Параметризуемый метод встраивания цифровых водяных знаков в младшие биты с адаптивным формированием ключа

А.Ю. Баврина[1,2], В.В. Мясников[1,2], Р.Р. Юзькив[2]

[1]Институт систем обработки изображений РАН – филиал ФНИЦ "Кристаллография и фотоника" РАН, Молодогвардейская 151, Самара, Россия, 443001
[2]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Предлагается параметризуемый метод встраивания цифровых водяных знаков в младшие биты с адаптивным формированием ключа. Метод использует заранее заданное число младших бит для встраивания цифровых водяных знаков особого класса. Для повышения качества сокрытия информации могут быть использованы несколько функциональных параметров (алгоритмов). Проведенные эксперименты по оценке стеганографической стойкости водяных знаков позволяют выработать рекомендации по выбору параметров для предлагаемого метода.

1. Введение

В современном мире большое внимание уделяется обеспечению безопасности цифровой информации. Рассматриваемые в статье системы защиты информации предполагают внедрение (встраивание) секретной (или защитной) информации в некий цифровой носитель (контейнер, покрывающий объект). Согласно наиболее авторитетным работам в этой области [1, 2], выделяют два независимых деления: по существованию связи между контейнером и секретной информацией (цифровые водяные знаки и нет) и по скрытности самого факта встраивания (стеганографические и нет). Несмотря на такое деление, все эти подвиды могут использовать схожие техники встраивания и один и тот же математический аппарат [1-6].

Наиболее популярным в настоящее время направлением являются цифровые водяные знаки (ЦВЗ). Основной целью встраивания ЦВЗ является подтверждение авторских прав и целостности самого документа. Кроме того ЦВЗ применяют и как способ размещения дополнительной цифровой информации, предназначенной для авторизованного использования.

При разработке методов встраивания необходимо анализировать возможные несанкционированные действия по обнаружению, модификации или извлечению скрытого сообщения, чем и занимается стегоанализ (стеганоанализ) [1-5, 7]. Действия (атаки) могут быть пассивными (выявление факта встраивания: визуальное или статистическое) и активные (извлечение или разрушение скрытого сообщения). По отношению к цифровым водяным знакам говорят об их стойкости по отношению к активным атакам и выделяют стойкие (робастные), полухрупкие и хрупкие ЦВЗ.

Теоретический аппарат построения стегосистем и ЦВЗ достаточно развит [1-6, 8]. Выделяют методы встраивания в пространственной и частотной областях. Методы встраивания в

пространственной области основаны на изменении непосредственно значений яркости (или цветовых составляющих) пиксела изображения-контейнера. К достоинствам этой группы методов относят простоту реализации (как правило) и незначительные искажения контейнера. Значимыми недостатками является плохая устойчивость к различным операциям над изображением и довольно большая вероятность обнаружения факта вложения статистическими методами.

Методы встраивания в частотной области изменяют спектр изображения-контейнера после некоторого преобразования (косинусное преобразование, преобразование Фурье, вейвлет-преобразование). Достоинствами этих методов является повышенная устойчивость к активным атакам, однако за это следует "платить" более высокой вычислительной сложностью.

Широко известной техникой, применяемой при встраивании в пространственной и частотной областях, является встраивании в наименее значимые биты (НЗБ). Методы встраивания в НЗБ в пространственной области базируются на том, что информация, содержащаяся в наименее значимых битах отсчетов изображения, носит шумовой характер и изменения в ней не будут заметны глазу человека. Несмотря на то, что предложено большое количество НЗБ-алгоритмов, эта область все еще нуждается в исследованиях, целью которых является разработка методов, сохраняющих вышеупомянутые достоинства методов встраивания в пространственной области и преодолевающих их недостатки.

