

## АНАЛОГІЇ МІЖ ЗОБРАЖЕННЯМИ ТА ТЕКСТАМИ: ЯВИЩЕ «СПАЛАХІВ» У ТЕКСТАХ І ЦИФРОВИХ ЗОБРАЖЕННЯХ

І. Я. Довгань, О. С. Кушнір, Ю. М. Фургала, Г. І. Паночко

*Львівський національний університет імені Івана Франка,  
вул. Ген. Тарнавського, 107, 79017 м. Львів, Україна  
[mozmarto@gmail.com](mailto:mozmarto@gmail.com), [oleh.kushnir@lnu.edu.ua](mailto:oleh.kushnir@lnu.edu.ua)*

Проаналізовано низку способів перетворення двовимірних зображень на одновимірні лінійні ланцюжки пікселів і розроблено аналогії між текстами та цифровими зображеннями, де відповідником лінгвістичних елементів (наприклад, слів) є рівні яскравості пікселів. З'ясовано можливості застосування методів статистичної лінгвістики для пошуку ключових слів у текстах до цифрових зображень. Проаналізовано явище спалахів у текстах, які є відповідниками низки зображень. На цій основі запропоновано метод розрізнення інформативних і суто шумових зображень.

*Ключові слова:* статистична лінгвістика, обробка природної мови, спалахи, ключові слова, опрацювання зображень.

**Вступ.** В останні десятиліття успішно розвиваються галузі цифрового опрацювання зображень і розпізнавання образів, з одного боку, та комп'ютерної лінгвістики та опрацювання природної мови, з іншого [1, 2]. Переважно цей розвиток відбувається незалежно та із застосуванням загалом різних ідей, підходів і методів, а приклади запозичень напрацювань однієї з галузей іншою все ще рідкісні (див. [3, 4]). Метою цієї праці є розвиток аналогій між текстами та зображеннями, а також нових простих підходів до оцінювання інформаційної та шумової складових зображень на основі запозичень методів статистичної лінгвістики.

**Аналогія між текстами та зображеннями.** Базова ідея такої аналогії проста: текстові як послідовності деяких лінгвістичних елементів (літер, знаків, слів тощо) можна поставити у відповідність цифрове зображення як послідовність пікселів, кожен з яких описується деяким значенням яскравості – її квантованим рівнем. У разі цифрових зображень обидві системи дискретні, а тому загалом допускають таку відповідність.

Перша проблема побудови такої аналогії – той факт, що тексти – це одновимірні, а зображення – двовимірні структури. Формально цей фактор можна усунути так:

(1) розглядати лише окремі рядки (або стовпці) цифрових зображень;

(2) представити все зображення як одновимірну структуру, яка є лінійною послідовністю рядків (або стовпців); при цьому послідовності пікселів у всіх рядках (стовпцях) можна відраховувати зліва направо (зверху донизу) – або, як альтернатива, в непарних рядках (стовпцях) відраховувати пікселі зліва направо (зверху донизу), а в парних – навпаки, справа наліво (знизу догори).

(3) розглядати будь-яку іншу «лінійну» послідовність пікселів у зображенні, наприклад спіраль, яка виходить із центру (або з периферії) зображення та послідовно проходить крізь усі пікселі.

Спосіб перетворення зображення (1) до тексту потерпає від збіднення статистики пікселів (оскільки  $M, N \ll MN$ , де  $M$  і  $N$  – відповідно кількість рядків і стовпців у зображенні). Це можна частково виправити усередненням даних подальшого статистичного аналізу по всіх стовпцях (або рядках). Водночас, деяким недоліком способів (2) і (3) є поява «граничних ефектів» (або «розривів»), коли окремі пікселі, які формально «сусідні» в «тексті» (тобто в «зображенні», розтягнутому в одновимірну структуру), насправді не є сусідами у вихідному зображенні. Навпаки, окремі пікселі, які є «сусідніми» у вихідному зображенні, можуть виявитися вже не такими в кінцевому лінійному ланцюжку («тексті»). Цей фактор означає втрати «зв'язності» деяких елементів вихідного зображення, які неминучі при зведенні двовимірної структури до одновимірної.

Надалі, як приклад, ми оберемо спосіб (2) перетворення зображення на одновимірну структуру (текст). Тоді і звичайний текст, і розгорнуте в лінійний ланцюжок зображення фактично є часовими рядами – з можливістю запозичення відповідного апарату опису. Роль дискретного часу  $t_i$  тут відіграє позиція лінгвістичного елемента в тексті ( $t_i = 1, \dots, t_L$ , де  $t_L$  – довжина тексту) і номер пікселя в зображенні, який відраховують спочатку в першому рядку справа наліво, а потім по наступних рядках зверху донизу ( $t_i = 1, \dots, t_N, t_{N+1}, \dots, t_{2N}, \dots, t_{(M-1)N+1}, \dots, t_{MN}$ , де  $t_L = t_{MN} = MN$  – «довжина» зображення), або по стовпцях, аналогічно до попереднього способу.

Наступне питання – які саме лінгвістичні елементи (літери, знаки, слова тощо) вважати аналогами рівнів пікселів – нетривіальне. Адже слова в тексті, на відміну від літер або знаків, є найменшими лінгвістичними одиницями, які мають семантичне (змістове) навантаження. Проте наперед невідомо, чи коректно приписувати деяку «семантику» окремим рівням яскравості пікселів у зображенні. Проста логіка швидше схиляє до негативного висновку: «семантика», мабуть, притаманна, як мінімум, певним групам пікселів, які відповідають деякому змістовному кластеру на зображенні (обличчю, предметові тощо). Більше того, інтуїтивно зрозуміло, що «семантику» зображення бажано формулювати в термінах, які не прив'язані до способу його конкретного поділу на пікселі.

