Встраивание информации в сжатые JPEG-изображения с минимизацией проявления демаскирующих признаков

О.О. Шумская, О.О. Евсютин
shumskaya.oo@gmail.com|eoo@keva.tusur.ru
Томский государственный университет систем управления и радиоэлектроники, Томск, Россия

Представлено исследование информативности признаков в пространственной и частотной областях сжатых JPEG-изображений при стегоанализе. Проведен отбор признаков с применением жадного алгоритма с исключением и получен набор информативных признаков, на основе которых определена целевая функция стеганографического встраивания. Встраивание осуществляется по алгоритму на основе операции замены. Зная, какие признаки несут информацию о наличии вложения в цифровом изображении, можно стремиться к минимизации искажений этих параметров с целью максимизации незаметности встраивания как визуально, так и для стегоанализаторов. Эксперименты показали, что такой подход позволяет при встраивании более чем 4000 бит в полутоновое изображение с разрешением 256×256 пикселей достичь минимальных искажений изображения, держа метрику визуального качества PSNR не ниже 52дБ, и обеспечивая устойчивость перед стегоанализатором.

Ключевые слова: *стеганография, стегоанализ, цифровые изображения, сжатые JPEG-изображения, демаскирующие признаки.*

Information embedding into compressed JPEG-images with minimization of manifestation of the unmasking features

O.O. Shumskaya, O.O. Evsutin
shumskaya.oo@gmail.com|eoo@keva.tusur.ru
Tomsk state university of control systems and radioelectronics, Tomsk, Russia

The research of informational content of features in the spatial domain and the frequency domain of compressed JPEG-images at the stegoanalisys is presented. Selection of features with application of the greedy algorithm with an exception is made and the set of informative features on the basis of which criterion function of steganographic embedding is defined is received. Embedding is carried out by means of the algorithm on the basis of replacement operation. Knowing what features contain information about existence of the secret message in the digital image, it is possible to seek for minimization of distortions of these parameters for the purpose of maximizing obscurity of embedding as visually, and for stegoanalyzer. Experiments have shown that such approach allows when embedding more than 4000 bits into the grayscale image with size of 256×256 pixels to reach the minimum distortions of the image, holding a metrics of visual quality PSNR not lower than 52 dB, and stability before the stegoanalyzer.

Keywords: *steganography, steganalisys, digital images, compressed JPEG-images, unmasking features.*

1. Введение

Сегодня трудно представить какую-либо сферу деятельности человека без использования цифровых изображений. Ежедневно в Интернет загружают миллионы изображений различного характера и содержания. Этот факт позволяет пользователям обмениваться информацией с использованием методов цифровой стеганографии, предназначенных для незаметного встраивания дополнительной информации в цифровые объекты.

Эффективность стеганографического встраивания информации в цифровые изображения оценивается с помощью различных показателей. Основным требованием является обеспечение устойчивости перед стегоанализом, что выражается в статистической неразличимости стегоизображений и изображений, не содержащих вложений. Для обеспечения устойчивости перед стегоанализом при встраивании информации в цифровое изображение необходимо, чтобы встраивание не приводило к проявлению демаскирующих признаков.

Выявление демаскирующих признаков может быть произведено посредством анализа массивов изображений (стегоизображений и изображений, не содержащих вложений). В свою очередь, для построения надежного стеганографического алгоритма необходимо проведение анализа выявленных признаков и выбор наиболее информативных из них.

2. Признаки цифровых изображений, используемые в стегоанализе

В качестве признаков в задаче стегоанализа используются разнообразные статистические характеристики цифровых изображений. Известны методы стегоанализа, основанные на анализе признаков изображения в пространственной области, и методы, основанные на анализе признаков в частотной области. Также встречаются комбинированные методы, использующие признаки и в пространственной, и в частотной областях. В каждом методе стегоанализа значения некоторого набора признаков объединяются в один вектор, с которым далее работает классификатор.

Ключевой задачей является выбор информативных признаков, анализ которых позволяет отделять стегоизображения от чистых изображений.

Кратко рассмотрим наборы признаков, представленные в некоторых современных исследованиях, посвященных стегоанализу JPEG-изображений.

