UDC 004.056.5

Informatics and Mathematical Methods in Simulation
Vol. 6 (2016), No. 4, pp. 357-364

METHOD OF DETECTION THE FACT OF COMPRESSION IN DIGITAL IMAGES AS AN INTEGRAL PART OF STEGANALYSIS

A.V. Akhmametieva

Odesa National Polytechnic University,
1 Shevchenko Str., Odesa, 65044, Ukraine; e-mail: anna-odessitka@mail.ru

The new method of detection the fact of compression in color digital images, allowing separate images, initially stored in lossless formats, from images, resaved into lossless format from losses formats, which analyzes spatial domain of digital contents, that allows avoid additional accumulation of a computing errors, is presented. The developed method is based on the accounting sequential color triads of triplets in the matrix of unique colors, and can be used to define the format of the original container in steganalysis.

Keywords: digital image, compression, losses format, lossless format, spatial domain, sequential color triads of triplets

Introduction

Nowadays developments in the field of steganography are widely used, providing transmission of confidential information via open channels of communication, hiding at the same time the fact of its presence at a container, that can be used by criminals. Digital images both in losses formats and in lossless formats can be used as containers (cover images) in steganography. After embedding of additional information into spatial domain of the container by LSB method the stego, that is the result of embedding of additional information into container, is saved in lossless format, so in the process of steganalysis, aimed at the detecting the fact of the presence/absence of hidden information in any digital content [1], as a rule, there is digital content in lossless format in the absence of an original container. The possibility of separating digital images, stored in lossless formats initially, from digital imaged, resaved into lossless format from losses formats, can increase efficiency of steganalysis, having specified the action which is carried out with digital image.

The problem of determining the format of the original container is essential in different sources: for method developed in [2-4] selection of different thresholds depending on the container format is required; proposed in [5-6] method is designed for containers in losses formats [7]; different conditions for digital images in losses/lossless formats are applied in method proposed in [8]. In [9] an approach for separation of digital images, stored in lossless formats initially, from digital imaged, resaved into lossless format from losses formats, is proposed, however, as a computational experiment conducted by the author showed, opposite results are received for various technical devices with different matrixes of quantization, what requires availability of data on matrixes of quantization of various devices and the ratios corresponding to them for ensuring effective work of the algorithm, that implements this approach. Besides, the approach proposed in [8] implies the transfer of a digital content to transformation domain that leads to the additional accumulation of computational errors, so the development of a similar method, that carry out the analysis of the spatial domain, is an actual problem.

The aim of the paper is development of the method that uses spatial domain of digital content for the analysis and allows separate digital images, stored in lossless formats initially, from digital imaged, resaved into lossless format from losses formats, which is a component of steganalysis.

Main part

In [5-6] was proposed steganalytic method, aimed to detect the presence/absence of the additional information in digital images stored in losses format initially, based on the accounting of sequential color triads of triplets in the matrix of unique colors UCT of size $U \times 3$, containing unique color triplets (r_i, g_i, b_i), $i = \overline{1,U}$, which at least once occur in the analyzed digital image, where r - brightness value of red color component, g - brightness value of green color component, b - brightness value of blue color component, U - number of unique color triplets.

Under a sequential Red- (Green-, Blue-) triad of triplets in the matrix of unique colors we will call triads:

$$(r_i, g_i, b_i) \in UCT \ \& \ (r_i + 1, g_i, b_i) \in UCT \ \& \ (r_i - 1, g_i, b_i) \in UCT, \ i = \overline{1,U}; \quad (1)$$

$$(r_i, g_i, b_i) \in UCT \ \& \ (r_i, g_i + 1, b_i) \in UCT \ \& \ (r_i, g_i - 1, b_i) \in UCT, \ i = \overline{1,U}; \quad (2)$$

$$(r_i, g_i, b_i) \in UCT \ \& \ (r_i, g_i, b_i + 1) \in UCT \ \& \ (r_i, g_i, b_i - 1) \in UCT, \ i = \overline{1,U}; \quad (3)$$

respectively. When counting the sequential triads of triplets we will associate sequential triad with $(r_i, g_i, b_i) \in UCT$, for which execution of the conditions (1), (2) or (3) depending on the type of triad is carried out. Triplet $(r_i, g_i, b_i) \in UCT$ we will call middle if there is a sequential triad for him.

