

# Steganalysis of digital contents, based on the analysis of unique color triplets

**Anna Akhmametieva**

Department of Informatics and Management of Information Systems Security,  
Odessa National Polytechnic University, Ukraine  
[anna-odessitka@mail.ru](mailto:anna-odessitka@mail.ru)

*Submitted June 24, 2016 — Accepted December 22, 2016*

## **Abstract**

The new steganalytic algorithm for detection of the presence of additional information that embeds into digital images and digital videos by LSB Matching method with a small hidden capacity (not more than 0.5 bpp) is presented. The proposed steganalytic algorithm analyses digital content in the spatial domain and is based on the accounting of sequential color triads in the matrix of unique colors of the digital content. Steganalytic algorithm has a high effectiveness of detecting the additional information embedded into one arbitrary color component of the container with a small hidden capacity.

*Keywords:* Steganalysis, LSB Matching, the spatial domain of the container, a digital image, a digital video

*MSC:* 68U10, 94A08, 68P30

## **1. Introduction**

The rapid development of information and communication technologies leads to their wide distribution in the state, public and household sectors, it is possible easily and quickly to transfer any information to long distances. If in the state activities secure channels of communication applies, then such open channels as e-mail, social networks allows to exchange externally innocuous data, therefore they are often used with criminal intentions. Open access to the Internet and scientific resources allows you to track newest developments in the field of information security, steganography and steganalysis. The use of steganographic methods

and algorithms allows to transfer confidential information via open communication channels by hiding the fact of its presence in the transmitted content. In a competitive environment, in the conditions of the spread of terrorism the hidden communication can lead to significant losses for businesses and to the catastrophic consequences of terrorist attacks for the society in general.

To prevent the criminal acts with using steganography it is extremely important to develop steganalysis aimed at the detecting the fact of the presence/absence of hidden information in any digital content (see [1]). Digital images, audio or video sequences can be used as containers in steganography.

Ones of the most widespread steganographic methods are different variations of the method of modification of the least significant bit (LSB Matching, LSB Replacement etc.) due to the simplicity of realization and the possibility of its use both in the spatial domain and in the transformation domain. Nevertheless, continuous improvement of steganographic developments impedes using of LSB method with a high hidden capacity, because it is easy to detect such embedding. Therefore, to ensure concealing of the secret communication LSB method is often used with a small hidden capacity (less than 0.5 bpp), which greatly complicates the process of detecting the presence/absence of the additional information.

A large number of steganalysis developments aimed at the detection of the presence/absence of additional information, embedded by LSB Matching method into digital images. There are enough effective steganalysis methods and algorithms (see [2, 3, 4, 5, 6, 7, 8]), that analyse digital images in the transformation domain (frequency domain, the singular/spectral decompositions of the corresponding matrices, etc.), however transfer of a digital content to transformation domain and back leads to the additional accumulation of computational errors, which considerably complicates the process of detecting presence of additional information that embeds with a small hidden capacity.

The steganalytic methods that analyse spatial domain of digital contents allows to avoid both additional time expenses, and accumulation of computational errors, however existing developments (see [9, 10, 11, 12, 13]) often have low effectiveness when embedding of additional information is carried out by LSB Matching method with a small hidden capacity.

The use of a digital video as a container in steganography allows to transfer a significant amount of data due to the large number of frames, using at the same time a small hidden capacity. It is very problematic to detect the presence of additional information in these conditions. However, despite advantages of using of digital videos, in open access there are not a lot of works devoted to video steganalysis. Widely spread are methods that analyses spatial (see [14]) and temporal (see [15, 16]) domains of digital video, as well as specialized methods directed against steganographic tools such as embedding of additional information to the motion vectors (see [17]) or H.264/AVC (see [18]).

Taking into account the advantages of steganalysis in the spatial domain, the aim is to develop steganalytic algorithm aimed to detect embedding of additional information by LSB Matching with a small hidden capacity (no more than 0.5 bpp)

into digital containers, which are color digital images and digital videos.

## 2. Research essence

As containers we consider color digital images and digital videos stored according to the RGB color scheme. We will use term "cover" (cover-image or cover-video) for the unfilled containers. Each video sequence  $V$  consists of frames  $F_l(m, n)$ , where  $l = \overline{1, K}$ ,  $K$  - number of frames,  $m$  - frame height,  $n$  - frame width. Additional information, which is a binary sequence, embeds by LSB Matching into the spatial domain of the randomly selected color component of the container. Result of embedding of additional information to the container we will call stego (stego-image or stego-video). It should be noted that if as a container the image in a losses format is used, it will be resaved in a lossless format after embedding of additional information.

LSB Matching method is realized according to the formula (see [10]):

$$p_s(i, j) = \begin{cases} p_c(i, j) + 1 & \text{if } b \neq LSB(p_c(i, j)) \text{ \& } r > 0, \\ p_c(i, j) & \text{if } b = LSB(p_c(i, j)), \\ p_c(i, j) - 1 & \text{if } b \neq LSB(p_c(i, j)) \text{ \& } r < 0, \end{cases} \quad (2.1)$$

where  $p_c(i, j)$ ,  $p_s(i, j)$  - the brightness value of the pixel of the color matrix of the original digital image/frame and stego respectively,  $b$  - bit of the secret message,  $r$  - a random value in the range  $[-1, +1]$ ,  $LSB(p)$  - the least significant bit of  $p$  (see [10]). Thus, embedding of additional information or will increase pixel's brightness value of an original matrix on 1 (+1), or will reduce it on 1 (-1), or will remain it unchangeable (0).

