Полухрупкие водяные знаки для аутентификации изображений с возможностью восстановления, адаптированные к HGI-компрессии

А.Ю. Баврина [1,2], В.А. Федосеев [2, 1]
[1] ИСОИ РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН,
443001, Россия, г. Самара, ул. Молодогвардейская, д. 151;
[2] Самарский национальный исследовательский университет имени академика С.П. Королёва,
443086, Россия, г. Самара, Московское шоссе, д. 34

Аннотация

В статье предлагается новая полухрупкая система встраивания цифровых водяных знаков для защиты изображений, обладающая возможностью локализации несанкционированных искажений и восстановления искаженных фрагментов изображения. Система адаптирована к методу компрессии HGI и, подобно ему, использует иерархическую структуру изображения при встраивании информации. Отличие заключается в замене этапа квантования постинтерполяционных остатков, присутствующего в HGI, специальным квантователем, основанным на методе Quantization Index Modulation. В результате встроенный водяной знак становится устойчивым к компрессии HGI вплоть до заданного пользователем уровня искажений, вносимых при сжатии. Проведенные эксперименты демонстрируют необходимость выбора баланса между качеством защищенного изображения и точностью локализации возможных изменений. Помимо отыскания маски искажённых областей, предлагаемая система позволяет восстанавливать искажённые фрагменты с приемлемым качеством за счёт встраивания данных исходного изображения в качестве водяного знака на других иерархических уровнях. Разработанная система может использоваться для защиты изображений дистанционного зондирования и медицинских изображений от злонамеренных искажений.

Ключевые слова: цифровая обработка изображений, цифровые водяные знаки, компрессия изображений, метод иерархической сеточной интерполяции.

Цитирование: **Баврина, А.Ю.** Полухрупкие водяные знаки для аутентификации изображений с возможностью восстановления, адаптированные к HGI-компрессии / А.Ю. Баврина, В.А. Федосеев // Компьютерная оптика. – 2022. – Т. 46, № 1. – С. 103-112. – DOI: 10.18287/2412-6179-CO-1021.

Citation: Bavrina AY, Fedoseev VA. Semi-fragile watermarking with recovery capabilities for HGI compression method. Computer Optics 2022; 46(1): 103-112. DOI: 10.18287/2412-6179-CO-1021.

Введение

На текущий момент уровень развития методов и средств обработки изображений позволяет производить злонамеренные искажения данных, которые очень трудно распознать [1, 2]. Такие искажения недопустимы в стратегических и жизненно важных областях, таких как данные дистанционного зондирования (ДДЗ) и медицина [3, 4].

Изображения со спутников и дронов все чаще используются в различных сферах промышленности, сельского хозяйства, предотвращения стихийных бедствий, военной сфере и СМИ [5]. К способам защиты изображений от фальсификации относятся пассивные методы, основанные на компьютерном зрении и машинном обучении [6]. Другой способ – это использование активного подхода, а именно встраивание в изображение цифрового водяного знака (ЦВЗ) [7 – 9], который обладает свойством хрупкости по отношению к типовым искажениям, которые может привнести злоумышленник. Таким образом, разница между исходным ЦВЗ и водяным знаком, извлечённым из изображения на этапе аутентификации, может указывать на несанкционированные изменения. В данной статье мы предлагаем систему встраивания ЦВЗ, реализующую этот подход.

Как правило, для хранения ДДЗ требуется большой объём дискового пространства, и для его уменьшения используются методы сжатия данных. Следовательно, система водяных знаков, направленная на защиту ДДЗ, должна быть устойчивой к искажениям, вызванным сжатием.

В работе предлагается метод встраивания ЦВЗ, совместимый с методом компрессии изображений HGI (Hierarchical Grid Interpolation – иерархическая сеточная интерполяция). Этот метод показал высокую производительность в системах обработки ДДЗ за счёт возможности иерархического доступа к данным и контролируемой ошибки сжатия [10, 11]. В предлагаемой системе встраивание водяных знаков выполняется на этапе квантования постинтерполяционных остатков с использованием квантователя на

основе метода Quantization Index Modulation (QIM) [12]. Для решения задачи обнаружения локальных искажений предлагается алгоритм, позволяющий найти компромисс между величиной искажений, вносимых встраиванием, и точностью их локализации. Также предлагается алгоритм восстановления изображения в обнаруженных областях локальных изменений.

На сегодняшний день известно множество примеров систем ЦВЗ, позволяющих обнаруживать и устранять несанкционированные изменения [13 – 16]. Ключевые характеристики производительности таких систем включают незаметность водяного знака, качество обнаружения области искажений, качество восстановления исходных данных и устойчивость к различным атакам. Любой алгоритм встраивания ЦВЗ всегда находит баланс между этими категориями.

Примеры полухрупких систем ЦВЗ, адаптированных для различных форматов сжатия, также описаны в литературе. В частности, в обзоре [17] приводятся методы встраивания для формата JPEG, некоторые из которых поддерживают восстановление после изменений. В работе [18] приводится алгоритм встраивания ЦВЗ для одной из версий MPEG, а в [19] – алгоритм, адаптированный к формату JPEG 2000. Формат компрессии HGI не является столь же распространенным, однако ввиду его практической значимости для компрессии данных ДЗЗ задача разработки системы встраивания для этого формата является актуальной.

Ключевыми и уникальными особенностями предлагаемой системы ЦВЗ являются устойчивость к сжатию HGI в заданном диапазоне уровней сжатия и возможность контролировать соотношение между уровнем искажения и точностью обнаружения и восстановления. Кроме того, следует подчеркнуть, что встраивание информации осуществляется именно в несжатое изображение, после встраивания оно также остаётся несжатым. Таким образом, предложенный метод может использоваться и без применения HGI-сжатия, то есть для изображений, представленных в форматах TIFF, BMP, PNG.

Оставшаяся часть статьи организована следующим образом. В параграфе 1 представлена предлагаемая система водяных знаков. В параграфе 2 оцениваются искажения, вносимые встраиванием в изображение. Параграф 3 содержит исследование эффективности предлагаемой системы ЦВЗ в разных условиях и режимах использования.

1. Система ЦВЗ для метода компрессии HGI

Алгоритм иерархической сеточной интерполяции основан на представлении изображения $I(m,n)$ в виде объединения иерархических уровней.

$$I=\bigcup_{l=0}^{L-1}\boldsymbol{I}_l,$$

где $\boldsymbol{I}_{L-1}=\{I(2^{L-1}m,2^{L-1}n)\}$ – отсчёты старшего уровня, взятые с шагом 2^{L-1} по каждой координате. Следующее равенство определяет младшие уровни:

$$\boldsymbol{I}_l=\{I\left(2^l m,2^l n\right)\}\setminus\{I\left(2^{l+1}m,2^{l+1}n\right)\}.$$

Предлагаемая система встраивания использует иерархическое представление изображения и схему компрессии метода HGI с изменённым квантователем.

