

*МОЛЧАНОВСЬКИЙ О.І.,
МИСТЕЦКИЙ В. А.,
НГІСМ ЛЕ КУАН,
БУЙ ХИУ ДАТ*

ЗАДАЧА ПОШУКУ ФАЙЛІВ ЗОБРАЖЕНЬ

В даній статті описується проблема вирішення задачі пошуку файлів зображень. Проведено аналіз існуючих рішень та розглянуті відомі системи, що дозволяють знаходити схожі зображення. Описана методологія тестування таких систем та наведені результати тестування системи Retrievr. Розглянуті підходи до проектування системи пошуку зображень, що розробляється. Наведено метод представлення зображень у базі даних та механізм порівняння вхідного зображення з зображеннями в базі даних на предмет схожості. Розглядається механізм порівняння зображень за спектрально-частотною характеристикою наявності кольорів. Описано підхід до оцінки сегментної структури зображень та використання цієї структури при порівнянні зображень. Також наведені висновки, щодо проведеної роботи та сформульовано проблеми для подальшого дослідження.

In this paper we describe the problem of image file search. We analyzed existed methods and systems that allow the user to find similar images. We propose testing methodology for such systems and provide the test results for Retrievr system [2]. In the second part firstly we describe the presentation of images in the corresponding database. Secondary we describe the method of image comparison based on spectral-frequency color characteristics. The estimation approach for segmental structure of images is provided as well as search method based on this approach. At the end we provide overall summary and the trends for the feature work.

Вступ

Цілями даної статті є розгляд існуючих інструментів пошуку схожих зображень, а також розробка оригінального інструментарію для розв'язання вказаної проблеми. При розгляді існуючих інструментів важливо висунути припущення щодо методів, які використовуються ними. Це можна зробити на основі інформації, яка надається в процесі експлуатації цих інструментів. Результати таких припущень є важливими з точки зору побудови власного інструментарію, адже дозволяє з'ясувати, що є найбільш важливим, а що другорядним при пошуку схожих зображень.

В умовах бурхливого розвитку та популяризації інформаційних технологій значно збільшується кількість електронної інформації. У зв'язку з цим виникає необхідність створювати засоби її каталогізації та пошуку.

На сьогоднішній момент існує велика кількість систем пошуку текстової інформації, що здатні враховувати популярність ресурсу, на якому ця інформація знаходиться, максимізувати відповідність результатів пошуку до запиту користувача, використовуючи при цьому геолокаційні дані про місцезнаходження користувача, інформацію про його попередні пошуки, вподобання тощо.

Проте значну частину інформації у всесвітній мережі складають мультимедійні файли, такі як музика, відео та зображення. Пошук музикальних файлів спрощений насамперед тим, що для більшості файлів існують закладені в нього метадані – теги, завдяки яким можна сказати про назву пісні, альбому чи виконавця. Відео файли найчастіше зберігаються на спеціалізованих сервісах, в яких слідкують за тим, щоб назва та опис, що супроводжує відео файл відповідав йому.

Найбільшу складність для пошуку складають файли зображень, що найчастіше не мають тегів, назва файлів не несе інформації про їх склад (більшість фотокамер, хостингів та редакторів зберігають фотографії з службовими назвами). Тому на сьогоднішній момент основним типом пошуку файлів зображень є пошук по текстовій інформації, що супроводжує це зображення на інтернет сторінці. Проте такий пошук не може задовольнити користувачів, які хочуть знайти саме певне зображення, що, наприклад, схоже на деяке зображення, яке користувач вже має.

Об'єктом нашого дослідження стала саме така система, яка зможе знаходити у своїй базі зображень зображення, що схоже не те, яке задав користувач у вигляді іншого зображення.

