Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

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

АТЕСТАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

ДОСЛІДЖЕННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ЗОБРАЖЕНЬ НА ОСНОВІ ДЕСКРИПТОРІВ ЛОКАЛЬНИХ ОСОБЛИВОСТЕЙ
(тема)

Виконав:
студент 2 курсу, групи ІНФм-19-1

Ткаченко Д.А
(прізвище, ініціали)

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

Освітня програма Інформатика
(повна назва освітньої програми)

Керівник доц. Творошенко І.С
(посада, прізвище, ініціали)

Допускається до захисту

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

2020 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)

Освітня програма Інформатика
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«____»_____ 20 ___ р.

ЗАВДАННЯ
НА АТЕСТАЦІЙНУ РОБОТУ

студентові Ткаченку Дмитру Андрійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів інтелектуального аналізу зображень на основі дескрипторів локальних особливостей

затверджена наказом по університету від « 23 » жовтня 2020 року № 1428Ст.

2. Термін подання студентом роботи до екзаменаційної комісії 24 листопада 2020 р.

3. Вихідні дані до роботи *Математичні моделі перетворювання зображень, база даних WordNet,теоретичні відомості про методи сегментації зображень*

4. Перелік питань, що потрібно опрацювати в роботі

1. Дослідження можливості використання автоматично отриманої навчальної вибірки для задач класифікації і порівняльний аналіз різних підходів до класифікації в цьому випадку

2. Розробка методу фільтрації пошукової видачі від нерепрезентативним примірників зображень

3. Порівняльний аналіз і розробка ефективних алгоритмів пошуку відповідностей між дескрипторами.

4. Розробка моделі представлення зображень у вигляді сегментів для задач структури з руху

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) *вхідні тестові зображення, таблиці результатів*

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

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на атестаційну роботу	23.10.2020	**виконано**
2	Аналіз завдання, підбір літератури	24.10.20-26.10.20	**виконано**
3	Аналіз літератури з досліджуваної проблеми	27.10.20-29.10.20	**виконано**
4	Аналіз технічних засобів	29.10.20-30.10.20	**виконано**
5	Розробка методу	00.11.20-00.11.20	**виконано**
6	Програмна реалізація	01.11.20-10.11.20	**виконано**
7	Оформлення пояснювальної записки	10.11.20-15.11.20	**виконано**
8	Перевірка на плагіат	01.12.20	**виконано**
9	Рецензування	02.12.20	**виконано**
10	Підготовка презентації та доповіді	05.12.20	**виконано**
11	Занесення роботи в електронний архів	06.12.20	**виконано**
12	Попередній захист атестаційної роботи	07.12.20	**виконано**

Дата видачі завдання 23 жовтня 20 20 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Творошенко І.С.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ/ABSTRACT

Пояснювальна записка до атестаційної роботи: 83 с., 7 табл., 18 рис., 35 джерел.

ОБРОБКА ЗОБРАЖЕНЬ, СЕМАНТИЧНА КОРРЕКЦІЯ, КЛАСИФІКАЦІЯ ЗОБРАЖЕНЬ, МОДЕЛЬ BAG-OF-WORDS, КОМП'ЮТЕРНИЙ ЗІР, АНАЛІЗ ЗОБРАЖЕНЬ.

У роботі увага концентрується на двох аспектах використання дескрипторів: задачі класифікації зображень і задачі вилучення геометрії з наборів зображень. У задачі класифікації зображень увага звертається на проблеми, пов'язані з ручним механізмом формування навчальної вибірки і на проблеми, пов'язані з відсутністю відносин класів між собою. У задачі добування інформації про геометрію об'єктів увага звертається на проблему надлишкової фільтрації відповідностей дескрипторів при пошуку співвідношень зображень між собою.

Застосовувалися методи комп'ютерної графіки, методи теорії обробки сигналів і теорії графів, методи математичної статистики і теорії ймовірності.

У результаті роботи проаналізовані негативні побічні ефекти, пов'язані з аналізом зображень шляхом опису дескрипторами локальними візуальних особливостей. Запропоновано підходи до використання дескрипторів локальних візуальних особливостей, що дозволяють поліпшити якість аналізу зображень.

IMAGE PROCESSING, SEMANTIC CORRECTION, IMAGE CLASSIFICATION, BAG-OF-WORDS MODEL, COMPUTER VISION, IMAGE ANALYSIS.

The paper focuses on two aspects of the use of descriptors: the problem of image classification and the problem of extracting geometry from sets of images. In the problem of image classification, attention is paid to the problems associated with the manual mechanism of formation of the educational sample and the problems associated with the lack of relationship between classes. In the problem of obtaining information about the geometry of objects, attention is paid to the problem of excessive filtering of descriptor correspondences when searching for the ratios of images among themselves.

Methods of computer graphics, methods of signal processing theory and graph theory, methods of mathematical statistics and probability theory were used.

As a result, the negative side effects associated with image analysis by describing descriptors of local visual features are analyzed. Approaches to the use of descriptors of local visual features are proposed, which allow to improve the quality of image analysis.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

BoW – Bag Of Words

NBNN – Naive Bayesian Nearest Neighbor

SVM – Support Vector Machine

ВСТУП

Область комп’ютерного зору може бути охарактеризована як молода, різноманітна і динамічна. І хоча існують більш ранні роботи, інтенсивне вивчення цієї проблеми почалося, коли комп’ютери змогли управляти обробкою великих наборів даних, таких як зображення. Однак ці дослідження зазвичай починалися з інших областей, і, отже, немає стандартної формулювання проблеми комп’ютерного зору. Також, і це навіть більш важливо, немає стандартної формулювання того, як повинна вирішуватися проблема комп’ютерного зору. Замість цього, існує маса методів для вирішення різних строго певних завдань комп’ютерного зору, де методи часто залежать від завдань і рідко можуть бути узагальнені для широкого кола застосування. Багато з методів і додатків все ще знаходяться в стадії фундаментальних досліджень, але все більше число методів знаходить застосування в комерційних продуктах, де вони часто складають частину більшої системи, яка може вирішувати складні завдання (наприклад, в області медичних зображень або вимірювання і контролю якості в процесах виготовлення). У більшості практичних застосувань комп’ютерного зору комп’ютери попередньо запрограмовані для вирішення окремих завдань, але методи, засновані на знаннях, стають все більш загальними.

Важливу частину в області штучного інтелекту займає автоматичне планування або прийняття рішень в системах, які можуть виконувати механічні дії, такі як переміщення робота через деяку середу. Цей тип обробки зазвичай потребує вхідних даних, що надаються системами комп’ютерного зору, що діють як відеосенсор і надають високорівневу інформацію про середовище і роботі. Інші області, які іноді описуються як належать до штучного інтелекту і які використовуються щодо комп’ютерного зору, це розпізнавання образів і навчальні методи. В результаті, комп’ютерний зір іноді розглядається як частина області штучного інтелекту або області комп’ютерних наук взагалі.

Ще однією областю, пов’язаної з комп’ютерним зором, є обробка сигналів. Багато методів обробки одновимірних сигналів, зазвичай тимчасових сигналів, можуть бути природним шляхом розширені для обробки двовимірних або багатовимірних сигналів в комп’ютерному зорі. Однак через своєрідну природи зображень існує багато методів, розроблених в області комп’ютерного зору і не мають аналогів в області обробки одновимірних сигналів. Особливою властивістю цих методів є їх нелінійність, що, разом з багатомірністю сигналу, робить відповідну подобласть в обробці сигналів частиною області комп’ютерного зору.

Крім згаданих підходів до проблеми комп’ютерного зору, багато досліджуваних питань можуть бути вивчені з чисто математичної точки зору. Наприклад, багато методи грунтуються на статистиці, методах оптимізації або геометрії. Нарешті, великі роботи ведуться в області практичного застосування комп’ютерного зору – того, як існуючі методи можуть бути реалізовані програмно і апаратно або як вони можуть бути змінені з тим, щоб досягти високої швидкості роботи без істотного збільшення споживаних ресурсів.

1 СПЕЦИФІКА АНАЛІЗУ ЗОБРАЖЕНЬ НА ОСНОВІ ДЕСКРИПТОРІВ ЛОКАЛЬНИХ ОСОБЛИВОСТЕЙ

1.1 Класифікація та особливості задач комп’ютерного

Людина здатна з легкістю сприймати навколишній реальний світ за допомогою візуальної інформації одержуваної від органів зору. Наш мозок здатний добудовувати плоске зображення, що отримується очима, до тривимірного зображення в свідомості, наприклад, людина без праці розуміє, що ваза має форму тіла обертання, спостерігаючи її стільки з однієї точки. Дивлячись на фотографію, свідомість практично миттєво видає інформацію про людей, зображених на ній, включаючи весь спектр супутніх даних, наприклад, пов’язані з людьми події. Ми здатні добудовувати зображення на основі лише часткових, або навіть схематично зображених уривків, при цьому використовуючи як весь фізіологічний функціонал, так і знання, і досвід накопичені раніше за життя.

Комп’ютерне зір – розділ інформаційних технологій, який досліджує можливості машин витягувати інформацію із зображень, отриманих з різних сенсорів і таким чином в деякій мірі емулювати людське візуальне сприйняття. Даною науці вже кілька десятків років, і вона продемонструвала значні результати в завданнях одних класів, в той час як у багатьох інших прогрес відносно низький. Основний проблематикою даної області можна назвати те, що вона намагається емулювати поведінка, в загальному, маловивчених і слабо зрозумілих процесів, що відбуваються при сприйнятті людиною візуальної інформації. У той час як людина може використовувати весь багаж накопиченого досвіду і знань для прийняття рішень щодо інтерпретацій візуального зображення (і, отже, будь-яке таке сприйняття є суб’єктивним), переважна кількість алгоритмів в комп’ютерному зорі є детермінованими і результат їх роботи залежить тільки від вхідних даних. З цієї причини комп’ютерний зір на сьогоднішній день неможливо розглядати

як комплексну теорію, швидше за це множина підходів, методів і алгоритмів, спрямованих на вирішення різних теоретичних і прикладних задач, слабо пов’язаних між собою, або ж іноді використовують прямо протилежні підходи для досягнення результату в різних завданнях.

Складність, притаманна комп’ютерного зору, частково виникає з розташування цієї області на стику багатьох наук і сфер, серед яких можна перерахувати:

– фізика, в першу чергу оптика, а також всі інші пов’язані з поширенням світла розділи;

– нейробіологія, як наука займається вивченням принципів роботи людського сприйняття;

– штучний інтелект, наприклад, для використання елементів пошуку за шаблоном і методів навчання;

– обробка сигналів, багато методи комп’ютерного зору вимагають первісної обробки і трансформації вхідних даних;

– машинне навчання та прикладна математика. У комп’ютерному зорі вирішується широкий спектр обчислювальних задач, які вимагають ефективного математичного апарату для роботи з великими обсягами даних.

Так само слід окремо відзначити особливості сфери, що випливають з обсягів оброблюваної інформації. Великі розмірності зображень, помножені на потенційно чимале число пов’язаних між собою зображень і на необхідність працювати в різних масштабах, призводять до того, що практично будь-які методи, інтуїтивно здаються простими, не можуть використовуватися без глибокого перегляду з боку можливих оптимізацій. Практично будь-які методи, які працюють «в лоб», не подаються можливими до використання через неприйнятні тимчасових і ресурсних витрат. З цього протікає та особливість, що багато методів комп’ютерного зору були отримані «від противного» – тобто максимальна можлива результативність виходячи з поточного розвитку доступної обчислювальної техніки (нехай і з

множинними припущеннями і похибками), а не виходячи з прямого моделювання, наприклад, фізичних процесів.

Серед великих областей, що розглядаються комп'ютерним зором, можна умовно виділити наступні:

– завдання розпізнавання;

– завдання реконструкції сцени;

– завдання аналізу відео.

Цей поділ досить вільне, так як в областях іноді використовуються пересічні набори підходів і методів, так само перетинається фундаментальний інструментарій, пов'язаний з обробкою зображень. Розглянемо зміст даних областей більш докладно.

Завдання розпізнавання. Комп'ютерне зір виділяє певні розрізи серед завдань розпізнавання, які відрізняються цього приладу:

– ідентифікація об'єкта – пошук примірників об'єкта на представленому зображенні, можливо в спотвореному вигляді, проте зі збереженням візуальних особливостей об'єкта. Ці завдання можуть вирішуватися різними способами, в залежності від конкретної області застосування;

– завдання виявлення – пошук будь-яких областей за заданими критеріями, без чітких візуальних особливостей. Областями застосування можна назвати діагностику в медицині і системи відеоспостереження, системи пошуку людей і осіб на фотографіях;

– завдання сегментування зображень – виділення пов'язаних областей, можливо відділення фону від знаходиться перед ним предмета;

– завдання класифікації зображень – привласнення входять зображенням міток з набору класів, виходячи зі змісту зображень;

– завдання класифікації через локалізацію – одночасний пошук розташування примірників різних класів на зображенні з наступним присвоєнням міток виходячи зі знайдених примірників.

Завдання реконструкції сцени. У широкому розумінні під цими завданнями можна розуміти вилучення інформації зі зв’язаних між собою (можливо, невідомим чином) зображень. Серед розділів цього класу задач можна виділити наступні:

– «зшивання» зображень – найпростіший випадок даного класу задач, при якому не потрібно обчислювати геометрію сцени, а тільки взаємне розташування пересічних зображення і відповідне перетворення, що переводить координати одного зображення в координати іншого;

– структура з руху – найпростіший випадок вилучення інформації про геометрію об’єкта з набору його зображень з різних точок. Включає в себе завдання калібрування камери, завдання побудови розширеної реальності, завдання вилучення найпростішої геометрії заснованої на лініях і площинах;

– реконструкція 3D моделі з набору зображень – широкий спектр завдань реконструкції сцени, який передбачає вилучення інформації про об’ємної геометрії предметів з набору зображень цих предметів (сцени). Може мати відчутні відмінності в підходах виходячи з масштабу завдання – від моделювання невеликого предмета за допомогою звичайної камери, до моделювання великих просторів (large scale reconstruction) завдяки великій кількості непов’язаних між собою зображень отриманих з різних камер.

Завдання аналізу відео. У широкому розумінні ці завдання можна сформулювати як вилучення інформації з зв’язкового послідовного набору зображень. Специфічну складність цих завдань надає обсяг інформації, який необхідно обробляти, пропорційний кількості кадрів для offline систем і частоті кадрів в секунду для систем реального часу.

Можна виділити наступні завдання:

– виділення фону з відео, знятого з фіксованою точки. При вирішенні цього завдання окрему проблему представляють собою, крім переміщаються об’єктів, артефакти стиснення потоку і необхідність аналізувати потенційне зміна освітлення сцени;

– завдання виявлення і стеження за об’єктами. Вона включає в себе виявлення об’єкта в поле зору камери, зазвичай з використанням заздалегідь навчених класифікаторів, і подальше спостереження за переміщенням об’єкта, можливо з системою передбачення траєкторії руху;

– завдання розпізнавання дій. По набору ознак, зафіксованих в відео, і їх просторово-часової взаємозв’язку, використовуючи набір класифікаторів і евристик, можна робити припущення про «характер», що відбувається в потоці.

Можна навести множину прикладних областей з повсякденного життя, в яких методи комп’ютерного зору успішно застосовуються:

– розпізнавання тексту і даних. Конвертація растрових зображень в набір символів. Сюди ж можна віднести такі вузькоспеціалізовані випадки, як читання номерів автомобілів, поштових індексів на конвертах, штрих-кодів і QR-кодів;

– пошук осіб на фотографії. Сьогодні багато компактних камер здатні виконувати цю операцію на льоту (наприклад, для визначення області фокусування), в подальшому більш складні системи здатні виконувати пошук людей в наборі фотографій;

– системи відеоспостереження, як окремо розташованих будівель, так і автомобільних трас;

– біометрія та пов’язані з цим завдання, такі як ідентифікація людини за відбитками пальців;

– панорами міст – сьогодні картографічні сервіси (наприклад Google і Yandex) надають можливість подивитися панорами зі знімків отриманих при русі машин по дорогах. При обробці цих даних так само використовуються методи з комп’ютерного зору.

У той час як деякі завдання розпізнавання в рамках комп’ютерного зору можна вважати в тій чи іншій мірі дозволеними (як, наприклад, завдання локалізація конкретного об’єкта (object detection) або пошуку примірників об’єкта – instance recognition), завдання класифікації зображень (category

recognition) (в цілому, або частинами) залишається вкрай важкою. Причин цього кілька:

– візуальна мінливість предметів, що відносяться до певної категорії;

– допустимі структурні відмінності між предметами відносяться до однієї категорії;

– неможливість екстенсивного нарощування навчальної вибірки, тому що комбінаторний вибух призводить до перенавчання;

– зв’язок між категоризацією об’єкта і зовнішнім контекстом;

– нерозуміння, як працюють дані механізми в свідомості людини.

Окрему складність може представляти те, що можливі ситуації, коли жоден з класів, яким навчена система розпізнавання, що не представлений на уже згадуваному зображенні. В цьому випадку система розпізнавання повинна видати відсутність наявних класів, що призводить до використання деяких порогових значень для механізму прийняття рішень, підбір яких несе в собі окрему складність [1].