Пусть $I(m,n)\in[0,255]$ – исходное изображение (контейнер). $B(k)\in\{0,1\}$ – некоторая бинарная последовательность, полученная с помощью секретного ключа K (может выступать начальным значением для генератора псевдослучайных чисел), известного как при встраивании, так и при извлечении. Соответствие между отсчётами исходного изображения и последовательностью B устанавливается определенным отображением F, учитывающим иерархическую структуру. В результате отображения мы получаем следующую матрицу, которую и будем называть в дальнейшем цифровым водяным знаком:

$$W(m,n)=F(B)=\begin{cases}-1, & B(k)=0\\ 0, & \text{нет встраивания в } (m,n)\\ 1, & B(k)=1\end{cases}.$$

Во время встраивания на каждом иерархическом уровне сначала отсчёты текущего уровня интерполируются отсчётами более старшего уровня (была использована билинейная интерполяционная функция). Затем производится расчёт разницы между истинными значениями отсчётов и значениями, полученными путем интерполяции. Затем ошибки интерполяции квантуются квантователем на основе QIM Q_{qim} с использованием водяного знака W. Основным параметром квантователя QIM является ε_{qim}, определяющий устойчивость ЦВЗ и уровень искажений, вносимых встраиванием. Затем выполняется реконструкция отсчётов текущего уровня для использования их на более младших уровнях (обратное квантование постинтерполяционных остатков и суммирование их с результатами интерполяции).

В качестве квантователя QIM был выбран простейший квантователь семейства QIM [12]:

$$R_{qim}=Q_{qim}(R,W,\varepsilon_{qim})=$$
$$=\begin{cases}R_0, W=\{0,-1\}\\ R_0+\varepsilon_{qim}\cdot\operatorname{sign}(R-R_0), & W=1,\end{cases}$$

$$R_0=2\varepsilon_{qim}\times\operatorname{Round}\left(\frac{R}{2\varepsilon_{qim}}\right),$$

$$\operatorname{sign}(R-R_0)=\begin{cases}1, & R-R_0\ge 0\\ -1, & R-R_0<0\end{cases},$$

где R – значение постинтерполяционного остатка, R_{qim} – результат квантования (с одновременным встраиванием), ε_{qim} – половина шага квантования QIM.

Обратное квантование: $Q_{qim}^{-1}(R_{qim},\varepsilon_{qim}) = R_{qim}$.

Пусть I^W – изображение со встроенным ЦВЗ W. Величина искажений, вносимых встраиванием, оценивается с использованием критерия *PSNR*, который показывает пиковое отношение сигнал-шум между исходным изображением I и изображением со встроенным ЦВЗ I^W.

Изображение I^W может быть подвергнуто атаке, и принимающая сторона будет иметь изображение $\tilde{I}^W$, которое может отличаться от I^W. Различие ЦВЗ $\tilde{W}$, извлечённого из $\tilde{I}^W$, и исходного ЦВЗ W свидетельствует о факте несанкционированного изменения.

При извлечении ЦВЗ из изображения $\tilde{I}^W$ (возможно, изменённого в результате атаки) выполняются этапы, как при встраивании, но вместо этапа квантования выполняется восстановление значений ЦВЗ $\tilde{W}$:

$$\tilde{W}(m,n) = 2\cdot \mathrm{mod}\left(\mathrm{Round}\left(\frac{\tilde{R}_{qim}}{\varepsilon_{qim}} \right), 2 \right) - 1,$$

где $\tilde{R}_{qim}$ – значение постинтерполяционного остатка на этапе извлечения ЦВЗ.

Для регулирования доли отсчётов изображения, подвергаемых встраиванию, введём следующие параметры. Параметры $l_{\min}$ и $l_{\max}$ – номера минимального и максимального иерархических уровней соответственно, в которые производится встраивание. Параметр θ устанавливает процент отсчётов текущего уровня, которые подвергаются встраиванию.

2. Оценка величины искажений, вносимых встраиванием

Для оценки величины искажений, вносимых встраиванием, были проведены численные эксперименты на 10 изображениях в градациях серого из базы Waterloo Grayscale Set 1 и 2 [20]. Из каждого изображения были вырезаны фрагменты размера $M = N = 257$. Матрица W была сформирована следующим образом: нулевые значения для отсчётов, встраивание в которые не производится, и равновероятное появление значений 1 и –1 для остальных отсчётов.

Рис. 1 показывает зависимость $PSNR(I, I^W)$ от значения ε_{qim} (величины были усреднены по десяти исследуемым изображениям). Было использовано $L = 8$ иерархических уровней, $l_{\min} = 0$, $l_{\max} = 7$, $\theta = 100$. График показывает, что значение *PSNR* уменьшается с увеличением ε_{qim}.

Следующая часть исследований посвящена оценке искажений в результате встраивания, применённого к части иерархических уровней. В табл. 1 приведены усредненные значения *PSNR* для различных комбинаций $l_{\min}$ и $l_{\max}$ при $\varepsilon_{qim} = 20$. Анализ таблицы показывает, что чем больше отсчётов подвергается встраиванию, тем ниже становится значение *PSNR*. Минимального значения оно достигает при встраивании во все иерархические уровни ($l_{\min} = 0$, $l_{\max} = 7$) и максимального – при встраивании только в старший иерархический уровень ($l_{\min} = l_{\max} = 7$).

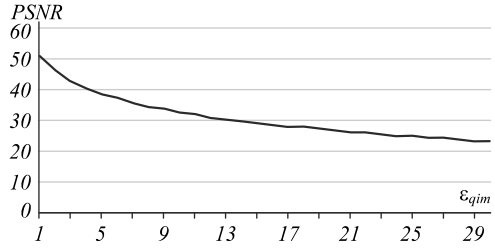


Рис. 1. Зависимость PSNR(I, I^W) от ε_{qim}

Табл. 1. Значения PSNR в случае встраивания ЦВЗ в иерархические уровни от l_{min} до l_{max} для $\varepsilon_{qim} = 20$

	***l*max**							
***l*min**	0	1	2	3	4	5	6	7
0	27,7	27,1	26,9	26,9	26,9	26,9	26,9	**26,9**
1		30,9	30,6	30,5	30,5	30,5	30,5	30,5
2			32,6	32,6	32,4	32,4	32,4	32,4
3				33,3	33,3	33,2	33,2	33,2
4					33,5	33,5	33,5	33,5
5						33,6	33,6	33,6
6							33,6	33,6
7								**33,6**

Ещё одним параметром, который влияет на величину искажений, является процент встраивания в каждый иерархический уровень θ. Рис. 2 показывает зависимость $PSNR(I, I^W)$ от θ при фиксированном $\varepsilon_{qim} = 20$. И снова, чем большее количество отсчётов используется для встраивания, тем меньше значение *PSNR*.