Each frame of the video sequence  $V$  is an image formed by three color components: red, green and blue matrices of size  $m \times n$ . Accordingly, each pixel of the frame/image is represented as a triplet of values  $(R, G, B)$ . All triplets  $(r_i, g_i, b_i)$ ,  $i = \overline{1, k}$ ,  $k = m \cdot n$ , that occur in a digital image/frame of a digital video form some matrix  $CT$  (color triplets) of size  $k \times 3$ . Triplets can repeat in the matrix  $CT$ , depending on that how often they appear in a digital image/frame of a digital video.

**Definition 2.1.** All of the triplet's various values  $(R, G, B)$  we will call unique colors, their number is denoted by  $U$ .

**Definition 2.2.** A matrix of size  $U \times 3$  of ordered unique colors  $(r_j, g_j, b_j)$ ,  $j = \overline{1, U}$  we will call the matrix of unique colors  $UCT$  (unique color triplets).

The matrix of unique colors in MathWorks MatLAB can be received by standard procedure  $UCT = unique(CT, 'rows')$ , where parameter 'rows' denotes that the matrix  $UCT$  will contain unique rows (triplets) of matrix  $CT$ . Thus, the matrix  $UCT$  is an ordered sequence of unique colors which at least once occur in the analyzed digital image/frame of a digital video. The matrix of unique colors does

not take into account the frequency of appearance of some triplet  $(R, G, B) \subset UCT$  in the container.

Consider what changes the matrix of unique colors will be undergone when additional information embeds to the container by LSB Matching according to the formula 2.1.

Let to some pixel of the container (for example, a digital image) forming the triplet  $(95, 116, 68)$  additional information embeds into a green color component. This triplet will be included in the matrix of unique colors of the container as it at least once occurs in the digital image. Embedding of the bit of information into a green color component of the pixel can change it value or on  $(95, 115, 68)$ , or on  $(95, 117, 68)$ , or to leave it without change. Let the pixel value will take  $(95, 117, 68)$ . If the triplet is founded in the digital image only once, then matrix of unique colors of stego after embedding of additional information will not contain original triplet  $(95, 116, 68)$ , but its modification  $(95, 117, 68)$  will appear. However, many different pixels in the container can have the same color, i.e. the triplet occurs in the container as a rule repeatedly. Therefore, after embedding of additional information into different pixels of the same color, they can be modified by all three ways:  $+1, -1, 0$ , that will lead to appearance of all three modifications of triplets in the matrix of unique colors of stego (in this example,  $(95, 115, 68)$ ,  $(95, 116, 68)$  and  $(95, 117, 68)$ ).

Similar modifications occur when additional information embeds into red or blue color components.

Thus, in case of embedding of additional information into container in the matrix of unique colors there will be additional triplets differing from original on  $\pm 1$  in that color component where embedding was carried out.

Introduce the following definitions.

**Definition 2.3.** Under a sequential Red-triad for the current triplet  $(r_j, g_j, b_j)$ ,  $j = \overline{1, U}$ , in the matrix of unique colors we will understand execution of the condition:

$$(r_j, g_j, b_j) \subset UCT \text{ AND } (r_j - 1, g_j, b_j) \subset UCT \text{ AND } (r_j + 1, g_j, b_j) \subset UCT, \\ j = \overline{1, U}.$$

The Red-triad corresponds to the red color component of the container.

**Definition 2.4.** Under a sequential Green-triad for the current triplet  $(r_j, g_j, b_j)$ ,  $j = \overline{1, U}$ , in the matrix of unique colors we will understand execution of the condition:

$$(r_j, g_j, b_j) \subset UCT \text{ AND } (r_j, g_j - 1, b_j) \subset UCT \text{ AND } (r_j, g_j + 1, b_j) \subset UCT, \\ j = \overline{1, U}.$$

The Green-triad corresponds to the green color component of the container.

**Definition 2.5.** Under a sequential Blue-triad for the current triplet  $(r_j, g_j, b_j)$ ,  $j = \overline{1, U}$ , in the matrix of unique colors we will understand execution of the condition:

$$(r_j, g_j, b_j) \in UCT \text{ AND } (r_j, g_j, b_j - 1) \in UCT \text{ AND } (r_j, g_j, b_j + 1) \in UCT, \\ j = \overline{1, U}.$$

The Blue-triad corresponds to the blue color component of the container.

**Definition 2.6.** The basic triad is a sequential triad corresponding to that color component of the container, into which embedding of additional information was carried out.

**Definition 2.7.** Concomitant triads are that sequential triads that corresponds to unfilled color components of the container.

I.e. after embedding of additional information into blue color component the basic triad is Blue-triad and concomitant triads are Red- and Green-triads.

The computational experiment that analyses the quantity of Red-, Green- and Blue-triads in unfilled digital containers was carried out.

There are the following digital images as containers:

1. Set 1: 203 color digital images from [19] in JPG format;
2. Set 2: 201 high-quality digital images from [20] in JPG format;
3. Set 3: 215 images received by non-professional photo cameras in JPG format;
4. Set 4: 200 color digital images from [19] in TIFF format;
5. Set 5: 200 images received by non-professional photo cameras in TIFF format.