В данной работе предложен метод встраивания ЦВЗ в младшие биты изображения в формате без сжатия с адаптивным формированием ключа. Предлагаемый метод обладает следующими преимуществами:

- устойчивость к визуальным и статистическим атакам;
- стойкость к искажениям, не изменяющим значения отсчетов изображения (поворот на углы, кратные 90°, масштабирование с использованием интерполяции по ближайшему соседу, обрезание краёв и вырезка фрагментов);
- большая вычислительная сложность извлечения встраиваемого изображения в случае, когда алгоритм встраивания известен, но не известен ключ.

Метод стеганографического встраивания ЦВЗ, предлагаемый в работе, может быть использован в различных областях: для подтверждения целостности изображения (в этом случае ЦВЗ может представлять "узорную сетку"), для размещения аннотаций к предметам на изображении (медицинские снимки) и других.

Статья организована следующим образом. Во втором разделе приводится описание предлагаемого метода. Третий раздел содержит перечисления параметров метода и их значений (алгоритмов), используемых в работе. Исследования стеганографической стойкости (пассивные и активные атаки, а также оценка максимальной вычислительной сложности попытки извлечения встраиваемого сообщения при известном алгоритме вложения) рассматривается в четвертом разделе. Заключение содержит выводы и направления дальнейших исследований.

2. Описание метода

Введем следующие обозначения. Пусть $C(n)$ – отсчеты контейнера (покрывающего объекта), $W(n)$ – отсчеты встраиваемого сообщения (ЦВЗ), $C^W(n)$ – отсчеты стегосигнала, $W^R(n)$ – отсчеты восстановленного встраиваемого сообщения, $K = \{K^j\}$ – стегоключ (секретный ключ).

В качестве контейнера будем рассматривать цифровые изображения, сохраненные в формате без сжатия, с целочисленными значениями отсчетов. То есть $C(n), C^W(n) \in \mathrm{Z}$, где Z – множество целых чисел, $n = \overline{0, M \cdot N - 1}$, где M, N – линейные размеры изображения.

В качестве встраиваемого сообщения будем рассматривать бинарное изображение того же размера $M \times N$ – $W(n), W^R(n) \in \{0,1\}$, содержащее "контурные" или "скелетные" элементы (будем считать, что единичным элементам соответствует контур, а нулевым – фон). Ключевыми особенностями рассматриваемого класса встраиваемых объектов являются:

- контурность – малое количество единичных отсчетов по сравнению с нулевыми, что в рамках предлагаемого метода дает потенциальную устойчивость к атакам определенного вида (визуальным, гистограммным);
- связность контура – все единичные отсчеты должны удовлетворять некоторому критерию связности (например, восьми-связности в локальном окне 3x3).

Предлагаемый метод основан на вложении информации в наименее значимые биты (НЗБ). Схема работы метода представлена на рисунке 1.

Рисунок 1. Схема вложения для метода встраивания ЦВЗ с адаптивным формированием ключа.

Вначале производится отбор отсчетов контейнера, подлежащих изменению согласно условию $\{C(n)\,|\,W(n)\neq 0\}$. Затем производится кластеризация значений b НЗБ отобранных отсчетов на определенное количество кластеров k. В результате кластеризации формируются центры кластеров $K=\left\{K^j\right\}$, $j=\overline{0,k-1}$, которые и являются стегоключом. Формирование стегосигнала производится путем замены (кодирования) b НЗБ отсчетов контейнера, подлежащих встраиванию, на соответствующие им центры кластеров (остальные отсчеты остаются неизмененными). При этом возникает ситуация, когда b НЗБ отсчетов контейнера, соответствующих нулевым значениям $W(n)$ (которые не нуждаются в изменении), оказываются равными значениям центров кластеров. Для обеспечения безошибочного извлечения b НЗБ таких отсчетов должны быть заменены на значения в окрестности центра кластера. Но данную замену производить не нужно, когда в окрестности текущего отсчета нет отсчетов $\{C(n)\,|\,W(n)\neq 0\}$, так как это будут единичные выбросы, которые можно легко устранить фильтрацией на этапе извлечения информации.

