МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра комп'ютерних наук

ПШИК Віталій Володимирович

Математичне та програмне забезпечення анотування зображень в пошукових системах / Mathematical Tools and Software for Images Annotation in Search Engines

спеціальність: 121 - Інженерія програмного забезпечення
освітньо-професійна програма - Інженерія програмного забезпечення

Кваліфікаційна робота

Виконав студент групи ІПЗм-21
В. В. Пшик
_____

Науковий керівник:
к.т.н., доцент Є. О. Марценюк
_____

Кваліфікаційну роботу
допущено до захисту:

"___" _____ 20___ р.

Завідувач кафедри
_____ **А. В. Пукас**

ТЕРНОПІЛЬ - 2022

ЗМІСТ

ВСТУП 9
РОЗДІЛ 1 ОСОБЛИВОСТІ СУЧАСНИХ МЕТОДІВ ТА АЛГОРИТМІВ АВТОМАТИЧНОГО АНОТУВАННЯ ЗОБРАЖЕНЬ 11
1.1. Аналіз відомих методів автоматичного анотування зображень 11
1.2. Пошукові методи 16
1.3. Порівняння методів автоматичного анотування зображень 17
1.4. Аналіз методів кластеризації даних 19
1.5. Аналіз відомих програмних систем для анотування зображень 20
Висновки до першого розділу 25
РОЗДІЛ 2 МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ СИСТЕМИ АНОТУВАННЯ ЗОБРАЖЕНЬ В ПОШУКОВИХ СИСТЕМАХ 27
2.1. Обчислення глобального візуального дескриптора 27
2.2. Обчислення глобального візуального дескриптора 29
2.3. Обчислення локальних дескрипторів кольорів 33
2.4. Метод створення текстового дескриптора 34
Висновки до другого розділу 38
РОЗДІЛ 3 ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ СИСТЕМИ АНОТУВАННЯ ЗОБРАЖЕНЬ В ПОШУКОВИХ СИСТЕМАХ 39
3.1. Загальна архітектура системи 39
3.2. Результати експериментальних досліджень обчислення візуальних дескрипторів 45
3.3. Дослідження параметрів алгоритму формування глобальних дескрипторів 48
3.4. Результати експериментальних досліджень автоматичного анотування зображень 52
Висновки до третього розділу 54

ВИСНОВКИ..... 56
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... 58
ДОДАТОК А ЛІСТИНГ ОСНОВНИХ МОДУЛІВ СИСТЕМИ ..... **Помилка! Закладку не визначено.**

ВСТУП

Актуальність теми. В останні десятиліття широке поширення пристроїв із вбудованими відеокамерами призвело до експоненційному зростанню кількості зображень у мережі інтернет, що викликало необхідність їхнього ефективного пошуку. Існуючі методи пошуку зображень можна розділити на три типи: пошук по текстовим анотаціям, аналіз зображень з візуального змісту та методи на основі автоматичного анотування. У пошукових методах першого типу зображенням вручну надаються суб'єктивні текстові описи, а пошук здійснюється як у текстових документах. Методи пошуку зображень за змістом, що вимагають зображення-запит, виконують пошук на основі аналізу та порівняння низькорівневих ознак зображення, такі як колір або текстури. Однак при цьому часто спостерігається проблема семантичного розриву – відсутності зв'язку між низькорівневими ознаками зображення та його інтерпретацією людиною.

Основною ідеєю методів автоматичного анотування зображень є формування семантичної моделі з навчальної вибірки зображень великого об'єму. За допомогою семантичної моделі ключові слова автоматично визначаються для нових зображень. Таким чином, методи автоматичного анотування припускають пошук по ключових словах, отриманих на основі аналізу змісту зображень та використовують переваги перших двох підходів.

Зв'язок роботи з науковими програмами, планами, темами

Напрямок виконаних досліджень безпосередньо пов'язаний з науково-дослідним напрямком кафедри “комп’ютерних наук” Західноукраїнського національного університету.

Мета і задачі дослідження

Метою роботи є підвищення ефективності автоматичного анотування зображень в інформаційно-пошукових системах.

Відповідно до поставленої мети у роботі потрібно вирішити такі основні завдання дослідження:

1. Провести аналіз методів та алгоритмів автоматичного анотування зображень, кластеризації даних, опис зображень за допомогою низькорівневих ознак.

2. Розробити алгоритм відновлення пропущених ключових слів в анотаціях навчальних зображень.

3. Створити алгоритм автоматичного анотування зображень на основі однорідних текстово-візуальних груп.

4. Розробити програмне забезпечення, що реалізує алгоритми обчислення дескрипторів, відновлення пропущених ключових слів, формування однорідних текстово-візуальних груп та автоматичного анотування зображень

Об'єкт дослідження – анотування зображень в інформаційно-пошукових системах.

Предмет дослідження – методи та програмні засоби анотування зображень в інформаційно-пошукових системах.

Методи дослідження

В роботі використовувалися методи теорії цифрової обробки зображень, теорії обробки інформації, методи теорії розпізнавання образів та аналізу даних, методи об'єктно-орієнтованого програмування.

Наукова новизна одержаних результатів

Запропоновано метод швидкого вилучення набору локальних дескрипторів, що описують усі частини зображення, що дозволяє істотно прискорити процес анотування та формувати більш повний глобальний візуальний дескриптор зображення.

Практичне значення одержаних результатів

В рамках магістерського дослідження розроблено експериментальне програмне забезпечення для автоматичне анотування зображень.

РОЗДІЛ 1
ОСОБЛИВОСТІ СУЧАСНИХ МЕТОДІВ ТА АЛГОРИТМІВ АВТОМАТИЧНОГО АНОТУВАННЯ ЗОБРАЖЕНЬ

1.1. Аналіз відомих методів автоматичного анотування зображень

Відомі методи автоматичного анотування зображень можна розділити на дві категорії, анотуючі зображення за допомогою одного та кількох ключових слів відповідно. Класифікація методів за категоріями наведена в таблиці 1.1.

Таблиця 1.1

Класифікація методів автоматичного анотування зображень

Категорії	Підходи	Методи
Анотування одним ключовим словом	Класифікаційний	– На основі невід'ємного матричного розкладання – На основі методу опорних векторів – На основі багатоваріантного навчання
Анотування декількома ключовими словами	Генеративний	- Модель спільної зустрічальності - Модель машинного перекладу – На основі моделей релевантності
	Пошуковий	– Joint Equal Contribution (JEC) - Tag Propagation (TagProp) – 2-Pass K-Nearest Neighbor (2PKNN)

Методи класифікаційного підходу розглядають процес анотування зображень як проблему категоризації зображень. Для цього ключові слова видаються у вигляді незалежних класів, на приклади яких навчається класифікатор. При анотуванні нового зображення класифікатор визначає клас, до якого воно відноситься, та надає відповідне ключове слово. Декілька ключових слів можуть бути отримані з припущення, що зображення належить кільком класам. Розглянемо докладніше деякі методи цього підходу.

Методи на основі невід'ємного матричного розкладання. Невід'ємне матричне розкладання (NMF, Non-negative Matrix) Factorization) є одним з методів розкладання матриць, завдяки обмеження на невід'ємність, що набув поширення для обробки даних (таких як текстові документи та зображення) на основі аналізу їх частин [1-5]. У роботі [4] метод NMF використовувався для класифікації зображень. Автори роботи створили колекцію, що складається з плиток (квадратних фрагментів) зображень і розділили її на 10 класів. З цієї колекції випадково вибиралося по 1000 плиток для формування навчальної та тестової вибірок. Під час навчання метод NMF формував підпростори для кожного класу, на яких надалі навчався класифікатор. При класифікації тестове зображення спочатку відображалось у кожне з 10 створених підпросторів, після чого вибирався клас, який отримав найбільше відгуків класифікатора.

У подальшій роботі дані автори порівнювали кілька різних метрик у просторах, одержаних за допомогою методу NMF. У своїх експериментах з класифікації об'єктів вони виявили, що у випадку, коли об'єкти частково перекривають один одного, метод NMF косинусною метрикою показує найкращі результати. Однак у роботі [7] було показано, що базис, отриманий за допомогою методу NMF, підходить для безпосереднього розпізнавання об'єктів за допомогою методів найближчого сусіда, вони запропонували проводити ортонормалізацію базису перед подальшим аналізом, внаслідок чого підвищувалася точність розпізнавання об'єктів.

Підходи на основі методу опорних векторів. Метод (машина) опорних векторів (SVM, Support Vector Machine) є одним із найбільш популярних методів для класифікації даних [9]. Основна ідея лінійного методу опорних векторів полягає в тому, що множин ознак, що належить двом класам, можна розділити оптимальною гіперплощиною. Оптимальна гіперплощина формує компактні множини з найбільшої кількості ознак одного і того ж класу, при цьому максимізуються відстані від обох класів до гіперплощини.

У роботі [10] автори одними з перших застосували метод опорних векторів для класифікації зображення. Для опису зображень використовувалися колірні гістограми, а метод опорних векторів, спочатку розроблений для класифікації двох класів, навчався за принципом "один проти всіх" для класифікації семи класів.

Також було запропоновано метод для класифікації областей зображень за допомогою ансамблю SVM-класифікаторів [11]. У даному методі на На першому етапі зображення розбивається сітками на прямокутні плитки з використанням кратних 8 пікселів масштабів. З кожної плитки витягується 90-мірний вектор ознак, після чого отримане 90-мірний простір різнорідних ознак (значення та діапазони одних ознак суттєво відрізняється від інших) розбивається на 9 однорідних підмножин. На другому етапі "слабкий" SVM-класифікатор навчається для кожної однорідної підмножини ознак. В результаті навчання вибираються найефективніші класифікатори, а також відповідні підмножини ознак та розміри плиток. На останньому етапі обрані «слабкі» SVM-класифікатори об'єднуються з використанням методу бустингу.

Також було запропоновано метод, у якому для автоматичного анотування зображень комбінуються два набори SVM класифікаторів [12]. Один набір класифікаторів навчається на ознаках областей зображень, отриманих за допомогою багатоваріантного методу навчання (MIL, Multiple Instance Learning) [13], а інший набір використовує глобальні ознаки

зображень на навчання. Результати роботи обох наборів класифікаторів об'єднуються для анотування нових зображень.

Підходи на основі багатоваріантного навчання. Багатоваріантне навчання є різновидом бінарного методу навчання з учителем. Даний метод замість навчання на наборі елементів, кожен з яких позначений як позитивний або негативний, отримує набір позитивних та негативних пакетів (Bags). Кожен пакет містить кілька елементів. Він позначається як негативний, якщо всі його елементи негативні, і як позитивний, якщо хоч один елемент пакета є позитивним (рисунок 1.1). Ціль методу MIL полягає у навчанні принципу, за допомогою якого можна правильно помічати окремі елементи.

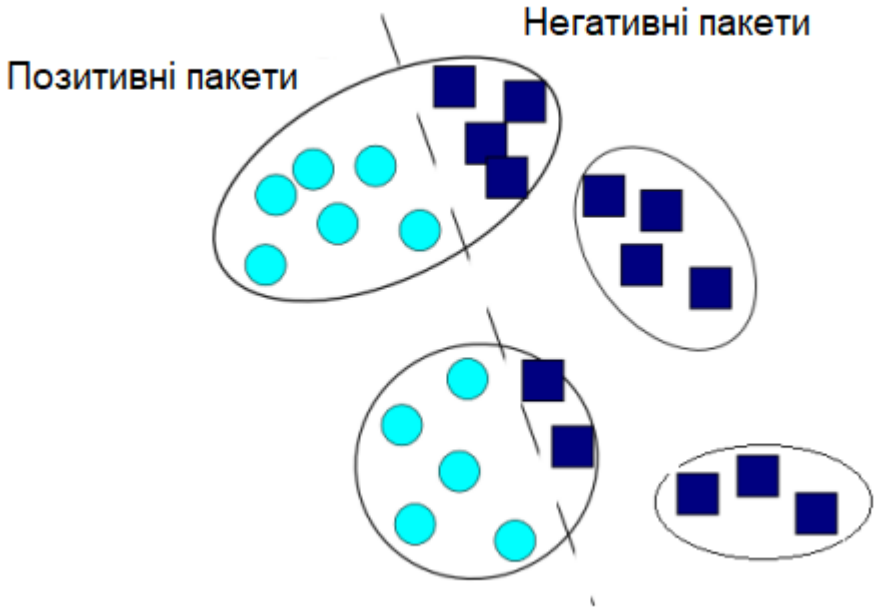

Рис. 1.1. Приклад пакетів, помічених як позитивні та негативні

Для вирішення цієї проблеми було запропоновано підхід, званий Diverse Density (DD). Основна ідея підходу полягає у обчисленні для кожного елемента DD-значення, що є мірою того, скільки різних позитивних пакетів мають елементи поблизу даного елемента та як далеко від цього елемента розташовані елементи негативних пакетів.

У деяких роботах зображення розглядається як пакет елементів, кожен з яких являє собою вектор ознак області зображення. По відношенню до певного ключового слова зображення, проанотовані цим ключовим словом, позначаються як позитивні, тоді як інші позначаються як негативні. У роботі [15] для класифікації зображень запропоновано метод DD-SVM, об'єднуючий метод Diverse Density із класифікатором SVM. На першому етапі даного методу зображення сегментуються на області, після чого кожній вилучається 9-мірний вектор ознак, що використовується як елемент пакета. На наступному етапі визначається набір елементів прототипів, використовуючи DD-функцію. Кожен елемент-прототип є представником класу елементів, які з більшою ймовірністю з'являться у пакетах з однією міткою. З використанням елементів-прототипів в як осі створюється новий простір, в який відображаються навчальні пакети (зображення). При цьому координата пакета на конкретної осі дорівнює відстані між відповідним прототипом та найближчим до нього елементом пакета. На останньому етапі SVM класифікатор навчається на основі розташування пакетів у створеному просторі. У роботі [17] використовувався аналогічний підхід, проте замість DD-функції для вибору прототипів та класифікації елементів був адаптований метод розріджених опорних векторів (Sparse Support Vector Machine). Згідно з отриманими результатами цей підхід більш ефективний.

