскравість дорівнює I_p. Встановлюється поріг яскравості t. Потім розглядаються 16 пікселів, що лежать на колі з центром в p. Воно називається колом Брезенхема з радіусом 3 (рис. 1.3). Якщо яскравість послідовних N точок на вибраному колі більша за $I_p + t$ або менша за $I_p - t$, тоді піксель p можна вважати точкою інтересу. Ці дії повторюються для усіх пікселів зображення.

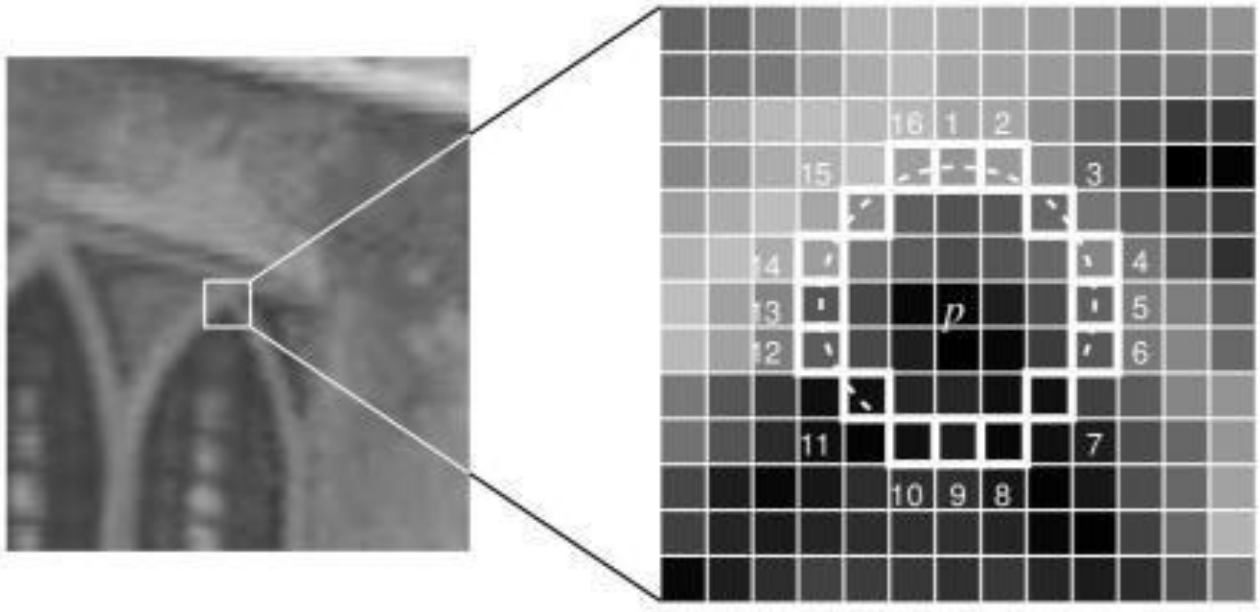

Рисунок 1.3 – Фрагмент зображення з потенційною КТ та з 16 точками, що розглядаються [14]

Крок 2. *Просіювання КТ*. Оскільки розрахунок кутової точки FAST працює лише з порівнянням різниці в яскравості між пікселями, це може давати значний відгук на прямі краї, що може призвести до зменшення інформативності виявлених точок. Через це, у алгоритмі ORB впроваджена додаткова процедура просіювання. Обчислюються відповіді Харріса для всіх

знайдених кутових точок FAST. Далі, множина сортується за цими значеннями та відбираються перші N точок. Формула розрахунку значення відгуку Харріса виглядає наступним чином [13]

$$R = \det(M) - k(trace(M))^2,\ M = \sum w(x,y)\begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}, \quad (1.1)$$

де R – значення відповіді Харріса;

M – матриця 2×2;

k коливається від 0,04 до 0,06 [15];

$w(x,y)$ – функція вікна зображення;

I_x – зміна точки в горизонтальному напрямку;

I_y – зміна характерної точки у вертикальному напрямку.

Крок 3. *Побудова піраміди зображень*. Піраміда зображень – це послідовність одного і того ж зображення, але з різним масштабом [11]. На кожному рівні піраміди застосовується алгоритм FAST. Отримані КТ зіставляються, тим самим набувши певної інваріантності масштабу. На рисунку 1.4 показаний приклад піраміди зображень.

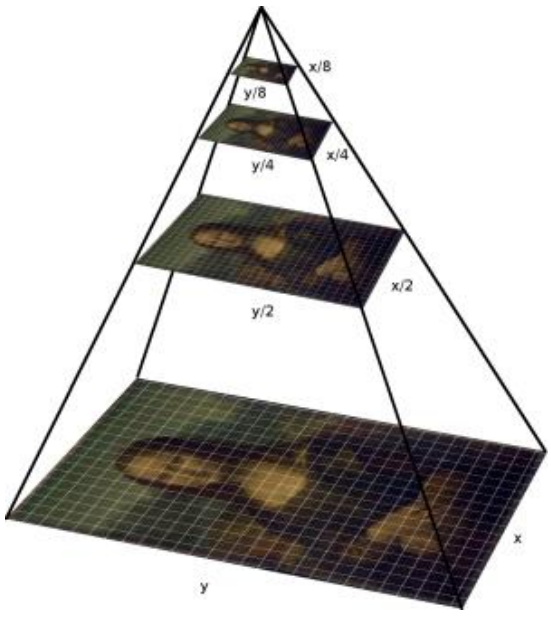

Рисунок 1.4 – Приклад піраміди зображень [13]

Крок 4. *Встановлення орієнтації КТ*. Для того, щоб витягнуті точки інтересу мали інваріантність до обертань, для кожної, за допомогою методу центроїда інтенсивності C, розраховується напрямок. Вважається, що C зміщений від ключової точки O, тож знайшовши вектор $\overrightarrow{OC}$, його можна використати для знаходження кута орієнтації. Формула координат центроїду наступна [13]

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right), \tag{1.2}$$

де m_{pq} розраховується як

$$m_{pq} = \sum_{x,y} x^p y^p I(x,y). \tag{1.3}$$

Тоді кут θ дорівнює

$$\theta = a\tan 2(m_{01}, m_{10}). \tag{1.4}$$

Завдяки наведеним вище крокам кутові точки FAST мають інваріантність масштабу та обертання, що значно покращує їх надійність на різних зображеннях.

Після виділення КТ алгоритм ORB використовує вдосконалений алгоритм *BRIEF (Binary Robust Independent Elementary Features)*, запропонований Майклом Калондером [16], спрямований на прискорення зіставлення та зменшення споживання пам’яті [17].

BRIEF – двійковий векторний дескриптор, вектор якого складається з чисел 0 і 1 [13]:

$$\tau(p;x,y)=\begin{cases}1, & p(x)<p(y),\\ 0, & p(x)\geq p(y),\end{cases} \tag{1.5}$$

де x,y – деякі області навколо випадково обраних пікселів;

$p(x), p(y)$ – середні значення яскравості відповідних областей.

Щоб зменшити вплив шуму, спочатку зображення піддається фільтрація Гауса. Базуючись на КТ p, взятій за центральну точку, розглядається вікно розміром $S\times S$ і випадковим чином обираються N пар пікселів у цьому вікні. Потім, значення яскравості кожної пари x,y порівнюється і виконується двійкове призначення (1.5). Зазвичай для ORB $S=31,\ N=256$, а розмір областей x,y дорівнює 5×5. Як результат, виходить N-вимірний вектор [13]

$$f_N(p)=\sum_{1\leq i\leq N}2^{i-1}\tau(p;x_i,y_i). \tag{1.6}$$

Знаходження дескриптору залишає велику кількість варіантів вибору пар $(x;y)$. Для вирішення цього завдання обирається один із п’яти способів відбору (геометрії вибірки) [16], проілюстрованих на рисунку 1.5. Вважається, що ключова точка знаходиться по центру області, що розглядається.

G I. Точки x та y обираються випадкового за рівномірним розподілом.

G II. Точки x та y обираються випадково за розподілом Гауса.

G III. Точки x та y обираються випадково у два етапи. Спочатку, за розподілом Гауса обирається x відносно координат центру. Далі обирається y відносно x.

G IV. Точки x та y обираються випадково за допомогою дискретної радіальної сітки.

G V. Точка *x* – це центр, а *y* випадково обрана за допомогою дискретної радіальної сітки.

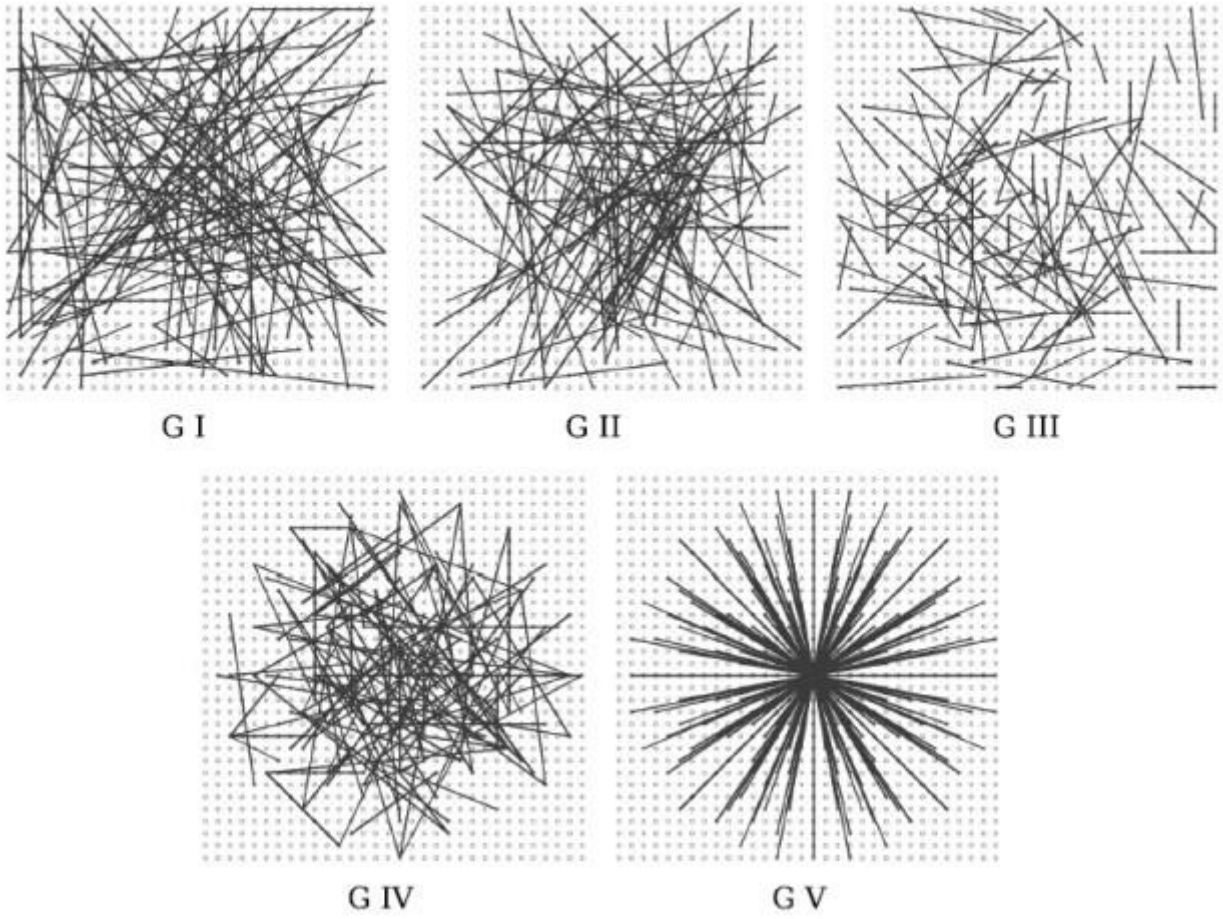

Рисунок 1.5 – П’ять способів вибору пар точок для формування дескриптора BRIEF [16]

У [16] дослідники провели експерименти з метою порівняння швидкості даних підходів. На рисунку 1.6 проілюстровано отримані результати.