Формально схему вложения можно представить в следующем виде. Рассмотрим представление отсчетов в виде последовательности L битовых отсчетов:

$$C(n)=\sum_{i=0}^{L-1}C_i(n)\,2^i,\ C_i(n)\in\{0,1\}.$$

Тогда формирование стегосигнала при вложении в b НЗБ производим по формуле:

$$C^W(n)=\sum_{i=b}^{L-1}C_i(n)\,2^i+\begin{cases}K^j,\ если\ W(n)=1,\ где\ K^j-центр\ кластера\ для\ значения\ \sum_{i=0}^{b-1}C_i(n)\,2^i\\ R^j,\ если\ W(n)=0\ и\ \sum_{i=0}^{b-1}C_i(n)\,2^i=K^j\ и\ в\ окрестности\ W(n)\ нет\ 1-ных\ значений\\ \sum_{i=0}^{b-1}C_i(n)\,2^i,\ иначе\end{cases}.$$

Здесь R^j – значение в окрестности K^j ($R^j\neq K^l\ \forall\, l=0..k-1$).

Извлечение информации производится следующие образом:

$$W^R(n)=\begin{cases}1,\ если\ \sum_{i=0}^{b-1}C_i^W(n)\,2^i=K^j\\ 0,\ иначе\end{cases}.$$

После чего необходимо произвести фильтрацию единичных выбросов (единичные значения $W^R(n)$ заменяются на нулевые если окрестность не содержит больше единичных значений).

3. Параметры метода

Параметрами предлагаемого метода являются следующие величины (алгоритмы):

- число наименее значимых бит для вложения b ;
- число кластеров k ;
- алгоритм формирования стегоключа $K = \{K^j\}$, $j = \overline{0,k-1}$ (кластеризации значений b наименее значащих бит отсчетов контейнера, подлежащих изменению);
- способ выбора центра кластера K^j для конкретного отсчета, подлежащего встраиванию (кодирования отсчета);
- способ замены отсчетов контейнера, соответствующие нулевым значениям $W(n)$, равных центрам кластеров (способ поиска значения R^j в окрестности центра кластера K^j).

При исследовании стеганографической стойкости метода рассматривались следующие значения параметров (алгоритмы):

- $b = \{1,2,3\}$;
- значение k выбиралось в зависимости от объема встраиваемых данных $k = \left[2^{bc} \frac{U(W)}{M\,N} \right]$, где c – количество цветовых плоскостей контейнера, $U(W)$ – количество единичных отсчетов встраиваемого сообщения, $[\cdot]$ – целая часть числа
- рассматривались алгоритмы кластеризации: *maxmin* (алгоритм максиминного расстояния), *maxmin+kmeans* (комбинация алгоритмов максиминного расстояния и k-средних), *ratio* (претендент на центр кластера выбирается в зависимости от количества равных ему значений контейнера среди точек подлежащих и не подлежащих встраиванию)
- рассматривались следующие алгоритмы выбора центра кластера K^j для конкретного отсчета, подлежащего встраиванию: *nearest* (ближайшее к K^j), *probability* (случайный выбор в соответствии с расстояниями от текущего отсчета до всех центров кластеров), *random* (случайный выбор одного из K^j)
- отсчеты контейнера, не подлежащие встраиванию, но равные одному из центров кластера K^j заменялись на ближайшие значения

Опишем более подробно некоторые алгоритмы.

Алгоритм кластеризации *ratio*. Для каждого значения $l = \sum_{i=0}^{b-1} C_i(n) 2^i$ среди отсчетов контейнера, подлежащих встраиванию, вычисляется величина J_l. равная отношению количества таких точек, что $\sum_{i=0}^{b-1} C_i(n) 2^i = l \; u \; W(n) = 1$ к количеству точек $\sum_{i=0}^{b-1} C_i(n) 2^i = l \; u \; W(n) = 0$. Значения J_l сортируются по убыванию и берутся центы кластеров, соответствующие первым k наибольшим значениям J_l. Преимуществом данного алгоритма является то, мы хотим оставить неизменными точки, подлежащие встраиванию и при этом поменять минимальное количество точек, не подлежащих встраиванию (но равных значению ключа).

