# МЕТОД СОЗДАНИЯ ЦИФРОВЫХ ВОДЯНЫХ ЗНАКОВ НА ОСНОВЕ ГЕТЕРОАССОЦИАТИВНЫХ СЖИМАЮЩИХ ПРЕОБРАЗОВАНИЙ ИЗОБРАЖЕНИЙ И ЕГО РЕАЛИЗАЦИЯ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

*А.А. Сирота [1], М.А. Дрюченко [1], Е.Ю. Митрофанова [1]*
*[1] Воронежский государственный университет, Воронеж, Россия*

***Аннотация***

Рассматривается метод и реализуемые на его основе алгоритмы создания цифровых водяных знаков, базирующиеся на применении обобщённых сжимающих преобразований при встраивании данных во фрагменты контейнеров - изображений. Принципиальной особенностью предлагаемого метода является использование гетероассоциативного сжимающего преобразования – взаимного отображения со сжатием двух соседних областей изображения произвольной формы. Проведены комплексные исследования для оценки показателей качества создаваемых цифровых водяных знаков, и показано, что предлагаемый метод и реализуемые на его основе алгоритмы позволяют достаточно гибко регулировать показатели уровня искажения контейнера и достоверности извлечения скрытых данных. Проводится анализ предлагаемых алгоритмов создания цифровых водяных знаков в условиях воздействий различных видов помех и преобразований, направленных на разрушение скрываемой информации, а также результаты сравнения с известными стегоалгоритмами.

*Ключевые слова:* сжимающие отображения, обработка изображений, нейронные сети, стеганография, цифровые водяные знаки.

*Цитирование:* **Сирота, А.А.** Метод создания цифровых водяных знаков на основе гетероассоциативных сжимающих преобразований изображений и его реализация с использованием искусственных нейронных сетей / А.А. Сирота, М.А. Дрюченко, Е.Ю. Митрофанова // Компьютерная оптика. – 2018. – Т. 42, № 3. – С. 483-494. – DOI: 10.18287/2412-6179-2018-42-3-483-494.

## *Введение*

Широкое распространение цифровых технологий обуславливает необходимость разработки и совершенствования методов обеспечения конфиденциальности, целостности, контроля несанкционированного использования и тиражирования информации. К числу перспективных технологий, используемых для решения данных задач, можно отнести технологии компьютерной стеганографии, позволяющие организовывать защищённое скрытное хранение и передачу данных по открытым каналам коммуникации, а также скрытное маркирование данных цифровыми водяными знаками (ЦВЗ) в интересах контроля их использования и распространения. В качестве носителей (контейнеров) скрываемой пользовательской информации, как правило, выступают цифровые объекты, обладающие психовизуальной избыточностью. В данной работе в качестве контейнеров будут рассматриваться изображения.

Подавляющее большинство разработанных на сегодняшний день методов и алгоритмов стеганографического скрытия информации (ССИ) в изображения реализуют встраивание данных в их пространственную или частотную составляющую. Алгоритмы, работающие с пространственным представлением контейнеров, чаще всего реализуют варианты LSB- [1–3] или PVD-методов [4], BPCS-методы разделения битовых слоёв растра на сегменты по уровню сложности с последующей заменой шумовых областей битами сообщения [5, 6], а также варианты «блочного» ССИ за счёт манипуляции значениями цветности/яркости пикселей в блоках фиксированного размера [7, 8]. Основным недостатком подобных алгоритмов является слабая устойчивость скрытых данных к последующим трансформациям контейнера. Алгоритмы, работающие с частотным представлением контейнеров, чаще всего реализуют ССИ путём модификации спектральных коэффициентов (дискретного косинусного или вейвлет-преобразований) [9, 10]. Как правило, алгоритмы ССИ в частотной области контейнера обеспечивают большую робастность скрытых данных, но также зависят от используемых алгоритмов кодирования и конечного формата контейнера.

Для эффективного применения технологий ССИ в интересах создания ЦВЗ необходимо выполнить ряд противоречивых требований, а именно: обеспечить визуальную незаметность скрываемых данных, сохранив, по возможности, качество контейнера, и одновременно обеспечить высокую достоверность извлечения ранее скрытого ЦВЗ. Указанные противоречия не снимаются в полной мере в известных методах и алгоритмах [1–10].

Возможным подходом к дальнейшему развитию методов ССИ, обеспечивающим минимальную визуальную и статистическую заметность результатов ССИ с сохранением высокой достоверности извлечения скрытых данных, является использование аппарата искусственных нейронных сетей (НС) [11–15]. В данной работе, развивающей предыдущие работы авторов [13–15], рассматривается метод построения алгоритмов ССИ на основе сжимающих преобразований общего вида, реализуемых на фрагментах изображений произвольной формы с использованием НС прямого распространения, проводится полномасштабное исследование их эффективности в условиях внешних негативных воздействий, а также сравнение с известными алгоритмами. Реализуемый подход к ССИ, на наш взгляд, обладает следующими достоинствами:

– возможностью создания алгоритмов с универсальной архитектурой, не зависящей от форматов контейнеров;
– обеспечением повышенной скрытности (визуальной и статистической) и устойчивости процедур ССИ за счёт близости нейросетевых алгоритмов анализа данных к статистически оптимальным;
– реализацией менее прозрачных процедур ССИ, воспроизведение которых сторонними лицами будет затруднено.

Принципиальной особенностью постановки задачи является использование при ССИ взаимного отображения со сжатием двух соседних областей изображения, имеющих произвольную форму, т.е. гетероассоциативного сжимающего преобразования (ГСП) как общего случая. Использование преобразования такого вида позволяет находить существенные взаимосвязи для различных областей изображений. Стандартный вариант, когда сжатие производится по отношению к одной области, – автоассоциативное сжимающее преобразование (АСП) является частным случаем предыдущего.

### *1. Метод ГСП и его применение для стеганографического скрытия информации*

Теоретическое обоснование возможностей построения сжимающих преобразований с использованием искусственных НС для решения задач, связанных с представлением данных и нахождением взаимосвязей между фрагментами случайных полей и изображений, было представлено авторами в работе [14]. Одной из особенностей рассмотренного в [14] метода ГСП и реализуемых на его основе алгоритмов является инвариантность к конфигурации обрабатываемых фрагментов изображений, что может быть успешно использовано на практике при формировании стеганографических ключей. Показано, что применение аппарата НС для реализации ГСП предоставляет преимущества с точки зрения сокращения времени выполнения процедуры сжатия при наличии ранее обученных преобразователей.

Далее рассмотрим подробнее схему ССИ, основанную на использовании сжимающих преобразований общего вида, реализуемых на фрагментах контейнеров - изображений. Скрываемое сообщение (ЦВЗ) представляется в виде бинарной последовательности $b^{(p)} \in \{-1,+1\}$, $p = \overline{1,P}$. На маркируемом контейнере - изображении $Z$ выделяются непересекающиеся области пикселей $\Omega^{(p)}$ с настраиваемой площадью и геометрической конфигурацией. В простейшем случае $\Omega^{(p)}$ представляют собой прямоугольные блоки пикселей с фиксированной шириной и высотой. В каждый такой блок реализуется встраивание одного бита из $b$.

При выполнении ГСП области $\Omega^{(p)}$ разбиваются на непересекающиеся подобласти – «входную» $z_1^{(p)}$ и «выходную» $z_2^{(p)}$, $z_1^{(p)} \cup z_2^{(p)} = \Omega^{(p)}$, $z_1^{(p)} \cap z_2^{(p)} = \varnothing$. Затем производится отображение данных входной части в данные выходной. При выполнении сжимающего преобразования автоассоциатвиного типа «входные» и «выходные» области совпадают $z_1^{(p)} = z_2^{(p)} = \Omega^{(p)}$ и в процессе сжатия происходит отображение данных на себя. В общем же случае области $z_1^{(p)}$, $z_2^{(p)}$ могут быть произвольной конфигурации. Для примера на рис. 1 представлены две конфигурации фрагментов полутонового изображения: области прямоугольной формы с размещением одной прямоугольной области внутри другой и решетки со случайным выбором ячеек внутри области прямоугольной формы. Серым цветом выделены пиксели входной части фрагмента $z_1^{(p)}$, белым цветом – пиксели выходной части $z_2^{(p)}$. Для цветных изображений конфигурация входной и выходной частей может быть также произвольной; при этом, например, входная часть может включать две цветовые компоненты фрагмента, которые отображаются на оставшуюся компоненту.

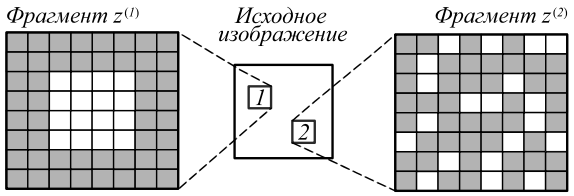


*Рис. 1. Примеры выходной части фрагмента изображения прямоугольной (а) и случайной (б) конфигурации*

В общем случае случайные векторы $z_1$, $z_2$ связаны соотношением:

$$z_2 = \mu_{2/1} + V = Hz_1 + V\ ,\ H = R_{z21}R_{z11}^{-1}\ , \qquad (1)$$

$$M[V] = 0\ ,\ M\left[VV^T\right] = R_{z22} - R_{z21}R_{z11}^{-1}R_{z12}\ ,$$

где $\mu_{2/1}$ – имеет смысл оптимальной (в классе линейных) оценки $z_2$ относительно наблюдения $z_1$; $V$ – стохастическая (маскирующая) составляющая, некоррелированная с $\mu_{2/1}$; $R_{z11} = M[z_1 z_1^T]$, $R_{z22} = M[z_2 z_2^T]$, $R_{z21} = M[z_2 z_1^T]$.