The basic triad is a sequential triad corresponding to that color component of the container, into which embedding of additional information was carried out.

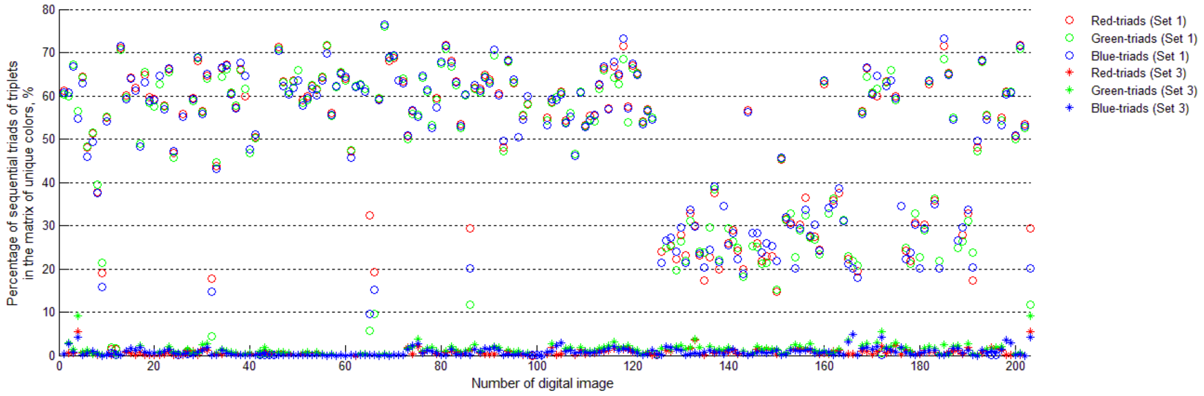
Concomitant triads are that sequential triads that corresponds to unfilled color components of the container.

In [7] the computational experiment, aimed to analyze the quantity of Red-, Green- and Blue-triads in unfilled digital containers stored in losses/lossless formats, was carried out. In computational experiments digital image database is used, which includes:

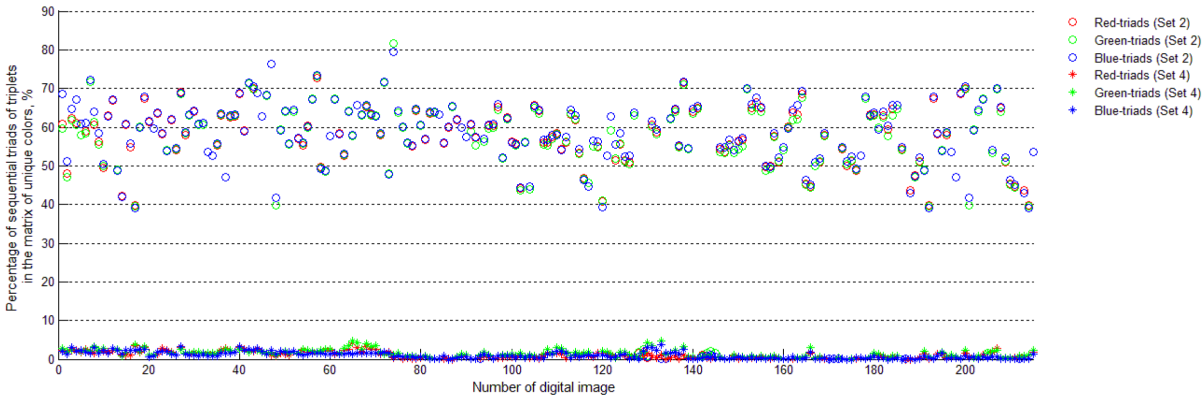
1. Set 1: 200 color digital images from [10] in TIFF format;
2. Set 2: 200 images received by non-professional photo cameras in TIFF format;
3. Set 3: 203 color digital images from [10] in JPG format;
4. Set 4: 215 images received by non-professional photo cameras in JPG format.