The quantity of sequential triads in the matrix of unique colors of unfilled digital containers stored in losses format (Set 1, 2, 3) does not exceed 3% of total number of unique colors, unlike containers stored in lossless format (Set 4, 5) where even in the absence of additional information the quantity of Red-, Green- and Blue-triads reaches 40-60% due to the lack of compression and, consequently, a large variety of unique colors. At the same time it is noted that the relative quantity of Red-, Green- and Blue-triads (relative to the number of unique colors) in unfilled digital images is comparable by values, i.e. the difference is not more than 1-1.5%.

Analyze how the number of sequential triads in the matrix of unique colors will change when additional information embeds into one arbitrary color component of the container with different values of hidden capacity. Consider as a container the digital image in losses format (Figure 1, a), into which additional information is embedded into a red color component by LSB Matching with different values of hidden capacity. Counting of Red-, Green- and Blue-triads in the matrix of unique colors of the formed stego was carried out. Counting of each kind of sequential triads dictated by the fact that in the process of steganalysis it is unknown, into what kind of color components additional information has been embedded. The percentage of Red-, Green- and Blue-triads in relation to the total number of unique colors of stego formed by embedding of additional information to the red color component with different values of hidden capacity is shown in Figure 1, b. Similarly,

the quantity of Red-, Green- and Blue-triads in matrices of unique colors of stegos formed by embedding of additional information only into green color component (Figure 1, c) and only into blue color component (Figure 1, d) is determined.

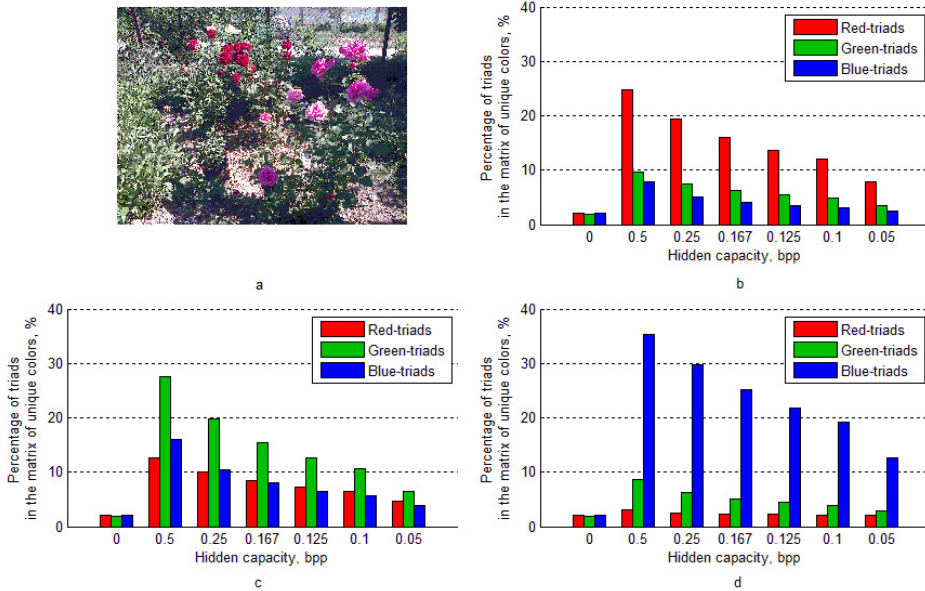


Figure 1: The quantity of sequential color triads in the image stored in losses format: a – original digital image from Set 3; b - percentage of sequential triads in the stego formed by embedding of additional information by LSB Matching into red color component; c - percentage of sequential triads in the stego formed by embedding of additional information by LSB Matching into green color component; d - percentage of sequential triads in the stego formed by embedding of additional information by LSB Matching into blue color component

As can be seen from Figure 1, embedding of additional information into one arbitrary color component causes a significant increase in the quantity of basic triads. At the same time the number of concomitant triads increases too, however less in comparison with the quantity of basic triads in the matrix of unique color of stego. This growth is associated with an increase in the number of unique colors when additional information has been embedded and is random. For example, embedding of additional information into blue color component leads to change of unique triplet (10, 158, 74) on Blue-triad (10, 158, 73), (10, 158, 74) and (10, 158, 75). If there are triplets (11, 158, 75) and (12, 158, 75) in the matrix of unique colors of stego then they with a new triplet (10, 158, 75) forms Red-triad, but embedding of additional information into red color component has not performed.

If as the container to use the image in lossless format then embedding of additional information practically does not influence the percentage of consecutive triads, what is shown in Figure 2, where into image in a lossless format (Figure 2, a) additional information embeds into a red color component (for example), the percentage of Red-, Green- and Blue-triads for which is shown in Figure 2, b. The similar situation is observed when additional information embeds into another color component and is typical for all digital images in a lossless format. Thus, further as containers only digital contents in losses format will consider.

