СТАТИСТИЧЕСКИЙ СТЕГАНОАНАЛИЗ ФОТОРЕАЛИСТИЧНЫХ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ГРАДИЕНТНЫХ ПУТЕЙ

Солодуха Р.А.[1]

Целью статьи *является экспериментальная проверка эффективности вектора признаков на основе градиентных путей в пространственной области изображения.*

Метод исследования *– сравнение стеганоаналитических векторов признаков на основе среднеквадратической ошибки и коэффициента детерминации, полученных с помощью SVM-регрессии в Matlab. Датасет сформирован путем автоматизации стеганопрограмм сегмента freeware, реализующих вложение в пространственную область изображения с последовательным и псевдослучайным выбором пикселей для внедрения.*

В результате исследования *экспериментально осуществлен подбор оптимальных, с точки зрения выявления вложения, параметров алгоритма нахождения градиентных путей. Получены и проанализированы результаты применения моделей машинного обучения, определен оптимальный масштаб ядра SVM-регрессора. Рассчитано время вычисления векторов признаков, обучения модели, распознавания контейнеров. Показано, что вектор признаков на основе градиентных путей целесообразно использовать для решения задач, где необходимо варьировать точность обнаружения вложения в зависимости от нагрузки на стеганоаналитическую систему, т.к. данный вектор признаков позволяет определить соотношение размерность/точность. Также путем эксперимента подобран комплексный 20D вектор из нескольких одномерных количественных стеганодетекторов и вектора признаков на основе градиентных путей, эффективность которого сопоставима с 686D вектором признаков SPAM.*

Ключевые слова: *вектор признаков, градиент яркости, стеганодетектор, машинное обучение, регрессия, машина опорных векторов, пространственная область изображения, наименее значащие биты, сегментация битовой плоскости по сложности.*

DOI: 10.21681/2311-3456-2022-1-26-36

Введение

Основой статистического стеганоанализа изображений является факт нарушения определенных закономерностей изображения, вызванный реализацией стегановложения. Одна из классификаций существующих методов стеганоанализа (сигнатурные, статистические, эвристические [1]) позволяет отнести к группе статистических те методы, которые основываются на сопоставлении характеристик обобщенных пустого и заполненного контейнеров, другими словами, оценивают близость исследуемого контейнера к естественному.

Статистические методы не являются средством, позволяющим гарантировано определять наличие скрытой информации. Они дают возможность аналитику с определенной вероятностью судить о том, используется стеганография или нет. Результаты работы методов зависят от стеганографического преобразования, используемого для встраивания скрываемых данных, а также от их объема. Как правило, выявление факта скрытия осуществимо при значительном заполнении контейнера. К тому же методы этой группы обычно построены на алгоритмах, требующих предварительного «обучения» на сериях из заполненных и пустых контейнеров.

Одним из признаков фотореалистичности является плавное изменение освещенности сцены. Таким образом, при прочих равных, стегановложение должно вносить более значительные искажения в группу пикселей, расположенных вдоль линии градиента яркости. Следовательно, возникает задача проверить данную гипотезу путем построения градиентных путей и проведения вычислительного эксперимента. Математическая модель, алгоритм и особенности программной реализации[2] формирования последовательности пикселей по направлению минимального градиента яркости с дальнейшим использованием в стеганоанализе подробно изложены в [2].

Следует отметить, что модели на основе градиента в стеганоанализе пространственной области изображения успешно применяются, в частности, для предсказания фоновых областей. В [3] приводится математическая модель и описан программный эксперимент, подтвердивший эффективность использования значений градиента, что вкупе со стеганодетектором

2 Солодуха Р.А. Формирование градиентных путей изображения // Свидетельство о регистрации программы для ЭВМ RU 2020660867, 15.09.2020. Заявка № 2020660198 от 04.09.2020.

1 Солодуха Роман Александрович, кандидат технических наук, доцент, доцент кафедры автоматизированных информационных систем органов внутренних дел Воронежского института МВД России, г. Воронеж, Россия, E-mail: standartal@list.ru

Weighted Stego улучшило точность стеганоанализа до 10%, по сравнению с модификацией WSPAM [4].

Таким образом, целью статью является анализ эффективности использования вектора признаков на основе градиентных путей и сравнение с известными векторами признаков, применяемыми в статистическом стеганоанализе пространственных областей изображений.

Формирование градиентных путей

Для нахождения градиента используется понятие производной дискретной функции[3]. В каждой точке изображения $f_{i,j} = 0,1,\ldots,255$, кроме краевых, вводятся разности, соответствующие оператору Робертса (кроме оператора Робертса можно использовать операторы Превитта и Собела), имеющие смысл градиента по направлениям: вертикальному, горизонтальному, главной и побочной диагоналям:

$$f^{|}_{i,j} = f_{i+1,j} - f_{i,j};\ f^{-}_{i,j} = f_{i,j+1} - f_{i,j};$$
$$f^{\backslash}_{i,j} = f_{i+1,j+1} - f_{i,j};\ f^{/}_{i,j} = f_{i-1,j+1} - f_{i,j}. \quad (1)$$

Каждому пикселю сопоставляется вектор-строка, показывающая направление градиента:

$$g_{i,j} = \left(\frac{f^{|}_{i,j}}{\left|f^{|}_{i,j}\right|}, \frac{f^{-}_{i,j}}{\left|f^{-}_{i,j}\right|}, \frac{f^{\backslash}_{i,j}}{\left|f^{\backslash}_{i,j}\right|}, \frac{f^{/}_{i,j}}{\left|f^{/}_{i,j}\right|} \right). \quad (2)$$