#### *1.1. Встраивание данных*

Предлагаемая схема стеганографического встраивания данных укрупнённо включает три этапа:

1. Подготовка контейнера для ССИ на основе «регулируемого сжатия» его фрагментов с формированием приближённой оценки «выходной части» $z_2^{(p)}$ по известным данным «входной части» $z_1^{(p)}$ со случайным или детерминированным выбором конфигурации областей локализации множеств пикселей $z_1^{(p)}$, $z_2^{(p)}$.
2. Формирование и выделение стохастической составляющей, маскирующей скрываемые данные.
3. Модификация «сжатых» фрагментов контейнера элементами встраиваемого сообщения и обратное добавление маскирующей составляющей.

При выполнении ССИ здесь используются две нейронные сети: НС-1 и НС-2, архитектура которых представлена на рис. 2*а, б*.

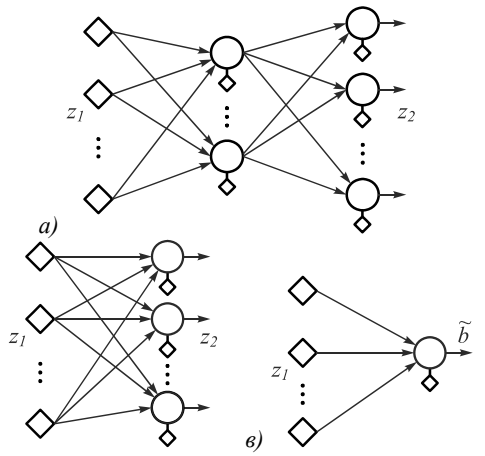


*Рис. 2. Архитектуры нейронных сетей, используемых при реализации ГСП и встраивании данных (а, б), а также при извлечении ранее скрытых данных (в)*

Здесь НС-1 (рис. 2*а*) выполняет сжимающее преобразование при количестве нейронов в скрытом слое $M<N_1$, $M<N_2$ ($N_1$, $N_2$ – количество входов и выходов сети) и применяется на первом этапе работы алгоритма для подготовки фрагментов контейнера перед ССИ. Обучение сети проводится по реализациям векторов $\{z_1^{(p)}, z_2^{(p)}, p=\overline{1,P}\}$, описывающим входные и выходные части всех используемых фрагментов контейнера. При $z_1^{(p)} \neq z_2^{(p)}$ НС-1 является гетероассоциативной, а при $z_1^{(p)} = z_2^{(p)}$ – автоассоциативной [16]. В случае реализации ГСП на вход НС-1 подаются данные «входной части» фрагмента $z_1^{(p)}$, при этом на выходе формируется сжатая оценка данных «выходной части» фрагмента $z_2^{(p)}$

$$\eta_{2/1}^{(p)} = W^{(2)} W^{(1)} z_1^{(p)}, \qquad (2)$$

где $W^{(1)}$, $W^{(2)}$ – матрицы весов первого и второго слоя НС-1. Оценки $\eta_{2/1}^{(p)}$ используются на последующих этапах работы алгоритма как элементы контейнера, в которые непосредственно и реализуется встраивание данных из $b$.

НС-2 (рис. 2*б*) применяется для формирования стохастической, маскирующей скрываемые данные составляющей. Обучение НС-2 также проводится по реализациям векторов $\{z_1^{(p)}, z_2^{(p)}, p=\overline{1,P}\}$. На выходе данной сети формируется реакция в виде квазиоптимальной оценки, соответствующей по структуре (см. (1)) оптимальной линейной оценке $\mu_{2/1}$ вектора $z_2$ относительно вектора $z_1$:

$$\nu_{2/1}^{(p)} = W z_1^{(p)} = \tilde{R}_{z21} \tilde{R}_{z11}^{-1} z_1^{(p)}, \qquad (3)$$

$$\tilde{R}_{z21} = \frac{1}{P-1} \sum_{p=1}^{P} z_2^{(p)} z_1^{(p)} = \left\| r_{jn}^{(2,1)} \right\|,$$

$$\tilde{R}_{z11} = \frac{1}{P-1} \sum_{p=1}^{P} z_1^{(p)} z_1^{(p)} = \left\| r_{in}^{(1,1)} \right\|,$$

где $\tilde{R}_{z21}$, $\tilde{R}_{z11}$ – выборочные матрицы ковариации случайных векторов, соответствующие матрицам $R_{z11}$, $R_{z21}$, $W$ – матрица весов НС-2. После этого для каждого фрагмента по аналогии с (1) можно выделить маскирующую, стохастическую составляющую прогноза как $V^{(p)} = z_2^{(p)} - \nu_{2/1}^{(p)}$.

С целью минимизации ошибки искажения контейнера при выполнении сжимающего преобразования далее (без ограничения общности) использовалась НС-1 при $M=N_2-1$, тогда $\eta_{2/1}^{(p)}$ будет отличаться от $\nu_{2/1}^{(p)}$ только отсутствием «высокочастотной» составляющей с дисперсией, соответствующей минимальному собственному числу выборочной матрицы ковариации $\tilde{R}_z = \tilde{R}_{z21} \tilde{R}_{z11}^{-1} \tilde{R}_{z12}$ оценки $\nu_{2/1}^{(p)} = \tilde{R}_{z21} \tilde{R}_{z11}^{-1} z_1^{(p)}$ [14].

Встраивание битов сообщения реализуется путём добавления к $\eta_{2/1}^{(p)}$ высокочастотной составляющей с амплитудой, изменяемой значением $b^{(p)}$, и обратным добавлением маскирующей составляющей

$$\bar{z}_2^{(p)} = \eta_{2/1}{}^{(p)} + A_m b^{(p)} \phi_{N_2} + V^{(p)}, \qquad (4)$$

$$\phi_{N_2} = r_{\min} \Big/ \sqrt{\left( r_{\min}{}^T r_{\min} \right)}, \; r_{\min} = \frac{1}{P} \sum_{p=1}^{P} \left( \eta_{2/1}{}^{(p)} - \nu_{2/1}{}^{(p)} \right),$$

где $A_m$ – задаваемая пользователем амплитуда встраиваемой последовательности. В (4) $\phi_{N_2}$ имеет смысл собственного вектора, соответствующего минимальному собственному значению матрицы выборочной матрицы $\tilde{R}_z$ [14].

В результате выполнения этих действий каждый заполненный фрагмент контейнера-изображения определяется вектором $\bar{z}^{(p)} = (z_1^{(p),T}, \bar{z}_2^{(p),T})^T$, т.е. состоит из нетронутой входной части и модифицированной выходной части. При этом заполненный контейнер в целом представляется совокупностью $\bar{Z} = \{\bar{z}^{(p)}, \; p=\overline{1,P}\}$.

Проведённый авторами в [14] анализ показал возможности построения ГСП как универсальных сжимающих преобразований спектрального типа с минимальным уровнем вносимых искажений на основе НС, архитектура которых представлена на рис. 2*а*, *б*. Установлено, что при выполнении ГСП каждый реально получаемый на выходе сети вектор квазиоптимальной оценки $\nu_{2/1}^{(p)}$, $p=\overline{1,P}$, может быть представлен в виде разложения по первым $M$ собственным векторам $\psi_i$, $i=\overline{1,M}$, выборочной матрицы ковариации $\tilde{R}_z = \tilde{R}_{z21} \tilde{R}_{z11}^{-1} \tilde{R}_{z12}$ такой оценки вектора $z_2$ относительно $z_1$. Показана возможность построения ГСП двумя способами: либо путём обучения НС, имеющей представленную на рис. 2*а* архитектуру; либо путём расчёта весовых коэффициентов автоассоциативного преобразователя линейной оценки $\nu_{2/1}$ на основе решения задачи на собственные числа и собственные векторы для матрицы $\tilde{R}_z$.

Укрупнённая схема работы алгоритма встраивания данных с использованием обученных НС-1 и НС-2 приведена на рис. 3. На первом шаге алгоритма

встраивания задаётся стеганографический ключ $K_1$. В общем случае его компонентами являются:

– параметры инициализации генераторов псевдослучайных числовых последовательностей (ПСЧП), определяющих структуру входных и выходных фрагментов в рамках каждого блока (псевдослучайный выбор пикселей $z_1^{(p)}$, $z_2^{(p)}$), а также порядок выбора блоков для модификации на всём изображении;
– параметры и значения весовых коэффициентов ранее обученных НС-1 и НС-2.