На сьогоднішній момент існує велика кількість вузько-направлених систем розпізнавання та обробки зображень. Такими системами є, наприклад, системи розпізнавання номерних знаків автомобілів, системи цифрового зору роботів, системи розпізнавання будинків на знімках супутників тощо. Головною особливістю таких систем є те, що в них наперед відомий формат вхідних зображень та об'єктів, що необхідно розпізнати. Це дає змогу використовувати найбільш підходящі алгоритми для обробки саме таких об'єктів. В поставленій нами задачі основною проблемою являється велика кількість типів зображень, між якими система повинна провести порівняння та прийняти рішення про схожість чи відмінність їх по деяким ознакам.

Насамперед необхідно визначитись з тим, саме за якими ознаками система повинна порівнювати зображення та визначати схожість між ними. За результатами проведеного дослідження [1] ми виділили наступні основні критерії, за якими можна порівнювати зображення: кольорова насиченість, сегментованість зображення, наявність деяких головних об'єктів (наприклад будинків, тварин, людей тощо) та жанр зображення (пейзаж, портрет, креслення і т.д.). Отже, система повинна враховувати ці критерії для пошуку та порівняння зображень.

Аналіз існуючих рішень

Зараз існує декілька систем, що намагаються вирішувати поставлену задачу: *Retrievev* [2], *Gazora* [3], *TinEye* [4] та деякі інші. Ми намагались відповісти на питання, яким чином можна оцінити роботу тієї чи іншої системи і створили деяку методологію оцінки системи пошуку зображень [1].

Для цього нами було запропоновано розбиття тестового корпусу зображень на групи (класи) за жанрово-семантичними ознаками. Таких груп було сформовано 12, серед яких: окремі об'єкти (автомобілі, літаки), креслення (блок схеми, таблиці), пейзажі, тварини, малюнки, фотопортрети. Більшість зображень можна віднести до деякої групи, проте присутня неоднозначність (деяке зображення можна віднести одразу до декількох класів).

Для кожного вхідного зображення виставляються оцінки від 0 до 10 балів по чотирьом параметрам: сегментованість (кількість яскраво виражених об'єктів), контрастність (оцінка чіт-

кості переходів між об'єктами), кольорова насиченість (кількість різних кольорів на зображенні), кількість відтінків (оцінка кількості відтінків одного кольору; для всіх кольорів зображення). Оцінка цих параметрів проводиться людиною експертом.

За цими оцінками були відібрані еталонні зображення, в яких певна характеристика оцінюється в 0, 3, 5, 7 та 10 балів. Еталони служать для спрощення оцінки зображень експертами.

До оцінки результуючих зображень, окрім цих чотирьох параметрів, додаються ще дві характеристики від 0 до 10 балів: семантична схожість результату з вхідним зображенням (сюжет, тематика, жанр) та кольорова схожість.

Для тестування була створена база зображень в 150 зображень, в якій були представлені зображення з усіх класів.

Після оцінки за вищевказаною схемою вхідних та результуючих зображень експертами проводиться статистична обробка отриманих даних (розрахунок та порівняння математичних сподівань та середньоквадратичного відхилення оцінок результуючих зображень від оцінок вхідних зображень і т.д.). Отримані залежності можна використовувати в якості оцінки пошуку системою взагалі, в розрізі окремих класів та характеристик.

За цією методологією була оцінена система *Retrievev*.

На рис. 1 зображений графік розподілення відносної кількості зображень у кожному класі в діапазонах оцінок семантичної схожості.

Для кожного вхідного зображення деякого класу значення середньої семантичної схожості розраховувалось як середнє арифметичне усіх результуючих зображень для цього вхідного зображення.

Як видно, майже у всіх класах для абсолютної більшості вхідних зображень середня оцінка критерію семантичної схожості лежить в діапазоні 0-2. Однак для класу №8 (пейзажі) ми бачимо ярко-виражений глобальний мінімум кривої для діапазону 0-2 та значне зростання кривої діапазону 9-10 та діапазонів 6-8 і 3-5. Ми пояснюємо це тим, що в базі зображень системою *Retrievev* знаходиться дуже велика кількість зображень на цю тематику (відносно зображень з інших класів).



Рис. 1. Розподілення оцінок результативних зображень за критерієм семантичної схожості

З цього можна зробити висновок, що система ніяк не враховує семантичну схожість при пошуку зображень.

