

УДК 621.397:004.056

КОРОЛЬОВ В.Ю., Інститут кібернетики НАН України,
ПОЛІНОВСЬКИЙ В.В., Відкритий міжнародний університет розвитку людини “Україна”,
ГЕРАСИМЕНКО В.А., Інститут кібернетики НАН України,
ГОРИНШТЕЙН М.Л., Інститут кібернетики НАН України

ДОСЛІДЖЕННЯ КОЛЬОРОВИХ ЦИФРОВИХ ФОТОГРАФІЙ МЕТОДАМИ RS-СТЕГОАНАЛІЗУ ТА СТАТИСТИКИ

Анотація. Про результати дослідження статистичних властивостей кольорових зображень за методом RS-стегоаналізу.

Аннотация. О результатах исследования статистических свойств цветных изображений по методу RS-стегоанализа.

Summary. On the results of researches of statistical properties of color images according to the method RS-stegoanalysis..

Ключові слова: кольорові зображення, цифрові зображення, RS-стегоаналіз, оцінка дисперсії шуму, метод ковзного вікна, статистика, середньоквадратичне відхилення оцінки шуму.

Дослідження статистичних характеристик цифрових зображень є основою для розробки нових методів фільтрації шумів, покращення методів двомірної інтерполяції, синтезу тримірних моделей віртуальної реальності тощо. Відомо, що багато методів цифрової обробки зображень було запозичене з оптики, кібернетики, спеціальних розділів математики, радіоелектроніки і використовують особливості сприйняття візуальних образів людиною.

Аналіз останніх досліджень та виділення не вирішених раніше частин проблеми. Сучасні прикладні дослідження характеристик зображень та сигналів ґрунтуються на використанні програмних комплексів комп'ютерної математики типу MATLAB, MathCAD, Mathematica та ін., які містять широкий набір програм цифрового аналізу і обробки сигналів та допускають створення нових [1] шляхом написання функцій користувачем у цих комплексах або на відомих мовах програмування. При оцінюванні параметрів отриманих даних завжди виникає проблема розрізнення шумів природного, технічного походження та артефактів реєстрації від корисної для даного дослідження інформації. Пошук методів розв'язання цієї проблеми призвів до створення класичних методів статистичного оцінювання параметрів експериментальних результатів, спектрального та вейвлет-аналізу, великого набору методів фільтрації тощо.

Вищезазначені методи ґрунтуються на роботі з дійсними або цілими числами, а значення, що знаходяться на межі розрядності процесора, трактуються як можлива арифметична похибка. RS-стегоаналіз [2, 3], навпаки, досліджує закономірності і зв'язки для найменш значимих біт даних з сусідніми бітами та тими бітами, які знаходяться на рівень вище. Математичний апарат RS-стегоаналізу включає не тільки статистичні, а й комбінаторні методи дослідження даних, що обумовило його унікальну здатність до розпізнавання і оцінки величин штучних змін у зображеннях, не доступних традиційним статистичним методам. Тому дослідження можливостей RS-стегоаналізу, а також розгляд концепцій похідних від нього алгоритмів для суміжних з стеганографією областей цифрової обробки сигналів, статистичного оцінювання параметрів експериментальних даних є актуальною проблемою. Зазначимо, що до запропонованої роботи ця тема ще не була предметом досліджень.

Метою статті є:

- дослідження масивів зображень за допомогою RS-стегааналізу;
- порівняння отриманих результатів з даними статистичних досліджень масивів зображень за відомими методами;
- визначення перспективи запозичення методики RS-стегааналізу для підвищення якості аналізу цифрових фотографій;
- опис методики перевірки придатності цього методу виявлення прихованих даних для створення нових методів фільтрації зображень після відповідних модифікацій.

Постановка задачі. У роботі розглянуто використання RS-стегааналізу для досліджень масивів зображень, їх реакції на застосування сучасних складних цифрових фільтрів, визначення чутливості статистичних характеристик зображень до малих стохастичних збурень, викликаних додаванням даних (стегобіт) у найменш значимі біти (далі – НЗБ) зображень. Дослідження властивостей кольорових зображень за методом RS-стегааналізу було виконано на декількох великих вибірках зображень різного класу, що підтверджує статистичну достовірність наведених даних. Подальші дослідження у вказаному напрямі можуть бути використані для побудови нових методів комп'ютерної обробки зображень.

Комплекс аналізу зображень. Для дослідження масивів зображень різними методами був розроблений універсальний комплекс аналізу зображень з модульною архітектурою, який дозволяє додавати нові формати зображень та алгоритмів їх аналізу, а також виконувати обробку без зміни самого комплексу. Отримана статистика зберігається в базі даних, структура якої дозволяє зберігати та обробляти статистику нових модулів аналізу даних. Також комплекс має досить широкі можливості вибірки та аналізу накопичених даних. Цим він відрізняється від наявних рішень, більшість з яких є вузькоспеціалізованими і виконують аналіз зображень лише одного типу та одним алгоритмом, результати обробки також представляються у своєму форматі.