В работе [7] предлагается набор признаков для стегоанализа, основанных на соотношениях между энергией, собранной в отдельных коэффициентах спектра дискретного косинусного преобразования (ДКП):

$$F_1 = \frac{E(f_0)}{E(f_{|\eta|=1})},$$

где $E(f_0)$ – среднее значение частот нулевых АС-коэффициентов изображения по блокам.

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18

$$F_2 = \frac{\sum_{|\eta|>1} E(f_\eta)}{E(f_{|\eta|=1})},$$

где $E(f_{|\eta|=1})$ – среднее значение частот АС-коэффициентов изображения, по абсолютной величине равных 1.

$$F_3 = \frac{En_{|\eta|>1}}{En_{|\eta|\leq 1})},$$

где $En_{|\eta|>1}$ – энергия АС-коэффициентов изображения, по абсолютной величине больших 1.

Для изображений характерна межблочная корреляция. Во время встраивания вносятся изменения в блоки изображения, что может привести к нарушению связи между блоками:

$$F_4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

где $\bar{x}$, $\bar{y}$ – средние значения АС-коэффициентов соседних блоков;

x_i, y_i – i-й АС-коэффициент соседних блоков.

Авторы работ [2, 4] используют 23 признака, как в частотной области, так и в пространственной:

1) общая гистограмма ДКП-коэффициентов изображения;
2) гистограммы первых пяти АС-коэффициентов, для каждого отдельно;
3) межблочные зависимости в разных направлениях (признаки в пространственной области);
4) двойные гистограммы АС-коэффициентов, значение которых находится в диапазоне [–5, 5]:

$$F_5, \dots, F_{15} = \frac{\sum_{k=1}^{B} \delta(d, d_k(i,j))}{\left\|\sum_{k=1}^{B} \delta(d, d_k(i,j))\right\|_{L_1}},$$

где d – фиксированное значение коэффициента, $d \in [-5,5]$;

B – количество блоков в изображении;

i, j – координаты положения коэффициента в блоке;

$\delta(u,v) = \begin{cases} 1, u = v, \\ 0, \text{иначе}; \end{cases}$

L_1 норма – максимальная из сумм элементов по столбцам.

Каждая величина рассчитывается дважды: для исследуемого изображения (J_1) и для изображения, которое получают путем обрезания исследуемого изображения сверху и слева на 4 пикселя (J_2). Подобное действие объясняется следующим образом: при обрезании изображения слева и сверху его разделение на блоки сдвигается, и повторное квантование ДКП-коэффициентов уничтожает имеющееся вложение. В результате характеристики нового изображения будут значительно отличаться от характеристик исходного изображения, если последнее содержало вложение. В противном же случае отличия будут незначительны.

Поэтому в качестве конечного значения каждого признака принимается значение функционала:

$$F = \|f(J_1) - f(J_2)\|_{L_1},$$

где L_1 норма – максимальная из сумм элементов по столбцам.

Для стегоанализа можно использовать не только специально формируемые признаки, но и произвольные признаки, используемые в иных задачах обработки цифровых изображений. В связи с этим в исследуемый набор признаков было включено 10 текстурных признаков $F_{16}, \dots, F_{25}$, приведенных в статье [12].

Метод, представленный авторами [1], основан на законе Бенфорда: вероятность появления цифры на первом месте в числе тем выше, чем меньше эта цифра. Основываясь на выводах работы [5], посвященной исследованию справедливости закона Бенфорда в отношении ДКП-коэффициентов JPEG-изображений до и после квантования, авторы предложили частный случай закона Бенфорда, так как квантованные ДКП-коэффициенты не подчиняются строго закону Бенфорда.

В набор признаков были включены признаки на основе закона Бенфорда для каждой из 9 цифр (1), а также признаки, объединяющие суть закона Бенфорда и идею смещения изображения на 4 пикселя из [2, 4] (2):

$$F_{26}, \dots, F_{34} = N \log_{10}\left(1 + \frac{1}{s + x^q}\right), \quad (1)$$

где $x = 1, \dots, 9$;

N, s, q – параметры, зависящие от качества JPEG-сжатия.

$$F_{35}, \dots, F_{43} = \|F_{26}, \dots, F_{34}(J_1) - F_{26}, \dots, F_{34}(J_2)\|_{L_1}. \quad (2)$$

3. Отбор информативных признаков

3.1 Методы исследования

Существует много алгоритмов отбора информативных признаков, основанных на выявлении статистических зависимостей (между элементами набора, между элементами набора и выходным значением), сравнительных экспериментах, сложных вычислениях. Например, жадные алгоритмы поиска, алгоритмы поочередного перебора, генетический алгоритм (ГА).