Також були розраховані середні арифметичні відхилення оцінок сегментованості, контрастності, кольорової насиченості та кількості відтінків і розрізі класів за формулою (1).

$$P_m = \frac{\sum_{i=1}^{n_m} |mark_res_i - \frac{\sum_{j=1}^{k_i} mark_inp_j}{k_i}|}{n_m} \quad (1)$$

де P_m – середнє арифметичне відхилення в m -му класі;

n_m – кількість вхідних зображень в m -му класі;

k_i – кількість результативних зображень для i -го вхідного зображення;

$mark_res_j$ – оцінка відповідного результативного зображення;

$mark_inp_i$ – оцінка відповідного вхідного зображення.

Отримані результати можна побачити на рис. 2.

Також були розраховані середнє арифметичне дисперсії оцінок сегментованості, контрастності, кольорової насиченості та кількості відтінків в розрізі класів за формулою (2).

$$D_m = \frac{\sum_{i=1}^{n_m} (mark_inp_i - \frac{\sum_{j=1}^{k_i} mark_res_j}{k_i})^2}{n_m}, \quad (2)$$

де D_m – середнє арифметичне дисперсії в m -му класі;

n_m – кількість вхідних зображень в m -му класі;

k_i – кількість результативних зображень для i -го вхідного зображення;

$mark_res_j$ – оцінка відповідного результативного зображення;

$mark_inp_i$ – оцінка відповідного вхідного зображення.

Отримані результати можна побачити на рисунку 3.

З поданих вище графіків можна зробити висновок, що результати мають досить великі відхилення та дисперсії від параметрів вхідного зображення. Виключеннями є значення дисперсій по параметру кольорової насиченості в 2 (блок схеми), 4 (тварини), 9 (портрети), 10 (графіка) та 12 (макрозйомка) класах. Причиною цього, на наш погляд, є кольорова монотонність більшості зображень в цих класах.

Таким чином, основним критерієм пошуку в системі Retrievr є кольорова схожість зображень. Всі інші параметри враховуються слабо, чи не враховуються взагалі.

Постановка задачі

Розглянувши деякі з існуючих систем пошуку зображень, нами була сформульована наступна задача. Необхідно розробити комплекс програмних модулів для пошуку файлів зображень, які розміщуються в локальному сховищі даних (локальний комп'ютер чи локальна мережа) або глобальній мережі Інтернет.

Реалізація даної задачі передбачає розгляд та реалізацію різних критеріїв пошуку. До них можна віднести порівняння зображень на основі кольорового спектру, на основі просторових характеристик (сегментів) тощо. Нижче ми послідовно розглянемо ці критерії.

