1. Тема
2. Актуальність роботи
   1. Полегшення пошуку і перегляду зображень – чим менше дублікатів – тим простіше знайти потрібне зображення візуально
   2. Видалення проміжних результатів обробки
   3. Зображення можуть займати значну кількість пам’яті персонального комп’ютера – потрібно регулярно видаляти/переміщати непотрібне.
   4. Запобігання завантаженню дублікатів до мережі
3. Постановка задачі

Метою дипломної роботи є побудова системи для пошуку схожих зображень.

Розроблена система повинна вважати схожими зображення з такими відмінностями:

* Використання фільтрів
* Поворот
* Зсув
* Зміна розмірів
* Додавання об’єктів
* Вхідні дані – зображення в форматі JPEG/PNG.
* Показувати користувачу інформацію про те які зображення можуть бути схожими
* Система повинна працювати на комп’ютері користувача

1. Деякі системи пошуку схожих зображень: Tineye, Dup Detector, Image Comparer, Google image search.
2. Порівняння

* Tineye – сервіс для виявлення дублікатів зображень. Система спрямована на виявлення спільного походження зображень, проте шукати серед власних дозволяє лише комерційна версія.
* Dup Detector – програма для пошуку схожих зображень на персональному комп’ютері. Має дещо незручний інтерфейс – виводить результат у вигляді логу.
* Image comparer – також виконує пошук на персональному комп’ютері. На відміну від попередньої, має зручніший інтерфейс, але використовує слабкіші методи.
* Google image search – і гугл. (сервіс для пошуку зображень на задану тему. Пошук виконується за схожістю і за ключовими словами.)

1. Методи

* Watermarking – додавання в зображення дані, які потім використовуються для визначення схожості зображення. Застосовується для захисту авторських прав. Зміни, які додаються, не змінюють зображення візуально, але потім легко виділяються відповідним програмним забезпеченням.
* Content-based (fingerprinting) – в основі лежить створення відбитка – невеликої кількості даних, що описують зображення. Як криптографічний хеш. Відбиток має бути стійким до незначних змін зображення. Ці методи не вимагають попередньої підготовки, проте роблять помилки

1. Схема системи

Дані можуть поступати в систему по запиту користувача, з мережі, або в фоновому режимі, всі створені/модифіковані зображення з жорсткого диска. З зображень створюються відбитки і необхідна інформація зберігається. Користувач переглядає звіт.

1. Алгоритм отримання відбитку. Етапи

* Перетворення Радона.

Воно застосовується через те що переводить зсув і обертання зображення в зсув перетворення. Масштабування зберігається.

* Автокореляція

Зсув зображення спричиняє зсув перетворення Радона, але розмір зсуву для кожної проекції є різним. Для забезпечення незмінності зсуву до кожної проекції перетворення застосовується нормалізована автокореляція.

* Log mapping – має перетворити масштабування вхідного сигналу в зсув вихідного.
* Перетворення Фур’є – має перетворити зсув вхідного в зміну фази.
* Оскільки основна сила сигналу на виході з перетворення Фур’є зосереджена на початку, початок залишається, а решта відкидається. Авторами методу експерементально було з’ясовано, що найкращі результати дають перші 20\*20 пікселів.
* В кінці автори алгоритму пропонують застосувати фільтр і об’єднати сигнали модуля і фази функцією XOR. (також експерементально підібрана авторами методу).

1. Алгоритм перетворює зображення у відбиток 20 на 20 (50 байт).
2. Створена система підтримує ввід за вимогою користувача директорії з зображеннями у форматі JPEG та PNG.

Зліва знаходяться всі оброблені зображення, справа згори – вибране, справа внизу – ті зображення, що є схожими до вибраного. В аудиторії яскраво, тому ось збільшені зображення.

1. Добре розпізнано розмиття
2. Додавання об’єктів
3. Поворот на 180. Також зміна розмірів. Повороти на 2, 10, 90 градусів хоча і дають співпадіння відбитків, але незначне. Зсуви не дали співпадінь.
4. Наступний скріншот.
5. Визначено схожими зображення пташки, з зсувом гілок і зміною пози пташки.
6. Хоча на 4 картинках зображено одне і те ж -
7. Програма не дала співпадінь, лише між двома найбільш схожими картинками відбиток співпадає на 62% - що скоріше говорить про те що зображення є різними.
8. Висновки

* На основі аналізу методів було виявлено, що для задачі пошуку схожих зображень на персональному комп’ютері найбільше підходять методи на основі відбитка
* Було реалізовано один з методів– на основі перетворення Радона. Відбиток є стійким до використання фільтрів, незначних афінних перетворень або змін зображення. Дає мало false-positives, середню кількість true-negatives.
* Було розроблено автоматизовану систему, яка використовує обраний метод.

1. Шляхи покращення

* Додавання нових способів введення
* Покращення алгоритму отримання відбитка, комбінування різних методів.
* Централізоване збереження звітів сканування.
* Синхронізація між різними приладами.

1. Дякую за увагу