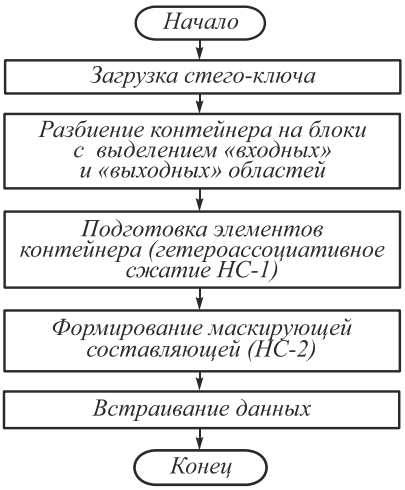


*Рис. 3. Обобщённая схема работы алгоритма встраивания данных*

Далее реализуется разбиение изображения-контейнера на фрагменты заданной конфигурации с выделением подмножеств «входных» и «выходных» элементов в каждом из них. В случае, если последующее использование маркированного контейнера не предполагает его серьёзных искажающих изменений, то более предпочтительным с точки зрения безопасности является вариант с псевдослучайным выбором координат пикселей «входных» и «выходных» элементов в рамках прямоугольных блоков небольшого размера. Отметим, что для эффективной работы алгоритма элементы $z_1^{(p)}$ и $z_2^{(p)}$ должны иметь ненулевую корреляционную зависимость, что исключает возможность наполнения $z_1^{(p)}$ и $z_2^{(p)}$ случайно выбираемыми по всему контейнеру пространственно разнесёнными пикселями.

Затем с использованием ранее обученных сетей НС-1 и НС-2 реализуется ГСП «выходных» фрагментов $z_2^{(p)}$ и формирование маскирующей встраиваемые данные составляющей, после чего реализуется (4) и элементы «заполненных» фрагментов $\overline{z}^{(p)}$ записываются на свои позиции в итоговом маркированном контейнере.

### *1.2. Извлечение данных*

При восстановлении ранее скрытых данных решается задача классификации вектора модифицированной выходной части фрагмента контейнера $\overline{z}_2$ по его принадлежности к одному из классов $H_1$, $H_2$. Каждый класс характеризуется различными векторами математического ожидания $H_1: m_+ = A_m \phi_{N_2}$, $H_2: m_+ = -A_m \phi_{N_2}$ и общей матрицей ковариации

$$R_\eta = W_{12} R_{z11} W_{12}^T + R_V ,$$

$$R_V = M[VV^T] = R_{z22} + W R_{z11} W^T - W R_{z12} - R_{z12}^T W^T .$$

Для построения решающего правила используется однослойная сеть НС-3 (рис. 2*в*) с линейной или нелинейной функцией активации. На вход НС-3 подаются модифицированные в результате встраивания выходные части фрагментов $\overline{z}_2^{(p)}$. Сеть обучается для решения задачи классификации входного для неё вектора с целью выделения значения ранее скрытого во фрагменте бита сообщения $\tilde{b}^{(p)}$, $p = \overline{1,P}$ $\| b^{(p)} - \tilde{b}^{(p)} \| \to \min$. Волнистая линия в обозначениях восстанавливаемой последовательности $\tilde{b}^{(p)}$ означает наличие возможных ошибок при классификации. Отметим, что вероятностный характер извлечения данных в значительной мере определяет варианты практического использования предложенной схемы ССИ, например, в приложениях для создания ЦВЗ.

Укрупнённая схема работы алгоритма извлечения ранее скрытых данных приведена на рис. 4.

Сначала выполняется загрузка стеганографического ключа $K_2$, опционально включающего параметры инициализации генераторов ПСЧП, определяющих структуру входных и выходных фрагментов в рамках каждого блока, порядок выбора блоков на всём изображении, а также параметры и значения весовых коэффициентов ранее обученной НС-3. Далее с использованием $K_2$ реализуется прогонка множества выделенных «выходных» элементов $\overline{z}_2^{(p)}$ через НС-3 с формированием выходного бинарного вектора $\tilde{b}^{(p)}$, $p = \overline{1,P}$, на основе которого реализуется восстановление скрытых данных.

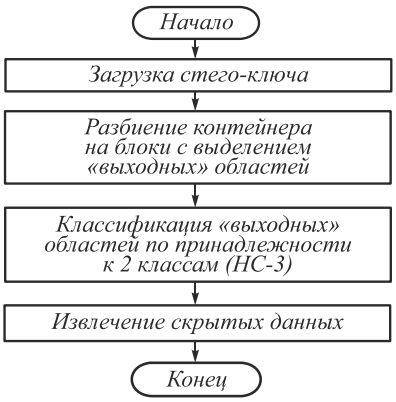


*Рис. 4. Обобщённая схема работы алгоритма извлечения данных*

## *2. Результаты экспериментальных исследований и их обсуждение*

### *2.1. Анализ показателей искажения контейнера при встраивании ЦВЗ*

Экспериментальный анализ алгоритма создания ЦВЗ проводился с использованием тестовых приме-

ров, формируемых на основе генерации реализаций однородных гауссовских случайных полей, а также реальных цветных изображений. В ходе анализа использовались показатели, определяющие степень искажения контейнера при встраивании ЦВЗ и связанные с визуальной заметностью вносимых искажений, а также вероятность ошибки классификации элементов двоичной последовательности $P_{er}$ при извлечении ранее встроенных данных. При проведении экспериментов использовался комплекс программных средств, разработанных в среде MATLAB.

Оценка степени искажения контейнера при встраивании ЦВЗ изначально проводилась по двум показателям: средней квадратичной ошибке (СКО, MSE) и максимальной абсолютной ошибке (МАО, MAE) при сравнении исходного $Z$ и заполненного контейнера $\bar{Z}$.

В ходе исследований выяснилось, что данные показатели не всегда позволяют связать свои значения с визуальной заметностью вносимых искажений. Например, при малых уровнях MSE на гладких изображениях искажения оказались более заметны визуально, чем на изображениях, имеющих фрагменты с малой пространственной корреляцией, при более высоких уровнях MSE. Поэтому для анализа качества заполненных контейнеров далее, чтобы не перегружать изложение, приводятся только оценки на основе MAE, определяющей уровень абсолютных искажений контейнера. Использование MSE по отношению к MAE качественно не изменяет ход полученных зависимостей и результаты сравнения алгоритмов.

Более объективный анализ уровня искажения контейнера, на наш взгляд, позволяет провести индекс структурного сходства (SSIM) Бовика [17]. Для полутоновых изображений в наших экспериментах он рассчитывался стандартным образом. Для цветных изображений он рассчитывался как среднее геометрическое показателей, полученных для каждой из трёх цветовых компонент RGB.

В ходе исследований проводились эксперименты с различными параметрами алгоритмов создания ЦВЗ. При создании ЦВЗ для полутоновых изображений, генерируемых как реализации гауссовских случайных полей $w(x,y)$, $x=\overline{1,800}$, $y=\overline{1,800}$ с различными функциями пространственной корреляции (в приведенных ниже примерах $R(x,x',y,y')=\sigma^2 e^{-a\sqrt{(x-x')^2+(y-y')^2}}$), рассматривались три типовых варианта построения алгоритма встраивания: на основе ГСП с использованием фрагментов квадратной формы со сторонами 8×8, выходной частью в виде прямоугольника со сторонами 5×6 и окружающей её входной частью ($N=64$, $N_1=34$, $N_2=30$); на основе ГСП с использованием фрагментов квадратной формы 8×8, выходной частью в виде 30 случайным образом размещённых в пределах фрагмента пикселей и соответствующей входной частью случайной конфигурации – оставшимися пикселями ($N=64$, $N_1=34$, $N_2=30$); на основе АСП с использованием фрагментов квадратной формы 8×8, в которых входная и выходная часть совпадают и полностью заполняют фрагмент $N=N_1=N_2=64$. Для упрощения представления дальнейших результатов эти варианты будем обозначать соответственно *vm*-1, *vm*-2, *vm*-3.

При создании ЦВЗ для цветных реальных изображений (формат *.bmp) также рассматривались три типовых варианта построения алгоритма встраивания: на основе ГСП с использованием в каждой цветовой компоненте фрагментов прямоугольной формы со сторонами 4×6, выходной частью в виде прямоугольника со сторонами 3×4 и окружающей её входной частью ($N=72$, $N_1=36$, $N_2=36$); на основе ГСП с использованием в каждой цветовой компоненте фрагментов 4×6, выходной частью в виде 36 случайным образом размещённых в «пироге трёх цветов» пикселей и соответствующей входной частью случайной конфигурации ($N=72$, $N_1=36$, $N_2=36$); на основе АСП с использованием в каждой цветовой компоненте фрагментов прямоугольной формы 4×6, в которых входная и выходная часть совпадают и полностью заполняют фрагмент $N=N_1=N_2=72$. При представлении дальнейших результатов эти варианты будут обозначены как *vc*-1, *vc*-2, *vc*-3.

Особое внимание уделялось анализу возможности использования нейросетевых преобразователей данных, которые обучались на одних изображениях для встраивания и извлечения ЦВЗ на других изображениях. Фактически это означает возможность использования алгоритма, адаптированного к определённому классу изображений, для применения по отношению к другим изображениям, т.е. контентную независимость алгоритма.