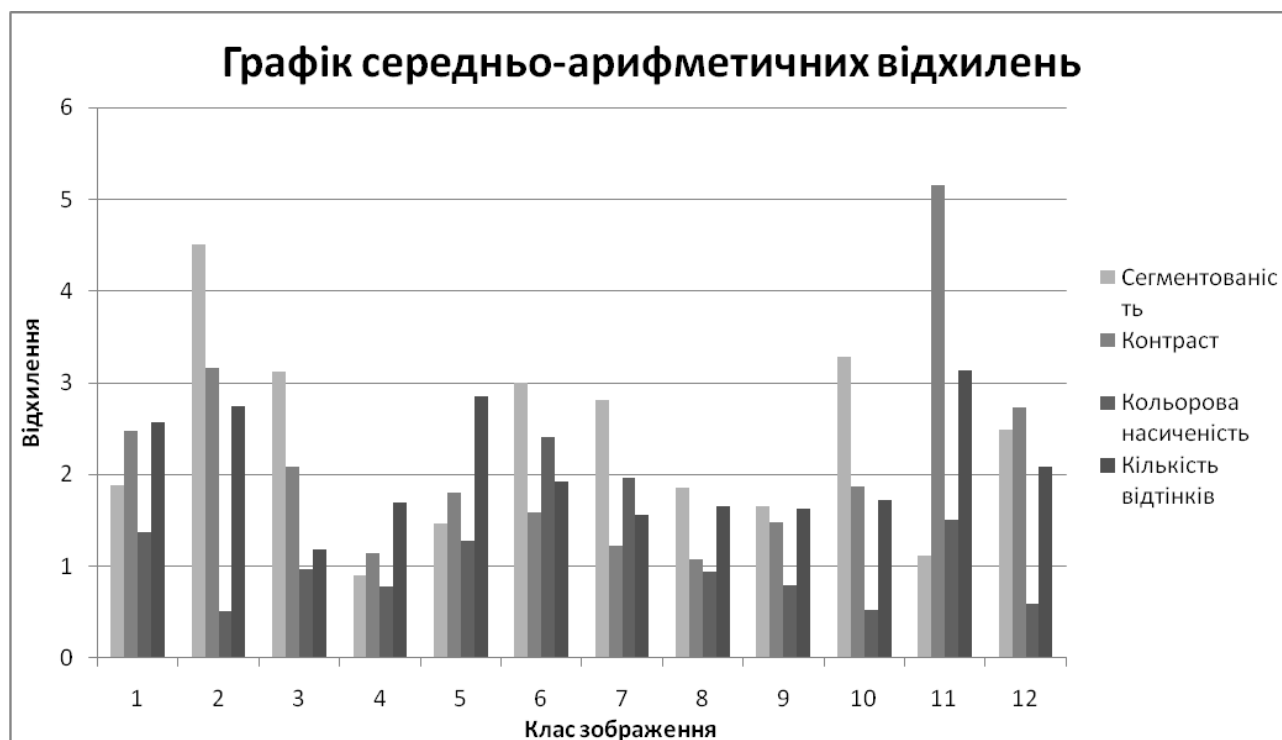


Рис. 2. Графік середньо-арифметичних відхилень по оцінкам сегментованості, контрасту, кількості кольорів та відтінків

