САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

На правах рукописи

Башмаков Даниил Андреевич

МЕТОДЫ И АЛГОРИТМЫ ВЫЯВЛЕНИЯ ВСТРОЕННЫХ СООБЩЕНИЙ В ПРОСТРАНСТВЕННОЙ ОБЛАСТИ НЕПОДВИЖНЫХ ИЗОБРАЖЕНИЙ ПРИ МАЛОЙ ПОЛЕЗНОЙ НАГРУЗКЕ

Специальность 05.13.19 – Методы и системы защиты информации, информационная безопасность

Научный руководитель: д.т.н., профессор

Коробейников Анатолий Григорьевич

Санкт-Петербург – 2018

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Актуальность темы. Методы встраивания информации в контейнеры всех видов (изображения, аудио, видео и другие) находят широкое применение в современном мире. Встраивание информации применяется как в легальных целях, так и в ходе противоправной деятельности.

К легальным применениям встраивания информации можно отнести цифровые водяные знаки в документах и объектах интеллектуальной собственности; организацию каналов скрытной передачи информации спецслужбами в рамках их деятельности; организацию каналов скрытной передачи информации в прочих целях (сохранение коммерческой тайны, тайны переписки и т. д.)

К наиболее распространённым способам противоправного применения техник встраивания информации относятся организация каналов скрытной передачи информации в целях организации и координации противоправной деятельности, в частности, террористических актов; организация каналов скрытной передачи информации с целью сокрытия факта выведения информации за охраняемый периметр, в частности, в рамках промышленного шпионажа [1, 2].

Выявление встроенных сообщений в контейнерах различной природы находит применение как в целях противодействия противоправным способам применения встраивания, так и вкупе с легальными применениями, в качестве инструмента контроля и проверки [3]. В частности, методы выявления встроенных сообщений могут применяться в задачах выявления цифровых водяных знаков с целью доказательства модификации контейнера в ситуации, когда извлечение сообщения штатными средствами невозможно (например, в случае недоступности ключевой информации встраивания); пассивного противодействия каналам скрытной передачи информации, используемых в нелегальных целях, в части обнаружения и классификации таких каналов;

противодействия промышленному шпионажу в части обнаружения факта попытки выведения охраняемой информации за периметр охраняемой зоны [4].

Неподвижные цифровые изображения являются одним из наиболее распространённых видов информации, передаваемой в сети интернет. Изображения могут выступать контейнером для встраивания сообщений. На сегодняшний день разработано множество способов встраивания в неподвижные цифровые изображения.

В условиях растущего разрешения изображений, отношение объёма встраиваемой полезной нагрузки и объёма изображения (далее также – отношение нагрузка-контейнер, ОНК) как стеганоконтейнера постоянно снижается. В данной работе рассматривается задача выявления факта встраивания информации методом стеганографии в неподвижные цифровые изображения и эффективность методов такого выявления (методов стеганодетектирования) при малых значениях полезной нагрузки.

Методы встраивания подразделяются в зависимости от пространства (домена) изображения [5, 6]. Пространственный домен изображения, то есть матрица уровней яркости его пикселей является одним из наиболее распространённых доменов для стеганографического встраивания [7]. Несмотря на то, что встраивание в плоскость НЗБ не лишено недостатков, высокая скрытность и большой объём контейнера обеспечивает его популярность и на сегодняшний день. Существует множество методов выявления в пространственном домене изображения в плоскости НЗБ [8]. Однако при малых значениях полезной нагрузки методы показывают низкую эффективность [9, 10]. Это, в свою очередь, не позволяет эффективно противодействовать скрытным каналам передачи информации, основанным на встраивании в пространственный домен изображения. Таким образом, повышение эффективности выявления факта встраивания в плоскость НЗБ является актуальной задачей [11].

Тенденция постоянного роста размеров стеганоконтейнеров вкупе с представлением о снижении эффективности выявления со снижением значения отношения нагрузка-контейнер определяет требование к методам выявления встроенных сообщений по повышению эффективности на малых значениях ОНК, как общей необходимой реакции на изменение условий противодействия каналам передачи данных, основанных на встраивании информации.

Задачи повышения эффективности детектирования факта встраивания в плоскость НЗБ рассматривались в работах как русских (В.Г. Грибунин, В. И. Коржик и другие), так и зарубежных учёных (А. Кер, Дж. Фридрич, М. Гольян). Таким образом, обеспечение высокой эффективности детектирования факта встраивания в плоскость НЗБ является актуальной задачей [3].

Степень разработанности темы. В работах В.И. Коржика обоснован общий подход по применению методов статистического стеганоанализа в задаче выявления встроенных сообщений в неподвижных изображениях. Основой для диссертационного исследования послужили работы Д. Фридрич, А. Д. Кера и М. Гольяна по разработке методов стеганодетектирования в пространственной области неподвижных цифровых изображений. В ходе работы выполнен анализ существующих методов повышения эффективности стеганодетектирования, предложенных в работах Р. Бёме, Ю. Сяо, Б. Нобору, П. Шоттле и других. Также заделом для диссертационного исследования послужили работы В. Г. Грибунина, Р. М. Юсупова, А. В. Аграновского, И. В. Туринцева, А.А. Молдовяна, Н.А. Молдовяна, И. Н. Окова, Г. Ф. Кохановича и других исследователей.

Целью работы является повышение эффективности выявления встроенных сообщений в НЗБ пикселей неподвижных изображений при малой полезной нагрузке в интересах обеспечения защищённости информации путём предотвращения её утечек по каналам передачи информации на основе стеганографии.

Для достижения поставленной цели в диссертационной работе решалась **научная задача** по разработке модели, алгоритмов и метода выявления встроенных сообщений в плоскости наименьших значащих бит пространственной области неподвижных изображений при малой полезной нагрузке.

Поставленная научная задача декомпозируется на следующие частные задачи:

1. Анализ существующих методов выявления встроенных сообщений в НЗБ неподвижных цифровых изображений при малой полезной нагрузке и выявление среди них наиболее эффективного на сегодняшний день.

2. Разработка модели выявления встроенных сообщений методом, определённым в задаче 1 в условиях малой полезной нагрузки, анализ модели и определение направлений по усовершенствованию метода.

3. Разработка алгоритмов, обеспечивающих повышение эффективности выявления встроенных сообщений в НЗБ неподвижных изображений при малой полезной нагрузке.

4. Разработка метода выявления встроенных сообщений в НЗБ фоновых зон неподвижных цифровых изображений с повышенной точностью при малой полезной нагрузке.

5. Экспериментальное подтверждение повышенной эффективности стеганодетектирования при применении метода выявления встроенных сообщений, разработанного в задаче 4.

Научная новизна положений, выносимых на защиту, состоит в следующем:

1. Разработанная модель выявления встроенных сообщений в наименьших значащих битах фоновых зон пространственной области неподвижных изображений отличается от существующих фокусом на особых семантических областях анализируемого изображения - фоновых зонах. Выделение самостоятельной модели выявления в фоновых зонах, анализ

зависимости эффективности детектирования от особенностей работы метода в фоновых зонах изображения проведены впервые.

2. Алгоритмы выявления встроенных сообщений в НЗБ фоновых зон неподвижных изображений обладают новизной за счёт задействования в алгоритмах крупных структур анализируемых пикселей, специфичных для фоновых зон естественных изображений. В отличие от существующих алгоритмов, разработанные оперируют способами выделения соседства пикселей в фоновых зонах изображения (в отличие от DIH, WS, SPAM и др.), сочетая это с задействованием накопленной статистики для использования в процессе анализа контейнера (в отличие от RS, SPA и др.)
3. Разработанный метод повышения эффективности выявления встроенных сообщений за счёт специальных алгоритмов прогноза значений пикселей в фоновых зонах, обладает новизной по сравнению с известными методами повышения эффективности за счёт:
   — прогнозирования значений пикселей анализируемого изображения с точностью, критичной при выявлении на малых значениях полезной нагрузки;
   — применения алгоритма выделения фоновой зоны изображения, специфичной в задаче выявления методом WS.

Теоретическая и практическая значимость работы. Использование метода выявления встроенных сообщений, предложенного в работе, в системах защиты информации, в частности, в компонентах пассивного противодействия каналам передачи данных, основанных на стеганографии в плоскости НЗБ неподвижных цифровых изображений, позволят повысить уровень защищённости информации за счёт снижения вероятности реализации риска её несанкционированной утечки по таким каналам.

Методология исследования заключается в постановке и формализации задач, связанных с оценкой эффективности методов и алгоритмов выявления встроенных сообщений, описании модели сущностей, используемых для

проведения оценок, разработке модели, методов и алгоритмов выявления встроенных сообщений в неподвижных изображениях, апробации полученных теоретических результатов посредством сравнительного анализа их реализаций с существующими решениями с получением количественных и качественных сравнительных оценок.

Методы исследований. Поставленные задачи решены на основе применения теории защиты информации, теории вероятности и математической статистики, методов дискретной математики.

В соответствии с заявленными целью и задачами работы, **объектом исследования** являются контейнеры для встраивания, являющиеся неподвижными изображениями с информацией, встроенной в наименьшие значащие биты пространственной области.

Предметом исследования являются методы и алгоритмы выявления встроенных сообщений в неподвижных изображениях при малой полезной нагрузке.

На защиту выносятся следующие основные положения:

1. Модель выявления встроенных сообщений в наименьших значащих битах фоновых зон пространственной области неподвижных изображений при малой полезной нагрузке обеспечивает оптимальный подход к выявлению встроенных сообщений в фоновых зонах.

2. Алгоритмы выявления встроенных сообщений в наименьших значащих битах фоновых зон пространственной области неподвижных изображений при малой полезной нагрузке обеспечивают повышенную точность прогноза пикселей анализируемого изображения в фоновых зонах.

3. Метод выявления встроенных сообщений в наименьших значащих битах пространственной области неподвижных изображений обеспечивает повышенную эффективность выявления встроенных сообщений при малой полезной нагрузке.

Достоверность полученных результатов достигается путём использования апробированного математического аппарата, использованием достоверных исходных данных, системным подходом при описании объекта исследования, проведением сравнительного анализа полученных результатов с существующими показателями, использованием проверенных практик в оценке эффективности методов выявления, результатами практических экспериментов.

Апробация результатов. Основные результаты работы представлялись на следующих конференциях:

- Всероссийская научно-практическая конференция с международным участием «Информационные технологии в профессиональной деятельности и научной работе», 2014 г.
- III Всероссийский конгресс молодых учёных, 2014 г.
- Всероссийский студенческий форум «Инженерные кадры - будущее инновационной экономики России», 2015 г.
- V Всероссийский конгресс молодых учёных, 2016 г.
- VI Всероссийский конгресс молодых учёных, 2017 г.

Публикации по теме диссертации. По результатам диссертационного исследования опубликовано 9 работ, из них 5 работ в журналах, входящих в перечень ВАК и 1 работа в журнале, индексируемом в международной базе цитирования Scopus.

Внедрение результатов работы:

- Институт земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова Российской академии наук ИЗМИРАН (северо-западный филиал в Санкт-Петербурге);
- Акционерное общество «Опытно-конструкторское бюро «Электроавтоматика» имени П.А. Ефимова»;
- Университет ИТМО, учебная дисциплина «Стеганографические методы защиты информации».

Структура диссертации. Диссертация состоит из введения, четырёх глав, заключения, списка литературы, состоящего из 104 пунктов, включающих труды автора, и двух приложений. Материал изложен на 150 страницах машинописного текста, содержит 39 рисунков и 10 таблиц.

Глава 1. В первой главе выполняется краткий обзор существующих на данный момент методов выявления сообщений в неподвижных цифровых изображениях. Приводится место выявления изображений в плоскости НЗБ в общей структуре методов выявления. Описываются требования, предъявляемые к методу выявления. Рассматривается вопрос оценки эффективности метода выявления, предлагается формальная метрика оценки, позволяющая сравнить эффективность нескольких методов выявления между собой. Предлагается подход к графическому представлению эффективности метода выявления, позволяющий наглядно отобразить эффективность нескольких методов на графике в единых осях. Приводится обзор существующих методов выявления сообщений в неподвижных цифровых изображениях, выбираются методы для дальнейшего углубленного исследования.

Глава 2. Рассматривается понятие домена цифрового изображения, анализируются характеристики пространственного домена цифрового изображения. Для оценки эффективности методов предлагается модель стеганографического встраивания и выявления встроенного сообщения в НЗБ неподвижного цифрового изображения и набора таких изображений. Определяются критерии применимости метода в задаче эффективного пассивного противодействия стеганографическому каналу скрытной передачи информации. С учётом предложенной модели, проведено углубленное исследование эффективности методов выявления, отобранных в Главе 1. Ключевыми выводами является недостаточность эффективности существующих алгоритмов выявления при малых значениях полезной нагрузки в задаче организации эффективного канала противодействия стеганографическим каланам скрытной передачи информации, а также определение наиболее

эффективного метода выявления из предложенных – метода Weighted Stego (WS).

Глава 3. Рассматриваются факторы, влияющие на эффективность выявления сообщения в цифровом изображении методом WS. Анализируется математический аппарат метода WS, выводится связь между эффективностью анализа и точностью предсказания пикселей анализируемого изображения. Вводится понятие точности прогноза пикселя изображения. Показывается связь между высокоуровневой семантикой анализируемого изображения и эффективностью анализа методом WS. Разрабатывается модель выявления в фоновых зонах однородного фона изображения. На основе разработанной модели показывается связь эффективности выявления в неподвижном изображении методом WS с точностью прогноза значения пикселя в фоновых зонах анализируемого изображения. Предлагаются алгоритмы прогноза значений пикселей в фоновых зонах с повышенной точностью по сравнению с известными, учитывающие высокоуровневую семантику анализируемого изображения, высокоуровневые статистики изображения и группы изображений.

Глава 4. Предлагается алгоритм выделения фоновых зон изображения. Предлагается метод выявления на основе метода WS с повышенной эффективностью за счёт использования предложенных в главе 3 алгоритмов прогноза значений пикселей фоновых зон. Приводятся результаты практических экспериментов с целью оценки эффективности предложенных методов выявления в пространственной области. Проводится сравнительный анализ эффективности существующих методов и предложенного метода. Подтверждается повышенная эффективность существующего метода при малых значениях полезной нагрузки.

Личный вклад. Положения, выносимые на защиту, отражают личный вклад автора в данную работу.

ГЛАВА 1. МЕТОДЫ ВЫЯВЛЕНИЯ ВСТРОЕННЫХ СООБЩЕНИЙ И МЕТОДИКА ОЦЕНКИ ИХ ЭФФЕКТИВНОСТИ

1.1.Введение в предметную область

1.1.1. Терминология

В настоящее время различные источники по-разному определяют понятия, используемые в работах по стеганографии и выявлению встроенных сообщений. В дальнейшем в данной работе будут использованы следующие понятия:

Стеганоконтейнер, *контейнер* – документ, существующий в виде цифрового файла, в который методами стеганографии внедряется информация. Стеганоконтейнер без внедрённой информации называется пустым контейнером. Стеганоконтейнер с внедрённой информацией называется стеганопосылкой. Термин стеганоконтейнер применяется в случае, если неизвестно, внедрялась ли в него информация.

Тип стеганоконтейнера определяется в зависимости от характера представления и восприятия пользователем информации, которую он представляет: неподвижные цифровые изображения, видеопоследовательности, аудиопоследовательности и т. д.

Сообщение (также – *внедрённое сообщение*, *нагрузка*) – информация, внедряемая методами стеганографии в стеганоконтейнер с целью скрытной передачи.

Тип сообщения, так же, как и тип стеганоконтейнера, определяется по характеру представления и восприятия пользователем информации, представляемой сообщением.

Встраивание, внедрение – операция добавления сообщения в стеганоконтейнер методами стеганографии [89].

Полезная нагрузка, отношение нагрузка-контейнер, ОНК – отношение размера сообщения к размеру контейнера, выраженное в процентах. Метод вычисления размеров сообщения и контейнера определяется индивидуально для каждого типа сообщения и контейнера [104].

Выявление встроенного сообщения (стеганодетектирование) – процесс определения факта наличия встроенного сообщения в контейнере, не включающий задачи его локализации в анализируемом контейнере и выделения сообщения из анализируемого контейнера [89].

Метод выявления встроенного сообщения (метод ВВС) – метод, позволяющий для стеганоконтейнера сделать вывод о наличии или отсутствии внедрённого сообщения, а также, опционально, получить оценку длины встроенного сообщения [90].

Система выявления встроенного сообщения (система ВВС) – программное решение, основанное на реализации метода выявления, позволяющее по документу отнести его к пустым контейнерам либо стеганопосылкам [24].

Стеганоканал – канал передачи данных в виде стеганоконтейнеров (пустых либо стеганопосылок) [24].

Эффективность метода выявления встроенного сообщения – мера достоверности при отнесении анализируемого стеганоконтейнера к пустым контейнерам либо к стеганопосылкам. Метод расчёта эффективности определяется в зависимости от типа контейнера, сообщения и метода выявления [24].

Пространственная область изображения – упорядоченная совокупность уровней яркости пикселей, составляющих изображение [62, 89, 103].

Неподвижное изображение – цифровое изображение, пространственная область распределения которого не изменяется с течением времени [89].

1.1.2. Типы стеганоконтейнеров

В настоящее время в сетях передачи данных циркулирует цифровой контент, представленный в самых различных формах. Высокие объёмы передаваемой информации позволяют использовать в качестве стеганоконтейнеров документы следующих типов [14, 12]: неподвижное изображение, видео, аудио, текст, бинарные данные.

Методы внедрения сообщения в контейнер основываются на эксплуатации специфических особенностей реализации того или иного формата представления цифрового контента, либо на особенностях контента как такового (например, на избыточности данных в представлении неподвижного цифрового изображения). Поэтому алгоритмы внедрения сообщения, и, следовательно, противостоящие им методы ВВС различаются от формата к формату и для различных типов цифрового контента.

Среди стеганоконтейнеров большое распространение получили неподвижные цифровые изображения [15, 13]. Неподвижное изображение представлено в виде матрицы чисел, представляющих уровни яркости точек (пикселей) того или иного цветового слоя изображения. Несмотря на то, что в последнее десятилетие доля трафика в сети Интернет, приходящаяся на аудио и видео, значительно возросла, неподвижное изображение остаётся наиболее популярным стеганоконтейнером. Это обусловлено следующим рядом причин [15, 16]:

- большие объёмы данных, передаваемых в сети Интернет в виде цифровых изображений, позволяют добиваться высокой пропускной способности стеганоканала даже при небольших значениях полезной нагрузки;
- цифровое представление неподвижного изображения имеет относительно большой объём, из-за чего возможно внедрение больших по объёму сообщений с сохранением малых значений полезной нагрузки;
- избыточность представления цифрового изображения позволяет использовать значительные области данных для внедрения сообщения при сохранении полной или практически полной незаметности для наблюдателя;
- большое разнообразие и хорошая проработанность методов внедрения сообщений в неподвижное изображение.

Следует отметить, что последняя причина обуславливает большое поле для научной работы в области выявления встроенных сообщений в неподвижных цифровых изображениях.

В настоящее время наблюдается постоянный рост разрешения изображений, передаваемых в сети интернет. Это позволяет, используя неподвижные цифровые изображения в качестве стеганоконтейнера, добиться высокой пропускной способности стеганоканала даже при малых значениях полезной нагрузки.

1.1.3. Области встраивания в неподвижные изображения

Методы встраивания и ВВС в неподвижных цифровых изображениях различаются в зависимости от формата изображения. Различные форматы

используют тот или иной способ декомпозиции содержимого изображения для представления его в цифровом виде для хранения, передачи и представления изображения в цифровом виде. В настоящее время наибольшее распространение (в общем и в стеганографии в частности) получили способы представления, основанные на следующих методах декомпозиции [17]:

- без декомпозиции (данные представляются в пространственном распределении) – используется такими форматами, как BMP, PNG;
- дискретно-косинусное преобразование – используется в стандарте JPEG;
- вейвлет-преобразование – используется в стандарте JPEG2000.

Несмотря на большое распространение форматов файлов, основанных на декомпозиции данных изображения, методы встраивания сообщения в пространственную область остаются популярными. Это связано с простотой представления данных, большим объёмом данных для встраивания, большим количеством и проработанностью методов внедрения сообщения. Также, популярность встраивания в пространственную область обусловлена тем, что, вне зависимости от используемого метода декомпозиции, любая система, представляющая изображение, способна представить его в виде пространственного распределения для показа пользователю. Таким образом, смена формата изображения в течение его жизненного цикла не влияет на способность представить изображение в виде пространственного распределения. Наконец, практическая применимость методов, основанных на внедрении сообщения в пространственную область, обуславливается распространённостью форматов представления изображения, использующих сжатие без потерь, не влияющих на область пространственного распределения (таких, как PNG).

При представлении изображения в виде пространственного распределения (также – пространственного домена) информация, составляющее изображение,

представлена в виде двумерной матрицы значений, означающих яркость точки изображения. Размеры матрицы соответствуют количеству отдельных точек изображения. Элемент матрицы пространственного распределения называется пикселем. Количество пикселей по ширине и высоте изображения называется разрешением изображения.

В случае представления цветного изображения, каждому цветовому слою изображения соответствует отдельная матрица пространственного распределения. В зависимости от используемой модели, цветное изображение может быть представлено различным набором цветовых слоёв. Наиболее распространённые модели RGB и CMYK используют 3 и 4 цветовых слоя соответственно.

Наибольшее распространение получил способ представления пространственной области, при котором каждый пиксель представляет собой число длиной в 1 байт, то есть, находящееся в диапазоне от 0 до 255, где 0 соответствует наименьшей яркости пикселя, а 255 – наибольшей. При такой точности представления яркости пикселя пространственная область избыточна – незначительное изменение значений яркости пикселей не приводят к проявлению видимых человеческим глазом артефактов в изображении. Этот факт используется рядом методов встраивания в пространственную область неподвижных цифровых изображений [18].

1.1.4. Методы внедрения в пространственную область цифрового изображения

Методы внедрения в пространственную область эксплуатируют избыточность данных при представлении неподвижного цифрового изображения в виде матриц плоскостей цветовых компонент. Такие алгоритмы оперируют значениями элементов этих матриц, изменяя их так, чтобы при извлечении из

абсолютных или относительных значений элементов матрицы можно было извлечь внедрённое сообщение. Алгоритмы различаются по конкретным способам, с помощью которых производится изменение значений пикселей.

Методы внедрения в наименьший значащий бит, или *LSB-методы* (Least significant bit – наименьший значащий бит) состоят в представлении внедряемого сообщения в виде последовательности бит и замены наименьших значащих бит пикселей одной из цветовых плоскостей изображения на биты сообщения. При сравнительно большой ёмкости контейнера при использовании такого метода, LSB-методы уязвимы к методам статистического ВВС [19, 20].

Методы, основанные на добавлении шума (Noise-adding Based Steganography), так же называемые в литературе методами стохастической модуляции, призваны снизить уязвимость к статистическому ВВС. Среди них выделяется метод *LSBM* (LSB Matching), являющийся простейшей модификацией LSB-методов и позволяющий значительно снизить уязвимость к методам статистического ВВС [21].

Группа методов, основанных на предсказании ошибки (Prediction error based steganography, BPCS) использует предварительный анализ изображения с целью выбрать области, встраивание в которые приведёт к наименьшей заметности сообщения для СЧЗ и статистического ВВС [22].

Методы, основанные на квантизации (Quantization based steganography), наиболее применимые в задачах встраивания цифровых водяных знаков, также могут быть использованы в стеганографии. Несмотря на ограниченную применимость в области пространственного распределения, в некоторых случаях использование методов этой группы позволяет добиться большой незаметности встроенного сообщения для методов статистического ВВС [23].

Среди представленных, LSB-методы и их модификации, такие, как LSBM, в настоящее время получили наибольшую распространённость в стеганографии в области пространственного распределения. Это связано с рядом причин, среди которых [19, 20]: большая ёмкость контейнера, простота и интуитивная понятность методов внедрения и извлечения сообщения, вычислительная простота методов внедрения и извлечения сообщения, высокая скрытность при малых значениях полезной нагрузки для методов ВВС, рост разрежений изображений напрямую влияет на рост стеганоконтейнера и позволяет встраивать более длинные сообщения с теми же значениями полезной нагрузки;

1.2. Выбор методов статистического выявления встроенных сообщений в неподвижных изображениях

1.2.1. Статистическое выявление встроенных сообщений

Выявление встроенных сообщений может быть определено как задача классификации по двум категориям [25]. Цель ВВС – определить, является ли анализируемое изображение пустым контейнером или стеганопосылкой. Наиболее общее деление методов ВВС основывается на их универсальности. Методы ВВС делятся на три следующие категории:

- специфичные – позволяют обнаруживать сообщение, внедрённое с использованием конкретного метода стеганографии (или семейства близкородственных методов);
- универсальные – эксплуатируют наиболее общие закономерности изменения изображения-контейнера в процессе встраивания и позволяют обнаруживать встроенное сообщение вне зависимости от использованного метода внедрения;

– полу-универсальные методы позволяют обнаруживать сообщение, внедрённое с применением широкого диапазона методов стеганографии, но неэффективные в частных случаях, особых областях распределения и т. п.

Общей закономерностью является то, что эффективность специфичных методов ВВС на области их «компетенции» оказывается выше, чем эффективность универсальных алгоритмов на той же области [25]. В дальнейшем в данной работе рассматриваются специфичные методы ВВС.

Понятие статистического ВВС связано с процессом вычисления характеристик высшего порядка анализируемого изображения, их анализа и определения отклонения их значений от стандартных [26]. Методы статистического ВВС эксплуатируют тот факт, что характеристики высшего порядка большинства неподвижных изображений, имеющих одинаковую природу, близки или вовсе одинаковы. Несмотря на, как правило, высокую вычислительную сложность, методы статистического ВВС демонстрируют высокую эффективность. В дальнейшем в данной работе рассматриваются методы статистического ВВС.

Методы статистического ВВС в пространственной области неподвижных цифровых изображений оценивают длину встроенного сообщения. Далее в зависимости от принятой пороговой длины, делается вывод об отнесении анализируемого стеганоконтейнера к чистым контейнерам либо стеганопосылкам.

1.2.2. Требования, предъявляемые к методам выявления встроенных сообщений

В различных источниках приводятся следующие требования к методам статистического ВВС [1, 6, 14, 15, 27]:

1. Эффективность. Алгоритм должен как можно с большей вероятностью корректно определять наличие или отсутствие внедрённого сообщения. Малый процент ошибочных определений часто рассматривается в качестве основного преимущества алгоритма статистического ВВС [24].