Робота з комплексом розбивається на декілька етапів:

1. Вибір файлу бази даних, в якому буде зберігатись статистика.
2. Додавання папок із зображеннями, які необхідно обробити. При цьому підтримується рекурсивна обробка підпапок та завдання списку масок файлів для обробки.
3. Завдання послідовності екземплярів фільтрів та аналізаторів, якими будуть обробляться та аналізуватись зображення. Фільтр – це модуль, який змінює зображення, наприклад, виконує його фільтрацію чи приховування даних певним алгоритмом. Аналізатор же зображення не змінює, а виконує його аналіз і повертає певний набір статистики. Для модулів обох типів можуть задаватись набір специфічних для них параметрів (наприклад, коефіцієнти роботи алгоритмів фільтрації та аналізу) – таким чином створюються екземпляри фільтрів та аналізаторів, які і додаються в послідовність.
4. Запускається на виконання завдання обробки файлів по даних попередніх етапів. Реалізація обробки виконана за схемою робочих потоків, які незалежно оброблюють файли зображень, завдяки чому підвищено ефективність паралельної обробки на SMP-системах. Підтримка розподілених систем планується в наступних версіях комплексу. Після закінчення обробки накопичена статистика доступна для вибірок та експорту у вигляді звітів трьох типів (Рис. 1).

Статистика по зображеннях дозволяє отримати результати аналізу зображень по кожному файлу окремо. При цьому можна вибрати необхідний набір даних, які будуть виводитись для зображень, а також задати умову фільтрації за ними, наприклад: ([Ширина] > 1000) and ([Висота] > 1000) and ([Г'мя] like 'nature%'). На другій вкладці можна вибрати набір даних статистики, яку необхідно вивести. Набір доступних даних статистики

відображається у вигляді дерева, яке містить послідовності екземплярів фільтрів та аналізаторів з вкладеними списками статистики, яку можна включити у звіт (рис. 2).

