УДК 681.322:621.391

А. С. Васюра, к. т. н., проф.; В. В. Лукичёв

МЕТОД ШАБЛОННОГО ВСТРАИВАНИЯ ДАННЫХ В ВЕЙВЛЕТ-КОЭФФИЦИЕНТЫ НА ОСНОВЕ КРИТЕРИЯ СТЕГАНОГРАФИЧЕСКОЙ СТОЙКОСТИ

Рассмотрены особенности встраивания, определяющие секретность и стойкость скрытых данных. С целью разработки эффективного стеганографического метода синтезирован критерий, который использован в качестве целевой функции при встраивании.

***Ключевые слова**: стеганография, JPEG-алгоритм, метод шаблонного встраивания, секретность, робастность, вейвлет-преобразование, метод опорных векторов.*

Стеганография изображений является областью, которая стремительно развивается на протяжении последних десяти лет. Ее цель может быть обозначена как секретное и стойкое к разнообразным преобразованиям сокрытие данных. Соответственно практические задачи, которые решаются в ее пределах, в большей или меньшей степени касаются аспектов секретности и робастности [1, 2].

Поскольку нарушение секретности может привести к полной потере сообщения, то именно это описанное качество определяет основные ограничения при проектировании стегосистемы. Надо отметить, что относительный характер этого показателя обуславливает существование большого количества критериев, эффективность которых неодинакова для разных методов встраивания.

Другим важным аспектом является требование стойкости. Поскольку широко распространенны схемы избыточного кодирования с защитой от ошибок, то вопрос робастности может быть решён с эффективностью, которая определяется достоверностью восстановленной информации [3].

Таким образом, проектирование любой стегосистемы можно рассматривать как задачу условной оптимизации, где целевая функция определенным образом связывает робастность со степенью секретности, а ограничения определяют область адекватности критерия. Такой универсальный подход позволит обеспечить высокую адаптивность к условиям непосредственного функционирования стегосистемы.

В стеганографии изображений особенно распространена схема слепого встраивания, где передается лишь стегоконтейнер. Это определяет особенности стегоанализа, задача которого состоит в бинарной классификации изображений на основе свойств, которые претерпевают наибольшие изменения при встраивании. Особенно перспективными являются критерии на основе метода опорных векторов (SVM – support vector machines) [4], наибольшим преимуществом которых является эффективность классификации точек-характеристик в многомерном пространстве признаков.

Среди методов обработки изображений наибольшей популярностью пользуются методы сжатия. Наибольший коэффициент уплотнения способны обеспечить методы сжатия с потерями [5]. Стандарт сжатия JPEG и до сих пор используется широко, несмотря на внедрение более эффективных форматов на основе вейвлет-преобразований (например, JPEG2000). Такая ситуация, очевидно, обусловлена инертностью концепций разработки программного обеспечения в этой сфере, которая в свою очередь разрешает прогнозировать значительную продолжительность перехода.

Поэтому в качестве основного фактора влияния на стегоконтейнер рассматривается обработка JPEG. С другой стороны, стеганографическое использование вейвлет-преобразований дает основания надеяться на незаметность внесенных изменений. С помощью разработанного критерия предлагается исследовать комплексную связь между секретностью и робастностью указанного использования в области вейвлет-преобразований.

Коэффициенты решено модифицировать соответственно распространенному подходу векторной квантизации. Его разновидностью является шаблонная схема встраивания, для которой значение тайной порции данных зависит от соотношений набора коэффициентов с определенным эталонным значением [6]. Основное преимущество шаблонной схемы – возможность многовариантного представления порции тайных данных, которая позволяет повысить их стойкость.

Во время разработки современных методов сокрытия в большинстве случаев оптимизируется одно из качеств секретности или робастности. Использование критерия, который объединяет определенные качества, должно повысить эффективность стегозащиты. Аспект актуальности не исчерпывается лишь данным критерием: предложен адаптивный путь его улучшения. Для этого учитываются свойства каждого объекта стеганографического манипулирования, который несет элементарную частицу тайных данных.

Предполагается, что особенности предложенного в статье подхода обеспечат высокую эффективность стегометода на его основе. Разработка такого метода является **целью** данного **исследования**.

