УДК 004.932

А.О. Коваленко, А.О. Різуненко

Військовий інститут телекомунікацій та інформатизації НТУУ „КПІ”, Полтава

ДОСЛІДЖЕННЯ КІЛЬКОСТІ ТА СЕРІЙ ПОМИЛОК У ДЕКОДОВАНОМУ ПОВІДОМЛЕННІ ДЛЯ СТЕГОАЛГОРИТМУ З ЦЕНТРУВАННЯМ ЗНАЧЕНЬ КОЛІРНИХ КОМПОНЕНТ ЗОБРАЖЕННЯ

Проведено аналіз стеганографічного методу центрування значень колірних компонент зображень. Визначено характер локалізації помилок, які виникають у декодованому повідомленні, а також залежність кількості помилок у пакеті від коефіцієнту стиснення та ширини інтервалу вбудовування. Обрано код, який необхідно застосовувати для виправлення помилок, що вносяться при компресії стего.

Ключові слова: *стеганографія, стего, контейнер, мультимедійні дані, ЦВЗ, робасний метод, JPEG-2000, пакетні помилки, коефіцієнт помилок.*

Вступ

На сьогоднішній день криптографічні методи не дозволяють в повній мірі забезпечити захист інформації від несанкціонованого доступу. Особливо це питання гостро постає в області захисту мультимедійних даних. Тому все частіше разом із засобами криптографії для захисту таких видів інформації додатково застосовуються стійкі до модифікацій стеганографічні протоколи.

На основі аналізу викривлень, які вносяться у зображення під час стиснення з втратами, було розроблено *метод центрування значень колірних компонент* [1]. Принцип методу заснований на тому, що 90–95% значень колірних компонент зображень під час компресії змінюють своє значення не більше, ніж на 3-5 рівнів градації яскравості.

Таким чином, змінюючи значення колірних компонент контейнера відповідно до біт інформаційного повідомлення, існує можливість після стиснення відновити значення цих біт.

Цей метод дозволяє (в залежності від обраної ширини інтервалу впровадження, коефіцієнту стиску та частоти колірного перепаду зображення) відновити від 70% до 95% початкового інформаційного повідомлення (табл. 1) [1].

Основний недолік розробленого методу – відсутність повного виправлення помилок, які вносяться в приховане повідомлення під час стиснення з втратами. Тому основна **мета дослідження –** виявлення характеру локалізації помилок, що вносяться, а також оцінка можливості їхнього виправлення та вибір завадостійкого коду.

Таблиця 1

Залежність значень $K_{пом}$ від інтервалу вбудовування, частоти колірного перепаду та коефіцієнту компресії

Частота колірного перепаду	Інтервал вбудовування	Коефіцієнт викривлення $K_{пом}$, %	
		$k_{стиску} = 5$	$k_{стиску} = 10$
0,8	w = 16	**5,36**	**28,83**
	w = 32	**1,022**	**19,044**
0,9	w = 16	**13,16**	**30,24**
	w = 32	**4,98**	**28,54**

Аналіз можливості виправлення помилок та вибір завадостійкого коду

Для визначення можливості виправлення помилок, утворених в результаті стиснення з втратами, необхідно спочатку виявити характер викривлень у відтвореному повідомленні.

Спочатку було проаналізовано просторову локалізацію помилок в самому зображенні, тобто, виявлено, в яких пікселях зображення виникло викривлення інформаційного біту. Сутність аналізу полягає у наступному: початкове і декодоване повідомлення побітно додаються за модулем 2 (елемент нерівнозначності, XOR). У сумі одиницями будуть позначені ті біти, в яких виникла помилка після стиснення з втратами. Потім отриману суму вбудовуємо у повністю біле зображення з розмірами, що дорівнюють розмірам контейнера. При цьому використовуємо метод вбудовування, аналогічний до методу центрування, але тепер ті пікселі, в які записуєть-

ся одиниця, перетворюємо на чорні. Таким чином буде отримано зображення, що відображає чорними крапками місця локалізації помилок (рис. 1).

У результаті аналізу просторової локалізації помилок було виявлено, що різні області зображення містять різну кількість помилок. Як бачимо з прикладу (рис. 1), у більш монотонних областях зображення з'являється менша кількість помилок. Така закономірність викривлень, що вносяться під час стиснення, пояснюється особливістю формату JPEG-2000: під час компресії відбувається розбиття зображення на квадратні блоки 16x16, 32x32, 64x64 або 128x128 пікселів (у даному випадку на блоки 128x128 пікселів) для їх подальшої обробки. Якщо частота колірного перепаду блока відносно велика (0,95-0,99), то вноситься більше помилок у приховане повідомлення, якщо менша (0,9 і менше) – то коефіцієнт викривлення буде меншим.