У роботі [18] було запропоновано модифікований метод DD, з допомогою якого визначалися області-зразки, відповідні конкретним ключовим словам. При анотуванні нового зображення воно поділяється на області, кожному з яких ставиться у відповідність найближча область-зразок та асоційоване з нею ключове слово. Так здійснюється інструкція лише на рівні об'єктів.

Загалом можна відзначити, що методи класифікаційного підходу дозволяють швидко та з досить великою точністю визначити зображення або області в ряд заздалегідь відомих категорій. Однак для цього потрібно збалансована навчальна вибірка (кількість прикладів для кожної категорії має

бути порівнянно), створення якої найчастіше здійснюється вручну. Також у роботі [18] показано, що зі збільшенням кількості категорій (ключових слів) точність класифікація значно знижується. Крім того, класифікаційні методи мають низьку масштабованість: щоразу при додаванні нових категорій або навчальних зображень необхідно навчати систему класифікації заново, що потребує значних обчислювальних витрат.

1.2. Пошукові методи

Пошукові методи автоматичного анотування зображень засновані на припущенні, що візуально схожі зображення мають анотуватись однаковими ключовими словами. Для нового зображення визначається набір візуально схожих зображень, які вже мають текстовий опис, після чого інструкція формується на основі значень схожість між зображеннями. Розглянемо докладніше основні пошукові методи.

Метод Joint Equal Contribution. У роботі [12] вперше запропоновано розглядати автоматичне анотування нового зображення як проблему пошуку візуально схожих навчальних зображень (найближчих сусідів). Для цього кожне зображення описується за допомогою 7 глобальних колірних та текстурних ознак, нормалізованих таким чином, щоб значення відстаней між парами ознак будь-яких двох зображень знаходилися в діапазоні [0; 1]. При порівнянні двох зображень спочатку окремо обчислюються відстані для кожного типу ознак, після чого отримані значення поєднуються з рівними вагами (JEC, Joint Equal Contribution).

Метод Tag Propagation. У роботі [16] запропоновано метод Tag Propagation, у якому на етапі навчання обчислюються ваги значимості окремих типів низькорівневих ознак зображень.

Таким чином, у пошукових методах відбувається процес анотування нового зображення можна розділити на два етапи: пошук невеликої кількості візуально схожих навчальних зображень і вибір їх ключових слів в якість анотації нового зображення. Завдяки цьому більшість пошукових методів анотування не вимагає списку заздалегідь відомих ключових слів, а також здатно продовжувати роботу при додаванні нових навчаючих зображень без повторної системи навчання. Також в якості навчальних наборів пошукові методи можуть напряму використовувати спеціалізовані веб-сайти, в яких розміщені фотоматеріали з користувацькими ключовими словами.

1.3. Порівняння методів автоматичного анотування зображень

Зазвичай для порівняння використовуються різні методи тестові бази зображень, наприклад:

– IAPR TC-12 містить 19627 фотографій різних сцен, включаючи людей, тварин, міста, ландшафти, а також інші аспекти життя. Значно вміст кожної бази зображень описано кількома пропозиціями, з яких автори статті вибирали створювані за допомогою програми TreeTagger. Ці іменники надалі використовувалися як ключові слова.Таким чином, був отриманий словник з 291 ключового слова;

– ESP Game містить 20770 зображень, які включають як фотографії, так і зображення з штучною графікою (анімаційні картинки, логотипи і т.п.). Ці зображення отримані за допомогою гри ESP, запропонованої в роботі [11]. У цій грі два гравця незалежно один від одного від одного присвоюють одному і тому ж зображенню ключове слово, якщо слова співпадають, то воно приймається в якості анотації зображення.

Загалом у базі використано 269 ключових слів. Автори роботи розділили зображення баз на навчальні та тестові вибірки із співвідношеннями 90% та 10% відповідно. Надалі такі вибірки використовувалися іншими дослідниками при публікації результатів роботи методів анотування. При цьому оцінка ефективності полягає у обчисленні середньої точності та повноти анотування, обчислених для ключових слів, а також підрахунку кількості використаних при анотуванні ключових слів. У таблиці 1.2 наведено опубліковані результати анотування деяких із розглянутих методів.

Як видно з таблиці 1.2, найкращі результати за точністю та кількістю використаних при анотуванні ключових слів показує метод SVM-DMBRM, тоді як метод TagProp-σML демонструє найкращі показники повноти для зазначених вище тестових баз. Обидва методи мають низький ступінь масштабованості, вимагаючи повторного навчання системи при додаванні нових ключових слів, а демонстровані найкращі результати анотування є недостатніми, що свідчить про необхідність подальшого розвитку методів.

Таблиця 1.2

Порівняльні оцінки ефективності методів анотування зображень

Метод						
	Точність, %	Повнота, %	N+	Точність, %	Повнота, %	N+
MBRM	24	23	223	18	19	209
JEC	28	29	250	22	25	224
TagProp-ML	48	25	227	49	20	213
TagProp-σML	46	35	266	39	27	239
2PKNN	49	32	274	51	23	245
SVM-DMBRM	56	29	283	55	25	259

1.4. Аналіз методів кластеризації даних

У розглянутому раніше методі автоматичного анотування зображень 2PKNN набір навчальних зображень запропоновано розділяти на семантичні групи, кожна з яких використовується для анотування нового зображення. Подібний поділ вибірки дозволило підвищити точність анотування, що свідчить про необхідності попереднього структурування навчального набору.

Одним із можливих способів структурування зображень є їх кластеризація за текстовим описом та візуальними ознаками. Розглянемо докладніше деякі з існуючих методів кластеризації.

Ієрархічні методи. Ієрархічні методи структурують вибірку векторів як системи вкладених розбиття, формуючи дерево кластерів, корінням якого є вся вибірка, а листям – окремі вектори.

Виділяють два основні типи методів ієрархічної кластеризації:

1. Східні алгоритми, що працюють за принципом «згори донизу». Спочатку всі вектори поміщаються в один кластер, який потім поділяється на все дрібніші кластери.

2. Висхідні алгоритми, що працюють за принципом «знизу нагору». На початку роботи кожен вектор переміщається в окремий кластер, після чого кластери об'єднуються у все більші до тих пір, поки всі вектори вибірки не будуть утримуватися в одному кластері. Даний тип методів є найпоширенішим.

В обох типах алгоритмів при ухваленні рішення про злиття (поділ) кластерів використовуються наступні способи обчислення відстаней між кластерами:

1. Метод одиночного зв'язку (відстань найближчого сусіда). Відстань між двома кластерами визначається як відстань між двома найбільш близькими векторами цих кластерів.

2. Метод повного зв'язку (відстань далекого сусіда). У цьому методі відстань між кластерами визначається найбільшою відстанню між будь-якими двома векторами цих кластерів.

3. Метод середнього зв'язку. Відстань між двома різними кластерами обчислюється як середня відстань між усіма парами векторів цих кластерів.

4. Центроїдний метод. У цьому методі відстань між двома кластерами визначається як відстань між їхніми центрами мас.

Ієрархічні методи кластеризації представляють вибірку векторів у вигляді деревоподібної системи кластерів, що в контексті завдання анотування дозволяє використовувати вибірку навчальних зображень на кількох рівнях деталізації. Однак ієрархічні методи мають високу обчислювальну складність, через що не застосовуються для вибірок, розмір яких більше 10000 зображень.

1.5. Аналіз відомих програмних систем для анотування зображень

На сьогоднішній день існує кілька програмних систем для автоматичного анотування зображень. Розглянемо деякі з них. Дослідницький програмний продукт "PiXiT", розроблений компанією PEPITe на основі роботи [15], призначений для класифікації зображень. Для того, щоб скористатися ним, необхідно написати розробникам лист із зазначенням найменування вашої організації та коротким описом проблеми, з якою збираєтеся працювати. Також необхідно додатково повідомити про результати проведені досліджень. Програма передбачає наявність навчальних зображень, з яких випадково витягується велика кількість блоків. З використанням цих блоків навчається ансамбль дерев рішень, за допомогою якого новим зображенням автоматично присвоюється одній мітці категорій (рисунок 1.1). Крім налаштувань за замовчуванням, користувачу надається

можливість налаштувати параметри розбиття зображень на блоки, отримання ознак та навчання класифікатора.

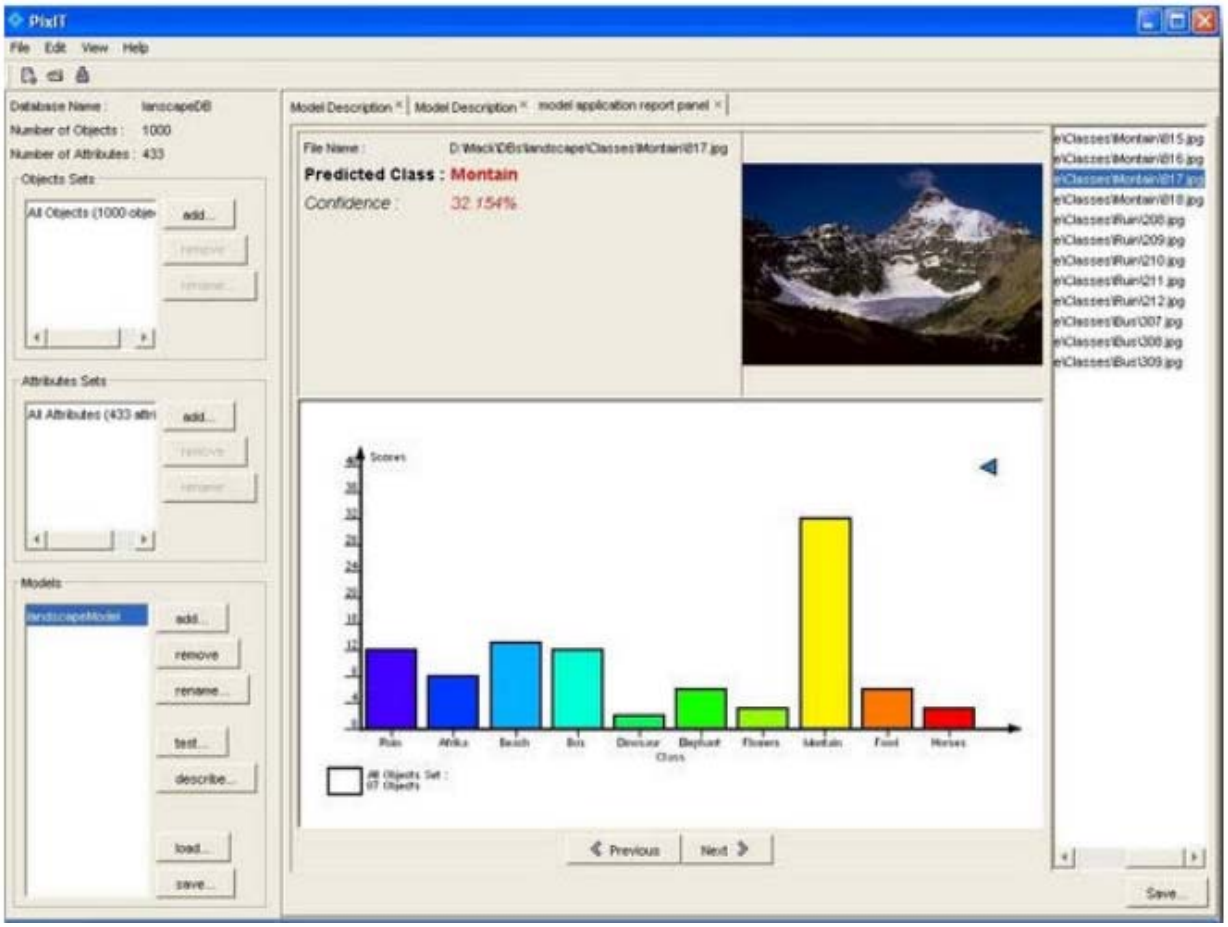

Рис. 1.1. Інтерфейс програми “PiXiT”

Програма "ImageTagger", розроблена компанією Attrasoft, є комерційним продуктом, що дозволяє анотувати зображення декількома ключовими словами. Для використання програми необхідно попереднє навчання, що вимагає від 1000 до 10000 тренувальних зображень для кожного ключового слова.

При цьому розмір словника ключових слів має бути невеликим (не більше 10 – 20 слів). Інтерфейс програми з прикладом анотування зображення представлений на рисунку 1.2.

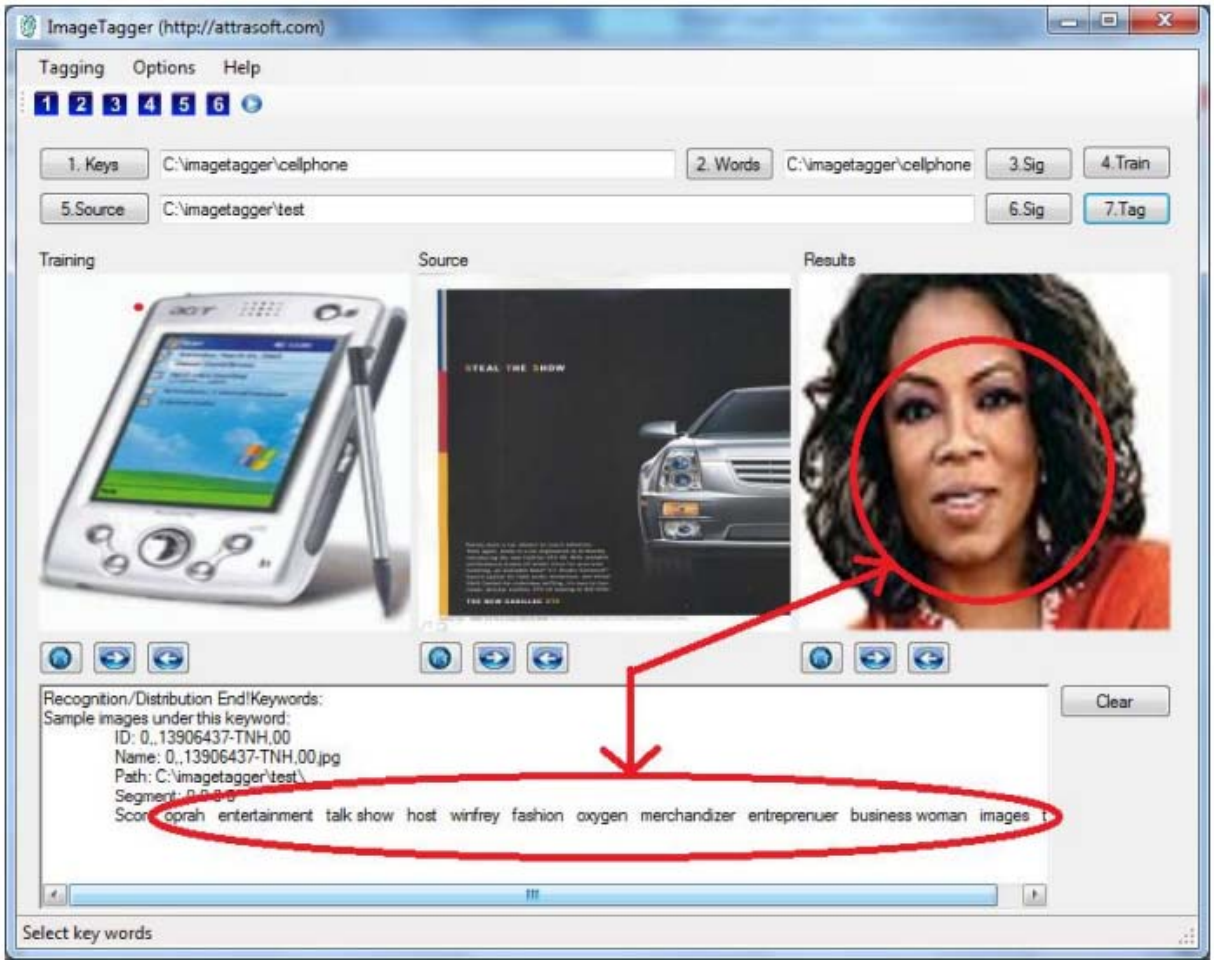

Рис. 1.2. Інтерфейс програми «ImageTagger»

Дослідницький веб-сервіс "MUFIN Image Annotation", розроблений лабораторією DISA (Laboratory of Data Intensive Systems and Applications), реалізує пошуковий метод анотування. Для завантажуваного зображення визначається набір візуально схожих зображень серед мільйонів зображень фотобанку Profimedia.