2. Универсальность по форматам анализируемого изображения и методам внедрения сообщения. Алгоритм должен позволять обнаруживать внедрённое сообщение для как можно большего числа форматов изображений и методов внедрения. В наши дни большое количество форматов представления цифровых изображений, а также обилие и проработанность методов цифровой стеганографии делают это требование важным.

3. Максимальная независимость от содержательных характеристик анализируемого изображения. Алгоритм не должен демонстрировать значительного снижения эффективности при анализе изображений со специфичными зонами, такими, как большие однородные зоны или, напротив, слишком зашумлённых изображений.

4. Вычислительная простота. Алгоритм должен справляться с задачей обнаружения встроенного сообщения за приемлемое время. Рост разрешений и количества изображений, передаваемых в сети Интернет и используемых в качестве контейнеров для стеганографии, сделал актуальной проблему «проклятия размерности» (Curse of dimensionality). Несмотря на то, что данное требование не относится к основным, в ряде работ такая характеристика алгоритма, как нечувствительность к «проклятию размерности» выставляется преимуществом алгоритма.

5. Устойчивость к модификациям методов внедрения. Алгоритм ВВС должен эксплуатировать наиболее фундаментальные зависимости и характеристики

процесса встраивания, чтобы оставаться робастным к незначительным модификациям метода встраивания.

Несмотря на то, что все приведённые требования могут быть учтены в задаче выбора оптимального метода ВВС, именно эффективность является определяющим требованием, поскольку именно эффективность (по своему определению) отражает способность достоверно выявлять встроенные сообщения, а прочие приведённые требования уточняют показатели эффективности в различных условиях применения метода ВВС. Под эффективностью метода ВВС понимается способность достигать определённого процента корректных классификаций в задаче различения стеганопосылки и чистого сообщения.

Независимость от семантики анализируемого изображения подразумевает под собой способность показывать близкие значения эффективности при анализе изображений, содержащих определённые специфические зоны, например, однородный фон, или, напротив, сильно зашумлённые зоны. Под семантикой изображения понимается наличие и распределение в пространственной области изображения определённых зон, удовлетворяющих требованиям по характеру распределения значений яркости пикселей в них (семантических зон). Факт падения эффективности алгоритма при анализе изображений определённой семантики может быть использован для организации атаки на алгоритм за счёт использования только изображений, содержащих зоны, при анализе которых алгоритм показывает низкую эффективность. Таким образом, независимость от семантики анализируемого изображения является значимым требованием к методу ВВС.

На основании анализа имеющихся публикаций сделан вывод о том, что наибольшую эффективность при анализе стеганопосылок, полученных при помощи определённого метода стеганографии показывают методы ВВС, разработанные под этот конкретный метод встраивания [26, 29, 41]. Иными словами, на сегодняшний день универсальные алгоритмы показывают меньшую

эффективность для определённого отдельно взятого метода встраивания, чем методы ВВС, разработанные конкретно для этого метода. То же верно для универсальности по форматам анализируемого изображения. Таким образом, сделан вывод что эффективность и универсальность по форматам и методам встраивания противоречат друг другу.

С учётом приведённых выше аргументов, в дальнейшем при выборе методов ВВС для дальнейшего исследования и улучшения проводится по следующим критериям:

- возможность обнаружения сообщения, встроенного методами LSB или LSBM, как наиболее распространённых методов встраивания в неподвижные цифровые изображения;
- отсутствие «крайних случаев», для которых алгоритм демонстрирует значительно меньшую эффективность, чем заявлена авторами;
- высокая или средняя эффективность по сравнению с аналогами для данного метода встраивания.

1.2.3. Выбор методов статистического выявления встроенных сообщений

Для оценки эффективности методов и предложения усовершенствований, требуется выбрать ряд методов, удовлетворяющих вышеприведённым критериям. Таблица 1 содержит методы, доступные из открытых источников, с приведением их сильных и слабых сторон.

По методам встраивания, на которые нацелен алгоритм, подходят Triples analysis, Chi-Square, RS-analysis, Sample pairs analysis, Weighted stego, HCF-COM, SPAM, Zhang-Cox, Difference Image Histogram, Zhang-Wang.

С учётом приведённых выше критериев, в дальнейшем будет проведён сравнительный анализ эффективности при различных параметрах изображений и различных значениях полезной нагрузки следующих методов: Triples analysis, Sample pairs analysis, RS-analysis, Weighted stego, Difference Image Histogram.

Таблица 1

Методы ВВС в пространственной области

Метод	**На какой метод нацелен**	**Преимущества**	**Недостатки**
1	2	3	4
Triples Analysis [28]	LSB	Простота реализации, вычислительная простота	Малая эффективность
Chi-square [29]	LSB	Также эффективен в областях преобразования	Малая эффективность
RS-analysis [30]	LSB	Высокая эффективность при малых значениях полезной нагрузки	Эффективность зависит от применяемой маски
Sample pair analysis [31]	LSB	Высокая эффективность при больших и средних значениях полезной нагрузки	Малая эффективность при малых значениях полезной нагрузки
Weighted Stego [32]	LSB	Высокая эффективность на широком диапазоне значений полезной нагрузки	Высокая вычислительная сложность
HCF-COM [33]	LSB, LSBM	Простота реализации, хорошая эффективность на цветных изображениях	Очень низкая эффективность на изображениях в оттенках серого
SPAM [34]	LSB, LSBM	Не подвержен «проклятию размерности»	Приемлемая эффективность только для больших значений полезной нагрузки

Таблица 1 - окончание

Zhang-Cox [35]	LSBM	Высокая эффективность	Только для изображений в оттенках серого
He-Huang (семейство алгоритмов) [36]	Методы стохастической модуляции	Не имеет альтернатив в анализе внедрения методами стохастической модуляции	Не работает для LSB, высокая эффективность только на ограниченных тестовых выборках
Niimi [37]	BPCS-методы	Высокая эффективность	Бесполезен при использовании LSB-методов встраивания, узкая направленность
Zhang-Wang [38]	Методы, основанные на предсказании ошибки	Также применим в LSB	Неприменим в LSBM
Difference Image Histogram [39]	LSB	Простота реализации, низкая вычислительная сложность	Малая эффективность при малых значениях полезной нагрузки, существенно сниженная эффективность при анализе цветных изображений

1.3. Методика оценки эффективности метода выявления встроенных сообщений

1.3.1. Понятие эффективности метода выявления встроенных сообщений

Для сравнения существующих методов ВВС, а также для практического подтверждения предлагаемых усовершенствований требуется определить метод оценки эффективности статистического ВВС, описать метод получения численной оценки эффективности. Поскольку эффективность как таковая – комплексная характеристика, разнящаяся в зависимости от многих параметров системы и анализируемых данных, требуется также определить метод наглядного представления эффективности метода ВВС, позволяющий показать различия в эффективности тех или иных методов ВВС без отсылки к объёмным таблицам числовых характеристик.

Метод статистического ВВС в неподвижном цифровом изображении может быть рассмотрен с двух позиций:

- Как инструмент оценки длины сообщения, встроенного в анализируемое изображение [40];
- Как инструмент бинарной классификации, позволяющий отнести анализируемое изображение к стеганопосылкам или к чистым изображениям [41].

1.3.2. Эксперимент по оценке эффективности метода выявления встроенного сообщения

Эффективность метода ВВС зависит не только от него самого, но и от характеристик данных, которые подвергаются анализу с его помощью. Для оценки эффективности требуется провести эксперимент, состоящий в выявлении сообщений в данных определённого набора исследуемым методом. На основании результатов эксперимента оценивается эффективность метода.

Следующий сценарий эксперимента используется в работе для определения эффективности метода ВВС:

1. Подготавливается выборка изображений. В качестве выборок использованы наборы изображений коллекции BOWS2. Изображения в тестовой выборке независимы друг от друга.
2. Для части изображений имитируется стеганографическое встраивание в LSB. При этом, в зависимости от цели эксперимента, может соблюдаться постоянство величины полезной нагрузки либо длины встраиваемого сообщения.
3. Изображения тестовой выборки анализируются с помощью метода ВВС.
4. Оценка длины встроенного сообщения либо факта наличия встраивания сличается с действительной длиной сообщения и фактом его наличия.
5. На основании результатов сличения строится статистика эффективности работы анализируемого метода ВВС.

С учётом вышеизложенного, разработан алгоритм проведения эксперимента по оценке эффективности метода ВВС. Блок-схема алгоритма приведена на рисунке 1.

1.3.3. Способы интерпретации результатов эксперимента по оценке эффективности метода ВВС

В качестве метрики оценки эффективности метода как средства оценки длины встроенного сообщения, предлагается использовать среднюю ошибку прогноза длины сообщения в пикселях при анализе набора изображений:

$$E_L = \frac{\sum_1^N |L_a - L_p|}{N}, \text{где}$$

$N -$ число анализируемых последовательно изображений,

$L_a -$ оцененная методом выявления длина встроенного сообщения,

$L_p -$ действительная длина встроенного сообщения

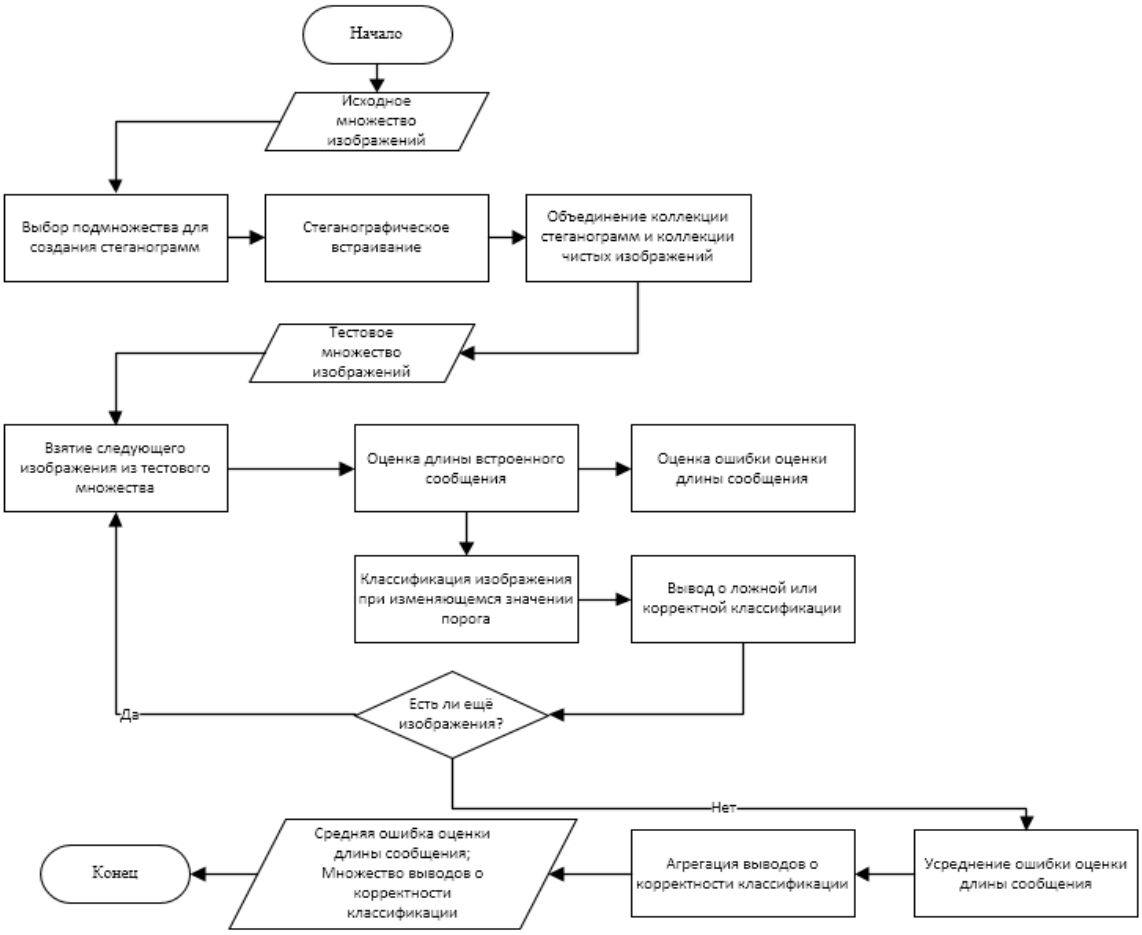

Рисунок 1 - Алгоритм эксперимента по оценке эффективности метода ВВС

Для оценки эффективности метода ВВС как бинарного классификатора используется существующий матаппарат математической статистики [43, 44]. Вводится понятие корректной и некорректной классификации:

- Если стеганоконтейнер, являющийся стеганопосылкой, классифицирован как пустой стеганоконтейнер, либо, напротив, пустой контейнер классифицирован как стеганопосылка, это – некорректная классификация.
- Если пустой контейнер либо стеганопосылка верно отнесены к классу пустых контейнеров и стеганопосылок соответственно, это – корректная классификация.

Также, для нужд дальнейшего анализа, вводятся более узкие понятия: ложноположительная и ложноотрицательная классификация [45]:

– Если стеганопосылка классифицирована как пустой контейнер, это – ложноотрицательная классификация.
– Если пустой контейнер классифицирован как стеганопосылка, это – ложноположительная классификация.

Понятие корректной классификации не нуждается в подобном разбиении на более мелкие подклассы.

События корректной и некорректной классификации составляют полное множество элементарных событий. Таким образом, можно считать, что проводимые эксперименты удовлетворяют условию экспериментов Бернулли. Эффективность алгоритма в данном случае определяется как вероятность успешной классификации. Вероятность успешной классификации определяется следующим образом:

$$p = \lim_{N \to \infty} \frac{S}{N}, \text{ где}$$

S – количество событий успешной классификации,

N – количество попыток классификации.

Очевидно, что бесконечное количество экспериментов недостижимо, таким образом, ставится задача выделения тестовой выборки. Репрезентативность тестовой выборки определяется соответствием оценок вероятности успешной классификации на выборке и на всём множестве изображений.

На настоящий момент в литературе не приводится исчерпывающего обоснования минимального объёма выборки, необходимого для достижения репрезентативности. Кроме того, в зависимости от анализируемого алгоритма, размер может варьироваться. Единой методики выборки также не разработано. Задача выделения репрезентативной выборки сложна, и обобщения её на всё множество методов ВВС не существует [46]. В работе в дальнейшем применяются

выборки различного объёма, в зависимости от конкретной рассматриваемой задачи оценки. Минимальное количество изображений в выборке – 1000. Качество и взаимная независимость характеристик изображений в выборке также имеет значение. В дальнейшем в работе используются выборки изображений из коллекций: BOWS2 [47], BOSS [48], eTrim [49], Places [50] и UCID [51], неоднократно применявшихся в исследованиях в областях стеганографии и цифровых водяных знаков [34, 51 – 54].

Алгоритмы ВВС находят практическое применение в обнаружении сообщений, скрытых методами стеганографии, при их передаче в реальных сетях связи настоящими пользователями. Разумеется, проводить исследование с использованием каналов передачи данных с агентами-людьми не представляется возможным – количество попыток анализа слишком велико, чтобы проводить работу вручную. Для проведения экспериментов в рамках исследования решалась задача симуляции стеганоканала и основных агентов.

Три основных участника моделируемого эксперимента: отправитель, получатель и перехватчик. Отправитель формирует стеганопосылки и посылает их посредством канала передачи данных получателю. Получатель получает стеганопосылки и извлекает содержащиеся в них сообщения. Перехватчик играет роль пассивного наблюдателя, анализируя документы, идущие от отправителя к получателю, пытаясь определить, какие из документов – стеганопосылки, а какие – пустые контейнеры.

Для симуляции системы, описанной выше, в работе используется метод Монте-Карло, широко применяемый в задачах определения статистических вероятностей ошибок [46]. Суть метода состоит в генерации псевдослучайных выборок и сообщений, имитирующих неизвестную перехватчику логику отправителя.

Псевдослучайным образом из доступного множества изображений формируется тестовая выборка. Отправитель, получая изображение из выборки, в соответствие с принятым решением либо встраивает при помощи стеганосистемы получаемое сообщение, формируя стеганопосылку, либо оставляет контейнер пустым. Контейнер или стеганопосылка отправляются перехватчику, который проводит анализ, производя классификацию полученного изображения как стеганопосылки или пустого контейнера. Получатель присутствует в системе номинально, не принимая участия в эксперименте. Результат классификации перехватчиком отправляется в блок сравнения, где сравнивается с подлинной классификацией. Совпадение означает корректную классификацию. Несовпадение означает ошибку классификации. Описанный процесс повторяется для каждого сообщения из тестовой выборки.

Таким образом, 3 источника данных для системы формируются псевдослучайным образом: тестовая выборка изображений-контейнеров, сообщения и решение о внедрении сообщения в текущий контейнер. Алгоритм внедрения является атрибутом отправителя и постоянен в рамках эксперимента.

Итоговый результат эксперимента – количество успешных и неуспешных классификаций. Тем не менее, подобная оценка не сообщает всех необходимых сведений о реальной практической эффективности алгоритма ВВС, лежащего в основе системы [46, 42].

Причина этого в том, что ошибки бинарного классификатора различаются по своей сути. Классифицируемое изображение может принадлежать к одному из двух подмножеств выборки: стеганопосылкам или пустым контейнерам. Результат классификации алгоритма также относит анализируемое изображение к одному из двух этих классов. При некорректных классификациях говорят о понятиях ложной положительной и ложной отрицательной классификации.

Суммарная вероятность ложной положительной и ложной отрицательной классификации даст вероятность ошибочной классификации. Тем не менее, при оценке реальной эффективности алгоритма ВВС следует учитывать *перекос выборки,* способный внести ошибку в интерпретацию результата, особенно при предположении о его равенстве нулю без математического обоснования [45]. Для исключения ошибки, в задачах оценки эффективности алгоритма ВВС, вероятности ложной положительной и ложной отрицательной классификации следует рассматривать независимо друг от друга [46].

1.3.4. Способы представления результатов оценки эффективности метода выявления встроенных сообщений

Как отмечалось ранее, алгоритмы статистического ВВС в процессе работы вычисляют некоторую оценку потенциального искажения, внесённого в анализируемый документ в процессе встраивания стеганопосылки. Для того, чтобы от этой оценки перейти к непосредственно классификации, вводится понятия *порога* классификации. Порог представляет собой величину, соответствующую такому значению числовой оценки внесённого искажения, при превышении которого анализируемое изображение считается стеганопосылкой. Очевидно, что для классификатора вероятности ложной положительной и ложной отрицательной классификации связаны через величину порога. Так как выбор порога во многом зависит от ситуации, в которой используется классификатор, для оценки его собственной эффективности требуется ввести независимую метрику, позволяющую наглядно сравнивать классификаторы по эффективности между собой.

В качестве такой метрики взята кривая доверительных интервалов (ROC curve). Кривая ROC выражает зависимость между вероятностью корректной классификации и вероятностью ложной положительной классификации [43].

По оси абсцисс откладывается вероятность ложной положительной классификации, а по оси ординат – значение $1 - P_{FN}$, где P_{FN} - вероятность ложной отрицательной классификации. Координаты точки на кривой выражают вероятности ложной положительной и корректной классификации при определённом значении порога классификации. График, построенный в таких осях позволяет оценить эффективность классификации при варьировании порога классификации. Чем ближе к осям проходит кривая графика доверительных интервалов, тем выше эффективность классификатора. В дальнейшем кривые доверительных интервалов будут использованы для наглядной демонстрации при сравнении эффективности алгоритмов ВВС.

В качестве численной оценки эффективности при этом используется значение вероятности ложноположительной классификации при определённом значении вероятности корректной классификации. В работе используется значение корректной классификации 95%, что является оптимальной Байесовой стратегией.

Численные оценки приводятся в виде таблиц, демонстрирующих значения численных оценок эффективности в зависимости от характеристик анализируемого изображения и используемых методов ВВС и их улучшений.

1.3.5. Связь эффективности выявления встроенных сообщений с защищённостью информации

В системе защиты информации, включающей в себя подсистему пассивного противодействия каналам передачи данных на основе стеганографии, связь

понятия защищённости информации с эффективностью метода ВВС, используемого в подсистеме пассивного противодействия, осуществляется через понятие риска утечки информации по такому каналу.

Для данной угрозы, по общей формуле, риск оценивается следующим образом [55]:

$$R = (P_{\text{реал}} - P_{\text{прот}}) * U = \left(P_{\text{реал}} - \sum P_{\text{сп}}\right) * U, \text{где}$$

R – риск, $P_{\text{реал}}$ – вероятность реализации угрозы, $P_{\text{прот}}$ – вероятность противодействия угрозе с использованием средств защиты информации, U – ущерб от реализации угрозы, $P_{\text{сп}}$ – вероятность противодействия с использованием конкретного средства защиты информации.

Рассматриваемое средство пассивного противодействия каналу передачи информации является одним из средств защиты информации. Вероятность корректного обнаружения встроенного сообщения определяет эффективность средства пассивного противодействия каналу передачи информации. Поскольку эффективность ВВС напрямую влияет на вероятность обнаружения угрозы, эффективность является основополагающей характеристикой при выборе метода ВВС.

Таким образом, повышение эффективности ВВС ведёт к снижению вероятности реализации угрозы утечки информации по каналу связи на основе стеганографии, и, в конечном итоге, к снижению риска этой угрозы. Снижение риска, в свою очередь, повышает защищённость информации [56].

Исходя из этого, именно эффективность метода ВВС, напрямую влияя на защищённость информации, является основополагающей характеристикой метода ВВС.

1.4. Выводы

1. Повышение эффективности ВВС ведёт к снижению риска реализации угрозы утечки информации по каналу передачи данных на основе стеганографии, при использовании метода ВВС в системе пассивного противодействия такому каналу передачи. Таким образом, эффективность метода ВВС является его основополагающей характеристикой.
2. Эффективность метода статистического ВВС неподвижного цифрового изображения вне задачи бинарной классификации может быть определена как среднее по выборке отклонение спрогнозированной длины встроенного сообщения от реальной длины. Эффективность метода статистического ВВС в неподвижном цифровом изображении в задаче бинарной классификации может быть определена как процент некорректной классификации при заданном значении ожидаемого процента корректной классификации.
3. Зависимость объёма контейнера при встраивании в пространственную область цифрового изображения от разрешения изображения на фоне роста средних разрешений изображений, передаваемых в сети, делает актуальной задачу ВВС на малых значениях полезной нагрузки.
4. Наиболее перспективными в задачах ВВС в пространственной области неподвижных изображений на малых значениях полезной нагрузки являются методы Triples analysis, Sample pairs analysis, RS-analysis, Weighted stego, Difference Image Histogram.

ГЛАВА 2. ЭФФЕКТИВНОСТЬ МЕТОДОВ ВЫЯВЛЕНИЯ ВСТРОЕННЫХ СООБЩЕНИЙ И ЕЁ ЗАВИСИМОСТЬ ОТ ХАРАКТЕРИСТИК КОНТЕЙНЕРА

2.1. Разработка эксперимента по оценке эффективности метода выявления встроенных сообщений

2.1.1. Характеристики пространственной области изображения как стеганоконтейнера

Неподвижное цифровое изображение представляется как набор матриц пикселей, каждая из которых соответствует определённому цветовому слою изображения. Их количество варьируется в зависимости от выбранной модели представления изображения [57].

Одна из наиболее распространённых моделей представления цветного изображения – RGB, манипулирует тремя цветовыми слоями, соответствующими красному, зелёному и синему цвету соответственно. В дальнейшем в работе цветное изображение как стеганоконтейнер будет рассматриваться в контексте цветовой модели RGB. В случае монохромного изображения его пространственное распределение представляет собой одну матрицу. Пиксель – наименьший элемент цветового слоя пространственного распределения изображения, представляет собой целое неотрицательное число, обозначающее яркость участка изображения, соответствующего этому пикселю. В зависимости от точности представления уровня яркости (глубины цвета), максимальное значение пикселя может варьироваться [58].

Пусть I – изображение, анализируемое методом ВВС, p – цветовая плоскость, а x, y – координаты пикселя в данной цветовой плоскости. Тогда I_p – матрица пространственного распределения изображения в цветовой плоскости, $I_p(x, y)$ – операция взятия значения пикселя с данными координатами в данной цветовой плоскости. В случае, если изображение монохромное, единственная

цветовая плоскость подразумевается неявно, её обозначение опускается, операция взятия значения пикселя записывается следующим образом: $I(x, y)$. Цветовые плоскости модели RGB обозначаются I_r, I_g, I_b для красного, зелёного и синего соответственно.

Каждый пиксель изображения можно представить в двоичном виде, переведя его значение систему счисления. В этом случае значение пикселя представляется набором бит, где младший разряд соответствует наименьшему значащему биту числа. В общем случае определим взятие i-го бита двоичного представления пикселя как $I(x, y)_n$, для наименьшего значащего бита, вне зависимости от номера его разряда, операцию взятие определим как $I(x, y)_L$.

Встраивание в LSB эксплуатирует избыточность пространственного распределения изображения, следовательно, чем более длинным является побитовое представление пикселя, тем более объёмным становится контейнер для стеганографического встраивания [61].

В дальнейшем в работе будет рассмотрена модель, ограничивающее значение пикселя числом 255. Это соответствует 256 уникальным значениям уровня яркости и 8 двоичным разрядам числа. Встраивание в LSB, таким образом, оперирует восьмым, младшим разрядом. С точки зрения стеганографического контейнера, пространственное распределение изображения представляет области для встраивания сообщений, выраженных в разных форматах. Встраивание в LSB, при этом, ограничивает объём информации на 1 пиксель одним битом, следовательно, встраиваемое сообщение при этом представляется в виде битовой строки.

При этом, фактически, в качестве стеганоконтейнера выступает плоскость наименьших значащих бит матрицы распределения цветового слоя, с которым работает метод встраивания. Совокупный контейнер при этом – набор плоскостей наименьших значащих бит пространственных распределений всех цветовых слоёв изображения [63].

Распределение значений пикселей в изображении определяется тем, что содержится в изображении. Искусственные изображения, полученные путём рисунка, либо сгенерированные при помощи определённых алгоритмов (например, фрактальные изображения), вносят закономерности в распределение значений (к примеру, рисунки могут содержать обширные зоны, в которых значения пикселей не меняются).

В противовес искусственным изображениям вводится понятие естественного цифрового изображения. Под естественным цифровым изображением понимается изображение, полученное путём фотографии объектов окружающего мира, позже оцифрованные методом, сохраняющим характер распределения значений яркости на любом отдельно взятом участке изображения.

Естественные изображения характеризуются высокой степенью случайности распределения значений пикселей, равно как и значений в определённых битовых плоскостях [65, 82]. Это делает естественные изображения практически применимым контейнером для стеганографического встраивания в пространственную область [67, 68]. Далее в работе под изображениями-контейнерами понимаются естественные изображения.