З результатів дослідження видно, що патерн G V значно програє своїм суперникам, тож його використання не має сенсу. G II має невелику перевагу над іншими трьома. Саме цей спосіб використовується у алгоритмі ORB.

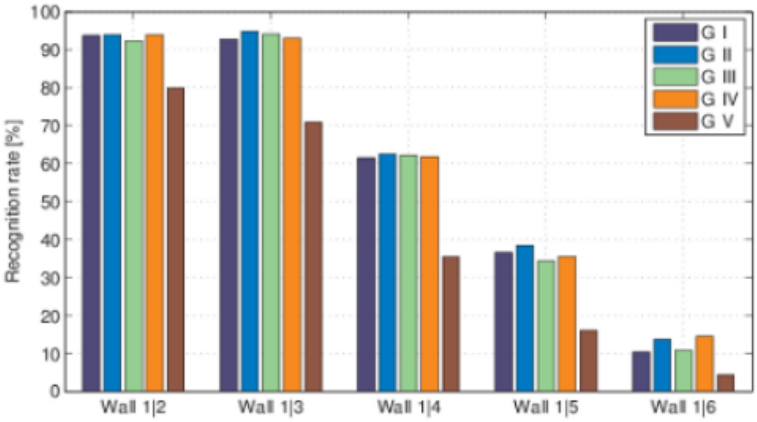

Рисунок 1.6 – Швидкість розпізнавання п’яти шаблонів вибору пар точок для формування дескриптора BRIEF [16]

Оскільки оригінальний дескриптор BRIEF не має інваріантності обертання, легко втратити дані, коли зображення обертається. Таким чином, алгоритм ORB використовує алгоритм «steered» BRIEF для знаходження основного напрямку кожної КТ. Встановлюється матриця повороту R_θ [13]:

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{1.7}$$

N пар пікселів формують матрицю Q:

$$Q = \begin{bmatrix} x_1, x_2, \ldots, x_N \\ y_1, y_2, \ldots, y_N \end{bmatrix}. \tag{1.8}$$

Далі, матриця корегується за допомогою R_θ:

$$Q_\theta = R_\theta Q. \tag{1.9}$$

Як результат, отримується дескриптор, що враховує напрямок:

$$g_N(p,\theta) = f_N(p) \mid (x_i, y_i) \in Q_\theta . \quad (1.10)$$

Підсумовуючи, на рисунку 1.7 зображена блок-схема алгоритму ORB.

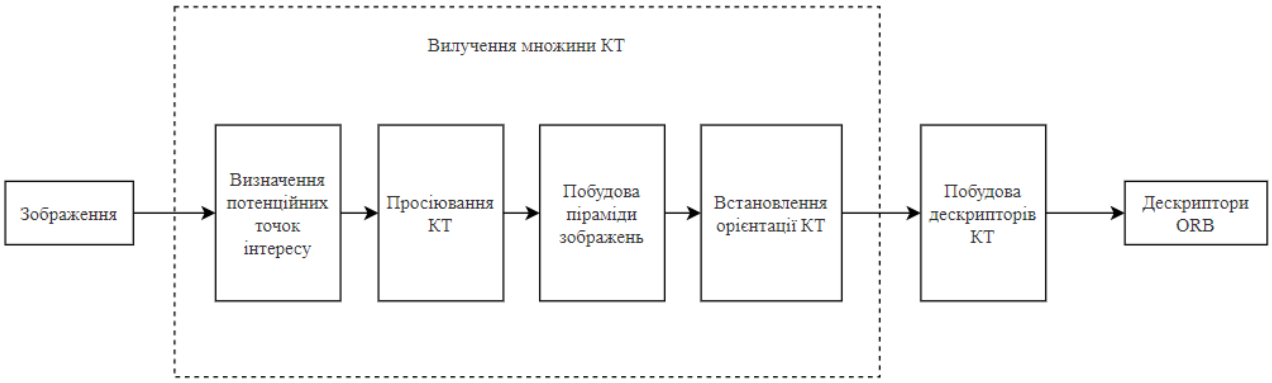

Рисунок 1.7 – Блок-схема алгоритму ORB

1.3 Постановка задачі дослідження

Дескриптори КТ формують простір ознак, що використовується у структурних підходах класифікації зображень з метою проведення аналізу на фундаментальному рівні. Більш поглиблений розбір візуальних об’єктів додає ваги побудованому класифікатору.

Об’єктом дослідження є методи класифікації зображень з використанням множин дескрипторів ключових точок у системах комп’ютерного зору.

Метою дослідження є впровадження технологій класифікації зображень на підставі багатокомпонентної моделі даних на множині бінарних дескрипторів для опису еталонних зображень.

Експерименти направлені на застосування цієї технології у структурних підходах прийняття рішень про клас об’єкту задля покращення класифікаційних властивостей.

Для досягнення мети необхідно вирішити такі завдання:

– провести аналіз та огляд теоретичного матеріалу щодо використання багатокомпонентних моделей даних для класифікації;

– підібрати зображення для тестової бази еталонів;

– впровадити та виконати дослідження способів побудови інформативних множин центрів класів;

– дослідити модифікацію існуючих алгоритмів класифікації зі структурним підходом для підтримки багатокомпонентних множин даних для описів еталонних зображень;

– вибрати технології та програмно реалізувати запропоновані методи;

– провести дослідження результатів класифікації та сформувати висновки щодо використання кількох центрів у порівнянні з одним.

2 МОДЕЛІ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ НА ОСНОВІ БАГАТОКОМПОНЕНТНИХ СТРУКТУР ОПИСІВ

2.1 Класифікація за статистичними розподілами компонентів

Розглянемо багатовимірний простір бінарних векторів B^n з розміром n [3, 4, 18–22]. Фактично, дана сукупність задає простір ознак для деякої бази еталонних зображень $E = \{E_1, E_2, ..., E_N\}$, кожен об'єкт якої представляє окремий клас у задачі класифікації. Описом кожного еталону є підмножина $E_i \subseteq B^n$, яка являє собою набір дескрипторів КТ [23–29]:

$$E_i = \{e_v(i)\}_{v=1}^{s}, \tag{2.1}$$

де i – індекс еталону;

$s = card\, E_i$ – загальна кількість побудованих дескрипторів.

Ця скінченна множина векторів формально є описом візуального об'єкта на зображенні.

Нехай M – фіксоване число центрів класів для кожного еталону, а $b_k(i)$ – k-й центр еталону E_i, де $k = \overline{1, M}$. Параметр M характеризує число компонентів у заданій моделі даних. Тоді, якщо кількість оброблюваних зображень N, то загальна кількість побудованих центрів дорівнюватиме $M * N$.

Для довільного дескриптора $e_v(j) \in E_j$ запровадимо процедуру формування матриці d_v, що представляє собою статистичний розподіл за множиною центрів $\{\{b_k(i)\}_{k=1}^{M}\}_{i=1}^{N}$:

$$d_v = \begin{bmatrix} d_{v,1}(1), d_{v,2}(1), ..., d_{v,M}(1), \\ d_{v,1}(2), d_{v,2}(2), ..., d_{v,M}(2), \\ ..., \\ d_{v,1}(N), d_{v,2}(N), ..., d_{v,M}(N) \end{bmatrix}, \quad (2.2)$$

$$\sum_{i=1}^{N} \sum_{k=1}^{M} d_{v,k}(i) = 1. \quad (2.3)$$

Значення $d_{v,k}(i)$ визначається шляхом знаходження деякої числової міри належності $\eta(e_v(j), b_k(i))$, що характеризує подібність певного дескриптору до центру класу та відтворенням умови (2.3), щоб відобразити значення у статистичній площині за допомогою відношення міри подібності до конкретної компоненти класу та суми значень до усіх компонентів оброблюваної бази загалом.

Дивлячись на результати досліджень, викладених у [3, 30], можна зробити висновок, що число M для кожного з класів має бути не великим та, на практиці, лежать на інтервалі від 2 до 5. Це обумовлено роботою введеної схеми розмиття інформації, тобто значень статистичного розподілу, між кожним компонентом даних. При їх значній кількості, суттєво збільшується вплив похибки обчислення, що призводить до не точності кінцевого результату класифікації.

Використовуючи (2.2), будується сукупність розміром s з матриць d_v, яка містить повну інформацію у заданому просторі про належність кожного дескриптора до відповідних компонентів даних. У результаті будується класифікатор $K \in [1, 2, ..., N]$, що відповідає еталону до якого віднесене досліджуване зображення. Варто зазначити, що отримані результати цілком залежать від початкових даних, тобто заданої бази описів еталонів.

На рисунку 2.1 зображена концептуальна схема даної моделі класифікації.

Рисунок 2.1 – Концептуальна схема моделі класифікації

Головною метою даного дослідження є підвищення результативності класифікації на підґрунті побудови ансамблю статистичних розподілів, шляхом використання багатокомпонентної моделі даних для тестової бази візуальних об’єктів.

Основна ідея полягає у програмному дослідженні декількох моделей класифікації з використанням різноманітних способів побудови багатокомпонентної моделі даних через виділення найбільш інформативних центрів класів. Розглядаються різні методи в основі яких лежить робота з апаратами теорії множин та векторного простору, метричними моделями визначення релевантності множин та механізмом кластеризації.

Розглянуті моделі класифікації базуються на ансамблях статистичних розподілів компонентів структурного опису візуальних об’єктів. Тож, початкові дії не будуть відрізнятися в залежності від обраного алгоритму.

