УДК 004.627(043.2)

ІНВАРІАНТНО-ПРОСТОРОВИЙ МЕТОД КОДУВАННЯ ДВІЙКОВИХ ПОСЛІДОВНОСТЕЙ

О. К. Юдін, д-р техн. наук, проф.; ***К. О. Курінь***, ***Р. В. Зюбіна***

kszi@ukr.net

Обґрунтовано доцільність використання методів кодування (відмінних від статистичних) у задачах стиснення даних. Запропоновано метод кодування даних, інваріантний до статистичних — метод інваріантно-просторового кодування (ІПК). Визначено кореляційну ознаку двійкових послідовностей: сумарну бітову кореляцію двійкової послідовності. Визначено процедуру кодування за визначеною кореляційною ознакою, згідно з якою кодуванням ІПК є привласнення двійковій послідовності її порядкового номера у кореляційній групі. Вперше описано та доведено правило розрахунку порядкового номера ІПК двійкової послідовності в кореляційній групі за значенням сумарної бітової кореляції двійкової послідовності. Вперше описано метод оцінки ефективності процедури стиснення, що забезпечується запропонованим методом кодування. Проведено оцінювання очікуваного мінімального ступеня стиснення для послідовностей різної довжини та виконано порівняння за даним параметром з методами кодування, що належать до класів методів нестатистичного стиснення. Здійснено оцінювання коефіцієнта стиснення для двійкових послідовностей у складі двійкового представлення трансформант дискретного косинусного перетворення. Отримані результати дають змогу зробити висновок про доцільність подальшого використання запропонованого методу кодування двійкових даних у технологіях стиснення.

Ключові слова: стиснення, нестатистичне кодування, кореляційні ознаки, бітова кореляція, кореляційна група, коефіцієнт стиснення.

The expedience of the appliyng of the coding methods which are different from statistical methods in the tasks of data compression is grounded. The offered method of data coding, invariant to statistical, is a method invariant-spatial code (ISK). The cross-correlation sign of binary sequences is defined: summary bit correlation of binary sequence. The definition of the procedure of codingaccording to the certain cross-correlation sign is formulated. In obedience to it the ISK code is an appropriation of sequence number of the binary sequence in a cross-correlation group to it. The rule of calculation of sequence number of ISK of binary sequence in a cross-correlation group by value of the total bit correlation of binary sequence is first described and proved. The method of efficiency estimation of the provided compression procedure which is provided with the offered coding method is first described. The estimation of the expected minimum compression coefficient for the sequences of different length is provided and comparison by this parameter with the methods of the classes of unstatistical compression is made. The assessment of compression coefficient for binary sequence in content of discrete cosine transformant is provided and comparison by this parameter with related coding method is provided. The received results allow making a conclusion about expediency of use of the offered method of coding in technologies of images compression.

Keywords: compression, non-statistical coding, correlation signs, binary correlation, correlation group, peak compression coefficient.

Актуальність дослідження

Одним з основних завдань сучасних інформаційно-телекомунікаційних систем (ІТКС) є швидкісна передача даних. Коли йдеться про скорочення часу передачі, то, як правило, мається на увазі два шляхи вирішення цієї проблеми. Перший — збільшення пропускної спроможності каналу передачі, що не завжди доцільно, оскільки частіше за все це пов'язано з технічним переоснащенням каналу передачі даних, що не допустимо з економічного погляду.

Тому більш доцільним є інший шлях вирішення проблеми скорочення часу передачі — зменшення об'єму даних, що передаються. Цей метод тотожний одному з основних завдань теорії інформації — кодуванню джерела або стисненню даних. Особливо це актуально, коли мова йде про передачу одного з найбільш інформативних типів даних — зображень, які, як відомо, характеризуються великим об'ємом.

Найпоширеніші на сьогодні технології стиснення зображень забезпечують високі ступені (із втратами інформації) стиснення за рахунок скорочення психовізуальної надмірності і подальшого статистичного кодування компонент трансформант ортогональних перетворень.

Психовізуальна надмірність скорочується в результаті обнулення високочастотних складових компонент трансформант за умови їх подальшого квантування.

Для кодування квантованих трансформант використовуються статистичні методи стиснення — такі, як кодування Хаффмена (для стандарту JPEG) й арифметичне кодування (JPEG-2000). Для даних методів кодування характерні такі недоліки [1; 3; 8]:

– кількість машинних операцій, що відводиться на обробку трансформованих зображень може становити до 70 % від загальної кількості операцій процедури їх стиснення; це зумовлено тим, що необхідно враховувати витрати кількості операцій на обчислення статистики, на побудову кодових таблиць і на організацію подвійного проходу по оброблюваним даним, а при відновленні весь фрагмент буде відновлений тільки після перекодування всіх нерівномірних кодових слів;

– кількість операцій, необхідних для виконання статистичного кодування для трансформант, сформованих на основі ДКП і ДВП, досягає 80 % від сумарної кількості операцій, що затрачуються на одержання стисненого зображення, й може перевищувати кількість операцій на виконання перетворень (для режимів, які забезпечують високу якість зображення);

– у зв'язку з необхідністю синхронізації й маркування нерівномірних кодових комбінацій на межах оброблюваних фрагментів трансформованого зображення значно підвищується складність програмно-апаратної реалізації процедури кодування;

– паралельна обробка статистичних кодів досить складно реалізується;

– при обробці сильнокогерентних зображень різко знижується ступінь стиснення, що зумовлено збільшенням значень високочастотних компонентів трансформанти, а також необхідністю використання кодових таблиць і маркерів;

– статистичне кодування не забезпечить додаткового стискання трансформанти при наявності серій нулів невеликих довжин.