Алгоритм выбора центра кластера *probability*. Для текущего отсчета контейнера вычисляются расстояния $D^j(n)$ от его b НЗБ до всех центров кластеров K^j. Величины $1 \Big/ (D^j(n) \sum_{i=0}^{k-1} D^i(n))$ располагаются последовательно на единичном отрезке. Генерируется случайная величина $r \sim U[0,1]$, распределенная равномерно на отрезке $[0,1]$. Решение о выборе

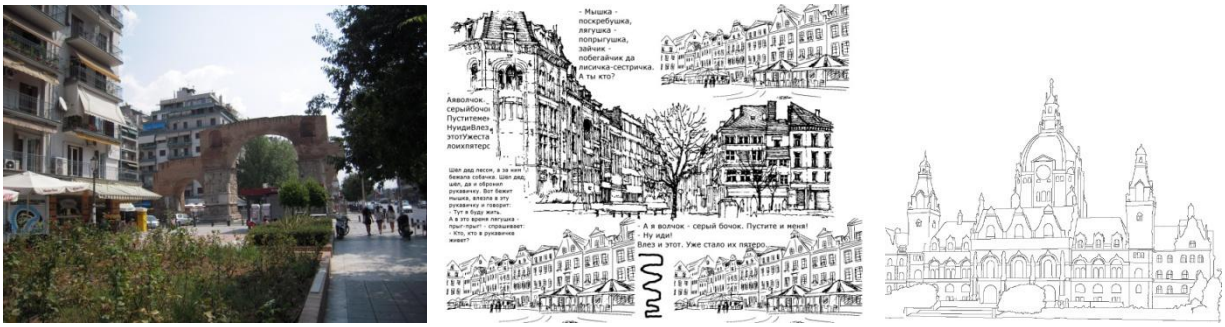
центра кластера для кодирования принимается в соответствии с тем, на отрезок, соответствующий какому K^j, попало значение r. Таким образом, с большей вероятностью выбирается ближайший центр кластера. Данный алгоритм позволяет имитировать шум в НЗБ, обычно присутствующий на изображениях, полученных с цифровых устройств регистрации.

4. Исследования стеганографической стойкости метода

Для исследования стеганографической стойкости рассматривались как пассивные атаки (визуальный, гистограммный анализ), так и активные (геометрические преобразования). Кроме того была оценена максимальная сложность попытки извлечения встраиваемого сообщения при известном алгоритме вложения.

В качестве контейнера рассматривалось RGB изображение, полученное с цифровой камеры, в формате BMP без сжатия (рисунок 2 (а)). Данное изображение содержит области с большой локальной дисперсией (растительность), области с плавным переходом яркости (здания), а также области с минимальным количеством шумов в младшем битовом слое (небо).

В качестве ЦВЗ было сгенерировано бинарное изображение, содержащее около 20% черных точек (единичных отсчетов) (рисунок 2(б)). Такой довольно большой процент вложения (в рамках предлагаемого метода) оправдан более выраженными визуальными эффектами и изменениями значений параметров при разного рода атаках. Следует заметить, что для практического применения встраиваемое изображение должно содержать гораздо меньшее количество единичных отсчетов (рисунок 2 (в)).



(а) (б) (в)

Рисунок 2. Изображения для исследований: (а) контейнер, (б) встраиваемое изображение с 20% единичных отсчетов, (в) встраиваемое изображение с 1,7% единичных отсчетов.