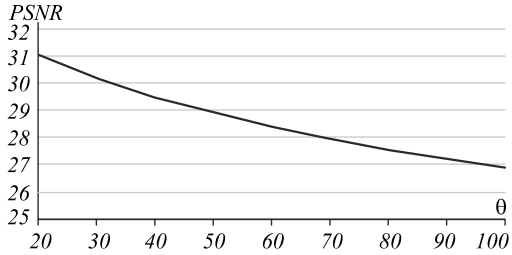


Рис. 2. Зависимость PSNR(I, I^W) от θ

Таким образом, можно сделать вывод, что степенью искажений изображения-контейнера, вызванных встраиванием, можно управлять с помощью параметров ε_{qim}, $l_{\min}$, $l_{\max}$ и θ.

Рис. 3 позволяет визуально оценить искажения, вносимые встраиванием ЦВЗ. На рис. 4 показаны гистограммы исходного изображения и защищённого изображения со встроенным ЦВЗ. Можно видеть, что в гистограмме защищённого изображения не просматривается никаких явных артефактов, которые могли бы указывать на существование водяного знака третьей стороне (например, характерная «гребёнка») [21].

3. Исследование работоспособности предлагаемой системы встраивания ЦВЗ в различных сценариях использования

В данном параграфе рассматриваются различные ситуации практического применения разработанной

системы встраивания ЦВЗ с целью исследования её работоспособности и эффективности.

3.1. Извлечение ЦВЗ после сжатия методом HGI

Как отмечалось во введении, HGI является эффективным методом сжатия ДДЗ. При этом его использование вносит некоторые искажения в данные (как и методы JPEG, JPEG 2000). Разумеется, эти искажения (до определённой величины погрешности) должны трактоваться как незначительные при аутентификации изображений. Проверим, обладает ли предложенная в параграфе 1 система таким свойством.

Рис. 3. Визуальные искажения, вносимые ЦВЗ: исходное изображение "Lena" (а); встраивание с ε_{qim}=10 (б); встраивание с ε_{qim}=20 (в)

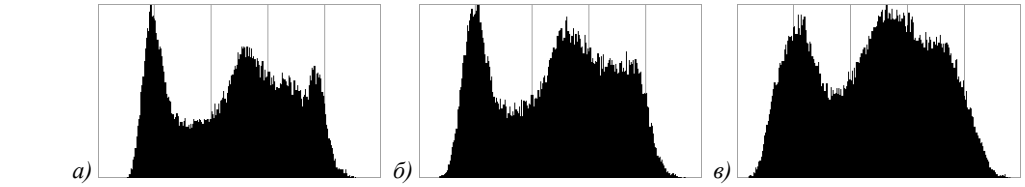


Рис. 4. Гистограмма исходного изображения и изображений с ЦВЗ: исходное изображение "Lena" (а); встраивание с ε_{qim}=10 (б); встраивание с ε_{qim}=20 (в)

Пусть изображение со встроенным ЦВЗ I^W подверглось HGI-компрессии с параметром ε_{hgi}, и $\tilde{I}^W$ – изображение, восстановленное после компрессии (из архива). На этапах квантования и восстановления в методе компрессии HGI были использованы следующие квантователь и деквантователь, обеспечивающие контроль максимальной ошибки восстановления (изображения после компрессии):

$$R_{hgi} = Q_{hgi}(R, \varepsilon_{hgi}) = \varepsilon_{hgi} \cdot \mathrm{Round}\left(\frac{R}{\varepsilon_{hgi}}\right),$$

$$Q^{-1}_{hgi}(R_{hgi}, \varepsilon_{hgi}) = R_{hgi},$$

где R_{hgi} – результат квантования постинтерполяционного остатка R, $\varepsilon_{hgi} = 2\varepsilon^{\max}_{hgi}$ – шаг квантования, равный удвоенному значению максимальной ошибки восстановления изображения после компрессии $\varepsilon^{\max}_{hgi}$.

Обозначим $\tilde{W}$ извлечённый из $\tilde{I}^W$ ЦВЗ (с использованием параметра ε_{qim}). В этом подпараграфе нас будет интересовать, как соотношение параметров HGI и QIM влияет на извлечение ЦВЗ, а также какие искажения в изображение вносит атака компрессией.

Для сравнения W и $\tilde{W}$ использовался критерий *BER* (Bit Error Rate). На рис. 5 показаны зависимости $BER(W, \tilde{W})$ и максимального отклонения $\varepsilon_{\max} = \max\{|I - \tilde{I}^W|\}$ от ε_{hgi} при фиксированном значении ε_{qim} (усреднённые для 10 исследуемых изображений).

Исследования показывают, что *BER* равен нулю для значений $\varepsilon_{hgi} \le \varepsilon_{qim}/2$, а также при $\varepsilon_{hgi} = \varepsilon_{qim}$. При $\varepsilon_{hgi} > \varepsilon_{qim}$ *BER* претерпевает резкий скачок и устанавливается на уровне около 0,5 (что говорит о разрушении ЦВЗ). При $\varepsilon_{hgi} = \varepsilon_{qim}$ ЦВЗ извлекается безошибочно, так как после встраивания все постинтерполяционные остатки кратны ε_{qim} и переквантование в HGI не изменяет их значений. Таким образом, предложенная система позволяет не учитывать при аутентификации искажения, вносимые процессом HGI-компрессии, пока $\varepsilon_{hgi} \le \varepsilon_{qim}/2$.

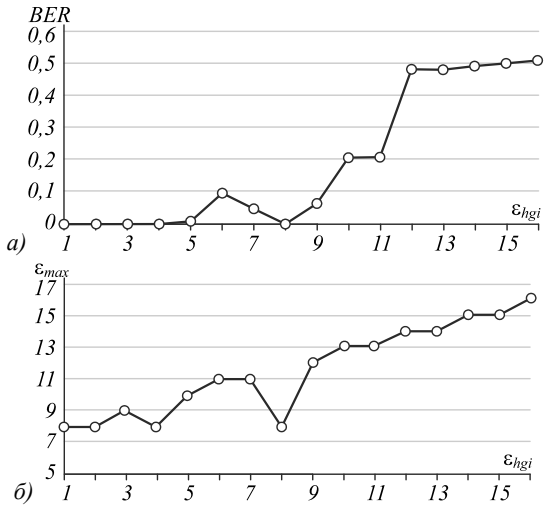


Рис. 5. Результаты атаки компрессией (ε_{qim} = 8): зависимость BER от ε_{hgi} (а); зависимость ε_{max} от ε_{hgi} (б)

Максимальное отклонение между исходным изображением и изображением, восстановленным после встраивания и компрессии, составляет:

$$\varepsilon_{\max} = \max\{|\, I - \tilde{I}^W \,|\} = \begin{cases} \varepsilon_{qim}, \varepsilon_{hgi} & \text{делитель } \varepsilon_{qim}, \\ \varepsilon_{qim} + [0{,}5\varepsilon_{hgi}], & \text{иначе.} \end{cases}$$

3.2. Реконструкция области локальных искажений

В этом подпараграфе анализируется точность определения области внесённых в изображение локальных искажений. Чтобы смоделировать локальные искажения, мы заменили отсчёты в пределах заранее определенной маски на случайные значения в диапазоне [0, 255], как показано на рис. 6. Обозначим результирующее изображение как $\tilde{I}^W$. Следует подчеркнуть, что столь заметные искажения были сделаны для упрощения визуализации результатов. Система же за счёт использования ЦВЗ позволяет обнаруживать и не различимые человеческим глазом искажения, и искажения, представляющие собой вставку фрагмента из другого места изображения (будет нарушена «истинная» последовательность бит в ЦВЗ, что обнаружит принимающая сторона).