Также каждому некраевому пикселю сопоставляется кортеж, элементы которого имеют смысл вторых производных в точке с координатами i, j по соответствующему направлению: $\left\langle f^{\|}_{i,j}, f^{=}_{i,j}, f^{//}_{i,j}, f^{\backslash\backslash}_{i,j} \right\rangle$ и формируют матрицу локальных экстремумов по правилу (t – пороговое значение):

$$E_{i,j} = true \left| \begin{array}{l} f^{\|}_{i,j} > t \wedge f^{=}_{i,j} > t \wedge f^{\backslash\backslash}_{i,j} > t \wedge f^{//}_{i,j} > t, \\ false - overwise \end{array} \right. . \quad (3)$$

Точки ветвления показывает булева матрица:

$$cross_{i,j} = true \left| \sum g_{i,j} > 1,\ false - overwise \right. . \quad (4)$$

Градиентным путем называется список пикселей по направлению градиента из каждого пикселя (кроме экстремума), до края, экстремума или ветвления. По мере прохода пикселя-ветвления в определенном направлении соответствующие позиции градиента обнуляются, т.е. каждый пиксел в одном направлении может быть пройден только один раз.

При построении путей можно варьировать такими параметрами, как:

- минимальная и максимальная длина списка;
- порог признания пикселя экстремумом;
- минимальное количество элемента списков одной длины для обеспечения статистической достоверности;
- вид оператора вычисления градиента.

Формирование вектора признаков

При формировании вектора признаком исходят из предположения, что стегановложение нарушает монотонность изменения яркости естественных изображений. При построении вектора признаков сравниваются параметры исходного изображения и изображения с обнуленными LSBs, в зависимости от глубины искажения контейнера стеганоалгоритмом. Перейдем к формальной модели.

Пусть L_{min} и L_{max} – минимальная и максимальная длины списков пикселей, соответственно. Списки пикселей хранятся в структуре типа список $L = \left\{ L^{len} \right\}_{len=L_{min}}^{len=L_{max}}$, где $L^{len} = \left\{ l_i^{len} \right\}_{i=1}^{i=C_{len}}$ – список размером C_{len} списков длиной len. В свою очередь $l_i^{len} = \left\{ p_j \right\}_{j=1}^{j=len}$, список пикселей (градиентный путь) длиной len, $i = (1,\ldots,C_{len})$.

Для каждого градиентного пути рассчитывается гладкость (с хранением в аналогичной списочной структуре)

$$G^{len} = \left\{ g_i^{len} = \sum_{k=1}^{len-1} \left| Ip_k - Ip_{k+1} \right| \right\}_{i=1}^{i=C_{len}}, \quad (5)$$

где Ip_k – значение p_k пикселя, $G = \left\{ G^{len} \right\}_{len=L_{min}}^{len=L_{max}}$.

Также гладкость рассчитывается для изображения с обнуленными LSBs – обозначим ее $\hat{G} = \left\{ \hat{G}^{len} \right\}_{len=L_{min}}^{len=L_{max}}$.

Для выявления вложения важно, сколько путей стали более гладкими после обнуления LSBs, а сколько менее. Запишем правило:

$$F_1^{len} = \frac{1}{C_{len}} \sum_{i=1}^{C_{len}} \begin{cases} 1, & g_i^{len} > \hat{g}_i^{len};, \\ 0, & otherwise. \end{cases} \quad (7)$$

$$F_2^{len} = \frac{1}{C_{len}} \sum_{i=1}^{C_{len}} \begin{cases} 1, & g_i^{len} < \hat{g}_i^{len}; \\ 0, & otherwise. \end{cases}$$

Также целесообразно включить в вектор признаков приведенное количество экстремумов до и после обнуления

$$f = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \begin{cases} 1, & E_{i,j} = true;, \\ 0, & otherwise. \end{cases} \quad (8)$$

$$\hat{f} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \begin{cases} 1, & \hat{E}_{i,j} = true; \\ 0, & otherwise. \end{cases}$$

3 Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2012. 1104 с.

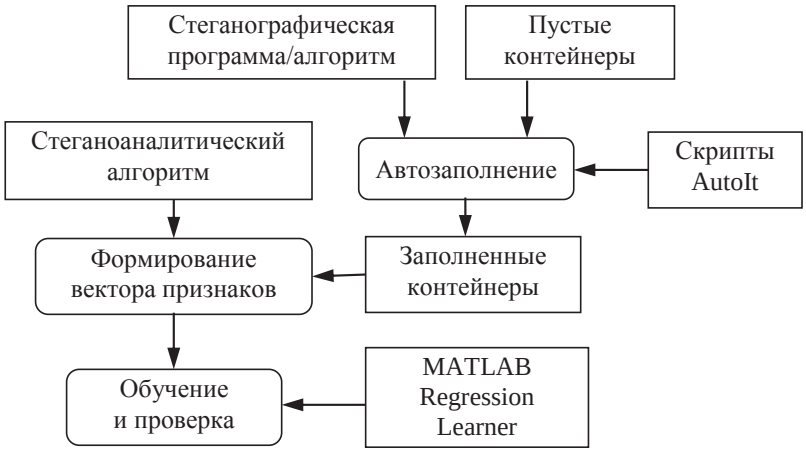

Рис. 1. Обобщенная схема проведения эксперимента

Таким образом, формируется следующий результирующий вектор признаков (GP) с размерностью $2\times\left(L_{max}-L_{min}+1\right)$:

$$F=\left\{f,\hat{f}\right\}\cup\left\{F_1^{len},F_2^{len}\right\}_{len=L_{min}}^{len=L_{max}}. \quad (9)$$

Экспериментальная часть

Целью эксперимента является проверка разработанного вектора признаков (9) на стеганопрограммах сегмента freeeware, реализующих вложение в пространственную область изображения с последовательным и псевдослучайным выбором пикселя для внедрения, и сравнение с известными стеганоаналитическими детекторами.