Нижче для визначеності ми, наразі умовно, вважатимемо рівні яскравості пікселів аналогами слів у тексті. Іншими словами, текстові як символічному часовому рядові, що складається зі слів (точніше, слововживань)  $w_i$  ( $t_i = 1, \dots, t_L$ ), ми ставимо у відповідність зображення як числовий часовий ряд рівнів яскравості пікселів  $W_i$  ( $t_i = 1, \dots, t_{MN}$ ).

Забігаючи наперед (див. обговорення результатів), зазначимо, що особливості просторового розміщення окремих рівнів яскравості по пікселях таки схиляють до висновку про те, що їхня аналогія із семантично наповненими словами небезпідставна. Нарешті, символічний часовий ряд, пов'язаний із текстом, можна перетворити на числовий ряд за допомогою низки стандартних алгоритмів (див., зокрема, описані нижче процедури). Стосовно ж зображення, розгорнутого в одновимірну структуру, воно одразу є числовим рядом. Правда, це не виключає застосування інших підходів до його перетворення на числові ряди, запозичених із аналогії з текстами. Наприклад, всі появи

деякого заданого рівня яскравості (скажімо,  $W_i = 255$ ) у лінійному ланцюжку можна перетворити на одиниці, а всі інші рівні яскравості ( $W_j \neq W_i$ ) у цьому ланцюжку – на нулі.

Аналогом абсолютної частоти  $F_j$  деякого  $j$ -го слова (кількості його випадів у тексті, тобто кількості слововживань) є кількість пікселів  $F_j$ , які мають деякий рівень яскравості в зображенні. Відповідно, аналогом словника  $V$  (кількості різних слів у тексті,  $j = 1, \dots, V$ ) є повна кількість різних рівнів яскравості в зображенні. Хоча словник тексту формально обмежений лише його розміром (довжиною)  $t_L$ , на практиці типово маємо  $V \ll t_L$ . На відміну від текстів, «словник» рівнів яскравості пікселів у зображенні обмежений найменшим із чисел, які задають «довжину» одновимірного «зображення»  $MN$  і його бітність  $\Gamma = 2^n$ , тобто  $V \leq \min(\Gamma, MN)$ . Наприклад, у 8-бітному зображенні ( $n = 8$ ) із  $\Gamma < MN$  цей «словник» ( $j = 1, \dots, V$ ) дорівнює  $\Gamma = 256$  або менший (якщо деякі з рівнів яскравості відсутні в зображенні, а інші повторюються в різних пікселях).

#### **Застосування методів статистичної лінгвістики для аналізу зображень.**

Розвинуті аналогії між текстами та зображеннями дають змогу застосувати до останніх підходи статистичної лінгвістики (див. [5–7]), пов'язані з ранговими залежностями  $F(r)$ , звичайним і кумулятивним розподілами ймовірності  $p(F)$  і  $P(F)$ , а також законами зростання словника зі зростанням довжини ( $V(t_L)$ ) або бітності ( $V(n)$ ) зображення. Надалі ми не розглядатимемо ці характеристики, а обмежимося лише аналізом властивостей зображень, пов'язаних із просторовим розподілом окремих рівнів яскравості. Відповідні властивості для текстів дають змогу знаходити ключові слова [8–11]. У свою чергу, словник  $V_{kw}$  ключових слів дає змогу представити текст у вигляді вектора образу його змісту. Усунувши більшість слів, які не мають стосунку до змісту тексту, можна понизити розмірність задачі аналізу цього тексту від  $V$  до  $V_{kw} \ll V$ .

На початку стисло нагадаємо відповідні результати зі статистичної лінгвістики (див., наприклад, [9, 12, 13]). Сформуємо на основі тексту часовий ряд, який містить позиції  $t_k$  слововживань деякого слова  $w_j$  зі словника:  $t_k = \{t_1, t_2, \dots, t_F\}$ , де  $F_j$  – частота цього слова (див. рис. 1). Для оцінювання часового розподілу цього слова в тексті на практиці зручно аналізувати не абсолютні позиції  $t_k$  слововживань, а т. зв. часи очікування слова  $\tau_k = t_{k+1} - t_k - 1$ . Дещо в інших («просторових») термінах,  $\tau_k$  – це «відстань» між двома сусідніми слововживаннями даного слова  $w_j$ , виражена в одиницях кількості інших слів  $w_r$  ( $r \neq j$ ).

Загалом найповнішу інформацію про розподіл слововживань даного слова в тексті дає відповідна масова функція розподілу ймовірності  $p(\tau_k)$ . Проте головні риси згаданого розподілу можна збагнути навіть за параметрами, пов'язаними з кількома першими статистичними моментами розподілу  $p(\tau_k)$ . Найпростіший з таких параметрів – це параметр «спалахів» (або «кластеризації»)  $R = \Delta\tau/\tau$ , де  $\tau$  і  $\Delta\tau$  – відповідно середнє значення та середньоквадратичне відхилення часів очікування  $\tau_k$  [9]. У розрахунках часів очікування використаємо т. зв. періодичні граничні умови ( $k = 1, \dots, F$

– див., наприклад, [14]), а для ігнорування вбогої статистики не розглядатимемо слів із частотами, нижчими за деяку мінімальну частоту (наприклад,  $F_{\min} = 5$ ).

Відомо, що розподіл слововживань функціональних слів (т. зв. «стоп-слів» на зразок «і», «не», «або»), які найчисельніші в текстах і несуть мінімальне семантичне навантаження, описується нульовою стохастичною гіпотезою. На рис. 1а наведено приклад для слова «все» в українському перекладі повісті Дж. Р. Р. Толкіна «Гобіт» за авторством О. Мокровольського. Статистику появ стоп-слів у тексті як рідкісних (екстремальних) і незалежних подій визначає розподіл імовірності Пуасона, а їхні часи очікування наближено описуються негативним експоненційним розподілом  $p(\tau) = \tau^{-1} \exp(-\tau/\tau)$  (із  $\tau = \text{const}$ ), для якого  $\Delta\tau = \tau$  [15]. Тому для таких слів маємо  $R \approx 1$ .