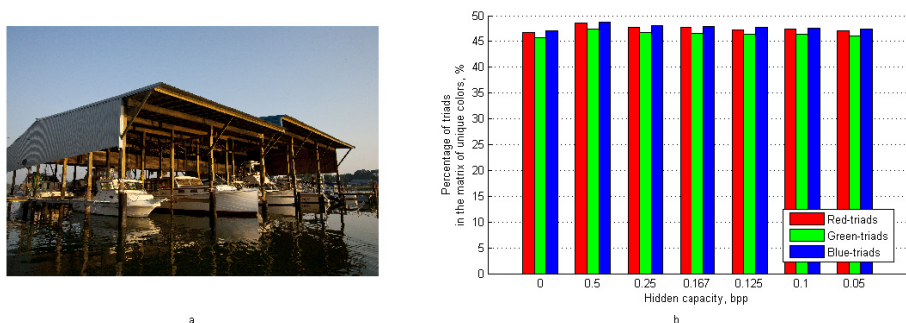


Figure 2: The quantity of sequential color triads in the image stored in lossless format: a – original digital image from Set 4; b - percentage of sequential triads in the stego formed by embedding of additional information by LSB Matching into red color component

Figure 1 shows that the percentage of concomitant triads in the matrix of unique colors of stego increases on average up to 8-12%. Accept preliminary thresholds of the quantity of sequential triads  $T_{low} = 2.5$  and  $T_{up} = 8$  for detection of the presence/absence of additional information in the digital content. On the basis of digital images from the Set 1, 2, 3 the computational experiment, determining percentage of Red-, Green- and Blue-triads in matrices of unique colors of original containers (hidden capacity 0 bpp) and stegos formed by embedding of additional information only into one arbitrary color component with different values of hidden capacity was carried out. Results of the experiment are shown in Table 1, where  $pc$  – the quantity (in %) of all sequential triads (and basic, and concomitant) in matrices of unique colors of digital contents,  $max(pc)$  – maximum percentage of sequential triads in matrices of unique colors of digital contents (for hidden capacity 0.05-0.5 bpp  $max(pc)$  corresponds to the maximum containing of basic triads in stego sets).

Thus, the computational experiment showed that an average of 90% of digital containers stored in losses format, contain no more than 2.5% Red-, Green- and Blue-triads, their number significantly increases after embedding of additional information. Therefore, the condition for the original container is performance of the relationship:

Set	Threshold	Hidden capacity, pbb						
		0.5	0.25	0.167	0.125	0.1	0.05	0
Set 1	$pc \leq T_{low}$	3.61	6.24	8.21	8.70	9.20	13.30	<b>94.42</b>
	$pc > T_{up}$	<b>89.66</b>	<b>80.79</b>	<b>72.58</b>	<b>64.37</b>	<b>56.49</b>	<b>35.80</b>	0.33
	$max(pc)$	56.94	47.44	47.68	45.33	43.22	31.36	9.08
Set 2	$pc \leq T_{low}$	3.15	5.31	6.30	8.62	11.61	18.24	<b>86.24</b>
	$pc > T_{up}$	<b>84.25</b>	<b>69.98</b>	<b>64.18</b>	<b>55.89</b>	<b>53.57</b>	<b>31.01</b>	0.83
	$max(pc)$	54.58	51.67	48.76	46.21	41.47	36.02	11.25
Set 3	$pc \leq T_{low}$	1.40	4.65	5.89	6.82	7.60	9.92	<b>92.25</b>
	$pc > T_{up}$	<b>90.70</b>	<b>85.74</b>	<b>77.36</b>	<b>68.22</b>	<b>61.24</b>	<b>46.98</b>	0
	$max(pc)$	55.08	51.65	48.28	45.32	43.80	38.53	4.69

Table 1: The percentage of sequential triads in the matrix of unique colors of digital images

$$(pR \leq T_{low}) \text{ AND } (pG \leq T_{low}) \text{ AND } (pB \leq T_{low}),$$

and back again for stego:

$$(pR > T_{low}) \text{ OR } (pG > T_{low}) \text{ OR } (pB > T_{low}),$$

where  $T_{low} = 2.5$ ,  $pR$  - the percentage of Red-triads in the matrix of unique colors,  $pG$  - the percentage of Green-triads in the matrix of unique colors,  $pB$  - the percentage of Blue-triads in the matrix of unique colors.

Threshold  $T_{up} = 8$  promotes to correctly detection of unfilled containers, providing the protection against the appearance of "false alarms".

Based on the established features of the changes in the number of color triads in the matrix of unique colors the steganalytic algorithm for detecting embedding of additional information by LSB Matching into spatial domain of digital containers (digital images and digital videos) is proposed. If digital images are as containers, it is necessary only first step of the algorithm for detecting.

## 2.1. Designations employed in the algorithm

$F_l(m, n), l = \overline{1, K}$  - frame of the analyzed video sequence  $V$  consisting of  $K$  frames of size  $m \times n$ .

$resultF$  - a matrix of size  $K \times 3$ , containing the result of the detection on each frame of the video sequence (sequence of digital images). In the case of a single digital image the matrix  $resultF$  is a matrix of size  $1 \times 3$  and is the end result of the detection. The first column of the matrix corresponds to red color component, the second - to green color component, and the third - to blue color component. Value 1 of the matrix corresponds to the presence of additional information in the corresponding color component, 0 - to its absence.

$UCT$  - a matrix of unique colors of the frame  $F_l$  (image  $I$ ) of size  $U \times 3$ , containing unique triplets  $(r_j, g_j, b_j), j = \overline{1, U}$ .



$countR$ ,  $countG$ ,  $countB$  - number of sequential triads in the  $UCT$  for red, green and blue color components, respectively.

$pR$ ,  $pG$ ,  $pB$  - percentage of sequential triads in the  $UCT$  in relation to total number of unique colors for red, green, blue color components, respectively.