Критерий стеганографической эффективности. Комплексную оценку эффективности стегометода предлагается осуществлять с использованием независимых показателей секретности и робастности. Мера робастности определена как частица сохраненных элементарных порций тайных данных после обработки стегоизображения. В качестве критерия секретности избран стегоаналитический критерий, предложенный в [4]. Он использует SVM для классификации изображений.

Основная идея метода опорных векторов – перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве (рис. 1). Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает на основе предположения, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.

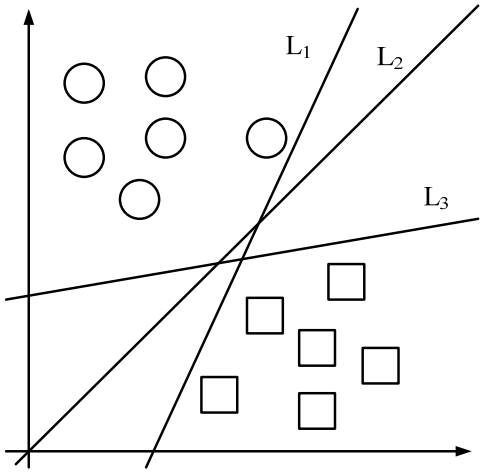

Рис. 1. Несколько классифицирующих прямых (гиперплоскостей)

Пусть точки имеют вид: $\{(x_1,c_1),(x_2,c_2),...,(x_n,c_n)\}$, где c_i принимает значение 1 или −1 в зависимости от того, к какому классу принадлежит точка x_i. Каждое x_i – это p-мерный вещественный вектор, обычно нормализованный значениями [0,1] или [-1,1]. Если точки не будут нормализованы, то точка с большими отклонениями от средних значений координат точек слишком очень повлияет на классификатор. Рассмотрим это как учебную коллекцию, в

которой для каждого элемента уже задан класс, к которому он принадлежит. Необходимо, чтобы алгоритм метода опорных векторов классифицировал их таким же образом. Для этого строим разделяющую гиперплоскость, которая имеет вид: $w \cdot x - b = 0$.

Вектор w – перпендикуляр к разделяющей гиперплоскости. Параметр b зависит от кратчайшего расстояния гиперплоскости до начала координат. Если параметр b равен нулю, гиперплоскость проходит через начало координат, что ограничивает решение.

При оптимальном разделении опорные вектора и гиперплоскости параллельны оптимальной (рис. 2). Можно показать, что эти параллельные гиперплоскости могут быть описаны следующими уравнениями (с точностью до нормировки): $w \cdot x - b = 1$, $w \cdot x - b = -1$.

Если учебная коллекция линейно разделима, то можем выбрать гиперплоскости таким образом, чтобы между ними не лежала ни одна точка обучающей выборки, и затем максимизировать расстояние между гиперплоскостями. Ширину полосы между ними легко найти из соображений геометрии, она равна $\frac{2}{\|w\|}$, таким образом минимизировать $\|w\|$. Чтобы исключить все точки из полосы, должны выполниться для всех i условия: $\begin{cases} w \cdot x_i - b \geq 1 \\ w \cdot x_i - b \leq -1 \end{cases}$

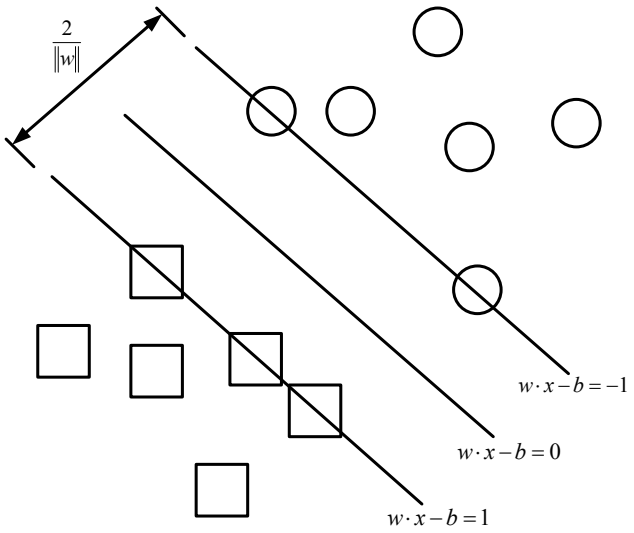

Рис. 2. Оптимальная разделяющая гиперплоскость для метода опорных векторов, построенная на точках из двух классов