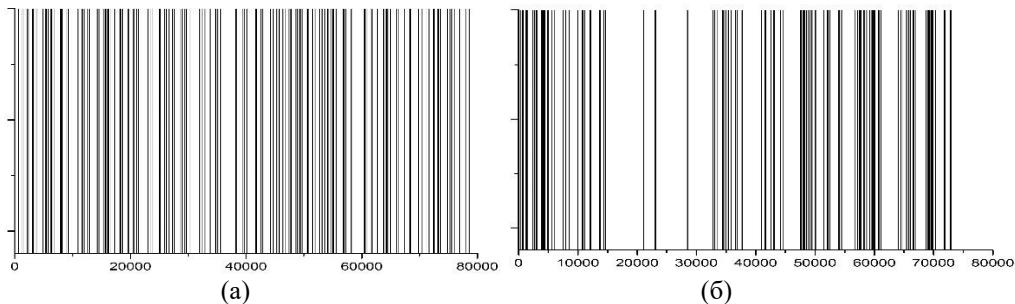


Рис. 1. Позиції слововживань (а) функціонального слова «все» (абсолютна частота появи  $F = 200$ , параметр спалахів  $R = 0,98$ ) і (б) змістового слова «Торін» ( $F = 187$ ,  $R = 2,18$ ) у тексті повісті «Гобіт» (довжина тексту  $t_L \approx 79,9$  тис. слів) [13].

Змістові слова, які мають істотніше семантичне навантаження, виявляють менш «однорідний» розподіл слововживань у тексті та помітні відхилення від нульової статистичної гіпотези. Тому для них маємо вище відхилення  $\Delta\tau$  за умови того ж середнього значення  $\tau$  і, відповідно, більший параметр  $R$  ( $R > 1$  – див. дані для слова «Торін» на рис. 1б). Нарешті, ключові слова тексту, які стисло охоплюють його семантичне поле, розподілені ще «неоднорідніше» ( $R > 1$  або навіть  $R \gg 1$ ). Їм притаманне виражене явище «спалахів» [12], яке означає концентрацію слововживань на одних ділянках і їхню відсутність на інших ділянках тексту (див. рис. 1б). Зазначимо, що гіпотетичний випадок  $R \rightarrow 0$  (тобто  $\Delta\tau \rightarrow 0$ ) відповідав би повній «однорідності» (точніше, періодичності) слововживань деякого слова в тексті. Проте навіть «квазіперіодичні» слова із  $R < 1$  у природних текстах украй рідкісні (наприклад, слово «розділ», якщо обсяги розділів книги приблизно однакові), а строго «періодичні» випадки  $R \rightarrow 0$  завжди відсутні.

Вивчимо тепер можливості параметра спалахів  $R$  в опрацьованні зображень.

**Методика досліджень.** Для аналізу було обрано два кольорові зображення:

(А) одне з них відповідало білому шумові та було генероване відповідним інструментом в Adobe Photoshop CC 2017 (розміри  $663 \times 626$  – див. рис. 2а);

(Б) інше зображення – це кадр із мультиплікаційного фільму «Tom and Jerry» з плоскими кольорами (розміри  $500 \times 500$  – див. рис. 2б).

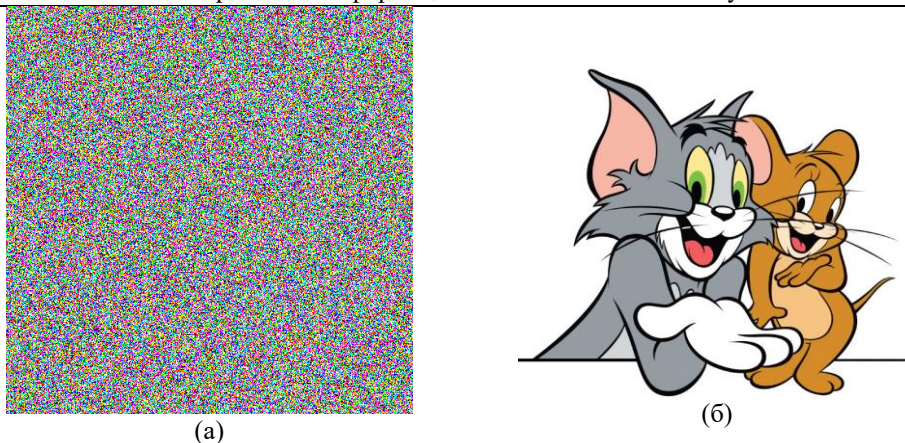


Рис. 2. Зображення, які відповідають білому шумові (а) та кадрові з мультфільму (б).

З метою оцінки впливу бітності зображення на одержані дані було проаналізовано два різновиди зображень (А) і (Б), квантовані на 16 і 256 рівнів. Згадані RGB-зображення перетворювали до зображень у відтінках сірого, а яскравості пікселів  $Y$  (тобто рівні відтінку сірого) визначали за однією зі стандартних схем під назвою «sRGB»:  $Y = 0,2126R + 0,7152G + 0,0722B$  (див. сторінку <https://www.w3.org/Graphics/Color/sRGB>).

Було розроблено програму мовою C# для перетворення цифрових зображень на текстові файли. Перетворення на текст відбувалося шляхом послідовного компонування горизонтальних рядків зображення. Наведемо як приклад початок «тексту», який відповідає 256-рівневному зображенню на рис. 2а:

0 134 182 184 72 251 195 182 193 236 182 236 72 200 22 72 ...