Далее приведены результаты, которые отражают типичный характер закономерностей, полученных в целом и для других случаев. На рис. 5 при использовании в качестве контейнеров полутоновых изображений, формируемых как реализации случайного поля (варианты *vm*-1, *vm*-2, *vm*-3), приведены зависимости для SSIM и $P_{er}$ восстановления ЦВЗ от амплитуды встраиваемой последовательности $A_m$, задаваемой в единицах уровня $i/255$, $i=\overline{0,255}$.

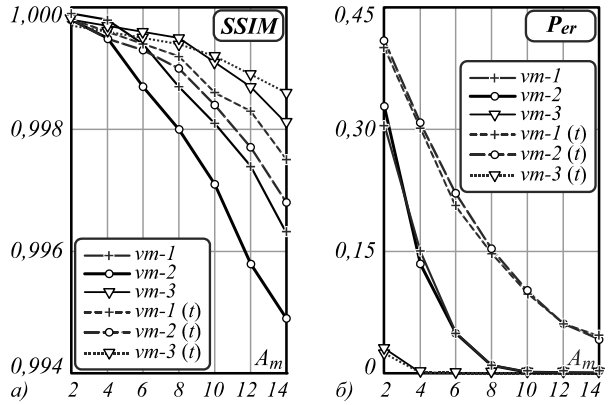


*Рис. 5. Зависимости показателей качества ЦВЗ для полутоновых изображений (реализаций случайных полей)*

Объём обучающей выборки для формирования НС $P_1=10000$, объём тестирующей выборки $P_2=10000$.

Здесь обучение и тестирование алгоритмов встраивания и извлечения ЦВЗ проводилось для гауссовских случайных полей с параметром функции корреляции $a_1 = 0{,}01$. Кроме этого, проводилось тестирование ранее обученных алгоритмов с использованием реализаций гауссовских случайных полей, генерируемых с параметром функции корреляции $a_2 = 0{,}05$ (кривые *vm*-1(*t*), *vm*-2(*t*), *vm*-3(*t*)). Это позволяет судить о степени «контентной независимости» обучаемых алгоритмов и оценки возможности их применения для создания ЦВЗ в других объектах.

Анализ зависимостей на рис. 5 показывает следующее. Значения показателя SSIM, естественно, лучше при использовании изображений с большим уровнем пространственной корреляции ($a_1 = 0{,}05$), чем у изображений с малой пространственной корреляцией ($a_1 = 0{,}01$). При этом величина MAE для вариантов с $a_1 = 0{,}05$ и $a_1 = 0{,}01$ примерно одинакова и варьируется от 1–2 при $A_m = 2$ до 7–12 при $A_m = 14$. При малой $A_m$ использование АСП приводит к большим искажениям контейнера (MAE порядка 6) по сравнению с вариантами, основанными на использовании ГСП (MAE менее 5), однако уже для значений $A_m > 4$ наблюдается обратная картина – для АСП MAE составляет порядка 6–9, а для ГСП – MAE порядка 7–12. Вероятность ошибки $P_{er}$ для вариантов с ГСП будет тем больше, чем больше пространственная корреляция изображений. Для АСП наличие отличной от теоретически нулевой вероятности ошибки при малых $A_m$ объясняется шумом квантования. При задании других значений параметра функции корреляции ($a_1 = 0{,}05$ для обучающего изображения, $a_2 = 0{,}01$ для тестирующего изображения) все зависимости располагаются в обратном порядке.

На рис. 6*а, б* при использовании контейнеров в виде цветных изображений (варианты *vc*-1, *vc*-2, *vc*-3) приведены аналогичные зависимости для SSIM и $P_{er}$ от амплитуды встраиваемой последовательности $A_m$.

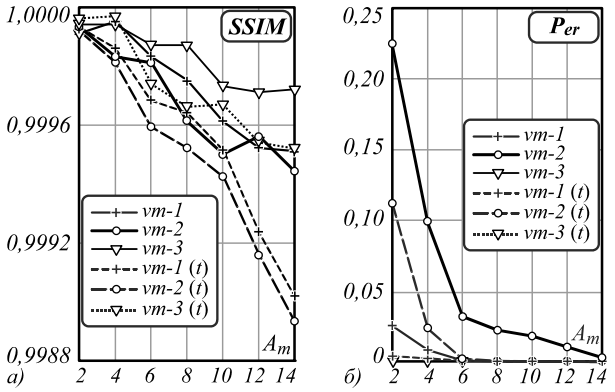


*Рис. 6. Зависимости показателей качества ЦВЗ для цветных изображений*

Здесь при обучении и тестировании алгоритмов использовалось изображение «lena.bmp», а при тестировании на ранее обученных алгоритмах – изображение «cars.bmp» (рис. 7*а, б*).

На рис. 8*а, б* в увеличенном масштабе приведены фрагменты исходного и заполненного контейнера (*vc*-2), а также разница между ними, умноженная на 100 (рис. 8*в*).


*Рис. 7. Примеры изображений-контейнеров*

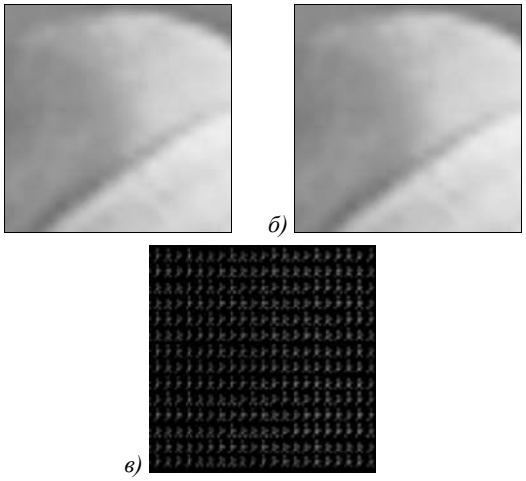


*Рис. 8. Фрагменты исходного и заполненного контейнеров и разница между ними, умноженная на 100*

Анализ зависимостей, приведенных на рис. 6, показывает в целом близость полученных здесь закономерностей с закономерностями, полученными для реализаций случайных полей.

### *2.2. Исследование стойкости ЦВЗ по отношению к негативным воздействиям*

Важным фактором, определяющим качество ЦВЗ, является стойкость в смысле сохранения скрытой в контейнерах информации по отношению к внешним негативным воздействиям (НВ) negative impact (NI) преднамеренного и непреднамеренного характера. Общий подход при выполнении подобных исследований состоял в воспроизведении всех этапов создания и восстановления ЦВЗ при введении определённых допущений относительно характера НВ и имеющейся у нарушителя априорной информации. Для различных типов НВ использован относительный уровень негативного воздействия $\varepsilon = 1, 2, ..., 9$, приведённый к диапазонам значений варьируемых параметров. В ходе экспериментов осуществлялся объективный и субъективный контроль уровня вносимых при НВ искажений, чтобы не допустить неприемлемой потери качества контейнера. Исследования проводились с использованием генерируемых реализаций случайных полей и реальных изображений. Следует отметить, что вопросы устойчивости алгоритмов создания ЦВЗ на основе искусственных нейронных сетей подобного типа были рассмотрены авторами ранее в работе [15] в существенно более ограниченном объёме.

В табл. 1 представлены типовые зависимости вероятности ошибки $P_{er}$ от ε при извлечении элементов последовательности ЦВЗ в условиях различных НВ для вариантов алгоритма *vc*-1, *vc*-2, *vc*-3. Амплитуда встраиваемой последовательности $A_m = 8$.

*Табл. 1. Вероятности ошибки при извлечении ЦВЗ после применения к маркированному контейнеру различных НВ*

| ε | Тип негативного воздействия | | | | | |
|---|---|---|---|---|---|---|
| | Адд. шум | Имп. шум | Unsharp mask | Low filter $k_h = 4$ | JPEG $k_h = 8$ | Подмена ЦВЗ |
| ГСП, *vc*-1, $A_m = 8$ | | | | | | |
| 1 | 0,0071 | 0,0494 | $<10^{-3}$ | **$<10^{-3}$** | **$<10^{-3}$** | $<10^{-3}$ |
| 3 | 0,0421 | 0,1310 | 0,0085 | **0,0021** | **$<10^{-3}$** | $<10^{-3}$ |
| 5 | 0,0688 | 0,2019 | 0,0221 | **0,0055** | **0,0086** | 0,0042 |
| 7 | 0,1083 | 0,2506 | 0,0379 | **0,0177** | **0,1221** | 0,0494 |
| 9 | 0,1396 | 0,2911 | 0,0492 | **0,1289** | **0,2716** | 0,4681 |
| ГСП, *vc*-2, $A_m = 8$ | | | | | | |
| 1 | 0,0084 | 0,0297 | $<10^{-3}$ | **0,0013** | **0,0212** | $<10^{-3}$ |
| 3 | 0,0483 | 0,0822 | 0,0084 | **0,0135** | **0,0305** | $<10^{-3}$ |
| 5 | 0,0810 | 0,1376 | 0,0227 | **0,0247** | **0,1043** | 0,0057 |
| 7 | 0,1252 | 0,1771 | 0,0378 | **0,0420** | **0,3829** | 0,0226 |
| 9 | 0,1491 | 0,2157 | 0,0498 | **0,1061** | **0,4802** | 0,4205 |
| АСП, *vc*-3, $A_m = 8$ | | | | | | |
| 1 | 0,0087 | 0,0779 | $<10^{-3}$ | **$<10^{-3}$** | **$<10^{-3}$** | 0,4772 |
| 3 | 0,0502 | 0,2008 | 0,0071 | **$<10^{-3}$** | **0,0027** | 0,4935 |
| 5 | 0,0907 | 0,2710 | 0,0166 | **$<10^{-3}$** | **0,1498** | 0,5107 |
| 7 | 0,1289 | 0,3284 | 0,0214 | **0,0012** | **0,4366** | 0,4930 |
| 9 | 0,1506 | 0,3688 | 0,0246 | **0,1406** | **0,4911** | 0,5023 |