4.1. Пассивные атаки

В рамках проведения пассивных атак выполнялись следующие исследования:

- вычисление значения PSNR (Таблица 1) как численной характеристики отклонения контейнера $C(n)$ и стегосигнала $C^W(n)$

  $PSNR = 10\log_{10}\frac{255^2}{MSE(C,C^W)}$, где MSE – среднеквадратическое отклонение;

- вычисление значений SamplePairs [9] как оценки доли вложения (относительное количество пикселей изображения, подвергшихся вложению). Значение вычислялось с использованием пакета [10] (Таблица 2);
- визуальное исследование битовых плоскостей (Рисунок 3);
- визуальное исследование гистограммы (Рисунок 4);
- визуальное исследование по алгоритму PixelValue, основанному на анализе распределения значений яркостей пикселей в локальном окне, реализованный в пакете [11] (Рисунок 5).

Таблица 1. Значения PSNR (а) b=1, (б) b=2, (в) b=3.

а)

	nearest	probability	random
maxmin	**58,47**	57,59	57,42
maxmin+ k-nearest ratio	**58,47**	57,59	57,42

б)

	nearest	probability	random
maxmin	**58,26**	52,47	51,38
maxmin+ k-nearest ratio	**58,26**	52,47	51,38
	57,85	52,74	51,78

в)

	nearest	probability	random
maxmin	**58,31**	47,07	45,77
maxmin+ k-nearest ratio	**58,31**	47,07	45,76
	57,33	47,06	45,88

Таблица 2. Значения SamplePairs (значение для контейнера 0,09) (а) *b=1*, (б) *b=2*, (в) *b=3*.

а)

	nearest	probability	random
maxmin	**9,86**	14,18	14,68
maxmin+ k-nearest ratio	**9,86**	14,12	14,67
	9,86	14,15	14,66

б)

	nearest	probability	random
maxmin	5,03	12,21	14,35
maxmin+ k-nearest ratio	5,03	12,12	14,24
	4,51	7,70	9,86

в)

	nearest	probability	random
maxmin	2,16	6,64	8,44
maxmin+ k-nearest ratio	2,16	6,55	8,35
	1,66	5,78	7,04

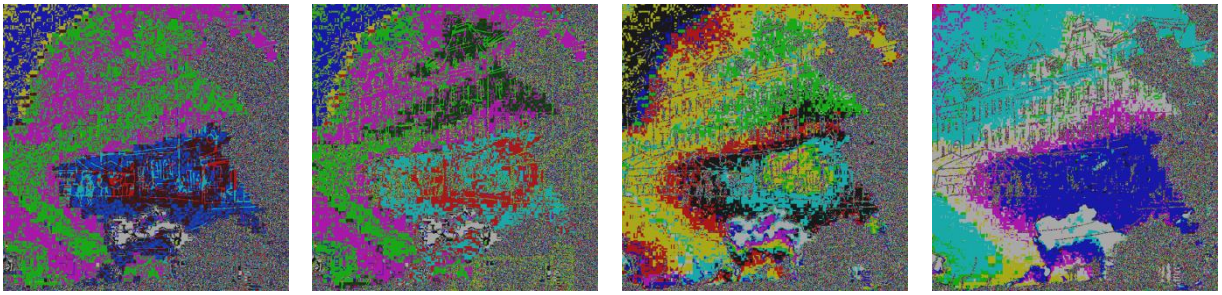
Таблица 1 показывает, что лучшие значения PSNR достигаются при алгоритме выбора центра кластера *nearest* и алгоритмах кластеризации *maxmin* и *maxmin+k-nearest*, близкие к ним значения при алгоритме кластеризации *ratio*. С увеличением значения *b* наилучшие значения PSNR почти не меняются.

Согласно таблице 2, лучшие значения SamplePairs достигаются при алгоритме выбора центра кластера *nearest* и алгоритмах кластеризации *ratio*. С увеличением значения *b* значения SamplePairs уменьшаются.

Следовательно, по рассматриваемым численным показателям стеганографической стойкости следует рекомендовать значения параметров *b=2, nearest, ratio*, или *b=3, nearest, ratio*.