As a result of computational experiment the characteristic difference between losses and lossless formats was revealed. As computational experiments showed, the matrix of unique colors of the unfilled containers stored in losses format contains insignificant quantity of middle triplets (generally up to 3%), containers stored in lossless format are characterized by high quantity of middle triplets in the matrix of unique colors (on average 40-60%) due to the lack of compression and, consequently, a large variety of unique colors, wherein percentage of Red-, Green-, Blue-triads of triplets differ slightly from each other.

Figure 1 shows the distribution of the quantity of sequential triads of triplets in the matrix of unique colors for digital images from Set 1 (TIFF) and 3 (JPG) from the NRCS Base [10] (Fig. 1a), and digital images from Set 2 (TIFF) and 4 (JPG) received by non-professional cameras (Fig. 1b).



a



b

Fig. 1. Comparison of the distribution of the quantity of sequential triads of triplets in the matrix of unique colors of original digital images: a – from Set 1 and 3; b – from Set 2 and 4

As can be seen from Fig. 1 the approach based on the analysis of sequential triads of triplets in the matrix of unique colors allows separate digital images, stored in lossless formats initially, from digital images, resaved into lossless format from losses formats. Especially significantly digital images from Set 2 and 4 differ, where there are practically no intersections.

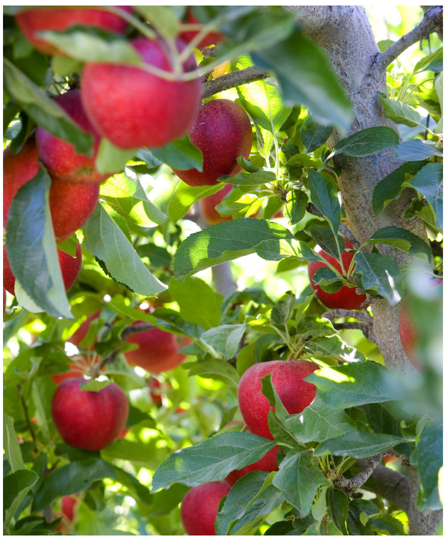
However there is a question: is it possible separate containers stored in losses format from containers stored in lossless format in case when the analyzed content is not original, but stego, formed by embedding of additional information by LSB method. To answer this question computational experiment on the basis of digital images from Set 1-4 was carried out that analyzes the percentage of sequential triads of triplets in the matrices of unique colors of stego, formed by embedding of additional information with varying degrees of filling and different values of hidden capacity: into one color component, into two color components, into all three color components. The comparative characteristic of quantity of sequential triads of triplets for the JPG and TIFF formats is provided in table 1.

An illustration of change of quantity of sequential triads in a matrix of unique colors serves Figure 2, where additional information embeds only into red color component (for example) of digital image with different values of hidden capacity. The bar charts (Fig. 2, b, d) reflects the change in the percentage of Red-, Green, Blue-triads of triplets in the matrix of unique colors of stego compared to unfilled containers (hidden capacity 0 bpp).

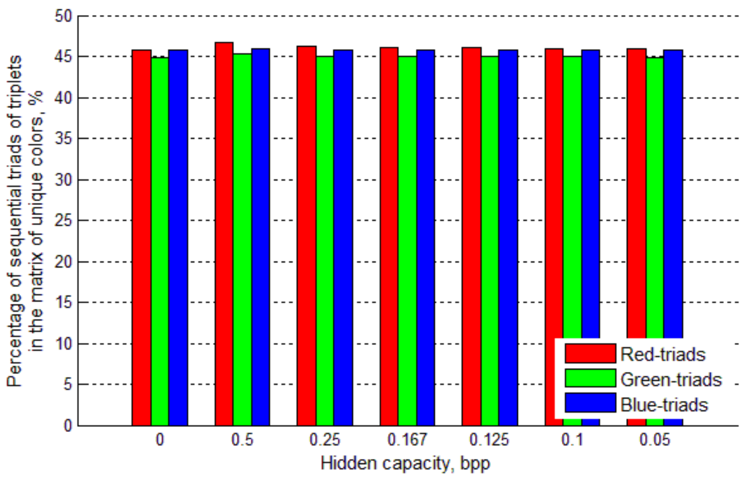
Table 1.