Перед початком класифікації необхідно певним чином трансформувати кожний опис еталонного об’єкту (2.1), щоб отримати набір центрів у вигляді векторів

$$b_k(i) = (b_{k,1}(i), b_{k,2}(i), ..., b_{k,n}(i)), \tag{2.4}$$

які обчислюються підставі множини дескрипторів E_i та відносно яких у подальшому будуватимуться розподіли. Отримання $\{b_k(i)\}$ є одним із

найголовніших складових класифікації. Позитивним моментом є те, що знаходження центрів відбувається в рамках попередньої обробки бази еталонних зображень, тобто не впливає на час та продуктивність самого процесу класифікації.

Своєю чергою, вектор $e_v(j) \in E_j$, як і будь-який дескриптор, можна представити як

$$e_v(j) = (e_{v,1}(j), e_{v,2}(j), ..., e_{v,n}(j)). \tag{2.5}$$

Наступним кроком, вхідні дані для аналізу, тобто множину дескрипторів, необхідно представити у вигляді деякої функції належності μ до заданих центрів класів, значення якої лежать на інтервалі від 0 до 1:

$$\mu : B^n \to [0,1],$$
$$\mu(e_v(j), b_k(i)) \in [0,1]. \tag{2.6}$$

Параметрами функції μ є дескриптор $e_v(j)$ та центр класу $b_k(i)$. Саму функцію належності μ можна визначити за допомогою основоположної характеристики у data science [31] – співвідношення значень мір, що задають число сприятливих випадків та загального числа випадків $M * N$, що відповідає загальній кількості центрів класів

$$\mu(e_v(j), b_k(i)) = \frac{\eta(e_v(j), b_k(i))}{\sum_{l=1}^{M*N} \eta(e_v(j), b_k(l))}, \tag{2.7}$$

де $\eta(e_v(j), b_k(i))$ – числове значення міри подібності для відповідного дескриптору та центру класу.

Вважаємо, що $b_k(i)$ – бінарний вектор розмірністю n. За даної умови, міру подібності η можна знайти за допомогою *Хемінгової метрики*. Хемінгова метрика, або *відстань Хемінга* – число позицій, в яких відповідні символи двох двійкових слів однакової довжини рівні. На рисунку 2.2 подана графічна інтерпретація даної відстані: 010 – 111 (результат 2) та 100 – 011 (результат 3).

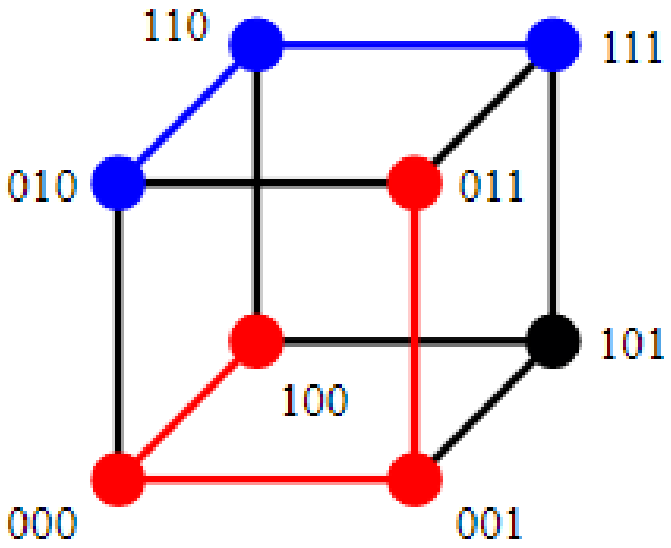

Рисунок 2.2 – Графічне подання Хемінгової метрики

Дана метрика якнайкраще підходить для визначення розрізнення для даного бінарного простору B^n. Потрібно лише знайти суму результатів операції XOR між двома двійковими дескрипторами, щоб мати можливість порівнювати їх [10].

Оскільки розмірність кожного вектору дорівнює n, то значення відстані лежатиме на інтервалі від 0 до n. Тоді міру подібності η можна розрахувати за формулою

$$\eta(e_v(j), b_k(i)) = n - \chi(e_v(j), b_k(i)). \qquad (2.8)$$

Знайшовши усі значення η, для дескриптора $e_v(j)$ та скориставшись виразом (2.7), можна визначити матрицю d_v. Повторивши це для всіх s дескрипторів, отримуємо повну інформацію, в рамках даного простору ознак, про статистичний розподіл компонентів даних.

Вищеописаний підхід дозволяє проводити додаткову обробку на етапах знаходження міри подібності η чи функції належності μ, тим самим, фільтруючи отримані дані для підвищення інформативності, або ж, стискаючи обсяг аналізованої інформації, при цьому зменшуючи час обчислення. Та варто зазначити, що точність класифікації цілком залежить від вхідних даних, тож виконуючи подібні процедури, є ймовірність значного впливу на кінцевий результат.

Далі розглянемо способи побудови класифікатору. Одна з найпростіших схем класифікації полягає у здійснені інтеграції розподілів в рамках кожного з класів та у подальшому виявленні класифікатору K шляхом знаходження максимального значення:

$$K : r = \arg\max_i \sum_{v=1}^{s} \sum_{k=1}^{M} d_{v,k}(i). \qquad (2.9)$$

Фактично це відтворює схему побудови класифікатору за інтегральним поданням розподілів з одним центром, результативність якого доведена дослідниками у [31, 32]. На рисунку 2.3 зображена схема даної моделі класифікації.

Базуючись на даному методі, можливо провести додатковий, більш глибокий аналіз, що впроваджується моделями опрацювання за множиною компонентів всередині класів як для окремих дескрипторів опису, так і за їх повною сукупністю. Варіантами оброблення є інтегрування розподілів в межах компонент окремих еталонів з прийняттям класифікаційного рішення за максимумом серед критеріїв для інтегрованих компонент, сумою критеріїв

для компонент окремих класів або шляхом логічного аналізу отриманих інтегрованих критеріїв.

Рисунок 2.3 – Схема моделі класифікації за інтеграцією розподілів класів

Інші варіанти аналізу розподілів полягають у реалізації поелементної класифікації окремо для кожного із дескрипторів об’єкта із подальшим узагальненням отриманих значень класів за моделлю бустінгу [33].

Друга схема побудови класифікатору, також, зводиться до інтегрування значень розподілів, але в рамках набору компонент. За заданою сукупністю матриць d_v, для $\forall e_v(j) \in E_j$ обчислюється значення критерію $\theta_{i,k}$:

$$\theta_{i,k} = \sum_{v=1}^{s} d_{v,k}(i). \qquad (2.10)$$

На основі цього, вводяться наступні варіанти прийняття класифікаційних рішень:

$$K : r = \arg \max_{i,k} \theta_{i,k}, \qquad (2.11)$$

$$K : r = \arg\max_i \sum_{k=1}^{M} \theta_{i,k},$$

$$K : r = \arg\max_i L(\{\theta_{i,k}\}, k \in E_i),$$

де L – деяка процедура на підмножині значень;

$k \in E_i$ – позначка, що означає деякий аналіз компонент даних класу E_i.

Прикладом L може бути визначення класу шляхом знаходження найбільшої чи складання двох найбільших компонентів класу, або ж введення інших додаткових перевірок та лімітованих значень на знаходження найбільш вагомих екстремумів, тим самим зменшуючи рівень помилок класифікації.

На рисунку 2.4 зображена схема моделі класифікації за інтегрування розподілів в рамках компонент.

Рисунок 2.4 – Схема моделі класифікації за інтеграцією розподілів компонент класів

Третя модель побудови класифікатору полягає у формуванні значення класу для кожного з дескрипторів окремо. Кінцевий результат визначається шляхом знаходження найбільшої суми голосів за відповідний опис. Цей

метод кардинально відрізняється від попередніх, так як кожний дескриптор самостійно обирає клас до якого він буде віднесений. Це дозволяє зменшити вплив помилкових дескрипторів, таким чином, зменшивши похибку класифікації.

Варіанти прийняття класифікаційних рішень для $\forall e_v(j) \in E_j$ мають вид:

$$r_v = \arg \max_{i,k} d_{v,k}(i),$$

$$r_v = \arg \max_{i} \sum_{k=1}^{M} d_{v,k}(i), \tag{2.12}$$

$$r_v = \arg \max_{i} L(\{d_{v,k}(i)\}, k \in E_i).$$

Отримані у результаті (2.12) значення складаються $\sum_i = \sum_i + 1$. Класифікатор відповідає класу, який набрав найбільшу кількість голосів:

$$K : r = \arg \max_{i} \sum_i . \tag{2.13}$$

Для оцінювання ефективності класифікації введене значення Δ_i, що дорівнює різниці між двома максимальними значеннями критерію класифікації $\theta_{i,k}$:

$$\Delta_i = (\theta_{i,\max_1} - \theta_{i,\max_2}) / \theta_{i,\max_1} . \tag{2.14}$$

На рисунку 2.5 зображена схема моделі класифікації на основі підрахунку голосів.

Рисунок 2.5 – Схема моделі класифікації за підрахунком голосів

2.2 Побудова багатокомпонентних моделей даних

У структурних методах класифікації зображень опис візуального об’єкту, що представлений у вигляді множини багатовимірних векторів даних, зіставляється з множинами векторів, що відображають інформацію про базу еталонів, тобто є центрами класів. Їх побудова, є важливим підготовчим етапом алгоритму, так як від них значною мірою залежить результат, тож важливо, щоб знайдені центри мали високі показники класифікаційних властивостей [34–43].

У роботі [31, 32] дослідники довели, що використання моделі класифікації (2.9) з одним *медоїдом* у якості центру показує певні результати та має перспективи у подальшому дослідженні даного механізму класифікації.

У кластерному аналізі, медоїд – об'єкт, що належить набору даних або кластеру, відмінність (наприклад, за координатами) якого від інших об'єктів у наборі даних або кластері мінімальна. Медоїд близький за змістом до центроїда, але, на відміну від нього, є об'єктом, що належить кластеру, і, як

правило, використовується в тих випадках, коли неможливо обчислити середні координати або центр мас кластера.

На рисунку 2.6 показана відмінність між середнім та медоїдом у двовимірному просторі.