Сначала формируется множество заполненных с определенным шагом стеганоконтейнеров. Для этого используется среда разработки AutoIt. Размер вложения рассчитывается в процентах от максимального, значение которого считывается из стеганопрограммы. Файл вложения вырезается из случайного места случайного jpeg-файл, затем архивируется с помощью WinRar. Таким образом, можно быть уверенным в том, что если в стеганопрограмме присутствует предварительная архивация вложения, то это не помешает эксперименту. Далее к контейнерам применяется стеганоаналитический алгоритм, сопоставляя каждому файлу трасологические данные. Эти данные можно комбинировать в различные по размеру и составу векторы признаков с дальнейшим обучением, что определяет схему эксперимента (рис.1).

В эксперименте используется традиционное машинное обучение с учителем, но в отличии от аналогичных работ по исследованию комбинаций <стеганографический алгоритм, вектор признаков, классификатор>, где стеганографическая модификация моделируется [5-6] или используется «чистый» алгоритм (HUGO, S-UNIWARD, WOW, etc.) [7-8], примененный подход позволяет учесть особенности стеганоалгоритма с неизвестными параметрами, реализованного в конкретной программе, что важно в рамках атак на уровне контейнера [9] на основании известной стеганопрограммы. К качеству выборки предъявляются высокие требования в контексте мнение, изложенного в [10] о том, что с точки зрения успеха в решении задач машинного обучения качество данных, как правило, намного важнее качества алгоритма обучения. При этом, в данной статье не ставится задача стеганоанализа с заданной достоверностью [11], поэтому методика подготовки датасета, изложенная в [12], не применяется.

Разработчики стеганографических алгоритмов стараются реализовать вложение с как можно меньшим искажением статистических свойств изображения. Особенно это касается стеганоалгоритмов, модифицирующих несколько битовых плоскостей, а не только плоскость LSB. Например, BPCS-стеганография[4] (Bit-Plane Complexity Segmentation steganography) использует для принятия решения о внедрении данных оценку сложности изображения, для чего осуществляют декомпозицию битовых срезов на блоки 8x8, подсчет числа переходов между черным и белым значением в каждой строке и столбце для блока. Подход основан на особенностях человеческого зрения, которое восприимчиво к внесению изменений в низкочастотные области изображения, но не может различить изменения в высокочастотных (шумовых) областях. Суть метода заключается в замене шумовых блоков на блоки со скрываемой информацией, с учетом их сложности. Если блок недостаточно сложен, то на него накладывается (xor) высококонтрастная маска (конъюгация). Пропускная способность контейнера при отсутствии видимых искажений достигает 3-4 бит/байт.

С стороны инсайдера BPCS-стеганография является наиболее предпочтительными для скрытой передачи значительных объемов информации. В контексте решения задачи стеганоанализа в ведомственной DLP-системе [13] выявление таких вложений приоритетно. С точки зрения стеганоанализа, блочные моди-

4 Kawaguchi E., Eason R., Principle and applications of BPCS-steganography // Multimedia Systems and Applications. 1998, vol. 3528. pp. 464–473.

Рис. 2. Модификация изображения программой Cryptarkan 1.0

Рис. 3. Модификация изображения программой The Third Eye 1.0

Рис. 4. Модификация изображения программой Qtech-Hide&View v02

фикации стеганоалгоритмов вносят больше искажений, поскольку встраиваемый блок может находиться в любой битовой плоскости, с другой – эти искажения носят локальный характер. При этом возникает вопрос о применении универсального метода стеганоанализа для BPCS и LSBs, для чего в тестовом наборе представлены программы, реализующие BPSC и LSB-replacement стеганографию в пространственную область файлов формата BMP 24 бит. Для наглядности на рис. 2-4 показаны измененные участки изображения (использована программа WinMerge[5]) при вложении 9, 49 и 99 % от максимально возможного:

1. Cryptarkan 1.0 – не требует установки, позволяет задействовать 1, 2 или 4 LSBs файла, шифровать стегановложение. Вложение реализуется последовательно (рис. 2).
2. The Third Eye 1.0 – не требует установки, задействует плоскость младшего бита. Байты для вложения выбираются случайно (рис. 3).
3. Qtech-Hide&View v02 (Qtech-HV02) – требует установки, реализует алгоритм BPCS, задействует 1-5 LSBs файла, в зависимости от размера вложения. Имеет предустановленный порог сложности 40%. Характер вносимых искажений – блочный (рис. 4).

В качестве источника контейнеров взята коллекция BOSSbase 1.01[6] (первые 1000 файлов), которая