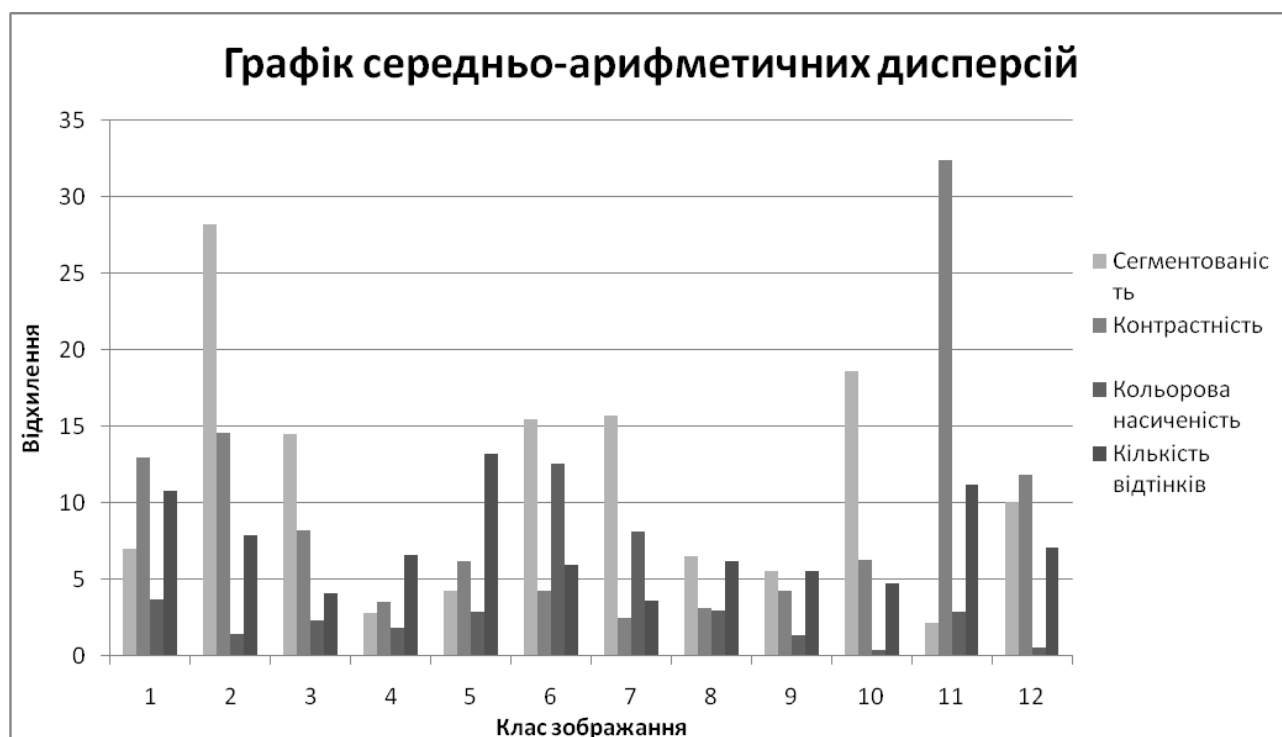


Рис. 3. Графік середньо-арифметичних дисперсій по оцінкам сегментованості, контрасту, кількості кольорів та відтінків

Представлення зображень

Одною з основних проблем при створенні системи пошуку зображень є великі об'єми інформації, що необхідно обробити та порівняти за фіксовано малий час запиту користувача. Для цього потрібно розробити механізм пред-

ставлення зображень бази даних в такому вигляді, щоб зберегти основні дані про кольори та їх розміщення на оригіналі зображення для подальшого порівняння цих даних з вхідним зображенням від користувача.

В якості такого механізму був обраний апарат модифікованого дискретного вейвлет-

перетворення [5]. Для зменшення обчислювальної складності не проводиться розрахунок часткових різниць на кожній ітерації та сумування із зміною частотно-часової локалізації та зріджування вхідного сигналу. Отриманий після перетворення об'єкт будемо вважати деяким ескізом зображення.

Перед створенням цього ескізу всі зображення нормалізуються до однакового квадратного розміру (наприклад 1024x1024). В якості вікна вейвлет-перетворення було обрано вікно розміром 32x32. Результатом сумування для кожного місцезнаходження вікна зводяться до середньо арифметичного значень кольорів всіх точок в цьому вікні. Отже, ми отримуємо 32x32 значення кольорів по 3 компоненти кожен (використовується стандартне кольорове представлення RGB). Таким чином вимірність ескізу буде 32x32x3.

Значення для компонент r , g та b вікна (i,j) буде розраховуватись за формулами (3), (4) та (5).

$$r_{i,j} = \frac{\sum_{l=32i}^{32i+32} \sum_{k=32j}^{32j+32} (r_{lk})}{32 \cdot 32}, \quad (3)$$

$$g_{i,j} = \frac{\sum_{l=32i}^{32i+32} \sum_{k=32j}^{32j+32} (g_{lk})}{32 \cdot 32}, \quad (4)$$

$$b_{i,j} = \frac{\sum_{l=32i}^{32i+32} \sum_{k=32j}^{32j+32} (b_{lk})}{32 \cdot 32}. \quad (5)$$

В якості міри відстані (оцінки схожості) між ескізами двох зображень була обрана Манхетенська відстань в кольоровому просторі RGB. Ця відстань розраховується, як відстань між двома точками в 32x32x3-мірному просторі.

Кожний об'єкт має 32x32x3 блоки (властивості) (всього 3072). Вважаємо, що 3 компоненти кольору кожного з 32x32 блоків являються тривимірними значеннями блоків, різниця яких усереднюється.

Кожна різниця нормалізується (ділиться на 255), а отримана відстань усереднюється (ділиться на 32x32).