Після цього ключові слова зображень набору кластеризуються за семантичним значенням та зважуються на підставі візуальної подібності зображень та семантичної схожості ключових слів. Як анотації вибирається 20 ключових слів із найбільшими вагами (рисунок 1.3).

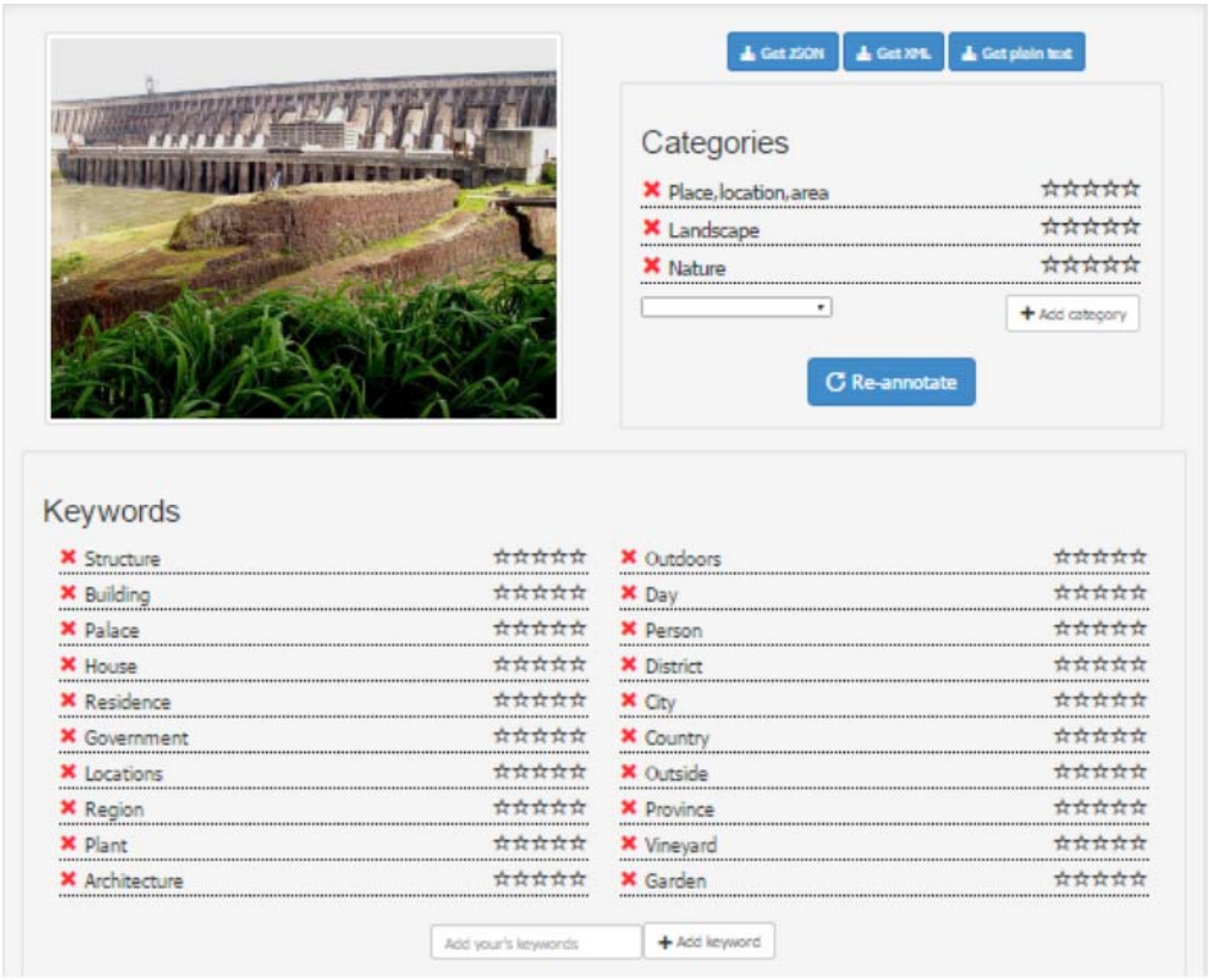

Рис. 1.3. Приклад анотування зображення у веб-застосунку «MUFIN Image Annotation»

Комерційний веб-сервіс «Imagga» включає API, що надає доступ до автоматичного анотування зображень. У відповідь на запит користувача, що включає ідентифікаційну інформацію та зображення, API надає список, що складається з пар ключове слово – рівень довіри (рис. 1.4).

Розробниками системи рекомендується фільтрувати отриманий список за рівнем довіри 30 %, що дозволяє скоротити кількість нерелевантних ключових слів, проте нерідко залишає в інструкції лише 2-3 слова.

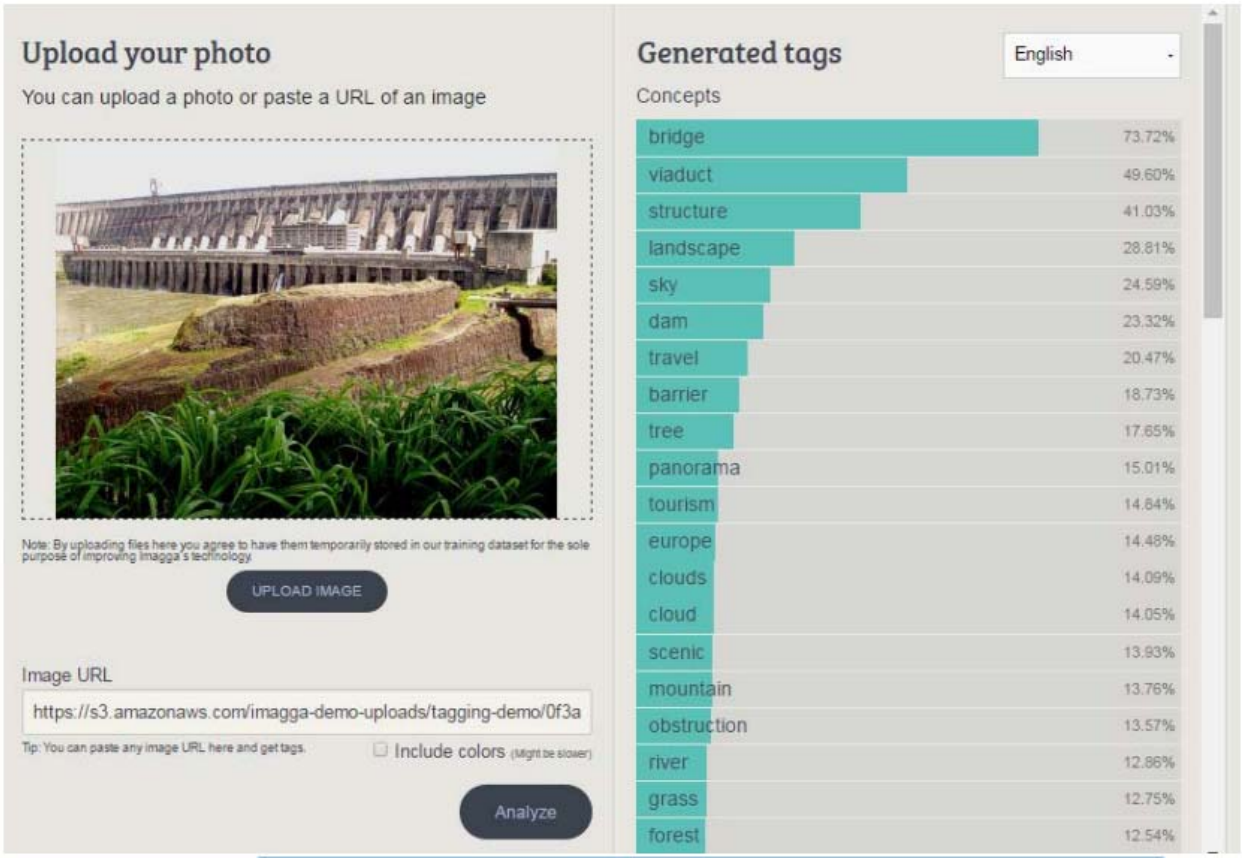

Рис. 1.4. Приклад анотування зображення на веб-сервісі “Imagga”

Безкоштовний веб-сервіс «Google Photos», розроблений та підтримуваний Google, призначений для завантаження, обробки та зберігання фотографій користувача. Відмінною рисою сервісу є використання нейронної мережі, що дозволяє розпізнавати частину найбільш помітних об'єктів, завдяки чому користувачам надається можливість пошуку фотографій за ключовими словами. Однак слід відзначити, що система не відображає надані ключові слова. Це ускладнює пошук фотографій, оскільки використовувані ключові слова не завжди очевидні. Також система не завжди розпізнає деякі об'єкти на майже схожі зображення (рис. 1.5).

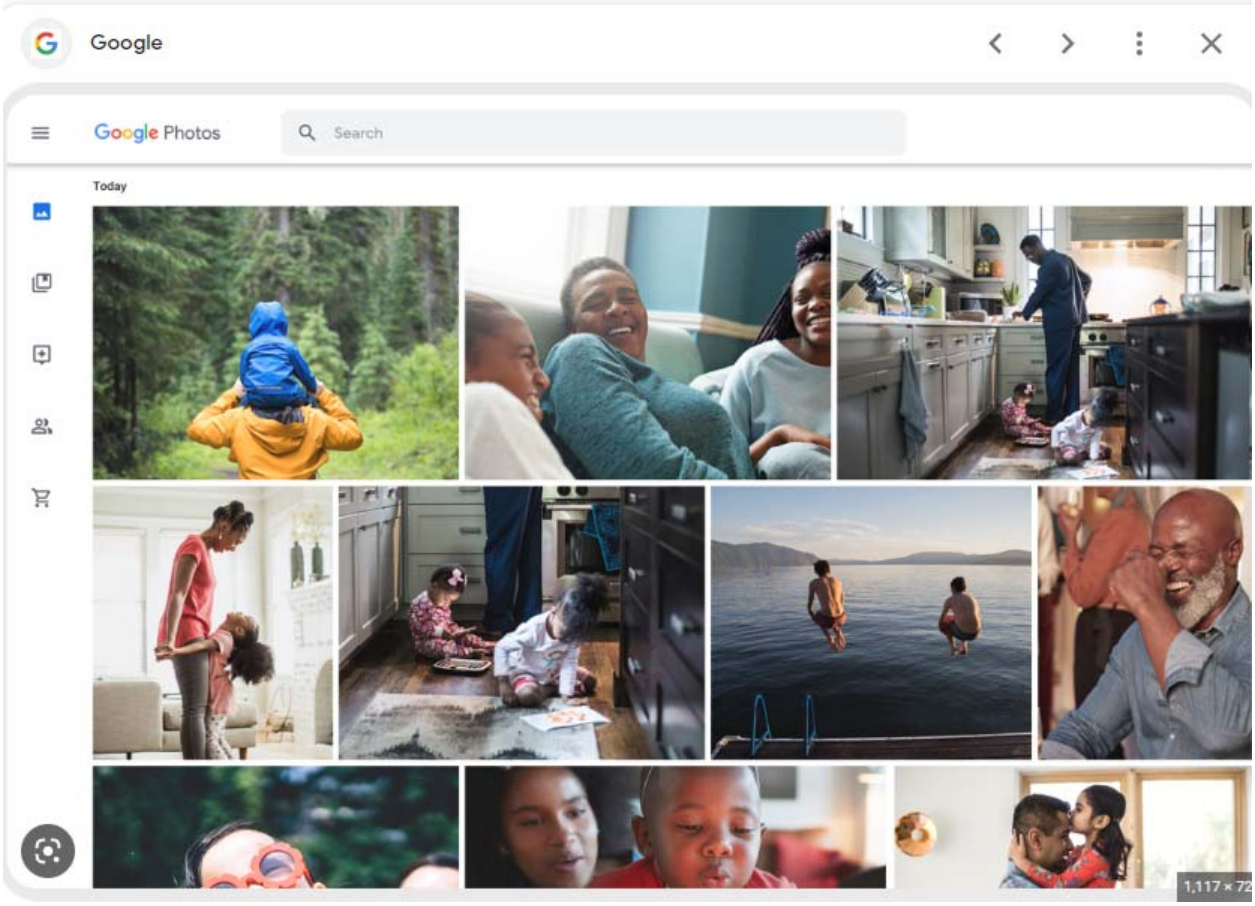

Рис. 1.5. Приклад анотування зображення на веб-сервісі “Google Photos”

Таким чином, незважаючи на наявність комерційних програмних продуктів, існуючі системи розпізнавання мають досить низькі показники точності та повноти анотування, у зв'язку з чим потрібний подальший розвиток методів автоматичного анотування зображень.