Мета статті — є розробка методу кодування, який забезпечуватиме більший ступінь стиснення порівняно з існуючими методами нестатистичного кодування при заданому рівні якості відтворення двійкових послідовностей, з урахуванням можливості його подальшого використання в технології стиснення зображень.

Запропоновано метод кодування, інваріантний до статистичного — *метод інваріантно-просторового кодування (ІПК).*

Завдання дослідження:

1. Сформулювати визначення кореляційної ознаки — сумарної бітової кореляції двійкової послідовності.

2. Встановити основні принципи представлення двійкових даних у кореляційному просторі згідно зі значеннями кореляційної ознаки.

3. Сформувати аналітичну модель методу ІПК.

4. Схарактеризувати структурно-аналітичну модель методу ІПК.

5. Провести порівняння запропонованого методу за ступенем стиснення з існуючими методами нестатистичного кодування.

Принцип інваріантно-просторового кодування

Матрицю колірної компоненти будь-якого зображення розмірністю $L \times C$ одночасно з інтенсивністю пікселів, які її складають, характеризує кореляція пар сусідніх пікселів (вважатимемо, що пікселі сусідні в рядку) — $(x_{i,j}, x_{i,j-1})_{x\in 0..L-1, y\in 0..C-1}$, яка визначається їх різницею $x_{i,j} - x_{i,j-1}$.

Якщо колірна компонента зображення має глибину оцифрування q, то піксель p_{ij} зображення можна представити у вигляді суми степенів двійки:

$$p_{i,j} = 2^0 \cdot (a_0)_{i,j} + 2^1 \cdot (a_1)_{i,j} + \ldots$$
$$\ldots + 2^{q-1} \cdot (a_{q-1})_{i,j} = \sum_{k=0}^{q-1} 2^k \cdot (a_k)_{i,j}.$$

Тоді кореляція між двома сусідніми пікселями:

$$\begin{aligned} x_{i,j} - x_{i,j-1} &= 2^0 \cdot \left((a_0)_{i,j} - (a_0)_{i,j-1}\right) + \\ &+ 2^1 \cdot \left((a_1)_{i,j} - (a_1)_{i,j-1}\right) + \ldots \\ &\ldots + 2^{q-1} \cdot \left((a_{q-1})_{i,j} - (a_{q-1})_{i,j-1}\right) = \\ &= \sum_{k=0}^{q-1} 2^k \cdot \left((a_k)_{i,j} - (a_k)_{i,j-1}\right). \end{aligned} \qquad (1)$$

Тут $(a_k)_{i,j}$ — k-й біт двійкового представлення пікселя $p_{i,j}$, ($k \in 0 \ldots q-1$).

Вираз $(a_k)_{i,j} - (a_k)_{i,j-1}$ визначимо як бітову кореляцію між сусідніми k-ми бітами двійкового представлення сусідніх пікселів.

Оскільки у виразі (1) коефіцієнти 2^k є незмінними величинами, можна вважати, що бітова кореляція $(a_k)_{i,j} - (a_k)_{i,j-1}$ так само однозначно характеризує зображення, як і кореляція між пікселями $x_{i,j} - x_{i,j-1}$.

З'ясуємо просторову організацію зображення (рис. 1). Як правило, матриця зображення інтерпретується представленням у одній з колірних моделей.

Будемо опиратися на одну з них — модель RGB, яка представляє зображення у вигляді трьох компонент — червоного R, зеленого G та блакитного B.

Кожна компонента описується матрицею цілочислових значень, співрозмірною із зображенням (розмірністю $L \times C$).