а б

Рис. 1. Залежність просторової локалізації помилок у декодованому повідомленні для сильно насиченого зображення (у синій колірній компоненті): а – контейнер; б – місця локалізації помилок у контейнері (місця, де виникли помилки, показані чорними крапками)

Така закономірність характерна лише для зображень, у яких частота колірного перепаду в різних областях різна (сильнонасичені зображення, у яких присутні монотонні області).

Для цілком монотонних зображень з відносно невеликою частотою колірного перепаду (0,7-0,8), а також для сильнонасичених зображень(0,95-0,99), у яких відсутні монотонні області, помилки розташовуються приблизно рівномірно, але з різними коефіцієнтами помилок (рис. 2, 3):

Рис. 2. Залежність просторової локалізації помилок у декодованому повідомленні для зображення з частотою колірного перепаду 0,7 ($К_{пом}$=13%)

Рис. 3. Залежність просторової локалізації помилок у декодованому повідомленні для зображення з частотою колірного перепаду 0,99 ($К_{пом}$=25%)

Треба також зауважити, що просторова локалізація помилок залежить від колірної компоненти, в яку вбудовано повідомлення. На рис. 4 зображено рівномірну просторову локалізацію помилок у червоній колірній компоненті для того ж самого зображення, що і на рис. 1.

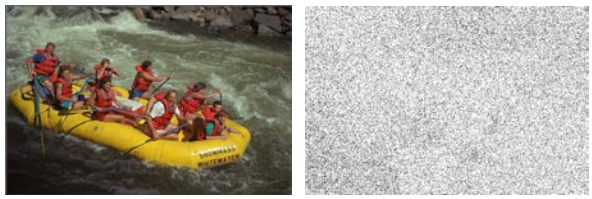

Рис. 4. Залежність просторової локалізації помилок у декодованому повідомленні для сильно насиченого зображення (у червоній колірній компоненті)

Проаналізувавши ці факти, можна зробити висновок, що при приховуванні даних у зображення необхідно враховувати насиченість блоків, з якими працює агент стиснення зображення (у данному випадку JPEG-2000), та приховувати дані у кожен блок окремо. Якщо величина блока заздалегідь невідома, то необхідно обирати розмір 16x16, як найменші, що застосовуються форматом.

На наступному етапі аналізу статистичних характеристик стего досліджувалася максимальна довжина серій пакетних (безперервних) помилок (табл. 2). Ця величина показує, яка відносна кількість помилок у відсотках припадає на одиночні, об'єднані у подвійні, потрійні і т.д.

Таким чином, сума значень відносної кількості пакетних помилок усіх довжин буде дорівнювати коефіцієнту помилки зображення.

Як і у випадку з коефіцієнтом помилок, відносна кількість викривлень, що припадає на довші пакети, зростає з ростом коефіцієнту стиснення та зменшенням ширини інтервалу вбудовування. Тобто, чим більше компресія і менше інтервал, тим більше помилок буде у подвійних, потрійних серіях і т.д.

Крім того, необхідно звернути увагу на те, що при відносно невеликій компресії (з $К_{ст}$=5) довжини серій у більшій мірі залежать від ширини інтервалу вбудовування (9-10 біт підряд для w=32, та 13-14 для w=16) ніж при більших коефіцієнтах стиснення ($К_{ст}$=10 і більше). Тому з ростом коефіцієнту компресії відпадає необхідність збільшувати інтервал вбудовування.

На основі даних з табл. 2 було побудовано графік, зображений на рис. 5. Виходячи з графіка можна зробити висновок, що розподіл помилок за пакетами в залежності від його довжини зменшується експоненціально, тобто аналогічно до рівномірного закону розподілу.

Таким чином, характер появи помилок у декодованому повідомлені не носить пакетний характер і тому при виправленні не потребує застосування кодів, що виправляють серії помилок.