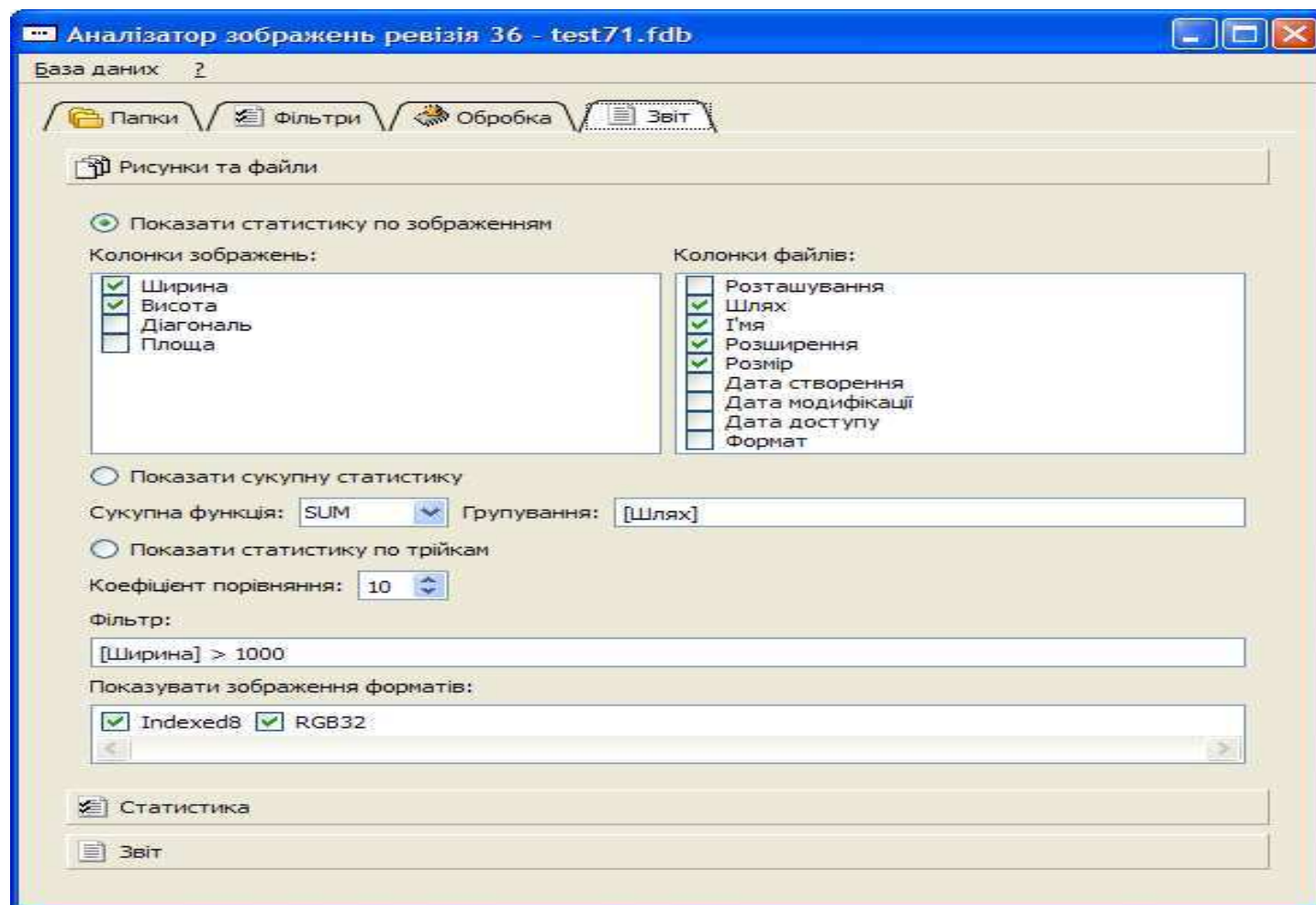


Рис. 1. Параметри отримання звіту зі статистикою

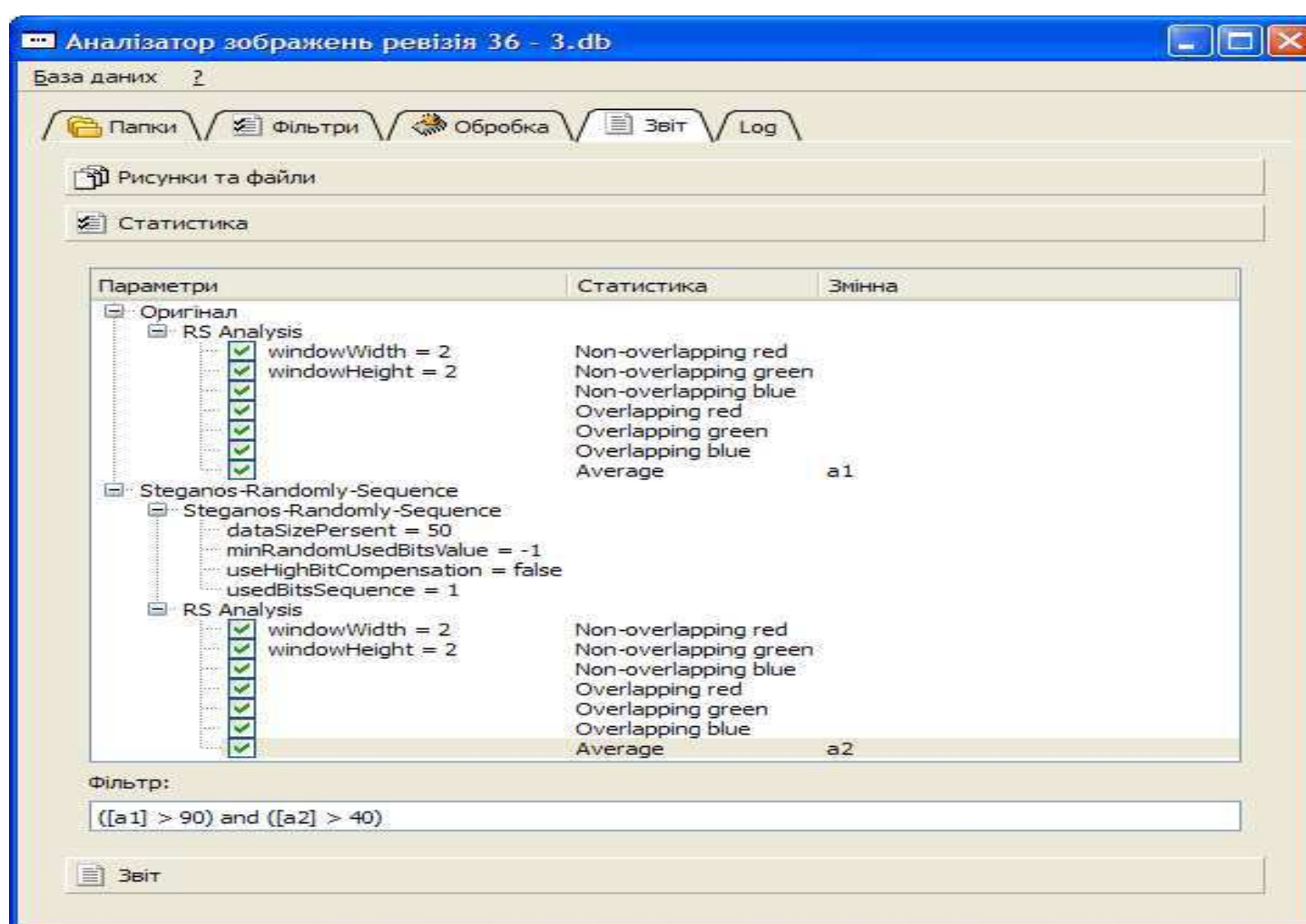


Рис. 2. Вибір статистики для звіту

Після завдання параметрів звіту можна переглянути результат на третій вкладці. Його можна експортувати в csv-файл для подальшої обробки в табличному процесорі типу Майкрософт Ексель.