Висновки до першого розділу

1. Проведено огляд відомих методів автоматичного анотування зображень, кластеризації даних та опис зображень за допомогою

низькорівневих ознак, наведена їх класифікація. Також розглянуто низку програмних систем, що реалізують автоматичне абстрактне зображення.

2. Відомі методи анотування можна розділити на три підходи: класифікаційний, генеративний та пошуковий. Класифікаційні методи представляють ключові слова у вигляді незалежних класів, приклади яких навчається класифікатор. Це дозволяє швидко присвоїти зображенням або їх областям мітки категорій, проте для досягнення достатньої точності необхідна збалансована навчальна вибірка. Також збільшення кількості категорій (ключових слів) призводить до значного зниження точності класифікації.

РОЗДІЛ 2
МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ СИСТЕМИ АНОТУВАННЯ ЗОБРАЖЕНЬ В ПОШУКОВИХ СИСТЕМАХ

2.1. Обчислення глобального візуального дескриптора

Аналіз існуючих підходів та методів анотування зображено показав доцільність анотування нових зображень на основі навчальних зображень, найбільш схожих візуально, а також необхідність попереднього структурування навчального набору зображень.

Складність у тому, що у навчальному наборі кожне ключове слово відноситься до всього зображення, а не окремим об'єктам, крім того анотаціях можуть бути відсутні релевантні ключові слова.

Обмеження та умови, що пред'являються до навчального набору та анотованих зображень, представлені в таблиці 2.1.

Таблиця 2.1

Обмеження та умови, що пред’являються до навчального набору та анотувані зображень

Критерій	Обмеження
Тип зображень	Фотографії
Розмір зображень	Не менше 480 × 360 пікселів
Розмір анотованих об’єктів	Не менше 5% площі зображення
Якість зображень	Рівномірно освітлені, контрастні, без розмиття, рівень шуму – не більше 2-3 дБ
Частота ключових слів	Не менше 0,005

Припустимо, що навчальний набір TS складається з зображень і відповідних їм текстових описів. Нехай $J = \{I_1, \dots, I_m\}$ – колекція зображень, а $K = \{K_1, \dots, K_m\}$ – словник, що складається з N ключових слів, тоді навчальний набір $TS = \{(I_1, K_1), \dots, (I_M, K_M)\}$, де $K_M \subseteq K$.

Також припустимо, що навчальний набір розділений на кілька однорідних текстово-візуальних груп (ОТВ-груп), що не перетинаються, $H = \{H_1, \dots, H_l\}$, а вибір ключових слів у процесі анотування тестового зображення A залежить від асоціації зображення з тією чи іншою групою.

Позначимо ймовірність приналежності зображення A ОТВ-групі H_l як $P(A|H_l)$. Також введемо ймовірність $P_l(A|k_n)$ для оцінки належності зображення A семантичній групі, сформованій із зображень ОТВ-групи H_l, що мають в описі ключове слово k_n. В цьому випадку анотування зображення моделюється як проблема пошуку апостеріорних ймовірностей:

$$P(k_n|A) = \frac{\sum_{H_l \in H}[P(H_l)P(A|H_l)P_l(A|k_n)P_l(k_n)]}{P(A)}, \qquad (2.1)$$

де $P(H_l)$ – апріорна ймовірність ОТВ-групи H_l; $P_l(k_n)$– апріорна ймовірність ключового слова k_n усередині ОТВ-групи H_l; $P(A)$ – апріорна ймовірність тестового зображення A.

Оскільки апріорна ймовірність $P(A)$ є константою, то будемо аналізувати чисельник:

$$P(k_n|A) \approx \sum_{H_l \in H}[P(H_l)P(A|H_l)P_l(A|k_n)P_l(k_n)], \qquad (2.2)$$

Використовуючи отримані значення $P(k_n|A)$, ключові слова ранжуються по спаданню. Як анотації використовується N_{kw} ключових слів з найбільшою ймовірністю. Таким чином, для анотування тестового зображення A необхідно оцінити ймовірності $P(H_l), P(A|H_l), P_l(A|k_n)$ та $P_l(k_n)$. Пропонований для вирішення цього завдання алгоритм

автоматичного анотування зображень на основі однорідних текстово-візуальних груп можна розділити на три етапи навчання та етап анотування:

I етап. Обчислення глобального візуального дескриптора:

- швидке обчислення набору локальних дескрипторів;

- обчислення колірних локальних дескрипторів;

- кодування набору локальних дескрипторів.

II етап. Створення текстового дескриптора:

- формування текстового дескриптора;

– відновлення пропущених ключових слів навчальних зображень.

III етап. Формування однорідних текстово-візуальних груп:

– первинний поділ зображень на основі спільної появи ключових слів;

– кластеризація зображень із використанням текстово-візуальних дескрипторів.

IV етап. Автоматичне анотування зображень.

Далі розглянемо докладніше алгоритмічну реалізацію кожного з поданих етапів.

2.2. Обчислення глобального візуального дескриптора

На першому етапі автоматичного анотування зображень для кожного зображення I з колекції навчальних зображень $\boldsymbol{J} = \{\boldsymbol{I_1}, \dots, \boldsymbol{I_m}\}$, а також набору анотованих зображень, обчислюється глобальний візуальний дескриптор $\boldsymbol{V} = \{\boldsymbol{V_1}, \dots, \boldsymbol{V_z}\}$/ Для цього із зображення витягується набір локальних дескрипторів, який кодується за допомогою словника візуальних слів в один глобальний дескриптор. При цьому обчислення локальних дескрипторів з використанням регулярної сітки ефективніше в порівнянні з іншими детекторами, а збільшення перетину областей, на яких обчислюються дескриптори, підвищує точність інструкції. Однак це

призводить до значного збільшення обчислювальних витрат. Для вирішення цієї проблеми запропоновано метод швидкого вилучення набору локальних дескрипторів, що описують усі частини зображення, а також алгоритм для їх кодування у глобальний візуальний дескриптор.

Пропонований алгоритм швидкого обчислення набору локальних дескрипторів (Fast Dense Speeded-Up Features - FD-SUF) [8] заснований на локальному дескрипторі SURF та складається з двох етапів: обчислення матриці частин дескрипторів $\boldsymbol{M}$, в яких кожен елемент $\boldsymbol{M_{rx,ry}}$ є вектором з 4 чисел, та побудови з її допомогою набору локальних дескрипторів (рисунок 2.1).

На першому етапі вихідне зображення $\boldsymbol{I}$ (рис. 2.2 а) перекладається з колірного простору RGB у відтінки сірого (рис. 2.2, б) для чого використовується компонента $\boldsymbol{Y}$ із колірної схеми $\boldsymbol{YUV}$:

$$Y = 0{,}299 \cdot R + 0{,}587 \cdot G + 0{,}114 \cdot B \qquad (2.3)$$

Отримане зображення розділяється сіткою на блоки$I_{rx,ry}$ розміром $5\sigma_{sc} \times 5\sigma_{sc}$, де σ_{sc}– масштаб (для зображень розміром 480 × 360 дорівнює 1) (рис. 2.2, в). Масштаб підбирається пропорційно до розміру зображень таким чином, щоб розміри матриць M, обчислених для будь-яких двох зображень, були зіставні. На наступному кроці в кожному блоці для 5×5 рівномірно розподілених точок обчислюються перші похідні

$$L_x(p,\sigma_{sc}) = I(p) * \frac{\partial}{\partial x} g(\sigma_{sc}),$$

$$L_y(p,\sigma_{sc}) = I(p) * \frac{\partial}{\partial y} g(\sigma_{sc}),$$

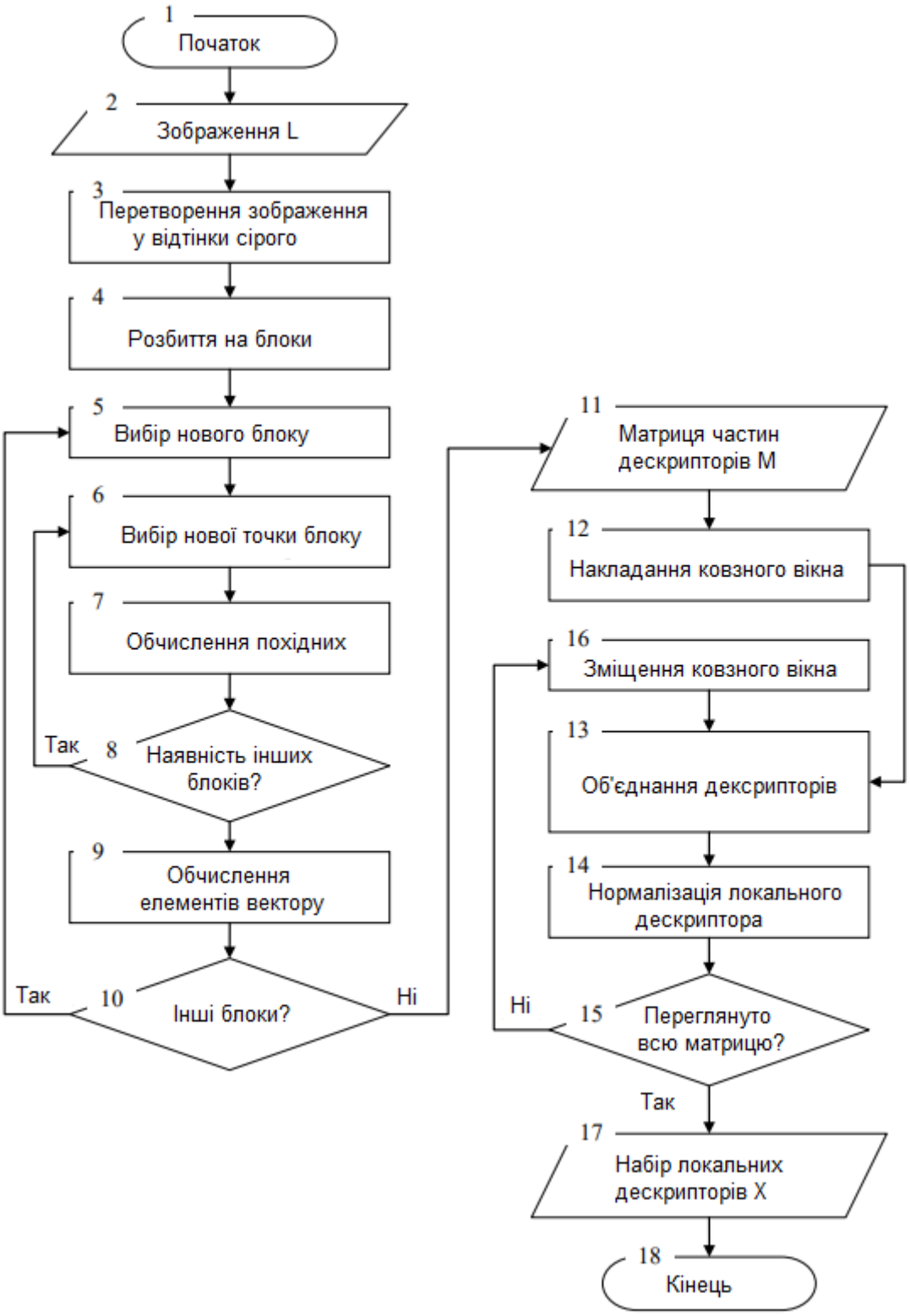

Рис. 2.1. Блок-схема алгоритму швидкого обчислення набору локальних дескрипторів

У запропонованому алгоритмі FD-SUF основні обчислювальні витрати потрібні на першому етапі, що включає два базових типу циклів – зовнішній

цикл за зображенням розміром $(w/5\sigma_{sc}) \times (h/5\sigma_{sc})$ та вкладені цикли по блоках, де (w і h – ширина та висота зображення. Таким чином, підвищити ефективність обчислень можна за допомогою розпаралелювання зовнішнього циклу [9].