Більшість існуючих на сьогоднішній день механізмів класифікації спираються на візуальні особливості зображень. Візуальні особливості (visual features) добре зарекомендували себе в задачах локалізації об’єкта і пошуку примірників об’єктів в 1990-х роках, в 2000-х роках були зроблені спроби використовувати цей же механізм для класифікації зображень. Використання даних про візуальних особливості відрізняється в різних методах класифікації зображень, але можна виділити якийсь загальний алгоритм:

а) складання навчальної вибірки зображень, розбитих (labeled) за класами;

б) витяг візуальних особливостей з навчальної вибірки;

в) перетворення отриманих даних для подальшої роботи;

г) використання оброблених даних для аналізу чергового вхідного зображення і прийняття рішень про приналежність його до певного класу.

Можна відзначити наступні проблеми, присутні в описаній вище схемі навчання і класифікації:

– бібліотеки «еталонних» зображень, що представляють окремі класи, формуються вручну, що знижує гнучкість і накладає необхідність підтримки бібліотеки в актуальному стані;

– вибір класифікатора може в значній мірі залежати від характеристик оброблюваних зображень, і, в разі появи невідповідних зображень, надійність класифікації різко падає;

– при значній кількості класів точність класифікації знижується, так як стає важко скласти однозначно характеризує модель кожного окремого класу;

– класи зображень не мають між собою зв’язків з семантичної навантаженням, що не дозволяє в подальшому інтерпретувати результати роботи.

Крім локальних візуальних особливостей, обчислених в околицях стійких опорних точок, вихідна інформація, яка описувала навчальну вибірку, може бути отримана іншими способами:

– обчисленням гістограм орієнтованих градієнтів (HOG) зображень. Традиційно такий підхід використовується для задач локалізації через класифікацію при обробці відео в режимі реального часу, наприклад, для розпізнавання автомобілів і пішоходів в системах відеоспостереження;

– обчисленням так званих «щільних» дескрипторів, які витягуються виходячи з регулярної сітки, на яку розбивається зображення, без урахування стійкості цих точок в просторі і масштабі;

– можливі комбінації зазначених вище підходів з використанням просторової піраміди, що дозволяє досягти деякого рівня інваріантності до змін масштабу зображень, однак не вирішує в цілому проблему відсутності інваріантності у подібного підходу вилучення дескрипторів [2].

При наявності заздалегідь підготовлених бібліотек, що містять різні текстури фонів зображень, можливе використання інформації про фон поряд традиційним застосуванням дескрипторів для класифікації зображення в цілому. Для цього окремо тренується візуальний словник відомих текстур на

основі наявних бібліотек, він застосовується для розпізнавання фонів аналізованих зображень. Потім він використовується для відділення фону зображення від предметів на передньому плані і використання цієї інформації як додаткового дискримінаційного фактора поряд з традиційним застосуванням методу візуальних слів. Незважаючи на перспективність аналізу текстур, мінусами в даному підході можна відзначити високий ступінь залежності від дозволу аналізованих зображень і необхідність ручного формування бібліотек текстур.

Були спроби використовувати пошукові системи для вилучення візуально узгоджуються результатів за однаковими пошуковим запитам. При цьому вхідний інформацією був текст на природній мові для пошукового запиту і проводилася спроба сформувати найбільш ймовірну модель зображень по даному запиту на основі імовірнісного латентносемантичного аналізу (pLSA – probabilistic Latent Semantic Analysis). На зображення по одному запиту накладалося обмеження візуальної узгодженості, при якій візуальні слова повинні розташовуватися в однаковому положенні відносно один одного виходячи з розбиття простору зображення на ділянки відповідно до методу Hough Transform. Даний підхід корисний для пошуку характерних візуальних уявлень класів, виходячи з їх назви на природній мові, однак накладає суттєві обмеження як на використання доступної з пошукових систем інформації, так і на подальше застосування знайдених зображень.

Можна виділити окремі напрямки, які стоять на стику задач класифікації зображень і завдань вилучення геометрії та суміщення зображень. Так, наприклад, вирішується завдання класифікації вхідного зображення між різними відомими місцями розташування (Landmarks). Для цього весь доступний масив зображень представляється у вигляді низькорозмірних GIST-дескрипторів, на основі яких проводиться первинний пошук потенційно пов’язаних зображень. Потім робиться спроба їх зіставлення з використанням відповідностей високоразмерних дескрипторів локальних особливостей і обчислення епіполярної геометрії. Ті зображення,

які мають найбільшу кількість дескрипторів, які відповідають обмеженням епіполярної геометрії, оголошуються портретними (iconic images) і всі пов’язані з ними зображеннями шикуються в граф. При подальшому розпізнаванні аналізованого зображення ведеться пошук найближчих портретних зображень, виходячи з методу Bag-Of-Words і проводиться тривимірна геометрична верифікація результату.

1.2 Аналіз проблем использование дескрипторів локальних особливо

Отриманий в 90-х роках апарат для представлення зображень у вигляді сукупності дескрипторів локальних особливостей добре зарекомендував себе для вирішення різних завдань комп’ютерного зору. Поділ характеризує інформації на властивості точки інтересу (крім координат в зображенні вони можуть включати в себе характерний масштаб точки, характерну орієнтацію цього фрагмента, величину відгуку функції детектора і так далі) і опис околиці даної точки у вигляді багатовимірного вектора виявилося зручним з точки зору використання і ефективним з точки зору продуктивності.Існує багато методів пошуку точок інтересу в зображенні і методів подання інформації про їх околиці у вигляді дескрипторів, найбільш відомим і універсальним дескриптором можна назвати SIFT. Даний дескриптор, як і більшість інших застосовуваних, оперує інформацією про зображення, представленої в градаціях сірого, тобто не використовує колірні компоненти вихідних зображень. При наявності великої кількості дескрипторів це може призводити до зниження його дискримінаційних властивостей, тому що різні фрагменти зображення здатні формувати схоже уявлення у вигляді багатовимірних векторів, і при їх зіставленні цілком імовірна багатозначність.

Використання значень дескрипторів (багатовимірних векторів) для пошуку найближчих елементів між зображеннями відрізняється для різних

областей застосування. У задачах класифікації поширене використання етапу квантування значень за візуальними словами, що дозволяє кардинальним чином знизити обчислювальну складність вибірки найближчих елементів і спростити модель класифікації. Однак даний підхід накладає обмеження на функцію відстані між дескрипторами, яка в цьому випадку може видавати значення нуль або нескінченність. У завданнях зіставлення зображень (для добування інформації про геометрію, для пошуку примірників, для складання панорам і так далі) необхідно точне значення відстані між значеннями дескрипторів. При цьому існують різні підходи до пошуку відповідностей між наборами дескрипторів: певні підходи використовують значення відповідних багатовимірних векторів поза інформації про точку інтересу, інші пропонують поєднане використання інформації про становище, розмір і орієнтації точки інтересу поряд з перебуванням найближчого значення дескриптора. Кожен з цих підходів має свої плюси і мінуси і не може розглядатися окремо від контексту його використання [3].

Дескриптори локальних особливостей можуть мати різну інваріантність до спотворень. Існують методи, що дозволяють сформувати дескриптори, що наближаються до інваріантності афінних перетворень. Щодо загальноприйнятим підходом до їх вилучення є метод Harris Affine, що спирається на еліптичні регіони отримані за допомогою обчислення матриць моментів другого порядку. Серед інших методів можна назвати так само виділення паралелограмів на основі отриманих за допомогою детектора Кенні країв, отримання еліпсоїдів на основі екстремумів інтенсивності кольору і аналізі радіально розходяться променів, використання регіонів, отриманих з обробленого пороговим фільтром зображення, і тим вимогою, щоб всі пікселі всередині регіону були або темніше, або світліше навколишніх регіонів. В цілому можна сказати, що застосовуються дескриптори локальних особливостей мають інваріантність до афінних і проективних перетворень.

Незважаючи на те, що відомі методи вилучення дескрипторів зазвичай супроводжуються методом пошуку відповідних цікавих точок, не можна сказати, що їх поєднання фіксоване. Різні детектори по-різному показують себе в залежності від наповнення сцени і кількісних параметрів зображення, в той час як різні дескриптори мають відмінні параметри інваріантності щодо різних спотворень зображення. Було показано, що їх комбінування здатне поліпшити показники розпізнавання, однак будь-якої консенсус про ефективний метод комбінування різних підходів відсутня [4].

Розглянемо задачу вилучення тривимірної геометрії об’єктів з набору їх зображень. Це завдання має назву пошуку структури з руху і має набір стійких методів, що дозволяють отримувати інформацію про геометрії об’єктів з якоюсь мірою наближення. Варто відзначити, що розглядається рішення задачі з використанням датчиків, що працюють у видимому діапазоні – існують спеціалізовані апаратні рішення використовують датчики глибини, в цьому випадку застосовуються кардинально відрізняються способи розрахунків. Сучасним прикладом використання систем з застосування датчиків глибини може бути пристрій Google Project Tango, в реальному часі послідовно будує 3D-модель навколишнього світу за допомогою поєднання інформації про глибину, інформації з датчиків акселерометра і гіроскопа, і камери працює у видимому діапазоні.

Найбільш поширений і традиційно використовується метод отримання інформації про геометрію об’єктів за їхніми зображеннями являє собою послідовність алгоритмів, які в цілому можна описати таким чином:

а) пошук стійких опорних точок, які повторюються на зображеннях об’єкта під іншим ракурсом;

б) обчислення дескрипторів областей навколо цих точок;

в) попарне порівняння дескрипторів для пошуку відповідників в зображеннях;

г) побудова моделі на основі епіполярних обмежень, відкидання невірних відповідників і подальша тріангуляція точок в просторі;

д) додавання нових зображень об’єкта для уточнення геометрії, паралельно вирішуючи завдання глобальної оптимізації помилки зворотної проекції отриманої структури для кожного з доданих видів.

Кожен з цих етапів представлений великою кількістю різних підходів зі своїми алгоритмами, особливостями і проблемними питаннями. В контексті даної цієї роботи звернемо увагу на наступні аспекти:

– переважна більшість методів пошуку і опису опорних точок (найчастіше це суміщений алгоритм) оперують із зображенням в градаціях сірого. Ця обставина пов’язана зі специфікою відповідного математичного апарату і моделей для обробки сигналів, так і з простотою їх програмної реалізації. Дослідження з пошуку рішень для обробки кольорових зображень ведуться, проте на даний час пропонуються досить громіздкі рішення, малопридатні для реального використання. В результаті при вирішенні завдання в цілому використовується не вся доступна вихідна інформація (отримана с датчиків фотоапаратів). У даній роботі робиться спроба використання інформації кольорового зображення поряд з відтінками сірого;

– якісний результат алгоритмів пошуку відповідників дескрипторів в парах зображень може погіршитися при збільшенні обсягу вхідних даних. Для ілюстрації – при роботі з зображеннями розмірі 4000 × 4000 пікселів алгоритми можуть видати гірші результати (наприклад, відкиданням відповідностей опорних точок, які могли бути отримані при меншому загальній їх кількості), ніж при роботі з зображеннями розміром 1000 × 1000 пікселів, при тих же настройках алгоритму вилучення опорних точок. Той же ефект може бути викликаний зміною настройки алгоритму вилучення опорних точок, в бік збільшення їх кількості. Це пов’язано зі збільшенням числа опорних точок і їх дескрипторів, і наступним збільшенням неоднозначності при пошуку відповідників дескрипторів, які є по суті хеш-сумами обчисленими в околиці опорних точок, тобто не можуть унікально ідентифікувати їх. Це питання може вирішуватися введенням більш агресивною фільтрації опорних точок (наприклад, за мінімальною

контрастності або максимальної купчастості), або ж підвищенням вимог критеріїв при пошуку відповідників дескрипторів один одному (це призводить до фільтрації, але на більш пізньому етапі), що так чи інакше не дозволяє використовувати весь потенційно доступний спектр дескрипторів.

Окремо, варто згадати, отримує розвиток в останні роки напрямок вилучення структури з руху в масштабі (Large Scale Structure from Motion). З розвитком загальнодоступних сховищ фотографій, потенційно з вбудованою геопозиційному інформацією, стало можливим рішення нових завдань, як наприклад Large Scale Reconstruction – отримання інформації про геометрію в великих масштабах на рівні цілих міст. Для даного напрямку так само актуальні зазначені вище аспекти, оскільки, знаходячи більшу кількість відповідностей дескрипторів між собою, можна добиватися сумірною деталізованості міської геометрії шляхом обробки меншої кількості зображень.

Інше завдання, яка розвинулася в даному напрямку, відноситься до геопозиціонування фотографії на основі наявного набору з відомими координатами – Worldwide Pose Estimation. З одного боку, подібні завдання надзвичайно складні через великий обсяг довідкової інформації (до якої можна віднести самі зображення і їх метадані), з іншого ж боку, можливість попередньої обробки цього масиву інформації дає можливість розширити традиційні методи роботи з дескрипторами опорних точок.

Так, наприклад, пропонується механізм пошуку відповідних дескрипторів аналізованого зображення з масивом дескрипторів наявних в базі зображень з огляду на їх зв’язку з заздалегідь побудованим хмарою тривимірних точок. Знаючи, які дескриптори в зображеннях були використані для формування хмари точок, можна скласти їх пріоритет і на його основі, а не основі рівномірного розподілу випадкових чисел, вибирати базові точки для роботи механізму RANSAC. З іншого боку, наявність підготовленого хмари тривимірних точок з набором відповідних дескрипторів дозволяє проводити зворотний пошук відповідників – шукати

відповідності в зображенні для кожної точки хмари; в цьому випадку мається на увазі обчислення деякого усередненого дескриптора для кожної точки, тому що кожній точці відповідає мінімум два різних дескриптора.

Інший варіант завдання Worldwide Pose Estimation – анотування довільного зображення. Як підходу використовується так само вилучення інформації з доступних колекцій фотографій, зроблених користувачами Інтернету і вільно доступними, наприклад, з сервісу Flickr [5].

Використовуючи, з одного боку, доступні анотації до зображень, і, з іншого боку, EXIF інформацію з геотеггінгом, можливо використовувати цей масив інформації як навчальну вибірку великого розміру. Маючи уявлення цих даних в зв’язковому вигляді, стає можливим проводити автоматичне анотування довільного вхідного зображення з посиланнями, наприклад, до відповідних статей Wikipedia.

Таким чином, завдання Worldwide Pose Estimation так само актуально дослідження способів пошуку найближчих дескрипторів між зображеннями. У розрізі використання доступних в мережі Інтернет зображень, варто згадати протилежність дескрипторів локальних особливостей зображень – глобальні дескриптори зображень (GIST). Незважаючи на очевидні недоліки глобальних дескрипторів (в першу чергу чутливість до геометричних спотворень через застосування довільної сітки сегментування), вони можуть успішно використовуватися для індексування та швидкого пошуку серед мільярдів зображень. На перший план тут виходить їх основна перевага – вони, з огляду на низьку розмірність, вимагають на багато менше пам’яті для зберігання і операцій, ніж множина локальних дескрипторів. Таким же чином глобальні дескриптори зображень можуть застосовуватися для початкової фільтрації навчальної вибірки, яка потім використовується для більш складних конструкцій на основі дескрипторів локальних особливостей.

1.3 Постановка задачі дослідження

Актуальність роботи. Комп’ютерний зір – важливий предмет в рамках емуляції деяких процесів, що відбуваються в свідомості живих істот. Великі вимірювання зображень і потенційно велике число пов’язаних між собою зображень призводять до того, що практично будь-які методи, інтуїтивно здаються простими, не можуть використовуватися без глибокого перегляду з боку можливих оптимізацій, обумовлених великими часовими і ресурсними витратами.

Різні області комп’ютерного зору мають відмінними джерелами складності, для такого важливого завдання, як класифікація зображень, ними певної категорії.

Можна виділити також характерні властиві їй проблеми: навчальна вибірка традиційно будується на основі вручну створених бібліотек; збільшення кількості класів призводить до зниження релевантності результатів; класи не пов’язані один з одним і відсутні методи аналізу семантичної залежності між ними. Спроба автоматизувати отримання навчальної вибірки була зроблена в роботах Fergus R. і Zisserman A., однак пропонований ними метод накладав істотне обмеження на візуальне уявлення об’єктів в плані відносного розташування частин. Таким чином завдання автоматичного отримання навчальної вибірки залишається відкритою.

Протягом останніх років було показано, що зручним поданням зображень для їх аналізу є сукупність векторів, що описують околиці точок інтересу дескрипторів локальних особливостей зображень. Найважливіші результати в розвитку концепції дескрипторів були отримані в роботах C.Harris, T.Lindeberg, C.Schmid, D.Lowe і G.Csurka. На сьогоднішній день дескриптори застосовуються в багатьох напрямках комп’ютерного зору.

Однак існуючі методи їх використання мають також недоліки: не застосовується інформації про контекст через суті дескрипторів як хеш-сум

множин і не використовується колірна інформація зображень, що призводить до зниження дискримінаційних властивостей. В силу компромісною специфіки принципів роботи дескрипторів і необхідності імітувати процеси, що протікають в свідомості живих істот, в даний час не існує розроблених методів остаточного вирішення завдань аналізу зображень із застосуванням дескрипторів.

Таким чином, актуальною є розробка ефективних методів аналізу зображень на основі дескрипторів з використанням більш високорівневою інформації про контекст.

