СТАТИСТИЧЕСКИЙ АНАЛИЗ АЛГОРИТМОВ ВСТРАИВАНИЯ ИНФОРМАЦИИ В ГРАФИЧЕСКИЕ ИЗОБРАЖЕНИЯ С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТ-АНАЛИЗА

В. И. Лобач

Белорусский государственный университет
Минск, Беларусь
E-mail: lobach@bsu.by

Исследуется проблема обнаружения встроенного в цифровое изображение сообщения статистическими методами на основе целочисленного вейвлет-преобразования.

Ключевые слова: вейвлет-преобразование, вейвлет Хаара, модифицированный стеганоконтейнер, скрытое сообщение.

ВВЕДЕНИЕ

Развитие средств вычислительной техники дало мощный толчок для развития компьютерной стеганографии. Появилось много новых областей применения. Большинство исследований так или иначе связаны с цифровой обработкой сигналов. Сообщения встраиваются в цифровые данные, имеющие цифровые данные, имеющие аналоговую природу и речь аудиозаписи, изображения, видео [1, 2]. Известны также работы по встраиванию информации в текстовые файлы и в исполняемые файлы программ. В данной работе рассматриваются методы статистического анализа алгоритмов встраивания в цифровые графические изображения с использованием целочисленного вейвлет-преобразования.

ЦЕЛОЧИСЛЕННОЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ ЦИФРОВОГО ИЗОБРАЖЕНИЯ

Цифровое изображение I представляет собой $(M \times N)$-матрицу действительных чисел

$$I = \begin{pmatrix} I_{11} & I_{12} & \cdots & I_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ I_{M1} & I_{M2} & \cdots & I_{MN} \end{pmatrix}. \quad (1)$$

Вейвлет-преобразование матрицы I получается следующим образом:

- дискретное вейвлет-преобразование применяется к каждой строке матрицы I, в результате чего генерируется новая матрица;

- дискретное вейвлет-преобразование применяется к сгенерированной на предыдущем шаге матрице, но теперь ко всем столбцам.

Получаются четыре матрицы, каждая из которых имеет размерность $M / 2 \times N / 2$

$$\begin{pmatrix} D^1 & C^1 \\ B^1 & A^1 \end{pmatrix}. \qquad (2)$$

Подматрица A^1 матрицы (2) представляет собой сжатое (огрубленное) исходное изображение (с так называемыми низкочастотными компонентами). Подматрица B^1 сохраняет горизонтальные детали изображения, подматрица C^1 аналогична подматрице B^1 за исключением того, что она сохраняет вертикальные детали изображения (уточняющие коэффициенты). Подматрица D^1 содержит диагональные детали изображения.

На втором шаге проводим те же операции с подматрицей A^1, полученной на первом шаге, в результате чего получаются матрицы второго уровня A^2, B^2, C^2, D^2 и т.д.

Приведем формулы, определяющие элементы матриц *A*, *B*, *C*, *D*, если в качестве базового вейвлета выбран вейвлет Хаара [3]:

$$\begin{aligned} A_{ij} &= \left(I_{2i,2j} + I_{2i+1,2j}\right)/2, \\ B_{ij} &= I_{2i,2j+1} - I_{2i+2j}, \\ C_{ij} &= I_{2i+1,2j} - I_{2i,2j} \\ D_{ij} &= I_{2i+1,2j+1} + I_{2i,2j}, \end{aligned} \qquad (3)$$

где $1 \le i \le M/2$, $1 \le j \le N/2$.

Очевидно, что формулы (3) обратимы и мы можем однозначно восстановить исходное изображение по вычисленным коэффициентам вейвлет-преобразования. Обратное вейвлет-преобразование задается следующими формулами

$$\begin{aligned} I_{2i,2j} &= A_{ij} - B_{ij}/2, \\ I_{2i,2j+1} &= A_{ij} + B_{ij}/2, \\ I_{2i+1,2j} &= I_{2i,2j+1} + C_{ij} - B_{ij}, \\ I_{2i+1,2j+1} &= I_{2i,2j+1} + D_{ij} - C_{ij}. \end{aligned}$$

