

СТЕГАНОГРАФІЯ ТА СТЕГОАНАЛІЗ / STEGANOGRAPHY & STEGANALYSIS

DOI: 10.18372/2225-5036.28.17370

МЕТОД ВІДОКРЕМЛЕННЯ ЦИФРОВИХ ЗОБРАЖЕНЬ В РІЗНИХ ФОРМАТАХ ПЕРВІСНОГО ЗБЕРЕЖЕННЯ, ЗАСНОВАНИЙ НА ВЛАСТИВОСТЯХ МАТРИЦІ НАЙМЕНЬШИХ СИНГУЛЯРНИХ ЧИСЕЛ БЛОКІВ

Бобок Іван

Національний університет «Одеська політехніка», Україна



БОБОК Іван Ігорович, д.т.н., доц.

Рік та місце народження: 1987 рік, м. Одеса, Україна.

Освіта: Одеський національний університет ім. І.І. Мечникова, 2009 рік.

Посада: доцент кафедри комп'ютеризованих систем та програмних технологій.

Наукові інтереси: математичні методи та моделі захисту інформації, стеганографія, стеганоаналіз, захист інформаційно-комунікаційних мереж, експертиза цілісності інформаційних контентів, соціальна інженерія, гібридні війни та інформаційне протиборство.

Публікації: понад 90 наукових публікацій, серед яких наукові статті, монографії, тези та матеріали доповідей на конференціях.

E-mail: onu_metal@ukr.net.

ORCID ID: 0000-0003-4548-0709.

Анотація. Цифрові зображення є сьогодні одними з основних носіїв інформації, а цілісність зображення є необхідною умовою для можливості його використання з нерозважальними цілями в будь-якій галузі людської діяльності. При несанкціонованих змінах зображення часто відбувається зміна його формату або зміна формату його частини, роблячи виявлення факту перезбереження в формат, що відрізняється від первісного, показником на порушення цілісності зображення, а задачу відокремлення зображень в різних (з/без втрат) форматах, підвищення ефективності якого шляхом розробки відповідного методу є метою даної роботи, актуальною задачею сучасної безпеки інформації. Запропонований метод заснований на аналізі властивостей матриці найменших сингулярних чисел блоків, що ставиться у відповідність зображенню. Алгоритмічна реалізація методу дозволила підвищити ефективність відокремлення зображень в різних форматах первісного збереження в порівнянні з аналогами, наслідком чого є підвищення в цілому ефективності виявлення порушень цілісності зображення, зокрема при стеганоаналізі, виявленні фотомонтажу тощо.

Ключові слова: цифрове зображення, формат з втратами, формат без втрат, сингулярне число, блокувальний метод.

Постановка проблеми

Цифрові зображення (ЦЗ) є на сьогоднішній день одними з основних носіїв інформації [1, 2], що мають значну перевагу в порівнянні з аналоговими [3], забезпечуючи можливість отримання будь-якої кількості копій в цифровому вигляді без втрати якості незалежно від часу, що минув після створення оригіналу. Але обробка таких зображень, маніпуляції з ними, неавторизовані зміни, які сьогодні просто і якісно робляться завдяки їх цифровому представленню з використанням комп'ютерної техніки за допомогою існуючих програмних середовищ [4],

з кожним днем підвищують актуальність задачі перевірки цілісності ЦЗ, виявлення результатів його несанкціонованих змін при їх наявності [5].

Цілісність зображення є необхідною умовою для можливості його використання з нерозважальними цілями в будь-якій галузі людської діяльності, особливо, коли це стосується області критичної інфраструктури держави, медицини, науки, банківської структури тощо [6-7], адже невиявлені тут своєчасно несанкціоновані зміни зображення можуть привести до катастрофічних наслідків як для окремої людини, організації, банку, так і для держави, людства в цілому.

Збереження ЦЗ може відбуватися з використанням двох принципово різних схем: з втратами та без втрат [3, 8, 9]. При несанкціонованих змінах ЦЗ часто відбувається зміна його формату або зміна формату його частини. Перша ситуація виникає, зокрема, при організації стеганографічного каналу зв'язку [10], коли як контейнер виступає ЦЗ в форматі з втратами, а вбудова додаткової інформації робиться стеганометодом, який не є стійким до атаки стиском, наприклад, методом модифікації найменшого значущого біта, що є одним з найпоширеніших стеганографічних методів в сучасному інформаційному просторі [11], що потребує збереження отриманого стеганоповідомлення в форматі без втрат, приводить до зміни первісного формату контейнера [12]. Друга ситуація має місце, коли порушення цілісності цифрового контенту відбувається локально, зокрема в результаті фотомонтажу, коли контенти, що використовуються, мають різні (з/без втрат) формати [13, 14]. Таким чином, виявлення факту перезбереження ЦЗ в формат, що відрізняється від первісного, чи наявність в ЦЗ частин, властивості яких відрізняються в сенсі формату збереження, є показником на порушення цілісності зображення, роблячи актуальною задачу відокремлення їх в різних (з/без втрат) первісних форматах.

Аналіз останніх досліджень і публікацій

Розв'язок задачі відокремлення ЦЗ в різних форматах збереження з метою використання отриманих результатів для виявлення фотомонтажу розглянутий в [13]. Формальним предметом дослідження тут є матриця нульових сингулярних чисел (СНЧ) блоків ЦЗ, отриманих стандартним чином [3], яка ставиться у відповідність зображенню. Кожний елемент матриці визначається як кількість нульових СНЧ відповідного блоку. Але, як показує практика, побудова та аналіз такої матриці для відокремлення ЦЗ в різних форматах первісного збереження часто приводить до ускладнень, пов'язаних з необхідністю визначення таких СНЧ, що точно дорівнюють нулю. Враховуючи те, що обчислення СНЧ відбуваються в системі чисел з плаваючою точкою, де при виконанні арифметичних операцій виникає і накопичується обчислювальна похибка, де наявні особливості машинної арифметики, встановлення рівності двох дійсних чисел, зокрема СНЧ і нуля, є нетривіальною задачею, розв'язок якої потребує додаткових досліджень, а відповідний використаний підхід для підвищення ефективності відокремлення ЦЗ в різних форматах збереження – удосконалення, що частково було проведено в [9], отримало подальший розвиток та покладене в основу методу, запропонованого в даній роботі.