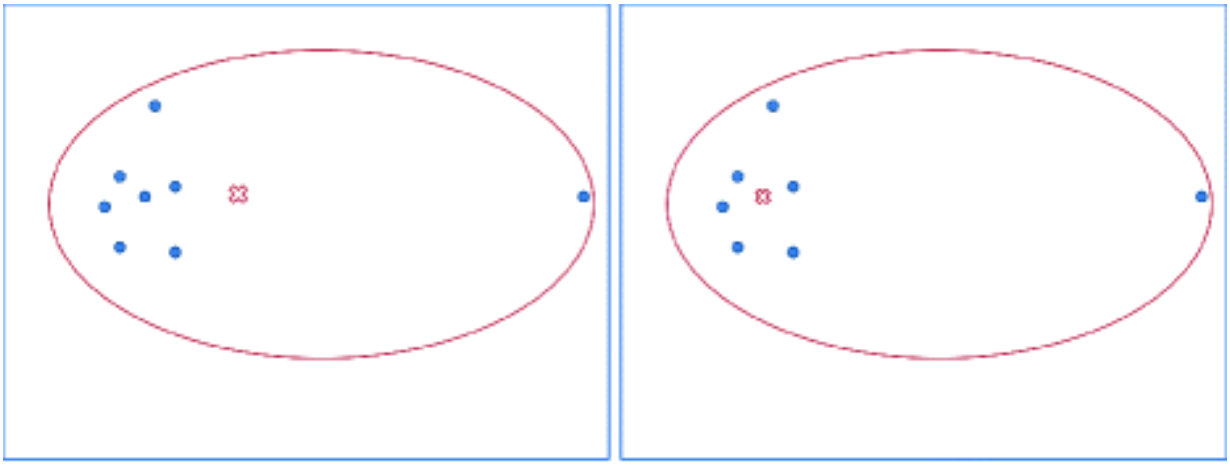

Рисунок 2.6 – Середнє і медоїд у двовимірному просторі [44]

На обох малюнках група точок ліворуч утворює скупчення, а крайня права точка є сторонньою. Червона точка на зображеннях – це центр, знайдений обома способами. Видно, що середнє значення має більший вплив помилкових або ж відсторонених даних, на відміну від медоїда, який є стійким до подібних значень та більш впевнено представляє центр.

У рамках нашого простору ознак, де кожний елемент – це бінарний вектор однакового розміру, медоїд m_i знаходиться шляхом розрахунку для кожного дескриптору суми відстаней Хеммінга до інших та виявлення мінімального значення, що відповідатиме медоїду:

$$D(\upsilon) = \sum_{v=1}^{s} \chi(e_v(i), \upsilon),$$

$$m_i = \arg \min_{\upsilon \in E_i} D(\upsilon). \tag{2.15}$$

Чим менша сума значень $D(\upsilon)$, тим менше відрізняється число елементів зліва та справа від значення m_i за відстанню до нього. Спираючись на це твердження, можна ввести процедуру сортування за зростанням всієї множини дескрипторів за сумарною відстанню до інших. Таким чином, на першій позиції буде медоїд, а на останній найвіддаленіший дескриптор від інших. Використовуючи ці данні, з'являється можливість виділення множини центрів спираючись на деякі логічні припущення. Прикладом, може бути множина центрів, що складається з медоїду та наступних двох, чи медоїду та найвіддаленішого.

Для використання повного контексту даних, можна вдатися до деяких перетворень, що охоплюють кожен дескриптор опису еталону. До такої процедури можна віднести знаходження середнього бінарного вектору $\overline{a}_i$, що представляє собою середнє значення для дескрипторів деякої підмножини $A_i \subseteq E_i \subseteq B^n$. Значення відповідних позицій двійкових слів складаються та діляться на загальний розмір множини A_i. Далі, працює правило, *якщо більше 0,5 то 1, якщо менше або дорівнює 0,5 то 0*, будується вектор $\overline{a}_i$. На рисунку 2.7 зображений приклад знаходження середнього вектору.

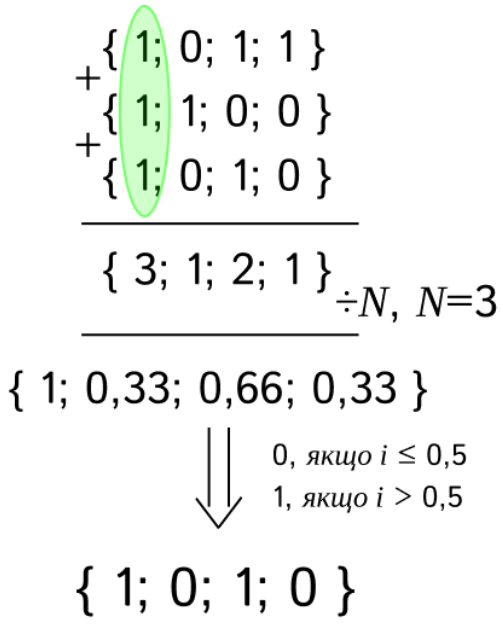

Рисунок 2.7 – Приклад знаходження середнього бінарного вектору

Таким чином, відсортовану за значеннями $D(\upsilon)$ множину дескрипторів E_i можна розбити на підмножини $A_{i,k}$ та розрахувати для них середні вектори, які являтимуть собою набір центрів класів. Це дасть змогу більш поглиблено дослідити весь набір дескрипторів, урахувавши кожен з них для побудови ансамблю опису структури еталону.

Способи виділення підмножин A_i також можуть бути різними. Найпростішим прикладом може бути розбиття всієї множини дескрипторів E_i на рівні за розміром підмножини.

Інший спосіб формування центрів може базуватись на обчисленні значення інформативності для кожного дескриптора еталонного зображення. Введення метричних критеріїв даних має суттєвий вплив у задачах класифікації. Це дає змогу зменшити загальний об'єм обчислень, а як результат оптимізувати час виконання алгоритму та, залежно від методів, підвищити результативність отриманого класифікатору, опрацьовуючи найбільш інформативні елементи множини. У рамках даного дослідження, застосовується варіант розрахунку, що запропоновано у [19].

Для $\forall e_v(j) \in E_j$ вводиться поняття інформативності $V(e_v(j), E)$ у складі бази еталонів E:

$$V(e_v(j), E) = \chi_{\min}(e_v(j), \bar{E}_j) - \chi_{\min}(e_v(j), E_j),$$

$$\chi_{\min}(e_v(j), \bar{E}_j) = \min_{k, i \neq j} \chi(e_v(j), e_k(i)), \tag{2.16}$$

$$\chi_{\min}(e_v(j), E_j) = \min_{k \neq v, i = j} \chi(e_v(j), e_k(i)).$$

Іншими словами, $V(e_v(j), E)$ дорівнює різниці між мінімальною відстанню від $e_v(j)$ до елементу бази E, що не належить до E_j та

мінімальною відстанню від $e_v(j)$ до елементу, що належить E_j, окрім самого себе. Це проілюстровано на рисунку 2.8.

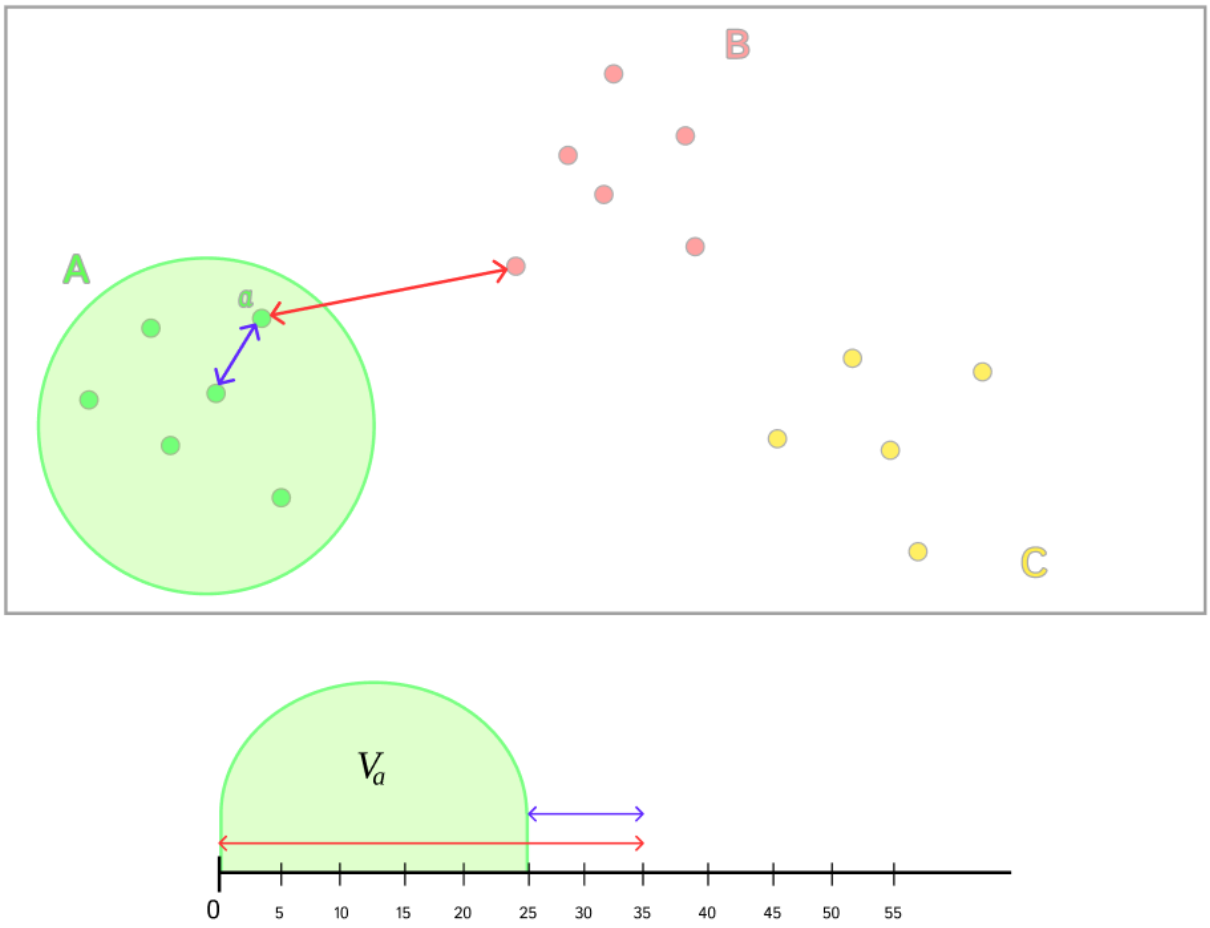

Рисунок 2.8 – Графічне подання характеристики інформативності (2.17)

Дана характеристика побудована на принципі «ближче до своїх, подалі від інших». Чим далі знаходиться найближчий елемент іншого класу, тим кращою вважається індивідуальність даного елементу.

3 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ РЕЗУЛЬТАТИВНОСТІ БАГАТОКОМПОНЕНТНИХ МОДЕЛЕЙ ДАНИХ

3.1 Вибір програмних технологій

Завданням даного дослідження є проведення експериментів з використанням багатокомпонентних моделей даних, наведених у розділі 2, у рамках задачі класифікації зображень. Для цього була обрана мова програмування C#, яка містить великий об’єм потужного інструментарію для розробки різнотипних застосунків. Для роботи з зображеннями була використана графічна бібліотека Emgu CV, що містить у собі прості засоби обробки медіа-файлів.

Мова програмування C# – це об’єктно-орієнтована мова програмування, а також мова кодування загального призначення, яку фахівці з обробки даних використовують для написання програм. Ця мова програмування походить від компанії Microsoft. Усі застосунки, написані фахівцями з обробки даних, написані в середовищі виконання .NET.