The comparative characteristic of quantity of sequential triads of triplets in the matrix of unique colors of original digital images and stego, formed with varying degrees of filling

		TIFF	JPG
Unfilled containers		High percentage (≈20-70%) of sequential triads of triplets, quantity of Red-, Green-, Blue-triads is comparable by values	Low percentage (up to 5%) of sequential triads of triplets, quantity of Red-, Green-, Blue-triads is comparable by values
Stego, formed by embedding of additional information into:	one color component	Quantity of Red-, Green-, Blue-triads is comparable by values and numerically practically does not differs from containers (not more than 1-2%)	Quantity of Red-, Green-, Blue-triads of triplets unevenly increases in comparison with a container (basic triads of ≈20-50%, concomitant triads – up to 15%)
	two color components	Quantity of Red-, Green-, Blue-triads is comparable by values and numerically practically does not differs from containers (not more than 1-2%)	Quantity of Red-, Green-, Blue-triads of triplets unevenly increases in comparison with a container (basic triads of ≈5-50%, concomitant triads – up to 35%)
	three color components	Quantity of Red-, Green-, Blue-triads is comparable by values and numerically practically does not differs from containers (not more than 1-2%)	Quantity of Red-, Green-, Blue-triads of triplets unevenly increases in comparison with a container (all triads are basic, their percentage varies from 5 to 50%)

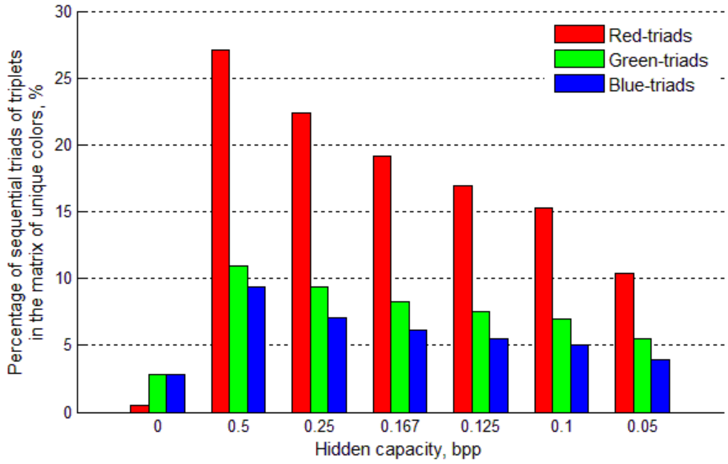


a



b

c



d

Fig. 2. The quantity of sequential triads of triplets in the matrix of unique colors of digital images: a – digital image stored in lossless format from Set 1; b – quantity of triads in the container and stego, formed by embedding of additional information into red color component of digital image from Set 1; c – digital image stored in losses format from Set 4; d – quantity of triads in the container and stego, formed by embedding of additional information into red color component of digital image from Set 4

From Table 1 and Fig. 2 implies that the embedding of additional information into containers initially stored in losses formats leads to unevenly increase of quantity of sequential triads of triplets, such stego are characterized by large differences of quantity values of sequential triads, even in cases when additional information was embedded into two or three color components. Stego, formed on the basis of containers in lossless formats, are characterized by stable, practically unchanged, high quantity of sequential triads of triplets regardless of the degree of filling, what is feature of digital images in a lossless format, which can be used to identify the format of the original digital content-container.

According to the results of the experiments we will formulate main steps of the method aimed to detect the fact of compression in digital images, which allows separate color images, stored in lossless formats initially, from digital imaged, resaved into lossless format from losses formats. For color digital image I of size $M \times N$ with color components R, G, B:

Step 1. Forming of the matrix UCT of size $U \times 3$ of unique colors (r_i, g_i, b_i), $i = \overline{1,U}$ for digital image I.

Step 2. Counting of Red-, Green-, Blue-triads.