В отличие от частотного распределения изображения, полученного дискретным преобразованием Фурье или вейвлет-преобразованием, пространственное распределение естественного изображения в высокой степени случайно [70]:

- Значения пикселей в той или иной области изображения полностью зависят от высокоуровневой семантики изображения, определяемой, в свою очередь, тем, что на изображении запечатлено. Следовательно, для набора независимых изображений распределение значений пикселей в пространственной области случайно.
- Распределение значений в плоскости наименьших значащих бит случайно в рамках одного отдельно взятого изображения в следствие сильного влияния

на значения в этой плоскости даже небольших колебаний яркости пикселей изображения.

Высокая случайность распределения значений в области наименьшего значащего бита изображения обуславливает высокую популярность и эффективность методов встраивания, основанных на модификации наименьшего значащего бита – детектирование факта встраивания битовой строки с условно случайным распределением в область случайного распределения величин является нетривиальной задачей [59, 60].

2.1.2. Метод встраивания в плоскость LSB изображения

Несмотря на случайный или почти случайных характер распределения значений как в плоскости наименьших значащих бит контейнера, так и во встраиваемом сообщении, характер распределения может отличаться. Статистические методы ВВС с высокой эффективностью определяют факт наличия встраивания на таких значениях полезной нагрузки.

На практике изменению подвергаются значительно меньшие объёмы бит изображения. Поскольку плоскость НЗБ сама по себе характеризуется большим объёмом как контейнера для стеганографического встраивания, обычно изменяются единицы процентов бит, что уже позволяет использовать метод для организации канала скрытной передачи информации. Со снижением значения полезной нагрузки «заметность» факта встраивания для статистических алгоритмов ВВС падает [71].

Пусть монохромное изображение I подвергается стеганографическому встраиванию в плоскость наименьших значащих бит. Встраивается сообщение M, представленное в виде битовой строки длиной L. В таком случае операция взятия бита на позиции i в изображении M обозначается $M(i)$. Единичная операция встраивания бита изображения при этом описывается так:

$$I(x,y)_L = M(i);$$

Метод встраивания в LSB определяет порядок и условия применения единичной операции к пикселям изображения-контейнера.

Простейший метод – *метод последовательного встраивания в LSB*, определяет начальные координаты x_s, y_s, с которых начинается встраивание. Далее, для каждого следующего пикселя изображения, слева направо и с переносом на следующий ряд пикселей по окончании текущего, производится единичная операция встраивания. Для последующего извлечения сообщения необходимо знать начальную позицию.

Метод рассеянного встраивания определяет функцию выбора пикселя контейнера для -го бита встраиваемого сообщения:

$$(x_i, y_i) = F(i)$$

Для последующего извлечения сообщения необходимо знать функцию.

Более сложные методы встраивания в LSB эксплуатируют значения других плоскостей изображения, например, плоскости, следующей за плоскостью LSB.

Метод последовательного встраивания, несмотря на простоту, широко используется на практике, позволяя добиться высокой скрытности, совмещая её с простотой выполнения и минимумом передаваемых дополнительных данных. В случае, если характер встроенного сообщения известен стороне приёмнику либо может быть детектирован автоматически (например, с помощью алгоритмов контроля целостности), метод может быть использован вообще без передаваемых дополнительных данных – извлечение сообщения может быть проведено путём перебора вариантов начальных координат на стороне-приёмнике [72]. В дальнейшем в работе в эксперименте, описанном в п. 1.3.2. будет использоваться метод последовательного LSB-встраивания.

Блок-схема алгоритма встраивания приведена на рисунке 2.

Рисунок 2 - Блок-схема алгоритма встраивания в LSB

Чтобы исключить влияние начальной позиции на эффективность исследуемого метода ВВС, начальная позиция выбирается псевдослучайным образом для каждого изображения.

Координаты следующего пикселя (x_n, y_n) берутся следующим образом:

– Если в текущем ряду есть пиксель справа от данного, его координата берётся следующей: $(x_n, y_n) = (x + 1, y)$;

- Если в текущем ряду нет пикселя справа от данного, но есть пиксель снизу от данного, берётся первый пикселя ряда непосредственно ниже данного: $(x_n, y_n) = (1, y + 1)$;
- Если в текущем ряду нет пикселя справа от данного и нет пикселя снизу от данного, берётся первый пиксель первого ряда: $(x_n, y_n) = (1, 1)$.

Таким образом, встраивание происходит по рядам слева направо сверху вниз, с переходом к первому элементу матрицы по достижению последнего, в случае необходимости. В этом случае, сообщение всегда может быть полностью встроено в изображение, если только его длина не превышает количества пикселей изображения.

2.1.3. Искажения, вносимые встраиванием в плоскость LSB изображения и модель их детектирования

Значение наименьшего значащего бита изображения, равно как и значение бита сообщения, может принимать одно из двух значений: 0 или 1.

Таким образом, единичная операция встраивания в LSB пикселя изображения оставляет его наименьший значащий бит неизменным, если бит сообщения равняется наименьшему значащему биту пикселя, либо меняет его на противоположный. Определим операцию смены бита на противоположный:

$$\overline{I(x,y)_L} = \begin{cases} 0, I(x,y)_L = 1, \\ 1, I(x,y)_L = 0 \end{cases};$$

Тогда единичную операцию встраивания описывает уравнение:

$$I'(x,y)_L = \begin{cases} I(x,y)_L, I(x,y)_L = M(i) \\ \overline{I(x,y)}_L, I(x,y)_L \neq M(i) \end{cases};$$

Изменение наименьшего значащего бита приводит к изменению значения пикселя в целом. Поскольку бит является разрядом двоичного представления

значения пикселя, а наименьший значащий бит соответствует разряду единиц, единичная операция встраивания изменяет значение пикселя сообщения следующим образом:

$$I'(x,y) = \begin{cases} I(x,y), I(x,y)_L = M(i) \\ I(x,y) \pm 1, I(x,y)_L \neq M(i) \end{cases};$$

Таким образом, встраивание в плоскость LSB изображения приводит к изменению ряда значений пикселей на 1 в большую или меньшую сторону, либо оставляет значение неизменным [73, 62].

Встраивание в плоскости старше наименьшей значащей приводит к изменениям в разрядах старше самого младшего. Поскольку в двоичной системе счисления значение по разрядам растёт со степенной зависимостью с основанием степени 2, с той же зависимостью растёт величина искажений, вносимых в контейнер. Высокие темпы роста величины искажений определяет практическую значимость встраивания именно в плоскость наименьших значащих бит.

Методы ВВС, направленные на детектирование факта встраивания в плоскость наименьших значащих бит изображения основывают свою работу на определении подобных аномальных для изображения изменений значений пикселей на 1, либо непосредственно (Weighted Stego image, RS-analysis), либо через анализ статистик высшего порядка, чувствительных к подобным изменениям (Sample Pairs analysis, Difference Image Histogram).

ВВС сводится к подсчёту аномальных участков, включающих пиксели с изменёнными значениями. Поскольку встраивание может не изменить значение пикселя в случае совпадения его наименьшего значащего бита с битом встраиваемого сообщения, простой подсчёт количества аномальных участков вносит ошибку в прогноз длины встроенного сообщения [64].

Исходя из предположения о случайном распределении значений в плоскости наименьшего значащего бита и в сообщении, предполагается, что

встраивание изменяет половину подвергшихся обработке битов. Таким образом, итоговое значение подсчёта умножается на 2.

Рисунок 3 иллюстрирует блок-схему общего вида метода выявления факта встраивания в область наименьшего значащего бита изображения. Общий вид не обязательно соответствует конкретному алгоритму реализации любого взятого метода ВВС, отражая общее направление работы метода.

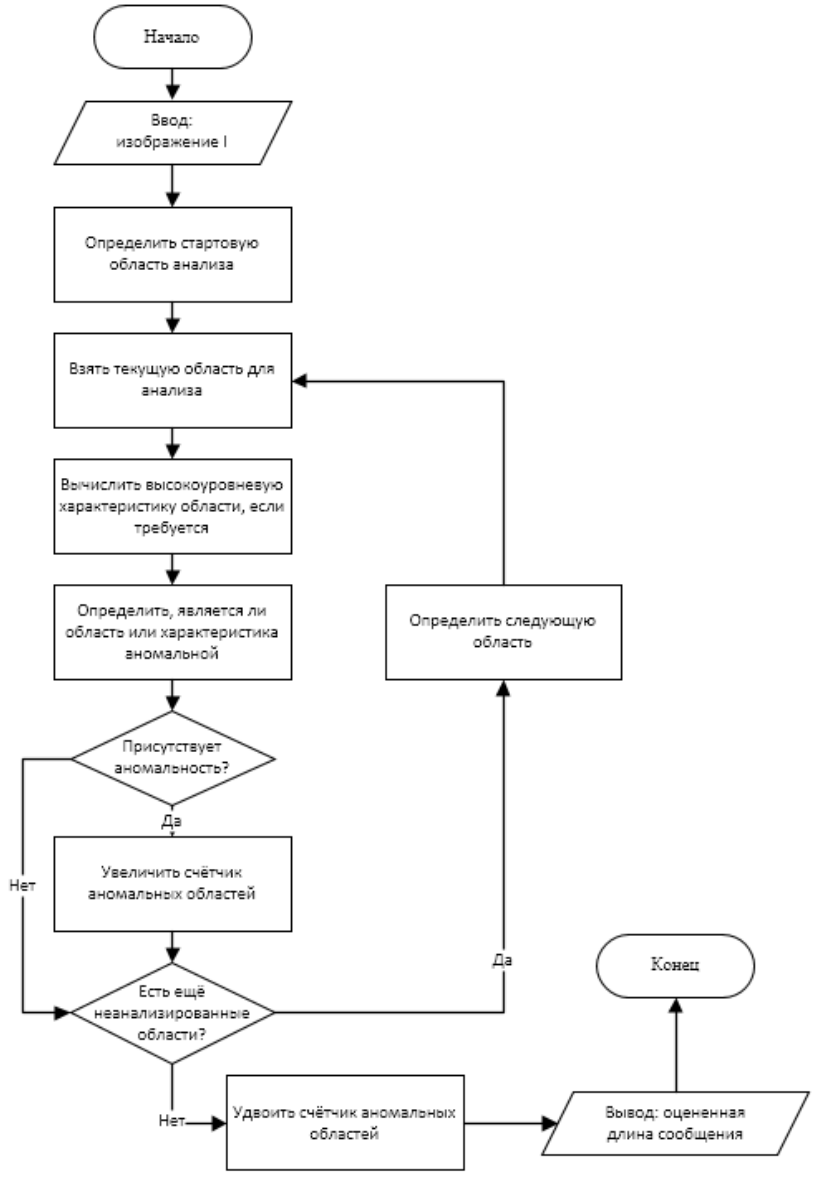

Рисунок 3 - Блок-схема общего алгоритма выявления факта встраивания в плоскость LSB

2.2. Анализ эффективности современных методов выявления встроенных сообщений в плоскости LSB изображений

2.2.1. Понятие отношение нагрузка-контейнер для встраивания в плоскость LSB изображения

Эффективность методов ВВС зависит от масштаба модификаций, вносимых в изображение-контейнер. Поэтому возможности современных методов ВВС следует рассматривать в контексте конкретных значений объёмов встраивания. Абсолютные значения, выраженные длиной сообщения в битах, неприменимы: широкий разброс разрешений изображений-контейнеров приводит к тому, что сообщение одной и той же длины может оказывать разное влияние на большие и маленькие изображения.

Вводится понятие отношение нагрузка-контейнер (ОНК), позволяющее оценить нормированную по размеру контейнера степень масштаба внесения искажений, вызванных стеганографическим встраиванием.

Для встраивания в LSB значение ОНК имеет смысл определить, как отношение максимального количества элементов, подходящих для изменения при встраивании, к фактическому количеству изменённых элементов. При этом:

- при встраивании изменяется один бит одного пикселя изображения, при этом разряд бита фиксирован;
- количество вносимых изменений равно длине встраиваемого сообщения в битах;
- общее количество элементов, подходящих для изменения соответствует общему количеству пикселей всех цветовых слоёв изображения-контейнера;
- если изображение цветное, встраивание производится во все цветовые слои изображения-контейнера;

– цветовые слои изображения обладают одинаковыми размерами, совпадающими с размерами изображения в пикселях.

С учётом приведённых утверждений, для цветного изображения значение ОНК определяется следующим образом:

$$P = \frac{L}{3AB}, \text{где}$$

L – длина встраиваемого сообщения,

A, B – размеры изображения-контейнера по каждому измерению.

Для монохромного изображения формула схожая, с поправкой на количество цветовых слоёв:

$$P = \frac{L}{AB}.$$

Формулы не содержат поправки на то, что в среднем половина пикселей не изменяются при встраивании. Это допущение оставлено специально, поскольку в контексте данной работы величина ОНК рассчитывается специально для соотнесения с эффективностью методов ВВС, которые уже закладывают множитель 2 в формулы оценки длины встроенного сообщения, как показано в п. 2.1.3. Для удобства, в дальнейшем в работе значение ОНК выражается в процентах. По определению, длина сообщения не может быть больше, чем количество пикселей в изображении, поэтому значение ОНК лежит в интервале от 0 до 100%.

2.2.2. Условия проведения эксперимента

Для оценки возможностей современных методов ВВС, отобранных в п. 1.2.3, проводится эксперимент в соответствии с алгоритмом, описанным в п. 1.3.2.

Условия проведения эксперимента следующие:

- тестовая выборка представляет собой набор изображений, случайным образом отобранных из используемых коллекций изображений;
- количество изображений в выборке: 39000;
- минимальное разрешение изображений: 392x550 пикселей;
- максимальное разрешение изображений: 5100x4025 пикселей;
- количество цветовых плоскостей: 1 (изображения монохромные)
- встраивание в LSB последовательное, по алгоритму, описанному в п. 2.1.2.
- встраиванию подвергается половина изображений, входящих в выборку;
- результаты усреднены для выборки.

2.2.3. Результаты оценки эффективности современных методов анализа в плоскости LSB

Оценена эффективность алгоритмов на значениях полезной нагрузки 1, 3 и 5%. Кривые доверительных интервалов приведены на рисунке 4. Кривые построены в соответствие с методом оценки Монте-Карло, описанным в п. 1.3.3.

Расшифровка обозначений на рисунке 4: DIH – метод Difference Image Histogram, WSI – метод Weighted Stego Image, SP – метод Sample Pairs analysis, RS – метод RS-analysis, TR – метод Triples analysis.

Из графиков на рисунке 4, а также из численных оценок в таблице А1 приложения А видно, что эффективность современных методов статистического ВВС на малых значениях отношения нагрузка-контейнер (ОНК) характеризуется большой долей ложноположительной классификации при доле корректного срабатывания 95%. Сделан вывод о необходимости усовершенствования существующих методов ВВС. Также из графиков видно, что вне зависимости от значения ОНК, наибольшую эффективность демонстрирует метод ВВС Weighted Stego (WS). В дальнейшем в работе рассматриваются методы увеличения эффективности ВВС методом WS.

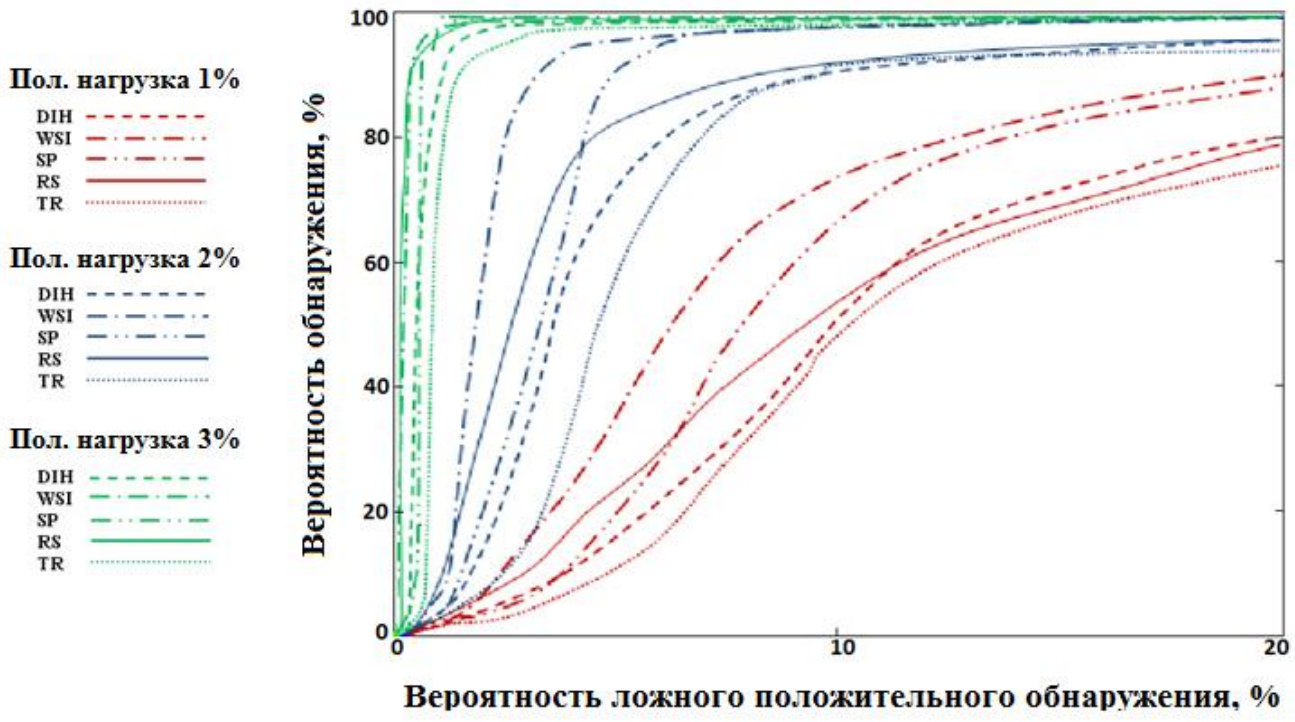

Рисунок 4 - Кривые доверительных интервалов для современных методов ВВС

Низкая эффективность современных методов ВВС в плоскости наименьшего значащего бита подтверждает практическую применимость методов встраивания в LSB и обосновывает актуальность задачи повышения эффективности методов ВВС в изображениях в области наименьшего значащего бита [9].

Следует отметить, что в эксперименте использовался самый простой из описанных в открытых источниках метод встраивания в плоскость наименьшего значащего бита. Тем не менее, его использование позволяет добиться высокой степени скрытности встроенного сообщения: на значении ОНК 1% результат работы всех представленных методов *мало отличим от угадывания*.

Тем не менее, к примеру, при разрешении изображения, соответствующем распространённому разрешению современных дисплеев 1920x1080 пикселей, ОНК 1% соответствует 62,2 Кб встраиваемой информации (для цветного изображения в формате RGB).

Также из графиков видно, что среди исследуемых методов, вне зависимости от объёма встраивания, наилучшую эффективность демонстрирует метод Weighted Stego Image (далее – WS). Далее в работе рассматриваются методы увеличения эффективности ВВС методом WS [11].

2.3. Анализ зависимости эффективности выявления встроенных сообщений в плоскости LSB от характеристик изображения

2.3.1. Значимые характеристики изображения

Для определения направлений работы по увеличению эффективности ВВС требуется определить, каким образом характеристики анализируемых изображений влияют на эффективность ВВС.

Методы ВВС в плоскости наименьших значащих бит базируют свою работу на определении аномальных изменений значений пикселей. При этом, аномальность значения оценивается по отношению к другим значениям окружающих пикселей, либо к значениям характеристик высшего порядка, вычисленных на других областях анализируемого изображения.

Пространственная область изображения не вносит глобальных закономерностей в распределение пикселей. В локальных, ограниченных небольшим пространством областях, закономерности могут проявляться. Из этого очевидно, что значения, собранные с соседних по отношению к анализируемому пикселей, имеют большее значение для оценки значения текущего пикселя [74, 75].

Следовательно, на эффективность анализа может влиять характер распределения значений пикселей по отношению друг к другу в небольших, локальных зонах изображения, соответствующих одному объекту или области окружающего мира, запечатлённому на естественном изображении. Очевидно,

что с ростом разрешения изображения размер таких зон в пикселях увеличивается.

Также, плоскость наименьшего значащего бита в большой соответствует шуму, присутствующему на изображении, и в большой степени подвержена изменению при внесении шума. Исследование зависимости эффективности ВВС от предварительной фильтрации изображения-контейнера проводилось в работах [66, 76, 77]. Сделан вывод о незначительном влиянии предварительной фильтрации на эффективность последующего ВВС. Также сделан вывод о том, что метод ВВС WS характеризуется наивысшей из рассматриваемых устойчивостью к предварительной фильтрации в контексте эффективности ВВС.

Исходя из приведённого выше, следующие характеристики изображения подлежат исследованию на предмет влияния на эффективность ВВС, разрешение изображения, монохромность изображения, присутствие обширных фоновых зон в изображении.

2.3.2. Анализ зависимости эффективности выявления встроенных сообщений от разрешения изображения

Зависимость эффективности статистического ВВС в неподвижном цифровом изображении от его разрешения отмечалась во многих работах. Наблюдается тенденция роста эффективности ВВС с ростом разрешения изображения при неизменном значении полезной нагрузки.

Актуальность задачи повышения эффективности ВВС на малых значениях полезной нагрузки обусловлена, прежде всего, постоянным ростом разрешения изображений, передаваемых в сети интернет. Рост эффективности ВВС при росте разрешения потенциально способен нивелировать проблему низкой эффективности анализа в случае, когда малое значение полезной нагрузки обусловлено именно большим объёмом стеганоконтейнера, а не малым объёмом

встраиваемого сообщения. Тем не менее, это предположение нуждается в экспериментальном подтверждении или опровержении.

Прежде всего, исследовано, подвержен ли метод BBC WS эффекту повышения эффективности при росте размера изображения-контейнера.

Эксперимент проведён по следующему сценарию:

- опытным путём установлен фиксированный объём f встраиваемого сообщения, позволяющих получить показательные результаты на большом интервале разрешений изображения-контейнера;
- отобран тестовый набор изображений контейнеров разных разрешений;
- для изображений из набора оценена эффективность BBC для метода WS при встраивании сообщения объёмом f (встраивание проводилось методом, описанным в п. 2.1.3);
- построен график зависимости эффективности BBC от разрешения изображения для каждого алгоритма (для наглядности графики построены в одних осях).

Количественные характеристики проводимого эксперимента следующие:

- фиксированный объём встраиваемого сообщения $f = 6400$ бит;
- разрешение изображений в диапазоне от 200x200 до 2000x2000 пикселей (изображения квадратные), выборки формируются с шагом 100 пикселей;
- размер каждой выборки – 1000 изображений;
- цветовая модель изображений – RGB, встраивание производится в случайно выбранный цветовой слой, каждая стеганопосылка полностью несёт сообщение в одном из своих цветовых слоёв;
- встраивание LSB, последовательное, стартовый пиксель выбирается псевдослучайным образом с равномерным характером распределения стартовой позиции в рамках каждой выборки фиксированного разрешения;

– для построения зависимости эффективности ВВС от разрешения взяты значения вероятности ПЛС при вероятности верного обнаружения 95%;
– для случая, когда вероятность верного обнаружения 95% недостижима, вероятность ПЛС принята равной 100%.

Для демонстрации экспериментального подтверждения факта роста эффективности ВВС при увеличивающимся разрешении в одних осях были построены две кривые:

– зависимость эффективности от разрешения при фиксированном объёме сообщения и растущем разрешении изображения;
– зависимость эффективности от полезной нагрузки при постоянном разрешении и с условием, что полезная нагрузка уменьшается в процентном отношении так же, как и при росте разрешения для первой кривой.

Рисунок 5 иллюстрирует полученный результат.

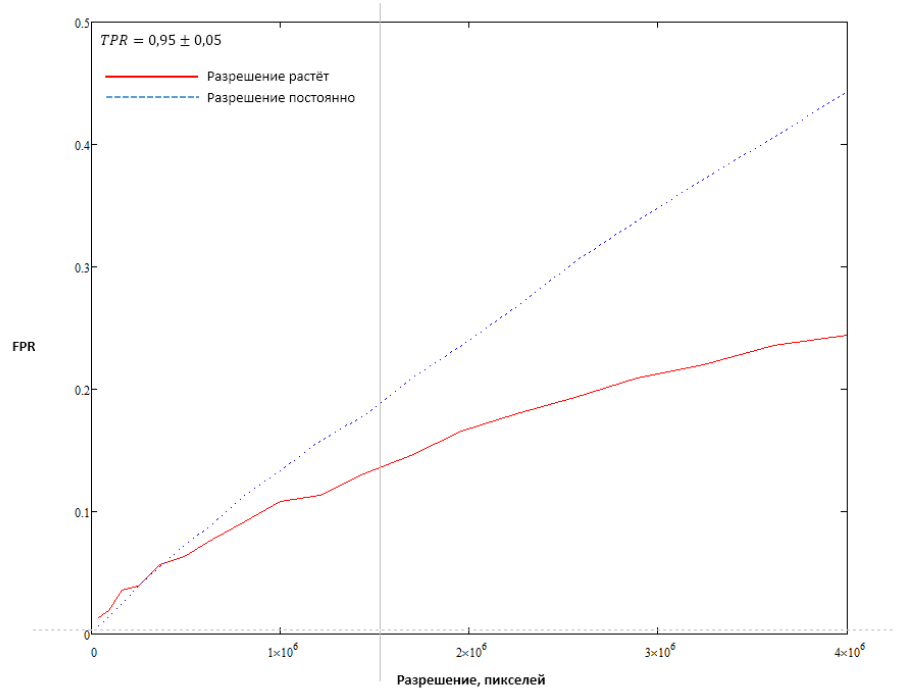

Рисунок 5 - Эффективность ВВС при росте разрешения и при постоянном разрешении изображения-контейнера

Из графиков на рисунке 5 видно, что рост разрешения действительно приводит к увеличению эффективности ВВС методом WS – при росте разрешений доля ложноположительной классификации растёт медленнее, чем при постоянном разрешении.

Общий рост доли ложноположительных срабатываний в данном случае обусловлен снижением значения полезной нагрузки в обоих случаях.

Следующий эксперимент ставит целью показать, что, несмотря на рост эффективности ВВС при росте разрешения анализируемого изображения, общее падение эффективности преобладает именно за счёт малых значений полезной нагрузки при встраивании сообщения постоянной длины в стеганоконтейнер с увеличивающимся разрешением.

Предположение следует подтвердить для всех исследуемых алгоритмов для подтверждения или опровержения факта того, что ни один из рассматриваемых алгоритмов не проявляет преимуществ в эффективности именно при анализе стеганоконтейнеров больших разрешений.