Тут перші позиції «слова» «72» дорівнюють  $t_k = 5, 13, 16, \dots$ , а відповідні перші часи очікування такі:  $\tau_k = 7, 2, \dots$ .

Для аналізу синтетичних «текстів», які є відповідниками зображень, було застосовано програму, написану мовою C#. Вона призначена для розрахунку часових позицій слововживань і часів очікування, а також визначення параметра  $R$  для усіх слів у тексті.

**Емпіричні результати та аналіз.** На рис. 3 показано приклади розподілів кількох слів на «довжині» обох досліджених зображень із 256 рівнями яскравості. Тут для слова «149» у тексті, що відповідає зображенню (А), фактично маємо ідеальний випадок стохастичного розподілу ( $R = 1$ ), слово «126» у тексті (А) і слово «167» у тексті (Б) виявляють слабе явище «спалахів» ( $R > 1$ ), слово «178» у тексті (А) – слабку квазі-періодичність ( $R < 1$ ), а слово «184» у тексті (Б) – яскраво виражені «спалахи» ( $R \gg 1$ ).

У таблицях 1 і 2 представлено ранги  $r$  (порядкові номери рівнів яскравості в їхньому списку за спадною частотою), а також відповідні частоти  $F$  і параметри  $R$  рівнів яскравості, які одержані для білого шуму (рис. 2а) відповідно у випадках 16- і 256-рівневих зображень. Тут для всіх рівнів яскравості значення  $R$  близькі до одиниці. Ба більше, середнє  $R_0$  і середньоквадратичне відхилення  $\Delta R_0$  (незаокруглених) значень  $R$  для всіх рівнів яскравості складають  $1,02 \pm 0,04$  і  $0,99 \pm 0,07$  відповідно для 4- і 8-бітних зображень. Це відповідає ситуації, коли фактично всі рівні яскравості (слова) розподілені по зображенню стохастично (за законом Пуасона), а їхні часи очікування  $\tau_k$

описуються негативним експоненційним розподілом, так що для них відсутнє явище спалахів.

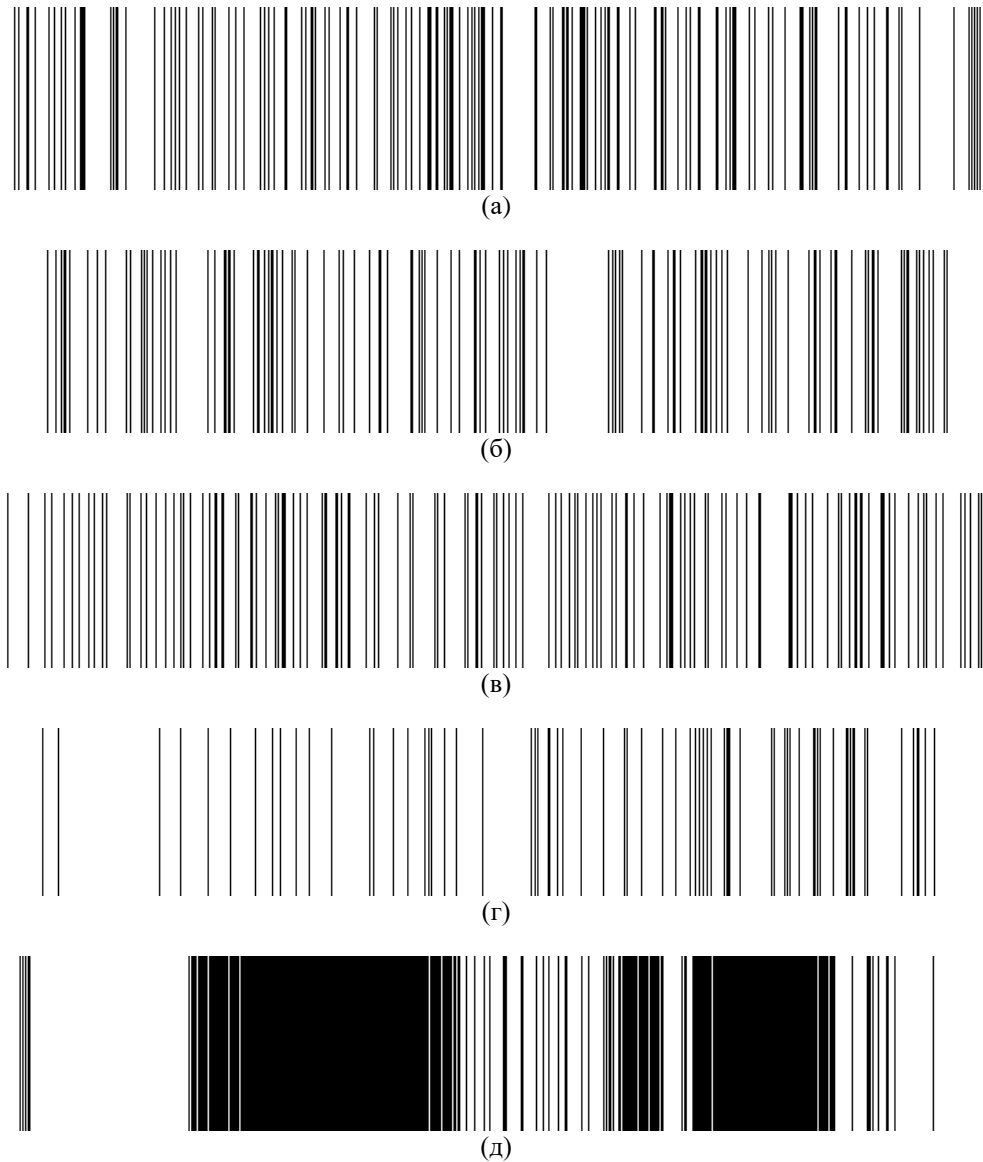


Рис. 3. Розподіли слововживань для «слів» (а) «149» (частота  $F = 163$ , параметр спалахів  $R = 1,00$ ), (б) «126» ( $F = 131$ ,  $R = 1,41$ ), (в) «178» ( $F = 140$ ,  $R = 0,78$ ), (г) «167» ( $F = 75$ ,  $R = 1,20$ ) і (д) «184» ( $F = 5817$ ,  $R = 14,67$ ) по «довжині» 8-бітних зображень, які відповідають білому шумові (а, б, в) і кадрові з мультфільму (г, д).