На стороне получателя ЦВЗ $\tilde{W}$ извлекается из $\tilde{I}^W$ и сравнивается с исходным ЦВЗ W, который генерируется с использованием того же секретного ключа, что и при встраивании (с теми же значениями l_{min}, l_{max} и θ). Разница между $\tilde{W}$ и W не равна D (так как биты ЦВЗ восстанавливаются случайным образом безошибочно из примерно половины искажённых отсчётов) (рис. 6*г*).

При использовании полухрупких ЦВЗ пользователь всегда вынужден находить компромисс между стойкостью и неразличимостью ЦВЗ. В случае встраивания ЦВЗ в выборочные иерархические уровни и при $\theta < 100\,\%$ мы получаем разреженную структуру ненулевых значений разности между W и $\tilde{W}$. Для реконструкции области локальных искажений $\tilde{D}$ предлагается следующий алгоритм.

0. Изначально $\tilde{D}(m,n) = 0$ для всех (m,n).
1. На <u>первом шаге</u>, двигаясь от отсчётов старшего иерархического уровня к отсчётам уровня l_{min}, при $\tilde{W}(m,n) \neq W(m,n)$ и $W(m,n) \neq 0$ мы заполняем «1» окрестность $\tilde{D}(m \pm (2^l - 1), n \pm (2^l - 1))$, где l – номер текущего уровня. Эта область может «пострадать» от искажений отсчёта *(m, n)* при последовательной реконструкции отсчётов уровней младше *l*.
2. На <u>втором шаге</u> мы обрабатываем $\tilde{D}$ двумя последовательными морфологическими фильтрами (max-min) с размером окна $win = 2^{l_{\min}+2} + 1$.

Для исследования качества предложенного алгоритма реконструкции маски изменений были использованы следующие величины: q_{01} – относительное количество ложных обнаружений отсчётов области локальных изменений, q_{10} – относительное количество пропусков и их сумма q:

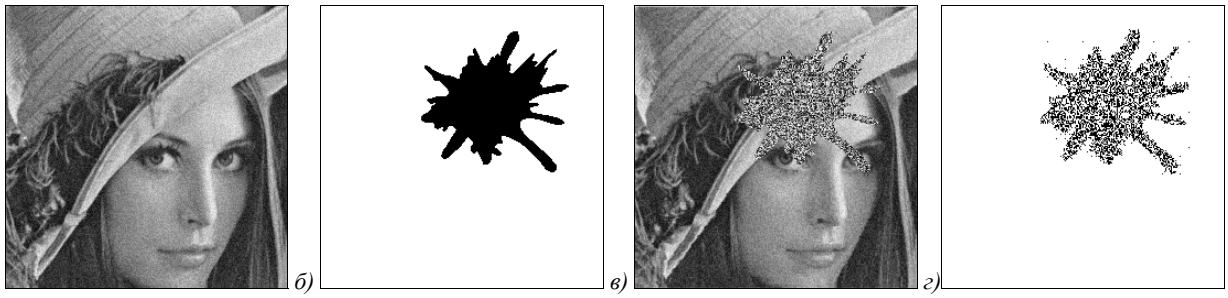
$$q_{01} = \frac{|\{D = 0, \tilde{D} = 1\}|}{|\{D = 1\}|},$$

$$q_{10} = \frac{|\{D = 1, \tilde{D} = 0\}|}{|\{D = 1\}|},$$

$$q = q_{01} + q_{10},$$

где $|\bullet|$ – мощность множества.

В качестве знаменателя используется размер области локальных изменений, чтобы значения критерия были сравнимы для областей большой и малой площади.



а) б) в) г)

Рис. 6. Локальное искажение изображения со встроенным ЦВЗ: изображение с ЦВЗ ($\varepsilon_{qim} = 20$, $l_{min} = 0$, $l_{max} = 7$, $\theta = 100\%$) (а); область локальных искажений D (б); искажённое изображение с ЦВЗ (в); разница между исходным и извлечённым ЦВЗ (г)

Следующая часть подпараграфа посвящена исследованию характера изменения значений этих величин для различных значений $l_{\min}$, $l_{\max}$ и θ. В этой части было использовано только изображение "Lena", а все представленные значения были усреднены для 100 наблюдений.

Рис. 7 показывает зависимость $PSNR(I, I^W)$ от θ для различных комбинаций $l_{\min}$, $l_{\max}$ и при фиксированном значении $\varepsilon_{qim} = 20$. График показывает, что для $l_{\min} = l_{\max} = 0$ кривая располагается слишком низко (искажения исходного изображения слишком очевидны). Встраивание ЦВЗ только во второй иерархический уровень обеспечивает лучший *PSNR*, но величина q для этого случая слишком высока. Таким образом, следует более детально рассмотреть случаи $l_{\min} = l_{\max} = 1$ и $l_{\min} = 1$, $l_{\max} = 2$.

На рис. 8 показаны кривые для q, q_{01}, q_{10} в зависимости от θ для $l_{\min} = l_{\max} = 1$ и $l_{\min} = 1$, $l_{\max} = 2$ при фиксированном значении $\varepsilon_{qim} = 20$. Сравнение может быть произведено следующим образом. Предположим, что

нас интересует значение относительного количества пропусков $q_{10}=0{,}05$. Тогда необходимо найти пересечение с кривой q_{10} и получить значения θ и q (обозначены тонкой сплошной линией). Таким образом, мы получим, что для $l_{\min}=l_{\max}=1$ значение критерия составляет $q=0{,}2$, а для $l_{\min}=1$, $l_{\max}=2$ значение критерия равно $q=0{,}3$. Следовательно, значения параметров $l_{\min}=l_{\max}=1$ и $\theta=70\div100\,\%$ могут быть рекомендованы для эффективного обнаружения локальных изменений.

На рис. 9 – 10 представлены некоторые результаты работы предложенного алгоритма реконструкции области локальных изменений для значений параметров, выделенных на рис. 8. Сравнение показывает лучшие результаты для рис. 9, чем для рис. 10.