Эксперимент проведён по следующему сценарию:

- Исходная тестовая выборка изображений разделена на подвыборки по разрешениям, от малых к большим. Всего сформировано 10 подвыборок, разрешение наименьшей: 200x200 пикселей, наибольшей: 2000x2000 пикселей. Таким образом, размер стеганоконтейнера варьируется от $4 * 10^4$ до $4 * 10^6$ бит.
- Для каждой подвыборки отдельно оценена эффективность ВВС по методу, описанному в Главе I. Получены численные оценки эффективности.
- Встраивание производилось по методу, описанному в п. 2.1.3.
- Длина встраиваемого сообщения постоянна и составляет 6400 бит.

Графики на рисунке 6 иллюстрируют полученный результат. Видно, что доля ложноположительных классификаций растёт при росте разрешения изображений,

это выполняется для всех исследуемых методов ВВС. Таким образом, подтверждено, что рост эффективности при росте разрешения изображения неспособен компенсировать падение эффективности от уменьшения значения полезной нагрузки при постоянной длине встраиваемого сообщения.

Более того, видно, что на всех разрешениях алгоритм WS показывает наименьшую долю ложноположительных классификаций.

Таким образом, подтверждена актуальность задачи увеличения эффективности ВВС методом WS на малых значениях полезной нагрузки.

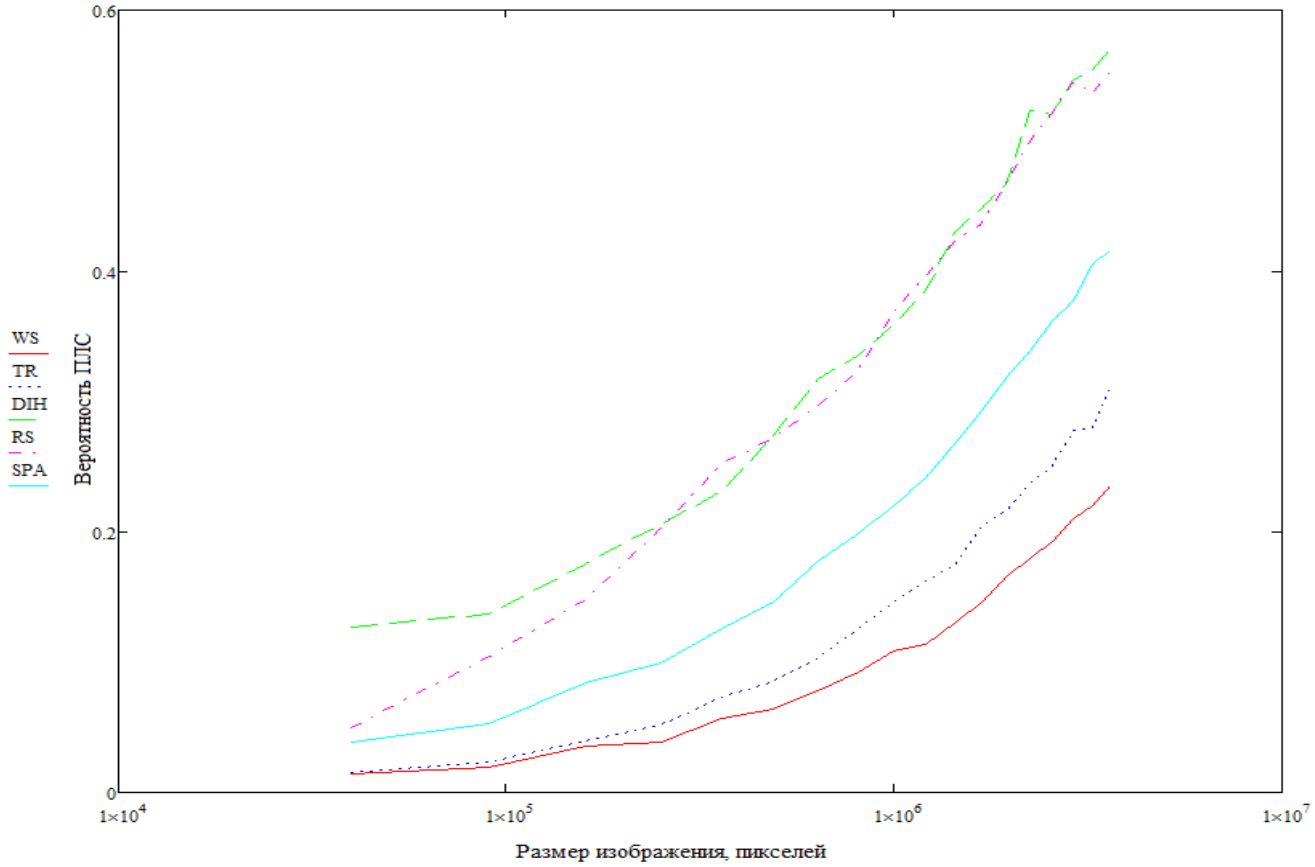

Рисунок 62 - Зависимость доли ЛПК от значения полезной нагрузки

2.3.3. Влияние доли однородного фона на эффективность выявления встроенного сообщения

Исходя из модели детектирования изменений, описанной в п. 2.1.3, разница между значениями соседних пикселей анализируемого изображения имеет значение при выявлении изменений, вносимых в плоскость наименьших значащих бит изображения. Поскольку методы ВВС в этой области ставят цель детектировать аномальные изменения пикселей на 1 по сравнению с ожидаемым, зоны, в которых разница в 1 по сравнению с соседними встречается часто, могут быть проблемой при ВВС. Метод может столкнуться с проблемой различения изменения значения пикселя на 1 как результата стеганографического встраивания и как естественного эффекта в данной зоне изображения [69].

В зонах, где средняя разница между соседними пикселями велика, изменение на 1 по сравнению с ожидаемым значением является аномалией, легко выделяемой методом ВВС. Тем не менее, существует общая семантическая зона изображения, повторяющаяся на большом наборе естественных изображений – однородные фоновые зоны.

Под однородными фоновыми зонами подразумеваются такие зоны изображения, в которых значения соседних пикселей изменяются не более чем на определённую величину в среднем для зоны и не более чем на определённую величину от среднего значения пикселей для той же зоны. Математически однородная фоновая зона определяется следующим образом: пусть I – множество пикселей изображения. Тогда однородная фоновая зона B – это совокупность таких зон изображения, для каждой из которых выполняется следующее:

$$\forall p \in B: |p - p_N| < T_N, |p - p_A| < T_A, \text{где}$$

p - значение яркости пикселя,

p_N – значение яркости соседнего пикселя, наиболее отличающегося по значению среди всех соседних пикселей данного,

p_A – среднее значение всех пикселей, принадлежащих множеству B

T_N, T_A – пороги, вводимые индивидуально для каждого изображения

В естественных изображениях однородные фоновые зоны встречаются часто: как правило, они соответствуют областям неба, моря либо искусственно создаваемого однородного фона, например, при студийной съёмке.

Характер искажений, вносимых встраиванием в LSB, определяет важность наличия и доли фоновых зон в анализируемом изображении. В фоновых зонах, где разница между значениями соседних пикселей невелика, естественный характер распределения значений пикселей мало отличим от следов встраивания в LSB. Таким образом, можно ожидать падения эффективности ВВС в изображениях с большой долей однородного фона.

Для подтверждения предположения проведён эксперимент по следующему сценарию:

- Исходная выборка изображений разделена на подвыборки HB и LB так, что в подвыборке HB доля однородного фона составляет не менее 40%, а в подвыборке LB доля однородного фона составляет менее 5%.
- Эффективность ВВС в каждой подвыборке отдельно оценена по методу, приведённому в Главе I.
- Встраивание производилось по методу, приведённому в п. 2.1.3, при этом при встраивании достигалось равномерное по подвыборке распределение областей фактического изменения наименьших значащих бит между фоновыми и нефоновыми зонами.

График на рисунке 7 показывает результат эксперимента для значения полезной нагрузки 1% для метода WS. Видно, что на подвыборке HB наблюдается значительное падение эффективности ВВС. На значениях полезной нагрузки 3, 5 и 10% наблюдается такой же эффект: эффективность на подвыборке HB стабильно ниже, чем на подвыборке LB.

Поскольку однородный фон – общая и часто встречающаяся семантическая зона изображения, а её выделение можно провести автоматически для изображения, падение эффективности ВВС методом WS при анализе в фоновых

зонах позволяет организовать атаку на наиболее эффективный метод из рассмотренных, встраивая сообщения в фоновые зоны с подходящими свойствами, тем самым снижая эффективность анализа подобных контейнером методом WS. Актуальна задача исследования причины падения эффективности при анализе в фоновых зонах и усовершенствования алгоритма WS для нивелирования эффекта падения эффективности при анализе в фоновой зоне [74, 78].

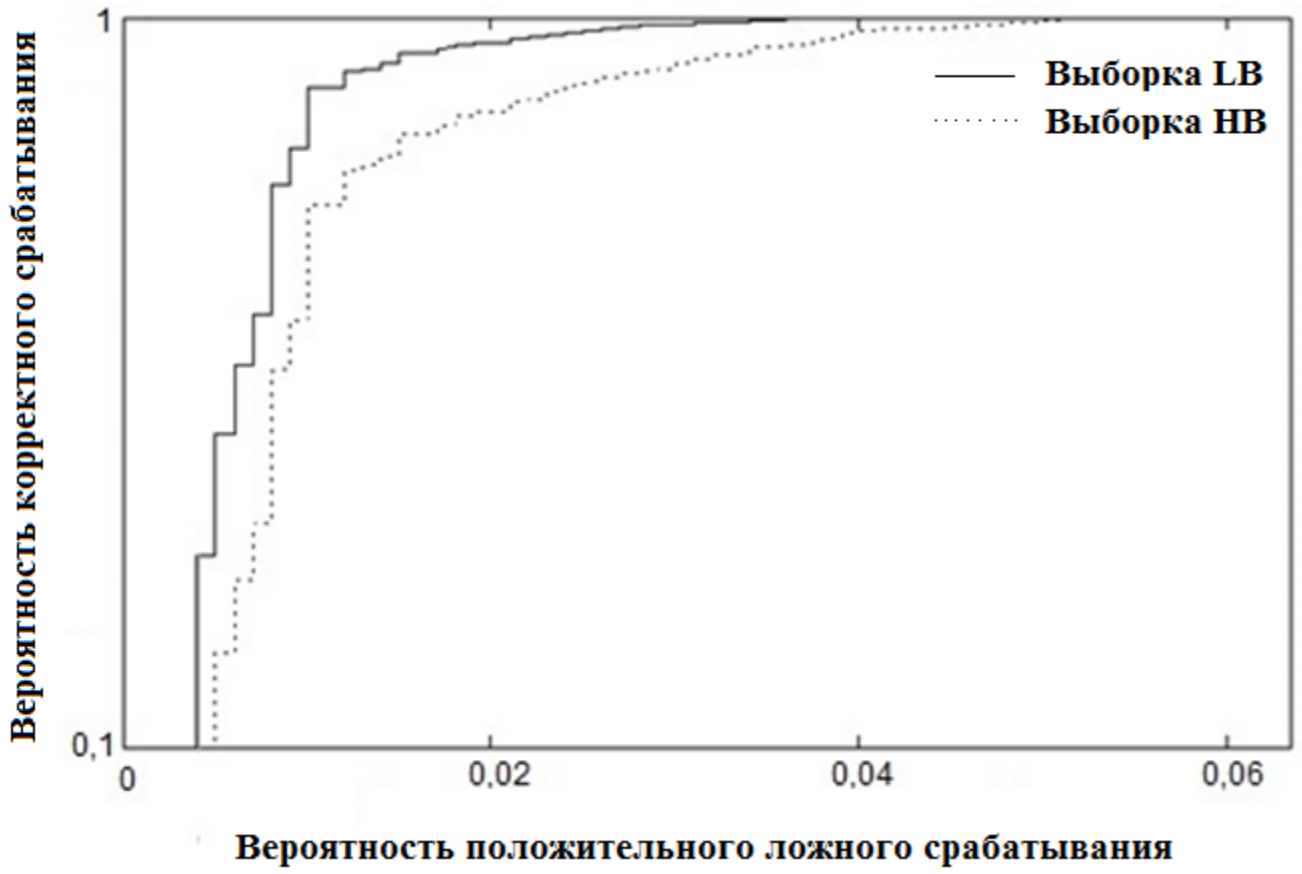

Рисунок 7 - Падение эффективности BBC методом WS в условиях большой доли однородного фона в изображении

2.4.Выводы

1. Плоскость наименьших значащих бит изображения – широко применимый на практике контейнер для стеганографического встраивания.

Эффективность использования плоскости наименьших значащих бит изображения для стеганографии достигается высокой случайностью распределения значений в этой плоскости, малой заметностью факта встраивания для наблюдателя и методов ВВС, большим объёмом зоны как стеганоконтейнера.

2. Современные методы ВВС демонстрируют низкую эффективность в задаче определения факта встраивания в плоскость наименьших значащих бит изображения.
3. Наибольшую эффективность среди исследуемых методов показывает метод Weighted Stego Image. Метод демонстрирует наибольшую эффективность вне зависимости от значения полезной нагрузки, цветовой плоскости изображения, предварительной фильтрации изображения и разрешения изображения.
4. Задача повышения эффективности ВВС в неподвижных цифровых изображениях на малых значениях полезной нагрузки актуальна в связи с постоянным ростом разрешений изображений, передаваемых в сети интернет, а также в связи с тем фактом, что рост эффективности ВВС за счёт повышения разрешения изображения не в состоянии нивелировать падение эффективности за счёт снижения значения полезной нагрузки при постоянной длине встраиваемого сообщения.
5. Предварительная фильтрация изображения-контейнера и выбор определённого цветового слоя для встраивания оказывает незначительное влияние на эффективность последующего ВВС в контейнере.
6. Эффективность ВВС методом Weighted Stego падает при анализе изображений с большой долей однородного фона.
7. Распространённость изображений с однородным фоном и возможность выделения фона из изображения автоматически создаёт возможность для атаки на метод WS с целью понижения эффективности анализа. Это обуславливает актуальность задачи разработки методов увеличения

эффективности ВВС методом WS, в том числе, в фоновых зонах изображений.

ГЛАВА 3. МОДЕЛЬ И АЛГОРИТМЫ ВЫЯВЛЕНИЯ ВСТРОЕННЫХ СООБЩЕНИЙ В ФОНОВЫХ ЗОНАХ

3.1. Модель выявления встроенных сообщений в фоновых зонах

3.1.1. Анализ причин падения эффективности выявления встроенных сообщений в фоновых зонах

Метод статистического BBC WS и его модификации, рассматриваемые в статье, в качестве результата работы имеют оценку длины встроенного в LSB пикселей сообщения. При построении систем детектирования факта стеганографического встраивания к результату работы алгоритма применяется бинарная классификация. Из превышения оцененной длинной сообщения определённого порога делается вывод о наличии либо отсутствии факта стеганографического встраивания в сообщение.

Идеальный классификатор всегда определяет оригинальные изображения как чистые (True Negative, TN) и модифицированные – как стеганограммы (True Positive, TP). В реальных условиях классификатор может допускать ошибки, классифицируя оригинальные изображения как стеганограммы (False Positive, FP) и модифицированные изображения как чистые (False Negative, FN). Распределение доли классификации по классам зависит от корректности оценки длины сообщения алгоритмом. В качестве оценки эффективности BBC используется процент некорректной классификации при заданном проценте корректной классификации.

При анализе точности функции предсказания пикселей в фоновых зонах в качестве метрики точности используется отклонение предсказанного значения от действительного, усреднённое по фоновой зоне изображения, и, далее, по всем изображениям выборки.

Метод WS в процессе ВВС оценивает длину встроенного в стеганоконтейнер сообщения [72]. Дальнейший вывод о наличии или отсутствии стеганографического встраивания делается исходя из заданных пороговых значений и того факта, что оцененная длина превышает (или не превышает) выбранный порог длины. При превышении пороговой длины делается вывод о наличии встраивания. Если оцененная длина меньше порогового значения, делается вывод, что встраивание отсутствует. Таким образом, итоговая эффективность классификации напрямую зависит от точности оценки длины встроенного сообщения в битах.

Как показано в Главе I, метод статистического ВВС в неподвижном цифровом изображении, вне зависимости от деталей его реализации, может быть рассмотрен в двух контекстах:

1. как инструмент оценки длины встроенного сообщения;
2. как бинарный классификатор, определяющий факт встраивания.

Таким образом, наблюдаемый эффект падения эффективности при ВВС в фоновых зонах, показанный в Главе II, может проявляться, в зависимости от контекста рассмотрения, одним из следующих образов.

1. При рассмотрении системы как бинарного классификатора – за счёт завышения доли положительных или отрицательных ложных срабатываний алгоритма при работе на изображениях с большой долей однородного фона;
2. При рассмотрении метода, оценивающего длину встроенного сообщения – за счёт системного завышения или занижения оцененной длины сообщения на выборке на выборке изображений с большой долей однородного фона по сравнению с таковой на выборке изображений с малой долей однородного фона.

Оценка обоих аспектов итогового падения эффективности важна для понимания причины наблюдаемого эффекта.

Падение эффективности ВВС как бинарной классификации состоит в завышении доли ложных классификаций (ложноположительной и ложноотрицательной как по-отдельности, так и одновременно) по сравнению с образцовыми значениями.

3.1.2. Условия анализа

Для оценки доли ложных классификаций при анализе изображений с различной долей однородного фона требуется провести два эксперимента на различных выборках изображений. Формируются три выборки.

- Образцовая выборка содержит набор изображений из тестовой коллекции без разделения на изображения с большой и малой долей однородного фона. При этом гарантируется, что выборка не имеет значительного перекоса в сторону изображений с большой или, напротив, малой долей однородного фона. Способ оценки величины перекоса приведён ниже.
- Выборка HB содержит только изображения, в которых доля однородного фона составляет более 40%.
- Выборка LB содержит только изображения, в которых доля однородного фона составляет менее 5%.

Доля однородного фона вычисляется по следующей формуле:

$$B = \frac{S_B}{S_T} * 100\%, \text{где}$$

S_B – количество пикселей, классифицированных как фон методом, описанным в Главе II, S_T – общее количество пикселей анализируемого изображения.

Для цветных изображений количество пикселей, как фоновых, так и общее, суммируется для всех цветовых слоёв изображения.

Перекос образцовой выборки оценивается как два следующих значения:

– Отклонение доли изображений, входящих в выборку, которые не могут быть отнесены ни к выборке HB, ни к выборке LB (иными словами – доля изображений, для которых доля однородного фона лежит в интервале от 5 до 40%) – неклассифицированных изображений от эталонной доли:

$$B_I = \left| B_{IE} - \frac{N_N}{N_{HB} + N_{LB} + N_N} \right| * 100\%, \text{где}$$

B_{IE} – эталонное значение доли неклассифицированных изображений, N_N – количество неклассифицированных изображений, N_{HB} – количество изображений, которые можно отнести к выборке HB, N_{LB} – количество изображений, которые можно отнести к выборке LB.

– Отклонение среднего значения доли однородного фона по выборке от эталонного значения:

$$B_B = \left| B_{BE} - \frac{\sum_1^N R_{Bn}}{N} \right|, \text{где}$$

B_{BE} – эталонное значение средней доли фона, N – число изображений в эталонной выборке, R_{Bn} – доля фона в n-ном изображении выборки.

Таким образом, образцовая выборка должна удовлетворять условиям:

$$B_I \leq B_{Imax}; B_B \leq B_{Bmax}, \text{где}$$

B_{Imax}, B_{Bmax} – максимальные допустимые значения перекоса выборки.

Эталонные значения взяты из соображений равномерного распределения характеристик, значимых в контексте эффективности ВВС на выборке.

Эталонное значение доли неклассифицированных изображений взято так, чтобы доли неклассифицированных изображений, изображений, удовлетворяющих выборке HB и изображений, удовлетворяющих выборке LB были равны. Эталонное значение средней доли фона оценено как среднее арифметическое пороговых значений отнесения изображений к выборкам HB и LB. Таким образом, численно эталонные значения взяты следующие: $B_{IE} = 0.33, B_{BE} = 22{,}5\%$.

На практике создание эталонной выборки произведено по следующему алгоритму:

1. Определены объёмы выборок HB и LB – 1000 изображений для каждой.
2. Выборки HB и LB набраны из тестового множества изображений.
3. Эталонная выборка образована как объединение выборок HB и LB.
4. В эталонную выборку добавляются неклассифицированные изображения из тестового множества, пока перекос по доле неклассифицированных изображений не станет меньше максимального допустимого значения.
5. В эталонную выборку добавляются последовательно по одному неклассифицированные изображения из тестового множества, пока перекос по средней доле фона не станет меньше максимального допустимого значения. Выбор изображения для добавления делается исходя из его доли однородного фона и потребности смещения средней доли однородного фона к эталонному значению. Если при добавлении перекос по доле неклассифицированных изображений выходит за рамки допустимого, в выборку добавляются изображения, удовлетворяющие критериям HB и LB и минимально меняющие среднее значение доли фона.

Максимальные допустимые значения отклонения выборки определены практически исходя из характеристик доступной тестовой выборки и составляют: $B_{Imax} = 3\%, B_{Bmax} = 1\%$.

Фактические значения перекоса эталонной выборки составили: $\mathrm{B}_I = 2{,}71\%, B_B = 0{,}98\%$.

3.1.3. Анализ непосредственных причин падения эффективности выявления встроенных сообщений

Для оценки завышения доли некорректных классификаций на трёх полученных выборках проводится следующий эксперимент:

1. Эффективность анализа методом WS оценивается для трёх выборок по алгоритму, приведённому в Главе I, при этом стеганографическое встраивание применяется *ко всем изображениям* выборок с одинаковым значением полезной нагрузки.
2. Эффективность анализа методом WS оценивается для трёх выборок по алгоритму, приведённому в Главе I, при этом стеганографическое встраивание *не применяется*.

В части 1 эксперимента все некорректные классификации будут относиться к ложноотрицательным классификациям. В части 2 эксперимента все некорректные классификации будут относиться к ложноположительным классификациям. Таким образом, можно получить точные непосредственные причины падения эффективности ВВС как бинарной классификации. Для исключения влияния значения полезной нагрузки на результат эксперимента, он проводится последовательно со значениями полезной нагрузки от 1 до 10%, с шагом 1%, после чего значения численных оценок эффективности усредняются.

Численные оценки эффективности ВВС приведены в таблице 2.

Из данных таблицы 2 видно, что доля ложноположительных классификаций значительно растёт при анализе изображений с большой долей однородного фона, доля же ложноотрицательных классификаций растёт незначительно.

Диаграмма на рисунке 8 иллюстрирует отношение роста доли ложноположительных срабатываний и роста доли ложноотрицательных срабатываний для двух выборок: HB и LB.

Отклонение спрогнозированной длины встроенного сообщения от реальной длины оценено исходя из части 1 эксперимента, поскольку метод WS не предусматривает отрицательного значения спрогнозированной длины, и в части 2 все оценки будут вести к завышению.

Таблица 2

Доли некорректных классификаций в зависимости от доли фона

Эксперимент / Выборка	**HB**	**Эталон**	**LB**
Часть 1 (оценена доля ложноотрицательных классификаций)	10,0%	9,8%	9.5%
Часть 2 (оценена доля ложноположительных классификаций)	22,4%	16,9	13,1%

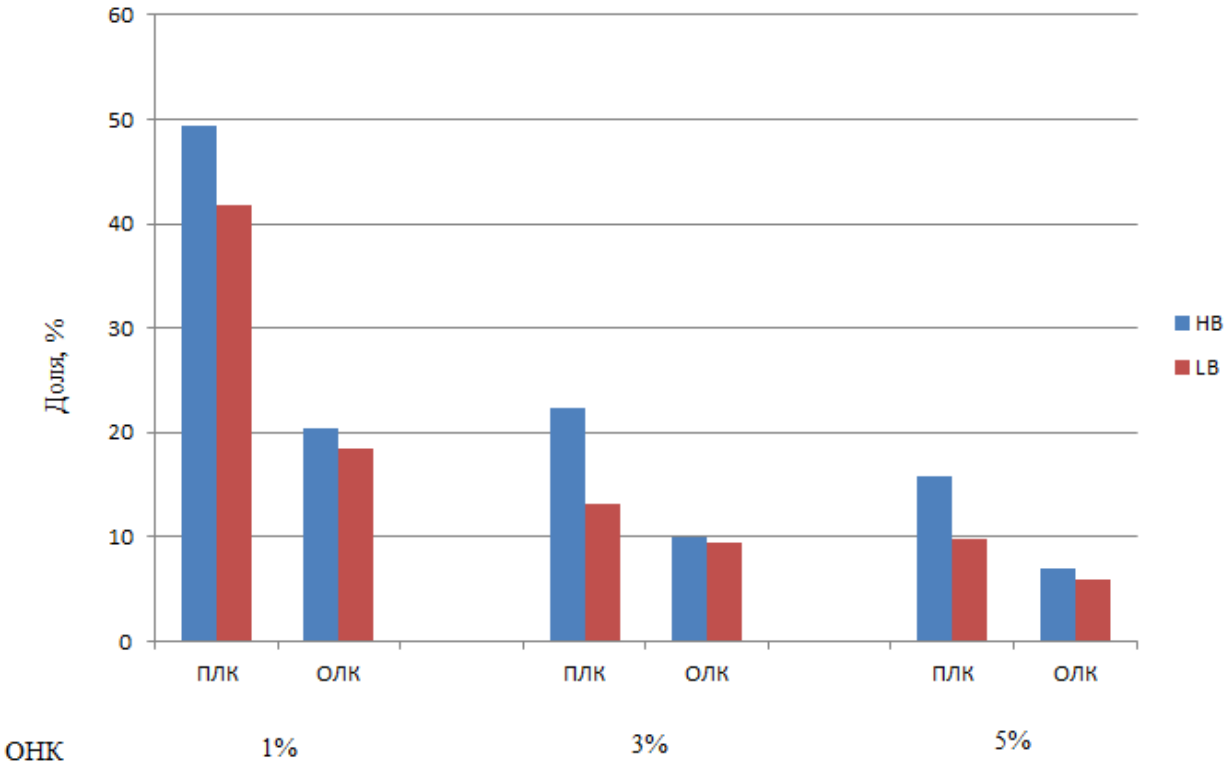

Рисунок 8 - Изменение доли некорректных классификаций в зависимости от доли фона

Отклонение оценено в процентах, при этом:

- положительное значение оценки означает завышение спрогнозированной длины по отношению к реальной на указанное количество процентов;
- отрицательное значение оценки означает занижение спрогнозированной длины по отношению к реальной на указанное количество процентов.

Численные оценки отклонения спрогнозированной длины встроенного сообщения от реальной приведены в таблице 3.