В [15] обґрунтована можливість розв'язку в межах просторової області ЦЗ задачі відокремлення зображень, збережених в форматі без втрат первісно, від таких, що були перезбереженні в формат

без втрат після первісного стиску, шляхом аналізу збурень значень яскравості пікселів в результаті перезбереження. Просторова область не вимагає додаткових обчислювальних витрат та позбавлена додаткової обчислювальної похибки, що виникає при використанні областей перетворень зображення (частотної, областей різноманітних розкладань матриць тощо) для обчислення аналізованих в ході експертизи параметрів, тому є пріоритетною для будь-яких методів, що працюють з ЦЗ, даючи їм перевагу в порівнянні з аналогами [16].

Принципово інший підхід для розв'язку задачі, що розглядається, запропонований в [17], де досліджуються послідовні кольорні тріади триплетів у матриці унікальних кольорів, сформованій на основі матриці ЦЗ також в просторовій області. Розроблений тут метод виявлення факту стиску з втратами в кольорових ЦЗ є ефективним (в силу чого використовується при порівняльному аналізі ефективностей методів-аналогів в даній роботі), може бути використаний для визначення формату вхідного контейнера при стеганоаналізі, але, оскільки для використовуваного підходу критичним є розмір досліджуваної частини ЦЗ, то його застосування очевидно не буде ефективним при виявленні фотомонтажу, де задіються зображення в різних (з/без втрат) форматах, але заміщуюча область має малі відносні розміри.

В [18] запропонований метод відокремлення ЦЗ в різних форматах збереження, заснований на аналізі кутів між нормованим вектором СНЧ блоків досліджуваного ЦЗ і першим вектором e_1 стандартного базису відповідного простору. Блоки розміру $l \times l$ є результатом стандартної розбивки матриці зображення. В ході експертизи відбувається перезбереження досліджуваного ЦЗ з матрицею A в форматі з втратами з різними коефіцієнтами якості $QF_i \in JP, i = \overline{1, t}$, де JP – множина всіх можливих значень параметра для використовуваного формату з втратами. В результаті отримуються відповідні (тобто такі, що відрізняються лише форматом (з/без втрат)) ЦЗ з матрицями $A_i, i = \overline{1, t}$. Для кожної пари відповідних ЦЗ виду $A, A_i, i = \overline{1, t}$, визначається значення

$$K = \frac{k(A)}{k(A_i)}, \quad (1)$$

де $k(A), k(A_i)$ – середні значення по $l \times l$ – блоках ЦЗ A, A_i відповідно кутів між нормованим вектором СНЧ і вектором e_1 . В ході експертизи значення співвідношення (1) порівнюється з одиницею, що дозволяє зробити шуканий висновок. Метод в цілому по ефективності перевищив існуючі аналоги, забезпечуючи лише 0.5% помилок першого роду

(ЦЗ, яке було Perezбережене у формат без втрат з формату з втратами, визначалося як оригінальне в форматі без втрат), 5% помилок другого роду (оригінальне ЦЗ визначалося як таке, формат якого був змінений), але необхідність Perezбереження поданого ЦЗ в формат з втратами з наступним дослідженням кожної пари відповідних ЦЗ робить цей метод обчислювально затратним в порівнянні з аналогами. Крім того, наявність помилок тут не дає можливості говорити про остаточний розв'язок задачі, що розглядається.

Таким чином, не зважаючи на численні існуючі методи і різноманітні підходи до задачі відокремлення ЦЗ в різних форматах збереження, ця задача не є вирішеною остаточо. Підвищення ефективності її розв'язку буде сприяти підвищенню ефективності виявлення порушення цілісності ЦЗ в цілому, зокрема в процесі стеганоаналізу, виявлення фотомонтажу тощо.

Мета статті та постановка задач

Метою роботи є підвищення ефективності процесу відокремлення ЦЗ в різних (з/без втрат) форматах первісного збереження шляхом розробки нового методу, заснованого на аналізі матриці найменших СНЧ блоків, введеної і частково дослідженої в [9].

Ефективність оцінюватиметься за допомогою показника точності ACC [19] (accuracy), яка визначається відповідно до формули:

$$ACC = (TP + TN) / (TP + FN + TN + FP), \quad (2)$$

де TP (True Positive) – число правильно виявлених ЦЗ в форматі з втратами, які були Perezбережені у формат без втрат; TN (True Negative) – число правильно виявлених оригінальних ЦЗ в форматі без втрат; FP (False Positive) – число оригінальних ЦЗ в форматі без втрат, помилково прийнятих за такі, формат яких був змінений; FN (False Negative) – число ЦЗ, Perezбережених у формат без втрат з формату з втратами, помилково визнаних оригінальними.

Для досягнення поставленої мети в роботі розв'язуються наступні задачі:

1. Забезпечити подальший розвиток підходу, заснованому на аналізі матриці найменших СНЧ блоків (МНСЧ), запропонованому в [9], для розробки методу відокремлення ЦЗ в різних форматах збереження, шляхом подальшого дослідження властивостей МНСЧ;

2. Розробити метод відокремлення ЦЗ в різних (з/без втрат) форматах первісного збереження та обґрунтувати вибір параметрів для його алгоритмічної реалізації;

3. Провести оцінку ефективності, в тому числі порівняльну, з сучасними аналогами алгоритмічної реалізації запропонованого методу.

Виклад основного матеріалу дослідження

Нехай F – $n \times m$ -матриця ЦЗ, яка розбивається стандартним чином на непересічні $l \times l$ -блоки, довільний з яких позначається далі

$$B_{ij}, i = 1, \dots, \left\lfloor \frac{n}{l} \right\rfloor, j = 1, \dots, \left\lfloor \frac{m}{l} \right\rfloor, \text{ де } \lfloor \cdot \rfloor - \text{ціла частина аргументу.}$$

Нехай $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_l \geq 0$ – СНЧ B_{ij} , які можуть бути отримані різними способами, зокрема шляхом побудови сингулярного розкладання матриці B_{ij} [20].