3.3. Восстановление искажённых фрагментов изображения при помощи ЦВЗ

В данном подпараграфе описывается модификация базового метода встраивания, позволяющая осуществлять не только аутентификацию изображения, защищённого ЦВЗ, но и восстановление искажённых областей.

Процесс встраивания схематично изображён на рис. 11. Данные для аутентификации и для восстановления встраиваются на различных иерархических уровнях: генерация отсчётов ЦВЗ на уровне l_{auth} производится с использованием секретного ключа, а отсчёты ЦВЗ на уровне l_{rec} используются для хранения информации о восстановлении изображения.

Исходное изображение разделяется на непересекающиеся блоки. Размер блока N_b зависит от значений l_{auth} и l_{rec}. В наших экспериментах были выбраны значения $N_b=4$, $l_{auth}=1$, $l_{\text{rec}}=0$ в качестве компромисса между точностью обнаружения локальных искажений и качеством восстановления.

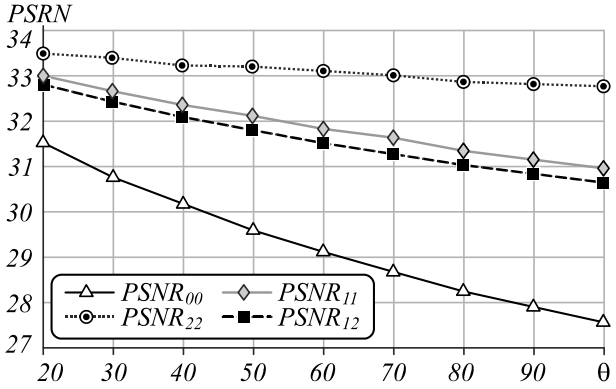


Рис. 7. Зависимость PSNR от θ для различных l_{min} и l_{max} ($PSNR_{00}$ – $l_{min}=l_{max}=0$, $PSNR_{11}$ – $l_{min}=l_{max}=1$, $PSNR_{22}$ – $l_{min}=l_{max}=2$, $PSNR_{12}$ – $l_{min}=1$, $l_{max}=2$)

Чтобы восстановить подделанный блок, его содержимое должно быть встроено в качестве ЦВЗ в другой блок, поэтому до процесса внедрения ЦВЗ выполняется перемешивание блоков.

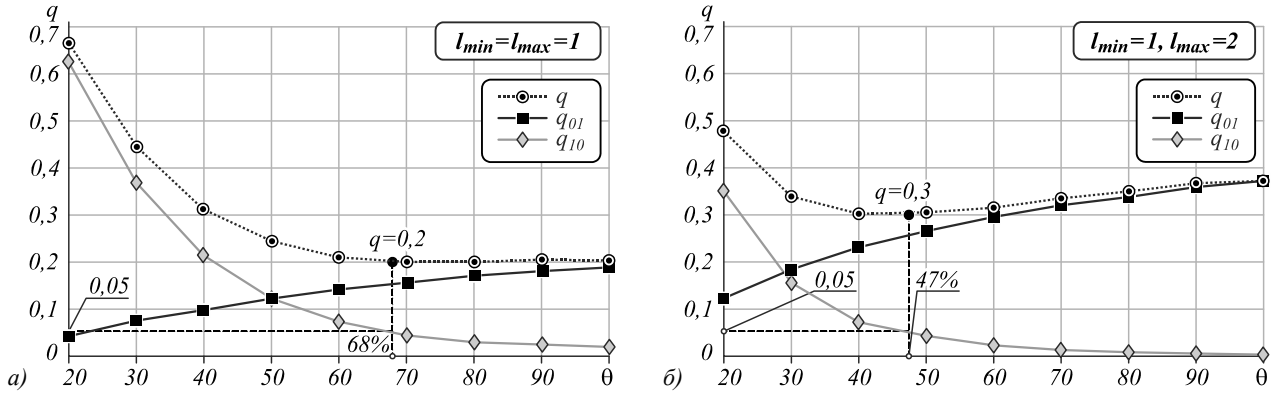


Рис. 8. Зависимость относительного количества ошибок восстановления области локальных искажений q от процента встраивания θ: $l_{min}=l_{max}=1$, круг соответствует $q=0{,}2$, $q_{10}=0{,}05$, $q_{01}=0{,}15$, $\theta=68\,\%$, PSNR = 31,65 (а); $l_{min}=1$, $l_{max}=2$, круг соответствует $q=0{,}3$, $q_{10}=0{,}05$, $q_{01}=0{,}25$, $\theta=47\,\%$, PSNR = 31,85 (б)

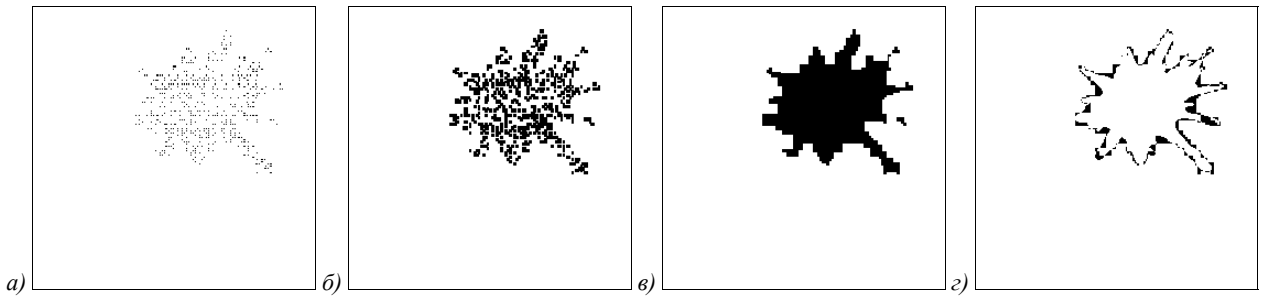


Рис. 9. Результаты работы алгоритма реконструкции области локальных изменений для $l_{min}=l_{max}=1$, $\theta=68\,\%$: разница между W и $\tilde{W}$ (а); результат шага 1 (б); результат шага 2 (в); ошибка реконструкции ($D\oplus\tilde{D}$) (г)