Таким чином, манхетенська відстань між двома ескізами лежить в діапазоні від 0 до 1 та обраховується за формулою (6).

$$P = \frac{\sum_{i=1}^{32 \cdot 32} \left(\left| \frac{R_{Ai} - R_{Bi}}{255} \right| + \left| \frac{G_{Ai} - G_{Bi}}{255} \right| + \left| \frac{B_{Ai} - B_{Bi}}{255} \right| \right)}{3 \cdot 32 \cdot 32}, \quad (6)$$

де P – манхетенська відстань між ескізами;

A – перший ескіз;

B – другий ескіз;

R, G, B – кольорові компоненти простору RGB відповідного блоку відповідного ескізу.

Тобто, чим менша отримана відстань, тим ескізі більш схожі.

Дана метрика була обрана в якості компромісу між «жорсткістю» оцінки (в порівнянні, наприклад, с Манхетенською відстанню між ескізами по параметру Hue кольорового простору HSV) та обчислювальною складністю (в порівнянні, наприклад, з евклідовою відстанню в кольоровому просторі RGB).

Наявність модулів в формулі розрахунку відстані дозволяє цьому механізму враховувати не лише кольорове співпадіння, а й враховувати розміщення тих чи інших кольорових блоків на оригінальних зображеннях.

Проте, використовуючи лише цей механізм, ще на етапі створення ескізу сильно різнокольорові зображення вироджувались в сіро-монотонні ескізи через усереднення кольору в кожному з блоків. В результаті порівняння для цих зображень апарат розраховував малі відстані з ескізами дійсно сіро-монотонних зображень в базі, що значно погіршувало результати пошуку.

Для уникнення цієї проблеми було розроблено підсистему розрахунку та врахування частотно-спектральної характеристики [6], яка характеризує процентне включення кольорів того чи іншого діапазону у зображенні.

Частотно-спектральний аналіз зображень

Для побудови частотного спектру діапазонів кольорів, що відповідають кольорам: червоний, оранжевий, жовтий, зелений, голубий, синій та фіолетовий, був застосований кольоровий простір HSV:

Компонента hue (тон) лежить в діапазоні значень 0-360° та відображає тон того чи іншого кольору.

Не дивлячись на математичну точність, у такої моделі є суттєвий недолік: на практиці кількість відтінків, що можна розрізнити змен-

шується при наближенні яскравості до нуля. Також на малих S і V з'являються суттєві похибки округлення при переводі компонент кольорового простору RGB в простір HSV та навпаки.

Тобто, якщо розглядати лише значення компоненти hue в кожному піддіапазоні ми будемо включати білі, сірі та чорні кольори, як показано на рисунку 4.

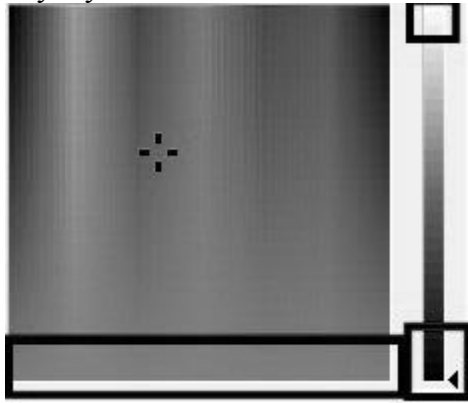


Рис. 4.

Тому для вилучення та врахування білого, сірого та чорного кольору спочатку проводиться перевірка кольору кожного пікселя на належність до цих під діапазонів в кольоровому просторі RGB.

Таким чином ми отримуємо 10 значень для наступних кольорів: червоний, оранжевий, жовтий, зелений, блакитний, синій, фіолетовий, чорний, білий та сірий. Очевидно, що сума цих значень дорівнює 1 (100% всього зображення).

Отже, розрахувавши ці характеристики ми отримуємо відображення кольорової насичення зображення, а саме його різнокольоровість чи монотонність. Тому, на основі цих характеристик, було створено спектральний фільтр, який для вхідних зображень, що мають різнокольорову структуру відкидає результати, що мають ярко-виражене домінування певного кольору (мають монотонну кольорову структуру).