У даній атестаційній роботі увага концентрується на двох аспектах використання дескрипторів: задачі класифікації зображень і задачі вилучення геометрії з наборів зображень.У задачі класифікації зображень увага звертається на проблеми, пов’язані з ручним механізмом формування навчальної вибірки і на проблеми, пов’язані з відсутністю відносин класів між собою. У задачі добування інформації про геометрію об’єктів увага звертається на проблему надлишкової фільтрації відповідностей дескрипторів при пошуку співвідношень зображень між собою.

Об’єктом дослідження є два аспекти використання дескрипторів: задачі класифікації зображень і задачі вилучення геометрії з наборів зображень. У задачі класифікації зображень увага звертається на проблеми, пов’язані з ручним механізмом формування навчальної вибірки і на проблеми, пов’язані з відсутністю відносин класів між собою. У задачі добування інформації про геометрію об’єктів увага звертається на проблему надлишкової фільтрації відповідностей дескрипторів при пошуку співвідношень зображень між собою.

Мета дослідження. Розробка методів поліпшення роботи дескрипторів в задачах пошуку структури з рухів і в задачах класифікації складних зображень шляхом використання інформації про контекст.

Для досягнення поставленої мети в роботі вирішуються наступні завдання:

– дослідження можливості використання автоматично отриманої навчальної вибірки для задач класифікації і порівняльний аналіз різних підходів до класифікації в цьому випадку;

– розробка методу фільтрації пошукової видачі від нерепрезентативних примірників зображень;

– порівняльний аналіз і розробка ефективних алгоритмів пошуку відповідностей між дескрипторами;

– розробка моделі представлення зображень у вигляді сегментів для задач структури з руху.

2 МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ЗОБРАЖЕНЬ НА ОСНОВІ ДЕСКРИПТОРІВ ЛОКАЛЬНИХ ОСОБЛИВОСТЕЙ

2.1 Особливості методів автоматичної побудови навчальної вибірки

У звичайній практиці системи класифікації навчаються на спеціально підготовленій вибірці зображень, розбитою по класах (можливо, із зазначеною ground truth). На сьогоднішній день існують такі бібліотеки, що дозволяє успішно оцінювати продуктивність тих чи інших класифікаторів на однакових наборах зображень. Однак є проблема в тому, що подібні бібліотеки, з одного боку, складаються за допомогою ручної праці, з іншого боку, представлені в них класи не мають зв'язків між собою. Це означає, по-перше, складність в їх підтримці і розширенні, по-друге, складність обробки зображень, що містять екземпляри різних класів.

Існують різні варіанти завдання класифікації і супутні їм вимоги до навчальної вибірки. В цілому можна виділити два типи таких задач – категоризація зображення в цілому і категоризація через локалізацію. Розглянемо задачу категоризації зображення в цілому. Дане завдання формулюється як знаходження класу, при якому функція ймовірності приналежності зображення до класу досягає максимального значення:

$$\hat{C} = \arg\max_C f(I, C), \tag{2.1}$$

де I – аналізоване зображення;

C – клас з наявного набору;

$\hat{C}$ – найбільш ймовірний клас для даного зображення;

f – функція ймовірності приналежності зображення класу;

argmax – значення аргументу при якому вираз досягає максимуму.

Для навчання відповідних класифікаторів потрібно тільки рознесення навчальної вибірки по різних класах. Розглянемо задачу категоризації через

локалізацію (categorization by localization). В цьому випадку потрібно не тільки віднести зображення до певного класу, а й локалізувати екземпляр цього класу на зображенні :

$$\hat{\mathrm{C}} = \arg\max_C f\ (\mathrm{I} \cap \mathrm{s}, \mathrm{C}), \tag{2.2}$$

$$\hat{\mathrm{C}} = \arg\max_C f\ (\overline{I \cap s}, \bar{C}), \tag{2.3}$$

де s – область в зображенні I.

Для можливості працювати з такими умовами завдання вводяться додаткові вимоги до навчальної вибірки – крім безпосередньо мітки класу для кожного зображення на зображеннях виділяється область присутності цього об’єкта (ground truth). У переважній більшості випадків вона представляється прямокутником, так як використання інших геометричних фігур пов’язане з подальшою складністю обробки. Для категоризації через локалізацію проводиться окреме навчання різних класифікаторів – для виділених об’єктів і для навколишнього їх фону. Надалі при аналізі вхідних зображень використовується функція, що враховує значення обох класифікаторів, і досягає максимуму в області передбачуваного розташування примірника певного класу. У найпростішому випадку ця функція може являти собою суму значень класифікаторів. Завдання, що формулюються як «знайти екземпляр класу або їх відсутність», зазвичай використовують саме цей підхід [6].

У даній роботі пропонується переглянути постановку задачі класифікації зображень, і рухатися не від наявних бібліотек зображень, розбитих за класами, а від уявлення людини про реальний світ, представленого у вигляді семантичного графа. У цьому графі пропонується виділяти поняття, що мають досить постійне візуальне уявлення, автоматичним чином проводити пошук відповідних зображень, і при аналізі вхідного зображення враховувати відстань між поняттями в семантичному графі. Для збору вибірки навчальних зображень пропонується

використовувати загальнодоступні пошукові системи. Таким чином, досягається, з одного боку, незалежність від наявності конкретних необхідних класів в наявних бібліотеках зображень і можливість гнучкого підстроювання під задачу, з іншого боку, з використанням пошукових систем можна добитися більшої зв'язності зображень в класах і поданням людини про поняттях (пошукова видача ранжируется виходячи з семантичної релевантності). Так само даний підхід привносить особливості, що стосуються того, що в навчальній вибірці відсутня інформація про розташування об'єктів на зображеннях (виділити їх автоматично не представляється можливим) і, отже, її можна використовувати для відповіді на питання «який з класів представлений на зображенні», але не «якийсь з класів, якщо який-небудь з них». У той же час, використання зв'язків між класами, представленими в навчальній вибірці, дозволяє аналізувати складні зображення, в яких можуть бути представлені екземпляри різних (ймовірно пов'язаних) класів.

Використання пошукових систем для формування навчальної вибірки дає такі позитивні ефекти:

– пошукова видача корелює з реальним поданням людства про візуальному представленні понять, тому що вона ранжируется з урахуванням посилань і цитування;

– набір необхідних класів може бути гнучко змінено відповідно кожну конкретну задачу, або відкоригований в процесі роботи;

– відсутність ручної праці при формуванні вибірки.

Для формування навчальної вибірки був використаний, в даному конкретному випадку, Google Search. Пошуковий сервіс видає набір зображень по даному запиту. Для кожного обраного поняття опціонально задається слово для запиту в пошукову систему, це необхідно через те, що певні слова можуть мати абсолютно різну видачу при різних формах цього слова в запиті. Традиційне кількість навчальних зображень в задачах класифікації може варіюватися в діапазоні від 10 до 30, таким чином,

поточний ліміт Google Custom Search на 100 перших результатів пошукового запиту не привносить обмежень на можливості навчання, що підтверджується результатами отриманих матрицями неточностей (confusion matrix) нижче. Варто також відзначити, що при такому формуванні навчальної вибірки, відсутня інформація про розташування об’єктів усередині зображень (ground truth), і ця вибірці не може бути використана для вирішення завдань categorization by localization [7].

Варто відзначити, що сучасні пошукові системи підтримують вказівку в пошуковому запиті типів зображень, це можуть бути, наприклад, фотографії і малюнок («синтетичні» зображення, або фотографії в значній мірі змінені графічним редактором). Для отримання навчальної вибірки використовуються запити із зазначенням пошуку фотографій, проте досвід даної роботи показав, що кліпарт зображення так само можуть аналізуватися класифікаторами, пропонованими в даній роботі, з прийнятними результатами. При цьому, зрозуміло, необхідно вибирати один з типів зображень, так як їх характеристики значно відрізняються і їх спільне використання в навчальній вибірці може привести до погано навченому класифікатором.

2.2 Аналіз та вибір методів класифікації

Були проаналізовані усталені на сьогоднішній день підходи до використання локальних візуальних особливостей для класифікації зображень, умовно їх можна розбити на деякі групи: Bag Of Words (BoW), Naive Bayesian Nearest Neighbor (NBNN), Part-Based і засновані на сегментації.

Part-based підхід, що враховує взаємне розташування візуальних особливостей, і підходи, засновані на сегментації, є спеціалізованими для

певних завдань і тому не можуть бути використані в загальному завданню класифікації.

Для роботи з навчальною вибіркою, отриманої автоматично, класифікатор повинен володіти, в першу чергу, стійкістю результату роботи при незначних змінах навчальної вибірки – при появі у вибірці «поганих» примірників загальний результат роботи не повинен кардинально погіршуватися. Розглянемо підходи BoW і NNBN с точки зору поставленої задачі.

В методі Bag Of Words всі візуальні особливості (які являють собою багатовимірні вектори – 128-мірні в разі SIFT) з усієї навчальної вибірки об'єднуються в загальний масив, який потім розбивається на задану кількість V візуальних слів. Значення візуальних слів в просторі дескрипторів приймемо за W_v [8]. Нехай візуальні слова w являють собою V-мірний вектор з один компонентом рівним одиниці і іншими рівними нулю:

$$\forall i \in [1,V] \forall v \in [1,V]: w_v^i = \begin{cases} 1, i = v, \\ 0, i \neq v. \end{cases} \quad (2.4)$$

Нехай зображення I представлено N дескрипторами $d = [d_0, d_1, \ldots, d_N]$, в такому випадку воно представимо у вигляді набору візуальних слів:

$$I' = [w_1, w_2, \ldots, w_N], \quad (2.5)$$

де кожне візуальне слово w_i вибирається виходячи з близькості дескрипторів до значень візуальних слів:

$$w_i^j = \begin{cases} 1, & j = \arg\min_n \|d_i - W^n\|, \\ 0, & j \neq \arg\min_n \|d_i - W^n\|. \end{cases} \quad (2.6)$$

Отримані набори візуальних слів використовуються для прийняття рішення про належність аналізованого зображення до певного класу. Для прийняття рішення застосовуються узагальнюючі або дискримінують методи. В якості узагальнюючого методу можна навести Naïve Bayesian класифікатор, а в якості дискримінуючого – класифікатор на основі методу опорних векторів. При використанні дискримінує класифікатора наступному етапі обчислюється представлення зображення у вигляді гістограми розподілу візуальних слів:

$$I'' = \frac{!}{N}\sum_{i=1}^{n} w_i. \tag{2.7}$$

Серед плюсів цієї моделі з точки зору запропонованого методу можна відзначити стійкість її результатів при незначних змінах навчальної вибірки. При цьому суть даного методу добре вписується в підхід з автоматично отриманої навчальної вибіркою. Серед недоліків варто відзначити принципову проблемність тих ситуацій, коли в уже згадуваному зображенні представлено більше одного примірника об’єкта з навчальної вибірки або цей об’єкт представлений частково. Так само серед мінусів можна виділити високу обчислювальну складність на етапі навчання, особливо процес кластеризації вихідного набору візуальних особливостей, і наявність етапу квантування, яке знижує дискримінаційну можливість дескрипторів [9].

Метод NBNN заснований на ідеї, що процес квантування оригінальних візуальних особливостей, застосовуваний в методі Bag of Words, знижує їх дискримінаційну можливість. З іншого боку, робиться припущення, що ймовірність знаходження окремого дескриптора d_i в класі C не залежить від імовірності знаходження інших:

$$P(I \mid C) = P(d_1, \ldots, d_N|C) = \prod_{i=1}^{n} P(d_n|C), \tag{2.8}$$

де I – аналізоване зображення, представлене дескрипторами d_i;

P (I | C) – ймовірність приналежності зображення *I* класу *C*.

У даному підході всі візуальні особливості навчальної вибірки для кожного класу об’єднуються в загальний масив. При аналізі вхідного зображень для кожної його візуальної особливості di шукається найбільш близька особливість серед кожного з класів *C*:

$$\forall d_i \forall C: NNc(d_i) = D_j^c, j = arg\ min \| D_n^C - d_i \|, \quad (2.9)$$

де D_j^c – дескриптори класу *C*.

Найбільш близьким класом вважається той, для якого сума відстаней буде мінімальна:

$$\hat{C} = arg\ min \sum_{i=1}^{n} \| d_i - NN_C(d_i) \|^2. \quad (2.10)$$

Плюсами даної моделі є відсутність етапу навчання і здатність працювати з об’єктами, представленими частково. До мінусів можна віднести нестійкість результатів роботи щодо змін навчальної вибірки – навіть незначна зміна вибірки для одного класу з «поганими» даними здатне кардинально погіршити загальну роботу системи (що пов’язано з відсутністю етапу квантування). Так само мінусом можна назвати підвищені вимоги до однотипності навчальної вибірки між різними класами – класи, для яких характерно більшу кількість видобутих візуальних особливостей, матимуть більші значення класифікатора в силу статистичних законів. Практика даної роботи показала, що ці мінуси роблять непридатним NBNN для роботи з автоматично отриманої навчальної вибіркою [10].

На основі мінливості навчальної вибірки і її неоднорідності (що випливає з використання пошукових систем), а також спираючись на практичні результати експериментів, було прийнято рішення для класифікації використовувати уявлення зображень у вигляді візуальних слів.

Після того, як для кожного з класів зібрана навчальна вибірка, в ній виділяється певна кількість зображень для подання класу в процесі кластеризації. В рамках даної роботи хороші результати були досягнуті при 15 зображеннях на один клас: зменшення цієї кількості веде до неоптимальному розташуванню центрів кластерів і зниження репрезентативності згодом привласнених візуальних слів, збільшення ж веде до зростання складності кластеризації, яка може займати значний час. Для запобігання появи надмірної кількості візуальних особливостей, зображення розміром понад 2048×2048 пропорційно зменшуються до 1024 по найбільшій стороні. Візуальні особливості зображень представлені SIFT-дескрипторами. Витягнуті з усіх обраних зображень всіх класів дескриптори, представлені 128-мірними векторами, потім кластеризуються по 900 кластерам за допомогою методу *K*-Means. Поведінка класифікатора SVM в залежності від кількості візуальних слів при 12 класах наведено на рисунку 2.1. Збільшення числа кластерів призводить до поліпшення дискримінаційних властивостей гістограм (гістограми мають кількість елементів дорівнює кількості кластерів), проте проблемно з точки зору продуктивності.

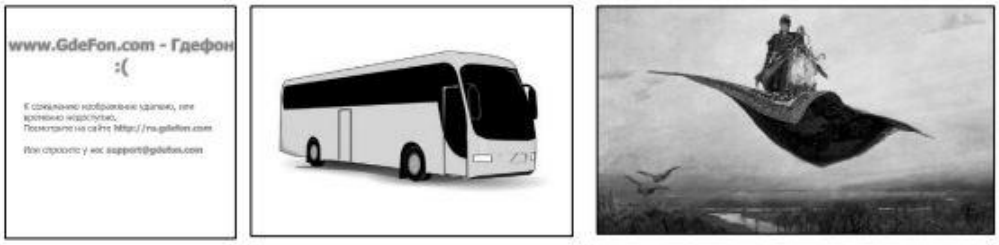


Рисунок 2.1 – Приклади викидів при формуванні навчальної вибірки з використанням пошукової системи

Для уявлення класів в процесі навчання класифікаторів вибирається певна кількість зображень з цих класів. Це кількість впливає на якість роботи класифікатора SVM. При цьому для обчислення характеристики confusion

matrix використовуються зображення класів, що не вибиралися для навчання класифікаторів – це дозволяє отримати незалежну оцінку продуктивності. Матриця *U*, що представляє confusion matrix, обчислюється таким чином:

$$\forall i \in [1,n] \forall j \in [1,n]: U_i^j = \sum_{I \in C_i} t(j,I), t(j,i) = \begin{cases} 1, f_j(I) = max_k f_k(I) \\ 0, f_j(I) \neq max_k f_k(I) \end{cases}, \quad (2.11)$$

де n – кількість класів;

i – рядки представляють зображення відповідних класів Ci;

j – стовпці представляють класифікатори;

f_i – (навчені для відповідних класів Cj);

I – зображення класів;

$\max_k f_k(I)$ – максимальне значення з усіх класифікаторів для зображення *I*.

Середній відсоток R вірності класифікації в даному випадку обчислюється таким чином:

$$R = \frac{1}{n} \sum_{j=1}^{n} \frac{U_j^i}{\sum_{j=1}^{n} U_i^j}. \quad (2.12)$$

Виходячи з отриманих результатів, було прийнято рішення навчати класифікатори на кількості зображень рівним 50.

2.3 Класифікація на основі моделі Bag-of-Words

В рамках даної роботи були протестовані два класифікатора, що використовують уявлення зображень у вигляді візуальних слів – наївний баєсів класифікатор і класифікатор на основі методу опорних векторів. Опис їх принципів роботи наведені нижче.

Наївний баєсів класифікатор являє собою найпростіший спосіб використання моделі BoW для прийняття рішення про належність зображення одного з класів. Він заснований на допущенні, що кожен клас має власне, незалежне від конкретних зображень, розподіл візуальних слів, і воно помітно відрізняється між класами. Таким чином, з'являється можливість розраховувати ймовірність знаходження окремих візуальних слів в різних класах незалежно один від одного. Шлях *N(t, I)* – кількість візуальних слів *t* в зображенні *I*.