$kR_{pos}$ ,  $kG_{pos}$ ,  $kB_{pos}$  - number of positive definite frames as containing embedded additional information in red, green and blue color components of the frame, respectively.

$kR_{neg}$ ,  $kG_{neg}$ ,  $kB_{neg}$  - number of negative definite frames as not containing embedded additional information in red, green and blue color components of the frame, respectively.

## 2.2. Steganalytic algorithm

**Step 1** (for digital images and frames of video sequence). For each image  $I$  (frame  $F_l$ ,  $l = \overline{1, K}$ ) the detection of the presence of additional information performs.

1. Forming of the matrix of unique colors  $UCT$  of the digital image  $I$  / frame  $F_l$  of video sequence  $V$ .
2. Counting of Red-, Green-, Blue-triads.
  - (a) If for the current triplet  $(r_j, g_j, b_j)$ ,  $j = \overline{1, U}$  in the  $UCT$  at the same time there are triplets  $(r_j + 1, g_j, b_j)$  and  $(r_j - 1, g_j, b_j)$  then  $countR = countR + 1$ ;
  - (b) If for the current triplet  $(r_j, g_j, b_j)$ ,  $j = \overline{1, U}$  in the  $UCT$  at the same time there are triplets  $(r_j, g_j + 1, b_j)$  and  $(r_j, g_j - 1, b_j)$  then  $countG = countG + 1$ ;
  - (c) If for the current triplet  $(r_j, g_j, b_j)$ ,  $j = \overline{1, U}$  in the  $UCT$  at the same time there are triplets  $(r_j, g_j, b_j + 1)$  and  $(r_j, g_j, b_j - 1)$  then  $countB = countB + 1$ .
3. To compute:

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

4. Detection of the presence/absence of additional information in the digital image  $I$  / single frame  $F_l$ ,  $l = \overline{1, K}$ , of video sequence  $V$ .
  - (a) if  $(pR = \max(pR, pG, pB))$  AND  $(pR > T_{up})$   
then  $result_{F_l, 1} = 1$ ,  
else if  
 $(pR > T_{low}$  OR  $pG > T_{low}$  OR  $pB > T_{low})$  AND  $(pR > 1.5 \cdot pG$  AND  
 $pR > 1.5 \cdot pB)$   
then  $result_{F_l, 1} = 1$ ,  
else  $result_{F_l, 1} = 0$ ;

- (b) if  $(pG = \max(pR, pG, pB))$  AND  $(pG > T_{up})$   
 then  $resultF_{l,2} = 1$ ,  
 else if  
 $(pR > T_{low}$  OR  $pG > T_{low}$  OR  $pB > T_{low})$  AND  $(pG > 1.5 \cdot pR$  AND  
 $pG > 1.5 \cdot pB)$   
 then  $resultF_{l,2} = 1$ ,  
 else  $resultF_{l,2} = 0$ ;
- (c) if  $(pB = \max(pR, pG, pB))$  AND  $(pB > T_{up})$   
 then  $resultF_{l,3} = 1$ ,  
 else if  
 $(pR > T_{low}$  OR  $pG > T_{low}$  OR  $pB > T_{low})$  AND  $(pB > 1.5 \cdot pR$  AND  
 $pB > 1.5 \cdot pG)$   
 then  $resultF_{l,3} = 1$ ,  
 else  $resultF_{l,3} = 0$ .

**Step 2** (for video sequences). Counting of positive and negative detection results in frames of a video sequence separately for each color component in the matrix  $resultF$ :

1. if  $resultF_{l,1} = 1$   
 then  $kR_{pos} = kR_{pos} + 1$ ,  
 else  $kR_{neg} = kR_{neg} + 1$ ;
2. if  $resultF_{l,2} = 1$   
 then  $kG_{pos} = kG_{pos} + 1$ ,  
 else  $kG_{neg} = kG_{neg} + 1$ ;
3. if  $resultF_{l,3} = 1$   
 then  $kB_{pos} = kB_{pos} + 1$ ,  
 else  $kB_{neg} = kB_{neg} + 1$ .

**Step 3** (for video sequences). Detection of the presence/absence of additional information in a digital video:

1. if  $kR_{pos} \geq kR_{neg}$   
 then additional information contains in a red color component,  
 else additional information is absent in a red color component;
2. if  $kG_{pos} \geq kG_{neg}$   
 then additional information contains in a green color component,  
 else additional information is absent in a green color component;
3. if  $kB_{pos} \geq kB_{neg}$   
 then additional information contains in a blue color component,  
 else additional information is absent in a blue color component.

### 3. Results of the experiment

In the computational experiment aimed at verifying the work of the proposed steganalytic algorithm, digital contents from the Set 1, 2, 3 and 367 video sequences of frame size  $320 \times 240$  obtained by the mobile cameras (Set V) have been used. Each video contains in average 250 frames.

It should be noted that color photos and videos obtained by the cameras of mobile devices (Set 3 and Set V) are the most probable containers as the most widespread due to permanent presence smart phones, iPad or other mobile devices with itself. Original mobile videos are stored in a losses format and has the extension \*.3gp or \*.mp4. After embedding of additional information video are saved as uncompressed video in \*.avi format.