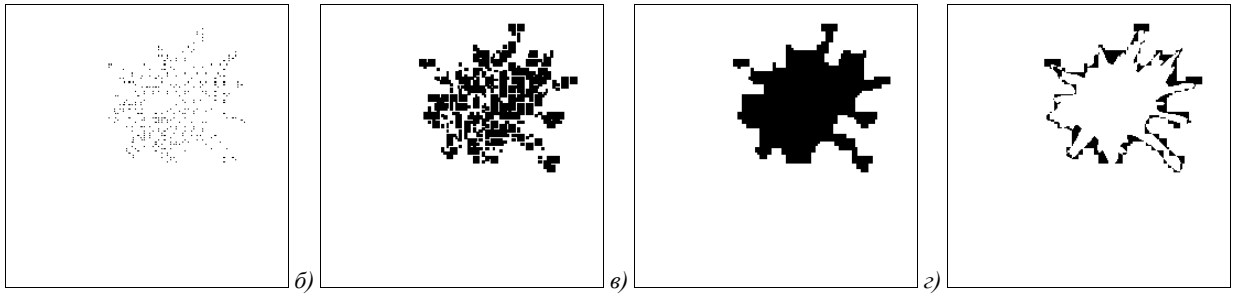


Рис. 10. Результаты работы алгоритма реконструкции области локальных изменений для $l_{min} = 1$, $l_{max} = 2$, $\theta = 47\,\%$: разница между W и $\tilde{W}$ (а); результат шага 1 (б); результат шага 2 (в); ошибка реконструкции ($D \oplus \tilde{D}$) (г)

На стороне получателя (рис. 12) водяной знак $\tilde{W}_{auth}$ извлекается из $\tilde{I}^W$ и сравнивается с водяным знаком W_{auth} (соответствующая часть W). При наличии несоответствий восстанавливается область локальных изменений $\tilde{D}$ с помощью алгоритма, описанного в подпараграфе 3.2. Для каждого изменённого блока изображения находится блок, содержащий информацию о нем (используется тот же алгоритм перемешивания блоков, что и при встраивании). Если найденный блок не был изменен, информация для восстановления извлекается и используется. Если для некоторого блока восстановление не может быть выполнено по причине искажения блоков, в которые встроена информация о нём, то данный блок восстанавливается приближённо по соседним блокам. Для этого выполняется интерполяция по блокам в окне размером 3×3 блока с использованием информации только из блоков, которые не были искажены или были успешно восстановлены.

На рис. 13 показано расположение отсчётов разных иерархических уровней внутри блока. Как видно из рисунка, этот блок содержит три отсчёта для аутентификации и 12 отсчётов для встраивания информации, используемой для восстановления.

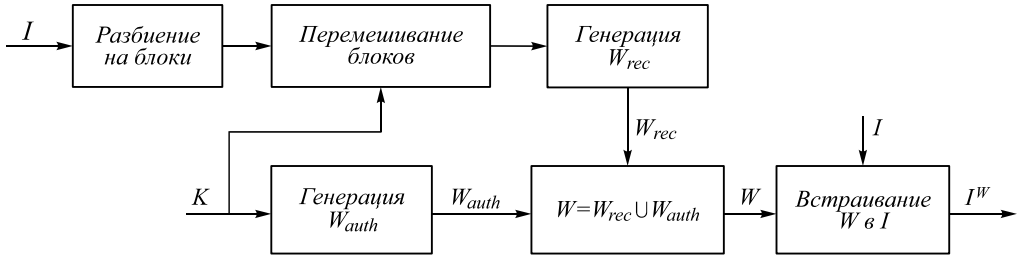


Рис. 11. Схема встраивания в изображение с возможностью аутентификации и восстановления

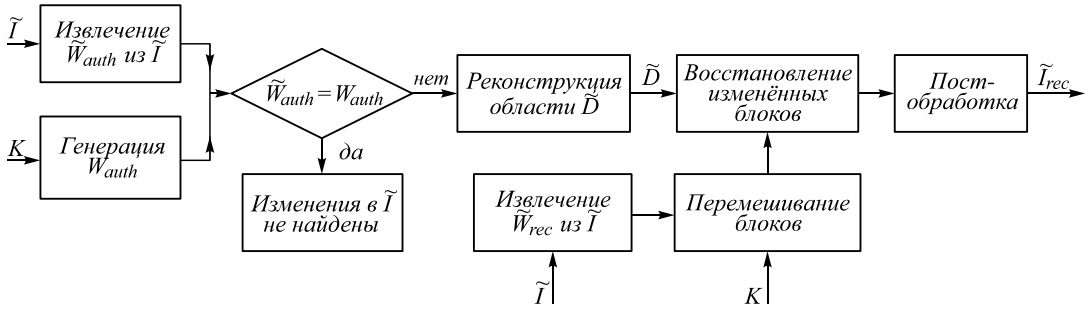


Рис. 12. Схема восстановления изображения после изменений

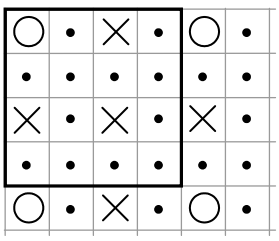


Рис. 13. Расположение отсчётов в блоке: точки – отсчёты уровня $l_{rec} = 0$, кресты – $l_{auth} = 1$, окружности – $l > 1$

Были рассмотрены две схемы генерации информации для восстановления. В первой схеме использовалась средняя интенсивность отсчётов блока (использовались 8 бит из 12 возможных). Такая схема имеет невысокую вычислительную сложность как на этапе встраивания ЦВЗ, так и на этапе аутентификации и восстановления изображения, но качество восстанавливаемых блоков не очень высокое.

Во второй схеме в качестве информации для восстановления сохранялись коэффициенты дискретного косинусного преобразования (ДКП) отсчётов блока. Общая схема приведена, например, в [14]: из отсчётов блока удаляется определенное количество младших бит, затем значения сдвигаются на половину

ширины диапазона значений. После этого выполняется ДКП с последующим квантованием коэффициентов преобразования. В качестве информации для восстановления сохраняется столько коэффициентов, сколько позволяет объем для встраивания. Преимуществом такой схемы является лучшее качество восстанавливаемых блоков, однако вычислительная сложность в этом случае возрастает.

Для исследования характеристик предлагаемого метода восстановления были проведены численные эксперименты. Как упоминалось выше, в каждый блок можно встроить до 12 бит информации для восстановления других блоков. В первом эксперименте были исследованы следующие схемы:

- *Среднее*: встраивается восемь (8) бит, определяющих среднее значение отсчётов внутри блока.
- *ДКП1*: берётся шесть (6) старших бит каждого отсчёта блока, выполняется ДКП, коэффициенты ДКП квантуются с использованием матрицы квантования JPEG [22]. Далее выбираются два максимальных коэффициента из четырёх в верхней левой подматрице. Для восстановления далее используются по 4 бита на каждое значение коэффициента (3 бита на модуль и 1 бит на знак), а также по 2 бита на его положение в подматрице.
- *ДКП2*: используется 6 старших бит на отсчёт. Сохраняются для последующего восстановления три коэффициента с координатами (1, 1), (1, 2), (2, 1) (нумерация начинается с 1), на хранение каждого коэффициента тратится 4 бита.
- *ДКП3*: используется 7 старших бит для каждого отсчёта блока. Хранится два максимальных коэффициента. На хранение первого отводится 5 бит, на хранение второго – 3 бита и по 2 бита на местоположение каждого из них.