Жадные алгоритмы поиска часто используются, так как достаточно быстры и дают хороший результат во многих задачах. Группа алгоритмов получила такое название из-за того, что если один из признаков был выбран в поднабор (или исключен), то в дальнейшем он остается в наборе (в случае жадного включения) или навсегда будет отсутствовать (в случае жадного исключения) [3].

В рамках настоящего исследования выбран жадный алгоритм с исключением, так как он является достаточно быстрым и эффективным согласно исследованиям [6, 9], учитывает все элементы набора признаков в совокупности, а не как отдельные составляющие. Применение исключения позволяет сравнить обновленный набор с тем, что был ранее и понять, показывает ли новый набор лучший результат при проведении стегоанализа.

Для классификации изображений использовался наивный байесовский классификатор.

3.2 Построение набора информативных признаков

Для формирования обучающей и тестовой выборок были взяты стандартные изображения (Lena, Peppers, Baboon и т.д.), а также изображения из баз UCID [8] и USC-SIPI ID [10] размером 256×256 пикселей. Обучающая (660 экземпляров) и тестовая (440 экземпляров) выборки содержат изображения:

1) без вложения;
2) с вложением по классическим алгоритмам F5, JSteg, PM1.

Так как в работе применен жадный алгоритм с исключением, то изначально рассматривался полный набор из 43 признаков, описанный ранее.

В табл. 1 частично представлены результаты экспериментов. Так как набор признаков достаточно объемный, то данная таблица включает три наибольших вероятности обнаружения на каждом шаге.

Шаг	Место		
	1	2	3
1	F_{25} (корреляция) 76,32	F_{33} (Бенфорд 8) 76,19	F_{24} (ковариация) 74,6
2	F_{17} (энтропия) 77,78	F_4 (корреляция межблочная) 76,19	F_{18} (однородность) 76,19
3	F_{30} (Бенфорд 5) 77,78	F_{27} (Бенфорд 2) 73,02	F_{19} (контраст) 71,43

The publication of the conference proceedings was supported by the RFBR, grant No. 18-07-20045\18

4	F_{19} (контраст) 77,78	F_{26} (Бенфорд 1) 77,78	F_{35} (Бенфорд+ смещение 1) 74,6
5	F_{26} (Бенфорд 1) 79,37	F_{11} (двойная гист-ма 1) 77,78	F_{28} (Бенфорд 3) 76,19
6	F_{24} (ковариация) 79,37	F_5 (двойная гист-ма –5) 74,6	F_{37} (Бенфорд+ смещение 3) 74,6
7	F_{27} (Бенфорд 2) 80,95	F_{14} (двойная гист-ма 4) 79,37	F_{32} (Бенфорд 7) 79,37
8	F_{31} (Бенфорд 6) 84,13	F_{36} (Бенфорд+ смещение 2) 84,13	F_{15} (двойная гист-ма 5) 82,54
9	F_7 (двойная гист-ма –3) 84,13	F_{10} (двойная гист-ма 0) 84,13	F_{13} (двойная гист-ма 3) 84,13
10	F_{35} (Бенфорд+ смещение 1) 84,13	F_8 (двойная гист-ма –5) 82,56	F_{14} (двойная гист-ма 4) 82,56
11	F_{15} (двойная гист-ма 5) 84,13	F_{32} (Бенфорд+ смещение 2) 82,56	F_1 (энергия част. 1) 82,56
12	F_{32} (Бенфорд 7) 84,13	F_{32} (Бенфорд+ смещение 2) 84,13	F_1 (энергия част. 1) 82,54
13	F_{13} (двойная гист-ма 3) 84,13	F_{10} (двойная гист-ма 0) 82,56	F_2 (энергия част. 2) 82,54
14	F_{14} (двойная гист-ма 4) 84,13	F_{40} (Бенфорд+ смещение 6) 84,13	F_{10} (двойная гист-ма 0) 82,54
15	F_6 (двойная гист-ма –4) 84,13	F_1 (энергия част. 1) 82,54	F_{12} (двойная гист-ма 2) 82,54
16	F_{33} (Бенфорд 8) 85,71	F_{40} (Бенфорд+ смещение 6) 85,71	F_3 (энергия част. 3) 84,13
17	F_{34} (Бенфорд 9) 87,3	F_{38} (Бенфорд+ смещение 4) 84,13	F_4 (корреляция межблочная) 82,54
18	F_{42} (Бенфорд+ смещение 8) 85,71	F_5 (двойная гист-ма –5) 84,13	F_{23} (дисперсия по *j*) 84,13