Рассмотрим визуальные атаки. Так как анализируемый объём информации, относящийся к этой части, довольно велик, в статье приводятся показательные примеры изображений и общие выводы.

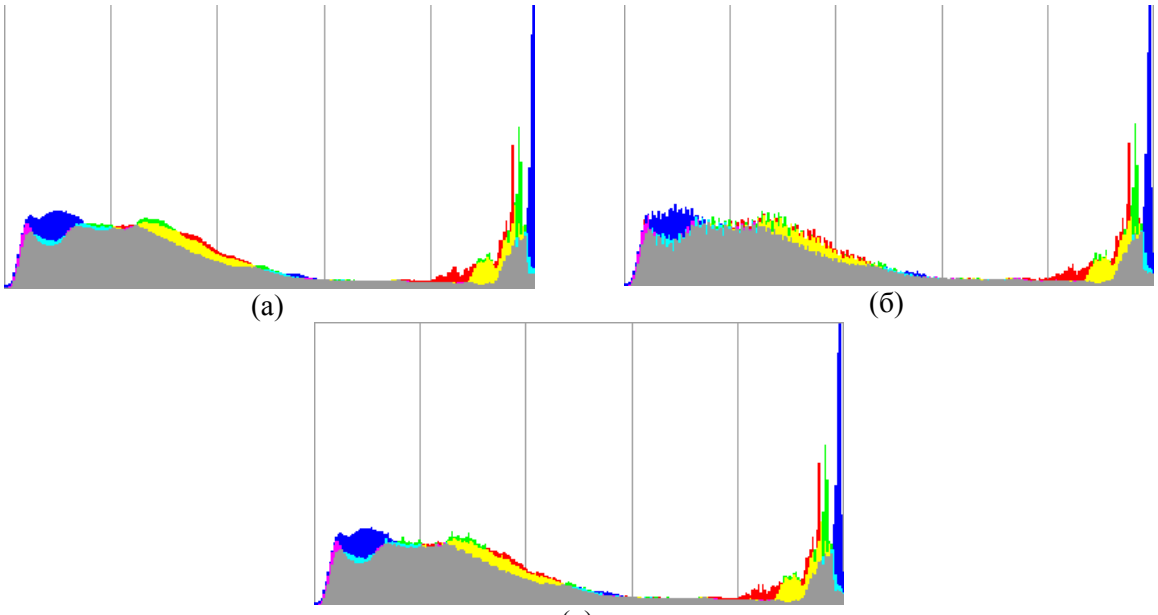
На рисунке 3 показаны некоторые артефакты, проявляющиеся при визуальном исследовании битовых плоскостей стегосигнала (нулевая битовая плоскость считается наименее значимой). Битовые плоскости визуализировались следующим образом: для каждой цветовой компоненты значение 0 отображается как 0, значение 1 отображается как 255. В результате анализа артефакты в той или иной степени были обнаружены при всех значениях параметров (хотя бы на одной из битовых плоскостей). Комбинации параметров при *b=2,3* и *nearest* дают не слишком выраженные артефакты, которые, возможно, будут практически незаметны для меньшей доли вложения.



(а) (б) (в) (г)

Рисунок 3. Артефакты при визуализации битовых плоскостей:
(а) плоскость № 0, *b=1, random, maxmin*, (б) плоскость № 0, *b=2, random, maxmin*,
(в) плоскость № 1, *b=3, random, maxmin+k-means*, (г) плоскость № 2, *b=3, random, ratio*.

На рисунке 4 показаны примеры гистограмм стегосигналов при различных значениях параметров. Анализ показывает, что при рассматриваемом проценте вложения (проценте единичных точек ЦВЗ) на всех гистограмма присутствует "гребенка" (характерные "зубцы") хотя бы для одной цветовой плоскости. Чем больше величина *b*, тем менее она выражена, для значения *nearest* "гребенка" выражена сильнее. При малом значении процента вложения (как на рисунке 2 (в)) и *b=3* "гребенка" практически незаметна.