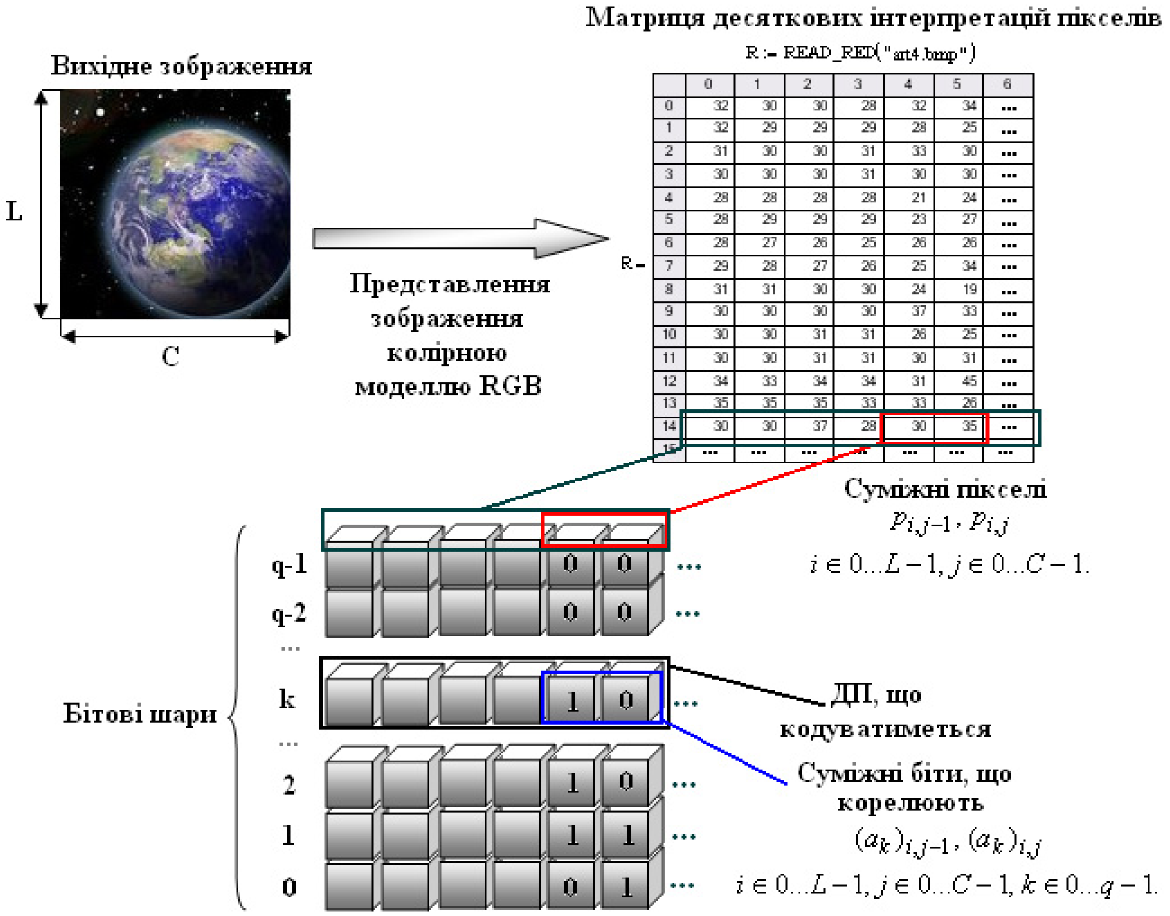

Рис. 1. Схема просторової організації зображення

Діапазон значень при глибині оцифровування q становить від 0 до 2^{q-1}, тобто кожне цілочислове значення може інтерпретуватися двійковою послідовністю завдовжки q. Сукупність бітів $(a_k)_{i,j}$ $\left(x \in 0...L-1, y \in 0...C-1\right)$ й складає k-й бітовий шар.

Двійкові послідовності (ДП) у складі k-го бітового шару складаються з послідовних суміжних бітів $((a_k)_{i,j}, (a_k)_{i,j-1})$.

Визначимо кореляційну ознаку — *сумарну бітову кореляцію* пікселів у ДП у складі бітового шару зображення. Оскільки для визначення принципу кодування немає значення орієнтація суміжних пікселів у бітовому шарі та порядок бітового шару, для спрощення викладення перейдемо від індексації бітів у зображенні до індексації у ДП.

Сумарна бітова кореляція пікселів у ДП $A = \{a_1, a_2, ..., a_n\}$ довжиною n становить:

$$S\left(\{a_i\}_{i=\overline{1,n}}\right) = \sum_{i=1}^{n} | a_{i-1} - a_i |, a_0 = 0.$$

Для спрощення викладення матеріалу позначатимемо $S\left(\{a_i\}_{i=\overline{1,n}}\right)$ через s.

Визначення 1. *Кореляційна група* — множина двійкових послідовностей, які характеризуються однаковим значенням кореляційної ознаки S. Кореляційну групу, яку визначає ознака s, позначатимемо $Y(s)$.

У табл. 1 наведені двійкові послідовності довжиною n в 4 біти. Для кожної з послідовностей вказане значення ознаки s. Зазначимо, що аналіз значень структурних ознак ведеться з урахуванням того, що в двійкову послідовність вводиться додатковий нульовий розряд (у таблиці — підкреслений). Таку послідовність називатимемо *розширеною*.

Таблиця 1

Значення порядкових номерів та кореляційної ознаки для двійкових послідовностей довжиною $n = 4$ біти

Двійкова послідовність	Кореляційна ознака, s	Порядковий номер у структурній групі $Y(s)$
00000	0	0
00001	1	0
00010	2	0
00011	1	0
00100	2	1
00101	3	0

Закінчення табл. 1

Двійкова послідов-ність	Кореляційна ознака, s	Порядковий номер у структурній групі $Y(s)$
00110	2	0
00111	1	0
01000	2	2
01001	3	1
01010	4	0
01011	3	0
01100	2	1
01101	3	1
01110	2	0
01111	1	0