5 https://winmerge.org/downloads/

6 http://dde.binghamton.edu/download/ImageDB/BOSSbase_1.01.zip

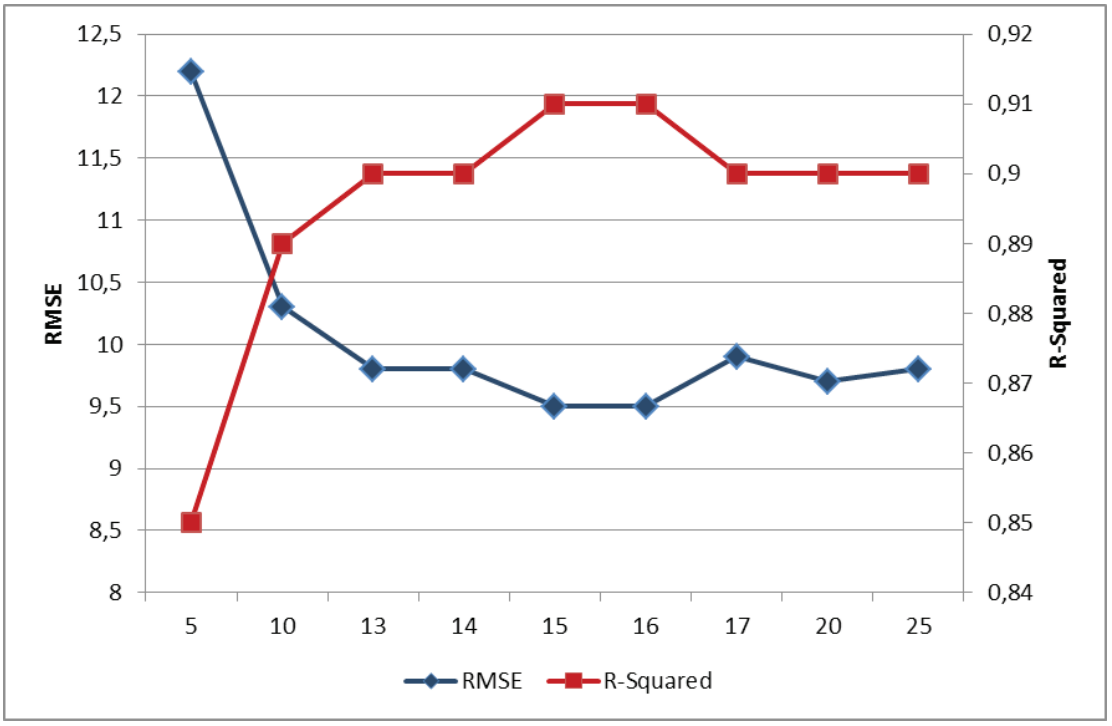

Рис. 5. Графики подбора оптимального значения t

содержит 10000 8-битных полутоновых изображений размером 512x512 пикселей в формате PGM (Portable Grey Map). Для дальнейшей работы файлы были преобразованы в формат BMP 8 бит с помощью XnConvert, затем в формат BMP 24 бит с помощью пакетного преобразования FastStone Image Viewer. Поскольку все три цветовых плоскости контейнера одинаковые, для анализа использовался только канал красного цвета. Контейнеры заполнялись с шагом 10 % от максимального размера вложения от 9 до 99 %. Таким образом, выборка составила 11000 контейнеров. Учитывая значительное количество классов – 11, в качестве прогнозной модели выбрана регрессионная, а именно, машина опорных векторов с гауссовским ядром. В качестве среды машинного обучения использовался MATLAB Regression Learner с регрессорами из группы SVM:

- Coarse Gaussian SVM (C), масштаб ядра – 4 $\sqrt{D}$,
- Medium Gaussian SVM (M), масштаб ядра – $\sqrt{D}$,
- Fine Gaussian SVM (F), масштаб ядра – $\frac{1}{4}\sqrt{D}$,

с настройками по умолчанию (D – количество предикторов). Все приведенные в таблицах результаты получены для лучших регрессоров (указаны в строке Learner). Выборка делилась на обучающую и тестовую в соотношении 50/50.

Основной настройкой при формировании градиентных путей является порог признания пикселя экстремумом *t*. Была проведена серия экспериментов (рис. 5), которые показали, что оптимальным порогом является 15.

Одним из параметров GP является длина максимальная списка пикселей («ext» - экстремумы, «ext, 3» - экстремумы и списки длиной 3, «ext, 3-4» - экстремумы и списки длиной от 3 до 4 и т.д.). На графиках (рис. 6) показана зависимость точности прогнозирования от этого параметра. Варьируя длину списка можно добиться оптимального соотношения размера вектора признаков и точности прогнозирования при решении конкретной задачи. В дальнейших вычислениях используется 16D вектор признаков.

В качестве стеганодетекторов для сравнения выбраны количественные 1D LSB–детекторы [14-16], реализованные лабораторией встраивания цифровых данных Бингхэмптоновского университета в виде программ для Matlab[7]: Weighted Stego[8] (WS), Sample Pairs[9] (SP), Triples analysis[10] (T), Asymptotically Uniformly Most Powerful detection[11] (AUMP), Pair of

7 http://dde.binghamton.edu/download/structural_lsb_detectors/

8 Ker A., Böhme R. Revisiting Weighted Stego-Image Steganalysis // Security, Forensics, Steganography, and Watermarking of Multimedia Contents, Proc. SPIE Electronic Imaging. 2008. Vol. 6819. pp. 0501-0517.

9 Dumitrescu S., Wu X., Memon D. On steganalysis of random LSB embedding in continuous-tone images // Proceedings of ICIP. 2002. pp.641-644.

10 Ker A., A general framework for the structural steganalysis of LSB replacement // Proceedings 7th Information Hiding Workshop. Vol. 3727, pp. 296-311, Barcelona, Spain, 2005.

11 Fillatre L. Adaptive steganalysis of Least Significant Bit replacement in grayscale natural images // IEEE Transactions on Signal Processing. 2012. Vol. 60(2), pp. 556-569.