(а) (б)

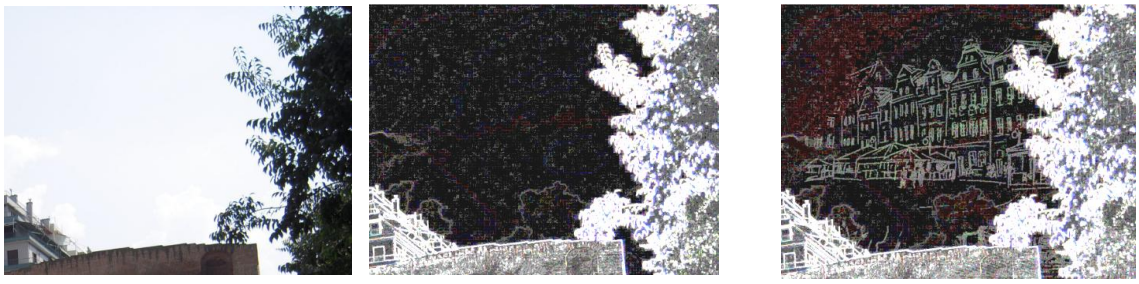
(в)

Рисунок 4. Анализируемые гистограммы: (а) для контейнера,
(б) для значений параметров *b=3, nearest, ratio,* процент вложения 20,
(в) для значений параметров *b=3, nearest, ratio,* процент вложения 1,7

На рисунке 5 показана визуальная атака PixelValue с использованием анализа распределения значений яркостей пикселей в локальном окне 3x3. Преобразование изображения *x* осуществляется по следующей формуле:

$$PixelValue(m,n) = \sum_{i,j=-1..1} \left| x_{m,n} - x_{m-i,n-j} \right| .$$

Вложение заметно на однородных областях контейнера при значениях *b=3, random, probability*.



(а) (б) (в)

Рисунок 5. Визуальная атака с использованием анализа распределения в локальном окне: (а) фрагмент изображения, (б) результат преобразования контейнера, (в) результат преобразования стегосигнала с параметрами *b=3, random, maxmin*.

По результатам визуальных атак можно рекомендовать значения параметров *b=3, nearest* (в сочетании с *maxmin*, *maxmin+k-nearest* и *ratio*).

4.2. Активные атаки

При встраивании ЦВЗ с помощью предлагаемого метода обеспечивается стойкость к искажениям, не изменяющим значения отсчетов изображения. После поворота на углы, кратные 90°, ЦВЗ восстановится без искажений. После масштабирования с использованием интерполяции по ближайшему соседу ЦВЗ будет также промасштабирован, однако возможно появление артефактов, связанных с фильтрацией единичных выбросов. После обрезания краёв и вырезания фрагментов в восстановленном ЦВЗ незатронутая часть останется без изменений (за исключением артефактов на краях, связанных с фильтрацией единичных выбросов).

4.3. Оценка сложности извлечения встраиваемого сообщения

Рассмотрим активную атаку, заключающуюся в попытке извлечения ЦВЗ, при условии, что известен алгоритм встраивания, но неизвестен стегоключ.

Уязвимым местом при данной атаке является связность встраиваемого ЦВЗ $W(n)$. При поиске значений параметров методом полного перебора некий критерий связности восстановленного вложения должен дать пик при верных значениях.

Пусть $V(M,N)$ – арифметическая сложность вычисления критерия связности восстановленного ЦВЗ при фиксированном значении стегоключа (и известном значении количества кластеров). Тогда сложность поиска верного значения стегоключа составляет $V(M,N)\cdot 2^{2bc}-2$ (напомним, что здесь b – количество бит для встраивания, c – количество цветовых плоскостей изображения).