Это может быть также записано в виде:

$$c_i(w \cdot x_i - b) \geq 1, \ 1 \leq i \leq n. \quad (1)$$

В случае линейной разделимости проблема построения оптимальной разделяющей гиперплоскости сводится к минимизации $\|w\|$, при условии (1). Это задача квадратичной оптимизации, которая имеет вид: $\begin{cases} \|w\|^2 \to \min \\ c_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n. \end{cases}$

По теореме Куна – Такера эта задача эквивалентна двойственной задаче поиска седловой точки функции Лагранжа

$$\begin{cases} L(w,b;\lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \lambda_i c_i (w_i \cdot x_i) \to \min_{u,b} \max_{\lambda} \\ \lambda_i \geq 0, \quad 1 \leq i \leq n \\ \begin{bmatrix} \lambda_i = 0 \\ w \cdot x_i - b = c_i, \end{bmatrix} \quad 1 \leq i \leq n \end{cases} \tag{2}$$

где $\lambda = (\lambda_1,...,\lambda_n)$ – вектор двойственных переменных.

Сведем эту задачу к эквивалентной задаче квадратичного программирования, содержащую только двойственные переменные:

$$\begin{cases} -L(\lambda) = -\sum_{i=1}^{n} \lambda_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j (c_i c_j (x_i \cdot x_j - b) - 1) \to \min_{\lambda} \\ \lambda_i \geq 0, \quad 1 \leq i \leq n \\ \sum_{i=1}^{n} \lambda_i c_i = 0. \end{cases} \tag{3}$$

Допустим, что решили данную задачу, тогда w и b можно найти по формулам: $w = \sum_{i=1}^{n} \lambda_i c_i x_i$, $b = w \cdot x_i - c_i$.

Алгоритм классификации может быть записан в виде:

$$a(x) = sign\left(\sum_{i=1}^{n} \lambda_i c_i x_i \cdot x - b \right) \tag{4}$$

Обратим внимание, что суммирование идет не по всей выборке, а только по опорным векторам, для которых $\lambda_i \neq 0$.

Для того, чтобы алгоритм мог работать в случае, когда классы линейно неразделимы, позволим ему допускать ошибки на учебной коллекции. Введем набор дополнительных переменных $\xi_i \geq 0$, характеризующих величину ошибки на объектах x_i, $1 \leq i \leq n$. Возьмем в качестве отправной точки (2), смягчим ограничения неравенства, так же введем в минимизируемый функционал штраф за суммарную ошибку:

$$\begin{cases} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \to \min_{\varpi,b,\xi_i} \\ c_i (w \cdot x_i - b) \geq 1 - \xi_i, \quad 1 \leq i \leq n \\ \xi_i \geq 0, \quad 1 \leq i \leq n. \end{cases}$$

Коэффициент C – параметр настройки метода, который позволяет регулировать отношение между максимизацией ширины разделяющей полосы и минимизацией суммарной ошибки.

Аналогично, по теореме Куна – Такера сводим задачу к поиску седловой точки функции Лагранжа:

$$\begin{cases} L(w,b;\xi,\lambda,\eta)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}\lambda_i(c_i(w_i\cdot x_i)-1)-\sum_{i=1}^{n}\xi_i(\lambda_i+\eta_i-C)\to\min_{u,b,\xi}\max_{\lambda,\eta} \\ \xi_i\ge 0,\ \lambda_i\ge 0,\ \eta_i\ge 0,\quad 1\le i\le n \\ \left[\begin{array}{l}\lambda_i=0\\ c_i(w\cdot x_i-b)=1-\xi_i,\end{array}\right. \quad 1\le i\le n \\ \left[\begin{array}{l}\eta_i=0\\ \xi_i=c_i,\end{array}\right. \quad 1\le i\le n. \end{cases}$$

По аналогии сведем эту задачу к эквивалентной:

$$\begin{cases} -L(\lambda)=-\sum_{i=1}^{n}\lambda_i+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j c_i c_j(x_i\cdot x_j)\to\min_{\lambda} \\ 0\le\lambda_i\le C,\quad 1\le i\le n \\ \sum_{i=1}^{n}\lambda_i c_i=0. \end{cases}$$

На практике для построения машины опорных векторов решают именно эту задачу, а не (3), так как гарантировать линейную разделимость точек на два класса в общем случае не представляется возможным.