Таблица 3

Отклонение спрогнозированной длины сообщения в зависимости от доли фона

ОНК / Подвыборка	**HB**	**Эталон**	**LB**
5%	21,0%	15,2%	8,3%
3%	26,8%	17,1%	10,7%
1%	34,9%	25,4%	15,4%

Данные таблицы 3 показывают, что имеет место системное завышение прогнозируемой длины встроенного сообщения по сравнению с действительной при анализе выборки HB. Это согласуется с наблюдаемым результатом завышения доли ложноположительных срабатываний при анализе этой выборки – завышенная оценка длины сообщения, встроенного в изображение, приводит к некорректному переходу через пороговое значение и обеспечивает ложноположительную классификацию.

Таким образом, причиной падения эффективности BBC методом WS изображений с большой долей однородного фона является системное завышение длины прогноза сообщения по сравнению с действительной длиной [74].

3.1.4. Анализ причины завышения прогноза длины встроенного сообщения

Наблюдаемый эффект завышения прогноза длины сообщения согласуется с предположением о повышенной трудности различения артефактов

стеганографического встраивания и естественных случаев распределения значений яркости в фоновых зонах изображения.

Как показано в Главе II, метод WS, как и ряд прочих методов статистического ВВС в пространственной области изображения, производит итоговую оценку длины встроенного сообщения как сумму (в случае WS – взвешенную сумму) показателей, вычисляемых из конкретных областей изображения.

Таким образом, завышение итогового прогноза длины складывается из завышения оценок характеристик определённых зон изображения. Поскольку завышение итогового прогноза наблюдается именно при анализе изображений со значительной долей однородного фона, следует выделить характеристику, участвующую в вычислении итогового прогноза, и оценить причины её завышения в фоновых зонах изображения.

Метод WS в качестве области анализа рассматривает каждый отдельный пиксель изображения. Характеристика области анализа следует из итоговой формулы оценки длины встроенного сообщения для случая последовательного встраивания методом WS:

$$M = 2\sum_{i=1}^{N}(s_i - \hat{c}_i)(s_i - \overline{s_i}), \text{где}$$

M – итоговый прогноз длины встроенного сообщения, N – размер изображения в пикселях, s_i – фактическое значение уровня яркости анализируемого пикселя, $\overline{s_i}$ – фактическое значение уровня яркости анализируемого пикселя с инвертированным наименьшим значащим битом, $\widehat{c_i}$ – значение уровня яркости пикселя, оцененное по формуле прогноза значения пикселя.

Формула прогноза значения пикселя является частью метода WS и составляет ядро анализатора. Из формулы итогового прогноза видно, что оцененное значение пикселя – единственная метрика, вносимая методом в процесс ВВС – все прочие переменные формулы следуют из самого

анализируемого изображения. Таким образом, именно отклонение в оценке значения пикселя должно вносить ошибку, приводящую к наблюдаемому эффекту падения эффективности ВВС.

Из формулы очевидно, что итоговая длина сообщения складывается из разности между предсказанным и действительным пикселем изображения-стеганограммы. Множитель $s_i - \overline{s_i}$ не зависит ни от чего, кроме одного отдельно взятого пикселя изображения для каждого слагаемого итоговой суммы, и, следовательно, распределение его значений очевидно не зависит от доли фона в анализируемом изображении. В дальнейшем сомножитель $s_i - \overline{s_i}$ называется *чётностью пикселя* E, при этом пиксель называется чётным, если $E = 0$, и нечётным, если $E = 1$.

Учитывая, что растёт именно доля положительных ложных срабатываний, то есть, эффект наблюдается в том числе при анализе только чистых изображений, можно предположить, что в фоновых зонах систематически завышается предсказанное значение пикселя по отношению к реальному для нечётных пикселей изображения. Для проверки этого предположения метод WS запущен на выборке HB без стеганографического встраивания, что позволяет оценить ошибку предсказания без влияния внешнего фактора, оценив точность самого метода предсказания пикселя.

Эксперимент по оценке эффективности ВВС, приведённые в Главе I модифицирован для возможности получения характеристик областей изображения, вычисляемых в процессе анализа. Совокупность характеристик всех областей анализируемых изображений в парах со значениями этих областей составляет результат эксперимента.

В случае метода WS, характеристика области – оцененное значение яркости пикселя, область – сам пиксель. Для каждого пикселя изображения вычислена чётность, далее для всех нечётных пикселей оценена ошибка оценки значения пикселя. Минимальная величина ошибки оценки значения пикселя равна нулю,

максимальная величина лежит в диапазоне $[1.1; 5.6]$ в зависимости от анализируемого изображения. Для исключения влияния характеристик изображения на результат эксперимента, вычислено распределение вероятности ошибки оценки значения пикселя для всей анализируемой выборки. График приведён на рисунке 9.

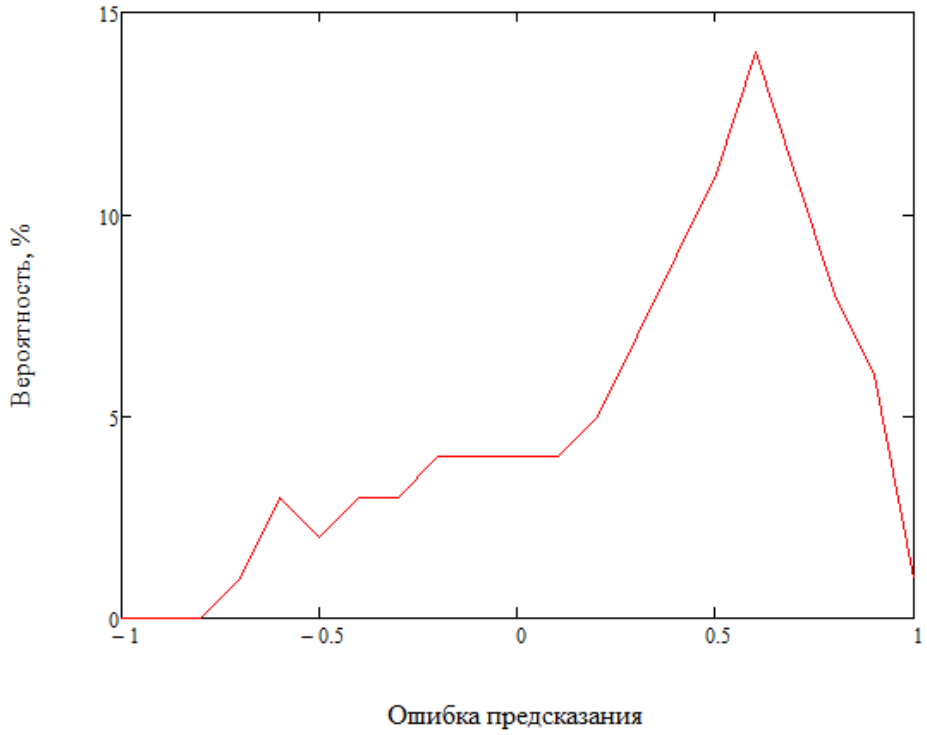

Рисунок 9 - Распределение вероятности ошибки оценки значения пикселя методом WS

Из графика на рисунке 9 видно, что при анализе изображений со значительной долей фона предсказанное значение действительно имеет пик вероятности в зоне завышения. Это объясняет высокую долю положительных ложных срабатываний.

Непосредственный результат оценки значения пикселя методом WS не обязательно является целым числом. Метод WS предусматривает округление оценки значения пикселя до целого [72]. Таким образом, имеет особое значение вероятность попадания оцененного значения в интервал $[0.5; 1.5)$, поскольку в этом случае результат округления даёт ошибку в 1, что соответствует артефакту,

наблюдаемому при стеганографическом встраивании в плоскость наименьших значащих бит изображения.

Из графика на рисунке 9 видно, что пик вероятности значения ошибки лежит в интервале $[0.5; 1.5)$, что объясняет завышение итоговой длины сообщения по сравнению с действительной.

Таким образом, методы повышения эффективности ВВС методом WS в фоновых зонах должны быть направлены на снижение ошибки предсказания значения пикселя в фоновых зонах, в частности, к сведению пиковой вероятности ошибки за пороговое значение ошибки 0.5 [74].

3.1.5. Точность оценки значения пикселя с использованием улучшенной формулы предсказания

Точность оценки, приведённая в п. 3.1.4, вычислена для формулы предсказания, приведённой авторами оригинального метода WS [32]. Процедура оценки значения пикселя при этом состоит в вычислении среднего арифметического значений четырёх пикселей, окружающих данный и лежащих сверху, снизу, справа и слева от него, как показано на рисунке 10 слева.

Таким образом, формула оцененного значения пикселя при оценке по четырём соседним следующая:

$$\widehat{c_F}(i,j) = \frac{c(i+1,j) + c(i-1,j) + c(i,j+1) + c(i,j-1)}{4}, \text{где}$$

$\widehat{c_F}(i,j)$ – оцененное значение пикселя на позиции $[i,j]$ в матрице пространственного распределения изображения, $c(i,j)$ – фактическое значение пикселя на позиции $[i,j]$.

Таким образом, при предсказании каждого конкретного пикселя полной информацией для функции предсказания являются четыре его соседних пикселя

(без учёта диагонально расположенных к данному). Ясно, что общее завышение предсказанного значения, так или иначе, складывается из завышения в каждом конкретном случае.

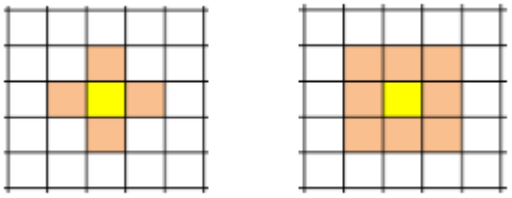

Рисунок 10 - Соседние пиксели

Выделение случаев, наиболее часто влияющих на точность предсказания может помочь нивелировать завышение. Попытка рассмотрения комбинаций значения предсказанного и окружающих пикселей для выборки тестовых изображений ожидаемо даёт огромное число различных комбинаций: фоновая зона изображения сама по себе может быть в целом светлее или темнее от изображения к изображению. Однако, анализатор каждый раз работает на одном конкретном изображении, и оттенок и яркость фона данного изображения по сравнению с другими из той же выборки не имеет значения. Следовательно, можно нормировать значения фоновых пикселей по отношению к среднему значению яркости фоновой зоны (для каждой цветовой плоскости отдельно):

$$b_n = \frac{c_n}{B_A}, \text{где}$$

b_n – нормированное значение пикселя, c_n – фактическое значение пикселя, B_A – среднее значение яркости фоновых пикселей изображения.

В результате нормирования и выделения фона для изображения получим множество нормированных фоновых пикселей B_N. Набор полной информации предсказателя в данном случае можно представить в виде четырёх соседних нормированных пикселей. Для оценки точности предсказания также следует добавить в набор действительное и предсказанное значение пикселя. Для определения типичных наборов, вводящих наибольшую ошибку предсказания,

оценено количество уникальных наборов четырёх соседей пикселя, ведущих к ошибке предсказания более чем на 0,5 относительно реального пикселя изображения. Для используемой выборки НВ в среднем для одного изображения количество таких уникальных наборов составляет 317. Такое количество не даёт возможности введения индивидуальных правил предсказания для каждого набора за разумное время. Следовательно, для улучшения точности предсказания следует изменить общую функцию предсказания пикселей изображения.
В работе [79] авторами предложен также усовершенствованный метод предсказания, оперирующий восемью соседними пикселями, как показано на рисунке 10 справа. Предсказание в этом случае осуществляется по формуле:

$$\widehat{c_E}(i,j) = 2 * \hat{c}_F(i,j) - \frac{c(i+1,j+1) + c(i-1,j+1) + c(i+1,j-1) + c(i+1,j+1)}{4}$$

График на рисунке 11 показывает характер распределения разности предсказанного и реального значения пикселя при предсказании по восьми соседним пикселям.

Как видно из рисунка 11, метод не позволяет значительно снизить разницу между предсказанным и действительным значением пикселя.

Улучшенная формула задействует все 8 соседних пикселей анализируемого. Из недостаточности всех соседних пикселей данного в задаче снижения ошибки оценки значения пикселя следует, что методы увеличения эффективности ВВС методом WS в фоновых зонах должны задействовать более дальние соседние пиксели анализируемого для более точного прогноза его значения [74].

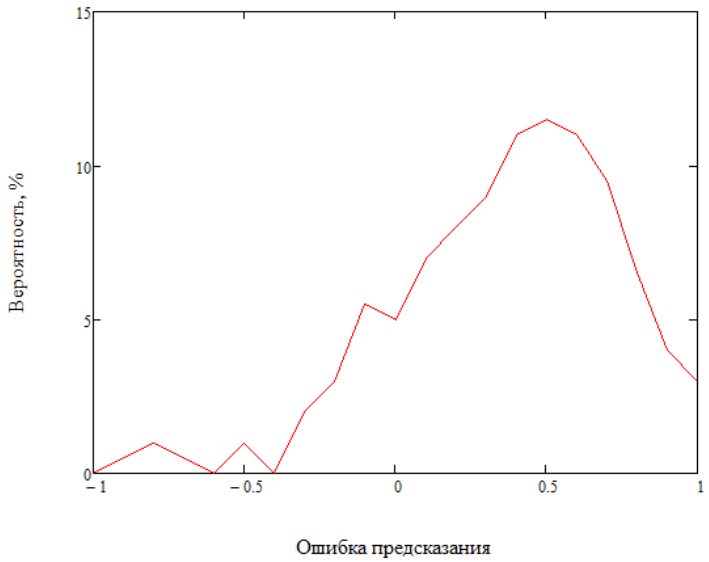

Рисунок 11 - Распределение вероятности ошибки предсказания пикселя методом WS с восемью соседними пикселями

3.1.6. Модель выявления встроенных сообщений в фоновых зонах

Как следует из п. 3.1.5, падение эффективности ВВС при анализе изображений с большой долей однородного фона связано с некорректным прогнозом значения пикселя исходя из доступных данных анализируемого изображения. Прогноз значения пикселя является центральной операцией ВВС методом WS, позволяя восстановить предполагаемое оригинальное изображение попиксельно с целью сравнения с действительным имеющимся изображением.

Модель ВВС в фоновых зонах (далее – МСФО) ставит целью увеличить точность прогноза в фоновых зонах за счёт использования более обширной информации об окружающих пикселях изображения в фоновой зоне.

В основу МСФО положено предположение о присутствии повторяющихся последовательностей пикселей в фоновой зоне – кортежей пикселей. Собрав информацию о кортежах фоновой зоны изображения до начала ВВС можно спрогнозировать значения яркости пикселей, удовлетворяющих кортежу в ходе ВВС. Если для изображения удаётся выделить кортеж, повторяющийся

значительное количество раз, прогноз по кортежу может быть точнее, чем по окружающим пикселям.

Предположение о наличии повторяющихся кортежей в фоновых зонах сделано на основании следующих фактов.

- Общее число возможных значений пикселя цветового слоя изображения в рассматриваемой модели невелико и составляет 255 значений.
- Фоновая зона, исходя из своего определения, характеризуется незначительным изменением значений соседних пикселей.

Есть возможность оценить вероятность возникновения в фоновой зоне изображения повторяющихся кортежей пикселей и привести формулу оценки ожидаемого числа повторений кортежа для данного значения пикселя.

Исходя из проблемы прогноза значения пикселя с ошибкой, лежащей в пределах от 0,5 до 1,5, очевидно, что значения соседних пикселей не должны отличаться от значения предсказываемого более, чем на 2. Исходя из предположения, что в фоновой зоне не наблюдается скачков значений пикселей, число вероятных кортежей для данного пикселя оценивается как количество возможных вариантов значений последовательности соседних пикселей в определённом направлении. При наличии допустимой разницы пикселя с соседним в 0, 1 и 2, количество уникальных кортежей длиной n оценивается:

$$U = 3^n$$

Без потери общности, для простоты рассмотрим квадратное изображение со стороной a пикселей. Общее количество пикселей изображения в цветовой модели RGB оценивается как $3a^2$. Рассматривая только фоновую зону изображения, следует ввести показатель доли фона изображения r. Таким образом, для квадратного изображения с долей фона r размером a в цветовой модели RGB с глубиной цвета 8 бит ожидаемое число повторений уникального кортежа длиной n в фоновой зоне изображения оценивается следующим образом:

$$k = r\frac{3a^2}{3^n} = r\frac{a^2}{3^{n-1}};$$

График поверхности на рисунке 12 показывает порядок величие ожидаемых значений количества повторяющихся градиентных кортежей для изображений размерами от 1000 до 2000 пикселей (диапазон взят на основе рабочего разрешения современного монитора 1920x1080 пикселей) и длины кортежей от 3 до 5 при допущении доли однородного фона в изображении 40% (пороговая доля для отнесения изображения к выборке НВ).

Исходя из оценки, доступной из графика на рисунке 12, можно ожидать значительного числа повторяющихся кортежей для данного пикселя в зонах, на которых метод BBC WS испытывает трудности с прогнозированием значений пикселя.

МСФО основывается на предположении о существовании среди множества повторяющихся кортежей для данного значения таких, которые обеспечены особенностями представления последовательных значений пикселей в пространственной области изображения. Искомые особенности могут быть вызваны деталями реализации запечатления изображения средствами фотографирования и особенностями распределения значений пикселей в изображениях при их последующей обработке.

Метод BBC WS, как и ряд прочих методов статистического ВВС, строится из предположения, состоящего в том, что на момент ВВС изображения-контейнера распределение пикселей в оригинальном изображении (том, на основании которого была получена потенциальная анализируемая стеганограмма), неизвестно. Неизвестность распределения значений в пространственной области предполагаемого чистого изображения-источника составляет основную проблему ВВС методом WS.

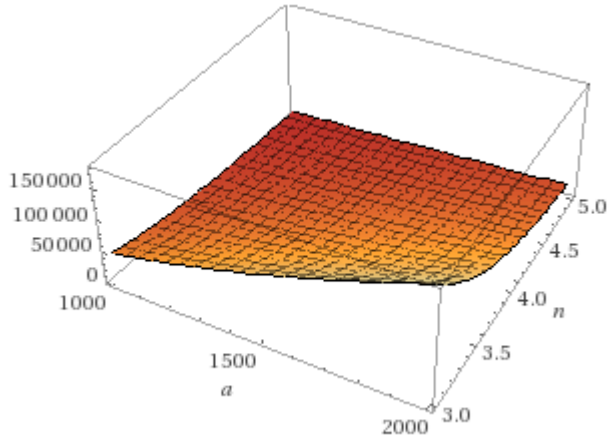

Рисунок 12 - Порядок величины ожидаемого количества повторяющихся кортежей для пикселя

В общем случае утверждение о неизвестности распределения значений в чистом изображении верно: как отмечено в Главе II, пространственная область изображения не характеризуется никакими универсальными закономерностями распределения в ней значений, стало быть, в общем случае анализа произвольно взятого изображения распределение значений в любой отдельно взятой области неизвестно.

Тем не менее, в фоновых зонах естественных изображений возможно выделение высокоуровневых структурных зон, характеризующихся системным переходом значений пикселей зоны от более тёмных к более светлым тонам – градиентов. Градиенты часто присутствуют в фоновых зонах естественных изображений за счёт особенностей освещения предметов, запечатлённых на изображениях. Несмотря на то, что градиенты, в общем случае, могут присутствовать не только в фоновых зонах изображений, именно в фоновых зонах, за счёт общей предсказуемости и равномерности распределения значений, градиентные области могут предоставлять дополнительную информацию для анализатора, предоставляя искомую общую закономерность распределения значений в пространственной области, которых не выделяется в общем случае.

Если выделить в анализируемом изображении достаточно большую область, характеризующуюся равномерным и постоянным переходом от более светлых к более тёмным тонам в изображении, то значение любого отдельно взятого пикселя анализируемой области можно спрогнозировать, не исходя из соседних пикселей, а исходя из общей характеристики изменения значений в градиенте. Значение при этом может быть предсказано точнее, чем по соседним пикселям, особенно в случае, когда направление градиента не совпадает с одним из ортогональных направлений, выбранных для соседних пикселей в методе прогноза, предложенном авторами оригинального метода WS.

Таким образом, предлагаемая модель ВВС в фоновых зонах неподвижных изображений заключается в:

1. Использовании для прогноза значений пикселей соседних пикселей с более высоким радиусом соседства.
2. Использовании для прогноза значений пикселей особенностей распределения значений пикселей в крупномасштабных областях, выделяемых в фоновой зоне.

Поскольку рассматриваемая модель ориентирована на специальную для фоновой области процедуру прогноза значения пикселей, она реализуется через алгоритмы прогноза пикселей в фоновой зоне и алгоритм выделения фоновой области в изображении.

3.2. Алгоритм прогноза значений пикселей в фоновой зоне по кортежам пикселей

3.2.1. Определение алгоритма прогноза по кортежам

Алгоритм предполагает предварительную обработку анализируемого изображения до начала выполнения процедуры ВВС. Предобработка заключается в построении матрицы кортежей пикселей фоновой области.

Пусть I – анализируемое изображение, элемент пространственного распределения которого определён как $I[i,j]$, где i,j – координаты пикселя. Тогда n-кортеж для пикселя $I[i,j]$ – кортеж $I[i \pm 1, j] \dots I[i \pm n, j]$. Таким образом, n-кортеж для пикселя представляет собой упорядоченное множество n соседних пикселей в определённом направлении. В зависимости от направления, такие кортежи обозначаются nL, nR, nT и nB (соответственно, для направлений влево, вправо, вверх и вниз).

Рисунок 13 иллюстрирует четыре кортежа пикселей для данного анализируемого в пространственном распределении анализируемого изображения. В примере на рисунке 13 длина кортежей равна 4. Таким образом, представлены 4L, 4R, 4T и 4B кортежи пикселя.

После построения всех четырёх кортежей для каждого пикселя изображения, для каждого уникального кортежа определяется количество пикселей, для которого он был построен. Уникальность кортежа определяется по его длине в пикселях, уровню яркости пикселей, составляющих его и его направлению. Самый часто встречающийся пиксель для данного кортежа – спрогнозированный пиксель. Набор пар «цепь – спрогнозированный пиксель» составляет *матрицу кортежей пикселей*. При прогнозе значения пикселя анализируемого изображения для него сроится n-кортеж. Если такой n-кортеж присутствует в матрице кортежей пикселей, значение пикселя берётся из матрицы для данной цепи. Если цепь отсутствует, значение прогнозируется как среднее четырёх соседних пикселей.

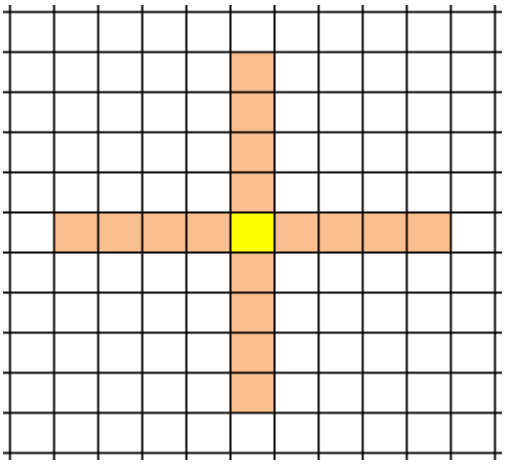

Рисунок 13 - 4L, 4R, 4T и 4B кортежи пикселя

Вводится понятие мощности кортежа. Мощность данного кортежа для данного значения пикселя – количество раз, которое данный кортеж был построен для данного пикселя при предобработке данного изображения. Таким образом, мощность как характеристика принадлежит не кортежу, а паре «кортеж – пиксель», то есть, элементу матрицы кортежей. Мощность включается в матрицу кортежей пикселей.

При предобработке изображения для одного значения пикселя может быть построено более одного уникального кортежа. После построения всех кортежей для всех пикселей матрица кортежей фильтруется по следующим правилам.

- Если для одного значения пикселя построено более одного кортежа, и мощности кортежей данного пикселя различны – в матрице остаётся кортеж наибольшей мощности, прочие исключаются из матрицы. Этот шаг фильтрации призван исключить случайные, несистемные цепочки пикселей, не встречающиеся в изображении большое число раз.
- Если для одного значения пикселя построено более одного кортежа, и присутствуют несколько кортежей максимальной и одинаковой мощности – в матрице остаётся кортеж наибольшей длины, прочие исключаются из матрицы. Этот шаг призван исключить короткие цепочки, встречающиеся значительное число раз за счёт ограниченности вариантов значений яркости пикселей анализируемого изображения.

Таким образом, после фильтрации каждому уникальному значению пикселя в матрице соответствует один кортеж.

Вводятся пороги доверия матрице соседства пикселей. При построении матрицы на изображении без явно выраженных структурных паттернов матрица может давать неадекватные результаты, которые могут отрицательно влиять на эффективность анализа. Для выборки размером K изображений для каждого кортежа вводятся пороги доверия, t – относительный, и a – абсолютный:

$$t = \frac{N_m}{N_s},$$

$$a = \frac{N_m}{K}, \text{где}$$

N_m – количество появления предсказанного пикселя для данного кортежа, N_s – количество появления данного кортежа для всей выборки, K – количество изображений в выборке.

Изменяя пороги доверия можно варьировать область применения матрицы соседства пикселей и анализа по соседним пикселям.

Рисунок 14 представляет собой визуализацию части матрицы кортежей пикселей, получаемой в ходе предобработки.

Различные матрицы соседства пикселей применяются как в статистическом ВВС, так и в ВВС на основе машинного обучения. Как правило, матрицы соседства пикселей строятся на обучающей выборке, после чего применяются при анализе тестовой выборки (в частности, такой подход применяется в методе ВВС в неподвижных цифровых изображениях SPAM). В общем случае метод не располагает обучающей выборкой, лишь изображением, предназначенным для анализа. Для применения метода в таком случае можно использовать само анализируемое изображение как обучающее. Это возможно при условии применения метода для малых величин полезной нагрузки. Учитывая, что в рассматриваемом случае значение полезной нагрузки невелико, искажений,

вносимых в изображение стеганографическим встраиванием, недостаточно, чтобы внести значительную погрешность в закономерности расположения пикселей в фоновых зонах, с учётом применения абсолютного и относительного порогов доверия матрице.