Теоретичний базис методу відокремлення ЦЗ в різних (з/без втрат) форматах збереження, що пропонується в роботі, заснований на результатах досліджень, отриманих в [9], де основним предметом дослідження є M_{F_T} – матриця найменших СНЧ блоків розміром $\left\lfloor \frac{n}{l} \right\rfloor \times \left\lfloor \frac{m}{l} \right\rfloor$ з елементами $m_{(ij)_{F_T}}$,

$i = 1, \dots, \left\lfloor \frac{n}{l} \right\rfloor, j = 1, \dots, \left\lfloor \frac{m}{l} \right\rfloor$, що ставиться у відповідність ЦЗ з матрицею F за наступним правилом:

$$m_{(ij)_{F_T}} = S_{B_{ij}(T)},$$

де $S_{B_{ij}(T)}$ – кількість СНЧ блоку B_{ij} , для яких має місце співвідношення:

$$\sigma_i < T, \quad (3)$$

а T – параметр, що задовольняє умові:

$$0 < T < 1. \quad (4)$$

В [9] отримані наступні властивості МНСЧ:

1. Мода гістограми значень елементів МНСЧ тим більше, чим менше значення коефіцієнта якості, що використовувався в процесі стиску контенту;

2. Для цифрових контентів в форматі без втрат значення моди не більше, ніж для контентів в форматі з втратами; моди гістограм елементів МНСЧ блоків для ЦЗ в форматі без втрат практично не виходять за межі $\{0,1,2,3\}$, тоді як у випадку зображення в форматі з втратами ця множина розширюється до $\{0,1,2,\dots,7\}$;

3. Встановлено, що для 97% протестованих ЦЗ максимальне значення МНСЧ блоків для зображення в форматі з втратами є неменшим за відповідне значення відповідного ЦЗ у форматі без втрат.

Встановлено, що на практиці можливі ситуації, коли для відповідних ЦЗ в гістограмах значень елементів МНСЧ співпадають і моди, і максимальні значення (рис. 1), що означає, що використання лише цих двох параметрів, на які зверталася основна увага в [9], не дає змоги ефективно відокремлювати ЦЗ в різних форматах збереження, потрібні додаткові дослідження властивостей МНСЧ блоків.

Якщо для ЦЗ ми маємо МНСЧ блоків, яка містить елементи, значення яких виходять за межі

$\{0,1,2,3\}$, то такі значення відповідають блокам, де практично відсутня високочастотна (і, можливо, середньочастотна) складова, тобто блокам, де немає перепадів значень яскравості пікселів. Візуально наше око сприймає сукупність таких блоків як фон [3]. Але не будь-який «фоновий» блок є таким, для якого буде наявною значна кількість СНЧ, порівняних з нулем: для ЦЗ в форматі без втрат незначні перепади в 1-3 градації присутні навіть в фонових блоках, що приводить до того, що їх найменші СНЧ будуть більшими за значенням, ніж відповідні їм СНЧ в відповідних блоках відповідних ЦЗ. Більше того, як встановлено в [21], більшою буде і швидкість їх зміни, при тому, що швидкість зміни найменших СНЧ в блоках ЦЗ з втратами близька до нуля. Для ЦЗ в форматі без втрат наявність блоків, де значна кількість СНЧ буде близька до 0, визначається лише сценою самого зображення, тоді коли для ЦЗ в форматі з втратами такі блоки – це є результат стиску (квантування частотних коефіцієнтів і округлення результату до цілого числа [3]), вони присутні незалежно від сцени, відтвореної на ЦЗ, з чого випливає,

що для ЦЗ в форматі з втратами вони будуть в більшій кількості незалежно від сцени. Чим менше буде використаний коефіцієнт якості стиску, тим більше таких блоків буде в ЦЗ. Ілюстрацією такого висновку є рис.1 для конкретних відповідних зображень, де гістограми МНСЧ блоків, хоча і мають однакові значення мод і максимальних елементів МНСЧ, але все одно є такими, що якісно розрізняються за своїми властивостями: для ЦЗ в форматі без втрат (рис.1(a)) переважна кількість блоків (розмір 8×8) має незначну кількість СНЧ, менших T , на відміну від ЦЗ в форматі з втратами. Для кількісної оцінки цієї різниці треба визначити поріг P для значень елементів МНСЧ блоків. Кількість блоків в ЦЗ, для яких кількість СНЧ, менших T , буде більше P , позначатиметься $k_{>P}$.

Враховуючи вищевикладене, а також результати, отримані в [9], основні кроки методу відокремлення ЦЗ в різних (з/без втрат) форматах збереження виглядають наступним чином.