Таблиця 1

Деякі статистичні параметри для зображення, що відповідає білому шумові, з 16 рівнями яскравості (сортування слів за спаданням  $R$ )

Ранг $r$	Слово (рівень яскравості)	Частота $F$	Параметр $R$
1	15	41138	1,11
2	12	31556	1,08
3	14	30091	1,08
5	3	24786	1,07
4	4	29636	1,07
6	11	24104	1,06
7	1	22121	1,05
8	0	18776	1,05
12	7	2868	1,03
9	13	6497	1,02
15	5	2714	1,02
11	6	2871	1,02
14	10	2797	1,01
16	9	2678	1,01
10	2	4544	1,01
13	8	2823	0,99

Таблиця 2

Деякі статистичні параметри для зображення, що відповідає білому шумові, з 256 рівнями яскравості (сортування слів за спаданням  $R$ ; представлено слова із першими 8 найбільшими та 7 найменшими значеннями  $R$ ): жирний шрифт відповідає словам, дані для яких наведено на рис. 3

Ранг $r$	Слово (рівень яскравості)	Частота $F$	Параметр $R$
236	<b>126</b>	131	1,41
77	229	414	1,16
238	136	126	1,15
211	101	160	1,15
146	131	250	1,14
173	154	194	1,14
228	127	144	1,13
138	87	264	1,13
...	...	...	...
192	156	178	0,87
249	158	112	0,87
133	39	273	0,86
79	56	412	0,85
189	91	180	0,85
244	125	115	0,84
232	<b>178</b>	140	0,77

У таблицях 3 і 4 наведено ті ж результати для зображення з мультфільму (рис. 2б). Тут спостерігаємо винятково високий рівень кластеризації більшості слів (рівнів яскравості) по «довжині» зображення, а тому параметр  $R$  для них більший або й істотно більший за одиницю. Зокрема, 11 рівнів яскравості у 8-бітному зображенні описуються величинами  $R > 10$  і ~ 150 рівнів яскравості – величинами  $R > 2$ . Лише 20 рівнів яскравості в ньому мають параметри  $1 > R > 1,5$ , що відповідає слабким спалахам, і немає жодного рівня із  $R < 1$  (див. Таблицю 5). Що цікаво, максимальні величини  $R$  є своєрідним інваріантом, який майже не залежить від бітності зображення (порівн. дані Таблиць 3 і 4). Порівнюючи дані для суто шумового зображення рис. 2а та інформативного зображення рис. 2б, бачимо, що значення  $R_{\max}$  дає кількісну оцінку вмісту інформації в зображенні. Його перевагою є порівняно невеликий обсяг потрібних обчислень, який істотно нижчий за складність, скажімо, фур'є-перетворення.

Таблиця 3  
Деякі статистичні параметри для зображення з мультфільму  
з 16 рівнями яскравості (сортування слів за спаданням  $R$ )

Ранг $r$	Слово (рівень яскравості)	Частота $F$	Параметр $R$
4	7	32491	25,67
2	9	52104	25,50
5	11	20959	20,58
1	15	247434	17,84
6	13	8584	13,18
7	6	3264	8,02
8	8	2962	7,13
10	14	2087	6,57
9	1	2481	6,51
11	2	2084	5,97
12	3	1980	5,89
15	5	1643	5,82
14	10	1793	5,48
13	4	1900	4,78
3	0	32502	3,99
16	12	770	3,81

Показово, що й усереднений за всіма словами (рівнями яскравості) параметр  $R_0$ , і його відхилення від середнього  $\Delta R_0$  різко відрізняються від відповідних значень, одержаних для шумового зображення. Так, для 16- і 256-рівневих варіантів інформативного зображення маємо відповідно  $10,42 \pm 7,64$  і  $3,30 \pm 3,31$ . Зазначимо, що причини значної відмінності цих параметрів загалом зрозумілі: через істотно вбогіший «словник» рівнів яскравості в 4-бітному зображенні в ньому частіше зустрічаємо «виродження» яскравості різних пікселів, а тому й більше кластерів пікселів із однаковими яскравостями. Отже, крім максимальної величини  $R_{\max}$ , статистичні параметри  $R_0$  і  $\Delta R_0$  можна теж пропонувати як кількісні мірила інформаційної ємності («семантичного наповнення») зображення. Іншими словами, ці параметри дають змогу розрізнити шумові та інформативні зображення – або й навіть сигналізувати про співвідношення сигнал/шум у зображенні.