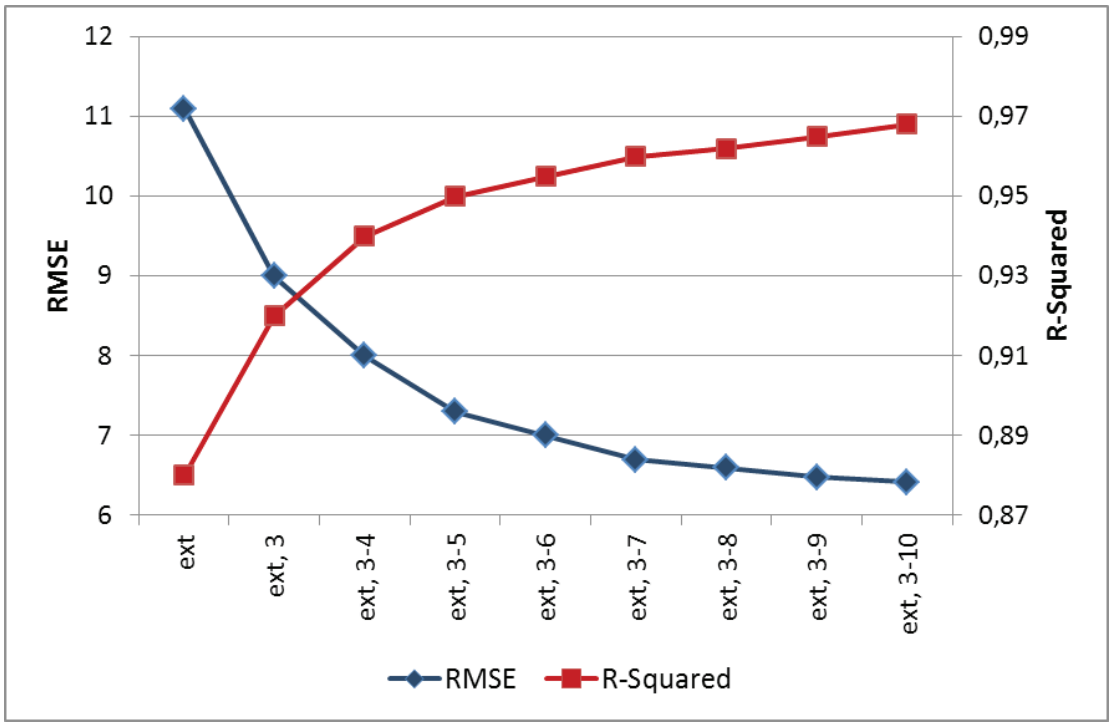

Рис. 6. Графики динамики точности в зависимости от количества признаков

Values[12] (PoVs) и 686D набор признаков SPAM[13]. WS использовался в модификации без коррекции смещения (WSn) и с коррекцией смещения (WSy), выбиралась модификация с лучшим результатом. Параметры AUMP: размер блока пикселей – 16, степень полинома – 6.

Результаты экспериментов[14] приведены в таблицах (табл. 1-4). Сначала приводятся результаты работы каждого метода, затем ряд «опорных» комбинаций, по которым можно понять вклад каждого метода. Дополнительная цель сравнения – определить наиболее результативную комбинацию и проверить, возможно ли достижение результативности SPAM вектором признаков меньшей размерности.

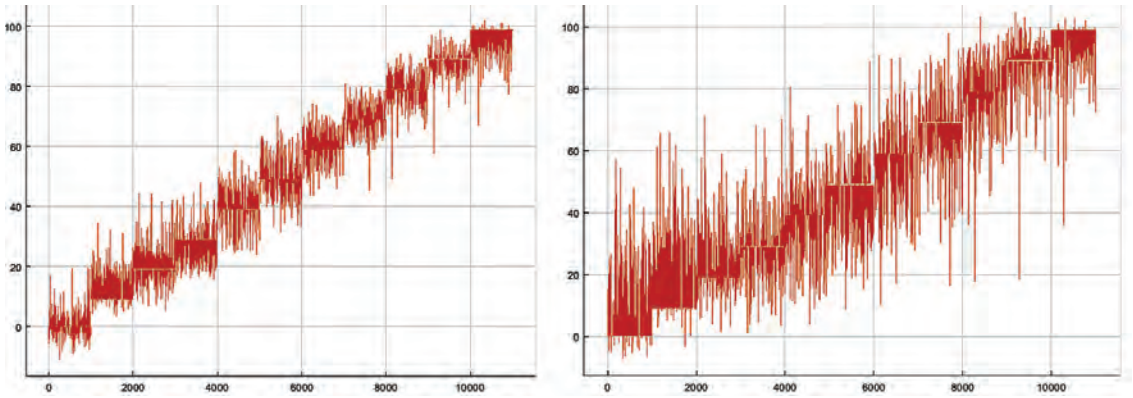

Рис. 7. Ошибки прогнозирования размера вложения в процентах от максимально возможного для векторов признаков SPAM (слева) и GP (справа)

Разброс ошибки прогнозирования SPAM и GP в зависимости от размера вложения для Qtech-HV02 проиллюстрирован с упорядочением контейнеров по размеру стегановложения (рис. 7).

12 Westfeld A., Pfitzmann A. Attacks on Steganographic Systems: Breaking the Steganographic Utilities EzStego, Jsteg, Steganos and S-Tools-and Some Lessons Learned // International Workshop on Information Hiding. 1999. pp.61-76.

13 Pevny T., Bas P., Fridrich J. Steganalysis by Subtractive Pixel Adjacency Matrix // IEEE Transactions on Information Forensics and Security. 2010. Vol. 5(2), pp. 215-224.