Мова програмування C# має багато переваг, зокрема:

– *C# – це об’єктно-орієнтована мова програмування*, яка дає змогу розробляти модульний, керований і багаторазово використовуваний код. Професіонали в області даних можуть вказати тип структурованих даних, які можна використовувати для застосування функцій до об’єктів. Це допомагає розбивати застосунки, які легко тестувати та читати;

– *збирач сміття*. C# забезпечує дуже ефективну систему для стирання та видалення всього сміття в системі. C# не спричиняє безлад у системі та не викликає її зависання під час виконання;

– *мова високого рівня, яка пропонує можливості доступу до пам'яті*. Оскільки мова C# схожа на людську мову, її класифікують як мову високого рівня. Іншими словами, вона далека від машинного коду, тому ми повинні

скомпілювати код C#, щоб апаратне забезпечення інтерпретувало його команди;

– C# є однією з найбільш часто використовуваних мов програмування у світі, а це означає, що є багато розробників C#, готових допомогти вам. Оскільки C# є продуктом Microsoft, то користувачі отримують переваги від підтримки компанії, яка включає професійну допомогу, додаткові ресурси та регулярні оновлення.

Вважається, що мова програмування C# ідеально підходить для наукових проектів. Ця мова кодування полегшує завдання вчених по створенню data science ініціатив. C# відома статичним кодуванням, що означає, що вона компілює кроки та має послідовний синтаксис.

Для кращої розробки проектів спеціалісти з даних можуть використовувати .NET для запису та повторного використання величезних обсягів даних у реальному часі. Її блискавична продуктивність і широке поширення роблять цю мову програмування ідеальною для обробки великих даних. Експерти з даних можуть використовувати цю мову кодування для наукових ініціатив на рівні виробництва.

Для простої роботи з масивами даних використаний LINQ. LINQ, що розшифровується як Language Integrated Query (вимовляється як «link»), – це розширення мови .NET, яке підтримує пошук даних із різних джерел даних, таких як XML-документ, бази даних і колекції. Воно було представлене у платформі .NET 3.5. Можливості LINQ підтримують такі мови, як VB, C# тощо. Існує різноманітність видів LINQ. Основними є [45]:

– *LINQ to objects* – дозволяє запитувати об’єкти в пам’яті, такі як масиви, списки та інші типи колекцій;

– *LINQ to XML* – дозволяє надсилати запити до XML-документа, перетворюючи документ на об’єкти XElement, а потім надсилаючи запити за допомогою локального механізму виконання;

– *LINQ to ADO.NET* – включає:

1) *LINQ to SQL* – використовується спеціально для роботи з базою даних SQL-сервера;

2) *LINQ to dataset* – дозволяє надсилати запити до будь-якої бази даних, до якої можна надсилати запити за допомогою ADO.NET;

3) *LINQ to entities* – це схоже на LINQ to SQL. Дозволяє розробникам запитувати модель даних концептуальної сутності;

– *Parallel LINQ* – це розширення LINQ для об’єктів, які мають бібліотеку паралельного програмування. Використовуючи це, можна розділити запит для одночасного виконання на різних процесорах.

LINQ приносить розробнику суттєву користь:

– не потрібно вивчати нову мову, оскільки вона схожа на SQL. LINQ простий і акуратний;

– LINQ є строго типізованим, тому помилки запиту перевіряються під час компіляції, а не старим способом пошуку помилок у запиті під час виконання. Коротше кажучи, це пришвидшує процес розробки;

– зменшує обсяг коду, який потрібно написати, що забезпечує кращу продуктивність і дає змогу приділяти більше уваги іншим технологіям;

– є підтримка IntelliSense для змінної діапазону під час введення решти операторів LINQ;

– у результаті запиту LINQ, можна використовувати вихідну послідовність як вхідні дані та змінювати її різними способами для створення нової вихідної послідовності (перетворення даних). Також, можна змінити саму послідовність, не змінюючи самі елементи шляхом сортування та групування.

Для роботи із зображеннями та побудови множини дескрипторів була використана Emgu CV. Emgu CV – кроссплатформна обгортка .NET з відкритим кодом для бібліотеки обробки зображень OpenCV [46]. Дозволяє викликати функції OpenCV із сумісних .NET мов, таких як C#, VB, VC++, IronPython тощо. Обгортку можна скомпілювати в Mono та запускати на пристроях Windows, Linux, Mac OS X, iPhone, iPad та Android.

Інструментарій даного продукту дозволяє вирішувати різноманітні завдання з 2D графікою, розпізнаванням об’єктів на фото чи відео та інші.

Також, використання Emgu CV дає можливість показувати ключові точки на вхідних зображеннях, бо бібліотека дозволяє з легкістю використовувати, обробляти та зберігати їх, підтримуючи різні типи (формати) статичних медіафайлів.

Для реалізації застосунку було використане програмне середовище Visual Studio 2022. Visual Studio – це провідне середовище розробки Microsoft для Windows і macOS. За допомогою Visual Studio можна розробляти, аналізувати, налагоджувати, тестувати, співпрацювати та розгортати своє програмне забезпечення. Visual Studio містить редактор коду, який підтримує IntelliSense (компонент завершення коду), а також рефакторинг коду. Вбудований налагоджувач працює як налагоджувач на рівні джерела, так і на рівні машини. Інші вбудовані інструменти включають профайлер коду, конструктор для створення застосунків графічного інтерфейсу користувача, вебдизайнер, конструктор класів і конструктор схем бази даних [47].

3.2 Опис бази еталонних даних

Для аналізу у якості бази еталонних зображень використовуються портрети українських письменників у форматі JPG та розміром 1024×1024 пікселів. Візуальний об’єкт знаходиться в центрі на світлому фоні. На рисунку 3.1 наведені використані зображення з прикладом виділених ключових точок.

З метою спрощення аналізу кожному еталону присвоєно порядковий номер:

– 1 – портрет Івана Карпенка-Карого;

– 2 – портрет Михайла Коцюбинського;

– 3 – портрет Миколи Хвильового.

а) б) в) г)

Рисунок 3.1 – База еталонних зображень:
а), б), в) – вхідні зображення; г) – виділені ORB дескриптори

Щоб дослідити правильність класифікації, тестова вибірка також складається з еталонних зображень. Класифікація вважається успішною, якщо кожне еталонне зображення було віднесене вірно.

Для кожного еталона виділено 500 бінарних дескрипторів ORB розміром 256 біт. Задля проведення певного порівняльного аналізу побудованих множини, знайдені значення метрики Хаусдорфа (типу множина – множина):

– для 1-го та 2-го еталонів – 89 (34,8% від максимуму 256);

– для 1-го та 3-го – 89 (34,8%);

– для 2-го та 3-го – 90 (35,2%).

Як бачимо, за критерієм відстані Хаусдорфа еталони у достатній мірі відрізняються один від одного. Для порівняння обчислено відстань Хемінга між медоїдами даних множин (у таблиці 3.1 наведене бітове подання медоїдів кожного еталонного зображення):

– для 1-го та 2-го еталонів – 66 (25,8% від максимуму 256);

– для 1-го та 3-го – 71 (27,7%);

– для 2-го та 3-го – 65 (25,4%).

Видно, що відстань Хаусдорфа краще розділяє дані множини у порівнянні з відстанню між медоїдами.

Таблиця 3.1 – Бітове подання медоїдів

№ еталону	Бітовий рядок
1	0101000000001101011111001011001111111110111010110001110111000010111100001011011001101011111001000011010101111001011001000010010001110111011101111111111000000101100001110111011000111010101011111010101111111001011001110000111110100000110000101101001111101
2	0100000010111011110101100111100111101110101010111111111011010011110111001101100100011101011100101001111000111101101111100000001100111111111011101111010011000001101100011110110010110110011111111111011010111100010110011101000110101000011110001010110111111111
3	1101000111111001111011111111101111111111111111111111111111111111111111000111100110111111111110111010010011111111111111011110100111111111000100111101111101111111011100111111111111110111111101111101001111101001001110111011010111111111

З метою проведення порівняльного аналізу, зображення класифіковані з використанням єдиного центру класу – медоїду. Використано дві моделі прийняття класифікаційних рішень: за інтегрованими статистичними розподілами (2.9) та за підрахунком голосів окремо для кожного дескриптора (2.13). До таблиць 3.2 та 3.3 занесені результати класифікації.

Бачимо ефективну класифікацію обраних зображень у обох варіантах. Максимум критерію відповідає «своєму» класу, тобто знаходиться на

діагоналі. Зауважимо, що модель (2.13) показує кращі значення Δ, тобто класифікація більш впевнена.

Таблиця 3.2 – Результат класифікації для моделі (2.9) з одним центром (медоїдом)

№ вхідного зображення	Критерій класифікації θ			Δ,%
	1	2	3	
1	167,479	167,166	165,355	0,19
2	164,338	167,87	167,792	0,05
3	164,276	165,815	169,909	2,41

Таблиця 3.3 – Результат класифікації для моделі (2.13) з одним центром (медоїдом)

№ вхідного зображення	Критерій класифікації θ			Δ,%
	1	2	3	
1	204	174	122	14,71
2	140	184	176	4,35
3	113	176	211	16,59

Було прийнято рішення, що розмір досліджуваних множин центрів класу дорівнюватиме трьом, так як, це найбільш оптимальне число. У разі його збільшення, значення статистичних розподілів буде значно зменшуватися та близитись до 0, що у свою чергу ускладнює обчислення.

3.3 Результати дослідження способів для вибору центрів

Розглянемо метод побудови ансамблю центрів класів, що базується на сортуванні множини дескрипторів за значеннями їх сумарної відстані до

інших (2.15). Як приклад, у таблиці 3.4 показана частина відсортованих дескрипторів.

Таблиця 3.4 – Відсортовані за відстанню дескриптори еталону 1

№	№ дескриптора	Сумарна відстань до інших
1	105 (медоїд)	53480
2	463	53508
3	474	53556
4	177	53818
5	428	53894
…	…	…
496	315	70898
497	57	72234
498	214	72484
499	45	73154
500	137	73790

Даний спосіб дозволяє експериментувати з варіантами множин центрів. Можна обирати дескриптори, які знаходяться ближче до медоїда, далі від нього, або впровадити додаткову логічну процедуру для виділення вдалих наборів.

На початку було перевірено найпростіший варіант відбору – дескриптори із зумовленими номерами: (1, 2, 3); (1, 25, 50); (1, 125, 250); (125, 250, 375); (1, 250, 500).

Використана модель класифікації (2.11), яка полягає у знаходженні максимального значення суми статистичних розподілів за цілими класами. У таблиці 3.5 показано принцип роботи даного алгоритму, де кольором відмічені значення статистичних розподілів та результуюча сума для рішення.