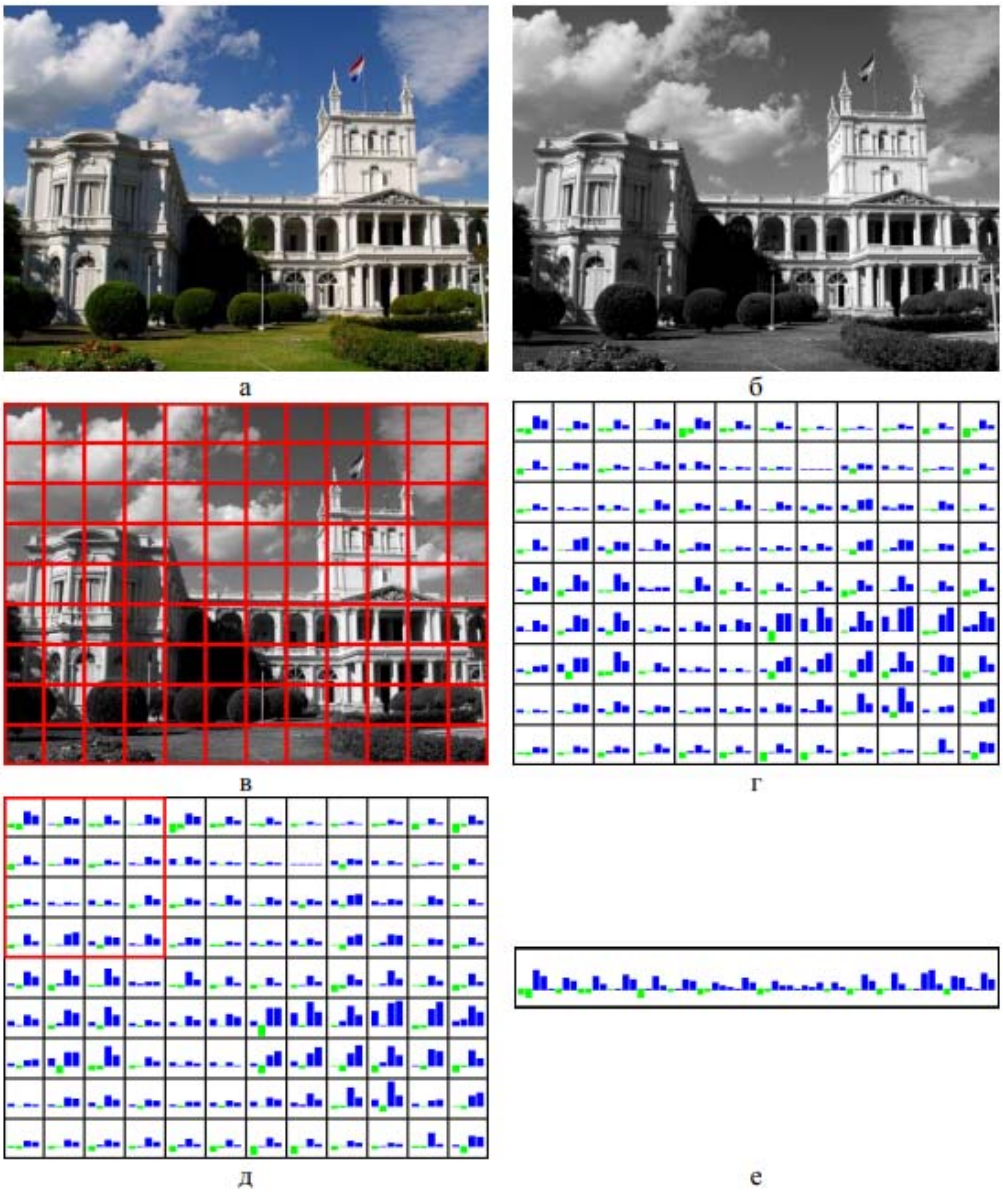

Рис. 2.2. Ілюстрація обчислення локального дескриптора для зображення 58 з бази IAPR TC-12: а) вихідне зображення; б) приведення зображення до відтінки сірого; в) розбиття на блоки із масштабом $\sigma_{sc} = 8$; г) візуалізація матриці частин дескрипторів (синім кольором позначені позитивні значення, зеленим – негативні); д) накладання ковзного вікна на матрицю частин дескрипторів; е) візуалізація сформованого локального дескриптора

Для цього зображення поділяється на смуги, обробка яких розподіляється між ядрами процесора обчислювальної системи. Для реалізації таких паралельних алгоритмів широкого поширення набув стандарт OpenMP для розпаралелювання програм мовами Сі, Сі++ та Фортран. Крім нього останнім часом розвивається розширення мови Сі ++, відоме як Intel Cilk Plus.

2.3. Обчислення локальних дескрипторів кольорів

Базовий алгоритм FD-SUF обчислюється тільки на зображеннях відтінках сірого, у зв'язку з чим схильний до сильного впливу умов освітленості, а також не враховує колірну інформацію. Для вирішення цієї проблеми запропоновано формувати колірні локальні дескриптори: дескриптори FD-SUF обчислюються для кожної компоненти колірного простору, після чого вони поєднуються (рисунок 2.3)

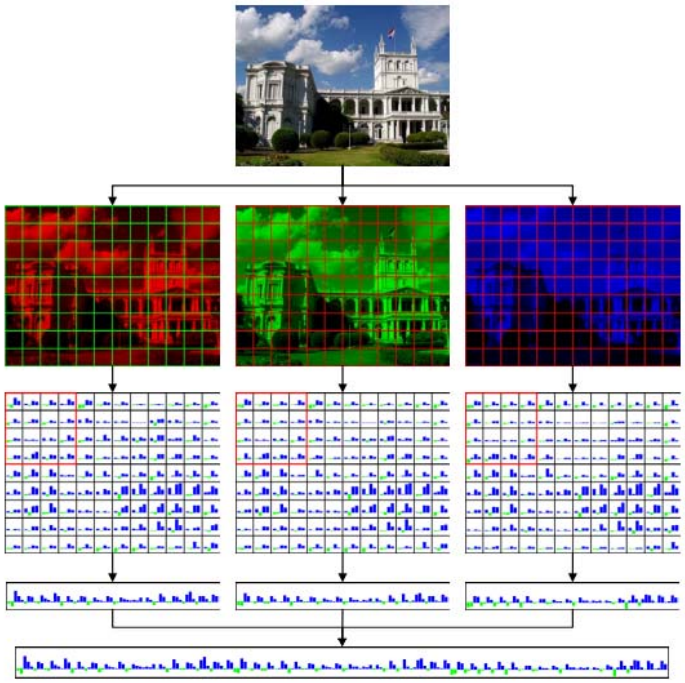

Рис. 2.3. Ілюстрація обчислення колірного локального дескриптора для зображення 58 з бази IAPR TC-12 у колірному просторі RGB

У дослідженні були використані такі поширені колірні гами, як HSV (Hue, Saturation, Value) та LUV (Lightness, Uniform chromaticity scale, Valence), а також простори, що володіють деяким ступенем інваріантності до зміни інтенсивності світла.

2.4. Метод створення текстового дескриптора

На другому етапі навчання системи анотування зображень для кожного зображення Im, яке належить навчальній множині $TS = \{(I_1, K_1)$, …,$(I_M, K_M)\}$, необхідно створити текстовий дескриптор $Tm = \{t_1, t_N\}$ довжина якого дорівнює розміру словника ключових слів. Для цього використовується як текстовий опис зображення K_m, так і частота появи кожного ключового слова в навчальній вибірці. Однак у зв'язку з тим, що інструкції у навчальних вибірках часто формуються вручну, то деякі зображення можуть бути проанотовані не всіма ключовими релевантними словами. Для вирішення цієї проблеми запропоновано метод відновлення ключових слів, який визначає кількість пропущених слів.

Оскільки деякі ключові слова є «спільними» (зустрічаються в описі великої кількості зображень різних категорій), то вони є менш корисними для опису зображень. У зв'язку з цим, при формуванні дескриптора вони отримують менше числове значення, ніж «характерні» ключові слова (ключові слова, що зустрічаються в описі невеликої кількості зображень) (рис. 2.5).

Для цього елементи текстового дескриптора обчислюються за допомогою статистичного зпідходу TF-IDF:

$$t_n^m = \frac{\partial(k_n \in K_m)}{|K_m|} \cdot log\left(\frac{M}{F(k_n)}\right), \qquad (2.4)$$

де $\partial(k_n \in K_m)$ позначає наявність/відсутність ключового слова k_n в описі зображення Im (приймає значення 1 і 0 відповідно); $|K_m|$ - кількість ключових слів в описі зображення Im; M – розмір навчальної вибірки; $F(k_n)$ – частота появи ключового слова k_n в навчальній вибірці.

При порівнянні двох зображень за допомогою їх текстових дескрипторів використовується косинусна метрика:

$$D_T(\mathbf{T}_i, \mathbf{T}_j) = \frac{\sum_{n=1}^{N} t_n^i \cdot t_n^j}{\sqrt{\sum_{n=1}^{N} (t_n^i)^2} \cdot \sqrt{\sum_{n=1}^{N} (t_n^j)^2}}.$$

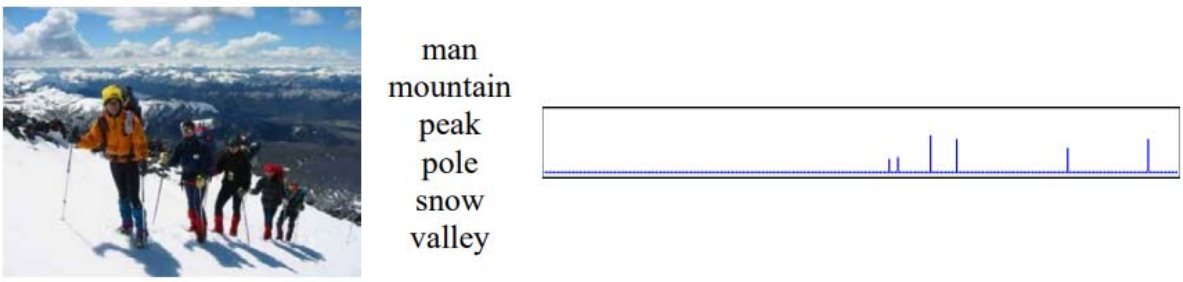

Рис. 2.4. Ілюстрація текстового дескриптора, сформованого для зображення 2731 з бази IAPR TC-12

В отриманих текстових дескрипторах деякі релевантні ключові слова можуть бути відсутніми. Пов'язано це з тим, що при складанні анотацій навчальних зображень вручну частина «очевидних» ключових слів часто пропускається. Наприклад, на рисунку 2.5 можна помітити, що ключове слово landscape підходить для всіх чотирьох зображень, але зустрічається в інструкції лише одного з них.

Рис. 2.5. Приклад зображень з бази IAPR TC-12 та їх анотацій

Анотація зображення Im поповнюється $Y(Im)$ ключовими словами з найбільшими значеннями ймовірності $P(kn|Im)$ (рисунок 2.6). У разі, якщо $Y(Im)$ негативне, інструкція залишається без змін.

Після обчислення для кожного зображення Im візуального Vm та текстового Tm дескрипторів, вони об'єднуються в один текстово-візуальний VTM. За його допомогою навчальний набір зображень пропонується розділяти на однорідні текстово-візуальні групи. Ідея полягає в тому, що навчальні зображення однієї ОТВ-групи формують контекст для анотованого зображення, іншими словами, якщо зображення віднесено до будь-якої групи, то воно анотується з обмеженого набору ключових слів цієї групи. Також передбачається, що тестове зображення може належати кільком ОТВ-групам, проте їх кількість обмежується візуальною подібністю.

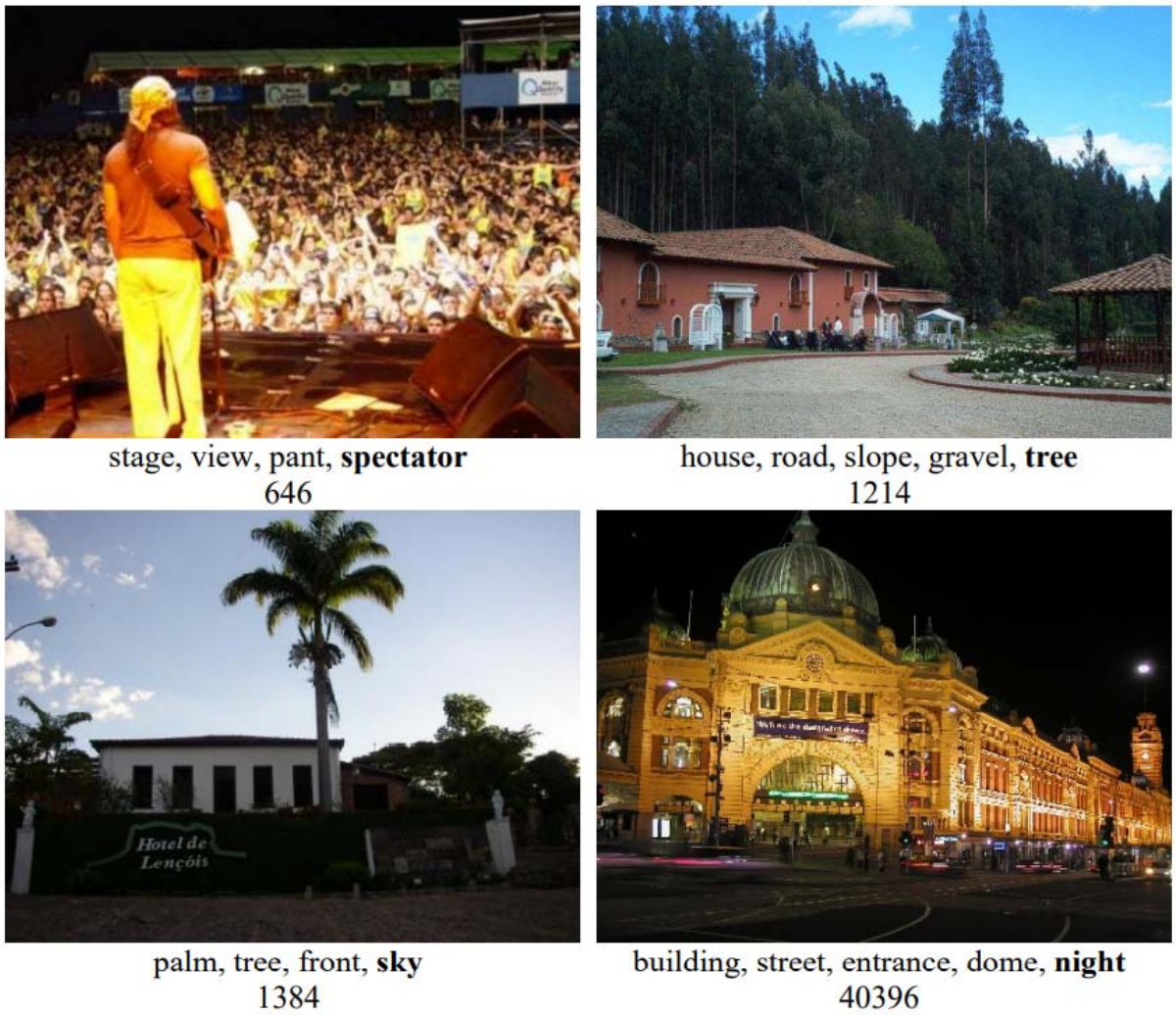

Рис. 2.6. Приклад зображень із бази IAPR TC-12 з відновленими ключовими словами (виділені напівжирним шрифтом)

Це дозволяє відсіяти свідомо нерелевантні ключові слова, не втративши релевантних, а також знизити кількість навчальних зображень, що беруть участь в анотуванні. Для цього кожна з ОТВ-груп має відповідати двом умовам:

1. Усі зображення групи у своїх анотаціях мають спільні "характерні" ключові слова.

2. Зображення групи мають суттєву візуальну подібність.

Це завдання вирішується у два етапи:

1. Проводиться первинний поділ зображень на групи на основі спільної зустрічальності ключових слів в описах зображень.

2. Зображення кластеризуються в автоматичне визначення кількість ОТВ-груп, використовуючи текстово-візуальні дескриптори.