14 Доступ к датасетам – https://www.kaggle.com/romansolodukha/some-bitplane-steganalysis-features

Таблица 1

Результаты работы количественных LSB-детекторов и их комбинаций для CryptArkan (4LSBs)

Детектор / Результат	all						SPAM	WSn+SP	WS+SP+ AUMP	GP+WSn	GP+WSn+SP	all\(PoVs, T)	all
	WSn	SP	T	AUMP	PoVs	GP							
RMSE	14.9	15.7	32.8	16.1	26.3	6.2	4.5	8.9	7.8	5.4	4.6	4.6	4.4
R-Squared	0.78	0.75	0	0.74	0.32	0.96	0.98	0.92	0.94	0.97	0.98	0.98	0.98
Learner	M	M	M	M	M	M	M	M	M	M	M	M	M

Таблица 2

Результаты работы количественных LSB-детекторов и их комбинаций для CryptArkan (LSB)

Детектор / Результат	all						SPAM	SP+AUMP	SP+ AUMP+T	GP+SP	all\(PoVs, WSy)	all
	WSy	SP	T	AUMP	PoVs	GP						
RMSE	25	12.1	19.2	16.6	27.5	9.5	4.4	11.3	8.9	7.7	6.7	6.9
R-Squared	0.23	0.82	0.63	0.72	0.22	0.91	0.98	0.87	0.92	0.94	0.95	0.95
Learner	F	F	F	F	F	F	M,C	F,M	F,M	M	M	M

Таблица 3

Результаты работы количественных LSB-детекторов и их комбинаций для The Third Eye (LSB)

Детектор / Результат	all						SPAM	GP+T	T+WS	T+WS+ AUMP	all\(PoVs, SP)	all
	WSy	SP	T	AUMP	PoVs	GP						
RMSE	25.1	27.5	19.4	25.5	21.4	11	3.1	10.9	9.6	8.2	7.1	6.9
R-Squared	0.37	0.25	0.62	0.35	0,48	0.88	0.98	0.88	0.91	0.93	0.95	0.95
Learner	F	M,F	F	F,M,C	F,M,C	F	M	F	F	F	M	M

Таблица 4

Результаты работы количественных LSB-детекторов и их комбинаций для Qtech-HV02 (BPCS)

Детектор / Результат	all						SPAM	SP+AUMP	SP+AUMP+WS	AUMP+GP	all\ (PoVs, T|WS)	SP+AUMP+GP	all\PoVs	all
	WSy	SP	T	AUMP	PoVs	GP								
RMSE	26.7	27.2	28	25	31	13	6.3	24.6	23.3	11.5	11	10.8	10.5	10.7
R-Squared	0.29	0.25	0.18	0.36	0.1	0.83	0.96	0.38	0.44	0.86	0.87	0.88	0.88	0.88
Learner	F	F,M	F	F,M	F	M	M	F,M	F	M	M	M	M	M

DOI:10.21681/2311-3456-2022-1-26-36

Таблица 5

Соотношение времени вычисления признаков

Детектор	SPAM	WSy	WSn	SP	T	AUMP	PoVs	GP
Вычисление	23	3.2	2.5	1.5	32	2	1	135

Таблица 6

Соотношение времени обучения регрессоров и скорости распознавания

Детектор		SPAM	WSy	WSn	SP	T	AUMP	PoVs	GP
Fine SVM	Обучение	70	11	10	10	11	10	4	25
	Распознавание	*0.56*	*14*	*15*	*11*	*15*	*11*	*29*	*11*
Medium SVM	Обучение	16	11	11	9	10	9	3	12
	Распознавание	*5.5*	*11*	*11*	*12*	*15*	*12*	*43*	*12*
Coarse SVM	Обучение	27	10	10	9	10	9	4	10
	Распознавание	*1.6*	*10*	*11*	*11*	*11*	*11*	*28*	*9.6*

Результаты и обсуждение

Одномерные детекторы работали весьма нестабильно и, как правило, малоэффективно. При этом их комбинации давали вполне приемлемые результаты (R-Squared>0.9) для алгоритмов семейства LSB, например: WS+SP+AUMP для Cryptarkan 4LSBs, SP+AUMP+T для Cryptarkan LSB, T+WS+AUMP для The Third Eye. Алгоритм BPCS перед одномерными детекторами и их комбинациями устоял (R-Squared<0.5).

SPAM показал стабильно высокие результаты во всех испытаниях (R-Squared>0.95). Эффективность GP можно охарактеризовать как стабильно среднюю (0.88<R-Squared<0.96). Неожиданным является разброс в результатах GP между Cryptarkan 4LSBs и Qtech-HV02. Оба алгоритма вносят соизмеримые искажения в битовые плоскости, Qtech-HV02 затрагивает даже 5LSB. Наиболее правдоподобным объяснение состоит в том, что искажения BPCS в рамках блока меньше влияют на положение экстремумов, и, как следствие, на градиенты.

При анализе эффективности Gaussian SVM с различными ядрами целесообразно учитывать результаты с R-Squared>0.7. Здесь можно наблюдать преимущество Medium SVM, за исключением использования LSB при размере вектора признаков не более 20 (табл. 2, 3), где результат лучше у Fine SVM.

Что касается скорости расчетов, то сравнение затруднено в связи с тем, что SPAM и одномерные детекторы работают в среде MatLab, а GP вычисляется приложением windows forms (C#). Примерная оценка времени обработки данных на офисном компьютере Intel Core i3 приведена в таблице 5, при этом использованы не абсолютные значения времени, а соотношение. Также приводятся сведения о времени обучения в секундах и скорости распознавания в тысячах объектов в секунду (усреднение по 3 испытаниям).

Анализ таблицы 6 показывает преимущество Medium SVM как во времени обучения, так и в скорости распознавания (за исключением WS). С увеличением количества признаков различие в производительности растет.

Заключение

В работе изложена процедура формирования вектора признаков GP на основе градиентных путей пространственной области изображения и исследована его эффективность путем сравнения с аналогами. GP показал лучшую результативность и стабильность по сравнению с одномерными детекторами, но проиграл SPAM для всех стеганопрограмм, участвующих в эксперименте.