Нехай F – $n \times m$ -матриця ЦЗ, що піддається експертизі

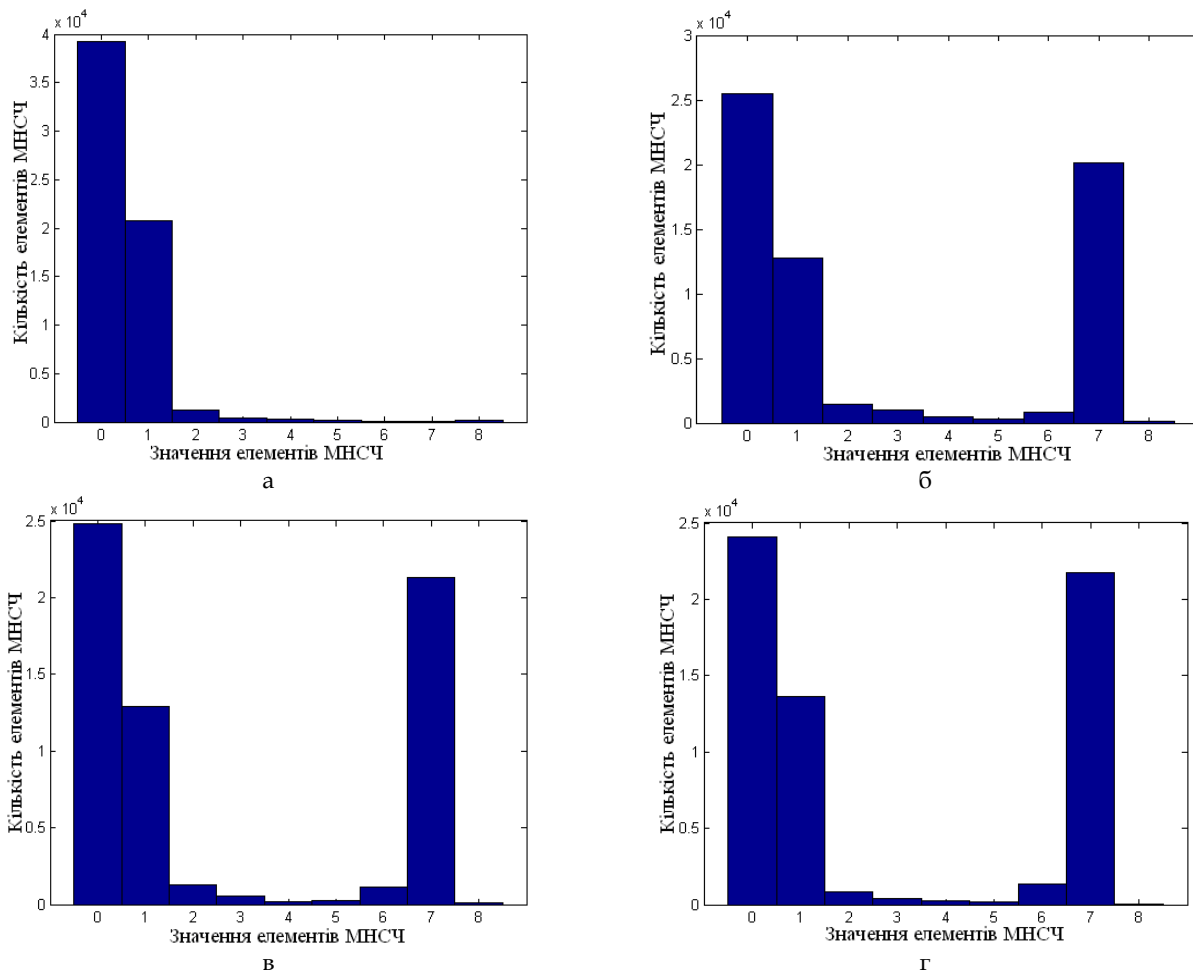


Рис.1. Гістограми значень елементів МНСЧ блоків відповідних ЦЗ:

а – ЦЗ в форматі без втрат; б, в, г – відповідні ЦЗ в форматі з втратами (Jpeg, $QF = 80, 75, 70$ відповідно)

Крок 1. Матрицю ЦЗ F розбити стандартним чином на непересічні $l \times l$ -блоки B_{ij} , $i = 1, \dots, \left\lceil \frac{n}{l} \right\rceil$,
 $j = 1, \dots, \left\lceil \frac{m}{l} \right\rceil$.

Крок 2. Для кожного блоку B_{ij} визначити кількість його СНЧ, що задовольняють умові (3) для кожного значення параметру $T \in \{T_1, T_2, \dots, T_r\}$, що задовольняє умові (4), при цьому: $T_1 < T_2 < \dots < T_r$.

Крок 3. Побудувати для досліджуваного ЦЗ МНСЧ блоків для кожного $T \in \{T_1, T_2, \dots, T_r\}$: M_1, M_2, \dots, M_r відповідно.

Крок 4. Побудувати гістограми $\Gamma_1, \Gamma_2, \dots, \Gamma_r$ значень елементів матриць M_1, M_2, \dots, M_r відповідно.

Крок 5. Знайти моди гістограм $\Gamma_1, \Gamma_2, \dots, \Gamma_r$: md_1, md_2, \dots, md_r та максимальні значення матриць M_1, M_2, \dots, M_r : m_1, m_2, \dots, m_r відповідно.

Якщо

$$\left(md_1 > \bar{P}_1 \right) \vee \left(md_2 > \bar{P}_2 \right) \vee \dots \vee \left(md_r > \bar{P}_r \right),$$

де $\bar{P}_1, \bar{P}_2, \dots, \bar{P}_r$ - параметри, що визначаються залежно від l , то ЦЗ в форматі з втратами;
 якщо

$$\left(0 \leq md_1 \leq \bar{P}_1 \right) \& \left(0 \leq md_2 \leq \bar{P}_2 \right) \& \left(m_1 \leq \overline{PP}_1 \right) \& \left(m_2 \leq \overline{PP}_2 \right),$$

де $\overline{PP}_1, \overline{PP}_2$ - параметри, що визначаються залежно від l ; то ЦЗ в форматі без втрат;
 якщо

$$\left(0 \leq md_1 \leq \bar{P}_1 \right) \& \left(0 \leq md_2 \leq \bar{P}_2 \right) \& \left(\left(m_1 > \overline{PP}_1 \right) \vee \left(m_2 > \overline{PP}_2 \right) \right),$$

то ЦЗ в форматі без втрат;

якщо

$$\left(m_1 > \overline{PP}_1 \right) \& \left(k_{>P} > K \right),$$

де P - поріг між значеннями елементів МНСЧ блоків, які є специфічними/неспецифічними для ЦЗ в різних форматах збереження, K - порогове значення (у відсотках) для кількості елементів МНСЧ блоків, що перевищують P , від загальної кількості елементів, для відокремлення ЦЗ в різних форматах; визначається експериментальним шляхом, залежить від l , то ЦЗ в форматі з втратами.

Для розробки алгоритмічної реалізації запропонованого методу відокремлення ЦЗ в різних (з/без втрат) форматах треба визначити значення

використовуваних їм параметрів. Враховуючи, що існуючі найпоширеніші алгоритми збереження ЦЗ з втратами (Jpeg, Jpeg2000) розбивають матрицю зображення на блоки розміром 8×8 пікселів, далі розглядається $l = 8$.

Поріг P , з урахуванням властивостей МНСЧ блоків, встановлених в [9] і згаданих вище в п. 2, має сенс взяти $P = 4$, тобто $k_{>P} = k_{>4}$.