Таким образом, для каждого стегоаналитического критерия связь между PSNR и энтропией детектирования $e^{\det}=-p\log p-\overline{p}\log\overline{p}$ является прямой, где $\overline{p}=1-p$, p – вероятность верной классификации. Для данного критерия экспериментально установлена высокая корреляция между этими показателями. Поэтому в дальнейшем эта связь рассматривается по умолчанию.

Оценка эффективности встраивания предусматривает учет следствий определенных характерных влияний со стороны третьего лица. В случае применения JPEG-компрессии, результат зависит от параметров сжатия, которые задаются пользователем. Квантование коэффициентов ДКП описывается зависимостью

$$dct_{i,j}^{jpeg}=\frac{Q_{i,j}}{q}\,\mathrm{round}\left(\frac{dct_{i,j}}{Q_{i,j}}q\right), \tag{5}$$

где $Q_{i,j}$ – соответствующий элемент матрицы квантования Q, $i,j=1\ldots8$, q – параметр, который задается пользователем и определяет качество и размер сжатого изображения [5]. Конечно, невозможно в каждом конкретном случае предусмотреть значение q, однако использование статистического распределения f_q позволяет перейти к обоснованной оценке. Рис. 3 отображает типичное распределение f_q. Поскольку результат обработки JPEG-алгоритмом (квантование) зависит от значений коэффициентов ДКП, то стойкость встроенных данных для разных блоков изображения будет разной. Такое же замечание касается количественной меры искажений встраивания. При JPEG-сжатии блоки изображения 8×8 обрабатываются независимо. Поэтому при условии независимого встраивания в эти блоки, можно получить адаптивную к требованиям секретности и робастности стегосистему.

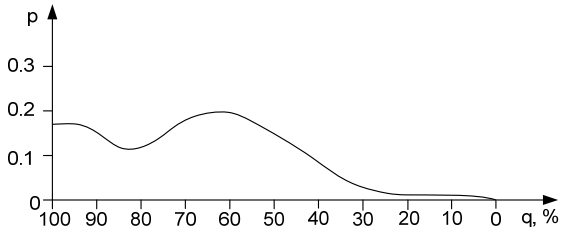

Рис. 3. Типичное распределение значений параметра q

Критерий эффективности встраивания должен быть интегральным, поскольку вместо значения q известно лишь распределение f_q. Таким образом определенному i-му условию квантования, которое полностью определяется q_i, соответствует вероятность f_{q_i} и комплексная характеристика эффективности системы z_i. Если z_i определить как произведение стегоаналитической энтропии детектирования $e_i^{\det}$ и показателя робастности $r_i = 1 - \mathrm{BER}_i$, где BER – показатель битовых ошибок, то критерий общей эффективности встраивания можно представить выражением:

$$E = \sum_i z_i f_{q_i} = \sum_i e_i^{\det} r_i f_{q_i} . \quad (6)$$

Учитывая, что значения показателей $e_i^{\det}$ и r_i являются зависимыми от энергии встраивания $d = \left\| I^{org} - I^{stg} \right\|^2$ (искажение стегоизображения I^{stg} по сравнению с оригинальным I^{org}), предыдущее выражение принимает вид:

$$E(d) = \sum_i e_{i,d}^{\det} r_{i,d} f_{q_i} . \quad (7)$$

Для случая непрерывного изменения условий квантования имеем:

$$E(d) = \int e^{\det}(q,d) r(q,d) f(q) dq . \quad (8)$$

Однако предложенный адаптивный подход требует дополнительного определения критерия эффективности встраивания. По вышеупомянутому предположению $e^{\det}(q,d)$ является однозначной функцией. Для большинства популярных стегометодов это касается и показателя робастности $r(q,d)$. В случае адаптивного встраивания, аргументов (q,d) недостаточно для адекватного представления уровня робастности, поскольку каждый из объектов стеганографического манипулирования может испытывать неоднозначное влияние. Поэтому ключевым моментом максимизации $E(d)$ будет поиск $r(q,d,\Omega)$, где $\Omega = \{\Omega_j\}$, $j = 1 \ldots m$, Ω_j – вектор состояния j-го объекта. Конечная задача проектирования стегометода формализируется:

$$\max_d \left(\max_\Omega \int e^{\det}(q,d) r(q,d,\Omega) f(q) dq \right). \quad (9)$$