Embedding of additional information was carried out into a randomly selected color component of digital images and videos with different values of hidden capacity: 0.5 bpp, 0.25 bpp, 0.167 bpp, 0.125 bpp, 0.1 bpp, 0.05 bpp. When additional information embeds into video sequence the selected color component is constant for all frames. Such embedding is caused by the fact that, as a rule, in case of steganography data transmission one component is used as the container and its choice is part of the secret key.

By results of experiment type I errors (False Negative  $FN$ ) – the pass stego in case of its presence and type II errors (False Positive  $FP$ ) – false detection stego in case of its absence (Table 2) were received.

Set	Errors	Hidden capacity, bpp					
		0.5	0.25	0.167	0.125	0.1	0.05
Set 1	$FN$	0	0	0	0	0	1.9704
	$FP$	0	0	0	0	0	0
Set 2	$FN$	0	0	0.4975	0	2.9851	7.9602
	$FP$	0	0	0	0	0	0
Set 3	$FN$	0	0	0	0	0	2.3256
	$FP$	0	0	0	0	0	0
Set V	$FN$	0	0	0	0.2725	0.8174	18.2561
	$FP$	0	0	0	0	0	0

Table 2: The effectiveness of detecting the presence/absence of additional information in digital contents, %

Table 2 shows that errors in detecting the presence of embedding of additional information in a digital content are very small, even for hidden capacity 0.05 bpp, indicating the high effectiveness of the proposed steganalytic algorithm.

For comparing of the effectiveness of the developed steganalytic algorithm with other existing methods for detection of embedding of additional information by LSB Matching in digital images stored in a losses format, the ROC-analysis is used.

ROC-curve analysis method applied to the steganalysis lies in realization of the testing of the group of digital images, that includes both unfilled containers, and stego, and it is known what each image is. Among analyzed digital images in a test group it is necessary to identify the stego (class  $V_1$ ) and original containers (class  $V_2$ ) using the developed steganalytic algorithm, as a result of which a positive  $\delta = 1$  (stego) or negative  $\delta = 0$  (cover) decision is adopted. Test results can be presented as Table 3 (see [3]), where  $TP$  is the quantity of correctly identified stego,  $TN$  is the quantity of correctly identified unfilled containers.

True state of digital image	Result of detection	
	$\delta = 1$ - stego	$\delta = 0$ - cover
class $V_1$ - stego	$TP$	$FN$
class $V_2$ - cover	$FP$	$TN$

Table 3: Results of detection of the test group of digital images

Obtained detection results are presented in the two-dimensional ROC-space, where the X-axis represents the specificity values that characterize the type II errors:

$$Sp = \frac{TN}{TN + FP},$$

and the Y-axis represents the sensitivity, characterizing the type I errors:

$$Se = \frac{TP}{TP + FN}.$$

In this study values  $Sp$  and  $Se$  are defined for different values of parameter  $T_{low}$ , based on which ROC-curves for values of hidden capacity 0.5, 0.25, 0.167, 0.125, 0.1 and 0.05 bpp shown in Figure 3, 4 are constructed. Figure 3 shows the ROC-space in the range  $1 - Sp \in \overline{0, 1}$ ,  $Se \in \overline{0, 1}$ , Figure 4 shows a fragment of ROC-space in the range  $1 - Sp \in \overline{0, 0.025}$ ,  $Se \in \overline{0.8, 1}$ .

Based on the constructed ROC-curves an integral parameter  $\rho$  characterizing the effectiveness of the studied steganalytic algorithm is obtained, where  $\rho = 2A - 1$ ,  $A$  - the area under the ROC-curve. The values of area  $A$  and parameter  $\rho$  for different values of hidden capacity are given in Table 4.

Comparison of the effectiveness of the steganalytic algorithm based on the analysis of color triads with other modern analogues for digital images (Ker's (see [7]), Liu's (see [8]), HGE, NDH COM, RLH COM, Fused feature, Joint feature set (see [9]), SAVV (see [3])) is carried out by comparing the integral parameters  $\rho$  for the corresponding values of hidden capacity. Visual comparison of the above methods is shown in Figure 5 (see [3]).

As can be seen from Figure 5, the results of detecting the presence/absence of embedding of additional information by LSB Matching in digital images are far superior previous solutions (see [3, 7, 8, 9]), especially in the case of a small hidden capacity (0.25 bpp or less). The most revealing is the comparison of the

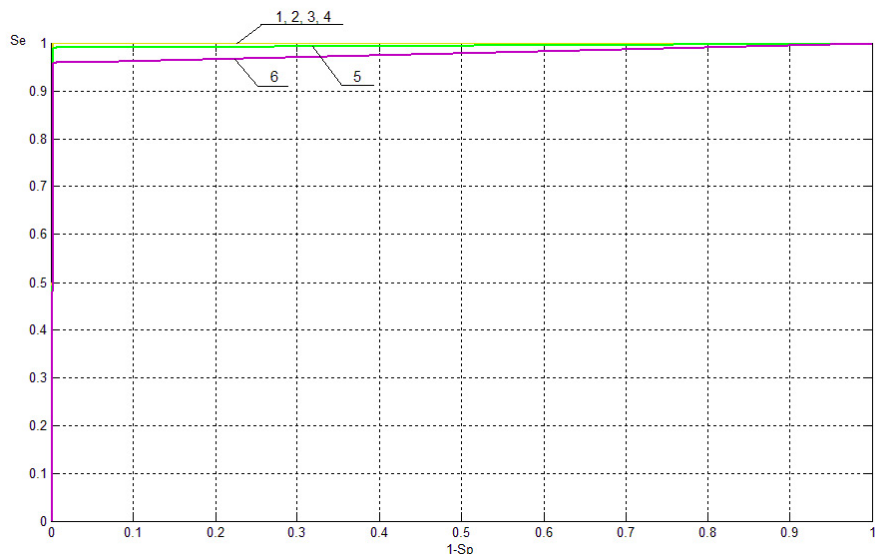


Figure 3: ROC-curves that characterizes the work of the steganalytic algorithm for detection of the presence of additional information embedded by LSB Matching into digital images with different values of hidden capacity: 1 - 0.5 bpp; 2 - 0.25 bpp; 3 - 0.167 bpp; 4 - 0.125 bpp; 5 - 0.1 bpp; 6 - 0.05 bpp