Зависимости приведены для изображения «cars.bmp» и являются типичными. С использованием стандартных функций и упомянутых далее опций среды MATLAB реализованы следующие типовые НВ:

- аддитивный гауссовский шум с относительной от интенсивности $I(x,y)$ каждого пикселя локальной дисперсией $\sigma_n^2 = 0,00025 \times \varepsilon \times I$;
- импульсный шум, реализующий подавление пикселей исходного изображения с вероятностью $p_n = 0,005 \times \varepsilon$;
- фильтрация на основе скользящего фильтра с Unsharp Mask (повышение резкости) с параметром усиления резкости $Amount = 2 \times \varepsilon$;
- низкочастотная фильтрация (Low filter) на основе скользящего фильтра в виде гауссианы с размером маски $h_L = \varepsilon$;
- применение сжатия изображений на основе преобразования JPEG со стандартным коэффициентом потери качества Quality $q = 100 - 10 \cdot \varepsilon$;
- уничтожение или подмена ЦВЗ при известной структуре и параметрах алгоритма путём встраивания другого ЦВЗ с амплитудой $A'_m = \varepsilon$.

Анализ полученных в ходе исследований результатов показывает достаточно высокую степень стойкости алгоритмов по отношению к НВ в виде аддитивных и импульсных помех, а также при применении фильтров Unsharp Mask. Существенный уровень вероятности ошибки при извлечении ЦВЗ здесь достигался при таком уровне НВ, который приводил к недопустимым искажениям контейнера, диагностируемым как на основе SSIM, так и визуально.

Для НВ, основанных на низкочастотном сглаживании с использованием масочных фильтров и JEPG-компрессии, уже при малых значениях ε, соответствующих размеру маски сглаживающего фильтра 3–5 пикселей и коэффициенту потери качества JPEG 80–90, происходит искажение ранее встроенных ЦВЗ.

Для противодействия подобным НВ была предложена следующая версия алгоритма. Она основана на переходе от модификации значений отдельных пикселей изображений при встраивании ЦВЗ к изменению средних значений групп рядом стоящих пикселей, образующих блоки размером $k_h \times k_h$ ($k_h = 2, ..., 8$). При этом исходное изображение $I$ разбивается на блоки $u_{ij}^{(r,g,b)}$, $i = \overline{1, I_x}$, $j = \overline{1, J_y}$ размером $k_h \times k_h$ пикселей, $I_x, J_y$ – количество блоков, размещаемых по оси OX и по оси OY. Затем вычисляются матрицы средних значений блоков $U^{(r,g,b)} = \|\overline{u}_{ij}^{(r,g,b)}\|$. Далее по отношению к полученному таким образом изображению $U$ выполняется стандартный алгоритм встраивания ЦВЗ на основе ГСП или АСП, а затем проводится соответствующая модификация средних значений яркости исходного изображения. Алгоритм извлечения ЦВЗ реализуется стандартным образом. Для обеспечения большей стойкости ЦВЗ по отношению к JPEG рекомендуется предварительно выполнить преобразование, переводящее исходное изображение из формата RGB в формат YCbCr. Следует отметить, что при таком способе создания ЦВЗ пропускная способность контейнера уменьшается в $k_h \times k_h$ раз.

Значения вероятности ошибок при реализации подобной схемы встраивания ЦВЗ выделены в табл. 1 жирным шрифтом. Здесь мы видим, что уже при $k_h = 4$ НВ в виде низкочастотного сглаживания фактически парируется. Для преобразования JPEG устойчивый эффект достигается для значений $k_h = 8$. При этом разрушение ЦВЗ происходит для таких уровней сжатия, которые соответствуют недопустимой степени потери качества исходного изображения.

В целом по представленным в табл. 1 зависимостям следует отметить более высокую стойкость алгоритмов с ГСП (варианты *vc*-1, *vc*-2) по отношению к типовым НВ по сравнению с алгоритмом, основанным на АСП (вариант *vc*-3).

Интересным представляется ещё одно свойство алгоритмов с ГСП: при их применении в рассматриваемой схеме внесения ЦВЗ произвести уничтожение или подмену ранее встроенной информации весьма затруднительно. Здесь возможны две ситуации.

В первой из них нарушитель знает только структуру алгоритма (т.е. фактически конфигурацию входной и выходной частей). Здесь при отсутствии у нарушителя точных значений элементов матричных операторов преобразования данных он может попытаться заново обучить нейронные преобразователи на контейнере с уже внедрённым в него ЦВЗ. Это в большинстве случаев не позволит уничтожить ЦВЗ,

так как ранее модифицированная высокочастотная составляющая меняет порядок своего расположения при выполнении сжимающего преобразования (в рассмотренных примерах уже при $A_m > 3$).

Во втором случае нарушителю известна не только структура, но и параметры преобразователей. И в этой ситуации уничтожить или подменить ЦВЗ затруднительно. Действительно, если для контейнера с ранее встроенным ЦВЗ абсолютно точно повторить всю последовательности преобразований (4) для любых вариантов алгоритма с ГСП, то при неизменной входной части значения оценок $\eta_{2/1}^{(p)}$, $\nu_{2/1}^{(p)}$, $p = \overline{1,P}$ очевидно сохранятся. При этом меняется только стохастическая составляющая прогноза, которая теперь будет содержать в себе разницу между модифицированной выходной частью и её оценкой $\tilde{V}^{(p)} = \tilde{z}_2^{(p)} - \nu_{2/1}^{(p)}$, $p = \overline{1,P}$. При её обратном добавлении во фрагменты контейнера в соответствие с (4) возвращается исходная ситуация. Встроенная нарушителем новая двоичная последовательность может изменить ЦВЗ только в том случае, если её амплитуда заведомо превышает амплитуду исходной последовательности (последний столбец табл. 1). А это, в свою очередь, ведёт к искажению контейнера и повышает заметность внесения в него новых данных. При использовании АСП маскирующая составляющая $V$ отсутствует и, соответственно, отсутствует возможность сохранения ранее встроенных данных. Поэтом в данном случае даже простое «зануление» высокочастотной составляющей приводит к удалению ЦВЗ (последний столбец табл. 1).

Следует также отметить, что для предлагаемого метода характерна также высокая скрытность по отношению к использованию сторонними лицами стеганографических ключей, используемых для извлечения данных. Использование неполной или частичной информации, особенно в алгоритмах со случайным размещением выходной части в компонентах цветного изображения, не позволяет обучить НС, способную восстановить ранее встроенный ЦВЗ.

### *2.3. Сравнение с известными алгоритмами встраивания ЦВЗ*

Для оценки эффективности предлагаемых в статье алгоритмов проводилось их сравнение с известными алгоритмами встраивания ЦВЗ, реализующими стегоскрытие данных в пространственную и частотную область изображений. Были выбраны алгоритмы блочного типа, работающие в пространственном представлении контейнера – алгоритм Брундокса (*Bruyndonckx*) [7] и Ленгелаара (*Langelaar*) [8], алгоритм Каттера (*Kutter*) [18], использующий значения цветности в локальных областях контейнера для прогноза значений восстанавливаемых бит ЦВЗ, алгоритм Коха–Жао (*Koch–Zhao*) [19], реализующий ССИ в частотной области контейнера путём относительной замены величин коэффициентов дискретного косинусного преобразования (данный принцип является базовым для значительного числа алгоритмов ССИ в частотной области), а также два относительно новых алгоритма на основе комбинации дискретного вейвлет (DWT), дискретного косинусного (DCT) преобразований и сингулярного разложения (SVD) DCT-DWT-SVD [20]. Выбор первых четырёх алгоритмов для сравнения с предлагаемыми вариантами был обусловлен сопоставимыми значениями стеганографической пропускной способности, а также широкой известностью и доступностью их реализаций. Алгоритмы на основе DCT-DWT-SVD выбирались для сравнения, поскольку они являются достаточно современными, робастными, высокопроизводительными стегоалгоритмами, реализующими принципы ССИ, широко применяемые во многих других алгоритмах создания ЦВЗ. Отметим, что по пропускной способности алгоритмы DCT-DWT-SVD превосходят описываемые в данной работе алгоритмы с ГСП. Однако для извлечения ЦВЗ эти алгоритмы требуют наличия исходного незаполненного контейнера, что является их недостатком по сравнению с предлагаемыми.