Користуючись результатами досліджень, проведених в [9], в алгоритмічній реалізації запропонованого методу використовується: $r = 4$, $T_1 = 0.1$, $T_2 = 0.3$, $T_3 = 0.5$, $T_4 = 0.7$; для 8×8 -блоків: $\bar{P}_1 = 1$, $\bar{P}_2 = 1$, $\bar{P}_3 = 2$, $\bar{P}_4 = 3$, $\overline{PP}_1 = 4$, $\overline{PP}_2 = 4$. Визначення СНЧ для кожного блоку B_{ij} на кроці 2 робиться за допомогою сингулярного розкладання матриці B_{ij} .

Для визначення K проведемо обчислювальний експеримент. При проведенні обчислювальних експериментів скрізь в роботі використовується ЦЗ без втрат (Tif) з бази [22], які для отримання відповідних ЦЗ в форматі з втратами Perezберігаються в формат Jpeg з $QF \in \{70, 75, 80\}$ (найпоширеніші на практиці коефіцієнти якості). Результати обчислювального експерименту, що підтверджують зроблений вище теоретичний висновок, наведені на рис. 2. Кількісно на основі отриманих гістограм можна оцінити значення K для відокремлення ЦЗ в різних форматах збереження лише як:

$$K < 10. \quad (5)$$

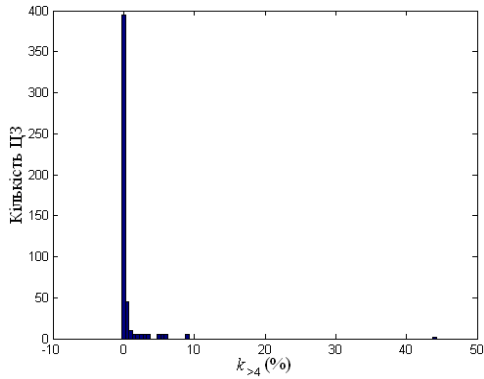
Але цього замало для алгоритмічної реалізації. Очевидно, що зміна значення K буде відображатися на співвідношенні помилок першого та другого роду при роботі відповідних алгоритмічних реалізацій методу: чим більше K , тим менше буде помилок другого роду, тим точніше будуть визначатися оригінальні ЦЗ в форматі без втрат, але тим більшою буде кількість помилок першого роду, що свідчить про пропуск при роботі алгоритму зображення, яке було Perezбереженим у формат без втрат, що відрізняється від його первісного формату - з втратами. Для встановлення «компромісного» значення K був проведений обчислювальний експеримент, в ході якого були протестовані алгоритмічні реалізації розробленого методу з різними значеннями K (5), результати чого наведені в табл. 1, рис. 3. З рис. 3 очевидно, що значення $K \approx 6.3$ забезпечить порівнянність помилок першого і другого роду, що є бажаним на практиці. Тестування алгоритму при $K = 6.3$ забезпечило: 3% - помилки першого роду, 4% - помилки другого роду. Для порівняльної оцінки ефективності алгоритмічної реалізації розроблено методу, результати якої наведені в табл. 2, були обрані сучасні аналоги: M1 (2013) [15], M2 (2016) [17], M3 (2017) [18].

Як видно з табл. 2, алгоритмічна реалізація запропонованого методу дещо поступається МЗ в сенсі помилок першого роду і, як наслідок в сенсі значення точності АСС (2).

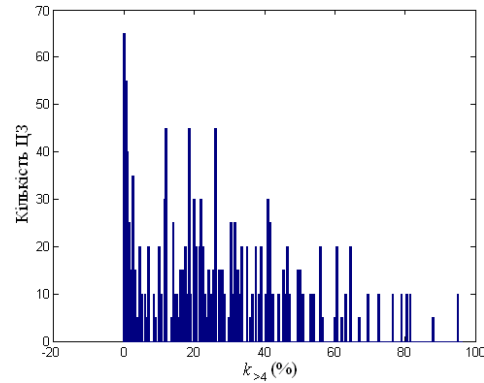
Цей показник можливо покращити наступним чином. МНСЧ блоків для ЦЗ в форматі з втратами дає можливість встановити місце розташування контурів об'єктів, про які свідчить значна зміна значень елементів, що знаходяться поруч в МНСЧ (від значень з множини $\{0, 1\}$, які відповідають блокам, що містять контури, деталі, оскільки найменші СНЧ

в них в більшості перевищують T , до значень з множини $\{2, \dots, 8\}$ - фонові блоки, високочастотна складова в яких порівнянна з нулем), що для МНСЧ блоків відповідного ЦЗ без втрат не буде виглядати настільки «яскраво», навіть, може взагалі не проявитися (рис. 4).

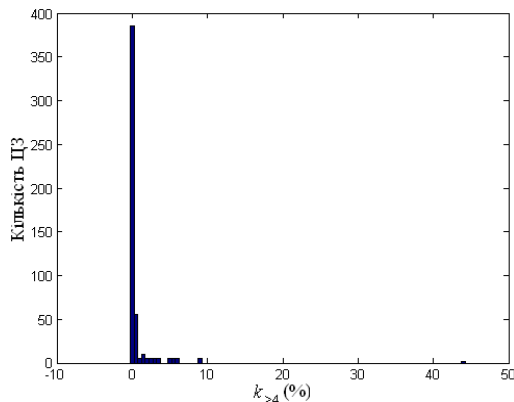
Отже, врахування такої відмінності між МНСЧ блоків дає додаткову можливість перевірки ЦЗ для відокремлення, отже сприятиме підвищенню ефективності розробленої алгоритмічної реалізації методу.



