

МЕТОД ПОШУКУ ДУБЛІКАТИВ ЦИФРОВИХ ЗОБРАЖЕНЬ**Філатова Г.Є., Грічанюк А.О.***Національний технічний університет**«Харківський політехнічний інститут», м. Харків*

Багато власників комп'ютерів і ноутбуків, які захоплюються фотографією або зберігають на цих пристроях велику кількість зображень, фото або скріншотів, рано чи пізно стикаються з тим, що на жорсткому диску зберігаються десятки схожих файлів, які займають великий об'єм пам'яті. Сумарний розмір вторинних файлів може досягати кількох гігабайт. Пошук дублікатів цифрових зображень – це перший крок для звільнення місця на диску. При цьому проводити пошук та видалення однакових фотографій, картинок та скріншотів у ручному режимі – це досить стомлююче та часовитратне завдання. Тому метою дослідження є створення автоматизованої системи пошуку дублікатів цифрових зображень.

Для досягнення поставленої мети було запропоновано метод пошуку дублікатів, заснований на методах цифрової обробки зображень.

Основою методу пошуку дублікатів є дискретне косінусне перетворення (ДКП) відповідних кольорових площин цифрових зображень, а також півтонової складової. Для зменшення часу розрахунків та приведення зображень до одного розміру пропонується зменшення розмірів зображень до 100×100 пікселів. Перетворення кольорових зображень в напівтонові виконується наступним чином: $Y = 0.299R + 0.587G + 0.114B$, де Y – інтенсивність на півтонового зображення; R , G , B – червона, зелена та блакитна складові кольорового зображення відповідно.

Пошук однакових зображень пропонується за допомогою розрахунку вектора значень коефіцієнтів ДКП півтонової та кожної кольорової складових, при цьому використовується лише перші 21 коефіцієнт ДКП для на півтонової складової та перші 6 коефіцієнтів ДКП для кожної кольорової складової. Значення кожного коефіцієнту ділиться на 10 та округляється до цілого значення для зменшення об'єму інформації. Таким чином кожне зображення описується вектором розмірності 39 значень: $\vec{x} = (x_1, x_2, \dots, x_j, \dots, x_{39})^T$.

Класифікація нового зображення виконується за допомогою методу Фікса-Ходжеса (метод «найближчих сусідів») для $k=1$, тобто нове зображення відноситься до того класу Ω_l ($l \in \overline{1, N}$), для якого в базі даних (БД) знаходиться «найближчий сусід» нового зображення ω : $R(\omega, \omega_l) = \min_{i \in \overline{1, N}} R_i(\omega, \omega_i)$, де

$$R_i = \sqrt{\sum_{k=1}^{39} (x_j - x_{ij})^2} \text{ – відстань між новим зображенням } \omega \text{ та зображенням } \omega_i \text{ в}$$

БД; x_j , x_{ij} – значення j -ї ознаки, виміряні у зображень ω та ω_i відповідно; N – кількість зображень в БД.

Таким чином, у роботі запропонований метод пошуку дублікатів цифрових зображень, заснований на методах цифрової обробки зображень та методах розпізнавання образів.