Таблиця 3.5 – Принцип роботи моделі (2.11) з інтеграцією у рамках класу

№	Критерій класифікації θ								
	1			2			3		
1	0,114	0,119	0,099	0,097	0,1	0,101	0,119	0,124	0,127
2	0,1	0,113	0,112	0,115	0,114	0,113	0,11	0,112	0,111
3	0,098	0,109	0,102	0,124	0,118	0,112	0,116	0,11	0,11
4	0,124	0,113	0,105	0,113	0,11	0,114	0,11	0,107	0,104
5	0,125	0,11	0,113	0,114	0,107	0,107	0,108	0,105	0,111
…	…			…			…		
∑	167,693			167,535			164,772		

До таблиці 3.6 занесено результати класифікації за даною моделлю для визначених наборів центрів, де кольором позначені максимальні суми розподілів за якими приймається рішення.

Таблиця 3.6 – Результат класифікації моделі (2.11) з інтеграцією у рамках класу з трьома центрами

Варіант центрів	№ вхідного зображення	Критерій класифікації θ			Δ,%
		1	2	3	
1	2	3	4	5	6
(1, 2, 3)	1	167,693	167,535	164,772	0,09
	2	165,07	168,066	166,863	0,72
	3	164,249	166,379	169,372	1,77
(1, 25, 50)	1	167,653	165,952	166,395	0,75
	2	166,347	166,111	167,542	0,71
	3	166,161	164,552	169,287	1,85
(1, 125, 250)	1	167,791	167,409	164,801	0,23

Продовження таблиці 3.6

1	2	3	4	5	6
	2	165,717	168,686	165,596	1,76
	3	164,569	167,832	167,599	0,14
(125, 250, 375)	1	167,881	168,028	164,091	0,09
	2	165,449	168,787	165,763	1,79
	3	164,122	168,658	167,22	0,85
(1, 250, 500)	1	166,775	168,175	165,05	0,83
	2	164,433	168,447	167,09	0,82
	3	163,233	167,852	168,915	0,63

Видно, що тільки один набір центрів впорався з класифікацією зображень – це множина, що складається з медоїду та найближчих до нього дескрипторів: (1, 2, 3). Якщо порівнювати ефективність класифікації Δ, а саме (0,19%, 0,05%, 2,41%) для єдиного центру та (0,09%, 0,72%, 1,77%) для трьох, то можна помітити, що суттєвого покращення немає. Тільки 2-й еталон виділяється більше впевнено у порівнянні з попередніми результатами, але при цьому результативність інших падає.

Модель (2.11) може мати інші варіації. Наприклад, коли класифікаційне рішення прийняте шляхом знаходження максимального інтегрованого значення розподілів для кожного центру окремо. Таким чином, є можливість нехтувати найменш інформативними центрами відносно інших у наборі, але, при цьому, з’являється небезпека при наявності аномально схожого компонента з іншого класу.

Аналогічно попередній моделі, у таблиці 3.7 показаний принцип роботи даного алгоритму, а у таблиці 3.8 показано результати класифікації для різних множин центрів.

Таблиця 3.7 – Принцип роботи моделі (2.11) з інтеграцією у рамках кожної компоненти окремо

№	Критерій класифікації θ								
	1			2			3		
1	0,114	0,119	0,099	0,097	0,1	0,101	0,119	0,124	0,127
2	0,1	0,113	0,112	0,115	0,114	0,113	0,11	0,112	0,111
3	0,098	0,109	0,102	0,124	0,118	0,112	0,116	0,11	0,11
4	0,124	0,113	0,105	0,113	0,11	0,114	0,11	0,107	0,104
5	0,125	0,11	0,113	0,114	0,107	0,107	0,108	0,105	0,111
…	…	…	…	…	…	…	…	…	…
∑	56,045	55,766	55,883	55,859	55,385	56,291	55,27	54,74	54,762

Таблиця 3.8 – Результат класифікації моделі (2.11) з інтеграцією у рамках кожної компоненти окремо з трьома центрами

Варіант центрів	№ зображення	Критерій класифікації θ									Δ,%
		1			2			3			
1	2	3	4	5	6	7	8	9	10	11	12
(1, 2, 3)	1	56,045	55,766	55,883	55,859	55,385	56,291	55,27	54,74	55,762	0,44
	2	54,935	55,197	54,939	56,045	56,066	55,955	56,039	55,418	55,407	0,04
	3	54,875	54,712	54,662	55,339	55,53	55,51	56,745	56,457	56,169	0,51
(1, 25, 50)	1	56,86	55,421	55,372	56,467	54,958	54,527	55,942	54,149	56,304	0,69
	2	56,01	54,326	56,01	56,9	54,946	54,265	56,977	55,333	55,231	0,13
	3	56,368	53,774	56,019	56,644	54,027	53,881	58,19	56,442	54,655	2,66
(1, 125, 250)	1	58,433	54,998	54,36	58,445	55,783	53,181	57,669	55,238	51,894	0,02
	2	57,523	53,883	54,312	58,89	55,601	54,195	58,669	54,709	52,188	0,32
	3	57,38	54,155	53,035	57,995	55,381	54,456	59,355	54,874	53,371	2,29

Продовження таблиці 3.8

1	2	3	4	5	6	7	8	9	10	11	12
(125, 250, 375)	1	57,188	55,97	54,723	57,496	54,655	55,876	57,103	53,761	53,227	0,54
	2	56,56	56,677	52,213	57,971	56,441	54,375	57,226	54,623	53,914	1,29
	3	56,878	55,085	52,159	57,614	56,482	54,562	57,309	55,812	54,099	0,53
(1, 250, 500)	1	62,598	58,012	46,165	62,767	56,343	49,065	61,763	55,39	47,897	0,27
	2	61,693	58,206	44,534	63,341	57,57	47,566	62,887	55,675	48,528	0,72
	3	61,54	56,606	45,086	62,293	57,654	47,905	63,507	56,905	48,503	1,91

У рамках заданої бази еталонних зображень та представлених наборів центрів, модель класифікації (2.11) для кожної компоненти окремо виявилась недієвою. Жоден з випадків не показав коректного результату. Цікаве спостереження, що частіше максимальна сума належала медоїду чи найближчому до нього дескриптору, як у випадку з набором центрів (125, 250, 375). Тобто, це ще раз доводить, що *медоїд значної мірою впливає на результат класифікації*.

Вирішено впровадити додаткову процедуру для підвищення інформативності даних, а саме інтегрування найбільших двох значень розподілів для кожного дескриптору у рамках класу (табл. 3.9).

Таблиця 3.9 – Принцип роботи моделі (2.11) з інтеграцією двох найбільших значень у рамках класу

№	Критерій класифікації θ								
	1			2			3		
1	2	3	4	5	6	7	8	9	10
1	0,114	0,119	0,099	0,097	0,1	0,101	0,119	0,124	0,127
2	0,1	0,113	0,112	0,115	0,114	0,113	0,11	0,112	0,111
3	0,098	0,109	0,102	0,124	0,118	0,112	0,116	0,11	0,11
4	0,124	0,113	0,105	0,113	0,11	0,114	0,11	0,107	0,104

Продовження таблиці 3.9

1	2	3	4	5	6	7	8	9	10
5	0,125	0,11	0,113	0,114	0,107	0,107	0,108	0,105	0,111
…	…			…			…		
∑	115,346			113,386			111,131		

Результати описаної моделі класифікації занесені до таблиці 3.10.

Таблиця 3.10 – Результат класифікації моделі (2.11) з інтеграцією двох найбільших значень у рамках класу

Варіант центрів	№ вхідного зображення	Критерій класифікації θ			Δ,%
		1	2	3	
(1, 2, 3)	1	115,346	113,386	111,131	1,7
	2	113,423	113,677	112,449	0,22
	3	112,523	112,505	114,2	1,47
(1, 25, 50)	1	118,355	115,495	116,077	1,92
	2	117,299	115,701	117,143	0,13
	3	117,139	114,681	119,169	1,7
(1, 125, 250)	1	119,569	118,943	116,302	0,52
	2	117,549	120,035	116,91	2,07
	3	117,236	119,038	118,23	0,68
(125, 250, 375)	1	121,771	119,954	117,664	1,49
	2	119,127	120,065	118,418	0,78
	3	119,115	119,723	119,568	0,13
(1, 250, 500)	1	125,129	125,07	119,856	0,05
	2	123,531	125,073	120,954	1,23
	3	122,493	124,251	122,963	1,04

Результати дещо схожі з тими, що було отримано під час першого досліду з інтеграцією усіх значень класу. Як і у попередньому випадку, тільки набір центрів (1, 2, 3) зміг вірно класифікувати зображення. Значення Δ дорівнюють (1,7%, 0,22%, 1,47%), що являється найкращим результатом з отриманих: (0,09%, 0,72%, 1,77%) для єдиного центру та (0,19%, 0,05%, 2,41%) для трьох із загальною інтеграцією за класом.

Можна зробити декілька попередніх висновків. Перший полягає у тому, що модель (2.11) може використовуватися для класифікації зображень з використанням багатокомпонентних множин описів класів. Як мінімум, для використаної бази еталонів, варіації з інтеграцією усіх розподілів та двох найбільших у рамках класу показали вірні результати для одного з наборів центрів, тобто їх варто продовжити перевіряти для інших. Модифікація з інтеграцією розподілів для кожного центру окремо виявилася найменш вдалою, тож надалі в рамках даної класифікаційної роботи використовуватись не буде.

Другий висновок стосується використання розглянутої примітивної процедури побудови центрів класу, а саме те, що вона виявилася недостатньо дієвою для моделі класифікації (2.11). Тільки один набір центрів (1, 2, 3) із запропонованих показав вірний результат, але не можна сказати, що використання трьох класів замість одного значною мірою покращило впевненість побудованих класифікаторів.

Далі, вказані набори центрів досліджувалися з використанням різних варіацій моделі, що базується на підрахунку голосів кожного дескриптора (2.13). Спочатку розглянуто метод, коли голос віддається тому класу, що має найбільшу суму розподілів.

Таблиця 3.11 показує принцип роботи даної процедури, де кольором помічені вектори розподілів за які віддаються голоси та загальна сума голосів за якою приймається рішення.

Таблиця 3.11 – Принцип роботи моделі (2.13) з голосуванням за найбільшу суму розподілів

№	Критерій класифікації θ								
	1			2			3		
1	0,114	0,119	0,099	0,097	0,1	0,101	0,119	0,124	0,127
2	0,1	0,113	0,112	0,115	0,114	0,113	0,11	0,112	0,111
3	0,098	0,109	0,102	0,124	0,118	0,112	0,116	0,11	0,11
4	0,124	0,113	0,105	0,113	0,11	0,114	0,11	0,107	0,104
5	0,125	0,11	0,113	0,114	0,107	0,107	0,108	0,105	0,111
…	…			…			…		
∑	180			172			148		

До таблиці 3.12 записано результати дослідження даної моделі класифікації для обраних наборів центрів.