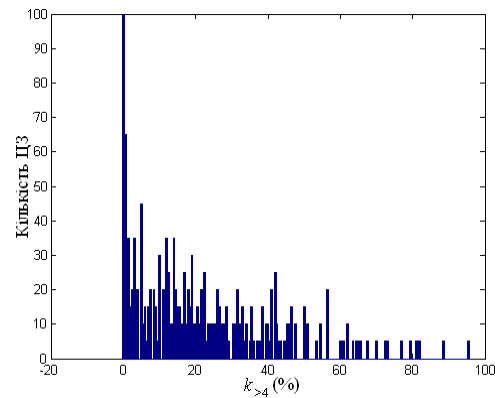
а



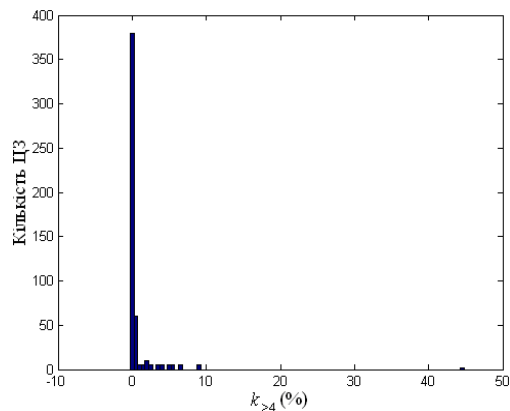
б



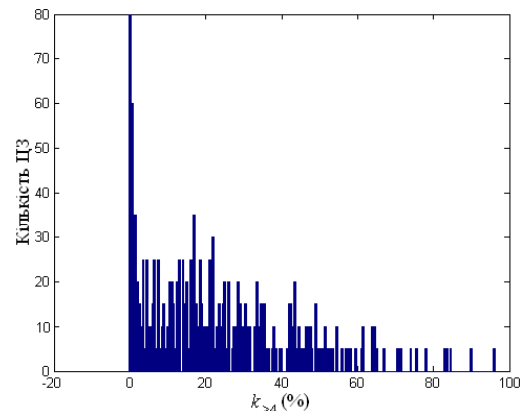
в



г



д



е

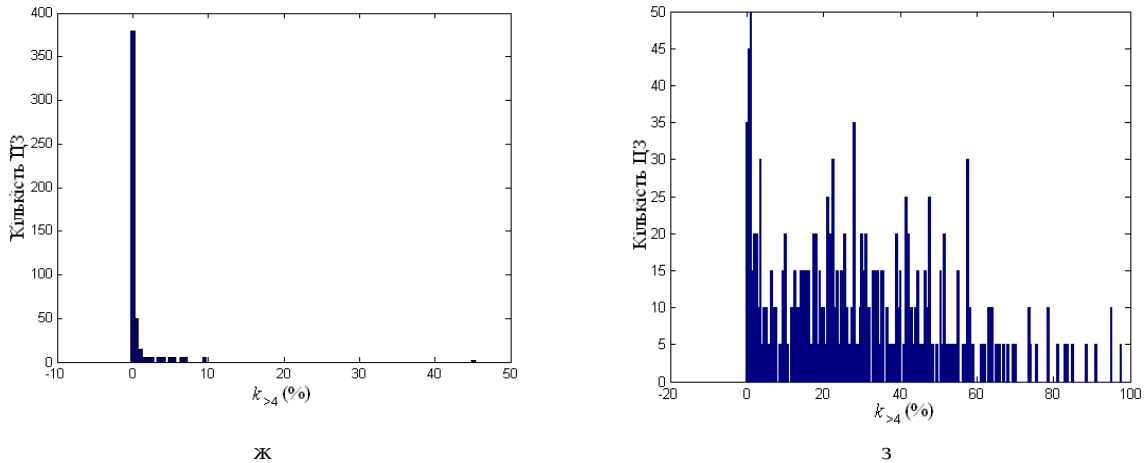


Рис.2. Гістограми значень $k_{>4}$ для відповідних ЦЗ, задіяних в обчислювальному експерименті: а - ЦЗ в форматі Tif ($T=0.1$); б - ЦЗ в Jpeg ($T=0.1$); в - ЦЗ в Tif ($T=0.3$); г - ЦЗ в Jpeg ($T=0.3$); д - ЦЗ в Tif ($T=0.5$); е - ЦЗ в Jpeg ($T=0.5$); ж - ЦЗ в Tif ($T=0.7$); з - ЦЗ в Jpeg ($T=0.7$)

Таблиця 1
 Залежність ефективності алгоритмічної реалізації запропонованого методу від параметру K

K (%)	Помилки першого роду (%)	Помилки другого роду (%)
3	1	9
5	2.7	6
7	4.7	3

Таблиця 2
 Порівняльна оцінка ефективностей методів відокремлення ЦЗ в різних форматах збереження

Метод	M1 (2013)	M2 (2016)	M3 (2017)	Розроблений метод (2023)
Помилки I роду (%)	4.6	3	0.5	3
Помилки II роду (%)	7.1	4.5	5	4
ACC	0.9415	0.9625	0.9725	0.965

З урахуванням останнього зауваження, крок 5 запропонованого методу можна вдосконалити наступним чином:

Крок 5 (удосконаленого методу):

якщо

$$\left(0 \leq md_1 \leq \bar{P}_1\right) \& \left(0 \leq md_2 \leq \bar{P}_2\right) \& \left(\left(m_1 > \overline{PP}_1\right) \vee \left(m_2 > \overline{PP}_2\right)\right),$$

то ЦЗ в форматі без втрат;

якщо

$$\left(\left(m_1 > \overline{PP}_1 \right) \& \left(k_{>p} > K \right) \& \left(\begin{array}{l} \text{МНСЧ блоків мас} \\ \text{перепали значень елементів, які стоять поряд, що} \\ \text{відповідають наявним} \\ \text{контурам ЦЗ} \end{array} \right) \right),$$

то ЦЗ в форматі з втратами.

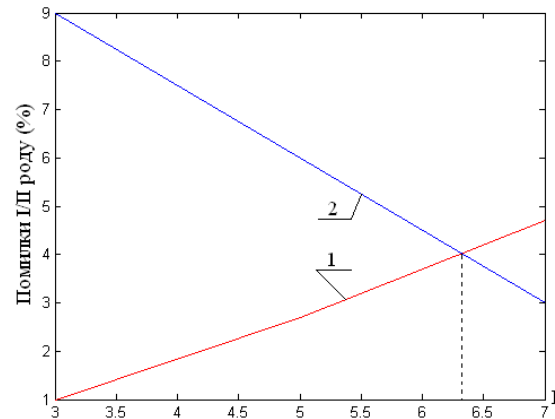


Рис.3. Графіки залежності помилок першого/другого роду від параметра K : 1 - помилки першого роду; 2 - помилки другого роду

Додаткова перевірка дала можливість знизити помилки першого роду у відповідному алгоритмі до 1.3%, ACC відповідно зростає до 0.9735, що на 0.1% краще за значення цього показника для кращого з аналогів - M3.