Таким чином, для зображень I_i складових класу *C*, отримуємо ймовірність *P* належності до класу, виходячи з містяться в них візуальних словах:

$$P(C|I_i)\ \alpha\ P(C)P(I_i|C) = P(C)\ \prod_{t=1}^{V} P(w_t|C)^{N(t,I_i)}. \qquad (2.13)$$

Імовірність знаходження окремого візуального слова w_t в класі *C* обчислюється з використанням розмиття по Лапласа (additive smoothing), для запобігання появи нульових ймовірностей в добутку.

Найбільш близьким класом для зображення *I* вважається той, для якого більше спільна ймовірність знаходження візуальних слів:

$$C = \ arg\ max_C \prod_{t=1}^{V} P(w_{t|C})^{N(t,l)}. \qquad (2.14)$$

Метод опорних векторів (SVM) застосовується для поділу даних двох класів з максимальним зазором між ними. Зазор при цьому визначається як мінімальна відстань від розділяє дані гиперплоскости до найближчої вибірки [11]. Таким чином, для вибірки *x*, зазначеної класами *C*, приймають значення ± 1, функція класифікації виглядає наступним чином :

$$f\ (x)\ =\ sign\ (w^T x\ +\ b), \qquad (2.15)$$

де w, b – параметри відповідної гиперплоскости.

Для використання даного методу зображення з навчальної вибірки перетворюються в гістограми розподілу візуальних слів, виходячи з кількості візуальних слів, що зустрічаються в зображенні. Так як в даному випадку вибірка класів представлена багатовимірними векторами (розмірність визначається кількістю візуальних слів), ці дані не завжди є лінійно нероздільні. SVM вирішує цю проблему за допомогою двох підходів:

1) вводиться допустима міра помилки, що залежить від відстані вектора до розділяє гиперплоскости (м'який зазор між точками);

2) знаходиться таке відображення Φ, що переводить вибірку x в простір з більшої розмірністю. Застосовується нелінійна функція ядра, наприклад радіальна базисна функція.

Так як в завданні класифікації зображень є більше двох класів, використовується класифікація «цей клас або всі інші». Для m класів навчається m класифікаторів SVM, кожен з яких розрізняє зображення класу і від зображень інших m-1 класів [12].

За підсумками роботи описаних вище класифікаторів на реальних даних, отриманих автоматичним чином, було прийнято рішення використовувати класифікатор на основі методу опорних векторів, як показав найкращі результати на тесті confusion matrix.

2.4 Аналіз результатів автоматичної побудови навчальної вибірки

У зображеннях, отриманих в результаті роботи пошукових систем за запитами, з великою часткою ймовірності будуть присутні екземпляри незв'язані з основною масою – викиди. Приклади таких зображень наведені на рисунку 2.1. Причин цьому може бути декілька:

– семантична багатозначність пошукового запиту, наприклад, викликана лексичної багатозначністю;

– значна внутриклассовая візуальна мінливість – в цьому випадку візуальне уявлення поняття може приймати характеристично далекі значення;

– відмінності, викликані зміни ставленням об’єкта класу і фону (контексту, в якому він зображений). Фотографії, на якому об’єкт зображений з віддаленим фоном, або з фоном, що несе мінімум інформації, будуть значно відрізнятися в термінах моделі BoW;

– синтетичні зображення (спочатку представлені у векторному вигляді) в певних випадках можуть не бути так визначені пошуковою системою і потрапити в видачу поряд з фотографіями. Витяг дескрипторів і наступні операції не розраховані на роботу з синтетичними зображеннями і видаватимуть значно відрізняються результати;

– помилки технічного роду – випадки, коли сервери за запитом видають неправильне зображення, або зображення-заглушку (яке видається в разі неможливості надати запитувану зображення) [13].

Наведені як приклад викиди мають різні причини появи. Перше зображення було отримано внаслідок помилки на стороні веб, і являє собою заглушку, позбавлену сенсу в рамках візуального представлення класу. Друге зображення було отримано внаслідок помилки поділу зображень на стороні пошукової системи – в даному випадку воно було розцінено як фотографія, хоча по факту є синтезованим векторних зображенням. Третє зображення являє собою приклад семантичної багатозначності запиту, і, очевидно, це зображення погано характеризує візуальний клас «літак».

Як показала практика даної роботи, автоматичним чином, без наявності еталонного набору зображень, можна фільтрувати тільки синтетичні зображення і зображення, які є наслідком будь-якої помилки.

У даній роботі для фільтрації такого роду викидів використовується представлення зображень в моделі BoW, обчислюється середнє арифметичне значення гістограм, міра розкиду через середнє квадратичне відхилення і ставлення відстані кожного зображення до міри розкиду. Візуальні слова при

цьому витягуються для різних класів незалежно, це дозволяє домогтися більш рівномірного їх розподілу всередині класу. Кількість візуальних слів для цього завдання було знайдено достатнім рівною кількості зображень (близько ста зображень в рамках обмежень пошукових систем), збільшення цієї кількості призводить до збільшення значень разбросов, проте відносини між зображеннями при цьому залишаються практично незмінними.Зменшення ж кількості візуальних слів призводить до зменшення розкиду і до неможливості виділення викидів на цій підставі.

Нехай *n* зображень представлені гістограмами h^j розподілу *k* візуальних слів. В такому випадку середнє арифметичне значення обчислюється таким чином:

$$\forall i \in [1,k]: \overline{h_i} = \frac{1}{n}\sum_{j=1}^{n} h_i^j, \qquad (2.16)$$

де h_i^j – елемент гістограми h^j показує кількість візуальних слів *i*.

Середньоквадратичне відхилення для класу обчислюється з використанням міри відстані L_2 між середнім арифметичним і окремими зображеннями:

2.5 Особливості методів семантичної корекції у задачах класіфікації

Живі істоти, при вирішенні завдання віднесення того чи іншого зображення до певного класу (ідентифікації в широкому розумінні), спираються на їх наявний досвід і уявлення про навколишній світ. На відміну від механічно навчених систем класифікації, живі істоти співвідносять вивчення візуальних властивостей об’єкта або оточення с іншими властивостями, наприклад з категоризацією за різними осях (небезпека, приналежність до одного виду і т.д.). Людина при вивченні навколишнього

світу будує складні асоціативні зв’язки, що дозволяють отримувати цілий спектр пов’язаних абстрактних властивостей і характеристик. Використання цих пов’язаних знань дозволять уточнювати інформацію про спостережуваний об’єкт в разі недостатньої або надлишкової візуальної інформації, таким чином ми можемо ідентифікувати об’єкт і сцену по-різному в залежності від контексту і інших змінюються характеристик.

Таким чином, для наближення аналізу зображень автоматичним шляхом за допомогою навчених класифікаторів до того процесу, який відбувається в свідомості людини, необхідно поряд з даними класифікаторами використовувати деяку інформацію про розташування і взаємне відношення категорій в поданні знань про світ. Така інформація повинна бути представлена у вигляді структурованої бази даних з можливістю вилучення з неї необхідних відомостей. Наближено база даних може бути представлена у вигляді графа, вершини якого є абстрактними поняттями, відображеними в природній мові, а ребрами – відносини між поняттями. Поповнення такої бази знань відбувається поступово, відображаючи збільшення знань про навколишній світ. Зрозуміло, таке уявлення тільки щодо відображає ту множину асоціацій, якими користується людина [14].

В процесі візуального розпізнавання об’єктів людина використовує певний обсяг знань про навколишній світ – «модель світу». Узагальнена форма уявлень про світ може бути з тим або іншим ступенем достовірності витягнута з тлумачних словників природної мови, словників синонімів і тому подібних. У розпізнаванні складних об’єктів велику роль відіграють асоціативні зв’язки між об’єктами, однією з форм яких є контекстні асоціації, коли два об’єкти зустрічаються разом в одному контексті. Ідентифікація об’єкта може бути заснована на аналізі фрагментів, з яких складається цей об’єкт. В даному контексті під об’єктом або фрагментом ми розуміємо деякий семантичне поняття реального світу. Виникає необхідність в

універсальному інструменті представлення знань про світ, що дозволяє описувати відносини між об’єктами.

В рамках даної роботи була використана база даних WordNet, що розробляється Прінстонським Університетом, що представляє подібні відомості про природній мові – англійській. Його структура являє собою два взаємопов’язаних графа – граф слів і граф понять (synsets), при цьому поняттю може відповідати множині слів на природній мові. Для наочності в таблиці 2.1 наводиться приклад знаходження слова «house» в різних поняттях і відповідних наборах слів.

Таблиця 2.1 – Приклад використання слова в базі WordNet

house	a dwelling that serves as living quarters for one or more families; "he has a house on Cape Cod"; "she felt she had to get out of the house"
firm, house, business_firm	the members of a business organization that owns or operates one or more establishments; "he worked for a brokerage house"
house	the members of a religious community living together
house	the audience gathered together in a theatre or cinema;
house	an official assembly having legislative powers; "a bicameral legislature has two houses"

У термінології WordNet поняття може складатися як з окремих слів, так і з фразеологічних сполучень (collocations).Подібна деталізація хороша для формального включення всієї інформації про мову в базу даних, але в той же час призводить до складності застосування будь-яких метрик з цим графом – інформація виявляється надлишкової. Як приклад – відстані (виміряні кількістю проміжних кроків) між поняттями в графі можуть не збігатися з інтуїтивним уявленням про них.

На підставі цього був зроблений висновок про те, що графи, що містять надлишкову лінгвістичну інформацію, як наприклад граф WordNet, погано підходять для формування семантичних зв'язків між поняттями, що мають візуальне уявлення, і наступним визначенням контексту. У даній роботі використовується граф, заснований на автоматично побудованому на основі лінгвістичних словників. Даний граф зручний тим, що являє концентровану інформацію про мову, отриману з короткого тлумачного словника і таким чином позбавлений надмірності.

Введемо в розгляд семантичний граф $G = (V, U)$, побудований на основі тлумачного словника і словника синонімів, де V – множина всіх понять мови, U – зв'язку між словами, відповідні відносинам визначення, синонімії та асоціації. Ставлення визначення пов'язує слово з узагальнюючим його поняттям. Асоціативні зв'язки в графі G є ознакою використання відповідних понять в одній словникової статті. Зв'язки синонімії є точним відображенням визначень зі словника синонімів, проте в даній роботі вони розглядаються як контекстні асоціативні зв'язки. Контекстні зв'язки між об'єктами будемо використовувати в якості окремих етапів при структурному аналізі складного об'єкта. При такому підході будь-яке просте або складне ставлення розглядається як реалізація узагальнених відносин, заданого на частково впорядкованій множині понять. Очевидно, що тим далі один від одного в графі G знаходяться слова, тим менше вони пов'язані, а, отже, і менш пов'язані між собою розпізнані фрагменти зображення [15].

Граф G – зважений, кожна асоціативний зв'язок і зв'язок визначення має вагу. Це означає, що можна розглянути деяку околицю $\alpha(X)$ поняття X, в якій відстань від X до будь-якої вершини множини $\alpha(X)$ не перевищує заданої величини. Ваги ребер були обрані автором виходячи з таких міркувань:

а) якщо вага зв'язків визначення перевищує вагу зв'язків асоціації, то околиця $\alpha(X)$ кожного поняття повинна містити більше узагальнюючих термінів.

б) якщо більшу вагу мають асоціативні зв’язки, то в $\alpha(X)$ виявляються контекстні асоціації, що в повній мірі відповідає завданню виділення семантично пов’язаних об’єктів на зображенні.

У багатьох випадках поняття, відбите на природній мові, має стійке візуальне уявлення, властивості якого можуть бути описані класифікаторами. Прикладами таких понять можуть бути види тварин («кішка»), технічних пристосувань («автобус»), елементів побуту («стіл») і так далі. Для них спільним буде наявність схожих візуальних особливостей, що відносяться до екземплярів даного поняття в реальному світі. У набагато меншому ступені це відноситься до тих понять, які описують будь-який процес, якісне уявлення, або до абстрактних понять: вони не можуть мати постійного візуального представлення, прикладом можуть служити такі поняття, як «світанок», «краса» або «тепло». Якщо враховувати те, що візуальна класифікація зображень в рамках даної роботи проводиться на основі наявного набору «еталонних» зображень розділених по класах і представляють певні візуальні особливості (features) відповідних класів, природним чином з’являється вимога того, щоб цей клас представлявся візуально схожими зображеннями, інакше втрачається можливість навчання класифікаторів. Відповідно з’являється вимога вибору з семантичного графа тільки тих понять, які мають стійке візуальне уявлення [16].

З огляду на ті результати, які були отримані у другому розділі (формування навчальної вибірки з використанням пошукових систем), з’являється можливість використовувати вузли описаного вище семантичного графа як джерело пошукових запитів. Отриману таким чином вибірку пропонується використовувати для навчання класифікаторів, які будуть мати додатково семантичне навантаження в термінах розташування поняття в семантичному графі. Таким чином вирішується проблема відсутності зв’язків між класифікаторами і з’являється можливість подальшої обробки результатів класифікації аналізованого зображення.

В окремих випадках може мати попит використання стійких словосполучень в семантичному графі, спочатку представленому окремими словами. Для цього пропонується додавання в граф стійких понять включають ці слова і встановлення семантичного зв’язку з високим ступенем близькості з базовим поняттям. Стійкі словосполучення зручні для уточнення візуального представлення, хорошим прикладом може служити поняття «міст» і семантично близькі стійкі словосполучення: «пішохідний міст», «автомобільний міст», «залізничний міст» і так далі. Приклади пошукової видачі при використанні стійких словосполучень в даному випадку наведені на рисунку 2.2.

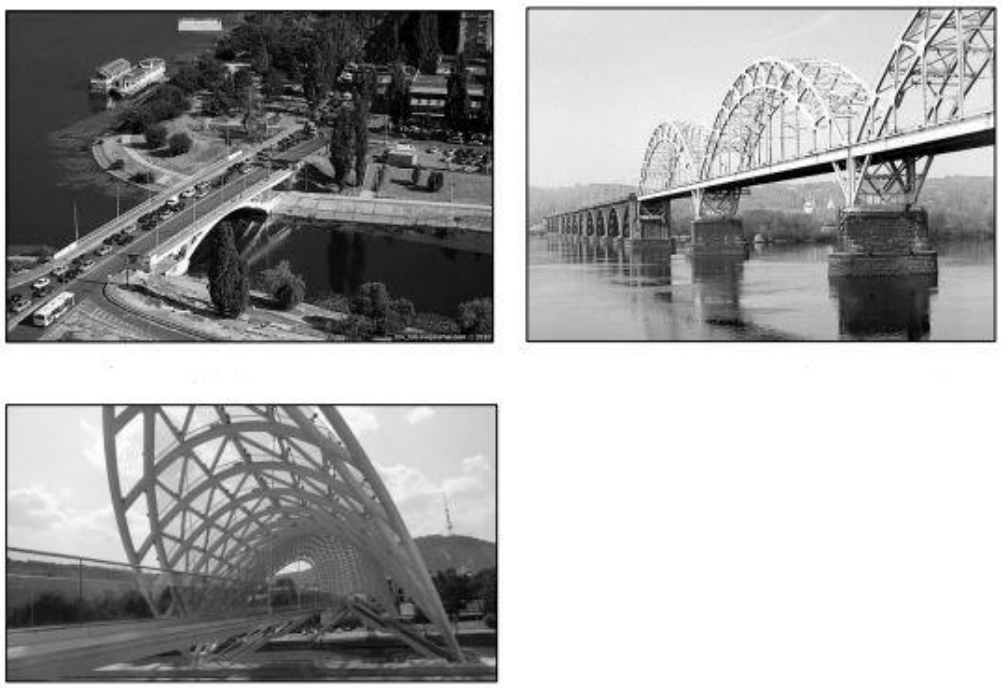


Рисунок 2.2 – Пошукова видача при використанні стійких словосполучень

При класифікації зображень такого візуально різноманітного класу можуть виникати проблеми з навчанням при неуточненої вибірці. Приклад такої поведінки наведено при класифікації зображення в даному випадку після навчання класифікаторів без використання стійких словосполучень, що отримується результат містить помилку. При формуванні ж навчальної

вибірки з використанням стійкого словосполучення «пішохідний міст» результати роботи класифікаторів більш релевантні.

Введемо поняття візуальної схожості класів для відображення зв'язку між відстаню в термінах семантичного графа і візуальними відмінностями між класами, представленими наборами зображень. Приклади зображень даних класів, отриманих в результаті роботи пошукової системи. Зображення класів представлені у вигляді набору візуальних слів – Bag of Words. Таким чином, на першому етапі візуальні особливості всіх зображень кластеризуються, і конкретні візуальні особливості замінюються на порядкові номери найближчих візуальних слів (в прикладі нижче, виходячи з малої кількості класів, число різних візуальних слів було прийнято рівним 100, як достатню для вираження відмінності між зображеннями) [17].

За аналогією з методом класифікації Naive Bayesian Nearest Neighbor введемо поняття візуального відстані $L_V(I, C)$ від окремого зображення I до набору зображень C як L_2 відстань до найближчого зображення в класі:

$$L_v(I, C) = \min_{J \in C} \sqrt{\sum_{i=1}^{k} ||h_i^j - h_i^I||^2}, \tag{2.17}$$

де h – нормовані гістограми розподілу k візуальних слів у відповідних зображеннях.