Таблиця 2

Залежність кількості помилок у пакеті від інтервалу вбудовування та коефіцієнту компресії, %

Довжина пакету	**k=5**		**k=10**	
	w=32	**w=16**	**w=32**	**w=16**
1	2,400	7,922	13,796	14,039
2	0,498	2,834	8,055	10,749
3	0,209	1,067	4,005	6,547
4	0,065	0,373	1,814	3,653
5	0,038	0,187	0,859	1,973
6	0,017	0,079	0,394	1,052
7	0,008	0,030	0,182	0,556
8	0,003	0,012	0,077	0,288
9	0,001	0,007	0,034	0,156
10		0,005	0,016	0,086
11		0,002	0,015	0,045
12		0,004	0,006	0,031
13		0,005	0,005	0,016
14		0,001	0,003	0,011
15			0,004	0,005
16				0,002
17				0,001

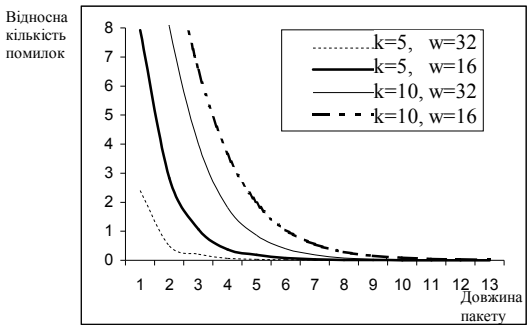

Рис. 5. Графік залежності відносної кількості помилок від довжини пакета для різних коефіцієнтів компресії та ширини інтервалу вбудовування

Отже, виходячи з результатів аналізу просторової локалізації помилок та розподілу пакетних помилок, можна зробити висновок, що з метою виправлення викривлень, які виникли в результаті стиснення зображення, доцільно застосовувати різновид циклічних виправляючих кодів – код Боуза-Чоудхурі-Хоквінгема (код БЧХ), як такий, що дозволяє виправляти помилки будь-якої кратності при відносно простій програмній реалізації [8].

При виборі різновиду коду БЧХ необхідно додержуватися наступних вимог [9,10]:

1) довжина кодового слова: $n = 2^h - 1$, де h – ціле число;

2) кількість перевірочних біт: $r = n - k \le ht$, де t – кратність помилки;

3) мінімальна кодова відстань: $d_{min} \ge 2t + 1$.

Виходячи з цих даних, було побудовано табл. 3, у якій відображаються залежності між виправляючою здатністю та кількістю перевірочних біт для кодів БЧХ з довжинами кодової послідовності 7, 15, 31 та 63 біт. У правій частині таблиці приведені значення середньої кількості помилок, які можуть виникнути в кодовому слові заданої величини при стисненні з коефіцієнтами 5 та 10 відповідно (див. табл. 1).

З табл. 3 видно, що для виправлення помилок, які виникають при стисненні контейнера з коефіцієнтом компресії не більше 5, найбільш доцільним є використання кодів БЧХ (7,4), БЧХ (31,6) та БЧХ (63,3) як таких, що дозволяють виправити помилки, і при цьому вносять найменшу надлишковість. Застосовувати коди з більшою довжиною кодового слова не доцільно через їхню нездатність виправити помилки, які виникають у кодовому слові такої довжини (наприклад, код БЧХ (127,1) здатен виправити помилку кратністю 18, але при $К_{пом}$ = 15% у такому кодовому слові виникає приблизно 19 помилок).

Таблиця 3

Залежність надлишковості кодів БЧХ від дожини кодового слова та виправляючої можливості

Довжина кодового слова	Кількість інформаційних біт	Кількість перевірочних біт	Мінімальна кодова відстань	Виправляюча здатність	Надлишковість коду, %	Кількість помилок у блоці	
						Кпом=15% (Кст=5)	Кпом=30% (Кст=10)
7	4	3	3	1	75,00%	1	2
15	7	8	5	2	53,33%	3	5
15	3	12	7	3	80,00%	3	5
31	11	20	9	4	64,52%	5	10
31	6	25	11	5	80,65%	5	10
31	1	30	13	6	96,77%	5	10
63	9	54	19	9	85,71%	10	19
63	3	60	21	10	95,24%	10	19
127	1	126	37	18	99,21%	19	38,1

Висновки

У результаті проведеної роботи було досліджено характер просторового розташування помилок у стего. Виявлено, що для зображень з відносно високою частотою колірного перепаду (0,9–0,99) помилки локалізуються в блоках, на які розбивається зображення під час компресії форматом JPEG-2000. Тому на етапі кодування необхідно проводити додатковий аналіз областей, в які вбудовується повідомлення та записувати повідомлення у квадратні блоки з такими розмірами, з якими працює агент стиснення.