Правило 1. *Про об'єм кореляційної групи*. Для кожної кореляційної групи $Y(s)$ кількість $VOL(s,n)=|Y(s)|$ послідовностей у групі визначається як кількість способів, якими можна розмістити *s* переходів між нулем та одиницею на *n* позиціях у розширеній послідовності довжиною $n+1$ біт, і розраховується як відповідний коефіцієнт бінома Ньютона *n*-го ступеня за формулою:

$$VOL(s,n)=\binom{n}{s}=\frac{n!}{s!(n-s)!}. \qquad (2)$$

Очевидно, що об'єм кореляційних груп буде меншим за об'єм множини двійкових послідовностей довжиною *n*:

$$VOL(s,n)<2^n.$$

Визначення 2. *Інваріантно-просторовим кодуванням (ІК)* називатимемо привласнення послідовності її порядкового номера у кореляційній групі $Y(s)$ — інваріантного коду. Значення інваріантного коду буде меншим за десяткове значення двійкової послідовності. *За рахунок цього й буде забезпечуватися стиснення*.

Сформуємо *аналітичну модель* методу інваріантно-просторового кодування.

Правило 2. Значення коду інваріантно-просторового кодування (ІПК) $NUM\left(\{a_i\}_{i=\overline{1,n}}\right)$ для двійкової послідовності довжиною *n* згідно зі значенням сумарної бітової кореляції s :

$$NUM\left(\{a_i\}_{i=\overline{1,n}}\right)=\sum_{i=0}^{\operatorname{dec}\left(\{a_i\}_{i=\overline{1,n}}\right)}\operatorname{sign}(1-\operatorname{sign}|s- \qquad (3)$$
$$-S(\operatorname{bin}(i,n))|)-1,$$

функція $S()$ — функція визначення сумарної бітової кореляції в розширеній двійковій послідовності:

$$S\left(\{a_i\}_{j=\overline{1,n}}\right)=\sum_{j=1}^{n}|a_{j-1}-a_j|, a_0=0;$$

s — сумарна бітова кореляція у двійковій послідовності $A=\{a_1,a_2,...,a_n\}$, яка чисельно дорівнює $s=S\left(\{a_i\}_{i=\overline{1,n}}\right)$; функція dec() — функція визначення десяткової інтерпретації двійкової послідовності довжиною *n*:

$$\operatorname{dec}\left(\{a_i\}_{i=\overline{1,n}}\right)=\sum_{j=0}^{n-1}2^{n-j-1}a_{j+1}=0;$$

функція $\operatorname{bin}(i,n)$ — функція визначення двійкової інтерпретації довжиною *n* цілого числа i :

$$\operatorname{bin}(i,n)=(a_1,a_2,...,a_n)=$$
$$=\left\{a_j : a_j=2\cdot\left\{\frac{\left[\frac{i}{2^{j-1}}\right]}{2}\right\}, j=\overline{1,n}\right\};$$

тут [] — знаходження цілої частини числа; {} — знаходження залишку від цілочислового ділення; sign() — функція виду:

$$\operatorname{sign}(x)=\begin{cases}-1, x<0,\\ 0, x=0,\\ 1, x>0.\end{cases}$$

Доведення. Згідно з визначенням 2.1 порядковий номер ІПК $NUM\left(\{a_i\}_{i=\overline{1,n}}\right)$ послідовності $A=\{a_1,a_2,...,a_n\}$ визначається кількістю ДП, які належать ті ж самій кореляційній групі, що й дана послідовність, тобто характеризуються тією ж кореляційною ознакою, і є меншими за $A=\{a_1,a_2,...,a_n\}$.

Визначення 3. Будемо вважати, що з двох послідовностей $A=\{a_1,a_2,...,a_n\}$ та $B=\{b_1,b_2,...,b_n\}$ більшою є послідовність $A_j=\{a_1,a_2,...,a_n\}_j$, якщо для неї виконується умова: $\sum_{i=1}^{n}\{a_i\}>\sum_{i=1}^{n}\{b_i\}$. Тобто з двох послідовностей $A=\{0;1;1;0\}$ та $B=\{1;0;0;1\}$ більшою є B.

Принцип підрахунку номера ІПК полягатиме в наступному: почергово аналізуються всі двійкові послідовності, довжиною *n*, які є меншими за A — це обмеження враховується за рахунок визначення верхньої межі для індекса знака суми у формулі (3) як $\operatorname{dec}\left(\{a_i\}_{i=\overline{1,n}}\right)$. Кожна послідовність інтерпретується своїм десятковим значенням *i*.

Для кожної із таких послідовностей підраховується значення кореляційної ознаки — сумарної бітової кореляції $S(\operatorname{bin}(i,n))$.

Це значення порівнюється зі значеннями відповідної кореляційної ознаки послідовності, для якої розраховується номер ІПК.

Збіг між $S\left(\mathrm{bin}(i,n)\right)$ та s фіксується за допомогою фрагменту формули (3) $\mathrm{sign}(1-\mathrm{sign}\,|\,s-S(\mathrm{bin}(i,n))\,|)$. У випадку, якщо кореляційні ознаки збігаються, даний вираз повертає значення 1.

Отже, послідовність i відноситься до тієї ж кореляційної групи, що й A_j. В цьому випадку поточний номер ІК нарощується на 1.

Правило 2 доведене.

На рис. 2 показано *структурно-аналітичну модель* методу інваріантно-просторового кодування.