Табл. 1. Отбор информативных признаков

В результате экспериментов был построен следующий набор из 26 информативных признаков:

$\{F_1 - F_5, F_8 - F_{12}, F_{16}, F_{18}, F_{20} - F_{23}, F_{28}, F_{29}, F_{36} - F_{43}\}$.

3.3 Использование построенного набора признаков для повышения незаметности стеганографического встраивания

Построенный набор признаков был использован для улучшения стеганографического алгоритма, полученного в [11] и работающего с JPEG-изображениями. Данный алгоритм встраивает один бит сообщения в один ДКП-коэффициент с помощью операции замены: коэффициент заменяется значением *x*, если встраиваемый бит равен 1, и значением *–x* — в противном случае. Коэффициенты, по абсолютной величине равные *x* и не используемые для встраивания, корректируются, чтобы извлечение было однозначным.

Выбор оптимальных позиций для размещения *n*-битового фрагмента сообщения в ДКП-блоке осуществляется с помощью ГА. За целевую функцию принимается значение метрики PSNR.

В настоящем исследовании в основу целевой функции был положен выход классификатора, осуществляющего стегоанализ изображений–контейнеров:

$$f(x^*) = \max_{n \in N, x \in X}(1 - \mathrm{acc}_{n,x}),$$

где $\mathrm{acc}_{n,x}$ — точность классификации стегоизображений с разрешением *N*×*N* пикселей при встраивании *n*-битовых фрагментов сообщения в ДКП-блоки с величиной замены *x*.

Таким образом, повышение незаметности стеганографического встраивания заключалось в минимизации проявления демаскирующих признаков, которая, в свою, очередь реализовывалась посредством максимизации ошибки классификатора, обученного с использованием данных признаков.

Результаты экспериментов по встраиванию сообщений в JPEG-изображения с минимизацией проявления демаскирующих признаков приведены на рис. 1. В качестве алгоритма оптимизации, как и в [11], был выбран ГА. Параметры ГА были заданы следующим образом: 30 особей, 30 поколений, 40 итераций.

Для оценки визуального качества стегоизображений использовалась стандартная метрика PSNR.

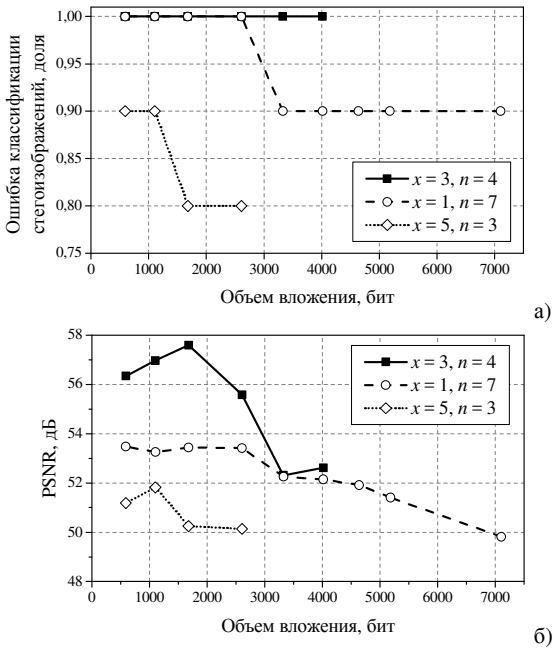

Рис. 1. Результаты встраивания: а) зависимость ошибки классификатора от объема вложения; б) зависимость значения PSNR от объема вложения

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18

4. Обсуждение результатов

Эксперименты с жадным алгоритмом показали, что набор из 43 признаков в пространственной и частотной областях можно сократить на 17 элементов, при этом повысив общую точность классификации на 19%.