2.1. *If* for the current triplet (r_i, g_i, b_i), $i = \overline{1,U}$ in the UCT at the same time there are triplets $(r_i + 1, g_i, b_i)$ and $(r_i - 1, g_i, b_i)$,
then $countR = countR + 1$, $countR$ – quantity of Red-triads in UCT;

2.2. *If* for the current triplet (r_i, g_i, b_i), $i = \overline{1,U}$ in the UCT at the same time there are triplets $(r_i, g_i + 1, b_i)$ and $(r_i, g_i - 1, b_i)$,
then $countG = countG + 1$, $countG$ – quantity of Green-triads in UCT;

2.3. *If* for the current triplet (r_i, g_i, b_i), $i = \overline{1,U}$ в UCT in the UCT at the same time there are triplets $(r_i, g_i, b_i + 1)$ and $(r_i, g_i, b_i - 1)$,
then $countB = countB + 1$, $countB$ – quantity of Blue-triads in UCT.

Step 3. To compute the percentage of middle triplets in relation to total number of unique colors (%):

$$pR = \frac{countR}{U} \cdot 100,\ pG = \frac{countG}{U} \cdot 100,\ pB = \frac{countB}{U} \cdot 100.$$

Step 4. Determining of format of original digital content:

If $((pR > T_{\lim}) \& (pG > T_{\lim}) \& (pB > T_{\lim}))$ when $pR \approx pG \approx pB$,

then original format of I is lossless format,

else original format of I is losses format,

where $T_{\lim}$ – threshold separating the digital content in a lossless format from digital content in a losses format. From Fig. 1, we can take the value $T_{\lim}$ from the range $[5,10]$.

The algorithm that implements the proposed method can be used as an integral part of steganalytic methods and independently. It should be noted that the method of detection the fact of compression in digital images is intended to detect the original format of only color digital content, for grayscale images its use is impossible.

Analyze efficiency of the developed method performing the analysis of spatial domain of digital contents by determination of type I errors (False Negative FN) – the pass of the fact of compression in case of its presence and type II errors (False Positive FP) – false determination of uncompressed digital content as undergone to compression.

In the computational experiment aimed at efficiency analysis of the proposed steganalytic algorithm, 400 digital images from Set 1-4 (200 images in losses format, resaved as lossless format, and 200 images in lossless format), of which ¼ of all the images is unfilled containers, ½ - stego formed by embedding of additional information into one arbitrary color component with different values of hidden capacity, ¼ - stego, formed by embedding of additional information into two or three color components with different values of hidden capacity. As a threshold $T_{\lim}$ separating digital content in lossless format from digital content in losses format, accept the $T_{\lim} = 7.5$, under condition $pR \approx pG \approx pB$ we will understand the case when $\max(pR, pG, pB) - \min(pR, pG, pB) \le 6$.

By results of experiment type I errors are $FN = 3\%$, type II errors are $FP = 4.5\%$, which allows to speak about high efficiency of the proposed method. It is noted that type I errors has been observed only in case of stego formed by embedding of additional information into all three color components of digital images initially stored in losses format. In the case when the fact of compression was detected in digital images initially stored in lossless format embedding of additional information into one arbitrary color component leads to similar perturbations of the matrix of unique colors as when using as containers digital images in losses format, that allows to determine the fact of presence/absence of additional information in such digital contents.

Conclusions

In this article the method of detection the fact of compression in color digital images analyzing spatial domain of digital contents, that allows avoid additional accumulation of a computing error, is proposed. The developed method is based on the accounting sequential color triads in the matrix of unique colors, and can be used to define the format of the original container in steganalysis.

As a results of computational experiments that conducted on the basis of both unfilled containers and stego formed under different conditions of embedding of additional information, the type I errors are 3%, the type II errors - 4.5%, indicating the high efficiency of this method.

The computational complexity of the algorithmic realization of proposed method of detection the fact of compression in digital image of size $M \times M$ is defined as $\underline{O}(M^4)$.