АЛГОРИТМЫ ВСТРАИВАНИЯ СООБЩЕНИЯ В ИЗОБРАЖЕНИЯ

На вход алгоритма встраивания поступает контейнер $I = \left(I_{ij}\right), i = \overline{1.M}, j = \overline{1.N}$, представляющий собой цветное изображение $M \times N$ пикселей, и скрываемое сообщение $m = \left(m_1,...,m_T\right), m_i \in \{0,1\}$, длины *T* бит. Общий алгоритм встраивания состоит из следующих шагов:

1) К изображению *I* согласно формулам (3) применяется дискретное вейвлет-преобразовние, глубина разложения *d* дается пользователем, рекомендуется брать не более 3–4-х уровней декомпозиции;

2) Выбирается подматрица коэффициентов, в которую будет встраиваться сообщение. По завершении декомпозиции на уровне глубины *d* возможно 4^d различных подматриц вейвлет-коэффициентов;

3) На основании полученных данных генерируется стегоключ *Key* = (*Y*, *T*, *K*), где $Y \in R^3$ – параметры генератора случайных величин, $K \in N^3$ – число уровней декомпозициии. Номер подматрицы для встраивания и размер блока коэффициентов;

4) На основании длины сообщения *T* задается число блоков изменяемых вейвлет-коэффициентов. Используя сгенерированный ранее параметр ключа *Y*, случайным образом выбираются номера блоков и их порядок, согласно которому будет производиться встраивание. Далее генерируется двоичный случайный образ, согласно которому будет

производиться встраивание. Далее генерируется двоичный случайный образ, согласно которому каждый блок E_i коэффициентов делится на два субблока E_{i0} и E_{i1}. Для каждого субблока вычисляются среднее значения l_{i0} и l_{i1}, выбирается некоторый порог α, и бит сообщения встраивается следующим образом:

$$l_{i0} - l_{i1} \geq \alpha, \text{ если } m_i = 1,$$
$$l_{i0} - l_{i1} < -\alpha, \text{ если } m_i = 0;$$

5) По формулам обратного вейвлет-преобразования вычисляется стеганограмма $I^* = \left(I_{ij}^*\right), i = \overline{1, M}, j = \overline{1, N}$.

Алгоритмы встраивания определяются выбором подматрицы коэффициентов вейвлет-преобразования.

ОЦЕНКА ЧИСЛА ПЕРЕХОДОВ ЗНАЧЕНИЙ МЛАДШИХ БИТ

Метод основывается на том факте, что между младшими битами соседних элементов, а также между ними и остальными битами в естественных контейнерах имеются корреляционные связи. При анализе графических файлов формата JPEG в качестве элементов анализируемой последовательности выбираются младшие биты соседних дискретных косинусных коэффициентов, отличных от 0 и 1.

Под «переходом» понимают переход значения i-го элемента последовательности в значение $(i + 1)$-го элемента последовательности x_t, $t = \overline{1, n-1}$, где n – длина последовательности. Так как последовательности являются двоичными, то анализируется четыре вида переходов: 0 → 0, 0 → 1, 1 → 0, 1 → 1. По полученным результатам строится гистограмма, где каждый столбец соответствует одному из переходов.