Анализируя результаты экспериментов, стоит отметить, что среди 17 удаленных признаков находится 7 признаков на основе закона Бенфорда (остались лишь для значений 3 и 4). Признаки на основе совмещения закона Бенфорда с идеей смещения изображения оказались более устойчивыми, из них выбыл всего 1 (для значения 1), что позволяет сделать вывод о большей информативности и большем вкладе модифицированной версии признаков.

Также в течение нескольких первых итераций жадного алгоритма большую точность обнаружения показывали наборы без структурных признаков в пространственной области (корреляция, энтропия, контраст, ковариация). Однако после удаления этих 4 признаков в табл. 1 практически не было замечено присутствие других элементов из данной группы.

Оставшиеся 5 выбывших признаков – двойные гистограммы. Тут стоит обратить внимание на то, для каких значений коэффициентов были эти признаки: –4, –3, 3, 4, 5. Известно, что после ДКП и квантования наибольшее количество элементов принимают значения в диапазоне [–1, 1], и чем дальше значение от границ этого диапазона, тем меньше вероятность встретить его в матрице ДКП-коэффициентов. Поэтому двойные гистограммы для данных значений выбыли из общего набора признаков.

Наборы с исключением первых 4 признаков: 3 энергетических признака в частотной области и межблочная корреляция между блоками с близостью DC-коэффициентов не более 3, – всегда показывали результат ниже среднего, что говорит о важном вкладе этих признаков в информацию о содержании изображения.

Применение результирующего набора из 26 информативных признаков позволило повысить незаметность стеганографического встраивания информации в цифровые изображения с помощью алгоритма, основанного на операции замены квантованных ДКП-коэффициентов.

Как можно увидеть из графика на рис. 1а, наилучший результат в части минимизации проявления демаскирующих признаков достигается при встраивании в один ДКП-блок 4 бит сообщения с использованием величины замены $x = 3$. В этом случае классификатор, обученный на наборе информативных признаков, пропускает все стегоизображения независимо от объема вложения. Уменьшение значения x с одновременным увеличением значения n, либо увеличение x с уменьшением n показывает худший результат: при повышении объема вложения 10–20 % стегоизображений детектируются верно.

Визуальное качество получаемых стегоизображений в сравнении с исходными изображениями было оценено с помощью метрики PSNR. Повышение визуального качества встраивания не было целью оптимизации, поэтому соответствующий график на рис. 1б не показывает явно выраженную зависимость значения PSNR от объема вложения. Однако даже при максимальном объеме вложения PSNR остается на уровне 50 дБ, что говорит об отсутствии заметных искажений на стегоизображении. При этом наилучший результат по качеству встраивания также обеспечивают значения параметров встраивания $x = 3$, $n = 4$.

Дальнейшего улучшения можно добиться, если усложнить задачу оптимизации и встраивать сообщение в ДКП-блоки изображения-контейнера фрагментами переменного размера.

5. Заключение

В настоящей работе проведено исследование информативности признаков сжатых JPEG-изображений, используемых в задаче стегоанализа. Сформирован набор, включающий признаки, представленные в литературе, и их модификации. Для проведения стегоанализа использован наивный байесовский классификатор. С помощью жадного алгоритма мощность набора уменьшена с 43 признаков до 26 с одновременным повышением точности классификации.

Сформированный набор информативных признаков использован для решения обратной задачи: повышение устойчивости стеганографического встраивания перед стегоанализом за счет минимизации проявления демаскирующих признаков. В основу целевой функции положен выход классификатора, осуществляющего стегоанализ изображений–контейнеров с использованием выбранных признаков.