Як правило, первинні повідомлення — мова, музика, зображення, вимірювання параметрів навколишнього середовища і т. д. — є функціями часу або інших аргументів — неелектричної природи (акустичний тиск, температура, розподіл яскравості на деякій площині тощо).

З метою передачі по каналу зв'язку ці повідомлення зазвичай перетворюються в електричний сигнал, зміни якого в часі $S(t)$, відображають передану інформацію.

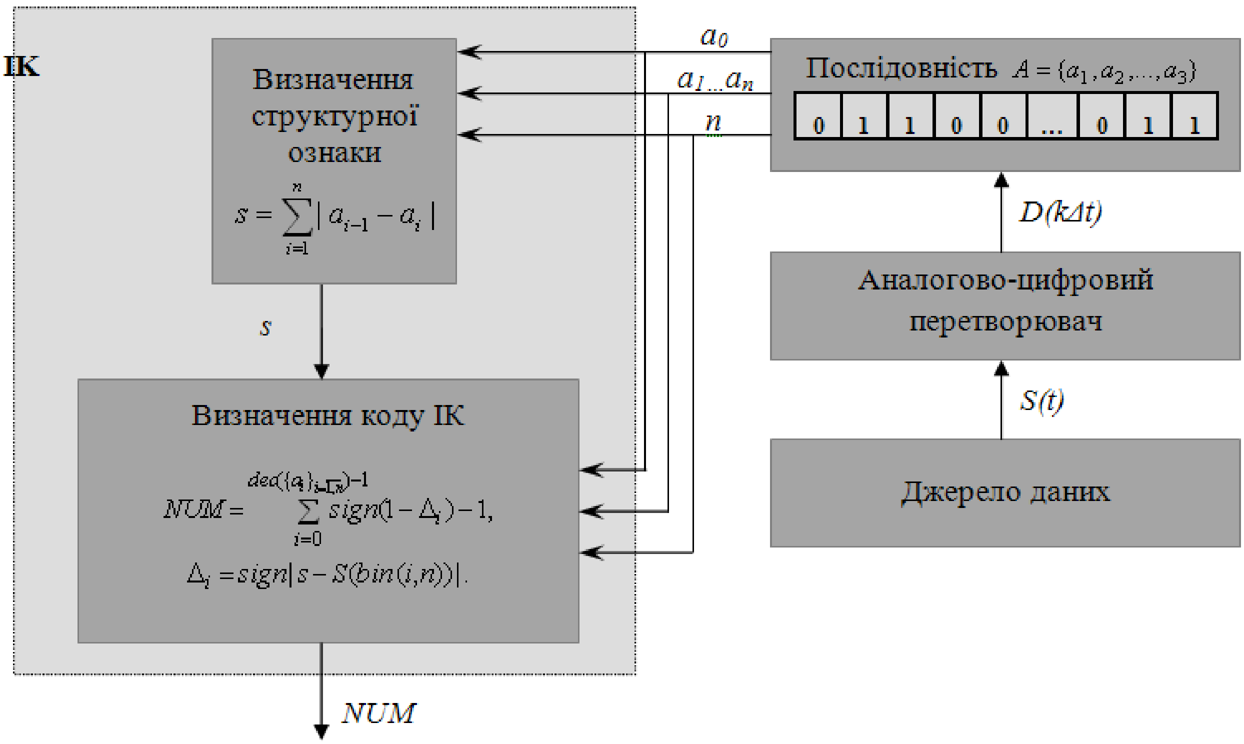

Рис. 2. Структурно-аналітична модель методу інваріантно-просторового кодування

Аналогово-цифровий перетворювач забезпечує перетворення інформаційного сигналу джерела повідомлення $S(t)$ у цифровий сигнал $D(k\Delta t)$ за допомогою дискретизації й квантування сигналу та його представлення у двійковій системі числення.

Порівняльна оцінка ефективності

З-поміж сукупності технологій компактного представлення даних виділяють клас методів стиснення, що ґрунтується на усуненні структурної надмірності шляхом виявлення структурних закономірностей у двійкових послідовностях за деякою ознакою — методи структурного кодування [5–7].

Проведемо оцінювання ефективності процедури стиснення, що забезпечується запропонованим методом кодування.

Оцінювання ефективності інваріантно-просторового кодування здійснюється згідно зі значенням коефіцієнта стиснення k_{ISK}:

$$k_{ISK}=\frac{n}{\left[\log_2 NUM\left(\{a_i\}_{i=\overline{1,n}}\right)\right]+1};$$

де n — довжина вихідної двійкової послідовності A; $NUM\left(\{a_i\}_{i=\overline{1,n}}\right)$ — інваріантний код послідовності A_j, що розраховується відповідно до формули (3), а отже $\left[\log_2 NUM\left(\{a_i\}_{i=\overline{1,n}}\right)\right]+1$ — кількість бітових розрядів, необхідних для подання інваріантного коду у двійковій формі.

Максимальне значення інваріантного коду двійкової послідовності $k_{ISK\_\min}$ є меншим на одиницю за об'єм $VOL(s,e,n)$ структурної групи

$Y(s,e)$, значення якого розраховується згідно з формулою (2) до якої входить дана послідовність $NUM\left(\{a_i\}_{i=\overline{1,n}}\right)_{\max} = VOL(s,n)-1$.