В комплексі реалізовано декілька модулів фільтрації та аналізу даних. Модулі фільтрації включають набір стандартних графічних фільтрів, таких як: GaussianBlur, Defocus, Highlight, Sharpen, BigEdge, Emboss, EmbossColor, EdgeDetect, Negative, RemoveChannel, Punch, а також п'ять фільтрів приховування даних, які є різновидами алгоритму LSB.

Також реалізовано чотири аналізатори:

1) математичної статистики – обраховує три статистичні характеристики: середнє значення, середньоквадратичне відхилення та медіану для декількох характеристик з різних кольорових просторів: R,G,B, RGB, середнє RGB, сірий, відстань RGB, H,S,L, H,S,V, C,M,Y,K, L,a,b;

2) статистики по шуму – розраховує дві шумові характеристики для тих же кольорових просторів;

3) порівняння кольорів – видає статистику за 13-ма співвідношеннями між значеннями інтенсивностей кольорових складових пікселів зображення (R,G,B);

4) RS-аналіз.

Дослідження закономірностей у каналах кольорових зображень за методом RS-стегааналізу. В результаті досліджень, проведених над 49678 зображеннями (оригінальними фотографіями без стегобіт), які доступні в Інтернеті, було визначено, що хибне визначення проценту виявлених стегобітів (ПВС) для них вище, ніж вказане авторами патенту для фотографій з цифрових камер і сканерів. На думку авторів, це пов'язано з тим, що над такими зображеннями було виконано обробку для підготовки до друку, яка включає високочастотну фільтрацію, підвищення яскравості і контрасту, вирівнювання значень гістограми і т.п. операції, які підвищують візуальну якість зображення. Це супроводжується збільшенням сингулярних груп у цифровій фотографії.

Також було помічено нерівномірне зашумлення по каналах для кольорових зображень. Символами R, G, B позначено значення коефіцієнту, отриманого за RS-стегааналізом для червоного, зеленого та синього каналів відповідно. Очевидно, що частотами співвідношень кольорів – нерівностями можна побудувати множину всіх можливих подій у вибірці. Елементарні події повної множини подані у табл. 3.

Вибірка № 3 з 1700 фотографій була зроблена цифровими камерами фотоапаратів, мобільних телефонів. Видно, що закономірність розподілу нерівностей зберігається і для фотографій, до яких не застосовувались цифрові перетворення та відсутні цифрові водяні знаки.

В результаті обчислень отримано наступні частоти появи нерівностей у відсотках, при цьому похибка для рівностей і нерівностей була встановлена в межах 1 %, щоб врахувати обмеження розрядності обчислень (табл. 1). Загальна кількість оброблених зображень становила 51378 файлів.

Таблиця 1

Відносні частоти появи нерівностей для двох вибірок

| Вибірки | Нерівності | | | | | | | | | | | | |
|---------|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | R>G>V | R>V>G | G>R>V | G>V>R | V>R>G | V>G>R | R=G>V | V>R=G | G=V>R | R>G=V | V=R>G | G>V=R | R=G=V |
| 1 | 13,46 | 18,26 | 8,29 | 9,43 | 21,99 | 19,43 | 1,03 | 2,36 | 1,11 | 1,72 | 1,28 | 1,35 | 0,28 |
| 2 | 16,12 | 18,95 | 11,23 | 10,95 | 19,05 | 16,97 | 0,86 | 1,33 | 0,97 | 0,92 | 1,25 | 0,97 | 0,43 |
| 3 | 10,71 | 12,18 | 7,82 | 11,65 | 25,29 | 30,06 | 0,12 | 0,59 | 0,59 | 0,24 | 0,35 | 0,24 | 0,18 |

Зазначимо, що висока повторюваність частот появи нерівностей дозволяє стверджувати, що явище має високу статистичну стійкість. Видно також, що нерівності, в яких переважає RS-коефіцієнт у червоному або синьому каналі, переважають за частотою появи інші (Табл. 2).

Таблиця 2

Відносна частота появи кольорів на першому місці у нерівностях

| Назва кольорового каналу | Значення максимуму, % | | |
|--------------------------|-----------------------|-----------|-----------|
| | Вибірка 1 | Вибірка 2 | Вибірка 3 |
| Блакитний (B) | 43,78 | 37,35 | 55,94 |
| Червоний (R) | 33,44 | 35,99 | 23,13 |
| Зелений (G) | 19,07 | 23,15 | 19,71 |

Нерівності, що відповідають RS-стегааналізу (Табл. 1, 2), можуть безпосередньо використовуватись для перевірки великих масивів зображень на предмет вбудовування прихованих даних або для створення стеганографічного алгоритму керування заповненням інформацією за НЗБ-методом для великих масивів зображень, що не впливає на сукупну статистику для набору зображень.