Таблиця 3.12 – Результат класифікації моделі (2.13) з голосуванням за найбільшу суму розподілів

Варіант центрів	№ вхідного зображення	Критерій класифікації θ			Δ,%
		1	2	3	
1	2	3	4	5	6
(1, 2, 3)	1	180	172	148	4,44
	2	124	186	190	2,11
	3	103	182	215	15,35
(1, 25, 50)	1	181	178	141	1,66
	2	156	180	164	8,89
	3	128	153	219	30,14
(1, 125, 250)	1	196	187	117	4,59
	2	156	198	146	21,21

Продовження таблиці 3.12

1	2	3	4	5	6
	3	134	188	178	5,32
(125, 250, 375)	1	185	189	126	2,12
	2	174	199	127	12,56
	3	139	187	174	6,95
(1, 250, 500)	1	134	177	189	6,35
	2	116	161	223	27,8
	3	93	16	241	31,12

З п’яти варіантів множин центрів вдалою виявилася тільки одна – (1, 25, 50). Порівнюючи результати з єдиним центром (14,71%, 4,35%, 16,59%), отримані результати дещо покращили впевненість класифікації для 2-го та 3-го еталонів. Значення Δ для них зросло майже у 2 рази, але при цьому значення 1-го зменшилось у 9. Інші ансамблі центрів виявилися недієвими.

По аналогії з попередньою моделлю, було розглянуто варіант голосування за найбільшим значенням розподілу. У таблиці 3.13 показаний принцип роботи алгоритму.

Таблиця 3.13 – Принцип роботи моделі (2.13) з голосуванням за найбільше значення розподілу

№	Критерій класифікації θ								
	1			2			3		
1	2	3	4	5	6	7	8	9	10
1	0,114	0,119	0,099	0,097	0,1	0,101	0,119	0,124	0,127
2	0,1	0,113	0,112	0,115	0,114	0,113	0,11	0,112	0,111
3	0,098	0,109	0,102	0,124	0,118	0,112	0,116	0,11	0,11

Продовження таблиці 3.13

1	2	3	4	5	6	7	8	9	10
4	0,124	0,113	0,105	0,113	0,11	0,114	0,11	0,107	0,104
5	0,125	0,11	0,113	0,114	0,107	0,107	0,108	0,105	0,111
…	…			…			…		
∑	256			124			111		

Даний спосіб виявився недієвим. Як і раніше, жодна множина центрів не показала коректний результат для усіх трьох зображень. Показ числових результатів не має сенсу.

Базуючись на попередній відносно успішний досвід з моделлю (2.11), далі була розглянута процедура з додатковою логічною обробкою, де голос віддається за найбільшу суму двох найбільших значень розподілу з класу (табл. 3.14).

Таблиця 3.14 – Принцип роботи моделі (2.13) з голосуванням за найбільшу суму двох найбільших розподілів

№	Критерій класифікації θ								
	1			2			3		
1	0,114	0,119	0,099	0,097	0,1	0,101	0,119	0,124	0,127
2	0,1	0,113	0,112	0,115	0,114	0,113	0,11	0,112	0,111
3	0,098	0,109	0,102	0,124	0,118	0,112	0,116	0,11	0,11
4	0,124	0,113	0,105	0,113	0,11	0,114	0,11	0,107	0,104
5	0,125	0,11	0,113	0,114	0,107	0,107	0,108	0,105	0,111
…	…			…			…		
∑	212			155			133		

Результати описаної моделі класифікації занесені до таблиці 3.15.

Таблиця 3.15 – Результат класифікації моделі (2.13) з голосуванням за найбільшу суму двох найбільших розподілів

Варіант центрів	№ вхідного зображення	Критерій класифікації θ			Δ,%
		1	2	3	
(1, 2, 3)	1	212	155	133	26,89
	2	156	176	168	4,55
	3	133	163	204	20,1
(1, 25, 50)	1	205	162	133	20,98
	2	176	171	153	2,84
	3	144	139	217	33,64
(1, 125, 250)	1	195	202	103	3,47
	2	158	226	116	30,09
	3	136	207	157	24,15
(125, 250, 375)	1	214	176	110	17,76
	2	190	185	125	2,63
	3	176	179	145	1,68
(1, 250, 500)	1	186	176	138	5,38
	2	164	177	159	7,34
	3	152	161	187	13,9

Отримано вірну класифікацію для двох множин центрів. Ансамбль (1, 250, 500) вперше показав коректний результат зі значеннями Δ (5,38%, 7,34%, 13,9%). Дивлячись на те, що Δ є слабкими, порівнюючи з єдиним центром – (14,71%, 4,35%, 16,59%), та посилаючись на попередні висновки про те, що медоїд досить сильно впливає на процес класифікації і кращими є близькі до нього центри, можна сказати, що даний результат є похибкою і не є стабільним. Тож, ним можна нехтувати.

Множина центрів (1, 2, 3) знову показала свою працездатність, при цьому змогла покращити значення Δ (26,89%, 4,55%, 20,1%) у порівнянні з єдиним центром.

Підбивши підсумки щодо модифікацій моделі (2.13), можна сказати, що даний метод може використовуватися для класифікації зображень з використанням декількох центрів для класу. Варіанти з голосуванням за найбільшу загальну суму та суму двох найбільших розподілів у рамках класу є працездатними.

Щодо використаних центрів, набори (1, 25, 50) та (1, 250, 500) виявилися дієвими, але неефективними у порівнянні з єдиним центром. Та варто зазначити, що ансамбль (1, 2, 3), при умові використання моделі з голосуванням за суму двох найбільших, дав більш впевненіший результат ніж один медоїд.

Попередні дослідження показали, що примітивного способу відбору центрів недостатньо для покращення результатів класифікації. Тож, наступні експерименти були пов’язані із розрахунком середніх бінарних векторів з використанням все тієї ж відсортованої множини дескрипторів.

Для початку був використаний набір центрів, що складався з медоїду та середніх векторів для дескрипторів з позиціями 1–250 та 251–500. Результати класифікації з використанням даних центрів наведено у таблиці 3.16.

Таблиця 3.16 – Результати класифікації з центрами (медоїд, середнє 1–250, середнє 251–500)

Модель класифікації	№ вхідного зображення	Критерій класифікації θ			Δ,%
		1	2	3	
1	2	3	4	5	6
(2.11) з інтеграцією у рамках класу	1	165,922	166,064	168,014	1,16
	2	163,294	168,519	168,187	0,2

Продовження таблиці 3.16

1	2	3	4	5	6
	3	164,062	166,838	169,099	1,34
(2.11) з інтеграцією двох найбільших у рамках класу	1	117,392	114,821	117,31	0,07
	2	115,423	116,931	117,584	0,56
	3	115,252	115,724	119,141	2,87
(2.13) з голосуванням за найбільшу суму	1	148	180	172	4,44
	2	104	227	169	25,55
	3	102	193	205	5,85
(2.13) з голосуванням за найбільшу суму двох найбільших	1	201	129	170	15,42
	2	144	186	170	8,6
	3	115	113	272	57,72

Заданий набір центрів показав свою ефективність тільки під час використання моделі (2.13) з голосуванням за найбільшу суму двох найбільших розподілів. Порівнюючи з результатами для одного центру (14,71%, 4,35%, 16,59%), бачимо суттєве покращення у виявленні класифікатору для 3-го еталону (майже у 3,5 разів) та певне покращення для інших. Для інших моделей даний ансамбль виявився неефективним.

Наступне дослідження полягало у знаходженні трьох середніх векторів, при цьому нехтуючи медоїдом у якості центру. Необхідно було розбити всю відсортовану множину дескрипторів на 3 рівні за розміром підмножини, але через те, що 500 не ділиться націло на 3, а відкидання деяких дескрипторів суперечило б принципу використання повного обсягу даних опису еталону, було прийняте рішення збільшити третю підмножину, котра містить найвіддаленіші від медоїду дескриптори. Таким чином, це дозволить певною мірою нівелювати вплив найвіддаленіших дескрипторів. Результати класифікації з використанням цих центрів вказані у таблиці 3.17.

Таблиця 3.17 – Результати класифікації з центрами (середнє 1–166, середнє 167–332, середнє 333–500)

Модель класифікації	№ вхідного зображення	Критерій класифікації θ			Δ,%
		1	2	3	
(2.11) з інтеграцією у рамках класу	1	167,543	166,329	166,128	0,72
	2	164,487	168,379	167,134	0,74
	3	164,055	166,725	169,22	1,47
(2.11) з інтеграцією двох найбільших у рамках класу	1	117,443	113,375	113,558	3,31
	2	114,533	114,618	113,887	0,07
	3	114,636	113,649	115,706	0,93
(2.13) з голосуванням за найбільшу суму	1	217	133	150	30,88
	2	115	225	160	28,89
	3	121	127	252	49,6
(2.13) з голосуванням за найбільшу суму двох найбільших	1	291	74	135	53,61
	2	182	168	150	7,69
	3	178	90	232	23,28

Отримані результати виявились найкращими з попередньо досліджених. Три з чотирьох варіацій класифікації показали коректний результат. Якщо порівнювати результати з одним центром, то для моделі (2.11) з обома варіантами модифікації покращення не спостерігалося. Значення Δ залишились майже на тому самому рівні, зміни не є суттєвими: (0,19%, 0,05%, 2,41%) для одного медоїду.

Варті уваги результати моделі (2.13) з голосуванням за найбільшу суму в рамках класу, так як впевненість класифікації зросла для всіх еталонів. З (14,71%, 4,35%, 16,59%) для одного центру до (30,88%, 28,89%, 49,6%) для трьох. Це є найкращим з отриманих результатів. Даний варіант можна вважати дієвим та ефективним.

Далі досліджено множини центрів, що складаються з дескрипторів із найбільшою інформативністю, що розрахована по формулі (2.16). Частина відсортованих за інформативністю дескрипторів зображена у таблиці 3.18.

Таблиця 3.18 – Відсортовані за інформативністю дескриптори 1-го еталону

№	№ дескриптора	Інформативність
1	404	51
2	458	50
3	402	48
4	214	47
5	137	46
…	…	…
496	89	-19
497	397	-20
498	178	-22
499	22	-23
500	25	-24

У результаті, найбільші значення інформативності дорівнюють:

– для 1-го еталону – 51, 50, 48 (для порівняння у медоїда -9);

– для 2-го еталону – 52, 50, 49 (у медоїда 15);

– для 3-го еталону – 58, 52, 51 (у медоїда 32).

Незважаючи на перспективність даного методу, він виявився недієвим. Жоден з варіантів класифікаційних моделей не показав правильного результату. У всіх випадках зображення були віднесені до 3-го еталону.

ВИСНОВКИ

У рамках кваліфікаційної роботи досліджено результативність використання багатокомпонентної моделі даних як опису класів у структурних підходах класифікації зображень.