Отже, мінімальне значення коефіцієнта стиснення інваріантного кодування:

$$k_{ISK\_\min} = \frac{n}{\log_2(VOL(s,n)-1)+1}.$$

Проведемо оціновання очікуваних значень мінімальних коефіцієнтів стиснення інваріантно-просторовогокодування для послідовностей різної довжини, що характеризуються різними значеннями структурних ознак.

У табл. 2 наведені значення мінімальних коефіцієнтів стиснення для послідовностей довжиною в 6, 8, 10 та 16 біт.

Виконаємо порівняння запропонованого методу кодування за значенням мінімального коефіцієнта стиснення методами кодування, що відносяться до класів нестатистичних методів — структурних — однознакових та двоознакових структурних кодувань (ОСК та ДСК) та нерівноважним позиційним кодуванням (НРПК) [3; 9].

Таблиця 2

Значення мінімальних коефіцієнтів стиснення $k_{ISK\_\min}$

для двійкових послідовностей довжиною *n* біт

Довжина послідовності, n	Кількість бітових переходів, s	Кількість одиничних елементів, e	Максимальне значення інваріантного коду, $NUM\left(\{a_i\}_{i=\overline{1,n}}\right)_{\max}$	Мінімальне значення коефіцієнта стиснення, $k_{ISK\_\min}$
6	2	1	4	2
		2	3	3
		3	2	3
		4	1	6
	3	2	3	3
		3	5	2
		4	5	2
	4	2	5	2
		3	5	2
		4	2	3
8	2	1	6	2,66
		2	5	2,66
		3	4	2,66
	3	2	5	2
		3	9	2
		4	11	2
		5	11	2,66
	4	2	15	1,6
		3	20	1,6
		4	18	2
		5	12	2
	5	2	10	2
		3	10	2
		4	18	1,6
		5	18	1,6
10	2	1	9	2,5
		2	8	3,33
		3	7	3,33
	4	2	28	2
		4	44	1,67
		5	39	1,67
		6	29	2
	5	3	21	2
		4	44	1,67
		5	59	1,67
		6	59	1,67

Закінчення табл. 2

Довжина послідовності, n	Кількість бітових переходів, s	Кількість одиничних елементів, e	Максимальне значення інваріантного коду, $NUM\left(\{a_i\}_{i=\overline{1,n}}\right)_{\max}$	Мінімальне значення коефіцієнта стиснення, $k_{ISK\_\min}$
10	6	3	35	1,67
		4	59	1,67
		5	59	1,67
		6	15	2,5
16	2	1	14	4
		2	15	4
	7	4	220	2
		7	1200	1,46
		8	1960	1,46
		9	1960	1,46
	8	4	495	1,78
		7	2520	1,33
		8	2450	1,33
		9	190	1,46
	9	5	330	1,78
		7	1890	1,46
		8	2450	1,33
		9	2450	1,33

Оцінювання ефективності ОСК проводиться згідно зі значенням коефіцієнта стиснення k_{OSK} [3]:

$$k_{OSK}=\frac{m}{\left[\log_2 N(m,\Lambda,\vartheta)\right]+1};$$

де m — довжина вихідної двійкової послідовності; $N(m,\Lambda,\vartheta)$ — код-номер ОСК, вирахуваний на підставі значень структурних ознак — вектора Λ, що визначає заборонені на появу одиничного елемента позиції у двійковій послідовності, та загальної кількості серій одиниць ϑ.

Мінімальне значення коефіцієнта стиснення ОСК:

$$k_{OSK\_\min}=\frac{m}{\left[\log_2 V(m,\Lambda,\vartheta)-1\right]+1}=$$

$$=\frac{m}{\left[\log_2\sum_{\xi=1}^{K}\prod_{z=1}^{Z}\frac{(m_z+1)!}{(2\vartheta_z^{(\xi)})!(m_z+1-2\vartheta_z^{(\xi)})!}-1\right]+1}.$$

Тут $V(m,\Lambda,\vartheta)$ — кількість двійкових послідовностей довжиною m з кількістю серій одиниць ϑ, у структурному просторі, заданому вектором обмежень Λ. Оцінювання ефективності ДСК проводиться згідно зі значенням коефіцієнта стиснення k_{DSK} [3]:

$$k_{DSK}=\frac{m}{\left[\log_2 N\left(m,\Lambda,\Theta^{(x)}\right)\right]+1},$$

де m — довжина вихідної двійкової послідовності; $N\left(m,\Lambda,\Theta^{(x)}\right)$ — код-номер ДСК, вирахуваний на підставі значень структурних ознак — вектора Λ, що визначає заборонені на появу одиничного елемента позиції у двійковій послідовності, та загальної кількості серій одиниць у кожній забороненій зоні $\Theta^{(x)}$.