Дослідження впливу цифрових фільтрів на значення ПВС. Відомо, що зображення з мереж загального доступу можуть мати цифрові водяні знаки, над ними можуть бути виконані операції покращення візуальної якості або спеціальні дизайнерські перетворення типу підготовки до друку. Тому було взяти 1700 фотографій з аматорської колекції, зроблені мобільним телефоном та цифровою камерою, над якими гарантовано не виконували будь-які перетворення. Далі вибірка була оброблена існуючими фільтрами (посилення кольору, зменшення/збільшення різкості, замулення і т.п.), і було виконано RS-стегааналіз, результати якого наведено у Табл. 3, 4.

Таблиця 3

Кількість зображень у групах, що відповідають фільтрам, які збільшують ПВС

| Назва фільтру і класифікація операції | Кількість зображень, % | | | | |
|---------------------------------------|------------------------|-----------|-------|--------|------------|
| | 2 % - max | | | | |
| | 0-2 % | 2 % - max | 2-5 % | 5-10 % | 10 % - max |
| Оригінал | 80,7 | 19 | 13 | 4,6 | 1,7 |
| Multiply_15 | 74,2 | 26,8 | 19,2 | 4,6 | 2 |
| Overlay_15 | 75 | 25 | 20 | 4 | 1 |
| High Pass_15 | 78 | 22 | 13 | 7 | 2 |
| Reduce_noise | 80,9 | 19,1 | 14,4 | 3,6 | 1,1 |
| AutoContrast+ FocusRestoration | 3,1 | 96,9 | 12,5 | 19 | 65,4 |
| Normalize | 79,27 | 20,73 | 13,7 | 5,15 | 1,88 |

За результатами експерименту було виявлено, що більшість зображень можна розділити на три групи:

1) зображення, на які перетворення суттєво впливають в напрямі зменшення або збільшення ПВС;

2) зображення, ПВС яких мало змінюється в результаті застосування фільтрів та перетворень будь-яких типів;

3) зображення, ПВС яких змінюється при застосуванні одних перетворень і мало змінюється при використанні інших.

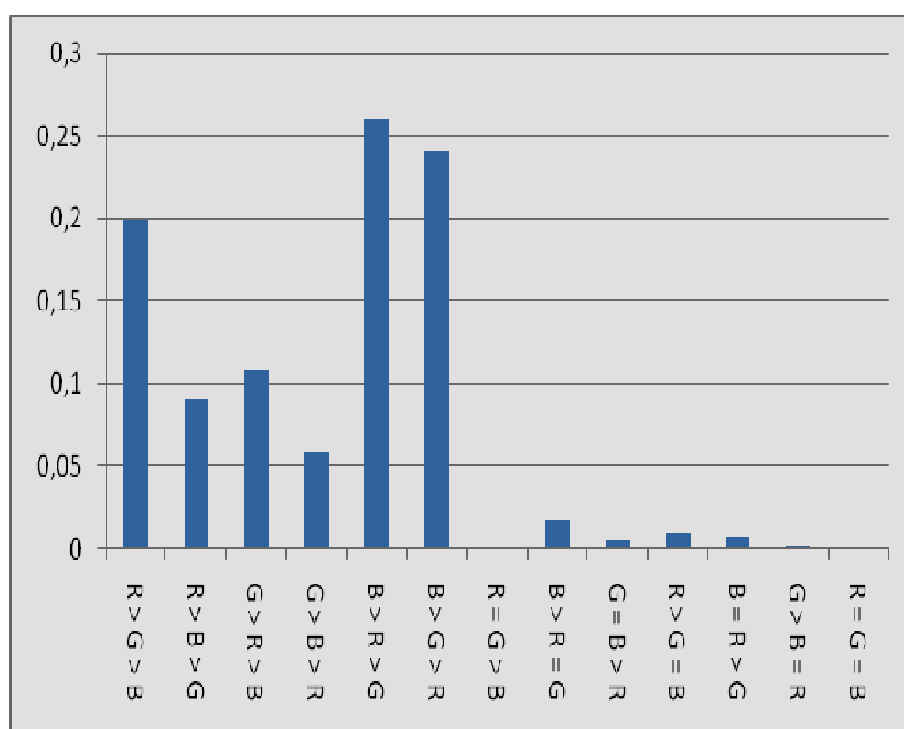
За впливом конкретного фільтру на вибірку можна виділити фільтри, що зменшують ПВС (табл. 4), та такі, що збільшують ПВС (табл. 3). Дослідження будуть продовжені в напрямі розширення номенклатури фільтрів і їх комбінацій, що будуть застосовуватись до вибірок цифрових фотографій.