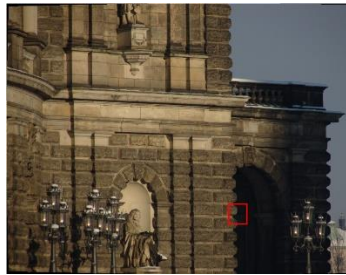
Для остаточного висновку в сенсі вибору з розглянутих аналогів такого, якому слід надати перевагу при розв'язку задачі, що розглядається, порівнюємо обчислювальні складності запропонованого методу і M3. Обидва методи є блоковими, тому для

експертизи ЦЗ з $n \times n$ -матрицею вимагають кількість операцій, що визначається кількістю отриманих при стандартному розбитті матриці зображення блоків, тобто становить $O(n^2)$.

Очевидно, при порівнянні обчислювальних складностей тут все буде залежати від коефіцієнта при n^2 . Цей коефіцієнт для розробленого методу визначається, головним чином, кількістю операцій для побудови сингулярного розкладання $l \times l$ -блоку для визначення його сингулярного спектру. Така ж операція обробки блоку є основною і в МЗ.

Тобто якщо б при аналізі ЦЗ воно досліджувалося в МЗ одноразово, без Perezберезження з наступним дослідженням отримуваних ЦЗ по аналогічній схемі, то обчислювальні складності МЗ і розробленого методу були б порівнянними.

Але необхідність в процесі аналізу методом МЗ Perezберезження ЦЗ збільшує його обчислювальну складність, в порівнянні з запропонованим, в стільки разів, скільки разів відбувається Perezберезження (при використанні як формату з втратами Jpeg, Perezберезження може відбуватися до 100 разів).



а



б

0	1	1	0	0	1	0	0	0	5	2	7	7	7	7	7	5	7	6	7	6	7	
0	0	1	1	0	0	1	0	0	3	2	7	6	7	7	7	5	7	7	7	7	6	3
0	0	1	1	0	0	1	0	0	0	1	2	3	6	7	7	7	7	6	7	6	4	4
0	0	1	0	0	1	0	0	0	1	0	0	4	7	7	7	7	7	7	7	6	7	7
0	1	0	1	1	0	1	0	0	0	0	1	7	7	7	7	7	5	4	5	6	6	6
1	0	0	1	0	0	0	1	0	0	0	0	7	7	7	7	6	6	7	7	4	7	7
1	0	0	0	0	0	0	0	0	0	0	1	1	6	7	7	6	4	5	7	7	7	7
1	1	0	1	0	1	0	1	0	0	0	0	7	7	7	7	6	7	6	7	6	6	6
0	0	0	0	1	0	1	0	1	0	1	2	0	1	6	7	6	2	6	7	7	6	6
0	0	0	0	0	0	0	0	1	0	1	0	2	6	7	7	3	6	4	6	7	7	7
0	0	0	0	0	0	0	1	0	0	0	4	7	7	6	5	5	4	3	7	7	7	7
1	0	0	0	1	0	0	0	0	0	2	3	7	7	7	3	7	3	1	4	4	4	4
2	0	0	0	1	0	0	1	2	7	6	7	7	7	7	3	7	4	4	4	4	4	4
0	0	1	0	1	0	0	0	3	7	3	7	7	7	7	7	7	6	7	7	7	7	7
0	1	0	1	0	1	0	1	0	3	3	7	3	7	7	7	7	4	6	7	7	7	7
1	0	0	1	0	0	0	0	1	0	2	7	4	7	7	4	7	3	7	7	7	7	7
0	0	0	0	0	0	0	0	0	1	1	5	6	7	7	3	6	6	7	7	7	7	7
0	0	1	1	1	0	0	0	0	1	0	3	7	7	7	7	6	6	6	6	6	6	6
1	1	1	0	1	1	1	1	0	1	1	1	6	7	7	6	5	4	5	6	6	6	6
0	0	1	0	0	1	1	0	0	1	1	1	7	7	7	6	3	4	6	7	7	7	7

в

0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1
1	0	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1
0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0
1	1	1	0	0	0	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	0
0	0	1	1	0	0	0	1	0	0	1	0	1	0	1	0	0	1	1	0	1	0	0
1	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	0	1	1
0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1
0	1	0	0	0	1	1	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0
0	0	1	1	1	0	1	0	0	0	0	1	0	1	1	1	1	0	0	0	1	0	1
0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	1	0	0	1	1	0	1	0
1	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	1	0
1	1	0	0	0	0	0	1	0	1	1	0	0	0	1	1	0	0	1	0	1	0	0
1	0	1	1	0	0	1	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1
1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0	0

г

Рис.4. Виявлення контурів на ЦЗ за допомогою МНСЧ блоків: а – подане ЦЗ; б – частина ЦЗ, для якої будується МНСЧ; в – МНСЧ блоків для ЦЗ в форматі Jpeg (QF=70); г – МНСЧ блоків для відповідного ЦЗ в форматі Tif

Таким чином, хоча абсолютне підвищення ефективності запропонованого методу в порівнянні з найкращим аналогом є незначним (0.1%), вказані вище недоліки методу МЗ, пов'язані з його значною обчислювальною складністю, роблять для розв'язку задачі, що розглядається, пріоритетним метод, розроблений в роботі.

Висновки. В роботі вирішено важливу науково-практичну задачу підвищення ефективності процесу відокремлення ЦЗ в різних (з/без втрат) форматах первісного збереження шляхом розробки нового методу, заснованого на аналізі властивостей матриці найменших СНЧ блоків.

В ході розробки теоретичного базису методу:

-отримав подальший розвиток підхід, заснований на аналізі МНСЧ блоків, запропонований в [9]: обґрунтовано та практично підтверджено, що для ЦЗ в форматі без втрат кількість блоків, що мають значну кількість СНЧ, близьких до нуля (тобто кількість елементів в МНСЧ, що перевищують порогове значення P), буде менше, чим аналогічна кількість для ЦЗ в форматі з втратами;

-встановлено: чим менше коефіцієнт якості QF , що використовується при стиску ЦЗ, тим більше в відповідній МНСЧ блоків буде елементів, значення яких перевищують P , тим більші абсолютні значення будуть мати елементи матриці.