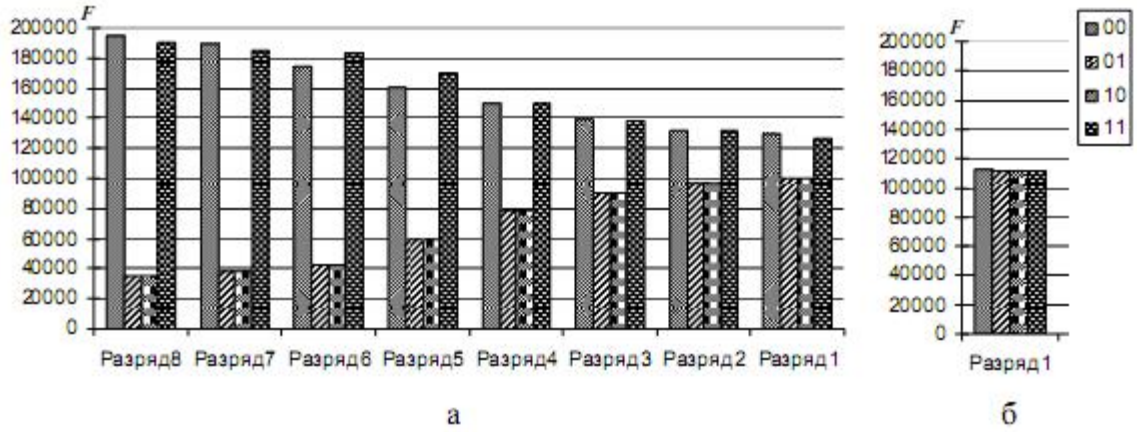


Рис. 1. Гистограмма частот переходов битовых значений:
а – пустого контейнера, б – стегоконтейнера

Для пустого и модифицированного контейнера число переходов в потоке НЗБ будет разным. Распределение НЗБ модифицированного контейнера имеет, как правило, случайный характер. Пустому контейнеру не свойственно примерно одинаковое число переходов в потоке НЗБ для всех состояний (рис. 1).

ОЦЕНКА ЧАСТОТ ПОЯВЛЕНИЯ БИТОВЫХ СЕРИЙ

Метод позволяет оценить равномерность распределения элементов в исследуемой последовательности на основе анализа частоты появления нулей и единиц, и серий, со-

стоящих из k бит. В битовом представлении исследуемой последовательности x_t подсчитывается, сколько раз встречаются нули и единицы (k = 1), серии-двойки (00, 01, 10, 11: (k = 2)), серии-тройки (000, 001, 010, 011, 100, 101, 110, 111: (k = 3)) и т. д. На основе результатов строится гистограмма. Для JPEG-изображений гистограмма строится по значениям частот появления битовых серий в потоке НЗБ дискретных косинусных коэффициентов, отличных от –1, 0, 1.

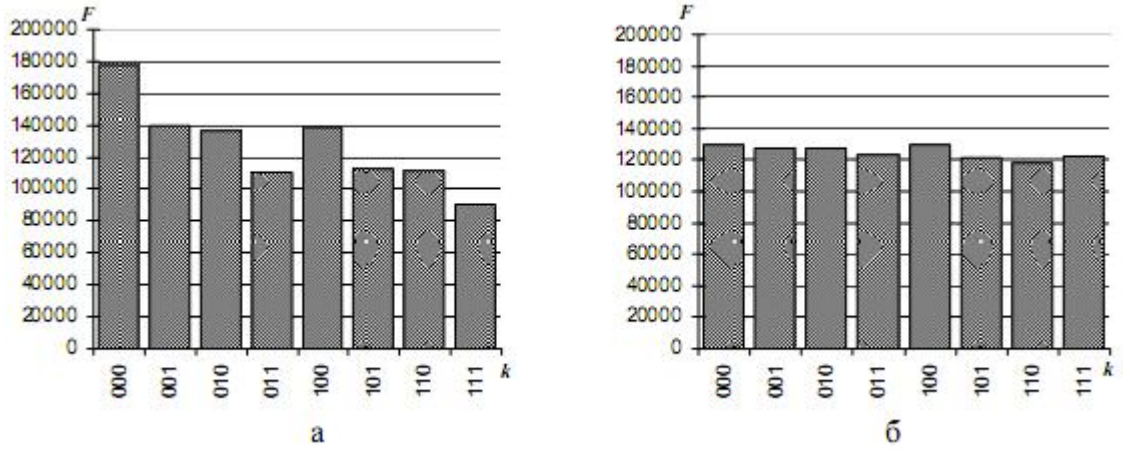


Рис. 2. Гистограмма частот серии-тройки (k = 3) в потоке НЗБ:
а – пустого контейнера, б – стегоконтейнера

Для незаполненных JPEG-изображений не является характерным, чтобы значения частот всех компонентов находились достаточно близко (рис. 2 а). При внедрении информации значения частот сближаются (рис. 2 б). Этот факт используется при анализе.