Приймемо за відстань $L_C(C_1, C_2)$ від візуального класу C_1 до класу C_2 суму візуальних відстаней до цього класу від зображень, що складають перший клас:

$$L_c\ (C_1, C_2) = \sum_{i=1}^{n} L_V(I_{C_1}^i, C_2), \tag{2.18}$$

де n – кількість зображень в класі C_1;

$I_{C_1}^i$ – зображення представляють клас C_1.

Варто зазначити, що через використання мінімального значення відстані між гістограмами, ця функція не симетричної в загальному випадку:

$$L_C(C_1, C_2) \neq L_C(C_2, C_1). \quad (2.19)$$

Візуальна відстань між класами, розрахована з використанням цього підходу та отриманий результат повністю відповідає інтуїтивному уявленню про їх взаємне розташування. Рядки лінійно нормовані в діапазоні [0,1].

Визначимо семантичні відстань між класами як величину, зворотний близькості між відповідними поняттями в семантичному графі. Для знаходження близькості між поняттями використовується обхід графа з знаходженням найкоротшого шляху між ними, враховуючи ваги ребер і коефіцієнт демпфірування при переході між поняттями. Значення близькості S чергового суміжного поняття b до попереднього a обчислюється за формулою:

$$S_b = S_a E\,(a, b)\, D, \quad (2.20)$$

де E – функція подібності при переході від одного поняття до іншого – виходить зі словника понять;

D – константних коефіцієнт демпфірування при кожному черговому видаленні від початкового поняття.

Таким чином, семантичне відстань між класами представляється послідовним твором заходів близькості понять і коефіцієнта демпфірування:

$$L_s(C_1, C_2) = \prod E\,(i, j)D\,, \quad (2.21)$$

де i, j – пари суміжних понять по найкоротшому шляху від поняття, відповідного класу C_1 до поняття, відповідного класу C_2.

Семантична близькість класів, розрахована таким способом, при коефіцієнті демпфірування D = 0,8.

Семантична близькість класів LS знаходиться в діапазоні [0,1], тому для зображення співвідношення з візуальним відстанню між класами, вводиться наступне нормування:

$$L'_S(C_1, C_2) = \frac{1 - L_S(C_1, C_2)}{\max(1 - L_S(\text{Ci}, \text{Cj}))}, j \in [1, n]. \qquad (2.22)$$

Таким чином, протягом поняття отримають відстань, рівну нулю, а найбільш віддалені – одиниці. Варто відзначити, що це співвідношення візуального і семантичного відстані багато в чому залежить від вхідних даних, що представляють обрані класи – з одного боку конкретні набори зображень, з іншого боку – структура використовуваного семантичного графа.

Зрозуміло, було б неправильно стверджувати, що поняття з семантичного графа, мають стійке візуальне уявлення, в кожному з випадків будуть мати схожі пропорції візуальних відстаней на пропорції семантичних відстаней. Подібна зв'язок візуального і семантичного відстані має місце на підставі, що семантичний граф відображає уявлення людини про світ, в якому в свою чергу закладена в тому числі різниця між поняттями в їх візуальному представленні [18].

На підготовчому етапі система навчається класифікації. Після завантаження семантичного графа, з нього виробляється вибірка понять на основі поточних вимог – ця інформація надається користувачем. У найпростішому випадку слова на природній мові, що представляють поняття в семантичному графі, можуть бути безпосередньо використані в якості пошукового запиту (це справедливо в більшості випадків). В окремому ряді випадків може знадобитися примусове вказівку пошукового запиту відмінним від уявлення поняття природною мовою. Причиною цього є те, що

сучасні пошукові системи реагують на форму слова в запиті навіть в разі пошуку зображень. Так, наприклад, слово «хмара» матиме набагато більш релевантну між класами комп'ютерному зору видачу від запиту «хмари», для якого навчання класифікатора буде більш осмисленим при фотографіях, на яких відображено небо в хмарах, а не окремі екземпляри. На основі відповідних запитів далі формується навчальна вибірка [19].

У процесі аналізу чергового зображення проводиться його представлення у вигляді гістограми розподілу візуальних слів. На основі отриманої гістограми і навчених класифікаторів відповідних класів проводиться незалежна візуальна класифікація (докладно розглянута у другому розділі). З огляду на взаємне розташування понять, відповідних візуальним класів, в семантичному графі, проводиться додаткова обробка отриманих результатів. З одного боку, проводиться верифікація результатів цієї класифікації, для спроби виявлення їх неузгодженості – вона може відбуватися через помилки роботи класифікаторів, або ж у разі відсутності сенсу класифікації зображення при наявному наборі класів (сюди ж включаються «безглузді» зображення). З іншого боку, в ряді випадків можливе отримання інформації про контекст при наявності двох і більше близьких понять, представлених на зображенні, що використовується для корекції результатів незалежної візуальної класифікації виходячи зі структури семантичного графа і взаємного розташування цих понять в ньому.

2.6 Аналіз результатів семантичної корекції у задачах класифікації

При наявності семантичних зв'язків між візуальними класами, між якими йде класифікація, з'являється можливість перевіряти «узгодженість» отриманих результатів. В окремих випадках класифікатори можуть видавати помилкові результати, що випливає з принципів їх функціонування, і можливість якимось чином сигналізувати про потенційний виникненні такої

ситуації була б корисна. Основою для такої верифікації може служити узгодженість результатів незалежної візуальної класифікації зі взаємними розташуванням відповідних понять в семантичному графі [20]. Позначимо результати незалежної класифікації за r^i:

$$\forall i \in [1, n]: r^i = f(I, C_i), \tag{2.23}$$

де f – функція класифікації;

I – аналізоване зображення;

C_i – клас з наявного набору;

n – кількість класів.

Таким чином, найбільш близький клас x вибирається як той, для якого класифікатор видав найбільше значення:

$$x = arg\max_i r^i \in [1, n]. \tag{2.24}$$

В багатьох випадках можна стверджувати, що семантично найближчий клас до класу x за підсумками візуальної класифікації не повинен отримувати найменше значення. Порушення цього затвердження може свідчити про помилку класифікатора [21].

Нехай y – самий семантично близький клас до x:

$$y = arg\, max L_S(C_x, C_i), i \in [1, n], i \neq x, \tag{2.25}$$

де L_S – функція семантичного подібності між поняттями відповідають класам.

Тоді твердження можна представити таким чином:

$$y \neq arg\ \min_l r^i, i \in [1, n]. \tag{2.26}$$

У процесі класифікації, при наявності високих результатів у класів, відповідних семантично близьким поняттям, можна говорити про виникає при цьому контексті. Контекст дозволяє посилювати результат цих понять, виходячи з тієї передумови, що близькі поняття при одночасному перебуванні мають велику важливість ніж окремі незалежні поняття. При аналізі складних з точки зору наповнення зображень можуть виникати ситуації коли класифікатори різних класів видають приблизно однакові значення. При цьому порівнянні значення можуть видавати як класифікатори класів, реально присутніх на зображенні, так і класифікатори, помилково видають високий результат. При наявності відомих семантичних зв'язках між класами подібні ситуації можна окремо обробляти, підвищуючи загальну релевантність видачі системи аналізу. Основна теза, на якому ґрунтується подальша логіка роботи, полягає в тому, що при наявності двох і більше семантично близьких класів з високою видачею їх класифікаторів, ці класи можна об'єднати в кластер з більш високим значенням, ніж у класів окремо.

Крім безпосередньої семантичної близькості між поняттями, для виділення контексту важлива їх околиця і ступінь її перетину, так як вона враховує структуру відносин між поняттями. Множина при цьому формується виходячи з семантичної близькості до понять. Перейдемо до визначення ступеня близькості між класами при аналізі конкретного зображення [22].

Нехай r_i – результат незалежної класифікації.

Введемо функцію $F(C_i)$ – ступінь подібності класу C_i відповідному поняттю s_i в семантичному графі, що залежить від r_i.

Зручним механізмом для визначення ступеня близькості класів з урахуванням їх околиці є обхід семантичного графа в ширину від кожного конкретного класу. Обхід в ширину враховує ваги ребер (в даній реалізації найкращий результат показали значення 0,3 для відносин визначення і 0,75 для асоціативних зв'язків), а також відстань від класу, для чого застосовується коефіцієнт демпфірування при проходженні чергового

поняття. При проходженні кожного чергового поняття розраховується його функція подібності з оригінальним поняттям-класом і, коли вона стає менше порогового значення, обхід в цьому напрямку завершується. Для обходу використовується алгоритм Дейкстри. Значення близькості S чергового суміжного поняття b до попереднього a обчислюється за формулою:

$$S_b = S_a E(a,b) D, S_a \geq T, \tag{2.27}$$

де E – функція подібності при переході від одного поняття до іншого – виходить зі словника понять;

D – константний коефіцієнт демпфірування при кожному черговому видаленні від початкового поняття;

T – порогове значення близькості до початкового поняття.

Значення ступеня подібності поняття x є фіксованим щодо класу значенням: при первинному обході воно приймається за одиницю, при обходах з урахуванням результату роботи візуального класифікатора, воно може бути відповідним чином скоригована [23]. Таким чином, візуальне відповідність транслюється в обхід семантичного графа.

Нехай $x \in [0.5; 1.0]$.

Позначимо множину S_i^C близькості понять до поняття, відповідного класу C:

$$S_i^C, i \in [1; N],$$

де N – кількість понять в словнику.

Введемо функцію ширини класу для заданої міри подібності:

$$W(C,x) = \sum_{i=1}^{N} s_i^C, \tag{2.28}$$

де S_i^C – розраховані при обході графа з поняття, відповідного класу C і подібності x.

Ширина класу буде служити основою для порівняння в термінах семантичних зв'язків.

Так як в семантичному графі поняття розташовуються з нерівномірною щільністю, потрібно система нормування для кожного окремого класу, інакше порівняння величин не матиме сенсу. Базою нормалізації прийнята ширина класу, отримана при обході зі значенням ступеня подібності поняття рівним одиниці, тобто $W(C,1)$.

Для отримання порівнянних величин в термінах семантичного графа, на основі ширини класів і з огляду на отриманий раніше результат незалежної візуальної класифікації $r_{\text{ш}}$, використовується наступне перетворення:

$$Qc = \frac{W(C,1-\frac{r_c-r_{min}}{r_{max}-r_{min}}0.5)}{\text{W(C,1),}}, \qquad (2.29)$$

де r_c – результат роботи класифікатора класу C;

r_{min} – мінімальний результат роботи всіх класифікаторів;

r_{max} – максимальний результат роботи всіх класифікаторів.

Таким чином, найбільш близький, виходячи з результатів незалежної візуальної класифікації, клас отримає ступінь подібності рівну 1,0, а найменш близький отримає ступінь подібності рівну 0,5.

Об'єднання пересічних понять в кластери На складних зображеннях можуть бути присутніми екземпляри різних семантично пов'язаних класів. При цьому класифікатори працюють незалежно, і кожен з них визначає наявність об'єктів свого класу на зображенні [24]. Нехай ймовірність появи об'єктів відповідних класів $P(C_1)$ і $P(C_2)$ пропорційна відповідним результатами незалежної класифікації r_{C_1} і r_{C_2}. Тоді при наявності пересічної

околиці, тобто при $P(C_1 \cap C_2) > 0$, ймовірність появи на зображенні суміщеного об'єкта з двох класів повинна бути вище ймовірностей окремо:

$$\mathrm{P}(C_1 \cup C_2) > \max(P(C_1), P(C_2)). \quad (2.30)$$

У процесі об'єднання понять проводиться порівняння класів «кожен з кожним» і обчислюється ступінь перетину на основі семантичних зв'язків і початкової видачі класифікаторів для цих класів [25]. При цьому йде перехід від окремих класів до кластерів класів, спочатку складаються з одних класів, потім, можливо, укрупнюються за рахунок об'єднання. Для визначення 66 необхідності об'єднання класів вводиться таку вимогу, при виконанні якого запускається механізм їх об'єднання:

$$\frac{\sum_{i=1}^{N} \sqrt{S_i^{C1} S_i^{C2}}}{W(C_1, x_{C_1}) + W(C_2, x_{C_2})} > 0{,}06, \quad (2.31)$$

де S_i^C – значення семантичної близькості понять до поняття класу *C*;

$W(C, x)$ – ширина класу *C* розрахована при ступеня подібності *x*.

При об'єднанні значеннями близькості понять S_i^C в новому кластері $C = (C_1 \cup C_2)$ приймаються набільші значення з оригінальних значень близькості:

$$S_i^C = \max\left(S_i^{C_1}, S_i^{C_2}\right). \quad (2.32)$$

Значення нормованої ширини Q_C об'єднаного кластера *C*, отриманого з C_1 і C_2, обчислюється за формулою:

$$Q_C = \max(Q_{C_1}, Q_{C_2}) \frac{1 + max(Q_{C_1}, Q_{C_2} min(Q_{C_1}, Q_{C_2}))}{max(Q_{C_1}, Q_{C_2})^3 + min(Q_{C_1}, Q_{C_2})^3}. \quad (2.33)$$

Таким чином досягається, з одного боку, плавне збільшення видачі для зв’язкових семантично понять з високою видачею, з іншого боку, коефіцієнти підібрані таким чином, щоб зменшити ймовірність виникнення ситуації, коли близькі семантично поняття з невисокою видачею , в результаті отримують значення вище спочатку правильно класифікованих. Після об’єднання двох понять в кластер (одне з них вже може бути кластером, отриманим на попередньому етапі – для алгоритму це не має різниці), одне з понять видаляється і процес триває до тих пір, поки існують класи зі ступенем перетину вище порогового значення.

По закінченню процесу об’єднань, вибирається кластер *C* з максимальним значенням нормованої ширини *Qc*, і він використовується в якості фінального результату класифікації [26].

Приклад роботи можна продемонструвати на зображенні, представленому на рисунку 2.3.

Рисунок 2.3 – Приклад класифікованого зображення

Найбільш близьким в даному випадку є клас «міст» (далі «автомобіль», «двері» і «вікно»), що помилково. Так як в даному складному зображенні є різні класи, пов’язані семантично, з’являється можливість скорегувати результати класифікації виходячи із взаємної близькості класів. Після роботи механізму обходу семантичного графа, нормалізації щодо ширини класів і

об’єднання пересічних класів. Найбільшого значення Q отримує кластер «автомобіль + колесо», далі йде кластер «двері + вікно», що більш релевантно ніж початкові результати класифікації.

Інший приклад показаний на рисунку 2.4, в цьому випадку візуальна класифікація помилково показує кращий результат для класу «кішка». В процесі семантичної коригування класи «вікно» і «двері» об’єднуються в один кластер, який має більш високе значення. Відповідні дані представлені в таблиці 2.2 та 2.3.

Таблиця 2.2 – Нормовані результати роботи класифікатора

Кластер	Стіл	Автомобіль	Літак	Квітка	Міст	Небо	Кішка	Хмара	Двері	Дім
Близкість	0,00	0,9	0,72	0,56	0,6	1,00	0,5	0,00	0,3	0,82

Таблиця 2.3 – Результат роботи після семантичного коригування

Кластер	Стіл	Автомобіль	Літак	Квітка	Міст	Небо	Кішка	Хмара	Двері	Дім
Близкість	0,00	1,00	0,37	0,28	0,51	0,25	0	0,16	0,96	0,42

Інший приклад показаний на рисунку 2.4, в цьому випадку візуальна класифікація помилково показує кращий результат для класу «кішка». В процесі семантичної коригування класи «вікно» і «двері» об’єднуються в один кластер, який має більш високе значення. відповідні дані наведені в таблиці 2.4 і в таблиці 2.5.

Рисунок 2.4 –Приклад класифікованого зображення.

Таблиця 2.4 – Нормовані результати роботи класифікатора

Кластер	Стіл	Автомобіль	Літак	Квітка	Колесо	Міст	Небо	Кішка	Хмара	Двері	Вікно	Дім
Близкість	0,0	0,72	0,63	0,64	0,79	0,28	0,69	1	0,14	0,69	0,86	0,93

Таблиця 2.5 – Результат роботи після семантичного коригування

Кластер	Стіл	Автомобіль	Літак	Квітка	Колесо	Міст	Небо	Кішка	Хмара	Двері, Вікно	Дім
Близкість	0,00	0,37	0,32	0,33	0,41	0,14	0,35	0,52	0,07	1,00	0,48

На рисунку 2.5 наводиться приклад зображень, що містять одну і ту ж сцену, відображену з різних ракурсів. В даному випадку класифікація передбачувано показує гранично схожі результати, що підтверджується в таблиці 2.6, і можливе вилучення інформації про геометрію об’єктів з подібних зображень шляхом вирішення завдання пошуку структури з руху.

Рисунок 2.5 – Зображення однієї сцени з різних ракурсів

Таблица 2.6 – Результати класифікації зображення, що має одну сцену з разних ракурсів

Кластер	Стіл	Автомобіль, Колесо	Літак	Квітка	Міст	Небо	Кішка	Хмара	Двері	Дім
Рис. 2.5(1)	0,52	0,37	0,85	0,5	1,00	0,82	0,69	0,75	0,45	0,93
Рис. 2.5(2)	0,5	0,49	0,88	0,5	0,94	0,76	0,56	0,68	0,42	1,00
Різниця	0,02	0,12	0,03	0,00	0,06	0,06	0,13	0,07	0,03	0,07

Використання додаткової інформації при аналізі зображень в плані приналежності до певних класів може підвищувати якість одержуваних результатів з точки зору релевантності. Як джерело інформації про зв’язки між класами зручно використовувати семантичний граф понять, витягнутий з тлумачного словника природної мови, що володіє добре структурованими і не перевантаженими зв’язками. Використання семантичного графа, як

джерела інформації про взаємозв’язках між класами, дозволяє проводити обробку результатів роботи візуальних класифікаторів [27]. Подібна обробка необхідна, в першу чергу, при аналізі складних зображень, в яких можуть бути представлені екземпляри різних класів, і відповідно, релевантність класифікації зображення в цілому в цьому випадку різко знижується.