Таблиця 4

Деякі статистичні параметри для зображення з мультфільму з 256 рівнями яскравості (сортування слів за спаданням  $R$ ; представлено слова із першими 8 найбільшими та 7 найменшими значеннями  $R$ ): жирний шрифт відповідає словам, дані для яких наведено на рис. 3

Ранг $r$	Слово (рівень яскравості)	Частота $F$	Параметр $R$
2	156	21254	24,44
3	155	19788	23,04
5	183	9849	18,30
4	124	11088	18,15
1	254	238942	16,96
6	123	8761	15,54
8	<b>184</b>	5817	14,67
18	214	2182	13,52
...	...	...	...
23	0	1608	1,27
249	231	37	1,267
245	195	49	1,21
226	<b>167</b>	75	1,20
238	243	57	1,19
250	239	36	1,16
253	196	30	1,16

Таблиця 5

Деякі статистичні параметри для природного тексту «Moby Dick» за авторством Н. Melville та інформативного 256-рівневого зображення з мультфільму (рис. 2б): окрім параметрів  $R_0$  і  $R_0$ , усі дані (у т. ч. відсоткові) стосуються слів (для тексту) та рівнів яскравості (для зображення)

Параметр тексту (зображення)	Текст	Зображення
Довжина $t_L$	211,8 тис.	415,0 тис.
Словник $V$	1907	256
Слова з параметрами $R > 10$	0%	4,3%
Слова з параметрами $R > 2$	6,2%	58,6%
Слова з параметрами $1 > R > 1,5$	55,2%	7,8%
Параметр $R < 1$	24,7%	0%
Параметр $R_0$	1,29	3,30
Параметр $\Delta R_0$	0,53	3,31

Щоправда, деяким недоліком параметрів  $R_0$  і  $\Delta R_0$  є помітна залежність від способу квантування яскравості. З іншого боку, підхід, заснований на параметрі  $R$ , фактично захищений від такого впливу. Наведемо єдину ілюстрацію останнього твердження: перший за частотою та рангом (і найвищий) рівень яскравості «15» у 4-бітному зображенні (Таблиця 3) і його відповідник у 8-бітному зображенні (найвищий рівень яскравості «255» – Таблиця 4) мають доволі близькі параметри спалахів (відповідно 17,84 і 16,96). Їхня незначна (5%) відмінність пояснюється впливом обсягу статистичної вибірки.

Корисно порівняти основні статистичні параметри текстів і зображень.

1. Масштаби явища спалахів в інформативному зображенні значно перевищують відповідні масштаби в природному (інформативному) тексті. Це видно із порівняння середніх значень  $R_0$  або відсотків слів чи рівнів яскравості, які мають великі ( $\gg 1$ ) і малі

( $\sim 1$ ) значення  $R$ . Тоді як природним текстам притаманні типові величини  $R_0 \sim 1,2-1,6$  і  $\Delta R_0 \sim 0,5-1,0$  (див. також дані Таблиці 5 для англomовного тексту «Moby Dick»), відповідні величини для інформативних зображень помітно більші. Щоправда, гігантські масштаби спалахів у зображенні рис. 2б частково є наслідком його схематичного характеру – і можна очікувати деякого послаблення цього явища та зменшення параметрів  $R_{\max}$ ,  $R_0$  і  $\Delta R_0$  для більш реалістичних («непласких») зображень.

2. На окреме обговорення заслуговує такий момент. Вище згадувалося про проблеми порушення статусу «сусідства» пікселів зображення при його перетворенні на одновимірний текст. Одним із наслідків може бути оманлива періодичність у результуючому тексті. Наприклад, якщо зображення містить вертикальну чорну смужку на білому фоні, а перетворення на текст відбувається по рядках, то низькі яскравості («чорні» пікселі) у відповідному тексті будуть періодично повторюватися з періодом, який дорівнює довжині рядка  $M$ , хоча реальний відповідник такої періодичності у вихідному зображенні відсутній. Такий артефакт тексту міг би призвести до помітного зменшення результуючого параметра  $R$  для слів (рівнів яскравості), які відповідають згаданім низьким рівням яскравості. Проте така інтуїтивно передбачувана паразитна періодичність, мабуть, притаманна лише простим модельним зображенням на зразок описаного вище. На практиці же ми не помітили цього явища на всіх проаналізованих нами «природних» зображеннях.

3. У природних текстах слова з найвищими частотами ( $F \sim F_{\max}$ ) типово мають параметри  $R \sim 1$ , тобто вони суто функціональні, і лише слова з помірно високими частотами можуть бути ключовими ( $R \gg 1$ ). В інформативних зображеннях така закономірність необов'язкова. Прикладом слугують уже згадані слова «15» і «255» із найвищими частотами відповідно в 4- і 8-бітних варіантах інформативного зображення. Очевидно, що вони відповідають світлому фонові, площа якого найбільша.

4. Рівні яскравості з найбільшими  $R$  можна розглядати як «ключові слова» зображення, які у стислій формі відображають основну інформацію, що міститься в цьому зображенні. Для перевірки цієї гіпотези можна скомпонувати «урізану» версію зображення, яка містить лише ті пікселі, яскравості яких відповідають кільком найвищим  $R$ -параметрам, і порівняти його з вихідним зображенням. Нарешті, той принциповий факт, що окремі рівні яскравості в зображеннях можуть виявляти явище спалахів і великі параметри  $R$ , вказує, що аналогія рівнів яскравості саме зі словами в тексті не безпідставна – адже «семантично порожнім» літерам у тексті спалахи не притаманні (див. [16]).

5. Білий шум є хорошим аналогом рандомних (або рандомізованих) текстів (див. [17]). Так, після  $10^9$  циклів рандомізації тексту «Moby Dick» за словами одержуємо текст, для якого  $R \approx 0,96 \pm 0,18$ , тобто явища спалахів і квазіперіодичності слів тут майже відсутні. Трапляються вони лише для слів із найнижчими частотами, статистика яких найменш надійна. Це якісно схоже до даних для шумового зображення (див. вище).