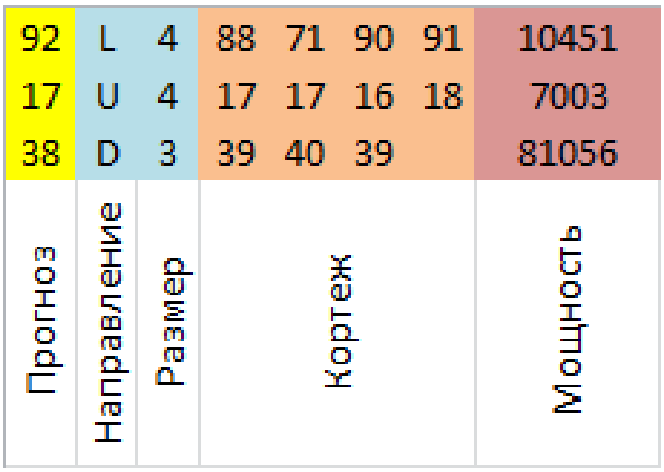

Рисунок 14 - Визуализация части матрицы кортежей пикселей

Матрицу следует применять и обучать только в фоновых зонах изображения. В нефоновых зонах, за счёт большего числа вариантов изменения значения соседнего пикселя по сравнению с текущим, количество повторяющихся кортежей для данного пикселя значительно меньше (как показано в п. 3.2.1, средняя мощность кортежа для данного пикселя характеризуется обратной экспоненциальной зависимостью от числа вариантов изменения значения пикселя по сравнению с текущим).

В частных случаях, в фоновой зоне также возможны значительные изменения значения пикселя по сравнению с соседним. Вводится интервал фона B – максимальная разница между значением пикселя и значением пика, при котором пиксель считается принадлежащим фону. Фоновой зоной считается та, для которой по меньшей мере для двух из трёх цветовых плоскостей разница между значением пикселя и пиком для данной цветовой плоскости не превышает B.

Таким образом, итоговый набор параметров усовершенствованного алгоритма включает в себя пороги доверия a, t и интервал фона B.

Усовершенствованная модель прогноза пикселя задаётся следующим уравнением:

$$pr[i,j] = \begin{cases} m[i,j], |pr[i,j] - P| > B \\ m[i,j], |pr[i,j] - P| < B, \nexists M(\mathrm{C}(p[i,j]), a, t, B) \\ M(\mathrm{C}(p[i,j]), a, t, B), |pr[i,j] - P| < B, \exists M(\mathrm{C}(p[i,j]), a, t, B) \end{cases}$$, где

$p[i,j]$ – пиксель анализируемого изображения,

$pr[i,j]$ – прогноз значения пикселя,

$m[i,j]$ – значение, предсказанное по среднему окружающих пикселей,

P – пиковое значение гистограммы для цветового слоя,

M – операция взятия предсказанного значения из матрицы кортежей,

C – операция вычисления кортежа для пикселя.

3.2.2. Блок-схемы алгоритма прогноза по кортежам пикселей

В соответствие с описанием в п. 3.2.1., алгоритм прогноза по кортежам пикселей включает в себя этап предобработки анализируемого изображения, состоящий в формировании матрицы кортежей, и этап прогноза значения пикселя изображения, используемый как часть общего алгоритма выявления встроенного сообщения.

Блок-схема на рисунке 15 иллюстрирует общий вид процесса формирования матрицы кортежей.

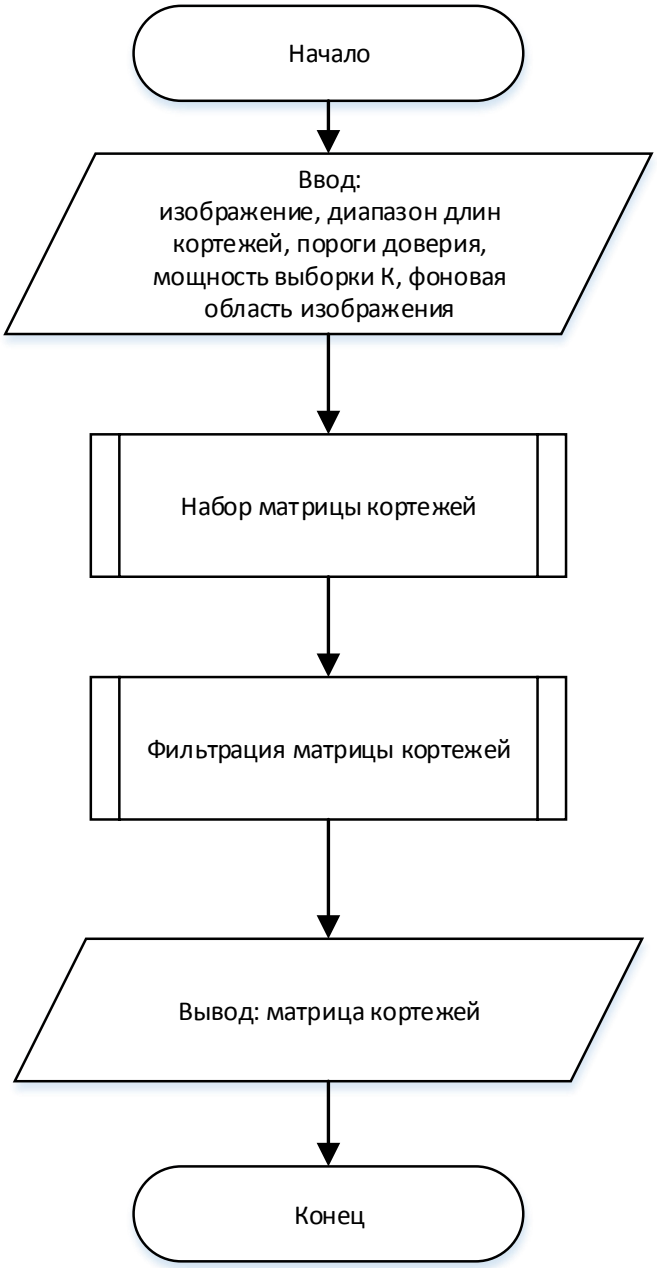

Рисунок 15 – Общий вид процесса формирования матрицы кортежей

Шаги набора и фильтрации матрицы выделены в отдельные подпроцессы для удобства представления в виде блок-схем. Рисунки 16, 17 иллюстрируют, соответственно, подпроцессы шагов набора и фильтрации матрицы кортежей.

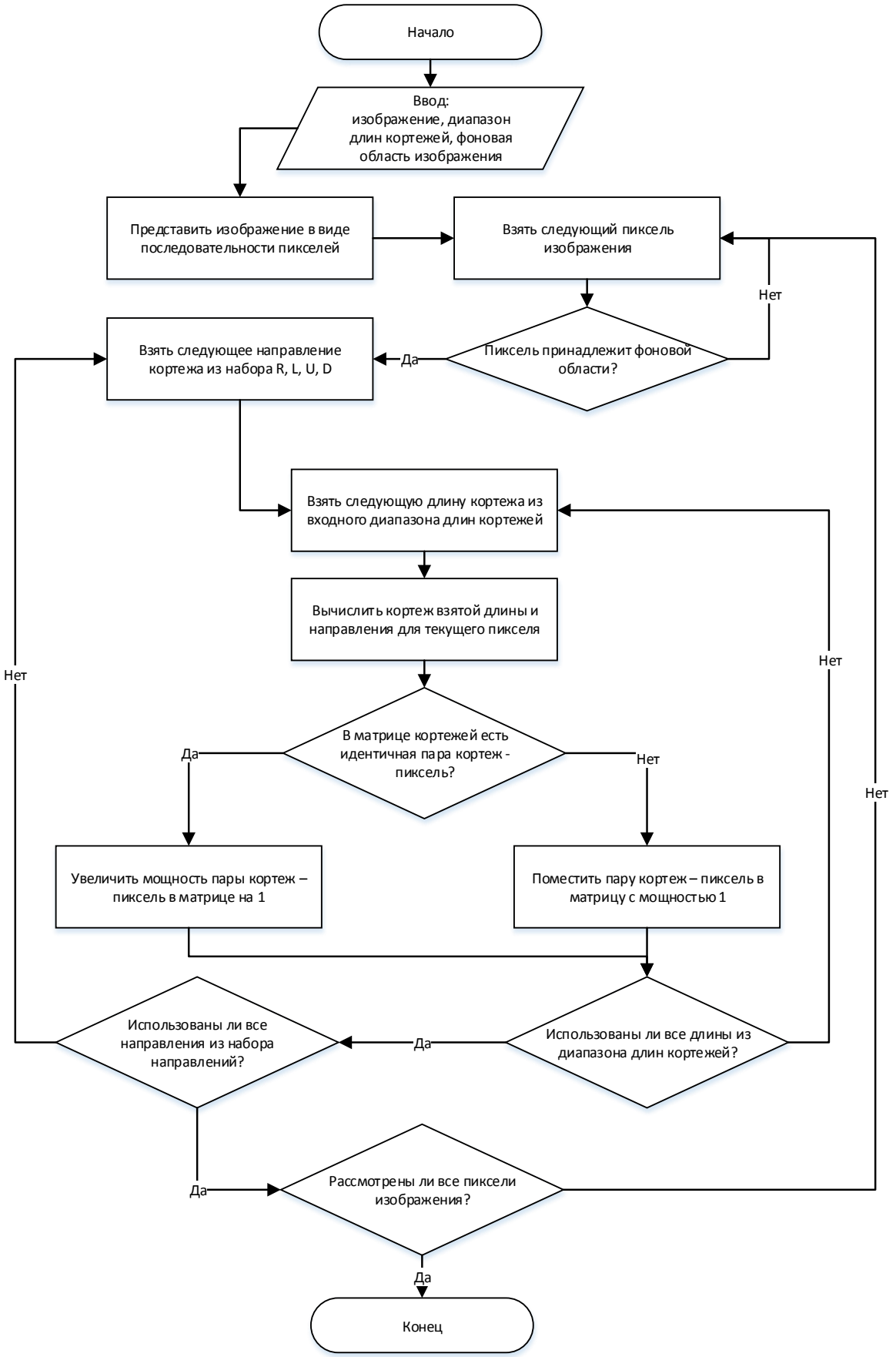

Рисунок 16 – Шаг набора матрицы кортежей изображения

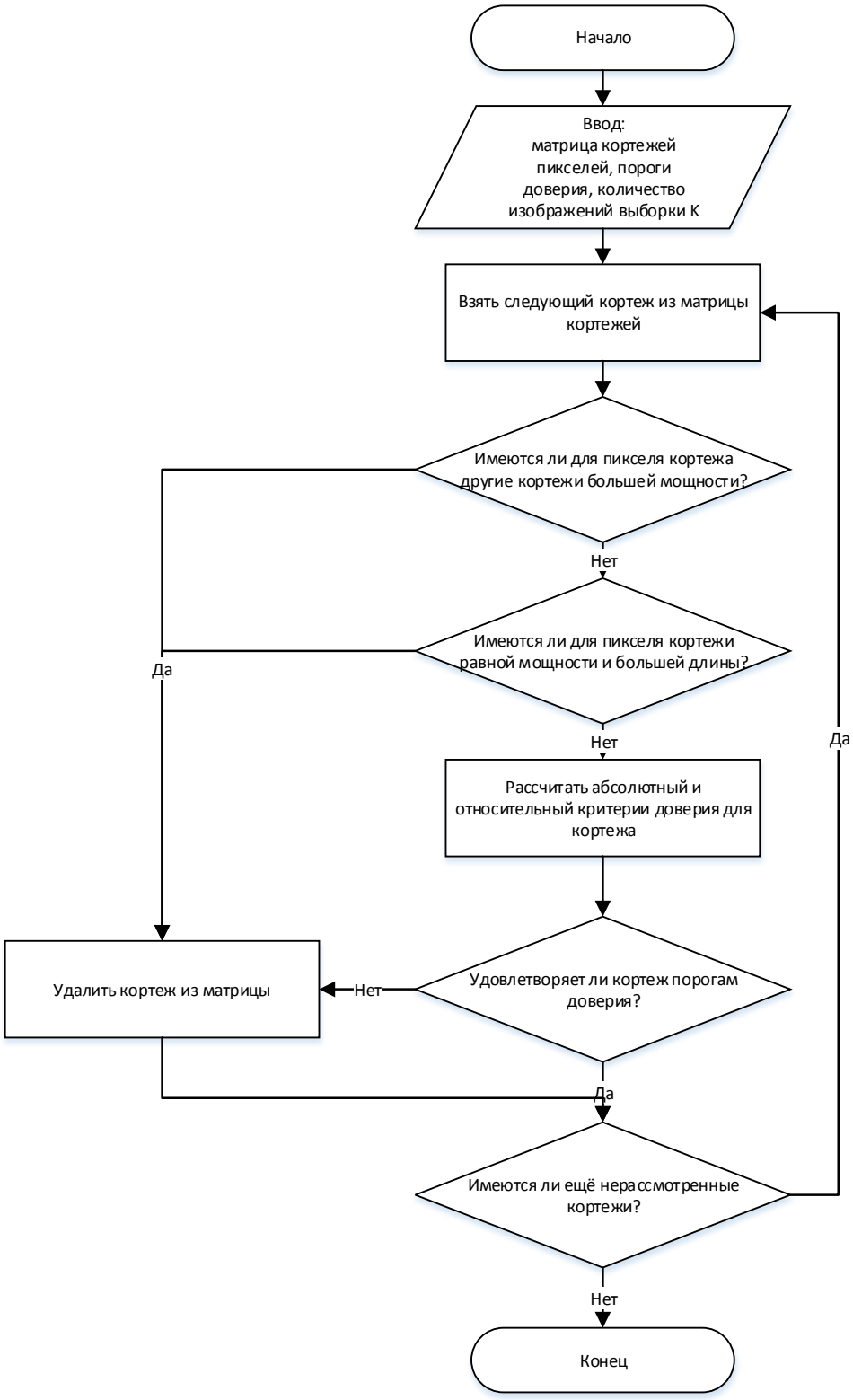

Рисунок 17 – Шаг фильтрации матрицы кортежей изображения

На выходе процесса формирования матрицы кортежей – готовая матрица, используемая для прогноза значений пикселей в ходе анализа. Процесс прогноза как части общего алгоритма ВВС иллюстрирует блок-схема на рисунке 18.

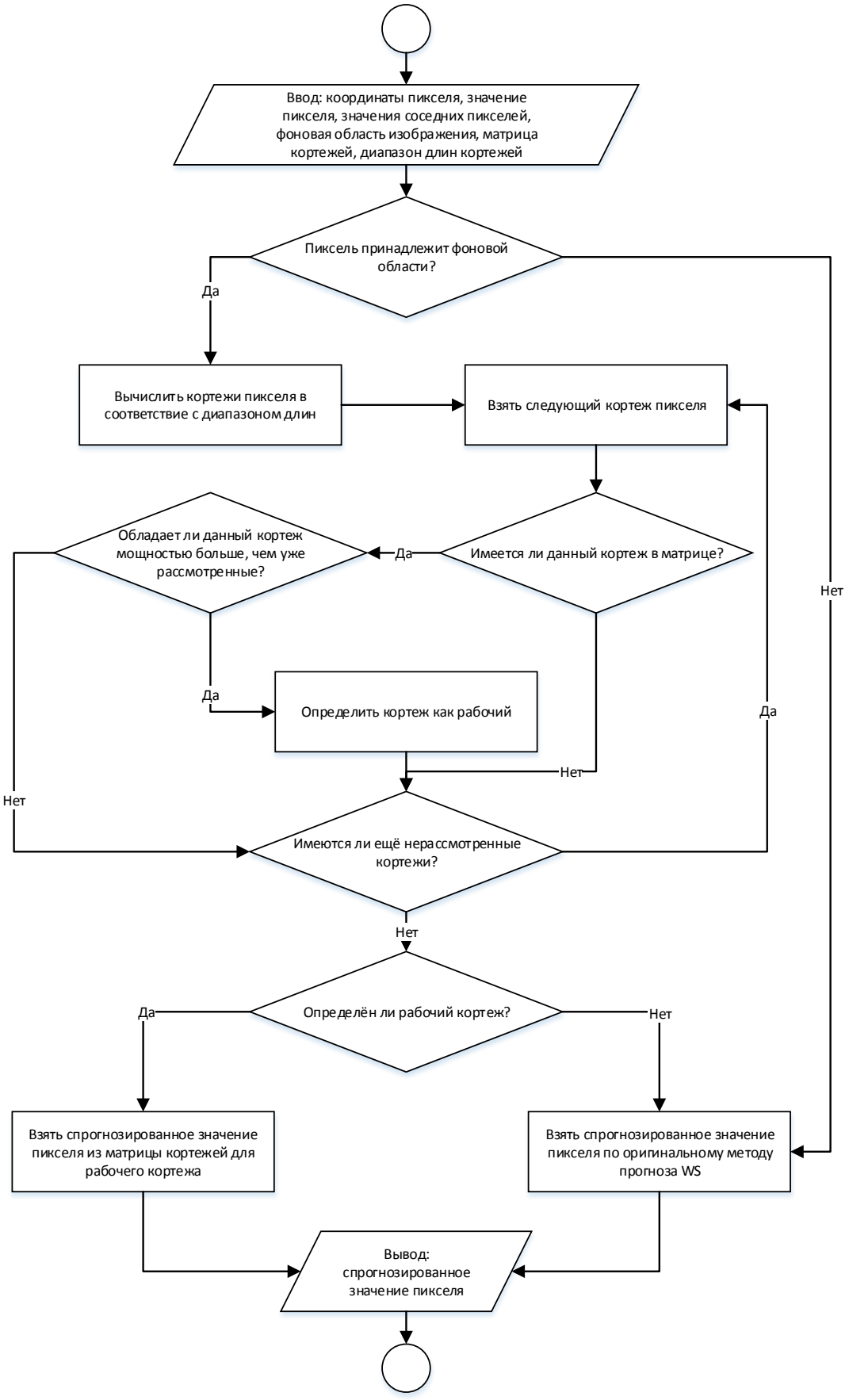

Рисунок 18 – Шаг фильтрации матрицы кортежей изображения

3.2.3. Практическая оценка эффективности применения и вычислительная сложность алгоритма прогноза по кортежам

Алгоритм прогноза по кортежам ставит целью снижение ошибки прогноза значения пикселя в фоновых зонах изображений.

Рисунки 19, 20 демонстрируют распределение вероятности ошибки прогноза пикселя в фоновых зонах для коллекций HB и LB соответственно. На графиках кривая PLAIN показывает распределение вероятности для прогноза по среднему соседних пикселей. Кривая CHAIN показывает распределение вероятности при использовании прогноза по кортежам пикселей.

Выборки и способ их формирования аналогичен описанному в п. 3.1. Использованы следующие пороги доверия: $t = 0.9, a = 200$.

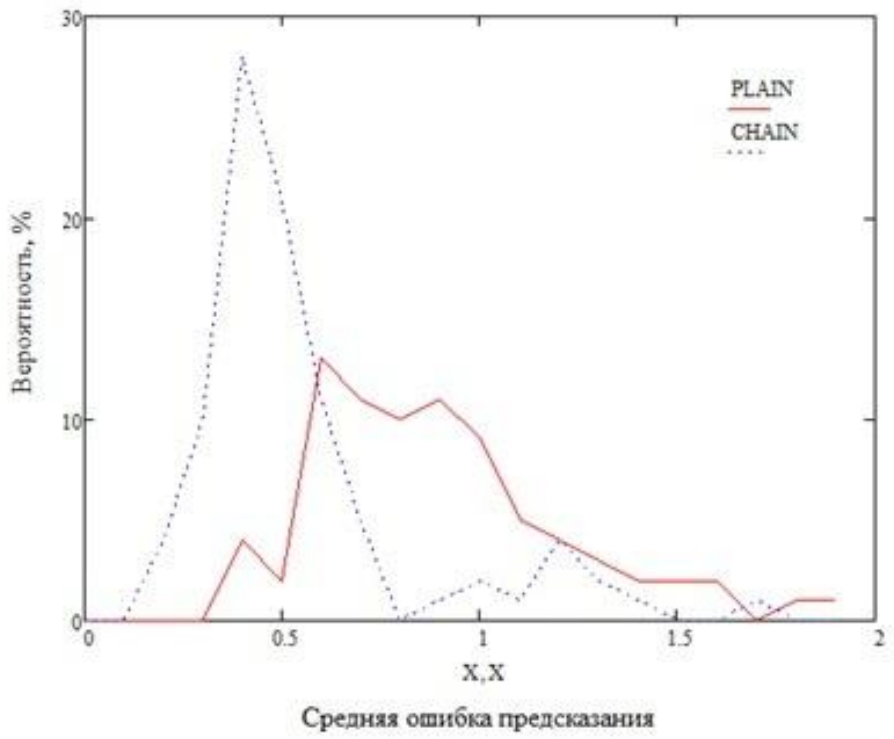

Рисунок 19 - Распределение вероятности ошибки прогноза для коллекции HB

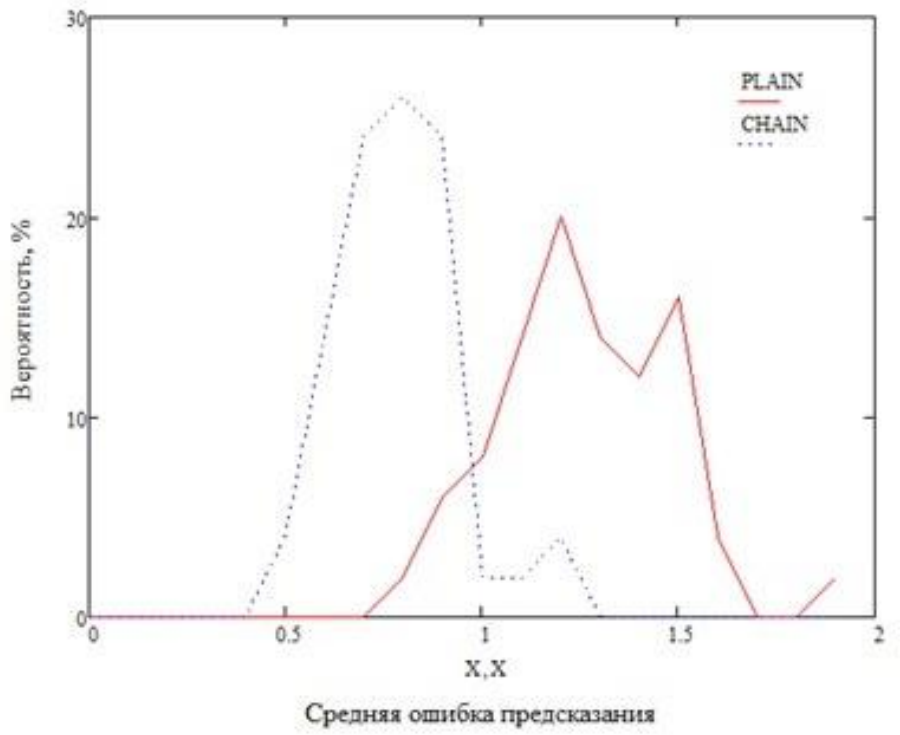

Рисунок 20 - Распределение вероятности ошибки прогноза для коллекции LB

Таблица 4 содержит численные оценки средней ошибки прогноза пикселя для различных коллекций и методов прогноза.

Из графиков на рисунках 19, 20 и данных таблицы 4 видно, что алгоритм прогноза по кортежам позволяет значительно снизить ошибку прогноза пикселя в фоновых зонах, выводя максимум распределения вероятности за интервал [0.5; 1.5].

Таблица 4

Средняя ошибка прогноза пикселей для различных коллекций и методов прогноза

Метод / Коллекция	**HB**	**LB**
По среднему	1.22	1.74
По кортежам	0.74	0.96

Предположение о том, что факт стеганографического встраивания не оказывает значительного влияния на статистику кортежей, также нуждается в экспериментальном подтверждении.

График на рисунке 21 демонстрирует среднюю ошибку прогноза по кортежам в зависимости от значения полезной нагрузки.

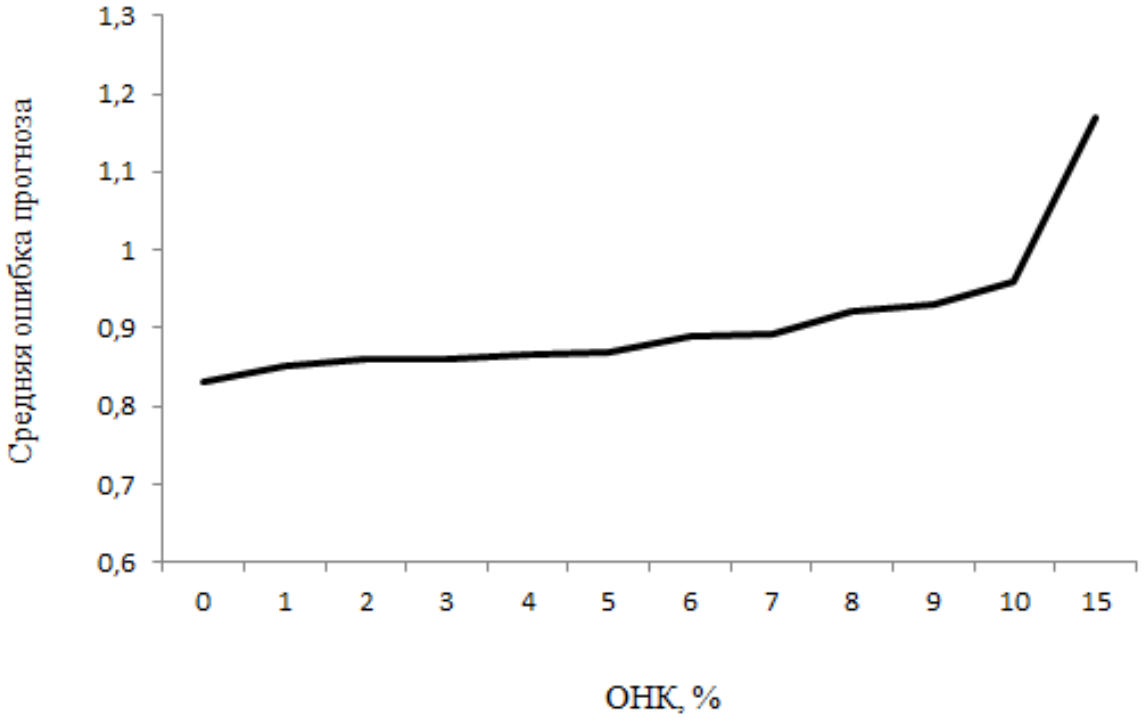

Рисунок 21 - Средняя ошибка прогноза по кортежам в зависимости от значения полезной нагрузки

Из графика на рисунке 21 видно, что при малых значениях ОНК ошибка предсказания растёт незначительно, демонстрируя заметный рост лишь при значениях полезной нагрузки, превышающих 10%.

Алгоритм прогноза по кортежам позволяет повысить эффективность ВВС в изображениях со значительной долей однородного фона за счёт снижения ошибки прогноза пикселей в фоновых зонах анализируемого изображения [80].

Вычислительная сложность оригинального алгоритма прогноза значения пикселя по среднему соседних пикселей имеет вычислительную сложность $O(n)$. Предлагаемый алгоритм состоит из следующих этапов.