В алгоритмах Брундокса, Ленгелаара и Коха–Жао встраивание одного бита информации осуществлялось в блоки пикселей или спектральных коэффициентов размером $8 \times 8$. В алгоритме Каттера энергия встраиваемого бита данных задавалась равной $\lambda = 0{,}05$, а размер области прогнозирования значений цветности пикселей при восстановлении информации (сторона «креста») $\tau = 3$. Порог разности между коэффициентами в алгоритме Коха–Жао $\delta = 25$. В зависимости от локализации выбираемых для модификации пар спектральных коэффициентов рассматривались два варианта алгоритма Коха–Жао – реализующие ССИ в среднечастотные (СЧ) или в низкочастотные (НЧ) коэффициенты. В алгоритмах DCT-DWT-SVD согласно [20] рассматривались две модификации. В первой (DCT-DWT-SVD-1) реализуется встраивание ЦВЗ путём модификации сингулярных значений матрицы, составленной из НЧ коэффициентов DCT в блоках $4 \times 4$ для низкочастотной составляющей двухуровневого DWT. Во второй (DCT-DWT-SVD-2) реализуется встраивание ЦВЗ за счёт модификации матрицы сингулярных значений разложения SVD, примененного к матрице коэффициентов DCT, вычисленных для низкочастотной составляющей двухуровневого DWT (не в блоках $4 \times 4$). Значение усиливающего коэффициента $\alpha$, влияющего на качество создаваемых заполненных контейнеров и на степень робастности скрытых данных [20], в алгоритмах выбиралось равным 0,05.

Тестирование алгоритмов проводилось на выборке из 30 полноцветных изображений из наборов Kodak Lossless True Color Image Suite [21] и TESTIMAGES [22]. Разрешение тестовых изображений варьировалось от 300×300 до 1200×1200 пикселей. Анализ алгоритмов проводился в части оценки искажений заполненных контейнеров и достоверности восстановления ранее скрытых данных в отсутствие и при наличии НВ. В табл. 2 для каждого алгоритма приведены усреднённые для тестовой выборки значе-

ния SSIM и вероятности ошибок восстановления скрытой информации $P_{er}$. Представленные в третьем столбце таблицы значения MAE определялись как максимальные значения, полученные для каждого алгоритма по всем примерам тестовой выборки $\mathrm{MAE} = \max(\mathrm{MAE}_i),\ i = \overline{1,30}$.

*Табл. 2. Результаты сравнения предлагаемых алгоритмов создания ЦВЗ с известными*

| Алгоритм | SSIM | MAE | $P_{er}$ |
|---|---|---|---|
| ГСП, *vc*-1, $A_m = 8$ | 0,999 | 6 | 0,012 |
| ГСП, *vc*-2, $A_m = 8$ | 0,998 | 8 | 0,014 |
| Bruyndonckx, $8 \times 8$ | 0,995 | 31 | 0,041 |
| Langelaar, $8 \times 8$ | 0,989 | 38 | 0,034 |
| Kutter, $\lambda = 0{,}05$, $\nu = 3$ | 0,996 | 13 | 0,062 |
| Koch-Zhao, СЧ, $8 \times 8$ | 0,993 | 10 | 0,006 |
| Koch-Zhao, НЧ, $8 \times 8$ | 0,991 | 13 | $<10^{-3}$ |
| DCT-DWT-SVD-1 | 0,992 | 8 | $<10^{-3}$ |
| DCT-DWT-SVD-2 | 0,956 | 32 | $<10^{-3}$ |

Анализ полученных результатов позволяет сделать вывод, что алгоритмы на основе ГСП в данной конфигурации обеспечивают минимальные искажения при ССИ по сравнению со всеми выбранными для сравнения алгоритмами. При этом усреднённые по тестовой выборке значения вероятности ошибок при восстановлении ЦВЗ в отсутствие НВ для предложенных алгоритмов *vc*-1 и *vc*-2 составляют порядка 0,014, что свидетельствует о надежном восстановлении ЦВЗ. В выбранных для сравнения алгоритмах, работающих в пространственном представлении контейнера, $P_{er}$ без НВ составила от 0,034 до 0,062. Наименьшая вероятность ошибки извлечения данных в отсутствие НВ была получена для алгоритма Коха–Жао, а также для вариантов алгоритма DCT-DWT-SVD (при больших искажениях контейнера по сравнению с ГСП).

В табл. 3, табл. 4 для сравниваемых алгоритмов приведены усреднённые вероятности ошибок восстановления скрытой информации в случае применения к маркированным контейнерам различных НВ. Алгоритмы на основе ГСП показали хорошую устойчивость встроенных ЦВЗ к зашумлению контейнеров аддитивным гауссовским шумом $P_{er} \approx 0{,}05$, что несколько лучше результата, полученного для алгоритмов Ленгелаара, Коха–Жао и DCT-DWT-SVD-2, существенно превосходит результаты для алгоритмов Брундокса и Катера, но уступает алгоритму DCT-DWT-SVD-2.

При негативном воздействии на заполненные контейнеры в виде импульсного шума с плотностью $p_n = 0{,}01$ наиболее робастным также оказался алгоритм DCT-DWT-SVD-2. Для алгоритмов на основе ГСП *vc*-1, *vc*-2, алгоритмов Катера, Ленгелаара и Коха–Жао вероятность ошибки восстановления примерно сопоставима, в то время как алгоритм Брундокса здесь показал вдвое больший процент ошибочно восстанавливаемых бит ЦВЗ.

При воздействии на заполненные контейнеры НВ на основе фильтра повышения резкости (Unsharp Mask) $P_{er}$, полученная для алгоритмов на основе ГСП, оказывается сопоставима или меньше, чем во всех других анализируемых алгоритмах, кроме алгоритма DCT-DWT-SVD-2.

*Табл. 3. Результаты сравнения предлагаемых алгоритмов создания ЦВЗ с известными*

| Алгоритм | $P_{er}$ в условиях НВ | | |
|---|---|---|---|
| | *Amount* = 6 ад. шум $\sigma_n^2 = 4 \cdot 10^{-4}$ | Имп. шум $p_n = 0{,}01$ | Unsharp Mask |
| ГСП, *vc*-1, $A_m = 8$ | 0,046 | 0,112 | 0,008 |
| ГСП, *vc*-2, $A_m = 8$ | 0,053 | 0,081 | 0,007 |
| Bruyndonckx, $8 \times 8$ | 0,138 | 0,224 | 0,086 |
| Langelaar, $8 \times 8$ | 0,075 | 0,139 | 0,048 |
| Kutter, $\lambda = 0{,}05$, $\nu = 3$ | 0,178 | 0,126 | 0,011 |
| Koch–Zhao, СЧ, $8 \times 8$ | 0,081 | 0,155 | 0,008 |
| Koch–Zhao, НЧ, $8 \times 8$ | 0,094 | 0,143 | 0,005 |
| DCT-DWT-SVD-1 | 0,077 | 0,165 | 0,118 |
| DCT-DWT-SVD-2 | 0,008 | 0,028 | 0,003 |

*Табл. 4. Результаты сравнения предлагаемых алгоритмов создания ЦВЗ с известными*

| Алгоритм | $P_{er}$ в условиях НВ | | |
|---|---|---|---|
| | Low filter $h_L = 3$ | JPEG, $q = 80$ | JPEG, $q = 40$ |
| ГСП, *vc*-1, $A_m = 8$ | 0,326 / **0,012** | 0,494 / **0,006** | 0,488 / **0,011** |
| ГСП, *vc*-2, $A_m = 8$ | 0,332 / **0,023** | 0,492 / **0,016** | 0,497 / **0,039** |
| Bruyndonckx, $8 \times 8$ | 0,388 | 0,192 | 0,305 |
| Langelaar, $8 \times 8$ | 0,436 | 0,141 | 0,288 |
| Kutter, $\lambda = 0{,}05$, $\nu = 3$ | 0,421 | 0,455 | 0,492 |
| Koch-Zhao, СЧ, $8 \times 8$ | 0,227 | 0,016 | 0,511 |
| Koch-Zhao, НЧ, $8 \times 8$ | 0,096 | 0,004 | 0,042 |
| DCT-DWT-SVD-1 | 0,344 | 0,271 | 0,363 |
| DCT-DWT-SVD-2 | 0,005 | $<10^{-3}$ | 0,007 |

Применение низкочастотной фильтрации одинаково негативно отражается на результатах работы всех алгоритмов, работающих в пространственном представлении контейнера. Для маски размером $h_L \geq 3$ вероятности ошибок извлечения встроенной информации для всех алгоритмов превышают 0,3, то есть восстановления ЦВЗ с приемлемым уровнем достоверности не происходит. Наилучшая достоверность восстановления при данном типе НВ также была получена в алгоритме DCT-DWT-SVD-2.