Для істотного кількості візуальних класів можна відзначити подібність візуальних відстаней між зображеннями представляють класи і семантичної близькості між відповідними класами. Запропоновано метод корекції результатів шляхом пошуку семантично близьких класів, представлених на зображенні, і виділення відповідного контексту шляхом об’єднання близьких класів до загального кластер. При аналізі складних за змістом зображень даний метод показав істотне підвищення релевантності результатів класифікації.

3 ДОСЛІДЖЕННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ЗОБРАЖЕНЬ НА ОСНОВІ ДЕСКРИПТОРІВ ЛОКАЛЬНИХ ОСОБЛИВОСТЕЙ

3.1 Вибір інструментальніх засобів та інформаційних технологій для аналізу зображень на основі дескріпторів локальних особливостей

Перейдемо тепер до задачі аналізу зображень на яких є різне уявлення одного і того ж об’єкта. У цьому випадку результати класифікації цих зображень повинні бути схожі і можливо вилучення інформації про геометрію зображеного об’єкта шляхом вирішення задачі реконструкції структури з руху. Виходячи із зазначених в першому розділі проблем використанням дескрипторів опорних точок для задач вилучення структури з руху, як варіант їх вирішення в даній роботі пропонується поєднання двох підходів – класичного (заснованого на опорних точках і їх дескрипторах) і псевдо-семантичного аналізу вихідного кольорового зображення. Використовуючи методи контуризаціі сегментування зображень можна витягти з зображення більш високорівневу інформацію, ніж просто множину дескрипторів опорних точок. Інформація може являти собою, наприклад, фрагменти зображення, отримані за допомогою стійких алгоритмів сегментації. Незважаючи на те, що завдання сегментування в загальному випадку також не може бути названа остаточно вирішеною, інформація про сегментах може бути використана з метою зниження комбінаторної складності методу при зростанні розмірності [28].

3.1.1 Схема роботи методу

Загальна ідея методу полягає в тому, щоб, використовуючи стійкі алгоритми, розбити вихідні зображення на сегменти, для вилучення

приблизно однакових об’єктів. З якоюсь часткою наближення можна стверджувати, що при різних ракурсах об’єктів зображення будуть схожим чином сегментовані. При цьому навіть якщо один і той же об’єкт потрапить в різні сегменти (або ж буде сегментований на різну кількість частин), на роботі методу це не позначиться негативно. Після отримання розбиття зображень на сегменти проводиться їх зіставлення шляхом підбору перспективного перетворення, яке могло б трансформувати координати опорних точок одного сегмента в координати опорних точок іншого. Такий підхід має право на існування, так як при вирішенні задач ракурси зображень відрізняються незначно – не більше 30-40 градусів нахилу площини, на цій же передумові грунтуються алгоритми вилучення дескрипторів опорних точок. В результаті залишаються тільки ті опорні точки, які при зміні положення камери не призводять до появи несумісних проекцій. Метод працює або з планарними об’єктами (сегментами), або з такими змінами камери, які не призводять до істотної зміни проекції об’єкта, в іншому ж випадку ці опорні точки будуть проігноровані.

Після того, як обрані відповідності сегментів і проведена верифікація їх правильності, опорні точки з цих сегментах передаються в наступні кроки як знайдені відповідності початкових зображень. При цьому вони є вірними з ймовірністю близькою до одиниці, що значно покращує показники подальших статистичних алгоритмів. Крім того, такий спосіб дозволяє збільшити детализированность отриманої в результаті геометрії, так як надає більшу кількість відповідностей точок, ніж у випадку з класичним підходом.

Послідовністю роботи метода є:

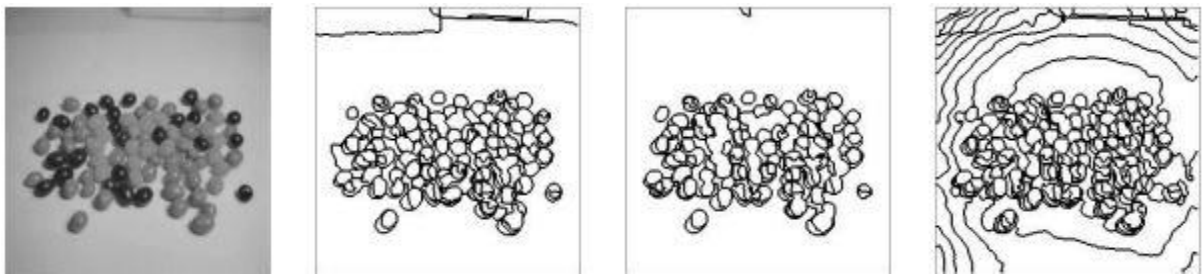
- сегментування вхідних зображень;
- пошук опорних точок;
- витяг дескрипторів;
- пошук відповідностей дескрипторів;
- побудова моделей перетворень сегментів;
- верифікація моделей перетворень на узгодженість;

– фінальна верифікація.

Метод тестувався на зображеннях з різною структурою і дозволяв отримати схожі результати. При тестуванні на зображеннях з різними якісними характеристиками, наведені нижче константи не вимагали змін, або були незначно змінені. Етапи обробки проілюструємо на одному зображенні, що дозволяє якісно оцінити ефективність пропонованої методики.

3.1.2 Сегментування зображень

Для розбиття вхідних зображень на сегменти використовується метод, заснований на алгоритмі MeanShift. Для того, щоб отримати результати сегментування, які найбільше підходять для цілей даної роботи (тобто по можливості сегменти, що складаються з окремих об’єктів), застосовується система EDISON, яка поєднує аналіз розподілу колірних компонентів з аналізом країв зображення . Інформація про розташованих в зображенні краях при цьому витягується з використанням оператора Кенні. Відмінність результату роботи системи EDISON від сегментування на основі алгоритму MeanShift без урахування інформації про краях показано на рисунку 3.1.



(а) (б) (в) (г)

Рисунок 3.1 – Сегментування зображення: (а) вхідне зображення; (б) сегментування EDISON; (в) сегментування MeanShift з тими ж параметрами; (г) надлишкове сегментування MeanShift

Основна мета – отримати якесь розбиття зображення на сегменти, що містять об’єкти, щоб мало сенс проективне перетворення одних сегментів в інші. Розмір сегментів при цьому повинен бути достатній, щоб забезпечити можливість отримання дескрипторів в кількості, що дозволяє ідентифікувати відповідність сегментів і їх взаємне положення. Проведені експерименти показали, що такій умові задовольняють сегменти розміром 100×100 пікселів і більше. Приклад розбиття зображення на сегменти показаний на рисунку 3.2.

Рисунок 3.2 – Початкове і розбите на сегменти зображення

Для вилучення опорних точок використовується алгоритм Scale-invariant Feature Transform (SIFT), як показав себе одним з найбільш стійких до різних спотворень. Ключовою особливістю цього методу є можливість працювати із зображенням в різних дозволах. Він відрізняється інваріантністю до повороту зображення і масштабування, суттєвого діапазону афінних перетворень, наявності шуму і зміни освітленості. Інваріантність до зміни розміру особливостей грунтується на аналізі зображення на різних рівнях масштабу, для цього застосовується модель простору масштабу (scale space) [29]. Існують інші методи пошуку і вилучення дескрипторів цього класу (RIFT, G-RIF, SURF, PCA-SIFT, ASIFT,

GLOH і т.д.), які можуть незначно перевищувати SIFT за характеристиками в певних ситуаціях, однак SIFT найбільш зручний в використанні так як є стандартом де-факто. Приклад отриманих дескрипторів в сегменті показаний на рисунку 3.3.

Рисунок 3.2 – Початкове і розбите на сегменти зображення

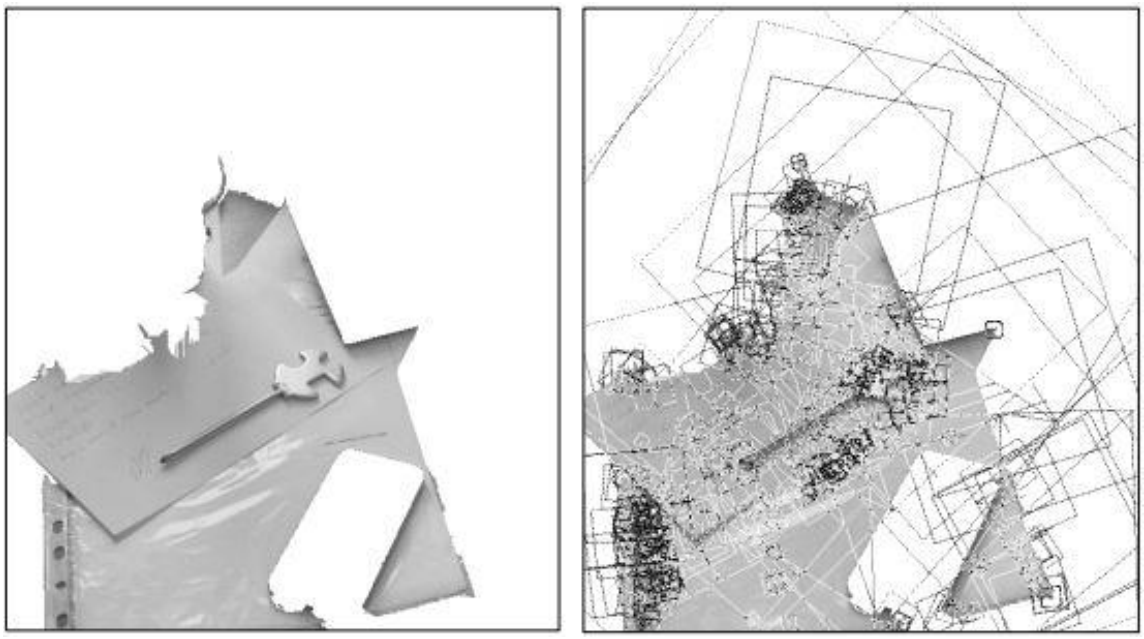


Рисунок 3.3 – Приклад сегмента і його дескрипторів

3.1.3 Попарне порівняння сегментів

Після вилучення дескрипторів з усіх сегментів проводиться пошук відповідників опорних точок аналізованих зображень. Так як на цьому етапі теоретично кожен сегмент першого зображення може відповідати кожному сегменту другого зображення, необхідно виконати попарне порівняння дескрипторів для кожної можливої пари сегментів: проводиться $N \times M$ перехресних порівнянь множин дескрипторів з пошуком відповідників, де N і M – кількість сегментів в першому і другому зображеннях. Так як дескриптори SIFT є 128-мірні вектора і в кожному сегменті можуть бути сотні опорних точок, а для порівняння використовується міра евклідового простору, підходи з повним перебором не є реалізованими на практиці.

Для більш ефективного перехресного порівняння традиційно використовуються алгоритми, засновані на kd деревах. Так як висока розмірність векторів дескрипторів викликає високу обчислювальну складність, при роботі з kd деревами для збільшення продуктивності застосовується наближення, хоча воно і кілька погіршує результати пошуку – в цьому випадку видається над гарантовано найближчий вектор [30].

У пропонованій роботі пошук оптимізується з використанням алгоритму Nene-Nayar (NN), який орієнтований на пошук найближчих значень в просторах з високою розмірністю. Він дозволяє звести пошук найближчого елемента у всій множині до пошуку серед елементів, що знаходяться в єокрестності, обмеженою гіперкубом. Для цього вектора множина послідовно упорядкована відповідно до її компонентів, і ці набори індексів застосовуються для послідовного відсікання пошуку. Іншими словами, метод вимагає, щоб найближчий вектор знаходився всередині гиперкуба зі стороною ε і з центром збігається з одним вектором, в цьому випадку гарантується перебування найближчого:

$$\forall i \in [1..128: |X_i - Y_i| < {}^{\varepsilon}/_{2}], \qquad (3.1)$$

де X і Y – порівнювані вектора дескрипторів;

X_i і Y_i – їх окремі компоненти;

ε – сторона гіперкуба.

Для методу фільтрації неоднозначних відповідностей Lowe спочатку запропонував використовувати міру відстані до найближчих:

$$\forall X \in A: \frac{L_1^X}{L_2^X} < 0{,}8, \tag{3.2}$$

де A – множина дескрипторів, що мають відповідність в порівнюваних зображенні;

L_n^X – евклідова відстань до n-ного найближчого дескриптора в другому наборі дескрипторів.

Таке ставлення дозволяє відкидати відповідності, які не можна використовувати для обчислення геометричних параметрів зображень, тому що такі відповідності можуть змінюватися при малих викривлення або шуми [31]. Додатково проводиться фільтрація потенційних відповідностей «один – до багатьох», коли два або більше дескрипторів першої множини мають найближчим один і той же дескриптор другої множини – в цьому випадку вибирається пара з найменшою відстанню. У певних випадках фільтрація ставленням найближчих може бути надмірною, особливо в ситуації великої кількості схожих дескрипторів.

У даній роботі використовується фільтрація за допомогою взаємного вибору (utual selection) дескрипторів як найближчих при перехресному пошуку найближчих елементів:

$$\forall (X;Y) \in A: \begin{cases} \min_i \left\| X - \overline{Y_i} \right\| = \|X - Y\|, \\ \min_j \left\| Y - \overline{X_j} \right\| = \|Y - X\|, \end{cases} \tag{3.3}$$

де *(X;Y)* – відповідні один одному дескриптори;

A – множина відповідностей дескрипторів;

$\overline{X_j}$ – множина дескрипторів першого зображення;

$\overline{Y_i}$ – множина дескрипторів другого зображення.

Такий підхід вимагає повторного проходження бази даних (пошук найближчих елементів для кожного елемента першої бази і такий же пошук для кожного елемента другий бази), проте дозволяє отримати більшу кількість відповідностей [32]. Співвідношення результатів роботи різних методів зіставлення дескрипторів наведено на рисунку 3.4, де показано вплив ε на кількість фінальних відповідностей після верифікації узгодженості сегментів. На них видно, що використання алгоритму NN дозволяє збільшити кількість відповідностей у фінальній видачу системи, що добре вписується в поставлену задачу.

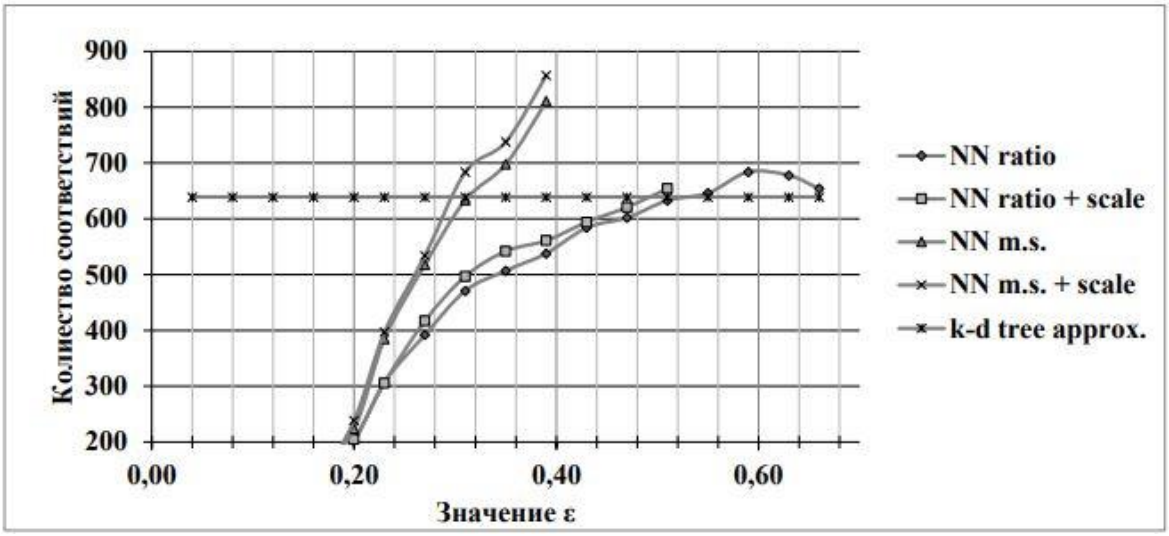


Рисунок 3.4 – Вплив ε на отримане кількість відповідностей

В якості значення ε була використана величина 0,39, як компроміс в балансі між повнотою пошуку з одного боку і продуктивністю системи з іншого. Ця залежність показана на рисунку 3.5. При цьому слід розуміти, що при збільшенні кількості відповідностей, отриманих з пари зображень, відсоток помилкових спрацьовувань також збільшується.

Для додаткової фільтрації позбавлених сенсу відповідностей і для підвищення продуктивності, вводиться обмеження за масштабом, на якому була отримана опорна точка і на якому був розрахований дескриптор. З огляду на використання перспективної моделі для подальшого встановлення відповідностей сегментів, було використано обмеження різниці розміру дескрипторів в 2 рази, тобто ті дескриптори, розмір яких відрізняється більш ніж в 2 рази, не розглядаються як можливі відповідності.