Достоверность выводов обеспечена формированием контейнеров непосредственно с помощью стеганографического приложения, значительным объемом тестовой выборки, использованием проприетарной программной среды.

Комбинациям GP и одномерных векторов признаков удается приблизится по эффективности к SPAM для алгоритмов семейства LSB. При этом количество признаков соотносится как 20/686. Таким образом, сфера применения GP – задачи, где неприменимы векторы большой размерности, а также класс задач, где необходимо варьировать размерность/точность в зависимости от загруженности для соответствия QoS (например, предварительный стеганоанализ в рамках DLP-системы). Также GP можно использовать в составе комплексного вектора признаков для повышения точности стеганоанализа.

Литература

1. Шниперов А.Н., Прокофьева А.В. Метод стеганоанализа статических изображений формата JPEG на основе искусственных иммунных систем // Вопросы кибербезопасности. 2020. № 2 (36). С. 22-31. DOI: 10.21681/2311-3456-2020-2-22-31.
2. Солодуха Р.А. Формирование градиентных путей изображения как предварительный этап стеганоанализа // Вестник Воронежского института МВД России. 2020. № 1. С. 97-106.
3. Башмаков Д.А. Точность предсказания пикселей фоновых областей цифровых изображений в задаче стеганоанализа методом Weighted Stego // Кибернетика и программирование. 2018. №2. С. 38-47. DOI: 10.25136/2306-4196.2018.2.25706.
4. Башмаков Д.А. Адаптивное предсказание пикселей пикселей в градиентных областях для улучшения точности стеганоанализа в неподвижных цифровых изображениях // Кибернетика и программирование. 2018. №2. С. 83-92. DOI: 10.25136/2306-4196.2018.2.25514.
5. Сивачев А.В., Прохожев Н.Н., Михайличенко О.В., Башмаков Д.А. Эффективность стеганоанализа на основе методов машинного обучения // Вопросы кибербезопасности. 2017. № 2 (20). С. 53-60. DOI: 10.21681/2311-3456-2017-2-53-60.
6. Сивачев А.В. Эффективность статистических методов стеганоанализа при обнаружении встраивания в вейвлет область изображения // Вопросы кибербезопасности. 2018. № 1 (25). С. 72-78. DOI: 10.21681/2311-3456-2018-1-72-78.
7. Сирота А.А., Дрюченко М.А., Иванков А.Ю. Стегоанализ цифровых изображений с использованием методов поверхностного и глубокого машинного обучения: известные подходы и новые решения // Вестник ВГУ, Серия: Системный анализ и информационные технологии. 2021. № 1. С. 33-52. DOI:10.17308/sait.2021.1/3369.
8. Монарев А.И., Пестунов В.А. Эффективное обнаружение стеганографически скрытой информации посредством интегрального классификатора на основе сжатия данных // Прикладная дискретная математика. 2018. № 40. С.59-71. DOI: 10.17223/20710410/40/5.
9. Макаренко С.И. Эталонная модель взаимодействия стеганографических систем и обоснование на ее основе новых направлений развития теории стеганографии // Вопросы кибербезопасности. 2014. №2 (3). С. 24-32.
10. Парасич А. В., Парасич В. А., Парасич И. В. Формирование обучающей выборки в задачах машинного обучения // Информационно-управляющие системы. 2021. № 4. С.61-70. DOI:10.31799/1684-8853-2021-4-61-70.
11. Atlasov I., Solodukha R. Sample Representativeness Estimation as a Preliminary Stage of Statistical Steganalysis / 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), 2021, pp. 78-84, DOI: 10.1109/SUMMA53307.2021.9632204.
12. Atlasov I., Solodukha R. The mathematical model of estimation the multidimensional steganalytical methods reliability / Journal of Physics: Conf. Series. 2019. Vol. 1202. DOI: 10.1088/1742-6596/1202/1/012022.
13. Солодуха Р.А. Концепция формирования системы противодействия стеганографическим каналам в компьютерных сетях органов внутренних дел // Вестник Воронежского института МВД России. 2021. № 1. С.131-142.
14. Вильховский Д.Э. Обзор методов стеганографического анализа изображений в работах зарубежных авторов // Математические структуры и моделирование. 2020. № 4(56). С. 75–102. DOI: 10.24147/2222-8772.2020.4.75-102.
15. Belim S.V., Vilkhovskiy D.E. Algorithm for detection of steganographic inserts type LSB-substitution on the basis of an analysis of the zero layer // Journal of Physics: Conf. Series. 2017. Vol. 944. DOI:10.1088/1742-6596/944/1/012012.
16. Грачев Я.Л., Сидоренко В.Г. Стегоанализ методов скрытия информации в графических контейнерах // Надежность. 2021. № 21(3). С. 39-46. DOI: 10.21683/1729-2646-2021-21-3-39-46.