Для экспериментов взят алгоритм встраивания, основанный на операции замены квантованных ДКП-коэффициентов. Минимизация проявления демаскирующих признаков проведена с помощью генетического алгоритма. В результате при встраивании более чем 4000 бит в полутоновое изображение с разрешением 256×256 пикселей удалось достичь минимальных искажений изображения, держа метрику визуального качества PSNR не ниже 52дБ.

Полученные результаты могут быть использованы для построения каналов скрытой передачи конфиденциальной информации с целью ее защиты от несанкционированного доступа. Уменьшение размерности признакового пространства приводит к уменьшению времени оптимизации, поэтому предлагаемое решение целесообразно использовать в приложениях, требовательных к ресурсам, например, в мобильных устройствах или в устройствах «интернета вещей».

6. Благодарности

Работа выполнена при финансовой поддержке гранта Президента Российской Федерации по государственной поддержке ведущих научных школ Российской Федерации № НШ-3070.2018.8

7. Литература

[1] Andriotis P. JPEG steganography detection with Benford's Law / P. Andriotis, G. Oikonomou, T. Tryfonas // Digital Investigation. – 2013. – Vol. 9. – P. 246–257.

[2] Chen M.-C. Alpha-trimmed Image Estimation for JPEG Steganography Detection / M.-C. Chen // Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics. – San Antonio, Texas, USA. – 2009. – P. 4581–4585.

[3] Cormen Th.H. Introduction to algorithms. Third edition / Th.H. Cormen, Ch.E. Leiserson, R.L. Rivest, C. Stein. – Third edition. – London: The MIT Press, 2013. – 1324 p.

[4] Fridrich J. Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes / J. Fridrich // Proceedings of the Sixth International Workshop on Information Hiding, Lecture Notes in Computer Science. – 2014. – Vol. 3200. – P. 67–81.

[5] Fu D. A generalized Benford's law for JPEG coefficients and its applications in image forensics / D. Fu, Y.Q. Shi, W. Su // Proceedings of SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX. – USA, San Jose. – 2007. – P. 1L1–1L11.

The publication of the conference proceedings was supported by the RFBR, grant No. 18-07-20045\18

[6] Guyon I. An introduction to variable and feature selection / I. Guyon, A. Elisseeff // Journal of Machine Learning Research. – 2003. – Vol. 3. – P. 1157–1182.
[7] Jia-Fa M. A steganalysis method in the DCT domain / M. Jia-Fa, N. Xin-Xin, X. Gang, Sh. Wei-Guo, Zh. Na-Na // Multimedia Tools and Applications. – 2016. – №75. – P. 5999–6019.
[8] Image Databases [Electronic recourse] // ImageProcessingPlace.com. – Electronic data. – URL: http://www.imageprocessingplace.com/root_files_V3/image_databases.htm.
[9] Molina L.C. Feature Selection Algorithms: A Survey and Experimental Evaluation / L.C. Molina, L. Belanche, A. Nebot // Proceedings of the 2002 IEEE International Conference on Data Mining, IEEE Computer Society. – 2002. – P. 306–313.
[10] SIPI Image Database [Electronic recourse] // USC University of Southern California. – Electronic data. – Los Angeles, CA, 2017. – URL: http://sipi.usc.edu/database/.
[11] Евсютин О.О. Алгоритм встраивания информации в сжатые цифровые изображения на основе операции замены с применением оптимизации / О.О. Евсютин, А.А. Шелупанов, Р.В. Мещеряков, Д.О. Бондаренко // Компьютерная оптика. – 2017. – Т. 41, № 3. – С. 412–421.
[12] Мицель А.А. Непараметрический алгоритм текстурного анализа аэрокосмических снимков / А.А. Мицель, Н.В. Колодникова, К.Т. Протасов // Известия Томского политехнического университета. – 2005. – Т. 308(1). – С. 65–70.

Об авторах

Шумская Ольга Олеговна, техник кафедры безопасности информационных систем факультета безопасности Томского государственного университета систем управления и радиоэлектроники. Ее e-mail shumskaya.oo@gmail.com.

Евсютин Олег Олегович, канд. техн. наук, доцент кафедры безопасности информационных систем факультета безопасности Томского государственного университета систем управления и радиоэлектроники. Его e-mail eoo@keva.tusur.ru.

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18