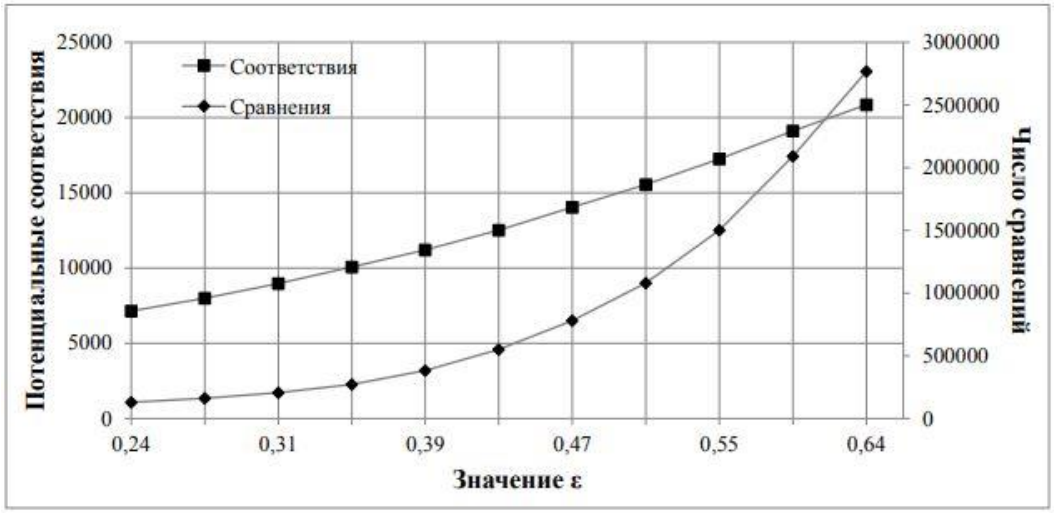


Рисунок 3.5 – Залежність результатів пошуку від величини

Графік поведінки такої фільтрації представлений на рисунку 3.6.

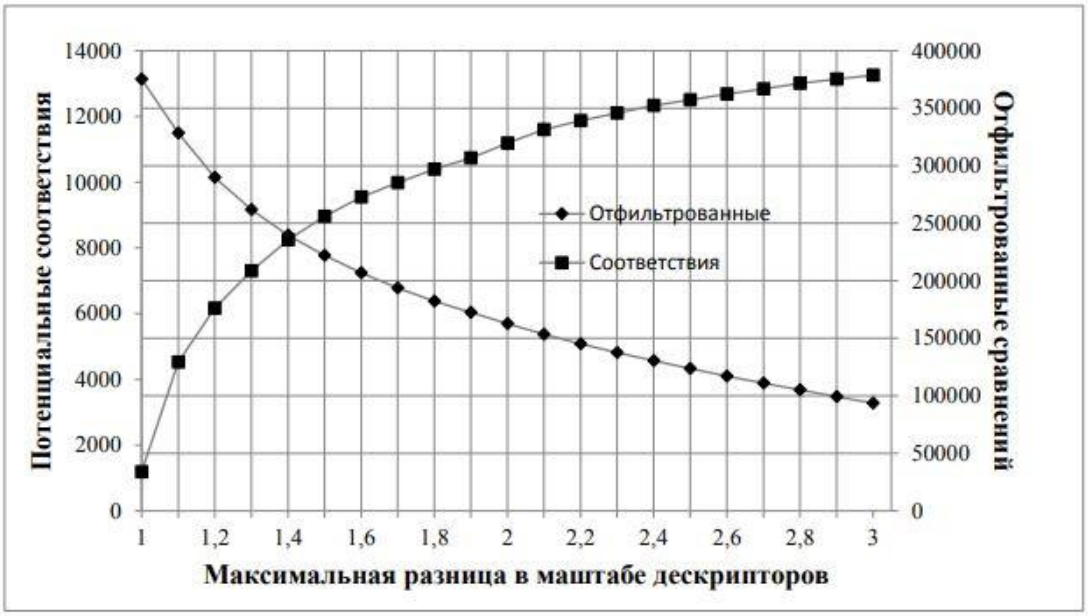


Рисунок 3.6 – Вплив обмеження масштабу на пошук

При такому підході установки не повинні змінюватися, або можуть змінюватися в обмеженому діапазоні в моменти зняття різних ракурсів об’єктів. Якщо ж таку умову задовольнити неможливо, ця фільтрація повинна бути відключена. Приклад відповідностей дескрипторів двох сегментів наведено на рисунку 3.7.

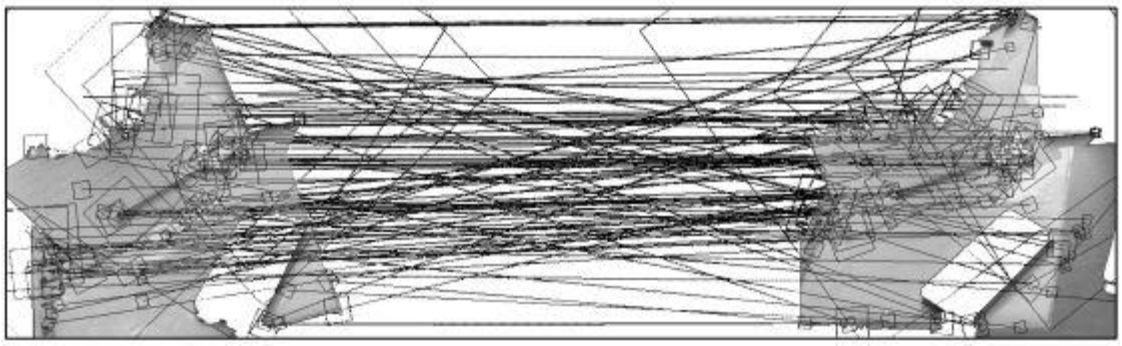


Рисунок 3.7 – Приклад відповідностей дескрипторів

3.2 Тестування розроблення програмні засоби

Існують в цілому два підходи до вирішення завдання пошуку перетворення одного набору векторів в інший при потенційному наявності неправильних відповідностей – повторюваний вибір базових векторів випадковим чином і голосування векторів за стан об’єкта.

Схема голосування використовує підхід Hough Transform, при якому складається багатовимірне (в найпростішому випадку чотиривимірний – дві координати позиції центру, орієнтація і масштаб) простір «кошиків для голосування» і кожен дескриптор голосує за можливі суміжні квантовані положення об’єкта (pose estimation) у другому зображенні, виходячи з його ставлення до центру об’єкта (зміщення, ставлення масштабу і різниця орієнтації) в першому зображенні. Такий підхід добре працює при завданні локалізації об’єкта (object localization), проте він має істотні мінуси. Вибір кроків «кошиків» і відповідного квантування в значній мірі впливає на те, чи

потрапить черговий дескриптор в модель і скільки невірних відповідників залишиться після голосування. Широкі кроки «кошиків» в даній задачі призводять до того, що неузгоджені сегменти можуть отримати більше кількість погоджень, з іншого боку, зменшення кроку призводить до надлишкової фільтрації вірних відповідностей в узгоджених сегментах. Таким чином, в даній роботі пропонується використовувати метод базових векторів з подальшою верифікацією, як більш гнучкий і точніше відповідний завданню – збільшення фінального кількості відповідностей між зображеннями.

На противагу методу Hough Transform (який розглядає множину дескрипторів як цілісний об’єкт), при використанні базових елементів на цьому етапі проводиться спроба пошуку перспективного перетворення координат дескрипторів (тобто не об’єкт в цілому, а окремі координати), що мають відповідність для кожної пари сегментів першого і другого зображення. Таке перетворення має перевести координати дескрипторів сегмента першого зображення в координати відповідних дескрипторів сегмента другого зображення. При цьому для обчислення гомографії на цьому етапі враховуються тільки координати дескрипторів, але не враховуються їх орієнтації і розміри. Ці дані можна використовувати для формування додаткової координати на кожен дескриптор, і фактично перетворення їх з однієї координати в вектор, і далі обраховувати перетворення, що переводить один набір векторів в інший. Однак, подібна додаткова інформація веде до додаткових вимог до перетворення, і в кінцевому підсумку до надмірної фільтрації.

Для обчислення гомографії використовується статистична модель RANSAC, суть якої полягає в тому, щоб N раз вибрати випадкові базові елементи, за якими буде будуватися максимально наближена модель перетворення. Мінімально необхідна кількість базових елементів для пошуку гомографії – 4 точки, але проведені експерименти показали набагато стабільніші результати при використанні 5 точок. Перетворення знаходиться

шляхом рішення системи рівнянь, приведених до однорідного виду, з використанням сингулярного розкладання (Singular value decomposition – SVD).

Нехай $X_i = (x_i, y_i)$ і $X_i' = (x_i', y_i')$ – відповідні один одному точки в зображеннях. Тоді задачу можна сформулювати як знаходження такої матриці перетворення H, яка переводить координати точок першого зображення в координати точок другого:

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \cong \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \tag{3.4}$$

де використовується представлення у вигляді однорідних координат і H визначається з точністю до масштабу.

Пошук значення h, що мінімізує значення помилки, проводиться за допомогою сингулярного розкладання матриці $A^T A$:

$$A^T A = U\ D\ U^T, \tag{3.5}$$

де D – діагональна матриця, що складається з сингулярних чисел;

U – матриця складається з сингулярних векторів.

Вектор h таким чином дорівнює сингулярному вектору, відповідному мінімальному сингулярному числу. Після того, як побудована матриця перетворення, усі відповідності опорних точок перевіряються на узгодженість з обмеженням моделі:

$$\|X' - HX\| < d\ , \tag{3.6}$$

де d – гранична величина похибки перетворення координат.

З огляду на, що координати точок нормовані в діапазоні [-1..1], *d* приймається рівним 0,03. Ті відповідності, координати яких після перетворення відрізняються від цільових менш ніж на максимальну похибку, вважаються, що задовольняють моделі – «не-викидами» (inliers), решта вважаються невірними – «викидами» (outliers) [33]. Вплив значення максимальної похибки моделі на загальну сумарну похибку показано на рисунку 3.8. Після всіх кидків кубиків, вибирається той варіант, який мав найбільшу кількість, що задовольняють моделі відповідностей, і він приймається як потенційне перетворення одного сегмента в інший. на рисунку 3.9 зображено приклад, що задовольняє моделі відповідностей дескрипторів.

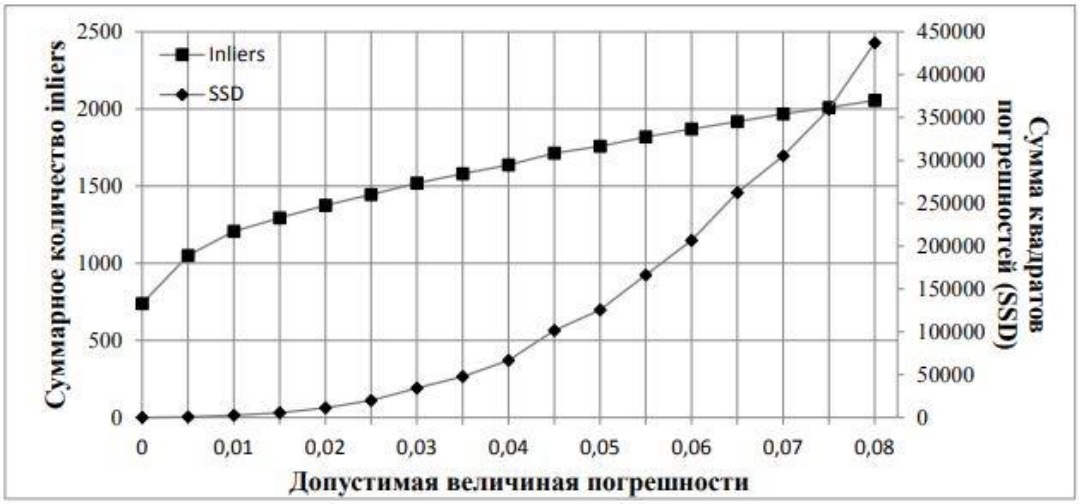


Рисунок 3.8 – Вплив максимальної похибки

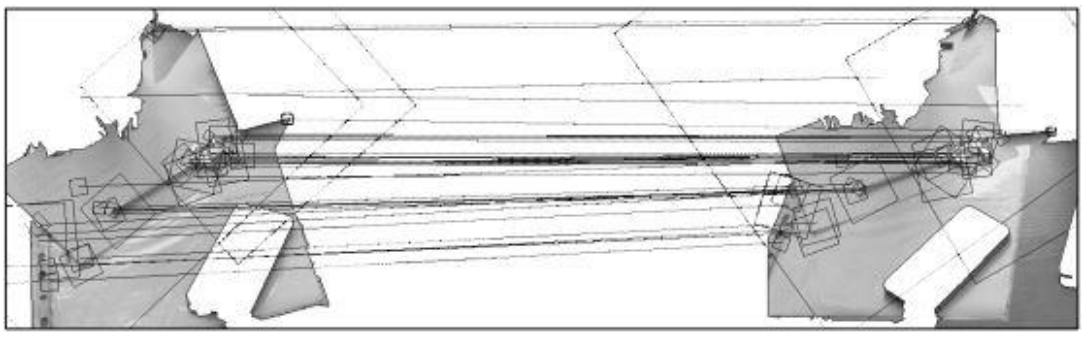


Рисунок 3.9 – Приклад, що задовольняє моделі відповідностей

3.3 Результати дослідження методів аналізу зображень на основі дескрипторів локальних особливостей

Самим нетривіальним завданням в запропонованому методі є верифікація того, що отримана модель перспективного перетворення має сенс, тому що, з великою часткою ймовірності, опорні точки одного сегмента можуть мати можливе перетворення в опорні точки невідповідного сегмента. Це пов’язано з тим, що дескриптори не можуть однозначно ідентифікувати місце в зображенні, а є, по суті, хешированием в загальному розумінні цього слова. Незважаючи на те, що в базі методу RANSAC закладено те, що невірні відповідності не зможуть між собою «домовитися» про задовольняє їх моделі, на практиці така ситуація може виникати і потребує вирішення. На рисунку 3.10 проілюстровано приклад, який не має сенсу перетворення сегментів одного в інший, при цьому зв’язками показано відповідності, які змогли потрапити в загальну модель (inliers).

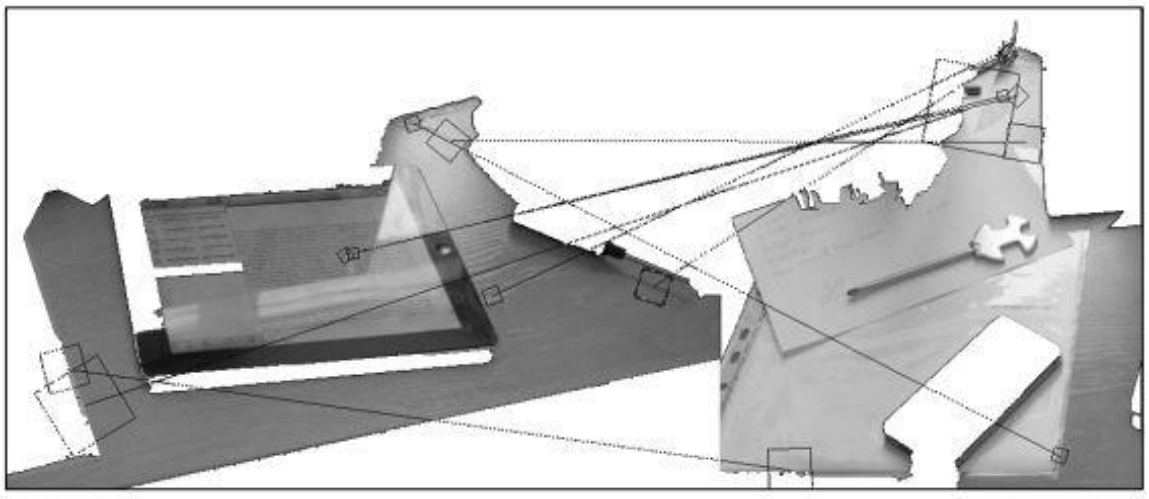


Рисунок 3.10 –Приклад не має сенсу перетворення

В даній роботі пропонується спосіб, що базується на перетворенні Hough, Суть його полягає в тому, щоб закодувати взаємне розташування опорних точок в першому сегменті і потім розкодувати його, використовуючи координати другого сегмента. Дескриптори SIFT (крім

безпосередньо 128-мірного вектора) містять таку інформацію, як розташування (X, Y), його масштаб (S – scale) і кут повороту (α):

$$P = \{X,\ Y,\ S, \alpha\}. \tag{3.7}$$

За основу перерахунку прийнята така величина, як центр маси опорних точок, тому що при невироджених перспективних перетвореннях взаємне розташування центру мас і опорних точок залишиться тим же самим. Таким чином, якщо перерахувати кожну точку першого сегмента у взаємне положення щодо центру мас, і потім зробити зворотний розрахунок для точок другого сегмента, можна отримати якусь міру узгодженості. В ідеальному випадку, при відсутності шумів і оптичних спотворень, після зворотного перетворення повинно вийти координата центру мас другого сегмента (якщо відповідність першого сегмента і другого вірно). У реальності, зрозуміло, має місце дисперсія розрахованих центрів мас, однак це не заважає використанню такої міри для оцінки узгодженості точок між собою.