Всі теоретичні висновки знайшли своє практичне підтвердження в ході обчислювальних експериментів.

Алгоритмічна реалізація розробленого методу дозволила підвищити ефективність відокремлення ЦЗ в різних форматах збереження в сенсі показника АСС на 0.1% в порівнянні з найкращим з аналогів, при цьому маючи обчислювальну складність набагато меншу, ніж складність аналога.

Отримані результати дають можливість говорити про наявність підвищення в цілому ефективності виявлення порушень цілісності ЦЗ, зокрема при стегааналізі, виявленні фотомонтажу.

ЛІТЕРАТУРА

[1] Deineko, Zh., Zeleniy, O., Lyashenko, V., Tabakova, I. Color space image as a factor in the choice of its processing technology. Abstracts of I International scientific-practical conference «Problems of modern science and practice» (September 21-24, 2021). Boston, USA, pp. 389-394.

[2] Dougherty, E.R. Digital image processing methods. CRC Press LLC, 1994. 504 p.

[3] Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.: Техносфера, 2006. 1070 с.

[4] Congbo, L., Yunhui, H., Zihe, T. Research on Digital Image Processing Technology and Its Application. Proceedings of the 2018 8th International Conference on Management, Education and Information (ME-ICI 2018) pp. 587-592.

[5] Marakumbi, P.R., Khanapuri, J.V. A Study on Image Authentication Methods. International Research Journal of Engineering and Technology. 2018. pp. 1719-1721.

[6] Cromey, D.W. Digital Images Are Data: And Should Be Treated as Such. Methods Mol Biol. 2013. pp. 1-27.

[7] Kobozeva, A., Bobok, I. Method for detecting digital image integrity violations due to its block processing. Radiotekhnika. 2019. pp. 130-141.

[8] Karthikeyan, N., Saravana Kumar, N.M., Mugunthan, S.R. Comparative study of lossy and lossless image compression techniques. International Journal of Engineering & Technology. 2018. pp. 950-953.

[9] Бобок, І.І., Кобозева, А.А., Маєвський, Д.А. Дослідження параметрів блоків матриці цифрового контенту в різних форматах збереження як теоретична основа для методів виявлення порушення його цілісності. Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. 2022. С. 41-58.

[10] Taher, M.M., Ahmad, A.R., Hameed, R.S., Mokri, S.S. A literature review of various steganography methods. Journal of Theoretical and Applied Information Technology. 2022. pp. 1412-1427.

[11] Aggarwal, A., Sangal, A., Varshney, A. Image steganography using LSB algorithm. International Journal of Information Sciences and Application. 2019. pp. 85-89.

[12] Dhawan, S., Gupta, R. Analysis of various data security techniques of steganography: a survey. Information Security Journal: A Global Perspective. 2021. pp. 63-87.

[13] Кобозева, А.А. Использование особенностей возмущений сингулярных чисел матрицы цифрового изображения для обнаружения его фальсификации. Штучный интеллект. 2008. С. 145-153.

[14] Jalab, H.A., Subramaniam, T., Ibrahim, R.W., Kahtan, H., Mohd Noor, N.F. New texture descriptor based on modified fractional entropy for digital image splicing forgery detection. Entropy. 2019. 371 p.

[15] Бобок, І.І. Выявление цифровых изображений, пересохраненных в формат без потерь из формата с потерями, как составная часть стегаанализа. Сучасна спеціальна техніка. 2013. С. 64-70.

[16] Костырка, О.В. Анализ преимуществ пространственной области цифрового изображения-контейнера для стегаанопреобразования. Информатика та математичні методи в моделюванні. 2013. С. 275-282.

[17] Akhmetieva, A.V. Method of detection the fact of compression in digital images as an integral part of steganalysis. Информатика та математичні методи в моделюванні. 2016. С. 357-364.

[18] Бобок, І.І. Метод виявлення зображень, Perezberzhenniy u format bez vtraty z formatu z vtratyami. Matematychny ta kompyuterny modelyuvannya. 2017. С. 5–14.

[19] Geetha, S., Sindhu, S., Kamaraj, N. Close color pair signature ensemble adaptive threshold based steganalysis for LSB embedding in digital images. Transactions on Data Privacy.2009. pp. 140–161.

[20] Деммель Д. Вычислительная линейная алгебра: теория и приложения. М.: Мир, 2001. 430 с.

[21] Кобозева А.А., Хорошко В.А. Анализ информационной безопасности. К.: ГУИКТ, 2009.251 с.

[22] Gloe, T., Böhme, R. The “Dresden Image Database” for benchmarking digital image forensics. Proceedings of the 2010 ACM Symposium on Applied Computing (SAC '10). New York, 2010. pp. 1585–1591.

УДК 004.056.5

Bobok I. Method of separation of digital images in different original storage formats based on the properties of the matrix of smallest singular values of blocks

Abstract. Today digital images are one of the main carriers of information, and the integrity of the image is a necessary condition for the possibility of its use for non-entertainment purposes in any field of human activity. Changing the format of an entire image or changing the format of a part of it is often the result of making unauthorized changes to the image, making the detection of the fact of re-saving to a format different from the original one an indicator of a violation of the integrity of the image. Thus, the task of separating images in different (with/without loss) formats, increasing the efficiency of which by developing a suitable method is the aim of this work, and urgent task of modern information security. The proposed method is based on the analysis of the properties of the matrix of the smallest singular values of blocks, which corresponds to the image. The algorithmic implementation of the method made it possible to increase the efficiency of separation of images in various original storage formats in comparison with analogues, as a result of which there is an overall increase in the efficiency of detecting violations of the integrity of the image, in particular during steganalysis, detection of photomontage, etc.

Keywords: digital image, lossy format, lossless format, singular value, block method.

Бобок Іван Ігорович, доктор технічних наук, доцент, доцент кафедри комп'ютеризованих систем та програмних технологій Національного університету «Одеська політехніка».

Ivan Bobok, Dc.S, Associate Professor, Associate Professor of the Department of Computerized Systems and Software Technologies of the Odesa Polytechnic National University.

Отримано 4 жовтня 2022 року, затверджено редколегією 14 листопада 2022 року