Слід зазначити, що ефективність запропонованого методу підвищується, якщо разом із тестовим зображенням будуть надані деякі ключові слова, отримані від користувачів або за допомогою вузькоспеціалізованих класифікаторів. В цьому випадку при оцінці ймовірностей використовується текстово візуальний дескриптор.

Висновки до другого розділу

1. Наведено метод автоматичного анотування зображень на основі навчального набору зображень, розділеного на однорідні текстово-візуальні групи, а також запропоновано алгоритм для реалізації даного методу, який відрізняється тим, що анотування нового зображення здійснюється за допомогою навчальних зображень невеликої кількості візуально схожих груп. Даний алгоритм включає три етапи навчання та етап анотування.

2. На першому етапі для всіх навчальних, а також анотованих зображень утворюється глобальний візуальний дескриптор. Для цього з зображення витягується набір локальних дескрипторів, який кодується за допомогою словника візуальних слів. Оскільки цей етап є найбільш обчислювально витратним, то запропоновано швидкий метод вилучення набору локальних дескрипторів, що дозволяє уникнути повторних обчислень при накладенні областей розрахунку дескрипторів, що істотно прискорює процес анотування. Також розглянуто процес обчислення колірних локальних дескрипторів, використання яких дозволяє підвищити точність анотування, та наведено алгоритми формування словника візуальних слів та кодування набору локальних дескрипторів у глобальний візуальний дескриптор.

РОЗДІЛ 3
ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ СИСТЕМИ АНОТУВАННЯ ЗОБРАЖЕНЬ В ПОШУКОВИХ СИСТЕМАХ

3.1. Загальна архітектура системи

Система являє собою модульний додаток, програмні модулі якого можуть бути використані як у сукупності, так і окремо для вирішення більш вузьких завдань, наприклад, кластеризації даних з можливістю ітеративного уточнення кластерів, обчислення візуальних дескрипторів для категоризації зображень та розпізнавання образів. Найменування розроблених модулів та їх функціональні показники наведені в таблиці 3.1, а структурна схема зображена на рисунку 3.1.

Таблиця 3.1

Розроблені програмні модулі та їх призначення

Назва модуля	Функціональна характеристика
Модуль перетворень зображення	
Модуль обчислення візуальних дескрипторів	Здійснює обчислення набору локальних дескрипторів, формування словника візуальних слів та кодування глобального візуального дескриптора
Модуль формування текстового дескриптора	Здійснює обчислення частот народження ключових слів та формування текстового дескриптора
Модуль відновлення ключових слів	Модуль формує семантичні групи та розширює інструкції навчальних зображень шляхом відновлення пропущених ключових слів
Модуль формування навчальних груп	Здійснює первинний поділ навчальних зображень та кластеризацію текстово-візуальних дескрипторів
Ядро системи	Реалізує автоматичне анотування зображень, використовуючи інші модулі системи
Інтерфейс користувача	Здійснює взаємодію користувача з ядром системи

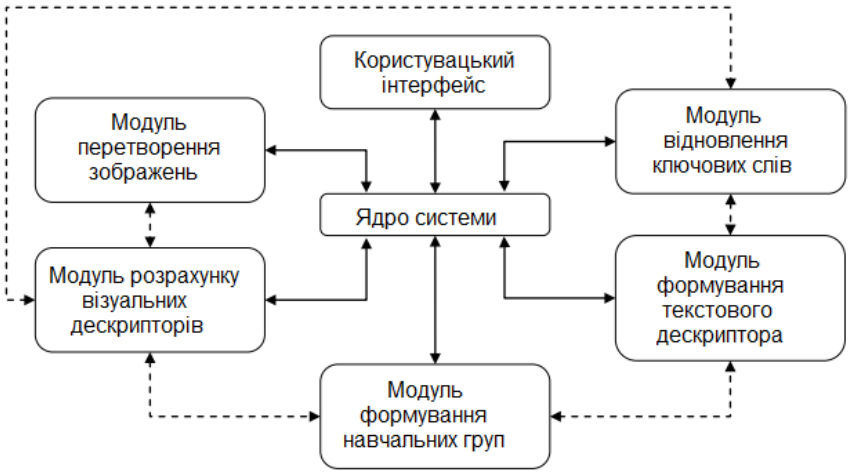

Рис. 3.1. Структурна схема експериментального програмного комплексу

Модуль перетворень зображення. Модуль забезпечує перетворення вхідних зображень на необхідні колірні простори та канали (RGB, nRGB, rg, Opponent, HSI, LUV, HSV, YUV). У кожному каналі допустима зміна значень наводиться до діапазону [0; 1], після чого обчислюються інтегральні зображення, що використовуються для обчислення локальних дескрипторів.

Модуль обчислення візуальних дескрипторів. Функції модуля обчислення візуальних дескрипторів реалізовані в програмному продукті «Система автоматичного формування візуальних слів», яка дозволяє здійснювати обчислення набору локальних дескрипторів, формування словника візуальних слів та кодування глобального візуального дескриптора.

Спрощена блок-схема алгоритмів модуля обчислення візуальних дескрипторів представлена на рисунку 3.2. На етапі навчання всі зображення навчальної колекції перетворюються на заданий колірний простір (блоки 2–3). Після цього з всієї колекції випадково вибирається 200 000 локальних дескрипторів, які використовуються для знаходження головних компонентів, а також формування словника візуальних слів (блоки 4-6). Потім з навчальної колекції випадковим чином вибирається 2048 зображень, для яких за допомогою візуальних слів обчислюються глобальні дескриптори (блок 7). Отримані дескриптори використовуються для знаходження основних

компонентів (блок 8). У випадку, коли задано декілька колірних просторів, блоки 2–6 виконуються для кожного колірного простору окремо.

Отримані візуальні слова та матриці перекладу у просторі основних компонентів використовуються для обчислень глобальних візуальних дескрипторів для всієї колекції навчальних зображень, а також анотованих зображень. Для цього кожне зображення також переводиться в заданий колірний простір (блок 13), після чого з зображення витягується набір локальних дескрипторів, розмірність кожного з яких зменшується (блоки 14–15). Використовуючи отримані локальні дескриптори, формується глобальний дескриптор, розмірність також скорочується (блоки 16–17). У випадку, коли задано декілька колірних просторів, блоки 13–15 виконуються для кожного колірного простору простору окремо.

Модуль формування текстового дескриптора. У модулі реалізовані функції для обчислення частот народження ключових слів у колекції навчальних зображень, а також формування текстових дескрипторів за допомогою статистичної величини TF-IDF.

Модуль відновлення ключових слів. У модулі реалізовані функції відновлення ключових слів навчальних зображень. Вихідними даними модуля є навчальний набір, у якому кожне зображення вже має візуальний та текстовий дескриптори.

Усі зображення розподіляються за семантичними групами на основі їх текстового опису, після чого зображення вибираються послідовно. Для кожного вибраного зображення (назвемо його вихідним) у кожній семантичній групі визначається по 2 візуально схожі навчальні зображення. Ці зображення поєднуються в набір, за допомогою якого оцінюється кількість пропущених ключових слів.

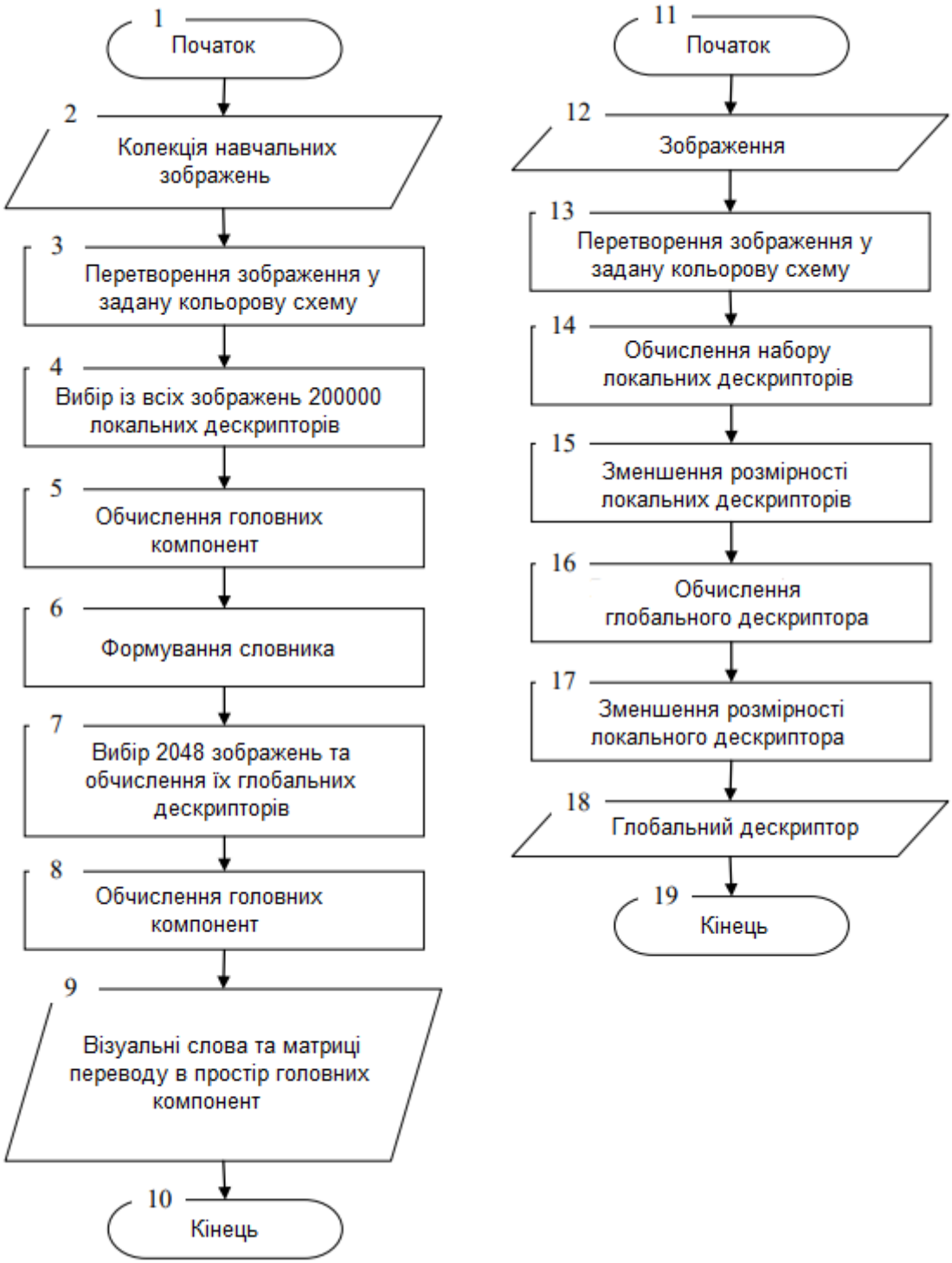

Рис. 3.2. Спрощена блок-схема алгоритмів модуля обчислення візуальних дескрипторів

Якщо отримане значення відсутніх ключових слів позитивне, то набір використовується для обчислення ймовірностей приналежності всіх ключових слів до початкового зображення. Після цього анотація вихідного

зображення поповнюється ключовими словами з найбільшими ймовірностями і алгоритм вибирає нове зображення.

Модуль ядра системи відповідає за взаємодію інших модулів програмного продукту між собою, а також реалізує розроблений алгоритм автоматичного анотування зображень.

Функції модулів ядра системи та формування текстових дескрипторів та навчальних груп реалізовані у програмному додатку «Система автоматичного анотування зображень», який реалізований з використанням мови програмування Java в середовищі NetBeans. Можливості налаштування модулів наведено на рисунку 3.3.

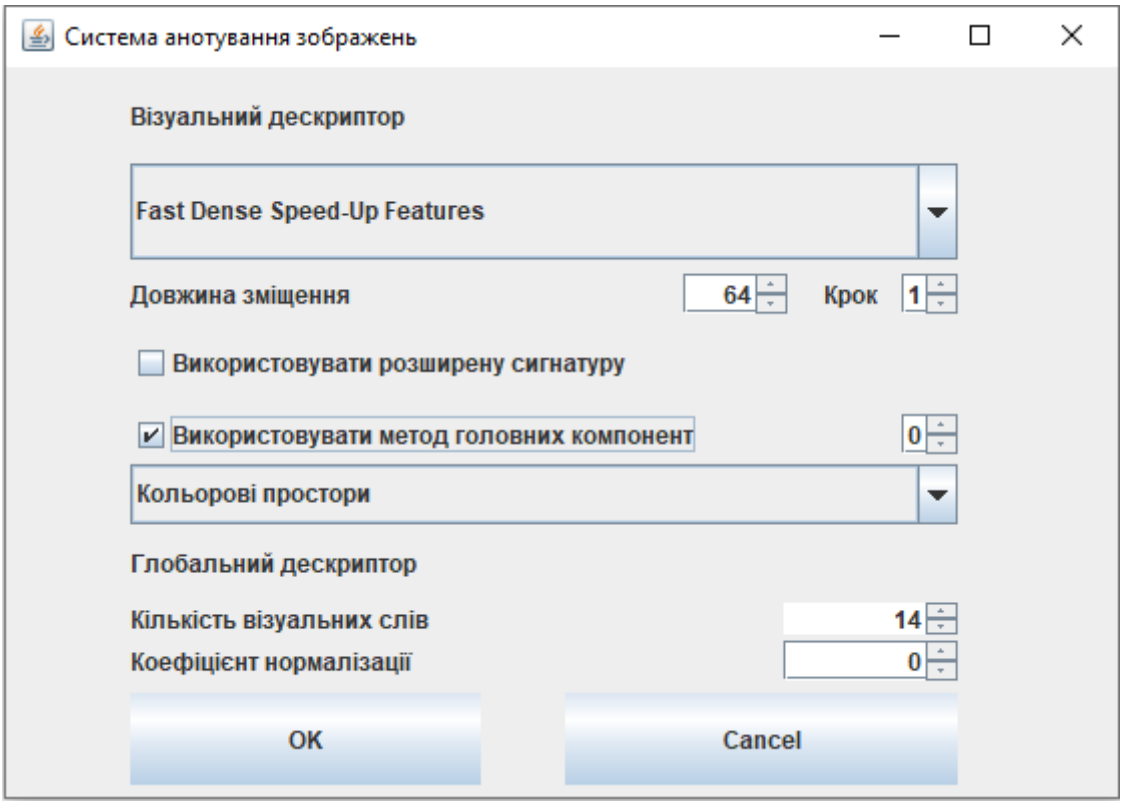

Рис. 3.2. Екранна форма налаштування модулів відновлення ключових слів, формування навчальних груп та ядра системи

Модуль інтерфейсу користувача дозволяє завантажувати в програму навчальні та тестові бази зображень, а також здійснювати налаштування параметрів реалізованих алгоритмів та оцінювати отримані результати. На

рисунку 3.3 представлено екранну форму головної форми системи анотування зображень.