Цей фільтр умовно поділений на дві частини. Перша частина визначає тип зображення:

- тип 1. Різнокольорове зображення (всі значення масиву спектру <0.3 , що означає, що немає кольору, який би займав більше 30% зображення);
- тип 2. Зображення що не має домінуючих кольорів (є хоча б одне значення масиву спектру, що >0.3 , але <0.4);
- тип 3. Монотонне зображення (є хоча б одне значення масиву спектру, що >0.4 , тобто є хоча б один колір, що займає більше 40% всього зображення і домінує в ньому).

Після цього він розраховує манхетенську відстань між спектральними характеристиками зображень (точки в 10×3 -мірному просторі), що порівнюються.

В залежності від результатів роботи першої частини спектрального фільтру формується комплексна оцінка схожості ескізів. В якості комплексної оцінки виступає середнє зважене оцінки вейвлет-ескізу зображень та їх частотно-спектральних характеристик. Вона обраховується за формулою (7).

$$P = a_1 \cdot w_mark + (1 - a_1) \cdot s_mark, \quad (7)$$

де a_1 – коефіцієнт, що визначений спектральним фільтром;

w_mark – оцінка схожості вейвлет-ескізів;

s_mark – оцінка схожості частотно-спектральних характеристик.

Вагові коефіцієнти задаються спектральним фільтром, як показано в таблиці 1 (конкретні значення коефіцієнтів були підібрані експериментальним шляхом).

Табл. 1. Значення коефіцієнта a_1 для різних комбінацій порівнюваних зображень

Типи	Тип 1	Тип 2	Тип 3
Тип 1	0,8	0,6	*
Тип 2	0,6	0,8	0,5
Тип 3	*	0,5	0,8

* – означає, що оцінка не розраховується взагалі і така пара зображень не порівнюється.

Таким чином система отримує комплексну оцінку кольорово-структурної та спектрально-частотної схожості вхідного ескізу, та деякого ескізу з бази даних. Причому вагові коефіцієнти окремих оцінок в комплексній оцінці змінюються в залежності від визначеного типу зображення.

Сегментація зображень

Іншою важливою характеристикою, що повинна враховуватись при порівнянні зображень є їх сегментна структура [7]. Сегмент – це деяка область зображення, що відокремлена від інших частин зображення достатньо вираженим контрастним переходом кольору (на границі сегменту є досить великий перехід тону чи яскравості). Найчастіше сегменти відповідають деяким об'єктам, що вирізняються на фоні зображення, чи іншого об'єкту. Також це можуть бути частини деякого об'єкту тощо. На наш погляд, сегментна структура є досить вагомою характеристикою зображення, оскільки людина звертає увагу на об'єкти, що відокремлюються від фону та інших об'єктів на зображенні. А

розмір, місце розташування та взаємозв'язок між цими об'єктами є важливою інформацією, яка повинна використовуватись при порівнянні зображень.

Таким чином, задача полягає в тому, щоб реалізувати механізм виокремлення сегментів в зображенні, причому вони повинні як найбільше відповідати реальним об'єктам зображення, а їх кількість повинна корелювати із експертними оцінками сегментованості зображення (наприклад так, як пропонується в описаній вище методології оцінки роботи систем пошуку зображень).

В якості математичного апарата для вирішення цієї задачі був обраний модифікований алгоритм Краскала побудови мінімального остового дерева. Кожний піксель зображення в цій задачі представляється в якості вершини графа, а вага ребра, що їх з'єднує визначається за формулою (8).

$$w(v_i, v_j) = |I(p_i) - I(p_j)|, \quad (8)$$

де $I(p_i)$ – інтенсивність (яскравість) пікселя p_i .

В ході роботи алгоритму Краскала, на проміжному етапі ми будемо мати декілька різних сегментів (підмножин пікселів) з мінімальною сумарною вагою ребер всередині: сегменти будуть об'єднані ребрами мінімальної довжини, тобто з мінімальними перепадами інтенсивностей між сусідніми пікселями, але тільки до деякого значення максимального ребра (перепаду інтенсивності).

В результаті роботи алгоритму ми отримуємо множину сегментів, їх корені (корені відповідних остових дерев), потужність (кількість пікселів в кожному сегменті) та координати усіх границь кожного сегменту.