При реализации JPEG-сжатия наиболее робастным также оказался алгоритм DCT-DWT-SVD-2. Предложенные алгоритмы на основе ГСП в этом случае не обеспечивают устойчивость встроенной информации. Однако в случае отсутствия жёстких требований по пропускной способности стегосистемы они также способны обеспечить высокую робастность встроенных меток к низкочастотной фильтрации и сжатию изображения при переходе от модификации значений отдельных пикселей к модификации средних значений пикселей в блоках заданного размера $k_h \times k_h$. Так, в табл. 3 жирным шрифтом приведены усреднённые значения вероятностей ошибок для Low filter и JPEG-преобразований, полученные для «блочных» вариан-

тов алгоритмов *vc*-1 и *vc*-2 на основе ГСП с $k_h = 8$. В данном случае вероятности ошибочного восстановления битов ЦВЗ для маркированных контейнеров, подвергшихся низкочастотной фильтрации с размером окна $h_L = 3$, не превышают 2,5%, а для контейнеров, подвергшихся сжатию с $q \geq 40$, – 4%.

### *Заключение*

Предложенный метод создания ЦВЗ основан на использовании гетероассоциативных сжимающих преобразований на фрагментах контейнеров произвольной формы с применением аппарата искусственных нейронных сетей. Метод обеспечивает форматонезависимость и универсальность реализуемых на его основе алгоритмов, что позволяет по общей схеме применять их для ССИ в различных типах контейнеров (изображения, звук, видео).

Проведенные исследования позволяют отметить следующие положительные качества предложенных алгоритмов: низкий уровень искажения контейнера в сочетании с высокой степенью устойчивости ЦВЗ по отношению к шумовым негативным воздействиям различных типов или воздействиям в виде высокочастотной пространственной фильтрации; наличие маскирующих свойств ГСП, которые затрудняют повторную перезапись ЦВЗ или подбор «ключа» для восстановления ЦВЗ. Для предлагаемого класса алгоритмов характерна также возможность гибкого варьирования показателей уровня искажения контейнера, достоверности извлечения ЦВЗ и его стойкости, что позволяет их использовать для решения различных задач ССИ со своими специфическими требованиями.

Одновременно следует отметить низкую устойчивость создаваемых ЦВЗ по отношению к низкочастотному сглаживанию и JPEG-компрессии. Для преодоления подобных негативных воздействий предложена модификация алгоритма, основанная на изменении по той же универсальной схеме обработки информации средних значений групп пикселей фиксированного размера. Подобная модификация алгоритмов на основе ГСП показала значительную устойчивость к указанным видам негативных воздействий.

Предложенные алгоритмы на основе ГСП обеспечивают минимальные искажения контейнеров по сравнению со всеми выбранными для сравнения стегоалгоритмами. В блочной модификации алгоритмы на основе ГСП превосходят выбранные для сравнения алгоритмы (за исключением DCT-DWT-SVD-2) по устойчивости скрытых данных к НВ типа импульсного и аддитивного гауссовского шума, повышения резкости, низкочастотной фильтрации и JPEG-компрессии. Стоит отметить, что оказавшийся более робастным алгоритм DCT-DWT-SVD-2 является «неслепым», т.е. требует наличия исходного незаполненного контейнера для извлечения ЦВЗ, что существенно ограничивает его применимость в реальных приложениях.

На практике исследованные алгоритмы на основе ГСП с учётом своей специфики могут быть достаточно эффективно использованы для создания полухрупких ЦВЗ и скрытного встраивания небольших объемов маркирующих данных в подвижные или статичные изображения высокого качества, обеспечивая минимальный уровень их визуальных искажений и сравнительно высокую достоверность восстановления.

Дальнейшее развитие предлагаемых алгоритмов может осуществляться в плане повышения их устойчивости на основе модификации по аналогичной схеме не только различных участков высокочастотных составляющих, но и среднечастотных составляющих, получаемых при выполнении гетероассоциативного преобразования фрагментов контейнера.

### *Литература*


1. **Younes, M.A.B.** A new steganography approach for image encryption exchange by using the least significant bit insertion / M.A.B. Younes, A. Jantan // International Journal of Computer Science and Network Security. – 2008. – Vol. 8, Issue 6. – P. 247-254.
2. **Hadhoud, M.M.** Secure perceptual data hiding technique using information theory / M.M. Hadhoud, N.A. Ismail, W. Shawkey, A.Z. Mohammed // International Conference on Electrical, Electronic and Computer Engineering (ICEEC '04). – 2004. – P. 249-253. – DOI: 10.1109/ICEEC.2004.1374433.
3. **Mandal, J.K.** Steganographic technique based on minimum deviation of fidelity / J.K. Mandal, M. Sengupta // 2011 Second International Conference on Emerging Applications of Information Technology. – 2011. – P. 298-301. – DOI: 10.1109/EAIT.2011.24.
4. **Wu, H.-C.** Image steganographic scheme based on pixel-value differencing and LSB replacement methods / H.-C. Wu, N.-I. Wu, C.-S. Tsai // IEE Proceedings – Vision, Image and Signal Processing. – 2005. – Vol. 152, Issue 5. – P. 611-615. – DOI: 10.1049/ip-vis:20059022.
5. **Kawaguchi, E.** Principle and applications of BPCS-steganography / E. Kawaguchi, R.O. Eason // Proceedings of SPIE. – 1998. – Vol. 3524. – P. 464-473. – DOI: 10.1117/12.337436.
6. **Maya, S.T.** Robust steganography using bit plane complexity segmentation / S.T. Maya, M.N. Miyatake, R.V. Medina // 1st International Conference on Electrical and Electronics Engineering. – 2004. – P. 40-43. – DOI: 10.1109/ICEEE.2004.1433845.
7. **Darmstaedter, V.** Low cost spatial watermarking / V. Darmstaedter, J.-F. Delaigle, J.J. Quisquater, B. Macq // Computers and Graphics. – 1998. – Vol. 22, Issue 4. – P. 417-424. – DOI: 10.1016/S0097-8493(98)00031-4.
8. **Langelaar, G.C.** Robust labeling methods for copy protection of images / G.C. Langelaar, J.C.A. van der Lubbe, R.L. Lagendijk // Proceedings of the SPIE. – 1997. – Vol. 3022. – P. 298-309. – DOI: 10.1117/12.263418.
9. **Provos, N.** Hide and seek: An introduction to steganography / N. Provos, P. Honeyman // IEEE Security and Privacy. – 2003. – Vol. 1, Issue 3. – P. 32-44. – DOI: 10.1109/MSECP.2003.1203220.
10. **Westfeld, A.** F5-A steganographic algorithm: High capacity despite better steganalysis / A. Westfeld // Proceedings of 4th International Workshop Information Hiding. – 2001. – P. 289-302.
11. **Kavitha, V.** Neural based steganography / V. Kavitha, K.S. Easwarakumar. – In: PRICAI 2004: Trends in Artificial Intelligence / ed. by Ch. Zhang, H.W. Guesgen, W.K. Yeap. – Berlin, Heidelberg: Springer-Verlag, 2004. – P. 429-435.
12. **Chang, Ch.-Y.** Using counter-propagation neural network for robust digital audio watermarking in DWT domain / Ch.-Y. Chang, W.-Ch. Shen, H.-J. Wang // 2006 IEEE International Conference on Systems, Man and Cybernetics. – 2006. – Vol. 2. – P. 1214-1219. – DOI: 10.1109/ICSMC.2006.384880.
13. **Sirota, A.A.** Neural network functional models and algorithms for information conversion in order to create digital watermarks

/ A.A. Sirota, M.A. Dryuchenko, E.Y. Mitrofanova // Radioelectronics and Communications Systems. – 2015. – Vol. 58, Issue 1. – P. 1-10. – DOI: 10.3103/S073527271501001X.
14. **Сирота, А.А.** Обобщённые алгоритмы сжатия изображений на фрагментах произвольной формы и их реализация с использованием искусственных нейронных сетей / А.А. Сирота, М.А. Дрюченко // Компьютерная оптика. – 2015. – Т. 39, № 5. – С. 751-761. – DOI: 10.18287/0134-2452-2015-39-5-751-761.
15. **Сирота, А.А.** Анализ устойчивости алгоритмов создания цифровых водяных знаков с использованием универсальных сжимающих преобразований по отношению к негативным воздействиям различных видов / А.А. Сирота, Е.Ю. Митрофанова, М.А. Дрюченко // Вестник воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2015. – № 3. – С. 111-121.
16. **Осовский, С.** Нейронные сети для обработки информации / С. Осовский; пер. с польск. – М.: Финансы и статистика, 2002. – 344 с. – ISBN: 5-279-02567-4.
17. **Wang, Z.** A universal image quality index / Z. Wang, A.C. Bovik // IEEE Signal Processing Letters. – 2002. – Vol. 9, Issue 3. – P. 81-84. – DOI: 10.1109/97.995823.
18. **Kutter, M.** Digital signature of color images using amplitude modulation / M. Kutter, F.D. Jordan, F. Bossen // Proceedings of SPIE. – 1997. – Vol. 3022. – P. 518-526. – DOI: 10.1117/12.263442.
19. **Koch, E.** Towards robust and hidden image copyright labeling / E. Koch, J. Zhao // Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing. – 1995. – P. 452-455.
20. **Divecha, N.** Implementation and performance analysis of DCT-DWT-SVD based watermarking algorithms for Color images / N. Divecha, N.N. Jani // International Conference on Intelligent Systems and Signal Processing (ISSP). – 2013. – P. 204-208. – DOI: 10.1109/ISSP.2013.6526903.
21. Kodak lossless true color image suite [Электронный ресурс]. – URL: http://r0k.us/graphics/kodak/ (дата обращения 03.01.2018).
22. Testimages. Free collection of digital images for testing [Электронный ресурс]. – URL: https://testimages.org/ (дата обращения 03.01.2018).