STATISTICAL STEGANALYSIS OF PHOTOREALISTIC IMAGES USING GRADIENT PATHS

Solodukha R.A.[15]

Abstract

Purpose *of the article is experimentally test the efficiency of the feature vector based on gradient paths in the spatial domain of the image.*

Research method *is a comparison of steganalytical feature vectors based on the mean square error and the coefficient of determination obtained using SVM-regression in Matlab. The dataset is formed by automating the freeware steganoprograms that implement embedding into the spatial area of the image with sequential and pseudo-random selection of a pixel for embedding.*

15 Roman A. Solodukha, Ph.D. (in Tech.), Associate Professor, Associate Professor of Department of Automated Information Systems of Interior Units, Voronezh Institute of the Ministry of the Interior of Russia, Voronezh, Russia, E-mail: standartal@list.ru

Results of the study: *the optimal parameters of the algorithm for seeking gradient paths from the point of view of embedding detection are experimentally obtained. The results of applying machine learning models are obtained and analyzed, the optimal scale of the SVM-regression kernel is determined. The computation durations of feature vectors obtaining, models training, recognizing containers are calculated. It is shown experimentally that the gradient paths feature vector is expedient to use for solving problems where it is necessary to vary the detection accuracy depending on functioning capacity of system, because the proposed feature vector allows to determine the dimension / accuracy ratio. Also, by experiment, a complex 20D vector is selected from several one-dimensional quantitative steganodetectors and the gradient paths feature vector. The effectiveness of result vector is comparable to the 686D feature vector SPAM.*

Keywords: *feature vector, brightness gradient, steganodetector, machine learning, regression, support vector machine, spatial domain of image, least significant bits, bit-plane complexity segmentation.*

References

1. SHniperov A.N., Prokof'eva A.V. Metod steganoanaliza staticheskih izobrazhenij formata JPEG na osnove iskusstvennyh immunnyh sistem // Voprosy kiberbezopasnosti. 2020. № 2 (36). S. 22-31. DOI: 10.21681/2311-3456-2020-2-22-31.
2. Soloduha R.A. Formirovanie gradientnyh putej izobrazheniya kak predvaritel'nyj etap steganoanaliza // Vestnik Voronezhskogo instituta MVD Rossii. 2020. № 1. S. 97-106.
3. Bashmakov D.A. Tochnost' predskazaniya pikselej fonovyh oblastej cifrovyh izobrazhenij v zadache steganoanaliza metodom Weighted Stego // Kibernetika i programmirovanie. 2018. №2. C. 38-47. DOI: 10.25136/2306-4196.2018.2.25706.
4. Bashmakov D.A. Adaptivnoe predskazanie pikselej pikselej v gradientnyh oblastyah dlya uluchsheniya tochnosti steganoanaliza v nepodvizhnyh cifrovyh izobrazheniyah // Kibernetika i programmirovanie. 2018. №2. C. 83-92. DOI: 10.25136/2306-4196.2018.2.25514.
5. Sivachev A.V., Prohozhev N.N., Mihajlichenko O.V., Bashmakov D.A. Effektivnost' steganoanaliza na osnove metodov mashinnogo obucheniya // Voprosy kiberbezopasnosti. 2017. № 2 (20). S. 53-60. DOI: 10.21681/2311-3456-2017-2-53-60.
6. Sivachev A.V. Effektivnost' statisticheskih metodov steganoanaliza pri obnaruzhenii vstraivaniya v vejvlet oblast' izobrazheniya // Voprosy kiberbezopasnosti. 2018. № 1 (25). S. 72-78. DOI: 10.21681/2311-3456-2018-1-72-78.
7. Sirota A.A., Dryuchenko M.A., Ivankov A.YU. Stegoanaliz cifrovyh izobrazhenij s ispol'zovaniem metodov poverhnostnogo i glubokogo mashinnogo obucheniya: izvestnye podhody i novye resheniya // Vestnik VGU, Seriya: Sistemnyj analiz i informacionnye tekhnologii. 2021. № 1. S. 33-52. DOI:10.17308/sait.2021.1/3369.
8. Monarev A.I., Pestunov V.A. Effektivnoe obnaruzhenie steganograficheski skrytoj informacii posredstvom integral'nogo klassifikatora na osnove szhatiya dannyh // Prikladnaya diskretnaya matematika. 2018. № 40. S.59-71. DOI: 10.17223/20710410/40/5.
9. Makarenko S.I. Etalonnaya model' vzaimodejstviya steganograficheskih sistem i obosnovanie na ee osnove novyh napravlenij razvitiya teorii steganografii // Voprosy kiberbezopasnosti. 2014. №2 (3). C. 24-32.
10. Parasich A. V., Parasich V. A., Parasich I. V. Formirovanie obuchayushchej vyborki v zadachah mashinnogo obucheniya // Informacionno-upravlyayushchie sistemy. 2021. № 4. S.61-70. DOI:10.31799/1684-8853-2021-4-61-70.
11. Atlasov I., Solodukha R. Sample Representativeness Estimation as a Preliminary Stage of Statistical Steganalysis / 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), 2021, pp. 78-84, DOI: 10.1109/SUMMA53307.2021.9632204.
12. Atlasov I., Solodukha R. The mathematical model of estimation the multidimensional steganalytical methods reliability / Journal of Physics: Conf. Series. 2019. Vol. 1202. DOI: 10.1088/1742-6596/1202/1/012022.
13. Soloduha R.A. Koncepciya formirovaniya sistemy protivodejstviya steganograficheskim kanalam v komp'yuternyh setyah organov vnutrennih del // Vestnik Voronezhskogo instituta MVD Rossii. 2021. № 1. S.131-142.
14. Vil'hovskij D.E. Obzor metodov steganograficheskogo analiza izobrazhenij v rabotah zarubezhnyh avtorov // Matematicheskie struktury i modelirovanie. 2020. № 4(56). S. 75–102. DOI: 10.24147/2222-8772.2020.4.75-102.
15. Belim S.V., Vilkhovskiy D.E. Algorithm for detection of steganographic inserts type LSB-substitution on the basis of an analysis of the zero layer // Journal of Physics: Conf. Series. 2017. Vol. 944. DOI:10.1088/1742-6596/944/1/012012.
16. Grachev YA.L., Sidorenko V.G. Stegoanaliz metodov skrytiya informacii v graficheskih kontejnerah // Nadezhnost'. 2021. № 21(3). S. 39-46. DOI: 10.21683/1729-2646.