References

1. Bohme, R. Advanced statistical steganalysis / R. Bohme. – Springer, 2010. – 302 p.
2. Кобозева, А.А. Стеганоаналитеческий метод для цифровых контейнеров, хранящихся в формате без потерь / А.А. Кобозева, А.В. Ахмаметьева, А.А. Ефименко // Інформаційна безпека. – 2014. – №1(13). – С. 31-42.
3. Ахмаметьева, А.В. Усовершенствование стеганоаналитического метода, основанного на анализе пространственной области цифровых контейнеров / А.В. Ахмаметьева // Інформатика та математичні методи в моделюванні. – 2015. – Т.5. – № 4. – С. 367-375.
4. Ахмаметьєва, Г.В. Стеганоаналітичний алгоритм для цифрових контейнерів, збережених в форматах з втратами / Г.В. Ахмаметьєва // Сучасна спеціальна техніка. – 2016. – № 3(46). – С. 31-38.
5. Ахмаметьева, А.В. Стеганоанализ цифровых изображений, хранящихся в формате с потерями / А.В. Ахмаметьева // Захист інформації. – 2016. – Випуск 23. – С.135-145.
6. Ахмаметьєва, Г.В. Детектування стеганоповідомлень, вбудованих методом LSB Matching у цифрові контейнери в форматах з втратами / Г.В. Ахмаметьєва // V Міжнародна науково-технічна конференція «Захист інформації і безпека інформаційних систем», Львів, 2-3 червня 2016 р. – С.106-107.
7. Ахмаметьева, А.В. Выявление области применения стеганоаналитического подхода, основанного на анализе пространственной области цифровых контентов / А.В. Ахмаметьева // Problemele energeticii regionale [Проблемы региональной энергетики]. Электронный журнал Академии наук Республики Молдова. – 2016. – № 2 (31). – С.104-111.
8. Узун, И.А. Стеганоанализ цифровых изображений, хранящихся в произвольных форматах / И.А. Узун // Інформатика та математичні методи в моделюванні. – 2013. – Т. 3. – № 2. – С. 179-189.
9. Бобок, И.И. Выявление цифровых изображений, пересохраненных в формат без потерь из формата с потерями, как составная часть стеганоанализа / И.И. Бобок // Сучасна спеціальна техніка. – 2013. – №3. – С.64-70.
10. NRCS Photo Gallery: [Електронний ресурс] // United States Department of Agriculture. Washington, USA. Режим доступа: http://photogallery.nrcs.usda.gov (Дата обращения: 26.07.2012).

МЕТОД ВИЯВЛЕННЯ ФАКТУ СТИСКУ У ЦИФРОВИХ ЗОБРАЖЕННЯХ ЯК СКЛАДОВА ЧАСТИНА СТЕГАНОАНАЛІЗУ

Г.В. Ахмаметьєва

Одеський національний політехнічний університет,
просп. Шевченка, 1, Одеса, 65044, Україна; e-mail: anna-odessitka@mail.ru

Запропоновано новий метод виявлення факту стиску у кольорових цифрових зображеннях, що дозволяє відокремлювати цифрові зображення, початково збережені в форматі без втрат, від цифрових зображень, перезбережених з формату з втратами в формат без втрат, який аналізує просторову область цифрових контентів, що дозволяє уникнути додаткового накопичення обчислювальної похибки. Розроблений метод оснований на врахуванні послідовних колірних тріад триплетів у матриці унікальних кольорів та може бути використаний для визначення формату оригінального контейнеру при стеганоаналізі.

Ключові слова: цифрове зображення, стиск, формат з втратами, формат без втрат, просторова область, послідовні колірні тріади триплетів

МЕТОД ВЫЯВЛЕНИЯ ФАКТА СЖАТИЯ В ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ КАК СОСТАВНАЯ ЧАСТЬ СТЕГАНОАНАЛИЗА

А.В. Ахмаметьева

Одесский национальный политехнический университет,
просп. Шевченко, 1, Одесса, 65044, Украина; e-mail: anna-odessitka@mail.ru

Предложен новый метод выявления факта сжатия в цветных цифровых изображениях, позволяющий отделять цифровые изображения, изначально хранящиеся в формате без потерь, от цифровых изображений, пересохраненных из формата с потерями в формат без потерь, который анализирует пространственную область цифровых контентов, что позволяет избежать дополнительного накопления вычислительной погрешности. Разработанный метод основан на учете последовательных цветовых триад триплетов в матрице уникальных цветов, и может использоваться для определения формата оригинального контейнера в стеганоанализе.

Ключевые слова: цифровое изображение, сжатие, формат с потерями, формат без потерь, пространственная область, последовательные цветовые триады триплетов