6. У галузі опрацювання зображень їхнє дозоване зашумлення здійснюють шляхом додавання до яскравостей окремих пікселів випадкової шумової компоненти з деякими середнім значенням і розподілом імовірності (наприклад, нормальним). З іншого боку, аналогії з текстами пропонують іншу схему зашумлення, яка передбачає випадкову просторову перестановку пікселів за умови незмінності їхніх яскравостей. Таку альтернативну рандомізацію зображення можна дозувати кількістю циклів перестановок, спостерігаючи за якісними (видимі зміни інформативності) та кількісними

(параметри  $R_{\max}$ ,  $R_0$  і  $\Delta R_0$ ) змінами зображення. Можна передбачити, що остаточний результат такого зашумлення приведе до границі  $R_0 \rightarrow 1$  і одночасного зменшення  $\Delta R_0$ .

Наостанок зазначимо, що ми здійснили також перші спроби порівняння кількісних характеристик автокореляцій окремих елементів у текстах і зображеннях. Для цього на основі тексту будують числовий ряд, у якому на позиціях, де трапляється наперед задане слово  $w_i$ , ставлять одиниці, а на всіх інших позиціях – нулі. Далі вивчають залежність середньоквадратичної флуктуації від ширини ковзного вікна, на якому підраховують кількість одиниць. Це відповідає методиці флуктуаційного аналізу для окремих слів (див., наприклад, [18–20]), вихідними даними якої є степеневий параметр флуктуацій  $\alpha$ . Надалі ми усереднювали параметри  $\alpha$  для всіх слів, які пройшли стандартний частотний фільтр  $F_{\min}$ . Нагадаємо, що значення  $\alpha = 1/2$  означає відсутність довгосяжних кореляцій у часовому ряді, а його перевищення ( $1/2 < \alpha < 1$ ) сигналізує про наявність таких кореляцій.

Наведемо приклад. Рандомізований текст «Moby Dick» описувався усередненим параметром  $\alpha = 0,49 \pm 0,03$ , а вихідний природний (тобто інформативний) текст – параметром  $\alpha = 0,58 \pm 0,06$ . Знайдені нами  $\alpha$ -параметри для 8-бітних зображень рис. 2а,б склали відповідно  $\alpha = 0,48 \pm 0,05$  і  $\alpha = 0,73 \pm 0,12$ . Отже, для розгорнутих у лінійний ланцюжок шумових зображень, як і для рандомних текстів, наближено одержуємо теоретичне значення  $1/2$  і відносно малі середньоквадратичні відхилення. Тобто такі стохастичні структури очевидно некорельовані. Водночас, тексти і зображення, які переносять корисну інформацію, виявляють довгосяжні кореляції ( $\alpha > 1/2$ ), а відповідні середньоквадратичні відхилення дещо більші за значення, притаманні їхнім рандомним аналогам. Схоже до явища спалахів (див. закономірності для параметра  $R$ ), кореляції в зображеннях виявляються дещо масштабнішими, ніж у текстах.

**Висновки.** Отже, в цій роботі запропоновано та проаналізовано аналогії між текстами та цифровими зображеннями. Запропоновано нові підходи до аналізу зображень, запозичені зі статистичної лінгвістики та засновані на вивченні наявності та істотності явищ спалахів і довгосяжних кореляцій. Такі підходи дають змогу кількісно відрізнити інформативні зображення від шумів.

Альтернативою вивченого в цій роботі рівня сірого як лінгвістичної одиниці можна обрати колірні характеристики (наприклад, відтінок чи насиченість кольору) або яскравість однієї з RGB-компонент. Ще одним предметом подальших досліджень могло би бути використання точніших за  $R$ -параметр (у т.ч. ентропійних) підходів до визначення мірила спалахів у зображеннях (див., наприклад, праці [8, 21]). Нарешті, сутність статистичного параметра  $R$  можна би було ліпше дослідити, порівнюючи зображення з різними гістограмами та з дозованими рівнями рандомізації. Цікаво було би проаналізувати, наприклад, можливості запропонованих методів у розпізнаванні зашифрованих за методом стеганографії зображень.

#### Список використаних джерел

- [1] *Gonzalez R., Woods R.* Digital image processing / Pearson, 2018. – 1020 p.
- [2] *Clark A., Fox C., Lappin S.* The handbook of computational linguistics and natural language processing / Wiley-Blackwell, 2010. – 802 p.
- [3] *Meriem H., Farida M. H.* Detection of a region of interest in the images based on Zipf laws // Proc. 7<sup>th</sup> Int. Conf. on Signal Image Technol. & Internet-Based Systems. – 2011. – P. 416–421.