Мінімальне значення коефіцієнта стиснення ДСК:

$$k_{DSK\_\min}=\frac{m}{\left[\log_2 V\left(m,\Lambda,\Theta^{(x)}\right)-1\right]+1}=$$

$$=\frac{m}{\left[\log_2\prod_{z=1}^{Z}\frac{(m_z+1)!}{\left(2\vartheta_z^{(x)}\right)!\left(m_z+1-2\vartheta_z^{(x)}\right)!}-1\right]+1}.$$

Тут $V\left(m,\Lambda,\Theta^{(x)}\right)$ — кількість двійкових послідовностей довжиною m з кількістю серій одиниць у заборонених зонах згідно з вектором обмежень $\Theta^{(x)}$, у структурному просторі, заданому вектором обмежень Λ. Мінімальне значення коефіцієнта стиснення НРПК [9]:

$$k_{NRPK\_\min}=\frac{q}{\sum_{\theta=1}^{\Theta}\left(\left[\log_2 u_\theta\right]+1+\mathrm{sign}(\Delta_\theta)\right)}=$$

$$=\frac{\sum_{\theta=1}^{\Theta}u_\theta}{\sum_{\theta=1}^{\Theta}\left(\left[\log_2 u_\theta\right]+1+sign(\Delta_\theta)\right)}.$$

Тут q — довжина вихідної двійкової послідовності; Θ — структурна ознака, що визначає кількість двійкових серій у послідовності; u_θ — довжина кожної окремої θ-ї двійкової серії; Δ_θ — різниця між величинами u_θ й $2^{[\log_2 u_\theta]}$, функція $\operatorname{sign}(\Delta_\phi) = \begin{cases} 1, \to \Delta_\phi \geq 1; \\ 0, \to \Delta_\phi = 0. \end{cases}$

У табл. 3 наведені значення мінімальних очікуваних коефіцієнтів стиснення для методі ОСК, ДСК та НРПК для двійкових послідовностей різної довжини [3; 9].

Оцінимо виграш, який забезпечує ІК порівняно з методами ОСК, ДСК та НРПК, розрахувавши відношення між мінімальними коефіцієнтами стиснення для послідовностей різної довжини (табл. 4).

На підставі значень, зведених у табл. 3, були побудовані діаграми та графіки, які ілюструють співвідношення значень $k_{OSK\_\min}$, $k_{DSK\_\min}$, $k_{NRPK\_\min}$ та $k_{IK\_\min}$ (рис. 3).

Таблиця 3

Значення мінімальних коефіцієнтів стиснення ОСК, ДСК та НРПК для двійкових послідовностей довжиною *n* біт

Довжина послідовності, n	Максимальне значення кодуномера ОСК, $V(n, \Lambda, \vartheta)$	Мінімальне значення коефіцієнта стиснення ОСК, $k_{OSK\_\min}$	Максимальне значення кодуномера ДСК, $V(n, \Lambda, \Theta^{(x)})$	Мінімальне значення коефіцієнта стиснення ДСК, $k_{DSK\_\min}$	Мінімальне значення коефіцієнта стиснення НРПК, $k_{NRPK\_\min}$
6	19	1,2	18	1,2	-
8	66	1,143	60	1,33	1
10	126	1,429	108	1,429	1
16	6054	1,231	4900	1,231	1

Таблиця 4

Порівняння мінімальних коефіцієнтів стиснення для різних методів структурного кодування для двійкових послідовностей довжиною *n* біт

Довжина послідовності, n	$\frac{k_{IK\_\min}}{k_{OSK\_\min}}$	$\frac{k_{IK\_\min}}{k_{DSK\_\min}}$	$\frac{k_{IK\_\min}}{k_{NRPK\_\min}}$
6	1,67	1,2	–
8	1,4	1,164	1,6
10	1,17	1,17	1,67
16	1,08	1,08	1,33

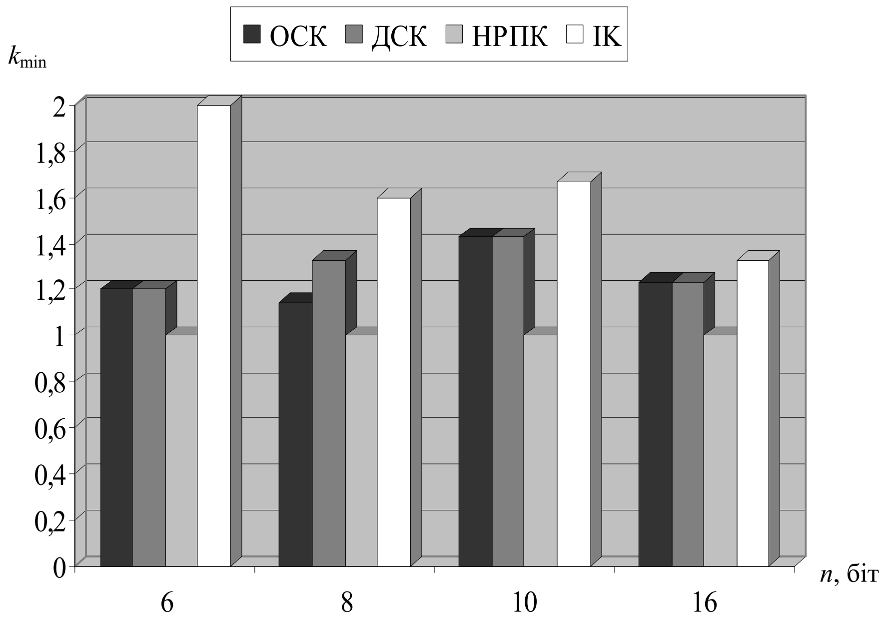

Рис. 3. Діаграми залежності значень мінімальних коефіцієнтів стиснення для методів ДСК, ОСК, НРПК та ІК від довжини двійкових послідовностей *n*