Очевидно, эффективность встраивания будет определяться не только методами оптимизации при решении поставленной выше задачи. Способ встраивания (схема) в первую очередь задает ограничение и существенно влияет на результат [2]. Хотя предложенный подход можно соединить с любой схемой, решено использовать шаблонную. Этот выбор объясняется высокой степенью свободы манипулирования.

Эксперимент. Целью эксперимента является сравнение эффективности встраивания данных разработанным методом и методами, которые широко используются на практике. Для сравнение избраны: метод последнего значущего бита (ПЗБ) [7], шаблонный метод на

основе целочисленого вейвлет-преобразования (IWT) [6] и метод, который оперирует в области ДКП [8]. В избранные изображения по единственному стегоключу были встроены секретные данные. Эффективность методов определялась по двум зависимостям: секретность стегоманипуляций и робастность встроенных данных от параметра q, который задает степень сжатия. Поскольку разработка метода велась на основе предложенного критерия эффективности встраивания, то сравнение с остальными методами по этому критерию и упомянутыми выше зависимостями позволит установить адекватность критерия.

Согласно описанным особенностям проектирования стегометода для постановки и решения задачи оптимизации встраивания необходимо предварительно определить распределение $f(q)$ и функцию стегоаналитической энтропии детектирования $e^{\det}(q,d)$. Зависимость $f(q)$ установлена путем экспертного распознавания популярных и широко используемых изображений в градациях серого размером 256×256, который в зависимости от потребностей рассмотренных web-страниц обрабатывались JPEG алгоритмом с разным значением параметра q. При определении $e^{\det}(q,d)$ для каждого q_i (значение q_i изменялись от 1 до 0.65 с шагом 0.05) были сформированы учебная и тестовая выборки. Первая использовалась для тренировки SVM согласно предложенному в [4] вектора характеристик, на второй определялась средняя вероятность верного детектирования в зависимости от значения искажений d. Изображение в учебной и тестовой выборках не совпадают. Каждая выборка наполовину состоит из оригинальных изображений (количеством 400), остальные – стегоизображения, полученные из оригинальных с помощью описанной шаблонной схемы встраивания.

Вследствие проведения описанных этапов оптимизации встраивания 2000 бит тайных данных в вейвлет-коэффициенты Хаара согласно критерию E, количественный показатель эффективности, который является средним для 20 изображений, составил 0.63. Для описанного в [6] метода на основе целочисленного вейвлет-базиса значение критерия составляет 0.48, стеганографическая эффективность метода [7] на основе ОЗБ – 0.28, эффективность встраивания в область ДКП [8] оценивается 0.42.

С целью демонстрации адекватности критерия и эффективности разработанного метода, приводятся два графика зависимостей вероятности детектирования $p^{\det}$ от q (рис. 4) и робастности встраивания r от q (рис. 5), которые наглядно показывают преимущества и недостатки каждого из оцененных выше методов.

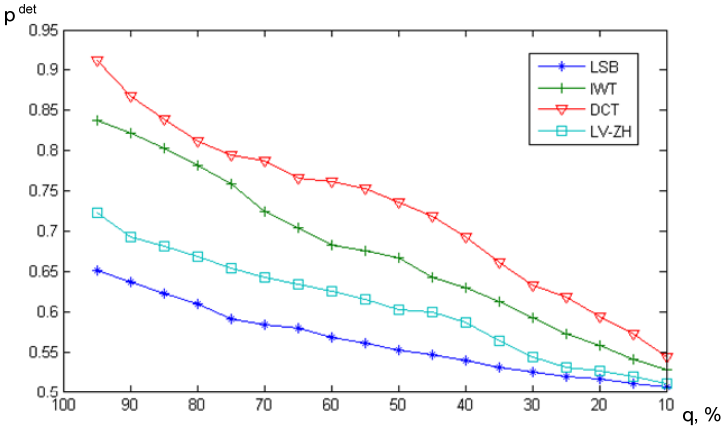

Рис. 4. Зависимость вероятности детектирования $p^{\det}$ от параметра качества JPEG-сжатия q