Суть перетворення полягає в перерахунку інформації про кожну опорної точки з чотирьох параметрів в два:

$$P = \{X,\ Y,\ S, \alpha\} \rightarrow \underline{P} = \{\theta, D/S\}, \tag{3.8}$$

де θ – кут між вектором орієнтації дескриптора і вектором з'єднує центр дескриптора і центр мас M;

D/S – відношення відстані між центром дескриптора і центром мас і масштабом дескриптора.

Взаємне розташування дескрипторів і центрів мас показано на рисунку 3.11, при цьому $P' = \{X',\ Y',\ S', \alpha'\}$ – відповідний дескриптор в іншому зображенні.

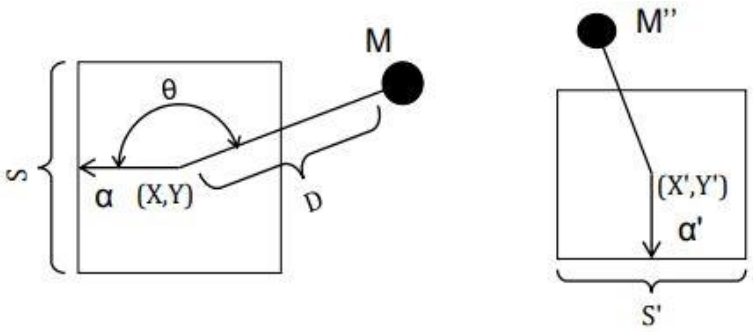


Рисунок 3.11 – Інформація для перетворення дескрипторів

Для кожної точки другого сегмента проводиться відповідний розрахунок очікуваного положення центру мас:

$$M'' = \begin{bmatrix} X' + \cos(\theta + \alpha')\frac{D}{S}S' \\ Y' + \sin(\theta + \alpha')\frac{D}{S}S' \end{bmatrix}. \quad (3.9)$$

Після того, як отримані координати нових центрів мас, можна оцінити їх скупченість, тим самим оцінити узгодженість взаємних положень опорних точок.Це можна робити множиною різних способів – як вводити загальну міру дисперсії, так і оцінювати кожну точку окремо. В рамках даної роботи найкращим чином проявив себе наступний варіант оцінки кожної точки:

$$\begin{cases} D = \|M' - M''\|, \\ F(S') = 2S', \\ D < F(S'). \end{cases} \quad (3.10)$$

де M' – центр мас другого сегмента розрахований прямим чином.

Таким чином, корелюється похибка зворотного обчислення центру мас і масштабу дескриптора, для якого був проведений розрахунок, що дозволяє домогтися інваріантності щодо масштабу. При цьому кращих результатів вдалося домогтися, коли дана перевірка узгодженості проводилася прямо в циклі RANSAC при складанні моделі перспективного перетворення, і

використовувалася для фільтрації inliers/outliers , а не проводилась наступним кроком після вже отриманої моделі перетворення. Різні результати зворотного розрахунку центру мас показано на рисунку 3.12.

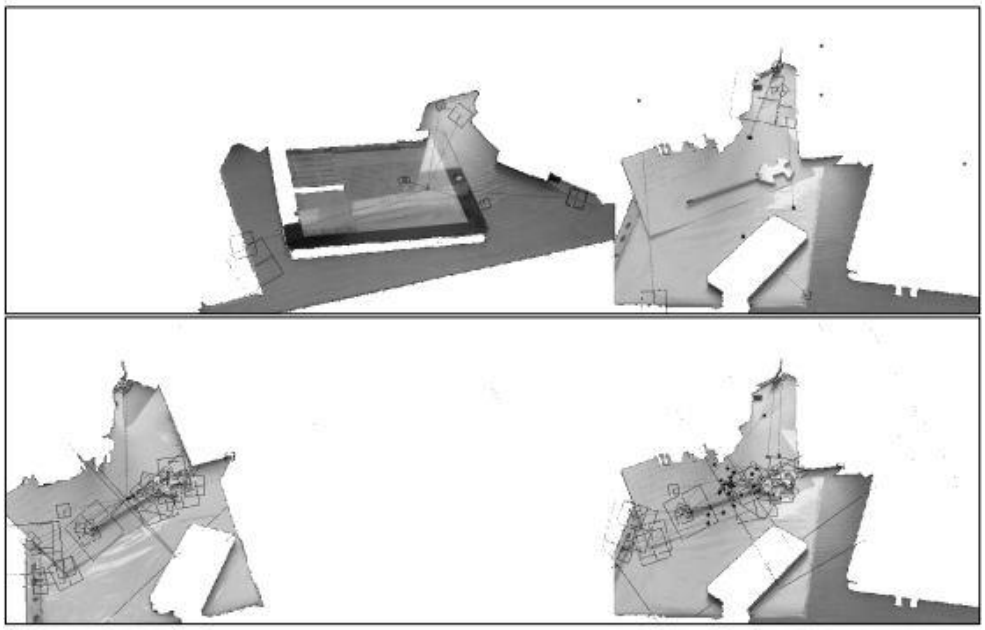


Рисунок 3.12 – Поведінка верифікації в різних випадках

Фінальною фільтрацією служить перевірка на мінімальну кількість inliers, що задовольняють і моделі перспективного перетворення та верифікації узгодженості, воно було прийнято рівним 7 (ілюстрація залежності вірних і помилкових прийнять рішень показана на рисунку 3.13).

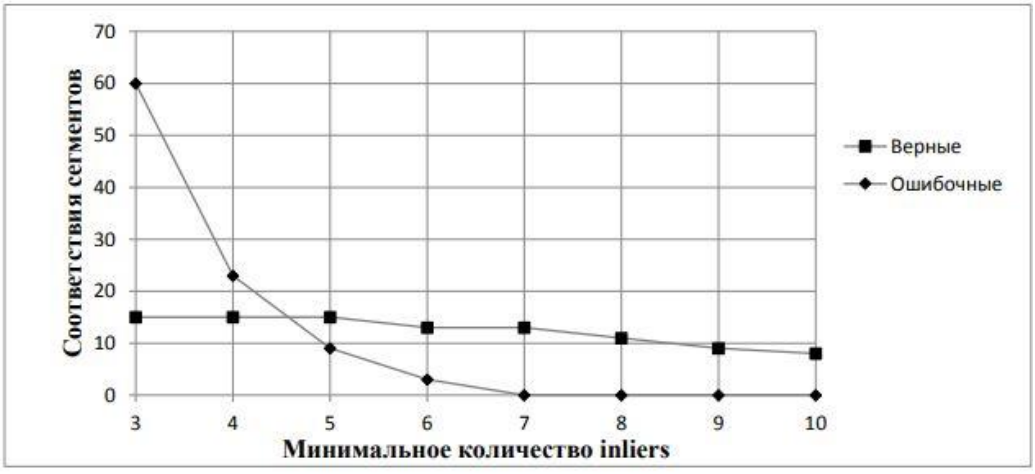


Рисунок 3.13 – Вплив мінімальної кількості inliers

ВИСНОВКИ

У рамках атестаційної роботи проаналізовані негативні побічні ефекти, пов’язані з аналізом зображень шляхом опису дескрипторами локальними візуальних особливостей. Запропоновано підходи до використання дескрипторів локальних візуальних особливостей, що дозволяють поліпшити якість аналізу зображень.

Запропоновано метод використання пошукових систем в якості джерела зображень для автоматизованого формування навчальної вибірки в задачах класифікації, розглянуті пов’язані з цим питання обробки. Автоматичне формування навчальної вибірки дозволяє знизити негативні ефекти, пов’язані з ручним формуванням наборів зображень, і дозволяє гнучко проводити підстроювання системи класифікації, яка навчається на даних наборах зображень. Докладно розглянуті питання підходів до використання отриманої навчальної вибірки для аналізу вхідних зображень і питання можливості фільтрації видачі пошукової системи в разі появи НЕ що відносяться до візуального класу зображень.

Запропоновано метод застосування семантичних графів понять для верифікації та корекції результатів візуальної класифікації. Дана обробка необхідна, в першу чергу, при аналізі складних зображень, в яких можуть бути представлені екземпляри різних візуальних класів, і відповідно, релевантність класифікації зображення в цілому в цьому випадку різко знижується. Запропонований механізм корекції результатів шляхом пошуку семантично близьких класів, представлених на зображенні, і виділення відповідного контексту шляхом об’єднання близьких класів в загальний кластер, показав істотне підвищення релевантності класифікації складних зображень.

Запропоновано метод зіставлення дескрипторів зображень з використанням проміжного представлення у вигляді фрагментів, дозволяє збільшувати детализированность геометрії, отриманої в Під час виконання

завдання структури з руху. Даний підхід знижує негативний ефект, пов’язаний з неоднозначністю дескрипторів локальних візуальних особливостей, і дозволяє враховувати колірні компоненти зображень при їх аналізі.

Результати даного дослідження апробовано на 24-му Міжнародному молодіжному форумі «Радіоелектроніка та молодь у XXI столітті» [34] та на IV Міжнародній науково-практичній конференції «Integration of scientific bases into practice» (Стокгольм, Швеція) [35].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Творошенко, І. С. (2015). Конспект лекцій з дисципліни «Цифрова обробка зображень»(для студентів 5 курсу денної та заочної форм навчання спеціальності 7.08010105–Геоінформаційні системи та технології).

2. Творошенко, І. С. Конспект лекцій з дисципліни «Цифрова обробка зображень» для студентів 4 курсу денної форми навчання напряму 6.080101–Геодезія, картографія та землеустрій.

3. Творошенко, І. С. (2018). Основи цифрової обробки зображень: конспект лекцій для студентів 4 курсу денної форми навчання напряму 6.080101–Геодезія, картографія та землеустрій.

4. Tvoroshenko, I. S., & Gorokhovatsky, V. O. (2019). Intelligent classification of biophysical system states using fuzzy interval logic. Telecommunications and Radio Engineering, 78(14).

5. Gorokhovatskyi, V. O., Tvoroshenko, I. S., & Peredrii, O. O. (2020). IMAGE CLASSIFICATION METHOD MODIFICATION BASED ON MODEL OF LOGIC PROCESSING OF BIT DESCRIPTION WEIGHTS VECTOR. Telecommunications and Radio Engineering, 79(1).

6. Tvoroshenko, I. S. (2010). Analysis of Decision-Making Processes in Intelligent Systems, Information Processing Systems, 2.

7. Gorokhovatskyi, V., & Tvoroshenko, I. (2020). Image Classification Based on the Kohonen Network and the Data Space Modification.

8. Kobylin, O. A., Gorokhovatskyi, V. O., Tvoroshenko, I. S., & Peredrii, O. O. (2020). THE APPLICATION OF NON-PARAMETRIC STATISTICS METHODS IN IMAGE CLASSIFIERS BASED ON STRUCTURAL DESCRIPTION COMPONENTS. Telecommunications and Radio Engineering, 79(10).

9. Gorokhovatskyi, V. O., Tvoroshenko, I. S., & Vlasenko, N. V. (2020). USING FUZZY CLUSTERING IN STRUCTURAL METHODS OF IMAGE CLASSIFICATION. Telecommunications and Radio Engineering, 79(9).

10. Szeliski R. (2010) Computer Vision: Algorithms and Applications, London, Great Britain: Springer-Verlag, 957 p.

11. Fergus R., Fei-Fei L., Perona P., Zisserman A. Learning object categories from Google's image search // Tenth IEEE International Conference on Computer Vision, ICCV. 2005. V.2, P.1816-1823.

12. Ferrari V., Tuytelaars T., Van Gool L. Simultaneous object recognition and segmentation from single or multiple model views // International Journal of Computer Vision. 2006. 67(2), P.159-188

13. Gorokhovatskyi V., Gadetska S., and Stiahlyk N. (2020) Image structural classification technologies based on statistical analysis of descriptions in the form of bit descriptor set, In CEUR Workshop Proceedings: Computer Modeling and Intelligent Systems (CMIS-2020), 2608, pp. 1027-1039.

14. Sonka M., Hlavac V., and Boyle R. (2014) Image Processing, Analysis, and Machine Vision, Atlanta, USA: Thomson-Engineering, 920 p.

15. 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), pp. 787-797. DOI: 10.1615/TelecomRadEng.v77.i9.40

16. Gorokhovatskyi V.O., Gadetska S.V., and Stiahlyk N.I. (2019) Study of statistical properties of the block representation model for a set of key image descriptors, Radio Electronics Computer Science Control, 2, pp. 100-107.

17. Duda R.O., Hart P.E., and Stork D.G. (2000) Pattern classification, Hoboken, USA: John Wiley & Sons, 738 p.

18. Tvoroshenko Irina, Ahmad M. Ayaz, Mustafa Syed Khalid, Lyashenko Vyacheslav, and Alharbi Adel R. (2020) Modification of Models Intensive Development Ontologies by Fuzzy Logic, International Journal of Emerging Trends in Engineering Research, 8(3), pp. 939-944. DOI: 10.30534/ijeter/2020/50832020

19. Mikolajczyk K., Schmid C. A performance evaluation of local descriptors // IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE. 2005.

20. Pantofaru C., Hebert M. A comparison of image segmentation algorithms / Robotics Institute, Carnegie Mellon University, 2005.

21. Yousef Ibrahim Daradkeh, and Iryna Tvoroshenko (2020) Application of an Improved Formal Model of the Hybrid Development of Ontologies in Complex Information Systems, Applied Sciences, 10(19). p. 6777. DOI: 10.3390/app10196777

22. Peters J.F. (2017) Foundations of computer vision: Computational Geometry, Visual Image Structures and Object Shape Detection, Cham, Switzerland: Springer International Publisher, 417 p.

23. Szeliski R. Computer vision: algorithms and applications / Springer Science & Business Media, 2010.

24. Tvoroshenko I.S., and Gorokhovatsky V.O. (2019) Intelligent classification of biophysical system states using fuzzy interval logic, Telecommunications and Radio Engineering, 78(14), pp. 1303-1315. DOI: 10.1615/TelecomRadEng.v78.i14.80.

25. Ahmad M. Ayaz, Tvoroshenko Irina, Baker Jalal Hasan, and LyashenkoVyacheslav (2019) Computational Complexity of the Accessory Function Setting Mechanism in Fuzzy Intellectual Systems, International Journal of Advanced Trends in Computer Science and Engineering, 8(5), pp. 2370-2377. DOI:10.30534/ijatcse/2019/77852019

26. Gorokhovatskyi V., and Tvoroshenko I. (2020) Image Classification Based on the Kohonen Network and the Data Space Modification, In CEUR Workshop Proceedings: Computer Modeling and Intelligent Systems (CMIS-2020), 2608, pp.1013-1026.

27. Gorokhovatskyi V.O., Tvoroshenko I.S., and Peredrii O.O. (2020) Image classification method modification based on model of logic processing of

bit description weights vector, Telecommunications and Radio Engineering, 79(1), pp. 59-69.

28. Sharma G., and Schiele B. (2015) Scalable Nonlinear Embeddings for Semantic Category-based Image Retrieval, Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV), pp. 7-13.

29. Gorokhovatsky V.A. (2016) Efficient Estimation of Visual Object Relevance during Recognition through their Vector Descriptions, Telecommunications and Radio Engineering, 75(14), pp. 1271-1283.

30. Matarneh Rami, Tvoroshenko Irina, and Lyashenko Vyacheslav (2019) Improving Fuzzy Network Models For the Analysis of Dynamic Interacting Processes in the State Space, International Journal of Recent Technology and Engineering, 8(4), pp. 1687-1693. DOI: 10.35940/ijrte.C5582.118419

31. Tvoroshenko I.S., and Gorokhovatsky V.O. (2019) Modification of the branch and bound method to determine the extremes of membership functions in fuzzy intelligent systems, Telecommunications and Radio Engineering, 78(20), pp. 1857-1868. DOI: 10.1615.

32. M. Ayaz Ahmad, Irina Tvoroshenko, Jalal Hasan Baker, and Vyacheslav Lyashenko (2019) Modeling the Structure of Intellectual Means of Decision-Making Using a System-Oriented NFO Approach, International Journal of Emerging Trends in Engineering Research, 7(11), pp. 460-465. DOI: 10.30534/ijeter/2019/107112019

33. Tvoroshenko I.S., and Gorokhovatsky V.O. (2020) Effective tuning of membership function parameters in fuzzy systems based on multi-valued interval logic, Telecommunications and Radio Engineering, 79(2), pp. 149-163. DOI: 10.1615/TelecomRadEng.v79.i2.70.

34. Ткаченко Д.А. Нормалізація проективно спотворених зображень з використанням інваріантних відображень. *Радіоелектроніка та молодь у XXI столітті*: тези доповідей 24-го Міжнародного молодіжного форуму (Харків, 7–9 квітня 2020 р.). Харків: ХНУРЕ, 2020. Т. 7. С. 52-53.

35. Tvoroshenko I., and Tkachenko D. (2020) Mechanisms of image classification based on descriptors of local features, *Abstracts of IV International Scientific and Practical Conference «Integration of scientific bases into practice» (October 12-16, 2020). Stockholm, Sweden*, pp. 443-448. DOI: 10.46299/ISG.2020.IV