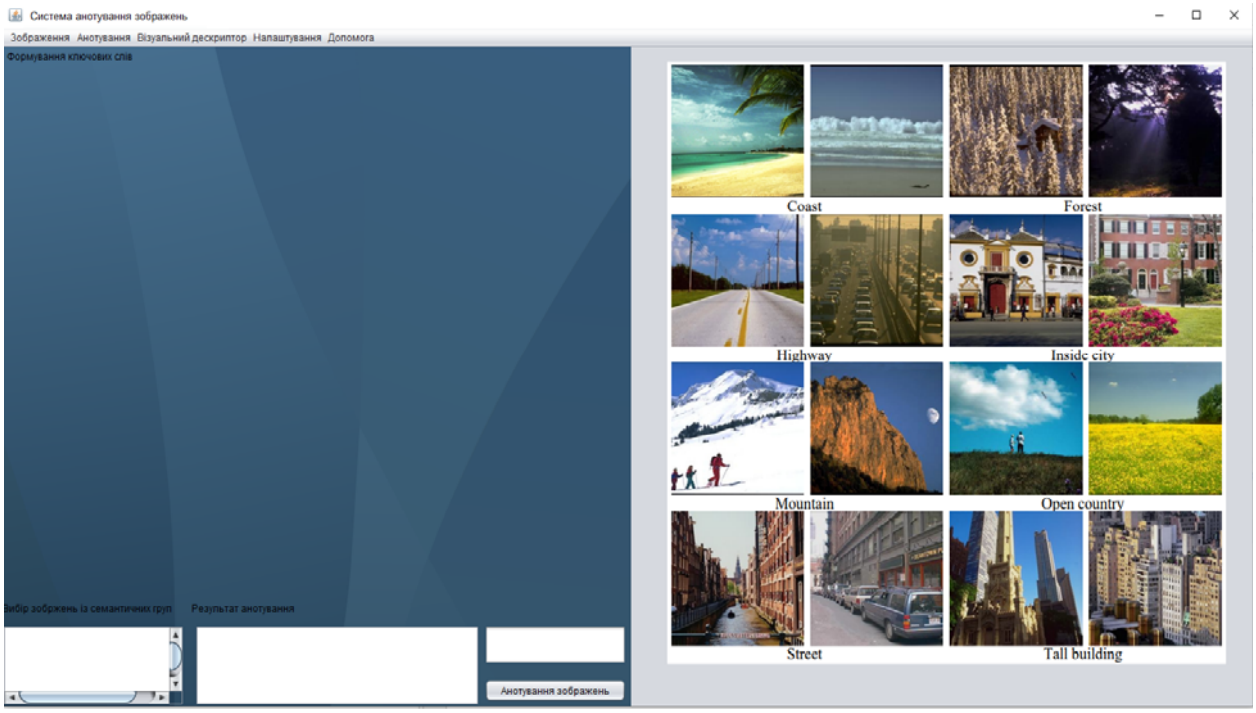

Рис. 3.3. Головна екранна форма системи анотування зображень в пошукових системах

Структура таблиць баз зображень описана у таблиці 3.2.

Таблиця 3.2

Структура таблиць бази даних анотованих зображень

Назва поля	Тип даних	Опис
Таблиця images містить навчальні, тестові або проанотовані зображення		
Id	Ключове поле	Ключове поле. Однозначно ідентифікує зображення
Name	Текстовий	Назва файлу зображення
Таблиця keywords містить список ключових слів		
Id	Ключове поле	Ключове поле. Однозначно ідентифікує ключове слово
Keyword	Тектовий	Назва ключового слова
Таблиця linking містить зв'язки між зображеннями та ключовими словами		
Id	Ключове поле	Ключове поле. Однозначно ідентифікує зв'язок ключового слова із зображенням
Image_ID	Числове	Ідентифікатор зображення
Keyword_ID	Числове	Ідентифікатор ключового слова

3.2. Результати експериментальних досліджень обчислення візуальних дескрипторів

Для перевірки запропонованих методів та алгоритмів обчислення візуальних дескрипторів використовувався набір зображень, що включає 8 категорій сцен (рисунок 3.4). Навчальна вибірка складається з 2688 зображень, розмір кожного зображення 256×256 пікселів. Тестування полягало у дослідженні впливу розміру словника візуальних слів та значення коефіцієнта нормалізації γ на точність категоризації зображень.

Додатково проведено дослідження для визначення оптимальної довжини глобального дескриптора Z, вибору найбільш інформативних колірних просторів та кількості потоків при паралельному обчисленні локальних дескрипторів. Для обчислення точності категоризації окремої категорії ctg використовувалася така формула:

$$accuracy_{ctg} = \frac{\mathrm{CC}_{ctg}}{CT_{ctg}} \tag{3.1}$$

де CC_{ctg} – кількість тестових зображень, правильно віднесених до категорії ctg; CT_{ctg} – кількість тестових зображень категорії ctg.

У проведених дослідженнях як класифікатор використовувалася машина опорних векторів, для навчання якої з кожної категорії випадковим чином вибиралося по 100 зображень, а решта використовувалися для тестування. Також ряд параметрів залишався незмінним для всіх експериментів:

- При обчисленні FD-SUF та інших локальних дескрипторів на точках інтересу, отриманих за допомогою регулярної сітки, масштаб дорівнював 1, а зсув точок інтересу становив 5 пікселів (1 блок).

- Для формування словника візуальних слів будь-якої довжини навчальної вибірки випадково вибиралося 200 000 локальних дескрипторів.

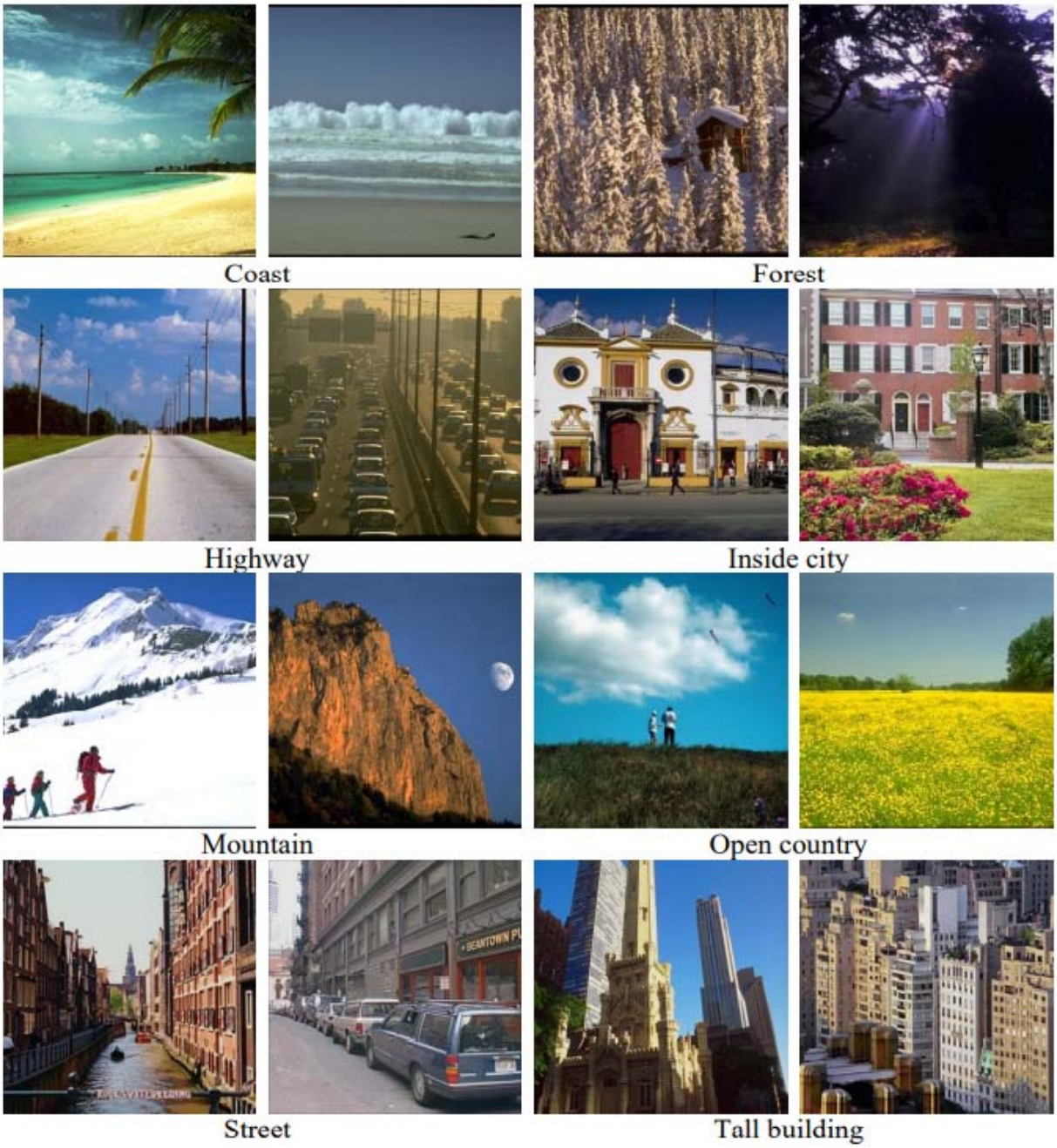

Рис. 3.4. Приклади зображень кожної категорії набору

- Для експериментів використовувався комп'ютер із процесором Intel Core i5-2430M 2,4 ГГц та оперативною пам'яттю Kingston 1333 МГц DDR3 8 Гб. Усі розрахунки повторювалися 5 разів, після чого результати усереднювалися.

У першому експерименті проводилося порівняння локальних дескрипторів FD-SUF, SURF та G-SURF. Для останніх двох обчислень здійснювалися на точках інтересу, отриманих за допомогою регулярної сітки та детектора «швидкий Гесіан».

Розмір словника візуальних слів обраний 128. При формуванні глобального дескриптора коефіцієнт $\gamma = 1$ (без нормалізації елементів), довжина дескриптора також не скорочувалася. Результати порівняння наведено у таблиці 3.3.

Таблиця 3.3.

Основні показники категоризації зображень під час використання різних локальних дескрипторів

	SURF	G-SURF	SURF	G-SURF	FD-SUF
Точність категоризації					
Coast	69,36	58,92	84,62	78,85	**85,51**
Forest	86,11	93,67	**94,30**	94,01	92,84
Highway	67,08	75,13	76,25	80,42	**82,08**
Inside city	78,52	80,85	90,06	87,02	**92,15**
Mountain	77,86	68,42	81,39	80,78	**82,36**
Open country	49,03	50,39	64,73	**70,43**	69,03
Street	69,96	70,23	82,12	80,38	**86,11**
Tall building	76,04	50,83	**87,50**	86,98	87,11
	71,75	68,55	82,62	82,36	**84,65**
Кількість обчислених дескрипторів					
Середнє	207	207	2304	2304	2304
Час обчислення дескрипторів, мс.					
Середнє	12,35	15,78	57,26	96,35	17,18

Обчислення виготовлялися з використанням одного процесорного ядра. Як видно з наведених даних, обчислення локальних дескрипторів на точках інтересу, отриманих за допомогою регулярної сітки, суттєво підвищує точність категоризації.

При цьому точність категоризації для дескрипторів SURF та G-SURF відрізняється для різних категорій, але у середньому однакова.

Також запропонований алгоритм обчислення набору локальних дескрипторів дозволяє підвищити точність категоризації зображень у середньому на 2% за рахунок виключення зважування елементів дескриптора за допомогою фільтра Гауса.

При цьому обчислення здійснюються у 3,3 та 5,6 разів швидше, ніж дескриптори SURF і G-SURF, які обчислюють кожен локальний дескриптор окремо.

3.3. Дослідження параметрів алгоритму формування глобальних дескрипторів

У другому експерименті визначався оптимальний розмір словника візуальних слів під час використання локальних дескрипторів FD-SUF. Так, як і в попередньому експерименті, коефіцієнт нормалізації γ встановлений рівним 1. Результати обчислень для окремих категорій наведено у таблиці 3.4, а графік для середньої точності категоризації представлений на рисунку 3.5.

Результати експерименту показали, що зі збільшенням розміру словника візуальних слів вище 128 елементів, точність категоризації зображень залишається практично без змін.

Це пов'язано з тим, що за велику кількість візуальних слів деякі з них можуть розташовуватися досить близько у просторі ознак, що призводить до фактичного дублювання окремих ділянок глобального дескриптора без внесення нової інформації. У зв'язку з цим, у всіх подальших експериментах використовується 128 візуальних слів.

Таблиця 3.4.

Точність категоризації зображень (%) в залежності від розміру словника візуальних слів

Категорії	Розмір словника візуальних слів						
	8	16	32	64	128	192	256
Coast	79,36	82,44	84,23	84,87	85,51	86,54	85,90
Forest	92,40	92,98	94,88	93,86	92,84	92,54	93,27
Highway	73,75	78,96	84,38	81,46	82,08	82,71	82,50
Inside city	89,10	87,34	89,42	93,11	92,15	92,63	91,99
Mountain	75,06	77,25	78,35	79,44	82,36	83,46	84,79
Open country	57,31	65,91	70,32	68,17	69,03	67,63	66,88
Street	71,88	75,35	80,56	84,20	86,11	86,98	87,67
Tall building	75,00	83,46	85,81	85,94	87,11	86,33	85,81
Середня точність	76,73	80,46	83,49	83,88	84,65	84,85	84,85

Рис. 3.5. Графік залежності середньої точності категоризації зображень від розміру словника візуальних слів ($\gamma = 1$)

У третьому експерименті досліджувалась залежність середньої точності категоризації від коефіцієнта нормалізації при фіксованому розмірі словник візуальних слів.