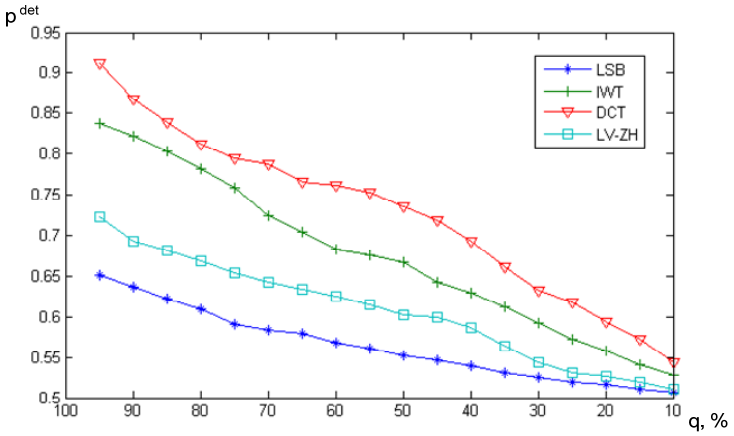

Рис. 5. Связь робастности *r* встроенных данных от параметра *q*

Выводы. Разработан стеганографический метод, который использует принцип шаблонного встраивания в области вейвлет-коэффициентов. Особенностью метода является учет требований секретности и робастности к JPEG-преобразованию, которая реализована путем их объединения с помощью предложенного критерия.

Предложенный подход позволяет повысить общую эффективность встраивания данных, которая подтверждена экспериментально при сравнении с популярными стегометодами. Недостатком метода является сложность, обусловленная дифференционной особенностью встраивания и, как следствие, необходимостью итеративного решения численных задач оптимизации.

СПИСОК ЛИТЕРАТУРЫ

1. Грибунин В. Г. Цифровая стеганография / В. Г. Грибунин, И. Н. Оков, И. В. Туринцев. – СПб.: Солон-Пресс, 2002. – 272 с.

2. Johnson N. F. Information Hiding: Steganography and Watermarking – Attacks and Countermeasures / N. F. Johnson, Z. Duric, S. Jajodia. – Berlin: Springer, 2001. – 160 p.

3. Glavieux A. Channel Coding in Communication Networks / A. Glavieux. – London: Hermes Science Pub. Ltd., 2007. – 416 p.

4. Zou D. Steganalysis based on Markov Model of Thresholded Prediction-Error Image / D. Zou, Y. Shi, W.Su, G. Xuan // IEEE ICME. – 2006. – № 1. – P. 1365 − 1368.

5. Pennebaker W. JPEG: Still Image Compression Standard / W. Pennebaker, J. Mitchell. – NY.: Kluwer Academic Pub., 1993. – 650 p.

6. Метод вбудовування даних на основі алгоритму вейвлет-стиснення зображень: праці конференції, 25 − 28 вер. 2006 р., Вінниця. Т. 1 / Відп. ред. В. М. Дубовой. – Вінниця: Универсум-Вінниця, 2007. – С. 491 − 495.

7. Wu H.C. Image Steganographic Scheme Based on Pixel-Value Differencing and LSB Replacement Methods / H.C. Wu, N.I. Wu, C.S. Tsai, M.S. Hwang // IEEE Transactions on Image and Signal Processing. – 2005. – № 5. – P. 611 − 615.

8. Quan L. Combination of DCT-Based and SVD-Based Watermarking Scheme / L. Quan, A. Qingsong // IEEE ICSP Conference Record. – 2004. – № 1. – P. 873 − 876.

Васюра Анатолий Степанович – директор института, профессор кафедры автоматики и информационно-измерительной техники;

Лукичёв Виталий Владимирович – соискатель кафедры автоматики и информационно-измерительной техники.

Винницкий национальный технический университет.