5. Заключение

Метод встраивания цифровых водяных знаков в младшие биты, предлагаемый в статье, основывается на адаптивном формировании ключа, используемом для кодирования отсчетов контейнера, подлежащих встраиванию. Метод предполагает использование различных параметров, таких, как алгоритм кластеризации для формирования ключа, размер ключа, число используемых битовых плоскостей и несколько других, что позволяет производить адаптацию встраивания в зависимости от типа изображения-контейнера и ЦВЗ. Дальнейшие исследования могут быть направлены на рассмотрение дополнительных алгоритмов в рамках предлагаемого метода и исследования их эффективности для различных типов изображений.

6. Благодарности

Работа выполнена при поддержке РФФИ (гранты 19-29-09045 мк, 18-01-00748 а, 17-29-03190 офи_м), а также при поддержке Министерства науки и высшего образования РФ в рамках выполнения работ по Государственному заданию ФНИЦ «Кристаллография и фотоника» РАН (соглашение № 007-ГЗ/Ч3363/26).

7. Литература

[1] Cox, I.J. Digital Watermarking and Steganography / I.J. Cox, M.L. Miller, J.A. Bloom, J. Fridrich, T. Kalker – Elsevier, 2008. – 624 p.
[2] Fridrich, J. Steganography in digital media: principles, algorithms, and applications / Cambridge University Press, 2010. – 437 p.
[3] Коржик, В.И. Цифровая стеганография и цифровые водяные знаки. Часть 1 Цифровая стеганография – Санкт-Петербург: СПбГУТ, 2016. – 225 с.
[4] Коржик, В.И.Цифровая стеганография и цифровые водяные знаки. Часть 2 Цифровые водяные знаки – Санкт-Петербург: СПбГУТ, 2017. – 197 с.
[5] Грибунин, В.Г. Стеганографические системы. Атаки, пропускная способность каналов и оценка стойкости / В.Г. Грибунин, В.Е. Костюков, А.П. Мартынов, Д.Б. Николаев, В.Н. Фомченко – Саров: ФГУП "РФЯЦ-ВНИИЭФ", 2015. – 217 с.
[6] Федосеев, В.А. Унифицированная модель систем встраивания информации в цифровые сигналы // Компьютерная оптика. – 2016. – Т. 40, № 1. – С. 87-98. DOI: 10.18287/ 2412-6179-2016-40-1-87-98.
[7] Westfeld, A. Attacks on Steganographic Systems / A. Westfeld, A. Pfitzmann // Information Hiding. Lecture Notes in Computer Science. – 2000. – Vol. 1768.
[8] Митекин, В.А. Алгоритмы встраивания информации на основе QIM, стойкие к статистической атаке / В.А. Митекин, В.А. Федосеев // Компьютерная оптика. – 2018. – Т. 42, №1. – С. 118-127. DOI: 10.18287/2412-6179-2018-42-1-118-127.
[9] Dumitrescu, S. Detection of LSB Steganography via Sample Pair Analysis / S. Dumitrescu, X. Wu, Z. Wang // Information Hiding. Lecture Notes in Computer Science. – 2003. – Vol. 2578.
[10] Digital Invisible Ink Toolkit [Electronic resource]. – Access mode: http://diit.sourceforge.net/ (26.12.2019).
[11] Simple Steganalysis Suite [Electronic resource]. – Access mode: https://code.google.com/archive/p/simple-steganalysis-suite/ (26.12.2019).

Parameterizable LSB watermarking method with adaptive key generation

A.Yu. Bavrina[1,2], V.V. Myasnikov[1,2], R.R. Yuzkiv[2]

[1]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001
[2]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. The method of parameterizable digital watermarks LSBs-embedding with adaptive key generation is proposed. It uses predefine number of least significant bits for secure hiding of digital watermarks of special class. Several functional parameters (algorithms) can be used in the method to improve secure of hiding information. Experiments on estimation of watermarks detectable property are conducted to obtain the recommendations of parameters selection for the proposed method.