Hidden capacity, bpp	$A$	$\rho$
0.5	0.99822294	0.99644588
0.25	0.99822294	0.99644588
0.167	0.997393263	0.994786526
0.125	0.99822294	0.99644588
0.1	0.993238873	0.986477747
0.05	0.977475004	0.954950008

Table 4: Values of the integral parameter  $\rho$  for evaluating the effectiveness of the steganalytic algorithm for detection of the presence of additional information embedded by LSB Matching

developed algorithm with method SAVV, analyzing digital images with a small hidden capacity.

In modern papers devoted to steganalysis of digital videos [14-17] in computational experiments the methods of embedding of additional information that are different from LSB Matching, are used, that does not allow correctly to compare the effectiveness of the algorithm based on the accounting of color triads in the matrix of unique colors with other analogues. However, as seen from results of the computational experiment (Table 2, set V), the developed steganalytic algorithm

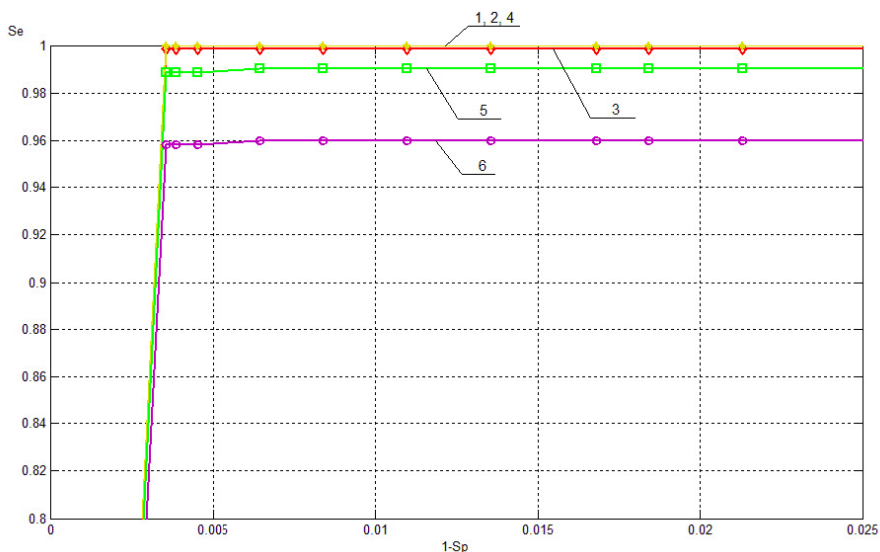


Figure 4: Fragment of ROC-curves in the ROC-space in the range  $1 - Sp \in [0, 0.025]$ ,  $Se \in [0.8, 1]$  that characterizes the work of the steganalytic algorithm for detection of the presence of additional information embedded by LSB Matching into digital images with different values of hidden capacity: 1 - 0.5 bpp; 2 - 0.25 bpp; 3 - 0.167 bpp; 4 - 0.125 bpp; 5 - 0.1 bpp; 6 - 0.05 bpp

is effective also in case of detection of the presence/absence of additional information in digital video sequences, at the same time both the frame analysis, and the analysis of digital video as a whole are possible.

## 4. Conclusions

In this paper a new steganalytic algorithm of detection of the presence/absence of additional information embedded by LSB Matching with a small hidden capacity (no more than 0.5 bpp) into one color component of digital images and digital videos stored in losses formats is proposed.

Comparison of the developed steganalytic algorithm with other modern tools of steganalysis showed that the algorithm based on the accounting of color triads in the matrix of unique colors is more effective than analogues, including for very small values of hidden capacity (0.1 and 0.05 bpp). The high effectiveness of the proposed algorithm for small values of hidden capacity is provided by its work in the spatial domain of digital contents, and as a result, there are no additional computational errors.