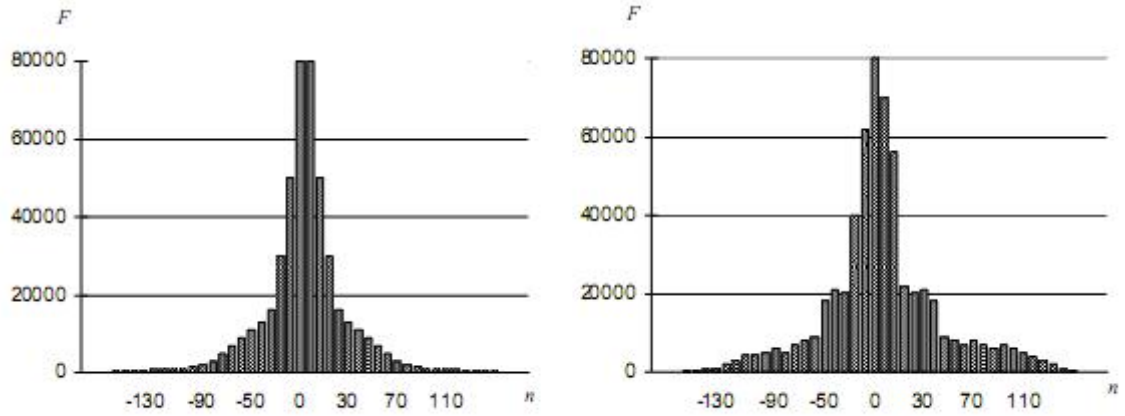
Результаты работы метода зависят от стеганографического преобразования, используемого для встраивания скрываемых данных, а также от их объема. Как правило, выявление факта скрытия осуществимо при заполнении контейнера на 60% и выше.

АНАЛИЗ ГИСТОГРАММ ЧАСТОТ ЭЛЕМЕНТОВ ИЗОБРАЖЕНИЯ

Метод позволяет оценить равномерность распределения элементов анализируемого изображения, а также определить частоту появления конкретного элемента. Если разброс частот появления элементов в цветовых составляющих изображения стремится к нулю, то контейнер содержит скрытые данные. В противном случае контейнер считается пустым.

Для изображений в JPEG-формате строится гистограмма частот квантованных дискретных косинусных коэффициентов. Экспериментально обнаружено, что огибающая гистограммы пустого изображения имеет более гладкий характер (рис. 3 а) по сравнению с гистограммами изображений, содержащими стеганографическое вложение (рис. 3 б).

Конечно, в зависимости от характера и степени сжатия изображения, гистограммы могут изменяться – в них могут появляться скачки и провалы, но важно то, что скрытие информации меняет общий вид гистограмм. Большинство стеганографических программ, работающих с JPEG, скрывают данные в младшие биты дискретных коэффициентов, отличных от 0 и 1. Как следствие, частоты 0-х и 1-х ДКП не изменяются, в то время как все остальные частоты либо уменьшаются, либо увеличиваются в зависимости от алгоритма встраивания. При значительных объемах скрываемой информации гистограммы часто приобретают ступенчатый характер, что нетипично для обычных JPEG-изображений.



а б

Рис. 3. Гистограмма частот дискретных косинусных коэффициентов:
а – исходного изображения, б – изображения, содержащего скрытую информацию

Проводилась компьютерная реализация указанных алгоритмов, в качестве контейнера использовались графические черно-белые изображения формата bmp. В качестве скрываемых данных использовались файлы формата txt размером от 40% до 80% от размеров контейнера – выявление факта скрытия данных в изображении было при объеме скрываемого сообщения более 50% от размеров сообщения.

ЛИТЕРАТУРА

1. *Грибунин, В. Г.* Цифровая стеганография / В. Г. Грибунин, И. Н. Оков. М. : Солон-Пресс, 2002. 272 с.

2. *Хорошко, В. А.* Введение в компьютерную стеганографию / В. А. Хорошко, М. Е. Шелест. Киев : Ми-Пресс, 2006. 178 с.

3. *Малла, С.* Вейвлеты в обработке сигналов / С. Малла. М. Мир, 2005. 671 с.