Таблиця 4

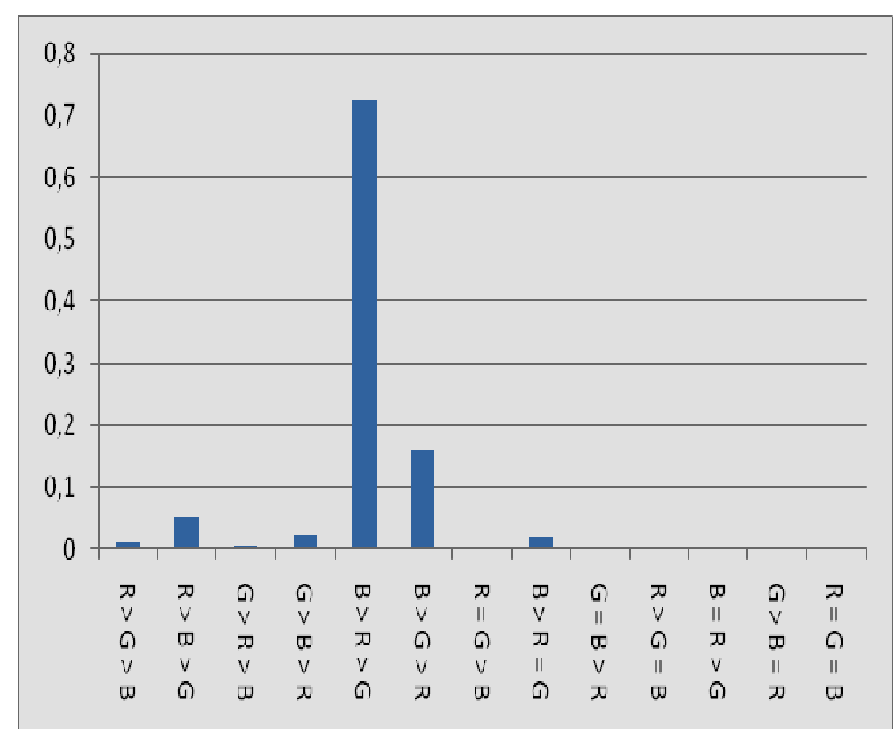
Кількість зображень у групах, що відповідають фільтрам, які зменшують ПВС

| Назва фільтру і класифікація операції | Кількість зображень, % | | | |
|---------------------------------------|------------------------|-------|--------|------------|
| | 2 % - max | 2-5 % | 5-10 % | 10 % - max |
| Оригінал | 19 | 13 | 4,6 | 1,7 |
| Медіанний | 0,06 | 0,06 | - | - |
| Гауссівський | 0,35 | 0,35 | - | - |
| Minimum_3x3 | 1,94 | 1,94 | - | - |
| Average_3x3 | 2,12 | 2,12 | - | - |

Нерівності за середньоквадратичним відхиленням оцінки шуму (СКВОШ) (Рис. 3 а, б) мають форму, подібну до нерівностей для ПВС при відсутності стегобіт, проте їх співвідношення практично не змінюється при додаванні стегобіт у НЗБ-рівень зображень, оскільки додавання стегобіт змінює величини СКВОШ кольорових каналів лише у сотих і десятих частках чисел [10 – 12].



а



б

Рис. 3. Гістограма для нерівностей СКВ- оцінки шуму

Видно, що зберігається залежність у зростанні частот нерівностей від синього до червоного і мінімум для зеленого. В той же час, за абсолютними значеннями СКВОШ у каналах немає закономірності для мінімумів, максимумів і середніх значень за вибіркою зображень, а лише має місце лінійне зростання в сотих частках цих величин при додаванні стегобіт.

За аналогією до попередніх статистичних досліджень було перевірено закономірності і для середніх значень, медіан та СКВОШ у вибірках зображень (Рис. 4). Як видно з рисунку, гістограми середніх значень мають максимуми у $R > G > B$ та у $B > R > G$.