Проте в даній реалізації алгоритму сегментування кількість отриманих сегментів виявляється дуже великою, тобто алгоритм розбиває реальний об'єкт на велику кількість малих сегментів. Таким чином інформація про кількість сегментів та їх розмір не відповідає кількості реальних об'єктів на зображенні.

Тому був розроблений алгоритм, що по закінченню роботи алгоритму сегментування проводить додатковий обхід отриманих сегментів та об'єднує їх з сусідніми сегментами, якщо різниця між кольоровими компонентами (манхетенська відстань) нижче деякого порогу. Да-

ний поріг є динамічним та визначається насамперед спектрально-частотними характеристиками зображення. Так якщо зображення є монотонним з домінуванням синього кольору, то алгоритм враховує це і збільшує пороги для об'єднання сегментів по компоненті Blue, тим самим не даючи об'єднати слабо виражений об'єкт з фоном.

Висновки та подальша робота

Універсальність системи, що розробляється накладає додаткові складності для підбору параметрів алгоритмів. Зміна параметрів в одну чи іншу сторону покращує результати порівняння для деяких типів зображень (наприклад для монотонних з великими об'єктами), але погіршує для інших (наприклад для сильно сегментованих, різнобарвних, з великою кількістю дрібних об'єктів). Тому ми здійснюємо підбір цих параметрів порівнюючи отримані результати (наприклад кількість сегментів на зображенні) з експертними оцінками та мінімізуємо середньоквадратичне відхилення отриманих системою даних із даними від експертів. Крім того, система намагається визначити тип зображення за спектрально-частотними характеристиками для того, щоб підібрати найбільш відповідні параметри алгоритмів, вагові коефіцієнти в комплексній оцінці схожості тощо.

В даний момент проводиться проектування високонавантаженої бази даних, що буде здатна оброблювати запити на великій базі зображень за прийнятний час. Ведеться розробка алгоритмів збереження інформації про структуру та взаємозв'язок сегментів для подальшого порівняння не тільки кількісних характеристик сегментів зображення (їх кількість та розмір), а й якісних (взаємне розташування, сусідство і т.д.). Також проводиться розробка модулю виявлення облич на зображення та збереження інформації про їх площу та розміщення. Це дозволить отримувати додаткову семантичну інформацію з зображення та проводити пошук, що направлений на сюжет зображення. Крім того це дозволить відокремити клас зображень – портрети, серед яких можна реалізовувати нейро-мережеві методи пошуку [8], що дають гарні результати для порівняння облич.

Список літератури

1. Молчановский А.И., Буй Хиу Дат. Методология тестирования систем поиска изображений: тези доповідей Міжнародної науково-практичної конференції [Інформаційні технології та комп'ютерна інженерія], (Вінниця, 19-21 травня 2010). – Вінниця: ВНТУ, 2010. – 487с.
2. Веб-сторінка системи Retrieve: <http://labs.systemone.at/retrieve/about>.
3. Веб-сторінка системи Gazopa: <http://www.gazopa.com>.
4. Веб-сторінка системи TinEye: <http://www.tineye.com>.
5. The engineer's ultimate guide to wavelet analysis: <http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html>.
6. Молчановский А.И., Нгием Лэ Куан. Спектральный анализ для задачи поиска изображений: материалы третьей Всеукраинской научно-практической конференции молодых ученых и студентов [Информационные процессы и технологии «Информатика – 2010»], (Севастополь, 26-30 апреля 2010). – Севастополь: СевНТУ, 2010. – 384с.
7. Молчановский А.И., Нгием Лэ Куан. Проблема сегментации для задачи поиска изображений: тези доповідей Міжнародної науково-практичної конференції [Інформаційні технології та комп'ютерна інженерія], (Вінниця, 19-21 травня 2010). – Вінниця: ВНТУ, 2010. – 487с.
8. Васильев В.И. Распознающие системы. – Киев: Наукова думка, 1983.-424с.