- [4] *Manzanera A.*  $\Sigma$ - $\Delta$  background subtraction and the Zipf law / In: "Progress in Pattern Recognition, Image Analysis and Applications (CIARP'07)". – Viña del Mar-Valparaíso, Chile, 2007. 10.1007/978-3-540-76725-1\_5. hal-01222660ff
- [5] *Zanette D. H.* Statistical patterns in written language / Centro Atómico Bariloche, 2012. – 87 p.
- [6] *Altmann E. G., Gerlach M.* Statistical laws in linguistics // Proc. Flow Machines Workshop: Creativity and Universality in Language (Paris, 2014). – arXiv:1502.03296 (2015).
- [7] *Kushnir O. S., Buryi V. O., Grydzhan S. V., Ivanitskyi L. B., Rykhlyuk S. V.* Zipf's and Heaps' laws for the natural and some related random texts // Електроніка та інформаційні технології. – 2018. – Вип. 9. – С. 94–105.
- [8] *Montemurro M. A., Zanette D. H.* Entropic analysis of the role of words in literary texts // Adv. Complex Syst. – 2002. – Vol. 05. – P. 7–17.
- [9] *Ortuno M., Carpena P., Bernaola-Galvan P., Munoz E., Somoza A. M.* Keyword detection in natural languages and DNA // Europhys. Lett. – 2002. – Vol. 57. – P. 759–764.
- [10] *Herrera J. P., Pury P. A.* Statistical keyword detection in literary corpora // Europ. Phys. J. – 2008. – Vol. 63. – P. 135–146.
- [11] *Carpena P., Bernaola-Galván P., Hackenberg M., Coronado A. V., Oliver J. L.* Level statistics of words: finding keywords in literary texts and symbolic sequences // Phys. Rev. E. – 2009. – Vol. 79. – P. 035102(R).
- [12] *Altmann E. G., Pierrehumbert J. B., Motter A. E.* Beyond word frequency: bursts, lulls, and scaling in the temporal distributions of words // PLoS ONE. – 2009. – Vol. 4. – P. e7678.
- [13] *Кушнір О. С., Волоско А. В., Іваніцький Л. Б., Рихлюк С. В.* Про статистику відстаней між словами в тексті та проблему розпізнавання змістових слів // Електроніка та інформаційні технології. – 2016. – Вип. 6. – С. 155–164.
- [14] *Carretero-Campos C., Bernaola-Galván P., Ivanov P. Ch., Carpena P.* Improving statistical keyword detection in short texts: Entropic and clustering approaches // Phys. Rev. E. – 2012. – Vol. 85. – P. 011139.
- [15] *Кушнір О. С., Альфавіцький М. А., Дзіковський В. С., Іваніцький Л. Б., Рихлюк С. В., Сокульський В. І.* Статистика появи слів у природних і рандомних текстах // Вісник нац. ун-ту «Львівська політехніка». Серія «Інформаційні системи та мережі». – 2017. – №872. – С. 162–178.
- [16] *Altmann E. G., Cristadoro G., Esposti M. D.* On the origin of long-range correlations in texts // Proc. Natl. Acad. Sci. (USA). – 2012. – Vol. 109. – P. 11582–11587.
- [17] *Kushnir O. S., Ivanitskyi L. B., Kashuba A. I., Mostova M. R., Mykhaylyk V. B.* Large-scale studies of the repetition characteristic for different models of symbolic sequences // Proc. 12<sup>th</sup> IEEE Int. Conf. on Electron. and Inf. Technol., 2021. – P. 61–66.
- [18] *Kantelhard J. W.* Fractal and multifractal time series / In: Mathematics of complexity and dynamical systems (Ed. by Meyers R. A.). – New York: Springer, 2011. – P. 463–487.
- [19] *Яремків В. В., Кушнір О. С., Іваніцький Л. Б.* Довгосяжні кореляції символів і послідовностей символів у текстах: метод флуктуаційного аналізу // Матер. ІХ Укр.-Польськ. наук.-практ. конф. «Електроніка та інформаційні технології». – Львів : Видавн. Львів. ун-ту, 2017. – С. 29–33.

- [20] Іваницький Л. Б., Кушнір О. С., Яремків В. В., Альфавицький М. А. Метод DFA для аналізу довгосяжних кореляцій у часових послідовностях // Матер. ІХ Укр.-Польськ. наук.-практ. конф. «Електроніка та інформаційні технології». – Львів : Видавн. Львів. ун-ту, 2017. – С. 59–62.
- [21] Carpena P., Bernaola-Galván P. A., Carretero-Campos C., Coronado A. V. Probability distribution of intersymbol distances in random symbolic sequences: applications to improving detection of keywords in texts and of amino acid clustering in proteins // Phys. Rev. E. – 2016. – Vol. 94. – P. 052302.

## ANALOGIES BETWEEN IMAGES AND TEXTS: ‘CLUSTERIZATION’ PHENOMENON IN TEXTS AND DIGITAL IMAGES

I. Ya. Dovhan, O. S. Kushnir, Yu. M. Furhala, H. I. Panochko

*Department of Optoelectronics and Information Technologies,  
Ivan Franko National University of Lviv,  
107 Tarnavsky Street, UA-79017 Lviv, Ukraine  
[mozmarto@gmail.com](mailto:mozmarto@gmail.com), [oleh.kushnir@lnu.edu.ua](mailto:oleh.kushnir@lnu.edu.ua)*

We develop a number of analogies between texts and images, where different pixel-brightness levels (or gray values) in an image correspond to different linguistic elements of some level (such as letters, symbols, words, etc.) in a text. This correspondence is possible only in the case of discrete structures such as digital images. A number of practical recipes for converting two-dimensional images into one-dimensional linear chains of pixels (i.e., ‘texts’) are discussed. In particular, this can be consideration of separate rows (or columns) in an image – or a series of sequential rows (or columns) in it, where a sequential number of pixel in a ‘one-dimensional image’ is analogue of a position of linguistic element in a text. The advantages and shortcomings of these recipes are analyzed. Besides, we introduce analogues of the statistical-linguistic notions of ‘vocabulary’ and ‘rank dependence’ in the case of images.

Then we clarify the possibilities for application to digital images of standard statistical-linguistic techniques for detecting keywords in texts. In particular, we remind of the phenomenon of clustering of words in a text, which is very weak or absent for so-called function words – and pronounced for keywords. A simplest clustering parameter  $R$  is introduced that relates to the first statistical moments of the probability distribution for the waiting times of a given word in a text. It quantizes the scale of the above phenomenon for this word. After that, we analyze the latter phenomenon in the ‘texts’ that correspond to two different digital images.

Our main result is that the ‘texts’ corresponding to a pure white noise and a simple informative image differ notably by their  $R$ -parameters associated with different ‘words’ (i.e., brightness levels). A possible meaning of ‘keywords’ in an image is discussed, which can be associated with some ‘semantic load’ of the appropriate brightness levels. We also advise to check the availability of long-range correlations among the same brightness levels in an image. This can be done according to one of standard techniques known for the studies of correlations (e.g., fluctuation analysis or detrended fluctuation analysis).

*Key words:* statistical linguistics, natural language processing, clustering, keywords, digital image processing.

*Стаття: надійшла до редакції 24.02.2022,  
прийнята до друку 25.02.2022.*