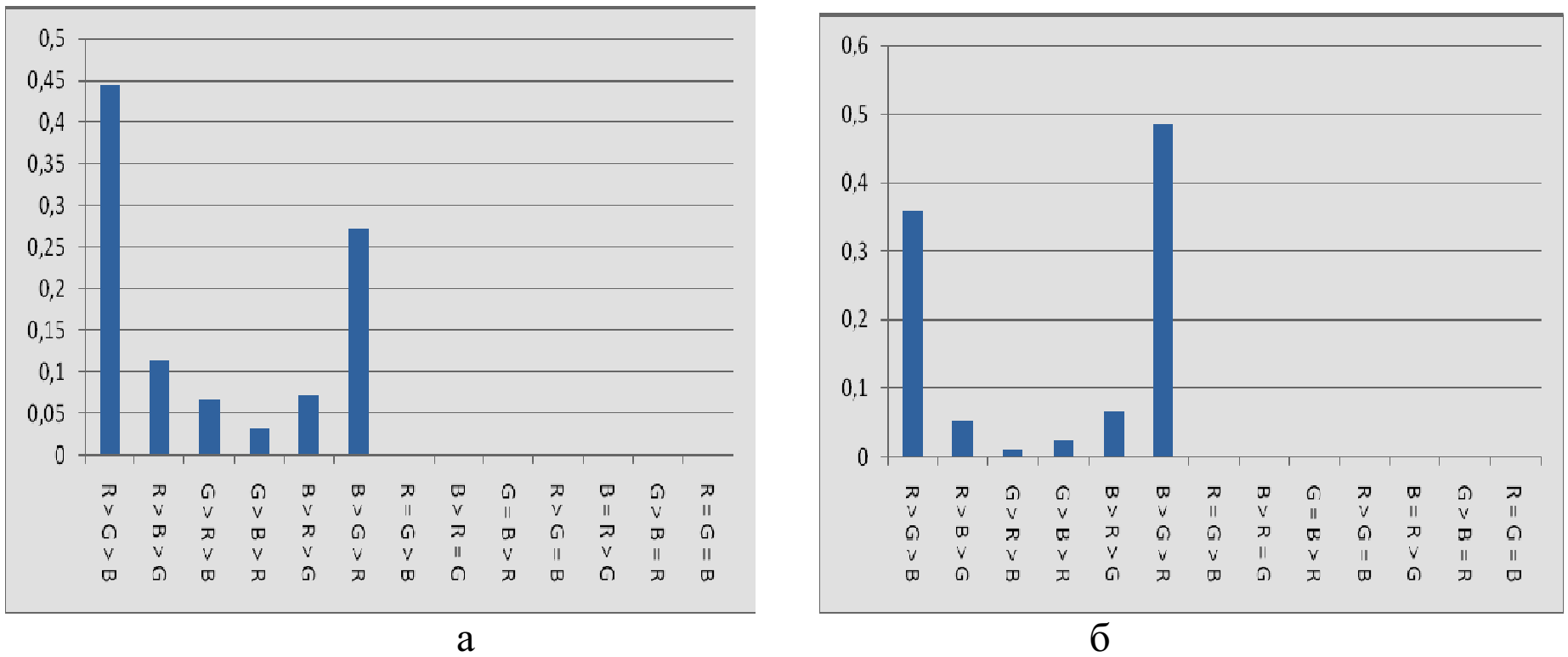


Рис. 4. Гістограми частот нерівностей для середнього значення у виборці.
(а) – Test, (б) – Datacr

Таку особливість можна пояснити специфікою побудови ПЗЗ-матриць сучасних цифрових фотоапаратів, які будуються на базі баєрівських матриць. Для медіан і СКВОШ у двох вибірках гістограми мають аналогічний вигляд з трохи більшими максимумами у таких же нерівностях. Подібні гістограми з екстремумами для середніх значень, медіан і СКВОШ, подані на Рис. 3 – 5 для системи RGB, були отримані у вибірках: Test, Datacr та Вибірка № 3 і для зображень у системі HLS.

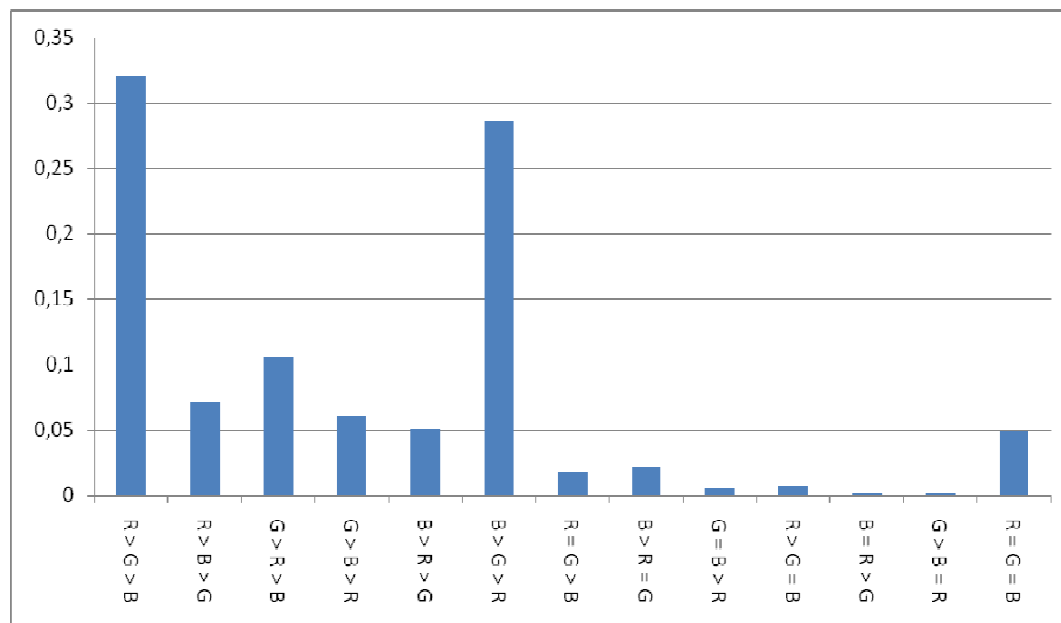


Рис. 5. Гістограма частот нерівностей для байтів кольорових каналів (53308 файлів)

Виявлений факт, на думку авторів, можна застосувати для покращення характеристик алгоритмів стиску великих масивів зображень шляхом групування зображень за нерівностями і створення методів доказової цифрової ідентифікації фотоапаратів, якими було зроблено серію зображень.