The proposed algorithm carries out an analysis of digital images and digital

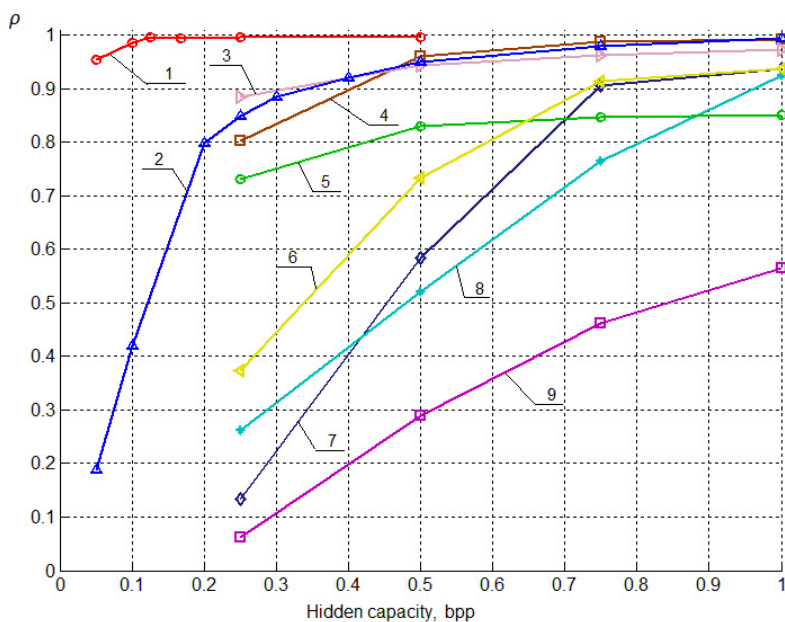


Figure 5: The results of comparing the effectiveness of the work of the steganalytic algorithm Color Triads with existing analogs for the detection of digital images: 1 - Color Triads, 2 - SAVV, 3 - Joint feature set, 4 - Liu's, 5 - RLH COM, 6 - Fused feature, 7 - Ker's, 8 - NDH COM, 9 - HGE

videos, for which its use can be expanded by analysis of single frames in the video sequence if additional information embeds not into all frames, but only into a small part of the total number of frames.

## References

- [1] BOHME, R., *Advanced statistical steganalysis*, Springer, 2010.
- [2] BOBOK, I.I., Steganalytic method for the digital signal-container stored in a losses format, *Modern Information Security*, Vol. 2 (2011), 50–60.
- [3] BOBOK, I.I., Application of ROC-analysis for integrated assessment of steganalysis method's efficiency, *Informatics and Mathematical Methods in Simulation*, Vol. 2, No. 3 (2012), 221–230.
- [4] ALIMORADI, D., The effect of correlogram properties on blind steganalysis in JPEG images, *Journal of computing and security*, Vol. 1, No. 1 (2014), 39–46.
- [5] VISAVALIA, S.R., GANATRA A., Improving blind image steganalysis using genetic algorithm and fusion technique, *Journal of computer science*, Vol. 1 (2014), 40–46.

- [6] YAMINI, B., SABITHA R., Blind steganalysis: to analyse the detection rate of stego images using different steganalytic techniques with support vector machine classifier, *International journal of computer applications*, No. 2 (2014), 22–25.
- [7] KER, A.D., Steganalysis of LSB matching in grayscale images, *IEEE Signal Processing Letters*, Vol. 12, No. 6 (2005), 441–444.
- [8] LIU, Q.Z., SUNG A.H., Image complexity and feature mining for steganalysis of least significant bit matching steganography, *Information Sciences*, Vol. 178, No. 1 (2008), 21–36.
- [9] ZHIHUA XIA, LINCONG YANG, A Learning-Based Steganalytic Method against LSB Matching Steganography, *Radioengineering*, Vol. 20, No. 1 (2011), 102–109.
- [10] JUN ZHANG, YUPING HU, ZHIBIN YUAN, Detection of LSB Matching steganography using the envelope of histogram, *Journal of computers*, Vol. 4, No. 7 (2009), 646–653.
- [11] GEETHA, S., SINDHU S., KAMARAJ N., Close color pair signature ensemble adaptive threshold based steganalysis for LSB embedding in digital images, *Transactions on Data Privacy*, Vol. 1, Iss. 3 (2008), 140–161.
- [12] MITRA, S., ROY T., MAZUMDAR D., SAHA A.B., Steganalysis of LSB Encoding in Uncompressed Images by Close Color Pair Analysis, *IIT Kanpur Hackers' Workshop 2004 (IITKHACK04)*, 23-24 Feb 2004, 23–24.
- [13] RUDNITSKIY, V., UZUN, I., Steganalysis algorithm for images that have been lossy compressed, *Information Security*, Vol. 15, No. 2 (2013), 122–127.
- [14] XIKAI XU, JING DONG, TIENIU TAN, Universal spatial feature set for video steganalysis, *Image Processing (ICIP), 19th IEEE International Conference*, Sept. 30 - Oct. 3 2012, 245–248.
- [15] PANKAJAKSHAN, V., HO, A.T.S., Improving video steganalysis using temporal correlation, *3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 26-28 Nov 2007, Kaohsiung, TAIWAN.
- [16] BUDHIA, U., KUNDUR, D., ZOURNTOS, T., Digital video Steganalysis exploiting Statistical Visibility in the Temporal domain, *IEEE Transactions on Information Forensics and Security*, Vol. 1, Iss. 4 (2006), 502–516.
- [17] TASDEMIR, K., KURUGOLLU, F., SEZER, S., Video steganalysis of LSB based motion vector steganography, *Visual Information Processing (EUVIP), 4th European Workshop*, 10-12 June 2013, 260–264.
- [18] SONGBIN LI, PENG LIU, QIONGXING DAI, XIUHUA MA, HAOJIANG DENG, Detection of Information Hiding by Modulating Intra Prediction Modes in H.264/AVC, *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE)*, 2013, 590-593.
- [19] NRCS Photo Gallery. Online: <http://photogallery.nrcs.usda.gov>.
- [20] WallpapersCraft. Online: <http://wallpaperscraft.ru/>.