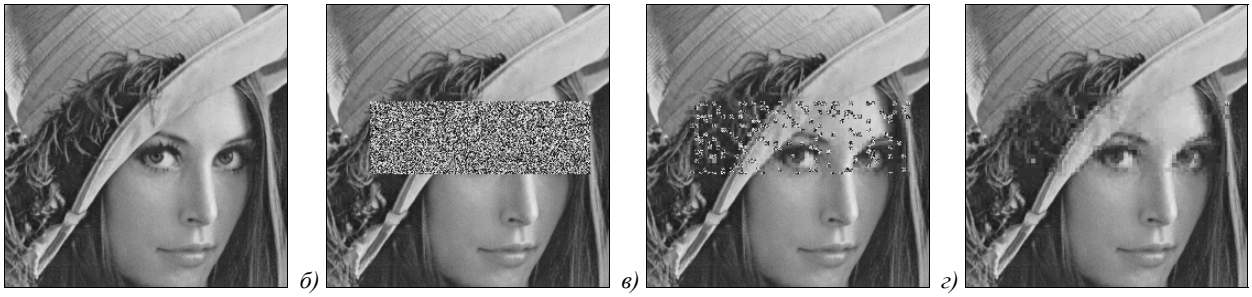
В табл. 2 приведены значения *PSNR*, характеризующие качество восстановления для каждой из этих схем. Следует отметить, что значение *PSNR* для восстановленного изображения вычислялось с использованием всех блоков изображения. Эти значения не зависят от площади изменений, параметра ε_{qim} и секретного ключа (начального значения генератора случайных чисел), и их сравнение позволяет оценить эффективность конкретной схемы представления блока изображения. Из таблицы видно, что *ДКП3* дает лучшие результаты для большинства рассмотренных изображений. По этой причине данная схема использовалась в дальнейших экспериментах.

Табл. 2. Значения PSNR для рассмотренных схем встраивания информации для восстановления

	Среднее	ДКП1	ДКП2	ДКП3
lena	24,67	27,14	27,60	**27,92**
barb	21,53	22,47	22,39	**22,81**
boat	23,37	26,38	26,74	**26,98**
bridge	22,41	25,48	**25,79**	25,78
goldhill	25,46	27,22	27,46	**28,22**
mandrill	25,29	26,62	26,86	**27,53**
mountain	16,75	18,57	**18,73**	18,46
peppers	26,52	28,74	29,13	**29,74**
washsat	30,67	29,93	30,05	**31,76**
zelda	27,56	29,11	29,47	**30,66**

На рис. 14 показаны основные этапы моделирования процесса восстановления от несанкционированного изменения. Данные для аутентификации и восстановления были сгенерированы и встроены в изображение. После этого отсчёты в заданной маске были заменены случайными значениями. В процессе аутентификации была реконструирована область искажений, и затем изменённые блоки были восстановлены либо при помощи ЦВЗ, либо по соседним значениям.

В табл. 3 представлены значения *PSNR* для изображений с ЦВЗ, а также для восстановленных изображений для различных значений ε_{qim} и процента встраивания. Значения усреднены для различных позиций маски изменений и значений секретного ключа. Как и ожидалось, чем больше значение ε_{qim}, тем больше искажений вносится в изображение и тем хуже качество восстановленного изображения (с другой стороны, тем большая степень сжатия может быть достигнута без разрушения ЦВЗ). Кроме того, процент изменений также влияет на качество восстановленного изображения: увеличение процента изменений уменьшает количество блоков, которые могут быть восстановлены, и мы вынуждены использовать усреднение по соседям, дающее более низкое качество восстановления. В целом, таблица показывает результаты, сопоставимые с данными, опубликованными в работе [14].



а) б) в) г)

Рис. 14.Обнаружение изменений и восстановление изображения: изображение с ЦВЗ (ε_{qim}=4) (а); внесенные изменения (20% изменений) (б); восстановленные блоки (в); восстановленное изображение (после постобработки) (г)

Табл. 3. Значения PSNR для изображения с ЦВЗ и восстановленного изображения для различных значений ε_{qim} и процента изменений (для изображения "Lena")

ε_{qim}	*PSNR* с ЦВЗ	*PSNR* восстанов. (10 % изменений)	*PSNR* восстанов. (20 % изменений)	*PSNR* восстанов. (30 % изменений)
4	40,98	34,62	31,39	29,09
8	35,19	32,63	30,48	28,62
12	31,81	30,63	29,31	27,87

Заключение

В статье предложена система цифровых водяных знаков, обладающая возможностью аутентификации, реконструкция области локальных изменений и восстановления искажённых областей. Система совместима с алгоритмом сжатия HGI: это означает, что сжатое при помощи HGI изображение с ЦВЗ остаётся защищённым, а его локальные искажения могут быть обнаружены и исправлены.

В исследовании было рассмотрено несколько схем встраивания информации о восстановлении. Таким образом, существует выбор между быстрым, но менее эффективным (с точки зрения качества восстановления изображений) усреднением и вычислительно затратным, но более эффективным ДКПЗ.

Дальнейшая работа может быть направлена на уменьшение ошибок пропусков (пропуск искаженных блоков) и исследование различных схем перемешивания блоков.

Благодарности

Работа выполнена при поддержке РФФИ (проект 19-29-09045 в части параграфа 3, проект 19-07-00357 в части введения и параграфа 1) и Министерства науки и высшего образования РФ в рамках выполнения работ по Государственному заданию ФНИЦ «Кристаллография и фотоника» РАН (соглашение 007-ГЗ/Ч3363/26) в части параграфа 2.

References

[1] Xuan X, Peng B, Wang W, Dong J. On the generalization of gan image forensics. In Book: Sun Z, He R, Feng J, Shan S, Guo Z, eds. Biometric recognition. Cham: Springer International Publishing; 2019: 134-141. DOI: 10.1007/978-3-030-31456-9_15.

[2] Westerlund M. The emergence of deepfake technology: A review. Technol Innov Manag Rev 2019; 9: 39-52.

[3] Nyeem H, Boles W, Boyd C. A review of medical image watermarking requirements for teleradiology. J Digit Imaging 2013; 26: 326-343. DOI: 10.1007/s10278-012-9527-x.

[4] Barni M, Bartolini F, Cappellini V, Magli E, Olmo G, Zanini R. Copyright protection of remote sensing imagery by means of digital watermarking. Proc SPIE 2001; 4540: 565-576. DOI: 10.1117/12.450706.

[5] Chuvieco E. Fundamentals of satellite remote sensing: An environmental approach. 2nd ed. Boca Raton: CRC Press; 2016.

[6] Rajalakshmi C, Germanux Alex M, Balasubramanian R. Copy move forgery detection using key point localized super pixel based on texture features. Computer Optics 2019; 43(2): 270-276. DOI: 10.18287/2412-6179-2019-43-2-270-276.

[7] Cox I, Miller M, Bloom J, Fridrich J, Kalker T. Digital watermarking and steganography. Burlington: Morgan Kaufmann; 2007. DOI: 10.1016/B978-0-12-372585-1.X5001-3.