Аналогічні статистичні дослідження були виконані для байтів кольорів цифрових фотографій на вибірці обсягом 53308 файлів (Рис. 5). Видно, що отримана гістограма подібна до Рис. 3, 4 та даних Табл. 3, 4.

Висновки.

Встановлено, що для кольорових каналів характерне різне значення хибно позитивного ПВС, розрахованого за методом RS-стегааналізу. Більш високі значення для блакитного каналу можна пояснити більш швидким релеєвським розсіюванням (поглинанням) синього спектра повітрям, порівняно з іншими кольорами, що відображається як підвищення рівня зашумлення у синьому кольоровому каналі. Високі значення для червоного кольору пояснюються більш високою контрастністю (детальністю) для червоного кольору, ширина спектра якого є найбільшою поміж трьох кольорів, а відносний ступінь поглинання є найменшим, оскільки з нього починається спектр видимого світла. Найнижче відносне значення зафіксовано для зеленого кольору, який займає проміжне місце у спектрі між червоним і блакитним.

Крім того, за результатами обробки зображень різними фільтрами було визначено, що їх можна розділити на три групи: зображення, в яких перетворення суттєво впливають на ПВС, зображення, ПВС яких мало змінюється в результаті застосування фільтрів, та зображення, ПВС яких змінюється при застосуванні одних перетворень і мало змінюється при використанні інших. Це дозволило виділити ті фільтри, які суттєво впливають на ПВС зображень, що може бути використано при створенні стеганоалгоритмів, стійких до стегааналізу.

Приховувати великий обсяг даних у масивах зображень, при якому кольорові канали фотографії рівномірно заповнюються на 10 – 15 % методами НЗБ-стегаграфії, не рекомендується, оскільки це змінює природну статистику нерівностей ПВС для зображень. Щоб збільшити стійкість приховування даних у масивах зображень до виявлення RS-стегааналізом необхідно неоднорідно вбудовувати стегобіти у кольорові канали зображень, тобто у відповідності з отриманими нерівностями для вибірок.

Отримані результати – нерівності для різних статистичних характеристик – можуть бути використані для розробки нових методів зменшення рівня шумів, перевірки зображень на предмет застосування методів цифрової обробки та вбудовування сторонніх елементів, а також для судово-доказової ідентифікації апаратних засобів реєстрації (digital forensic), якими були зроблені масиви цифрових фотографій.

Використана література

1. Корольов В.Ю. Концепція використання бібліотек комп'ютерної графіки для умовної деконволюції образів // Вісник Хмельницького національного університету. – 2006. – № 4. – С. 140 – 145.
2. J. Fridrich, M. Goljan, and R. Du, “Detecting LSB Steganography in Color and Gray-Scale Images”, Magazine of IEEE Multimedia, Special Issue on Security, October-November issue, 2001, pp. 22 – 28.
3. Fridrich J., Goljan M., Du R. Lossless Data Embedding – New Paradigm in Digital Watermarking // Special Issue on Emerging Applications of Multimedia Data Hiding. – 2002. – № 2. – P. 185 – 196.
4. Корольов В.Ю., Поліновський В.В., Герасименко В.А. Стегаграфічна персоналізація інформації на базі ПК // Вісті Академії інженерних наук України. – 2009. – № 2(39). – С. 18 – 24.

5. Корольов В.Ю., Поліновський В.В., Герасименко В.А. Дослідження стійкості НЗБ-стеганографії до RS-аналізу : матеріали IV Міжнар. конф. “Сучасні проблеми радіоелектроніки, телекомунікацій та приладобудування (СПРТП-2009)”. – (Ч. 1). – Вінниця : ВНТУ МОН України, 2009. – С. 53.

6. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография. – М. : СОЛОН-Пресс, 2002. – 272 с. – (Серия “Аспекты защиты”).

7. Основи комп’ютерної стеганографії : навч. посібник / [В.О. Хорошко, Л.Д. Азаров, М.Є. Шелест, Ю.Є. Яремчук]. – Вінниця : ВДТУ, 2003. – 143 с.

8. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. – К. : МК-Пресс, 2006. – 288 с.

9. Digital Watermarking and Steganography / [I.J. Cox, M.L. Miller, J.A. Bloom, J. Fridrich, T. Kalker]. – Morgan Kaufmann Publishers, 2008. – 593 p.

10. Корольов В.Ю., Поліновський В.В., Герасименко В.А. Визначення можливостей RS-стегааналізу для дослідження статистичних властивостей зображень // Вісник Хмельницького національного університету. – 2010. – № 4. – С. 102 – 110.

11. Корольов В.Ю., Полиновский В.В., Герасименко В.А. Стеганография по методу наименее значимого бита на базе персонализированных флеш-накопителей // Управляющие системы и машины. – 2011. – № 1 (231). – С. 79 – 87.

12. Корольов В.Ю., Поліновський В.В., Герасименко В.А. RS-стегааналіз. Принципи роботи, недоліки та концепція метода його обходу // Вісник Вінницького політехнічного інституту. – 2010. – № 6. – С. 66 – 71.

~~~~~ \* \* \* ~~~~~