Результати проведених обчислень для окремих категорій наведено у таблиці 3.5, а на рисунку 3.6 представлено графік середньої точності категоризації.

Таблиця 3.5.

Точність категоризації зображень (%) в залежності від коефіцієнта нормалізації γ

Категорії	Коефіцієнт нормалізації									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Coast	90,38	90,51	91,15	90,26	89,23	89,87	89,36	89,23	88,59	85,51
Forest	93,86	93,27	93,71	94,15	94,30	93,86	94,15	94,30	92,98	92,84
Highway	84,79	84,58	87,08	86,04	86,88	86,04	85,00	82,50	83,13	82,08
Inside city	93,43	93,91	94,55	94,71	94,55	94,23	95,03	94,07	93,27	92,15
Mountain	90,63	92,09	90,75	91,00	91,24	89,66	89,29	87,71	84,67	82,36
Open country	75,48	76,56	77,85	77,31	76,67	76,99	75,91	75,05	70,97	69,03
Street	88,54	89,24	90,28	89,58	89,93	88,89	88,54	88,37	85,94	86,11
Tall building	90,10	90,36	90,89	91,41	90,36	90,23	90,10	88,67	87,76	87,11
Середня точність	88,40	88,82	89,53	89,31	89,14	88,72	88,42	87,49	85,91	84,65

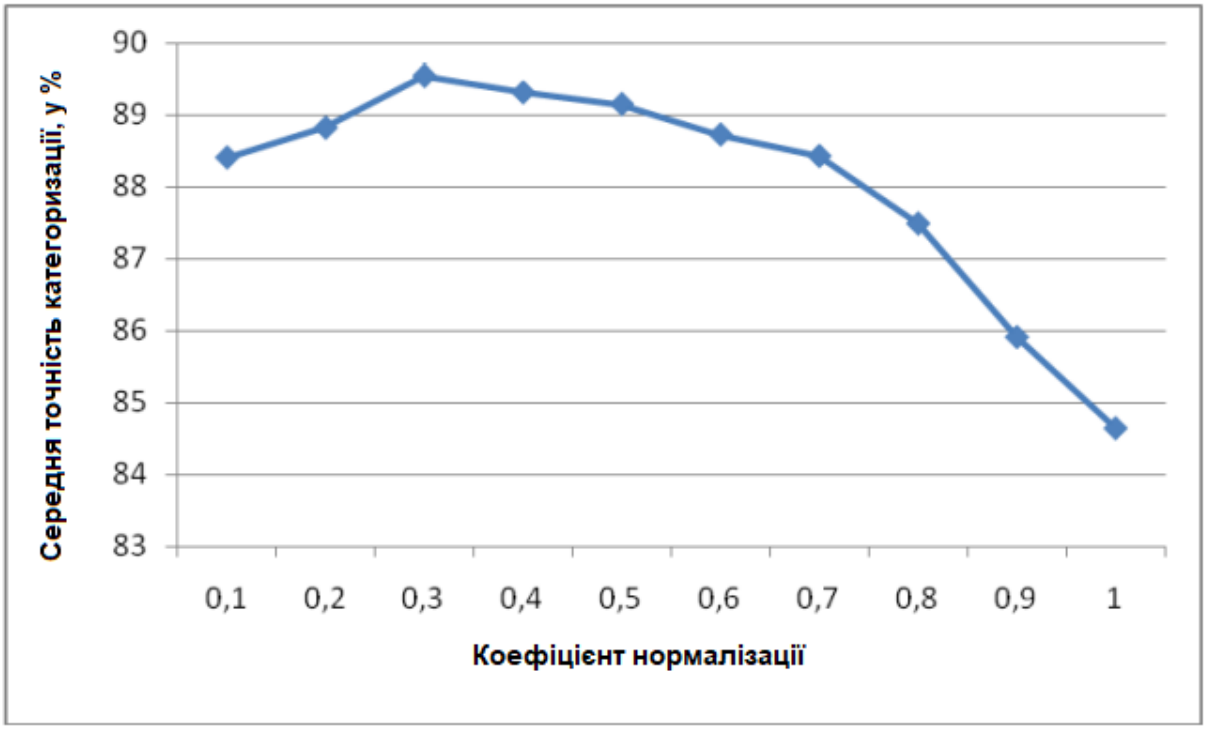

Рис. 3.5. Графік залежності середньої точності категоризації зображень від коефіцієнта нормалізації γ ($S = 128$)

Як видно з наведених даних, будь-яке значення коефіцієнта нормалізації γ, відмінне від 1, призводить до підвищення точності категоризації, оскільки це дозволяє знизити вплив локальних дескрипторів, обчислених на області із зображенням будь-якої структури, у випадку, коли ця область займає значну частину зображення. При цьому найкращий результат досягається при $\gamma = 0{,}3$. Це значення буде використовуватися в подальших експериментах.

Оскільки сформовані глобальні дескриптори мають великою розмірністю (для 128 візуальних слів довжина глобального дескриптора дорівнює 8192), то для швидких порівнянь зображень велике значення має скорочення розмірності за допомогою методу головних компонент.

Також було проведено додаткові експерименти для оцінки точності визначення окремих категорій зображень за допомогою колірних локальних дескрипторів. Оскільки довжина отриманих колірних дескрипторів змінюється в залежності від кількості використовуваних компонент колірного простору, то для порівняння всі глобальні дескриптори скорочувалися до 384 елементів. Результати проведених обчислень представлені у таблиці 3.6.

Результати дослідження показали, що найкращі результати в окремих категоріях досягаються з використанням трьох дескрипторів, обчислених у колірних просторах nRGB, HSV та LUV. Комбінація цих трьох дескрипторів дозволяє підвищити середню точність категоризації на 2,1% та 3% порівняно з дескрипторами, обчисленими тільки для nRGB та відтінків сірого (Y) відповідно.

Таблиця 3.6.

Порівняння точності категоризації зображень за допомогою різних кольорів дескрипторів (%)

	Кольорова гама (компонент)							
	Y	rg	Opponent	HSI	nRGB	HSV	LUV	nRGB + HSV + LUV
Coast	90,00	77,56	84,36	82,18	**90,13**	89,23	86,54	91,38
Forest	94,30	91,81	93,13	93,27	93,86	**95,61**	94,30	96,61
Highway	84,79	81,67	85,21	83,54	**85,63**	80,63	85,00	87,25
Inside city	93,27	87,34	89,26	85,42	**96,15**	90,38	88,94	94,27
Mountain	91,73	78,59	84,31	84,06	**94,16**	85,04	88,32	93,70
Open country	75,81	68,28	72,58	72,15	**81,29**	74,19	76,45	82,61
Street	89,06	81,25	89,24	88,02	85,42	88,54	**90,63**	94,23
Tall building	90,23	89,19	89,58	87,37	89,84	87,89	**91,41**	93,58
Середня точність категоризації	88,65	81,96	85,96	84,50	89,56	86,44	87,70	91,70

3.4. Результати експериментальних досліджень автоматичного анотування зображень

Для перевірки запропонованого методу автоматичного анотування зображень, використовувалася база зображень IAPR TC-12. Ця база містить 19627 зображень розміром 480 × 360 пікселів, кожне з яких описано кількома ключовими словами. Деякі показники бази представлені у таблиці 3.7.

Експерименти проводилися з метою дослідження впливу параметрів модифікованої мережі ESOINN та кількості вибраних зображень при відновленні ключових слів навчальних зображень та оцінці умовних ймовірностей приналежності нових зображень ОТВ-групам та семантичних груп на якість анотування цих зображень.

Таблиця 3.7

Статистичні дані бази зображень IAPR TC-12

Параметр		Значення
Кількість зображень	навчальних	17665
	тестових	1962
Кількість зображень на ключове слово	мінімальне	44
	медіанне	153
	середнє	347,7
	максимальне	4999
Кількість ключових слів	всього	291
	з частотою менше середнього	217
Кількість ключових слів на зображення	мінімальне	1
	медіанне	5
	середнє	5,7
	максимальне	23

Для оцінки якості анотування обчислювалися середня точність (precision), середня повнота (recall), а також спільна оцінка цих двох показників за допомогою F_β-міри:

$$precision = \frac{1}{N}\sum_{n=1}^{N}\frac{CA(k_n)}{AA(k_n)} \tag{3.2}$$

$$recall = \frac{1}{N}\sum_{n=1}^{N}\frac{CA(k_n)}{GT(k_n)} \tag{3.3}$$

$$F_\beta = \frac{(\beta^2+1)\cdot precision\cdot recall}{\beta^2\cdot precision+recall} \tag{3.4}$$

де $AA(k_n)$ – кількість зображень, автоматично анотованих ключовим словом k_n; $CA(k_n)$ – кількість зображень, правильно анотованих ключовим словом k_n; $GT(k_n)$ – кількість зображень, що містять у тестовій анотації ключове слово k_n; β – коефіцієнт, призначеннях [0; 1) дає більшу вагу точності, а при $\beta > 1$ – повноті анотування.

При тестуванні пріоритет віддавався точності анотування, у зв'язку з чим використовувалася $F_{0.5-\text{міра}}$ $(\beta = 0{,}5)$. Також як додатковою інформацією підраховувалася кількість ключових слів, використаних при автоматичному анотуванні $(N+)$.

У всіх експериментах при обчисленні візуальних дескрипторів використовувалися значення параметрів, отримані на етапі дослідження обчислення візуальних дескрипторів $(S = 128, \gamma = 0{,}3, Z = 384)$, а при формуванні навчальних груп текстовий дескриптор зображення має більш3 вагу $(\alpha = 0{,}75)$. Усі розрахунки повторювалися 5 разів, після чого результати усереднювалися.

Висновки до третього розділу

1. Розроблене експериментальне програмне забезпечення для автоматичного анотування зображень. Програмне забезпечення має модульну організацію та складається з семи модулів. Модулі, що реалізують роботу алгоритмів: модуль перетворень зображення, модуль обчислення візуальних дескрипторів, модуль формування текстового дескриптора, модуль відновлення ключових слів, модуль формування навчальних груп та частина ядра системи, що відповідає за реалізацію алгоритму автоматичного анотування зображень.

Для організації взаємодії з користувачем реалізований інтерфейс. Детально розглянуто схеми функціонування зазначених модулів та проведено відповідні експериментальні дослідження.

ВИСНОВКИ

У роботі представлені методи та алгоритми автоматичного анотування зображень в інформаційно-пошукових системах. Основні результати та висновки представлені нижче:

1. Проведено огляд відомих методів автоматичного анотування зображень, кластеризації даних та опис зображень за допомогою низькорівневих ознак, наведена їх класифікація. Також розглянуто низку програмних систем, що реалізують автоматичне абстрактне зображення.

2. Відомі методи анотування можна розділити на три підходи: класифікаційний, генеративний та пошуковий. Класифікаційні методи представляють ключові слова у вигляді незалежних класів, приклади яких навчається класифікатор. Це дозволяє швидко присвоїти зображенням або їх областям мітки категорій, проте для досягнення достатньої точності необхідна збалансована навчальна вибірка. Також збільшення кількості категорій (ключових слів) призводить до значного зниження точності класифікації.

3. Наведено метод автоматичного анотування зображень на основі навчального набору зображень, розділеного на однорідні текстово-візуальні групи, а також запропоновано алгоритм для реалізації даного методу, який відрізняється тим, що анотування нового зображення здійснюється за допомогою навчальних зображень невеликої кількості візуально схожих груп. Даний алгоритм включає три етапи навчання та етап анотування.

4. На першому етапі для всіх навчальних, а також анотованих зображень утворюється глобальний візуальний дескриптор. Для цього з зображення витягується набір локальних дескрипторів, який кодується за допомогою словника візуальних слів. Оскільки цей етап є найбільш обчислювально витратним, то запропоновано швидкий метод вилучення набору локальних дескрипторів, що дозволяє уникнути повторних обчислень

при накладенні областей розрахунку дескрипторів, що істотно прискорює процес анотування. Також розглянуто процес обчислення колірних локальних дескрипторів, використання яких дозволяє підвищити точність анотування, та наведено алгоритми формування словника візуальних слів та кодування набору локальних дескрипторів у глобальний візуальний дескриптор.

5. Розроблене експериментальне програмне забезпечення для автоматичного анотування зображень. Програмне забезпечення має модульну організацію та складається з семи модулів. Модулі, що реалізують роботу алгоритмів: модуль перетворень зображення, модуль обчислення візуальних дескрипторів, модуль формування текстового дескриптора, модуль відновлення ключових слів, модуль формування навчальних груп та частина ядра системи, що відповідає за реалізацію алгоритму автоматичного анотування зображень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Shimabukuro, Y.E.; Duarte, V.; Kalil Mello, E.M.; Moreira, J.C. Presentation of the Methodology for Creating the Digital PRODES; Technical Report; INPE: São José dos Campos, Brazil, 2000. [Google Scholar]
2. Vargas, C.; Montalban, J.; Leon, A.A. Early warning tropical forest loss alerts in Peru using Landsat. Environ. Res. Commun. 2019, 1, 121002. [Google Scholar] [CrossRef]
3. Van Leeuwen, W.J.; Casady, G.M.; Neary, D.G.; Bautista, S.; Alloza, J.A.; Carmel, Y.; Wittenberg, L.; Malkinson, D.; Orr, B.J. Monitoring post-wildfire vegetation response with remotely sensed time-series data in Spain, USA and Israel. Int. J. Wildland Fire 2010, 19, 75–93. [Google Scholar] [CrossRef]
4. Bouyerbou, H.; Bechkoum, K.; Benblidia, N.; Lepage, R. Ontology-based semantic classification of satellite images: Case of major disasters. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 2347–2350. [Google Scholar]
5. Du, L.; Tian, Q.; Yu, T.; Meng, Q.; Jancso, T.; Udvardy, P.; Huang, Y. A comprehensive drought monitoring method integrating MODIS and TRMM data. Int. J. Appl. Earth Obs. Geoinf. 2013, 23, 245–253. [Google Scholar] [CrossRef]
6. Singh, A. Review article digital change detection techniques using remotely-sensed data. Int. J. Remote Sens. 1989, 10, 989–1003. [Google Scholar] [CrossRef][Green Version]