### *Сведения об авторах*

**Сирота Александр Анатольевич**, 1954 года рождения, в 1976 году окончил Воронежский государственный университет по специальности «Радиофизика и электроника». Доктор технических наук (1995 год), профессор, заведует кафедрой технологий обработки и защиты информации Воронежского государственного университета. Область научных интересов: синтез и анализ систем сбора и обработки информации, методы и технологии компьютерного моделирования информационных процессов и систем, системный анализ в сфере информационной безопасности, компьютерная обработка изображений, нейронные сети и нейросетевые технологии в системах принятия решений. E-mail: *sir@cs.vsu.ru* .

**Дрюченко Михаил Анатольевич,** 1985 года рождения, в 2007 году окончил Воронежский государственный университет по специальности «Прикладная математика и информатика». Доцент кафедры технологий обработки и защиты информации Воронежского государственного университета. Область научных интересов: компьютерная стеганография и стегоанализ, компьютерная обработка изображений, программирование. E-mail: *m_dryuchenko@mail.ru* .

**Митрофанова Елена Юрьевна,** 1987 года рождения, в 2009 году окончила Воронежский государственный университет по специальности «Прикладная математика и информатика». Доцент кафедры технологий обработки и защиты информации Воронежского государственного университета. Область научных интересов: алгоритмы и технологии создания цифровых водяных знаков, нейросетевые технологии обработки информации. E-mail: *mitrofanova.e_yu@mail.ru* .

*ГРНТИ: 81.96.00, 28.23.37.*
*Поступила в редакцию 19 февраля 2018 г. Окончательный вариант – 8 мая 2018 г.*

## DIGITAL WATERMARKING METHOD BASED ON HETEROASSOCIATIVE IMAGE COMPRESSION AND ITS REALIZATION WITH ARTIFICIAL NEURAL NETWORKS

*A.A. Sirota [1], M.A. Dryuchenko [1], E.Yu. Mitrofanova [1]*
*[1] Voronezh State University, Voronezh, Russia*

### *Abstract*

In this paper, we present a digital watermarking method and associated algorithms that use a heteroassociative compressive transformation to embed a digital watermark bit sequence into blocks (fragments) of container images. A principal feature of the proposed method is the use of the heteroassociative compressing transformation – a mutual mapping with the compression of two neighboring image regions of an arbitrary shape. We also present the results of our experiments, namely the dependencies of quality indicators of thus created digital watermarks, which show the container distortion level, and the probability of digital watermark extraction error. In the final section, we analyze the performance of the proposed digital watermarking algorithms under various distortions and transformations aimed at destroying the hidden data, and compare these algorithms with the existing ones.

*Keywords:* data compression, image processing, neural networks, steganography, digital watermarks.

*Citation*: Sirota AA, Dryuchenko MA, Mitrofanova EYu. Digital watermarking method based on heteroassociative image compression and its realization with artificial neural networks. Computer Optics 2018; 42(3): 483-494. DOI: 10.18287/2412-6179-2018-42-3-483-494.

### *References*

[1] Younes MAB, Jantan A. A new steganography approach for image encryption exchange by using the least significant bit insertion. Inter J Comp Sci Network Security 2008; 8(6): 247-254.

[2] Hadhoud MM, Ismail NA, Shawkey W, Mohammed AZ. Secure perceptual data hiding technique using information theory. ICEEC '04 2004: 249-253. DOI: 10.1109/ICEEC.2004.1374433.

[3] Mandal JK, Sengupta M. Steganographic technique based on minimum deviation of fidelity. 2011 Second International Conference on Emerging Applications of Information Technology 2011; 298-301. DOI: 10.1109/EAIT.2011.24.

[4] Wu H-C, Wu N-I, Tsai C-S. Image Steganographic scheme based on pixel-value differencing and LSB replacement methods. IEE Proceedings – Vision, Image and Signal Processing 2005; 152(5): 611-615. DOI: 10.1049/ip-vis:20059022.

[5] Kawaguchi E, Eason RO. Principle and applications of BPCS-Steganography. Proc SPIE 1998; 3524: 464-473. DOI: 10.1117/12.337436.

[6] Maya ST, Miyatake MN, Medina RV. Robust steganography using bit plane complexity segmentation. 1st International Conference on Electrical and Electronics Engineering 2004: 40-43. DOI: 10.1109/ICEEE.2004.1433845.

[7] Darmstaedter V, Delaigle J-F, Quisquater JJ, Macq B. Low cost spatial watermarking. Computers and Graphics 1998; 22(4): 417-424. DOI: 10.1016/S0097-8493(98)00031-4.

[8] Langelaar GC, van der Lubbe JCA, Lagendijk RL. Robust labeling methods for copy protection of images. Proc SPIE 1997; 3022: 298-309. DOI: 10.1117/12.263418.

[9] Provos N, Honeyman P. Hide and seek: An introduction to steganography. IEEE Security & Privacy 2003; 1(3): 32-44. DOI: 10.1109/MSECP.2003.1203220.

[10] Westfeld A. F5-A steganographic algorithm: High capacity despite better steganalysis. Proceedings of 4th International Workshop Information Hiding 2001: 289-302.

[11] Kavitha V, Easwarakumar KS. Neural based steganography. In Book: Zhang Ch, Guesgen HW, Yeap WK, eds. PRICAI 2004: Trends in Artificial Intelligence. Berlin, Heidelberg: Springer-Verlag; 2004: 429-435.

[12] Chang Ch-Y, Shen W-Ch, Wang H-J. Using counterpropagation neural network for robust digital audio watermarking in DWT domain. 2006 IEEE International Conference on Systems, Man and Cybernetics 2006; 2: 1214-1219. DOI: 10.1109/ICSMC.2006.384880.

[13] Sirota AA, Dryuchenko MA, Mitrofanova EY. Neural network functional models and algorithms for information conversion in order to create digital watermarks. Radioelectronics and Communications Systems 2015; 58(1): 1-10. DOI: 10.3103/S073527271501001X.

[14] Sirota AA, Dryuchenko MA. Generalized image compression algorithms for arbitrarily-shaped fragments and their implementation using artificial neural networks. Computer Optics 2015; 39(5): 751-761. DOI: 10.18287/0134-2452-2015-39-5-751-761.

[15] Sirota AA, Mitrofanova EY, Dryuchenko MA. Analysis of the stability of the digital watermarking algorithms based on universal compression to the negative impact of various types [In Russian]. Proceedings of Voronezh State University, Series: Systems Analysis and Information Technologies 2015; 3: 111-121.

[16] Osowski S. Sieci neuronowe do przetwarzania informacji [In Polish]. Warsaw, Poland: Oficyna Wydawnicza Politechniki Warszawskiej; 2000. ISBN: 978-83-7207-615-1.

[17] Wang Z, Bovik AC. A universal image quality index. IEEE Signal Processing Letters 2002; 9(3): 81-84. DOI: 10.1109/97.995823.

[18] Kutter M, Jordan FD, Bossen F. Digital signature of color images using amplitude modulation. Proc SPIE 1997; 3022: 518-526. DOI: 10.1117/12.263442.

[19] Koch E, Zhao J. Towards robust and hidden image copyright labeling. Proc IEEE Workshop on Nonlinear Signal and Image Processing 1995: 452-455.

[20] Divecha N, Jani NN. Implementation and performance analysis of DCT-DWT-SVD based watermarking algorithms for Color images. International Conference on Intelligent Systems and Signal Processing (ISSP) 2013: 204-208. DOI: 10.1109/ISSP.2013.6526903.

[21] Kodak lossless true color image suite. Source: ⟨http://r0k.us/graphics/kodak/⟩.

[22] Testimages. Free collection of digital images for testing. Source: ⟨https://testimages.org/⟩.

### *Author's information*

**Alexander Anatolievich Sirota** (1954) graduated from Voronezh State University in 1976 majoring in "Radiophysics and Electronics". Professor, Doctor of Technical Sciences (since 1995). Currently head of Information Processing and Security Technologies chair at Voronezh State University. Research interests: analysis and design of information collection and processing systems, methods and techniques of information processes and systems computer modeling, system analysis in information security, digital image processing, neural networks and neural network technologies in decision-making systems. E-mail: *sir@cs.vsu.ru* .

**Mikhail Anatolievich Dryuchenko** (1985) graduated from Voronezh State University in 2007, majoring in Applied Mathematics and Informatics. Currently docent at the Information Processing and Security Technologies chair at Voronezh State University. Research interests: steganography and steganalysis, computer graphics processing, programming. E-mail: *m_dryuchenko@mail.ru* .

**Elena Yurievna Mitrofanova** (1987) graduated from Voronezh State University in 2009, majoring in Applied Mathematics and Informatics. Currently docent at the Information Processing and Security Technologies chair at Voronezh State University. Research interests: steganography and digital watermarking, neural networks and neural network technologies. E-mail: *mitrofanova.e_yu@mail.ru* .

*Received February 19, 2018. The final version – May 08, 2018.*