- Вычисление кортежей пикселей для каждого пикселя. С учётом фиксированной длины кортежа и количества кортежей сложность шага оценивается как $O(n)$.
- Добавление со слиянием кортежей в матрицу кортежей. Сложность оценивается как $O(1)$.
- Поиск значения пикселя для данного кортежа в матрице кортежей. Сложность оценивается как $O(1)$.

Таким образом, общая сложность предлагаемого алгоритма оценивается как $O(n)$ и не превышает сложность существующего алгоритма.

3.3. Алгоритм адаптивного прогноза в градиентных областях

3.3.1. Определение алгоритма адаптивного прогноза в градиентах

Как и алгоритм прогноза по кортежам, алгоритм адаптивного прогноза в градиентах предполагает предобработку анализируемого изображения и построение статистик, используемых в дальнейшем при анализе изображения.

Предобработка изображения состоит в выполнении следующих шагов:

1. Выделение градиентных областей в изображении
2. Расчёт параметров, необходимых для использования градиента в задаче прогноза значения пикселей.

Поскольку алгоритм состоит в использовании градиентов именно фоновых зон изображение, выделение фоновой зоне предшествует выделению градиентов.

В доступных источниках не приведено применимых в данной задаче способов выделения градиентов в изображениях. Алгоритм классификации градиентов строится из представления о том, как её впоследствии можно использовать. Под градиентом понимается фрагмент фона изображения с

выраженным и постоянным изменением яркости пикселей определённого цветового слоя в постоянном направлении. Таким образом, для цветных изображений градиенты следует рассчитывать для каждого слоя в отдельности. Градиентные области, которые для человеческого глаза выглядят едиными, для модели предсказания оказываются неоднородными, поскольку направление градиента различное в разных частях области.

За направление градиента примем направление любого из соседних пикселей данного. Таким образом, получим восемь направлений, обозначенных $U, D, L, R, UR, UD, LR, LD$ (Up, Down, Up-Right и так далее). Для выделения градиентов на изображении рассмотрим каждый пиксель зоны, ранее классифицированной как фоновая. Выделим градиентные области по следующему правилу: пиксель считается принадлежащим градиентной области направления D, если для любых P его соседей в данном направлении («продольные соседи») разница между двумя соседними пикселями не отличается от среднеквадратичного её значения на всём рассматриваемом наборе более, чем в k раз (первичный критерий), и для N его соседей в направлениях, перпендикулярных данному («поперечные соседи»), для данного направления выполняется то же условие (вторичный критерий). Таким образом, принадлежность пикселя с множеством перпендикулярных соседей B_N размером N множеству пикселей градиентных областей A_{gr} изображения I определяется условием:

$$\nexists a_i : |a_i - a_{i+1}| > k\sqrt{\sum_{1}^{P}(a_i - a_{i+1})^2} \cap \forall b \in B_n : |b_i - b_{i+1}| < k\sqrt{\sum_{1}^{P}(b_i - b_{i+1})^2}$$
$$\rightarrow a \in A_{gr}, \text{где } i \in [0, P], a_i(a_{mn}) = I[m \pm p][n \pm p], b_i(b_{qr}) = A[q \pm p][r \pm p]$$

Рисунок 22 иллюстрирует градиентные структуры так, как они определяются методом выделения градиента. На рисунке градиент слева

соответствует характеристике $UD, P = 4, N = 2$, градиент справа соответствует характеристике $URDL, P = 4, N = 2$. Желтый пиксель соответствует исходному, для которого строится градиент. Оранжевые пиксели соответствуют «продольным соседям», синие – «поперечным соседям», зелёные – «продольным соседям» «поперечных соседей».

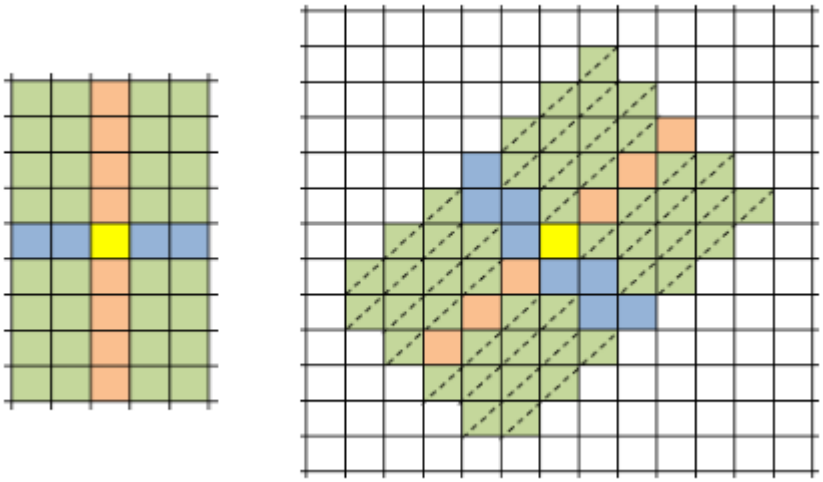

Рисунок 22 - Примеры градиентов в пространственной области изображения

P соседей пикселя в выбранном направлении обозначим как его *градиентный кортеж*. За силу градиента примем среднеквадратичное значение величины изменения значения пикселя по сравнению с предыдущим для его градиентного кортежа.

Результат работы алгоритма выделения – множество A_{gr}, с указанием для каждого силы и направления градиента, а также связи пиксель – градиентный кортеж, вместе составляющие матрицу градиентов.

Один пиксель изображения может удовлетворять условию принадлежности нескольким градиентам. Поскольку прогноз значения пикселя в процессе анализа должен быть получен исходя из определённого градиента и должен представлять одно число, конфликт принадлежности пикселя нескольким градиентам разрешается на этапе построения матрицы градиентов по следующим правилам.

1. В множестве градиентов остаётся тот, для которого показатель N градиента больше
2. В случае, если пиксель принадлежит нескольким градиентам с одинаковым показателем N, в множестве градиентов остаётся тот, для которого показатель P больше.
3. Если пиксель принадлежит нескольким градиентам с одинаковыми показателями P, N, выбирается любой из представленных случайным образом.

Рисунок 23 иллюстрирует часть получаемого множества градиентов. Множество для каждого пикселя, принадлежащего градиенту, содержит его градиентный кортеж с указанием силы и направления градиента.

1089	102	301	300	300	299	297	297	16	UL-DR
800	10987	0	0	1	1	1	2	5	D
7090	3	77	76	76	75	74	72	22	RL
Координата X	Координата Y	Градиентный кортеж						Сила	Направление

Рисунок 23 - Часть множества A_{gr}

Прогнозирование значения пикселя, принадлежащего кортежу, производится подбором значения, наименьшим образом меняющего силу кортежа. Таким образом, для данного кортежа K длиной P спрогнозированное значение пикселя должно удовлетворять условию:

$$a_f : S(P, a_f) = \mathrm{S}_{\min}(P, a_u), a_u \in U, a_f \in U, \text{где}$$

U – множество целых значений пикселей, из которого ведётся подбор данного. Множество значений определяется как среднее значение пикселей, окружающих данный прогнозируемый пиксель $\pm$ определённое значение допуска R. Допуск подбирается эмпирически. Множество выбора U вводится для облегчения задачи

прогнозирования пикселя за счёт уменьшения области определения функции прогноза. Подобное допустимо за счёт предположения, что значения яркости пикселей в фоновой зоне не меняется по отношению к соседним на значительную величину, что позволяет ограничить R малым числом.

В условиях стеганографического встраивания с малой полезной нагрузкой алгоритм прогнозирования работает за счёт того же эффекта, что и прогноз по матрице соседства пикселей: изменение отдельных значений в объёмных градиентах не позволяет размыть статистику по градиенту настолько, чтобы средняя ошибка прогноза увеличилась более, чем на единицу, что позволит алгоритму анализа сделать вывод о наличии стеганографического встраивания в LSB пикселей.

Несмотря на то, что алгоритм позволяет прогнозировать значения пикселей, использовать её в отрыве от прогнозирования по матрицам соседства пикселей нерационально, так как последняя хорошо справляется в фоновых зонах в целом. Прогнозирование по матрице соседства пикселей хорошо работает для градиентов доминирующего направления в данном изображении (несложно показать, что в идеальном градиенте прогнозирование по матрице соседства пикселей будет давать идеальную эффективность). Тем не менее, когда направления градиентов значительно изменяются по изображению и нельзя выделить единое доминирующее направление, прогнозирование по матрице соседства в градиентах становится менее точным, чем прогнозирование по кортежам. Для прогноза можно использовать взвешенное среднее прогнозов по матрице соседства пикселей и по кортежам, формула итогового прогноза пикселя выглядит следующим образом:

$$a_F = wa_{fm} + (w-1)a_{fg},$$

где a_{fm} – значение, спрогнозированное по матрице, a_{gf} – значение, спрогнозированное по градиентному кортежу, w – экспериментально подбираемый вес, $w \in [0; 1]$.

Таким образом, алгоритм адаптивного прогноза в градиентах рассматривается неразрывно с методом прогноза по кортежам пикселей, как его дополнение и расширение, использующее более сложную семантическую структуру фоновой зоны изображения.

3.3.2. Блок-схемы алгоритма адаптивного прогноза в градиентах

Как и алгоритм прогноза значений пикселей по кортежам, алгоритм адаптивного прогноза градиентов включает в себя фазу предобработки изображения, состоящую в формировании множества градиентов изображения, и фазу прогноза значения пикселя, являющуюся частью общего алгоритма ВВС.

Формирование множества градиентов, в свою очередь, подразделяется на шаги набора множества градиентов и фильтрации множества градиентов. На выходе у фазы – отфильтрованное множество градиентов, готовое к использованию в фазе прогноза значения пикселя.

Блок-схема на рисунке 24 иллюстрирует общий вид фазы формирования множества градиентов. Блок-схемы на рисунках 25, 26 подробнее иллюстрируют процессы шагов набора и фильтрации множества градиентов соответственно. Наконец, блок-схема на рисунке 27 показывает процесс прогноза значения пикселя как часть общего алгоритма ВВС.

Рисунок 24 – Общий вид процесса формирования множества градиентов

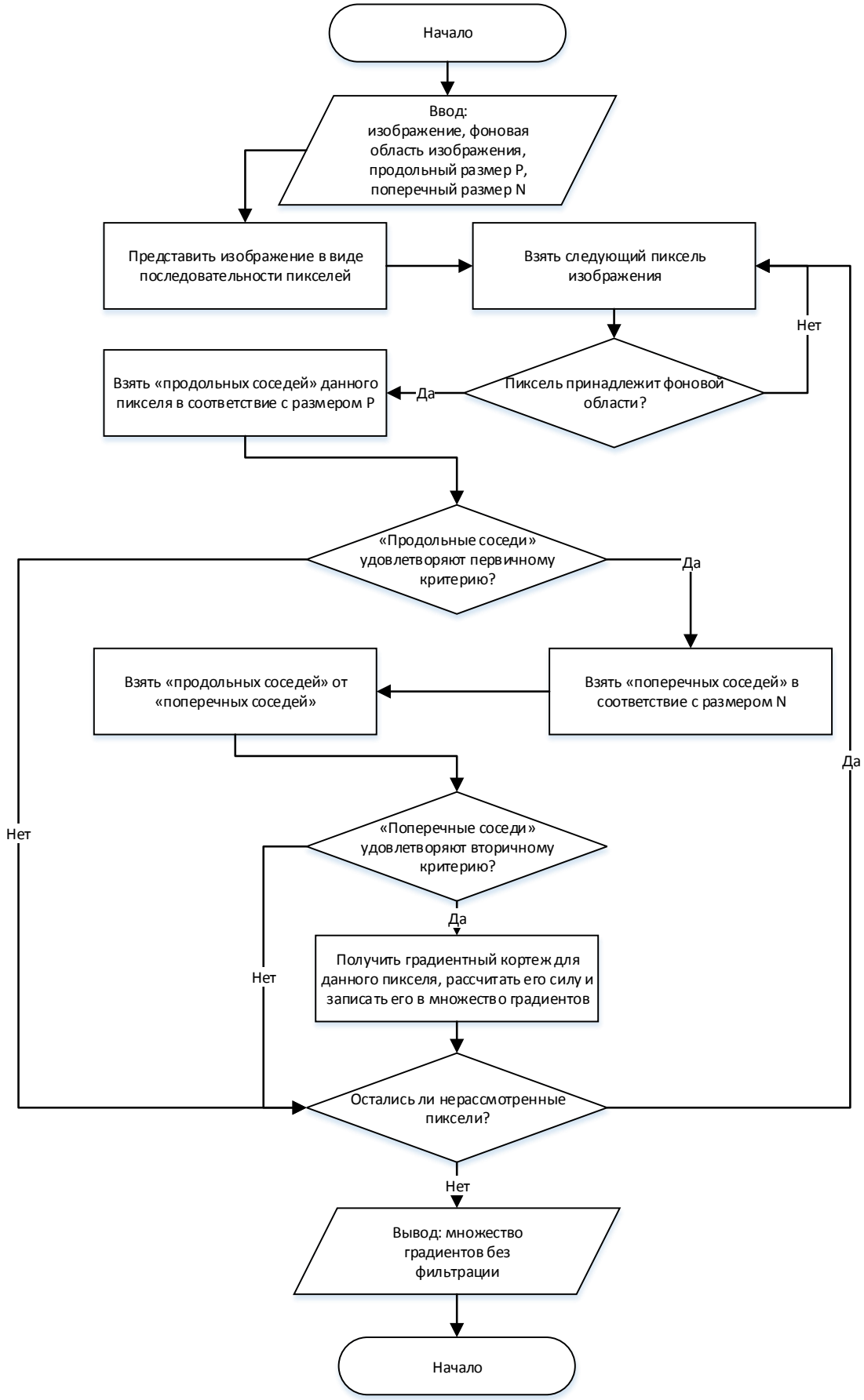

Рисунок 25 – Шаг набора множества градиентов

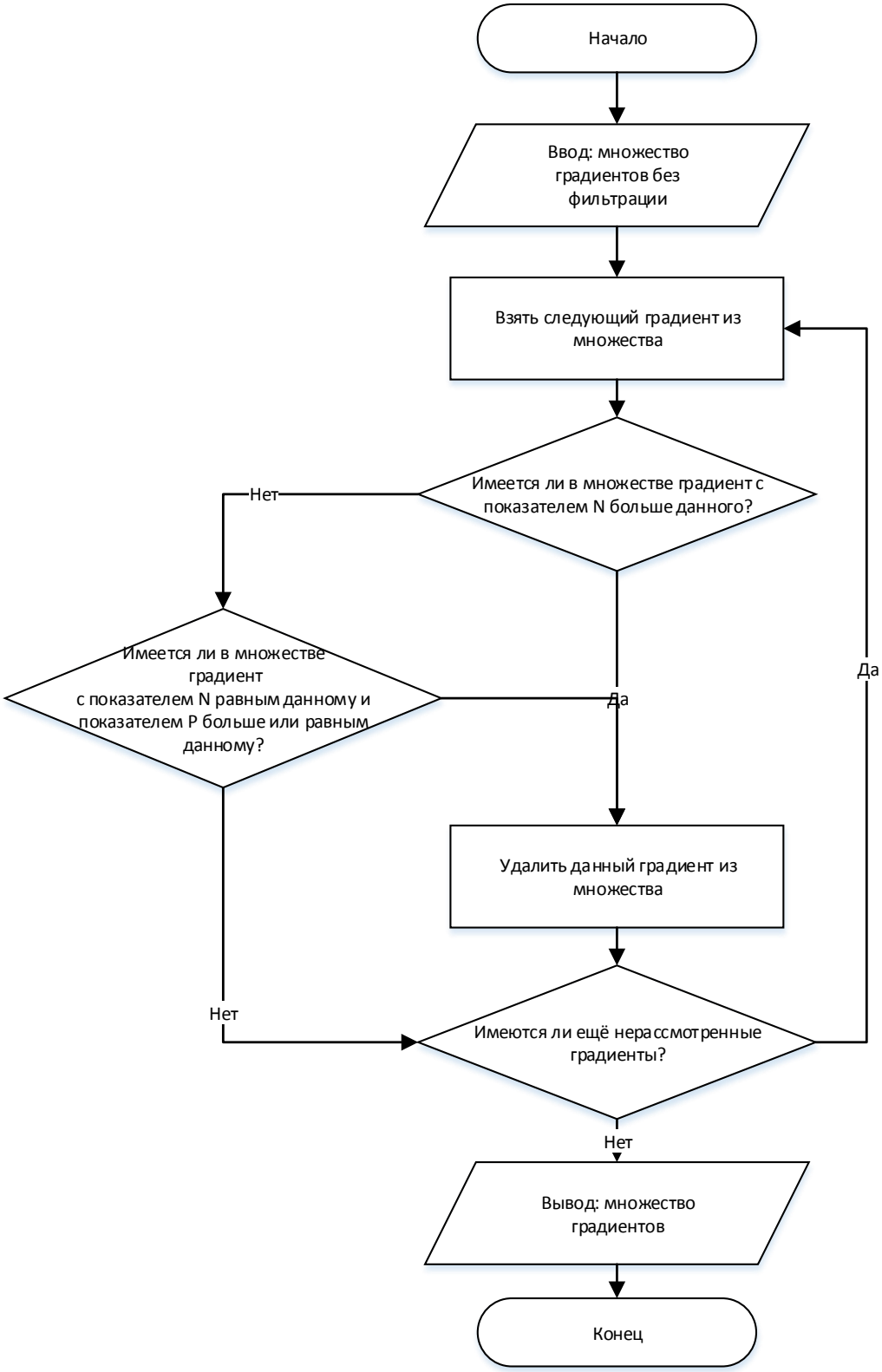

Рисунок 26 – Шаг фильтрации множества градиентов

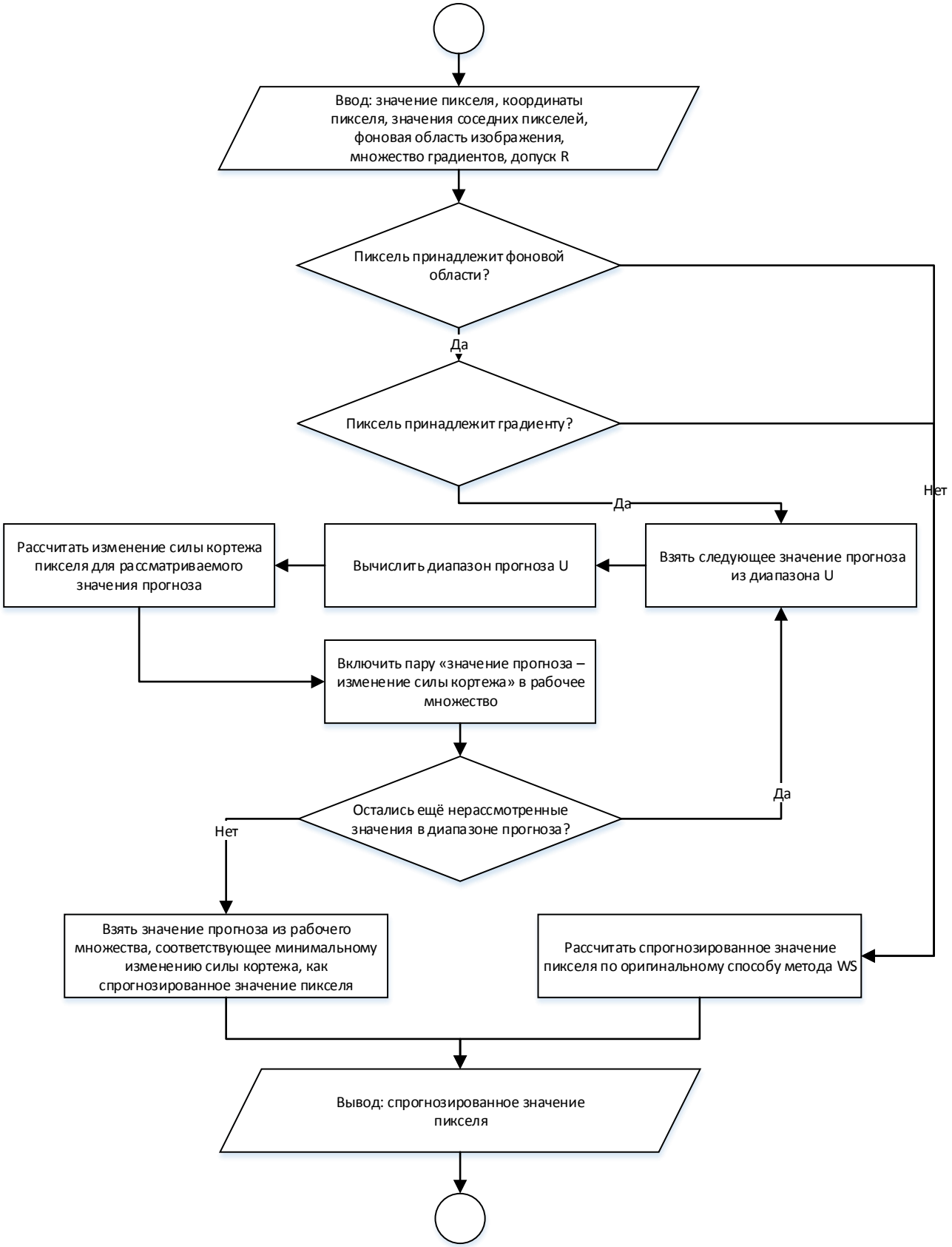

Рисунок 27 – Прогноз значения пикселя по множеству градиентов

3.3.3. Практическая оценка эффективности применения и вычислительная сложность алгоритма адаптивного прогноза в градиентных областях

Алгоритм адаптивного прогноза в градиентах ставит целью снижение ошибки прогноза значения пикселя в фоновых зонах изображений.

Рисунок 28 демонстрирует распределение вероятности ошибки прогноза пикселя в фоновых зонах для коллекци HB. На графиках кривая GC показывает распределение вероятности для прогноза по кортежам пикселей и адаптивном прогнозе в градиентах. Кривая CHAIN показывает распределение вероятности при использовании прогноза только по кортежам пикселей.

Выборки и способ их формирования аналогичен описанному в п. 3.1.

Таблица 5 демонстрирует среднее значение ошибки прогноза пикселя для коллекций HB и LB при использовании адаптивного прогноза в градиентах и при прогнозе только по кортежам пикселей.

Таблица 5

Средняя ошибка прогноза значения пикселя при использовании адаптивного прогноза в градиентах и без

Метод/Коллекция	**HB**	**LB**
По кортежам	0,74	0,96
По кортежам и в градиентах	0,63	0,82

Из графика на рисунке 28 и данных таблицы 5 видно, что использование адаптивного прогноза в градиентах позволяет снизить ошибку прогноза пикселя. Ошибка прогноза, соответствующая пиковому значению вероятности, уменьшается более чем на 50% [78].

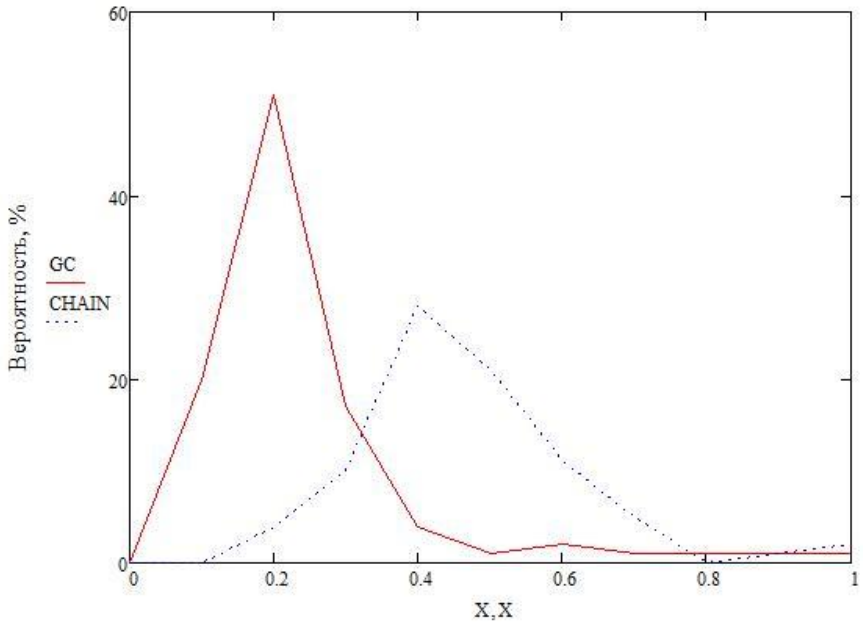

Рисунок 28 - Распределение вероятности ошибки прогноза при использовании адаптивного прогноза в градиентах и без

Предложенный алгоритм включает следующие этапы:

- Определение принадлежности пикселей изображения к градиентным областям. Сложность оценивается как $O(n)$.
- Расчёт статистической характеристики градиентов для всех пикселей, принадлежащих градиентам. Сложность оценивается как $O(n)$.
- Поиск рассчитанного градиента для анализируемого пикселя изображения. Сложность оценивается как $O(1)$.
- Подбор значения пикселя исходя из статистических особенностей его градиента. Сложность оценивается как $O(n)$ при использовании ограничивающего множества рассматриваемых пикселей U.

Таким образом, общая сложность алгоритма прогноза в градиентах оценивается как $O(n)$.

3.4. Алгоритм накопления статистики анализатора

3.4.1. Обоснование подхода

Алгоритмы увеличения эффективности ВВС, описанные в п. 3.2 и п. 3.3 подразумевают накопление статистики анализируемого изображения перед началом выполнения непосредственно процесса ВВС. Статистика, накопленная для анализируемого изображения, отражает устойчивые закономерности распределения значений пикселей в нём.

В наиболее изолированном случае ВВС, метод ВВС располагает одним анализируемым изображением в один момент времени и не имеет информации о прочих изображениях, анализируемых или планируемых к анализу в рамках жизненного цикла анализатора.

Тем не менее, в реальных системах пассивного противодействия скрытым каналам передачи информации с использованием стеганографических средств встраивания, задача ВВС не ставится относительно одного отдельно существующего изображения. Канал передачи данных на основе стеганографии подразумевает наличие множества сообщений, передаваемых по каналу последовательно, и каждое из них подвергается процедуру ВВС с целью различения чистых изображений и стеганопосылок.

Таким образом, выявитель работает на наборе изображений. В наиболее общем случае, набор изображений недоступен целиком в момент начала работы анализатора, а поступает к нему последовательно из обрабатываемого канала передачи данных.

Оригинальный метод ВВС WS не предусматривает использования факта ВВС предыдущих изображений анализируемого наборе при анализе текущего. Алгоритмы прогноза значений пикселей, предложенные выше, оперируют накопленной статистикой, которую возможно переиспользовать при анализе

следующих поступающих изображений при выполнении критериев применимости накопленной статистики для прогноза значений пикселей в следующем анализируемом изображении.