[8] Barni M, Bartolini F. Watermarking systems engineering. New-York, USA: Marcel Dekker Inc; 2004.

[9] Evsutin OO, Shelupanov AA, Meshcheryakov RV, Bondarenko DO. An algorithm for information embedding into compressed digital images based on replacement procedures with use of optimization. Computer Optics 2017; 41(3): 412-421. DOI: 10.18287/2412-6179-2017-41-3-412-421.

[10] Gashnikov MV, Glumov NI, Sergeev VV. A hierarchical compression method for space images. Autom Remote Control 2010; 71(3): 501-513. DOI: 10.1134/S0005117910030112.

[11] Gashnikov MV, Glumov NI. Hierarchical grid interpolation for hyperspectral image compression. Computer Optics 2014; 38(1): 87-93. DOI: 10.18287/0134-2452-2014-38-1-87-93.

[12] Chen B, Wornell G. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. IEEE Transaction on Information Theory 2001; 47: 1423-1443.

[13] Rakhmawati L, Wirawan W, Suwadi S. A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability. J Image Video Proc 2019; 61. DOI: 10.1186/s13640-019-0462-3.

[14] Singh D, Singh SK. Effective self-embedding watermarking scheme for image tampered detection and localization with recovery capability. J Vis Commun Image Represent 2016; 38: 775-789. DOI: 10.1016/j.jvcir.2016.04.023.

[15] Qin C, Wang H, Zhang X, Sun X. Self-embedding fragile watermarking based on reference-data interleaving and adaptive selection of embedding mode. Inf Sci 2016; 373: 233-250. DOI: 10.1016/j.ins.2016.09.001.

[16] Han Q, Han L, Wang E, Yang J. Dual watermarking for image tamper detection and self-recovery. Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing 2013: 33-36. DOI: 10.1109/IIH-MSP.2013.17.

[17] Egorova AA, Fedoseev VA. A classification of semi-fragile watermarking systems for JPEG images. Computer Optics 2019; 43(3): 419-433. DOI: 10.18287/2412-6179-2019-43-3-419-433.

[18] Lin C-Y, Chang S-F. Issues and solutions for authenticating MPEG video. Proc SPIE 1999; 3657: 54-65. DOI: 10.1117/12.344703.

[19] Sun Q, Chang S-F, Kurato M, Suto M. A quantitative semi-fragile JPEG2000 image authentication system. IEEE ICIP 2002; 2: 921-924. DOI: 10.1109/ICIP.2002.1040102.

[20] The waterloo fractal coding and analysis group. 2019. Source: ⟨http://links.uwaterloo.ca/Repository.html⟩.

[21] Mitekin V. A new key recovery attack against DM-QIM image watermarking algorithm. Proc SPIE 2017; 10341: 103411A. DOI: 10.1117/12.2268550.

[22] Quantization Matrix. 2021. Source: ⟨https://www.sciencedirect.com/topics/computer-science/quantization-matrix⟩.

Сведения об авторах

Баврина Алина Юрьевна, 1980 года рождения, в 2003 году окончила Самарский государственный аэрокосмический университет имени академика С.П. Королёва (ныне – Самарский университет), в 2006 году защитила кандидатскую диссертацию по техническим наукам. Работает научным сотрудником в ИСОИ РАН – филиале ФНИЦ «Кристаллография и фотоника» РАН, а также в Самарском университете в должности старшего научного сотрудника по совместительству. Области научных интересов: обработка изображений, распознавание образов, защита цифровых изображений. E-mail: *bavrina@mail.ru*.

Федосеев Виктор Андреевич, 1986 года рождения, в 2009 году Самарский государственный аэрокосмический университет имени академика С.П. Королёва (ныне – Самарский университет) по специальности «Прикладная математика и информатика», кандидат физико-математических наук (2012). В настоящее время работает доцентом кафедры геоинформатики и информационной безопасности Самарского университета и научным сотрудником в ИСОИ РАН – филиале ФНИЦ «Кристаллография и фотоника» РАН. Области научных интересов: анализ изображений, цифровые водяные знаки, стеганография. E-mail: *vicanfed@gmail.com*.

ГРНТИ: 28.23.15
Поступила в редакцию 08 августа 2021 г. Окончательный вариант – 30 сентября 2021 г.

Semi-fragile watermarking with recovery capabilities for HGI compression method

A.Y. Bavrina [1,2], V.A. Fedoseev [2,1]
[1] IPSI RAS – Branch of the FSRC "Crystallography and Photonics" RAS, 443001, Samara, Russia, Molodogvardeyskaya 151;
[2] Samara National Research University, 443086, Samara, Russia, Moskovskoye Shosse 34

Abstract

The article proposes a new semi-fragile watermarking system with the ability of tamper localization and recovery after distortions, adapted for the HGI image compression method. The system uses a hierarchical image structure when embedding and replaces the stage of post-interpolation residuals quantization with a special quantizer based on quantization index modulation. As a result, the protected image becomes resistant to HGI compression with an adjustable quality parameter. The proposed watermarking system allows an image to be restored after distortions with an acceptable quality. In this case, the authentication part and the recovery part operate at different hierarchical levels. The developed watermark system, compatible with the HGI compression method, may be used to protect remote sensing images and medical images from malicious distortion.

Keywords: digital image processing, digital watermarks, image compression, hierarchical grid interpolation method.

Citation: Bavrina AY, Fedoseev VA. Semi-fragile watermarking with recovery capabilities for HGI compression method. Computer Optics 2022; 46(1): 103-112. DOI: 10.18287/2412-6179-CO-1021.

Acknowledgements: This work was financially supported by the Russian Foundation for Basic Research under projects ## 19-29-09045 and 19-07-00357 and a state contract 007-GZ/Ch3363/26.

Authors' information

Alina Yurievna Bavrina (b. 1980) graduated from Samara State Aerospace University (presently, Samara National Research University) in 2003, received her PhD in Technical Sciences in 2006. At present she is researcher at the Image Processing Systems Institute of RAS – Branch of the FSRC «Crystallography and Photonics» RAS and part-time position as senior researcher at Samara University. Area of interests: digital image processing, pattern recognition, watermarking. E-mail: *bavrina@mail.ru* .

Victor Andreevich Fedoseev, (b. 1986), graduated from Samara State Aerospace University (presently, Samara National Research University) in 2009, majoring in Applied Mathematics and Computer Science. Candidate degree in Computer Science (2012). Currently he is an associate professor at the Geoinformatics and Information Security department at Samara National Research University and a research scientist at the Image Processing Systems Institute of RAS – Branch of the FSRC «Crystallography and Photonics» RAS. His scientific interests include image processing and analysis, digital watermarking and steganalysis. E-mail: *vicanfed@gmail.com* .

Received August 08, 2021. The final version – September 30, 2021.