Збільшення кількості центрів сприяє більш поглибленому аналізу, що у свою чергу підвищує ефективність побудованого класифікатору. Результативність прийнятого класифікаційного рішення цілком залежить від способу формування ансамблю центрів, використаної моделі класифікації та від самих вхідних даних.

За основу було взято дві моделі прийняття класифікаційних рішень: за максимальним значенням інтеграції статистичних розподілів у межах класу та формування значення класу для кожного дескриптору окремо з подальшим узагальненням числа голосів. Запропоновано декілька працездатних модифікацій для кожного методу задля впровадження багатокомпонентної моделі.

Проаналізовано різні способи формування центрів, а саме відбір з відсортованої за відстанню до інших множини дескрипторів, розрахунок середніх векторів та виявлення дескрипторів з максимальною інформативністю. Найбільш ефективним виявився ансамбль середніх векторів. Використавши модель класифікації з голосуванням за клас, вдалося покращити впевненість результату у кілька разів. Трішки гіршими виявилися результати використання медоїду та двох середніх для моделі класифікації з голосуванням за два найбільші розподіли класу. Варто зазначити, що медоїд та два найближчі до нього (1, 2, 3) також показали певний результат для моделей з інтеграцією двох найбільших та голосування за два найбільші розподіли.

Класифікація з роздільним голосуванням дескрипторів показує кращий результат з використанням багатокомпонентної моделі даних. Підхід з

вибором декількох центрів для класу є працездатним, але потребує подальшого дослідження для виявлення спектру більш стабільних ансамблів.

До практичної значущості належить розробка модифікованих структурних методів класифікації, результати експериментів із заданою базою еталонних зображень і їх дослідження та програмна реалізація запропонованих методів для подальшого використання у системах комп’ютерного зору.

Перспективи дослідження можуть полягати у знаходженні універсальних процедур виділення центрів, адаптованих до певного типу вхідних даних. Окремої уваги варто приділити оптимізації алгоритмів класифікації, так як зі збільшенням кількості центрів збільшується обсяг розрахунків.

Результати дослідження апробовано у вигляді двох наукових статей [23, 31], опублікованих у науковому журналі «Сучасні інформаційні системи» та двох тез доповідей [24, 32].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Daradkeh, Y. I., Gorokhovatskyi, V., Tvoroshenko, I., Gadetska, S., & Al-Dhaifallah, M. (2021). Methods of Classification of Images on the Basis of the Values of Statistical Distributions for the Composition of Structural Description Components. *IEEE Access*, *9*, 92964-92973.

2. Гороховатський, В. О., Гадецька, С. В., Стяглик, Н. І., & Власенко, Н. В. (2020). Класифікація зображень на підставі ансамблю статистичних розподілів за класами еталонів для компонентів структурного опису.

3. Гороховатський, В. О., & Пономаренко, Р. П. (2020). Класифікація зображень на підставі формування незалежної системи кластерів у складі структурних описів бази еталонів.

4. Гороховатський, В. О., & Гадецька, С. В. (2020). Статистичне оброблення та аналіз даних у структурних методах класифікації зображень.

5. Krig, S. (2016). Interest point detector and feature descriptor survey. In *Computer vision metrics* (pp. 187-246). Springer, Cham.

6. Khanna, P. (2019, November 23). *Harris Corner Detector-an overview of the original paper*. Medium. Retrieved October 16, 2022, from https://medium.com/swlh/harris-corner-detector-an-overview-of-the-original-paper-cf20c502ab0f.

7. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91-110.

8. Tyagi, D. (2020, April 7). *Introduction To Feature Detection And Matching.* Medium. Retrieved October 16, 2022, from https://medium.com/data-breach/introduction-to-feature-detection-and-matching-65e27179885d.

9. Alahi, A., Ortiz, R., & Vandergheynst, P. (2012, June). Freak: Fast retina keypoint. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 510-517). Ieee.

10. Işık, Ş. (2014). A comparative evaluation of well-known feature detectors and descriptors. *International Journal of Applied Mathematics Electronics and Computers*, *3*(1), 1-6.

11. Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision* (pp. 2564-2571). Ieee.

12. Rosten, E., & Drummond, T. (2005, October). Fusing points and lines for high performance tracking. In Tenth *IEEE International Conference on Computer Vision (ICCV'05) Volume 1 (Vol. 2, pp. 1508-1515)*. Ieee.

13. Luo, C., Yang, W., Huang, P., & Zhou, J. (2019, June). Overview of image matching based on ORB algorithm. In *Journal of Physics: Conference Series* (Vol. 1237, No. 3, p. 032020). IOP Publishing.

14. Tyagi, D. (2020, April 7). *Introduction to ORB (Oriented FAST and Rotated BRIEF)*. Medium. Retrieved October 16, 2022, from https://medium.com/data-breach/introduction-to-orb-oriented-fast-and-rotated-brief-4220e8ec40cf.

15. Xingteng, J., Xuan, W., & Zhe, D. (2015, October). Image matching method based on improved SURF algorithm. In *2015 IEEE International Conference on Computer and Communications (ICCC)* (pp. 142-145). IEEE.

16. Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010, September). Brief: Binary robust independent elementary features. In *European conference on computer vision* (pp. 778-792). Springer, Berlin, Heidelberg.

17. Liu, C., Xu, J., & Wang, F. (2021). A Review of Keypoints' Detection and Feature Description in Image Registration. *Scientific Programming*, 2021.

18. Гороховатський, В. О., Пупченко, Д. В., & Солодченко, К. Г. (2018). Аналіз властивостей, характеристик та результатів застосування новітніх детекторів для визначення особливих точок зображення.

19. Gorokhovatskyi, V., & Vlasenko, N. (2021). Редукція опису зображення у складі множини дескрипторів на основі метричного критерію інформативності. *Advanced Information Systems*, *5*(4), 10-16.

20. Gorokhovatskyi, V., Gadetska, S., & Ponomarenko, R. (2019, May). Recognition of visual objects based on statistical distributions for blocks of structural description of image. In *International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence"* (pp. 501-512). Springer, Cham.

21. Gadetska, S. V., Gorokhovatskyi, V. O., Stiahlyk, N. I., & Vlasenko, N. V. (2021). Statistical data analysis tools in image classification methods based on the description as a set of binary descriptors of key points. *Radio Electronics, Computer Science, Control*, (4), 58-68.

22. Гороховатский, В. А. (2014). Структурный анализ и интеллектуальная обработка данных в компьютерном зрении.

23. Гороховатський, В. О., Стяглик, Н. І., & Жадан, О. В. (2022). Застосування багатокомпонентної моделі даних для описів класів у задачі класифікації зображень.

24. Жадан, О. (2022). Класифікація зображень на основі багатокомпонентних моделей даних.

25. Гороховатський, В. О., Гадецька, С. В., & Стяглик, Н. І. (2019). Вивчення статистичних властивостей моделі блочного подання для множини дескрипторів ключових точок зображень властивостей моделі блочного подання для множини дескрипторів ключових точок зображень.

26. Gorokhovatskyi, V. O., & Gadetska, S. V. (2019). Determination of Relevance of Visual Object Images by Application of Statistical Analysis of Regarding Fragment Representation of their Descriptions. *Telecommunications and Radio Engineering*, *78*(3).

27. Gadetska, S., Gorokhovatskyi, V., Stiahlyk, N., & Vlasenko, N. (2022). Aggregate Parametric Representation of Image Structural Description in Statistical Classification Methods.

28. Ibrahim, D. Y., Gorokhovatskyi, V., Tvoroshenko, I., & Zeghid, M. (2022). Cluster representation of the structural description of images for effective classification.

29. Ayaz, A. M., Gorokhovatskyi, V., Tvoroshenko, I., Vlasenko, N., & Khalid, M. S. (2021). The Research of Image Classification Methods Based on the Introducing Cluster Representation Parameters for the Structural Description.

30. Gorokhovatskyi, V., Gorokhovatskyi, O., Yevgenyi, P., & Olena, P. (2018, August). Quantization of the space of structural image features as a way to increase recognition performance. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)* (pp. 464-467). IEEE.

31. Gorokhovatsky, V., Gadetska, S., Zhadan, O., & Khvostenko, O. (2021). Дослідження результативності класифікаторів зображень за статистичними розподілами для компонентів структурного опису. *Advanced Information Systems*, 5(1), 5-11.

32. Жадан, О. В. (2021). Метод стиснення множини ключових точок зображення.

33. Flach, P. (2012). *Machine learning: the art and science of algorithms that make sense of data*. Cambridge university press.

34. Gorokhovatskyi, V. A. (2018). Image classification methods in the space of descriptions in the form of a set of the key point descriptors. *Telecommunications and Radio Engineering*, *77*(9), 787-797.

35. Гороховатский, В. А., & Передрий, Е. О. (2009). Корреляционные методы распознавания изображений путем голосования систем фрагментов. *Радіоелектроніка, інформатика, управління*, (1 (20)), 74-81.

36. Gorokhovatskiy, V. A., & Zamula, A. A. (2016). Employment of intelligent technologies in multiparametric control systems. *Telecommunications and Radio engineering*, *75*(19).

37. Gorokhovatsky, V., Stiahlyk, N., & Tsarevska, V. (2021). Combination method of accelerated metric data search in image classification problems. *Advanced Information Systems, 5 (3)*, 5-12.

38. Daradkeh, Y. I., Gorokhovatskyi, V., Tvoroshenko, I., & Al-Dhaifallah, M. (2022). Classification of Images Based on a System of Hierarchical Features. *Computers, Materials & Continua*, *72*(1), 1785-1797.

39. Tvoroshenko, I., & Gorokhovatskyi, V. (2022). The Application of Hybrid Intelligence Systems for Dynamic Data Analysis.

40. Гороховатський, В. О., & Творошенко, І. С. (2022). Аналіз багатовимірних даних за описом у формі множини компонент.

41. Gorokhovatskyi, V., Tvoroshenko, I., & Chmutov, Y. (2022). Застосування систем ортогональних функцій для формування простору ознак у методах класифікації зображень. *Advanced Information Systems, 6*(3), 5-12.

42. Gorokhovatskyi, V. A., Rusakova, N., & Tvoroshenko, I. S. (2020). The application of image analysis methods and predicate logic in applied problems of magnetic monitoring. *Telecommunications and Radio Engineering, 79*(20).

43. Гороховатський, В. О., Творошенко, І. С., & Сидоренко, Д. (2021). Класифікація зображень із використанням кластерного подання.

44. Jin, X., & Han, J. W. (2010). K-Medoids clustering, Encyclopedia of machine learning (pp. 564–565).

45. *Why use LINQ in .NET*. Edureka. (2020, October 28). Retrieved October 20, 2022, from https://www.edureka.co/blog/unleash-the-power-of-linq-the-net-way/.

46. Shi, S. (2013). *Emgu CV Essentials.* Packt Publishing.

47. *What is Visual Studio?* Incredibuild. (2022, August 15). Retrieved October 24, 2022, from https://www.incredibuild.com/integrations/visual-studio.