Проаналізовано відносну кількість серій пакетних помилок. Виявлено, що характер появи помилок непакетний, і вони розподілені у безперервні серії так само, як і при рівномірному законі розподілу.

Проаналізовано можливість застосування кодів БЧХ для виправлення викривлень у декодованому повідомленні. Зроблено висновок, що існує можливість застосування кодів БЧХ (7,4), БЧХ (31,6) та БЧХ (63,3) для виправлення помилок, які виникають після компресії зображення з коефіцієнтом стиснення 5.

Для виправлення помилок після більш сильного стиснення необхідно застосовувати інші коди.

Подальші дослідження слід направити на практичне застосування виправляючих кодів та підбору коду, який дає можливість виправити помилки, які виникають після стиснення з коефіцієнтами, більшими за 5.

Список літератури

1. Коваленко А.О., Різуненко А.О. Метод приховування даних у зображення шляхом центрування значень колірних компонент // Системи обробки інформації. – Х.: ХУПС, 2006. – Вип. 7 (56). – С. 12-16.

2. Kutter M., Voloshynovskiy S., Herrigel A. The Watermark Copy Attack // Proceedings of SPIE: Security and Watermarking of Multimedia Content II. 2000. Vol. 3971.

3. Christian Cachin. An Information-Theoretic Model for Steganography // Proceedings of 2nd Workshop on Information Hiding, Lecture Notes in Computer Science, Springer, 1998.

4. Simmons G. The prisoner`s problem and the subliminal channel // Proc. Workshop on Communications Security, 1984. – P. 51-67.

5. Грибунин В.Г. Цифровая стеганография. – С.-Пб.: ВУС, 2000. – 272 с.

6. Прэтт У. Цифровая обработка изображений: Пер. с англ. Кн. 2. – М.: Мир, 1982. – 480 с.

7. Deepa Kundur. Multiresolution Digital Watermarking: Algorithms and Implications for Multimedia. 1999.

8. Fridrich J., Du R., Long M. Steganalysis of LSB encoding in color images // ICME, 2000.

9. Блейхут Р. Теория и практика кодов, контролирующих ошибки: Пер. с англ. – М.: Мир, 1986. – С. 187-238.

10. Жураковський Ю.П., Полторак В.П. Теорія інформації та кодування: підручник. – К.: Вища шк., 2001. – С. 134-157.

Надійшла до редколегії 16.07.2008

Рецензент: д-р техн. наук, проф. М.В. Галай, Полтавський національний технічний університет ім. Юрія Кондратюка, Полтава.

ИССЛЕДОВАНИЕ КОЛИЧЕСТВА И СЕРИЙ ОШИБОК В ДЕКОДИРОВАННОМ СООБЩЕНИИ ДЛЯ СТЕГОАЛГОРИТМА С ЦЕНТРИРОВАНИЕМ ЗНАЧЕНИЙ ЦВЕТОВЫХ КОМПОНЕНТ ИЗОБРАЖЕНИЙ

А.О. Ковалено, А.О. Резуненко

Проведен анализ стеганографического метода центрирования значений цветовых компонент изображения. Определен характер локализации ошибок, возникающих в декодированном сообщении, а также зависимость количества ошибок в пакете от коэффициента сжатия и ширины интервала внедрения. Выбран код, который необходимо применять для исправления ошибок, которые вносятся при компрессии в стего.

***Ключевые слова**: стеганография, стего, контейнер, мультимедийные данные, ЦВЗ, робастный метод, JPEG-2000, пакетные ошибки, коэффициент ошибок.*

INVESTIGATION OF THE NUMBER AND SERIES OF DISTORTIONS IN DECODED MESSAGE FOR STEGOALGORITHM WITH THE CENTERING OF THE COLOR COMPONENTS' MAGNITUDE USED IN IMAGES

A.O. Kovalenko, A.O. Rezunenko

The stegoalgorithm with the centering of the color components' magnitude has been analyzed. The character of distortions' location in decoded message has been defined. The functional dependence between the number of package distortions and compression ratio and embedding widthhas been defined. The antijamming code have been chosen.

***Keywords:** steganography, stego, container, multimedia data, DWM, robust method, JPEG-2000, package distortions, rate of distortions.*