Висновки

У даній статті обґрунтовано доцільність використання структурного кодування в задачах стиснення даних. Визначені структурні ознаки двійкових послідовностей: загальну кількість переходів від одиничного елемента до нульового та від нульового елемента до одиничного s в двійковій послідовності $A = \{a_1, a_2, ..., a_n\}$ довжиною n та загальна кількість одиничних елементів в двійковій послідовності. Сформульоване визначення процедури кодування за визначеними структурними ознаками, згідно якого кодуванням ІК є привласнення двійковій послідовності $A = \{a_1, a_2, ..., a_n\}$ її порядкового номера у структурній групі $Y(s, e, n)$.

Наукова новизна дослідження, описаного в статті полягає в такому:

– уперше описано та доведено правило розрахунку порядкового номера ІК двійкової послідовності в структурній групі за значенням кількості переходів між двійковими елементами послідовності;

– уперше описано метод оцінювання ефективності процедури стиснення, що забезпечується запропонованим методом кодування.

Практична цінність отриманих результатів полягає в такому:

– проведено оцінювання очікуваного мінімального ступеня стиснення для послідовностей різної довжини та проведено порівняння за даним параметром з подібними за класом методами кодування;

– запропонований метод інваріантного кодування забезпечує виграш за значенням мінімального коефіцієнта стиснення:

- порівняно з методом нерівноважного позиційного кодування: у 1,67 разу для послідовностей довжиною в 10 біт, у 1,6 разу для послідовностей довжиною 8 та в 1,33 разу для послідовностей довжиною в 16 біт;
- порівняно з методом ДСК: у 1,64 разу для послідовностей довжиною в 8 біт, у 1,17 разу для послідовностей довжиною 10 біт, у 1,08 для послідовностей у 16 біт, та в 1,2 разу для послідовностей довжиною 6 біт;
- порівняно з методом ОСК: у 1,4 разу для послідовностей довжиною в 8 біт, у 1,17 разу для послідовностей довжиною 10 біт, у 1,08 для послідовностей у 16 біт, та в 1,67 разу для послідовностей довжиною 6 біт;
- запропонований метод інваріантного кодування забезпечує виграш за значенням коефіцієнта стиснення порівняно з методом двоознакового структурного кодування — в межах від 1,26 до 2,33 разів, і для двійкових послідовностей у складі двійкового представлення трансформант дискретного косинусного перетворення коливається в межах від 3,01 до 5.

Отримані результати дозволяють зробити висновок про доцільність подальшого використання запропонованого методу кодування двійкових даних у технологіях стиснення.

ЛІТЕРАТУРА

1. *Урсул А. Д.* Нестатистические подходы в теории информации / А. Д. Урсул // Вопросы кибернетики. — 1967. — № 2. — С. 88–93.

2. *Шеннон К.* Работы по теории информации и кибернетике / К. Шеннон. — М. : Изд-во иностр. лит-ры, 1963. — 793 с.

3. *Юдін. О. К.* Методи структурного кодування даних в автоматизованих системах управління / О. К. Юдін. — К. : НАУ, 2007.

4. *Юдін О. К.* Кодування в інформаційно-комунікаційних мережах: монографія. — К. : НАУ, 2007. — 308 с.

5. *Баранник В. В.* Двупризнаковое структурное кодирование массивов двоичных данных / В. В. Баранник, А. К. Юдин // Автоматизированные системы управления и приборы автоматики. — Х. : ХНУРЭ, 2005. — № 133. — С. 64–72.

6. *Юдін О. К.* Обґрунтування ефективності двоознакового структурного кодування у двійковому поліадичному просторі / О. К. Юдін // Проблеми інформатизації та управління: збірник наукових праць. — К. :НАУ, 2006. — Вип. 2(17). — С. 137–141.

7. *Юдін О. К.* Метод кодування двійкових послідовностей за кількістю бітових переходів / К. О. Курінь, М. Г. Луцький, О. К. Юдін // Наукоємні технології. — К.: Вид-во Нац. авіац. ун-ту «НАУ-друк», 2012. — № 4 (16). — С. 98-102.

8. *Юдин А.* Методы и алгоритмы эффективного сжатия видеоданных на базе стандарта JPEG 2000 / А. К. Юдин, Д. А. Пуха // Защита информации: сборник научных трудов. — К.: НАУ, 2006. — Вып. № 13. — С. 209–214.

9. *Гулак Н. К.* Методи підвищення ступеня стиску відеоданих в інформаційних системах кодування зображень: дисертація на здобуття наукового ступеня кандидата технічних наук: 05.13.06/ МОН; Національний авіаційний університет / Н. К. Гулак. — Київ, 2011. — 191 с.

Стаття надійшла до редакції 14.10.2014.