Алгоритм накопления статистики состоит в определении способа оценки степени применимости накопленной статистики для анализа текущего изображения и способа учёта статистики, собранной при анализе текущего изображения при анализе следующих поступающих изображений.

3.4.2. Определение метода повышения эффективности выявления встроенных сообщений за счёт накопления статистики

Переиспользованию подлежит матрица кортежей, собираемая в ходе предобработки анализируемого изображения в рамках алгоритма предсказания по кортежам пикселей.

Поскольку алгоритм прогноза значения пикселя по кортежам предполагает использование единственной матрицы кортежей пикселей, предыдущие матрицы, собранные с проанализированных изображений, должны пройти процедуру объединения с образованием результирующей матрицы кортежей для переиспользования при анализе текущего изображения. Для того, чтобы матрицы, собранные с предыдущих изображений, способствовали увеличению точности прогноза, а не размыванию статистики, требуется определить степень подобия матриц кортежей и объединять только матрицы, обладающие степенью подобия выше определённой пороговой величины.

Матрица кортежей, построенная для текущего анализируемого изображения, дополняется информацией о том, для скольких случаев в изображении такая пара «пиксель – кортеж» встречена. Таким способом образуется гистограмма кортежей, в которой идентификатор столбца

гистограммы – уникальная пара «пиксель – кортеж», а высота столбца – количество таких пар *только в текущем* изображении.

Схожесть гистограмм кортежей двух изображений используется в качестве метрики схожести этих изображений. Это позволяет не использовать гистограммы для анализа изображений, обладающих сильно отличающейся статистикой кортежей.

Факт схожести гистограмм определяется следующим образом:

1. Для двух гистограмм G_1 и G_2 определяются пары «кортеж – пиксель», входящие в обе гистограммы. Множество таких пар G_M.
2. Для каждого элемента g множества G_M рассчитывается отношение схожести по формуле:

$$S = \begin{cases} \dfrac{m(G_1, g)}{m(G_2, g)}, m(G_1, g) < m(G_2, g) \\ \dfrac{m(G_2, g)}{m(G_1, g)}, m(G_2, g) < m(G_1, g) \end{cases}, \text{где}$$

$m(G, g)$ – количество вхождений элемента g в гистограмме G.

Определяется количество элементов, для которых отношение схожести S меньше определённого порога S_T. Если доля таких элементов в множестве G_M больше определённого порога T_G, считается, что гистограммы G_1 и G_2 схожи.

Рисунок 29 иллюстрирует процедуру определения подобия и объединения матриц кортежей двух изображений.

На гистограммах G_1 и G_2 на рисунке красным и зелёным цветом выделены пары «кортеж – пиксель», входящие в обе гистограммы. Синим выделены пары, уникальные каждая для своей гистограммы.

Гистограмма G_M на рисунке 29 показывает отношение схожести гистограмм по их общим элементам. Красным выделена доля мощности кортежей, входящих в гистограмму G_1. Зелёным выделена доля мощности кортежей, входящих в гистограмму G_2.

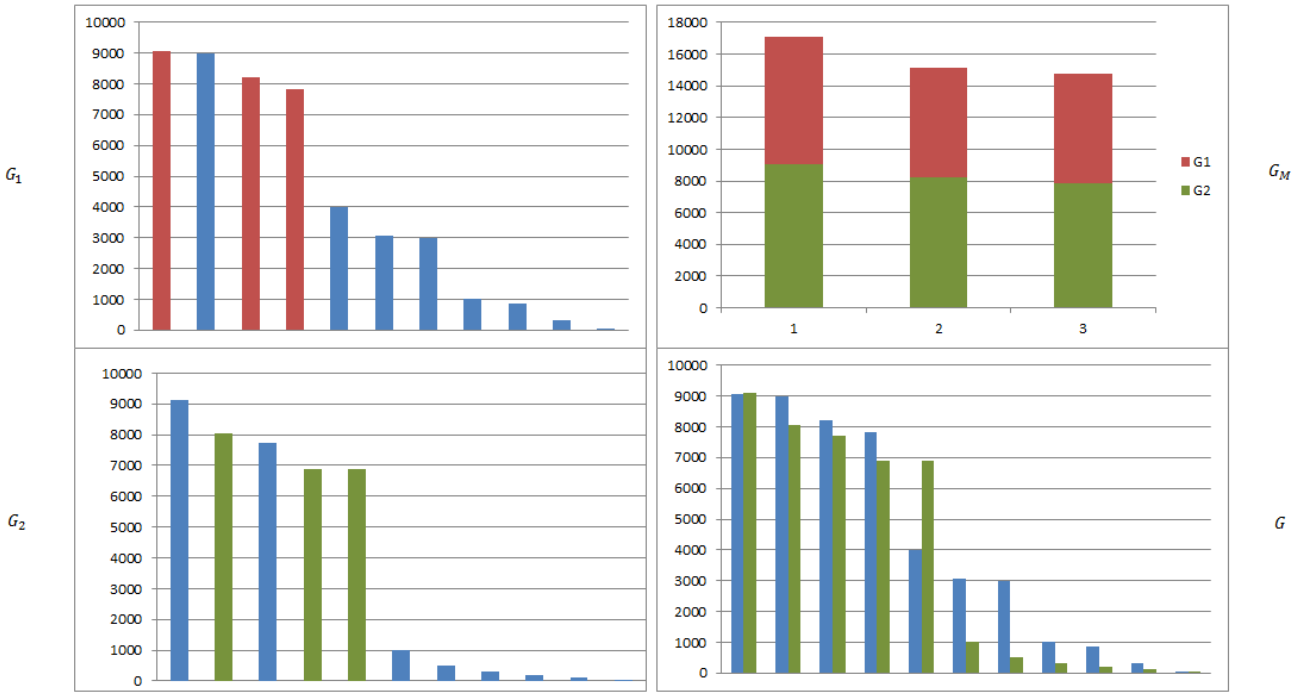

Рисунок 29 - Процедура объединения матриц кортежей

Гистограмма G на рисунке 29 – результат объединения гистограмм G_1 и G_2. Следует отметить, что гистограмма G содержит все элементы обеих гистограмм, а не только элементы, общие для двух исходных. Такой способ объединения гистограмм позволяет не терять уникальные статистические данные как текущего анализируемого, так и предшествующих изображений.

Для схожих гистограмм матрицы кортежей объединяются и при анализе изображения используется объединённая матрица, включающая информацию обеих матриц. Поскольку информация о частоте встречи определённого элемента матрицы не нужна в процессе ВВС, матрицы кортежей объединяются путём простого объединения множеств их элементов.

При анализе потока изображений образуется множество накопленных гистограмм кортежей $\mathbb{G}$. При анализе очередного изображения из множества определяется гистограмма, обладающая наибольшей степенью схожести с гистограммой данного изображения. Эти гистограммы объединяются. Результат объединения используется для анализа текущего изображения и замещает

гистограмму в множестве. Таким способом происходит постепенное накапливание статистики по мере прохода по множеству анализируемых изображений.

3.4.3. Блок-схемы алгоритма накопления статистики анализатора

Алгоритм накопления статистики анализатора не используется напрямую в процессе ВВС, отрабатывая перед ним. Алгоритм накопления статистики анализатора состоит в последовательном выполнении операции объединения матриц в процессе обработки множества анализируемых изображений.

При этом, непосредственно в процессе выявления встроенного сообщения, используется алгоритм прогноза значения пикселя, описанный для метода прогноза по кортежам пикселей.

Блок-схема на рисунке 30 иллюстрирует общий вид алгоритма объединения матриц кортежей.

Алгоритм включает в себя два подпроцесса – алгоритм определения подобия матриц и алгоритм слияния матриц. Для удобства представления в графическом виде, подпроцессы представлены в виде отдельных блок-схем.

Рисунки 31 и 32 содержат блок-схемы алгоритмов определения подобия матриц и слияния матриц кортежей соответственно.

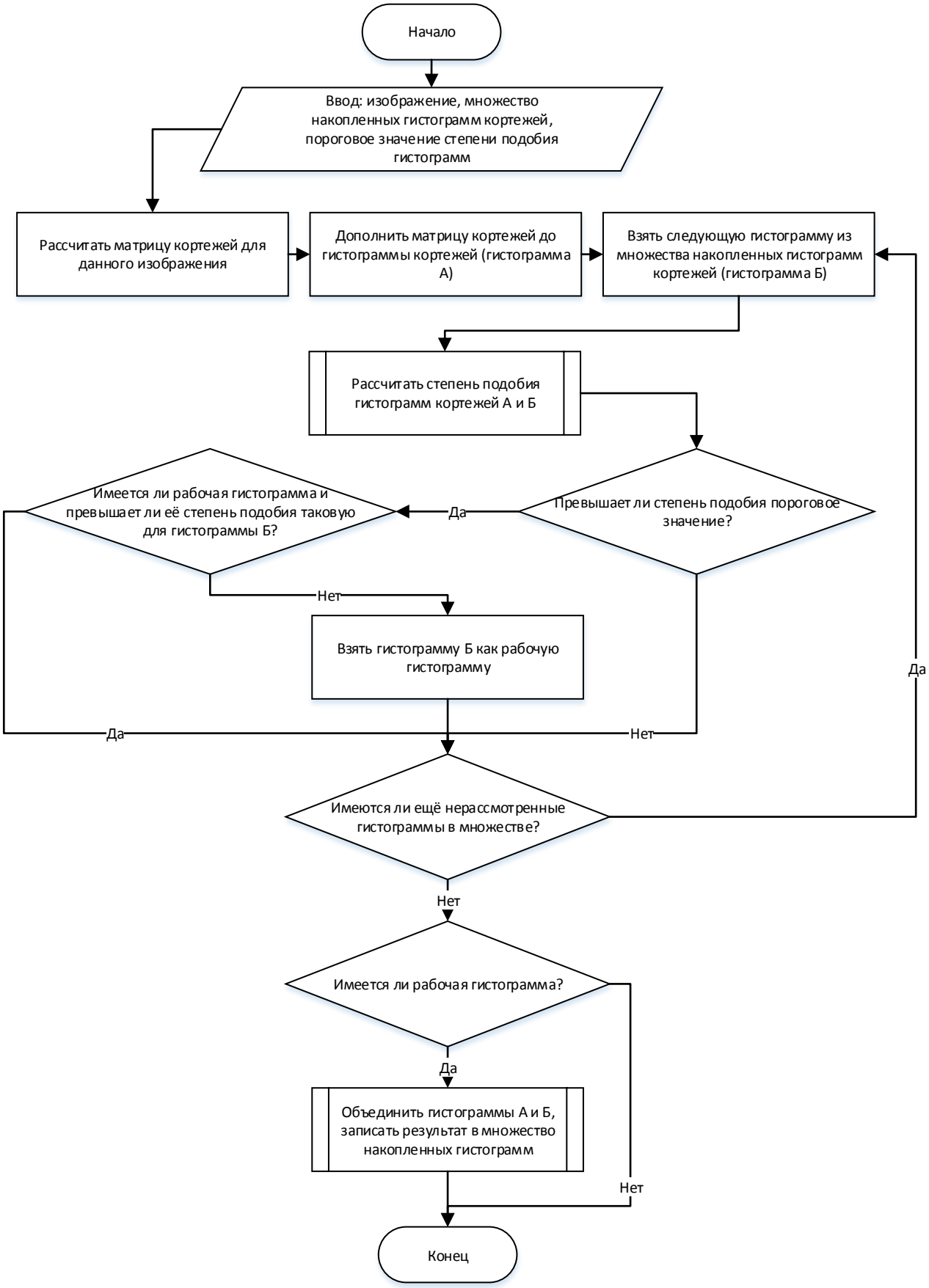

Рисунок 30 – Алгоритм объединения матриц кортежей

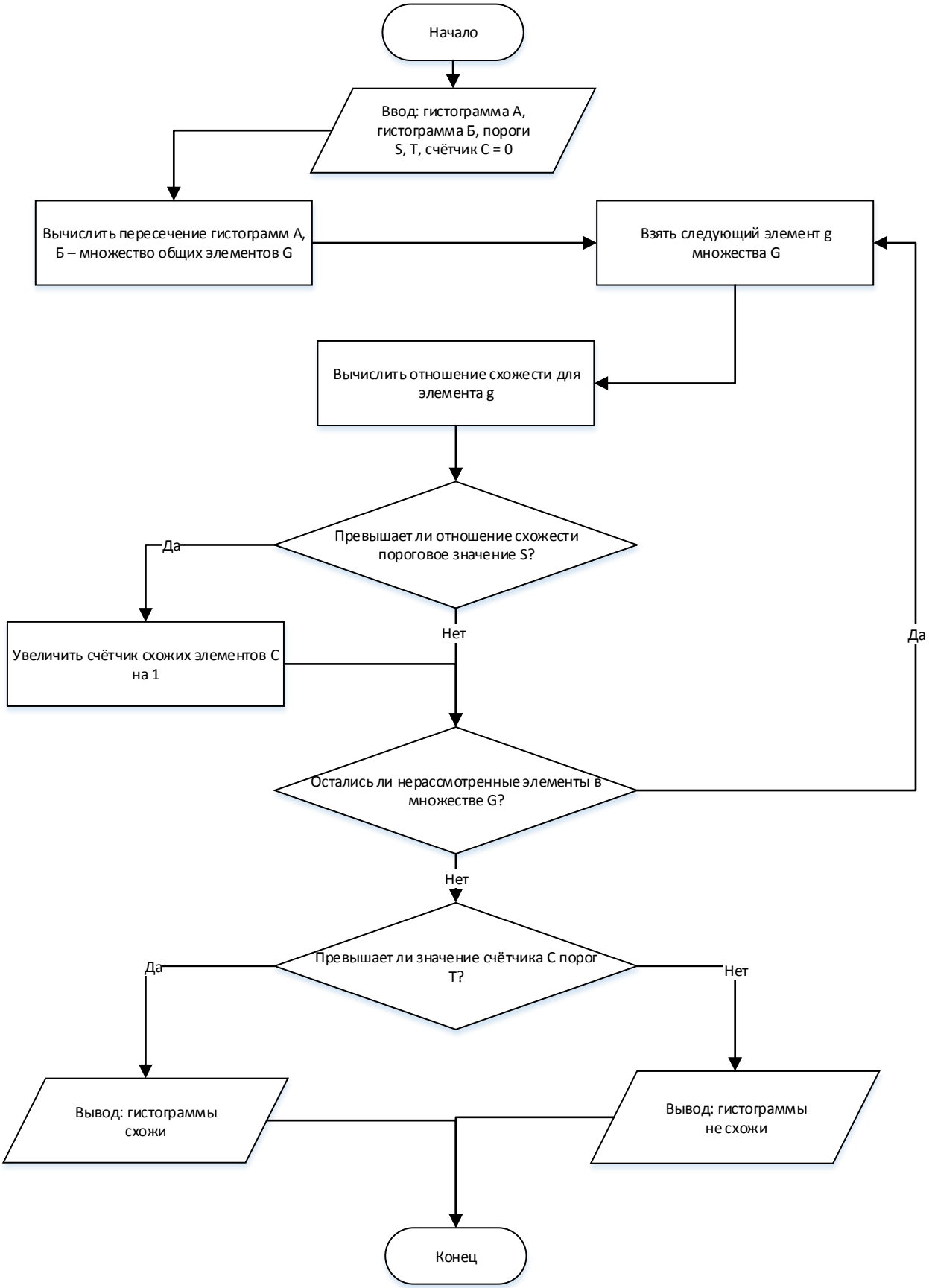

Рисунок 31 – Алгоритм определения подобия матриц кортежей

Рисунок 32 – Алгоритм слияния матриц кортежей

3.4.4. Практическая оценка эффективности применения алгоритма накопления статистики и его вычислительная сложность

В общем случае, поток анализируемых изображений может обладать различными характеристиками схожести. Для подтверждения применимости алгоритма накопления статистики требуется показать, что:

1. алгоритм позволяет добиться уменьшения ошибки прогноза пикселя на множестве схожих изображений;

2. алгоритм не ведёт к увеличению ошибки прогноза пикселя на множестве несхожих изображений.

Следующий эксперимент ставит целью подтвердить применимость алгоритма накопления статистики.

1. Из выборки HB исходного тестового множества изображений формируются две подвыборки SIM и DIF.
    a. Подвыборка SIM – максимальная по размеру подвыборка, для которой гистограммы кортежей всех входящих в неё изображений обладают степенью схожести, допустимой для объединения этих гистограмм при выполнении анализа.
    b. Подвыборка DIF – максимальная по размеру подвыборка, для каждых двух изображений которой гистограммы кортежей не обладают степенью схожести, допустимой для объединения этих гистограмм при выполнении анализа. Подвыборка DIF дополняется определённым процентом случайно выбранных из выборки HB изображений для активации объединения гистограмм в некоторых случаях. Это позволяет оценить возможный негативный эффект объединения гистограмм и анализа несхожих изображений именно объединёнными гистограммами.
2. Подвыборки SIM и DIF анализируются методом WS с предложенными методами повышения эффективности анализа. Для каждого анализируемого изображения оценивается средняя ошибка прогноза пикселя. После каждого проанализированного изображения оценивается средняя ошибка прогноза пикселя по выборке. Средняя ошибка по выборке оценивается как среднее арифметическое средних ошибок прогноза для каждого проанализированного на данный момент изображения выборки.

Следующие пороговые величины определены эмпирически и используются в эксперименте:

- Отношение схожести элементов матрицы кортежей $S_T = 0.88$
- Предельный порог схожести матриц кортежей $T_G = 0{,}72$.

Проведённый эксперимент подтверждает оба утверждения. График на рисунке 33 показывает изменение средней ошибки прогноза для выборок SIM и DIF по мере прохождения выборок.

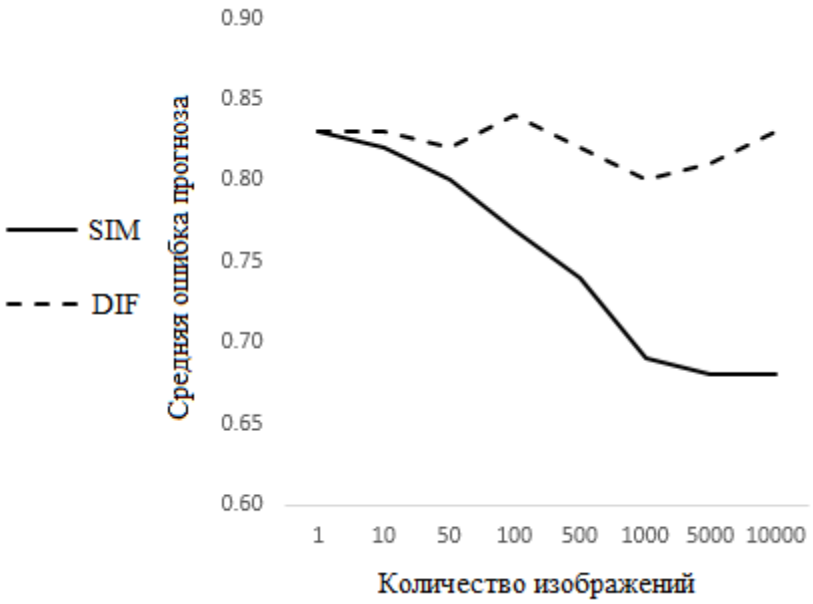

Рисунок 33 - Влияние накопления статистики на точность прогноза пикселей

Из кривых на графике на рисунке 33 видно, что накопление статистики анализатора способствует снижению средней ошибки прогноза для выборки схожих изображений и не влияет значительно на ошибку прогноза для выборки несхожих изображений DIF.

Таким образом, использование алгоритма накопления статистики анализатора может быть применено для снижения ошибки прогноза пикселя в ходе анализа методом WS, и, как следствие, для увеличения эффективности ВВС [74, 78, 80].

Предложенный алгоритм включает в себя следующие шаги (помимо шага создания матрицы кортежей, включённого в алгоритм прогноза по кортежам).

- Определение схожести матриц. В предельном случае уникального кортежа для каждого пикселя фоновой зоны анализируемого изображения, сложность оценивается не выше, чем $O(n)$.
- Слияние матриц. Сложность оценивается как $O(1)$.

Таким образом, итоговая сложность алгоритма оценивается как $O(n)$.

Количественные оценки и наглядная демонстрация степени увеличения эффективности ВВС методом WS при применении предложенных методов приведены в Главе IV.

3.5. Выводы

1. Эффективность ВВС методом WS напрямую зависит от точности прогноза значения пикселя по доступной информации анализируемого изображения.
2. Низкая точность прогноза в фоновых зонах анализируемого изображения ведёт к системному завышению оцененной длины встроенного сообщения по сравнению с реальным, и, как следствие, к повышению доли ложноположительной классификации, снижая итоговую эффективность анализа методом WS. Это позволяет использовать изображения с высокой долей однородного фона как основу для атаки на алгоритм WS.
3. В фоновой зоне изображения возможно использование более сложных структур пикселей для предсказания данного, чем в оригинальном алгоритме WS. Присутствие в фоновых зонах повторяющихся кортежей пикселей и особенности распределения значений в градиентных областях в фоновой зоне изображения позволяют выявить системные закономерности связи соседних пикселей в фоновых зонах изображения, что может быть использовано для снижения ошибки прогноза пикселей в фоновых зонах изображений.

4. Использование алгоритма прогноза пикселей изображения по кортежам и алгоритма адаптивного прогноза в градиентах позволяет снизить ошибку прогноза в фоновых зонах.
5. Накопление статистики анализатора применимо для увеличения точности прогноза по кортежам и неприменимо для улучшения точности адаптивного предсказания в градиентах.
6. Предложенные алгоритмы прогноза вводят в метод ВВС ряд параметров и пороговых значений, для которых необходим подбор оптимальных значений.
7. Подтверждено снижение средней ошибки прогноза значения пикселя анализируемого изображения при использовании предложенных алгоритмов прогноза значения пикселя. Для вывода о применимости предложенных методов в задаче противодействия каналам передачи информации на основе методов стеганографии требуется оценка влияния факта применения предложенных методов на фактическую эффективность ВВС методом WS.

ГЛАВА 4. МЕТОД ВЫЯВЛЕНИЯ ВСТРОЕННЫХ СООБЩЕНИЙ И ОЦЕНКА ЕГО ЭФФЕКТИВНОСТИ

4.1. Метод выявления встроенных сообщений с повышенной эффективностью

4.1.1. Алгоритм выделения однородного фона изображения

Метод выделения однородного фона, приведённый в главах II, III обладает определённой эффективностью выделения зон, представляющих трудности при ВВС и вносящих вклад в повышение доли ложноположительных классификаций при анализе изображения методом WS. Тем не менее, даже простым взглядом на результат выделения фона изображения (рисунок 34) несложно оценить, что выделение однородного фона может быть выполнено качественнее.

а б

Рисунок 34 - Оригинальное изображение (а) и изображение с выделенным фоном (б)

Особенности распределения яркости пикселей в фоновых изображениях, такие, как крупномасштабное изменение средней яркости пикселей, не позволяет выполнять выделение фона описанным выше методом достаточно эффективно.

В качестве альтернативного метода выделения фоновых зон предлагается использовать сегментационную нейронную сеть, описанную в [81]. Сегментационные нейронные сети применяются для выделения объектов из определённого множества на изображениях. Проблема обучения сети на конкретном множестве объектов классификации имеет место, однако, фоновая зона – достаточно универсальный объект, который может быть классифицирован на изображениях различной внутренней семантики.

Рисунок 35 демонстрирует результат выделения однородного фона сегментационной сетью, обученной на наборе 500 изображений, набранных из исходной тестовой выборки.

Рисунок 35 - Выделение однородного фона сегментационной нейронной сетью

На рисунке 35 серым выделен однородный фон, чёрным – прочие зоны изображений. В рамках реализации методов повышения эффективности ВВС,

предложенных выше, результатом работы сети является бинарная классификация каждого пикселя каждого цветового слоя анализируемого изображения.

Метрика определена исходя из предполагаемого способа последующего использования результатов выделения фона. Первоначальная решаемая проблема – некорректный прогноз пикселей изображения в фоновых зонах – сводится к эффекту системного завышения прогнозируемого значения по сравнению с реальным, такого, что эффект завышения сравним по порядку величины с эффектом наличия стеганографического встраивания.

Исходя из определения рассматриваемого метода стеганографического встраивания в плоскость LSB изображения, для оригинального метода прогноза значений пикселей, предложенного авторами метода WS, можно выделить множество пикселей, для которых результат прогноза и ожидаемый эффект от стеганографического встраивания схожи.

Пусть I – множество пикселей анализируемого изображения, $B(I)$ – множество фоновых пикселей изображения (выделенных определённом методом), $i \in I$ – пиксель изображения, $p(i)$ – результат прогноза пикселя оригинальным методом WS. Тогда $P(I)$ – множество спрогнозированных пикселей изображения, а $P_S(I)$ – множество таких пикселей, для которых результат прогноза схож с ожидаемым эффектом стеганографического встраивания.

Исходя из определения метода встраивания в плоскость LSB и особенностей округления значений при анализе методом WS, схожесть результата прогноза с эффектом встраивания для пикселя i определяется так:

$$|i - p(i)| < 1.5$$

Метод выделения фона должен помечать как фоновые, для последующего улучшенного анализа, пиксели множества $P_S(I)$. В таком случае, множество пикселей, корректно классифицированных, как фон, определяется следующим образом:

$$B_C = B(I) \cap P_S(I)$$

Множество пикселей B_{FP} – пикселей, не принадлежащих $P_S(I)$, классифицированных, как фон, называется множеством ложноположительной фоновой классификации. Множество пикселей B_{FN}, принадлежащих $P_S(I)$, не классифицированных, как фон, называется множеством ложноотрицательной фоновой классификации.

В таком случае, численную метрику эффективности выделения фоновой зоны в задаче последующего анализа изображения методом WS можно определить как отношение в процентах мощности множества корректной и суммы мощностей множеств всех фоновых классификаций:

$$A_B = \frac{|B_C|}{|B_C| + |B_{FP}| + |B_{FN}|} * 100\%$$

Проведённый эксперимент оценивает эффективность выделения фона для изображений тестовой выборки и усредняет результат для выборки. Для метода выделения фона, приведённого в главе II, показатель эффективности составляет 73%. Для метода выделения фона сегментационной сетью, описанным выше, показатель эффективности составляет 92%.

Таким образом, использование сегментационной нейронной сети для выделения фоновых зон позволяет повысить эффективность выделения фона и, следовательно, эффективность применения улучшенных методов прогноза значений пикселей в фоновой зоне [78].

4.1.2. Рекомендованные значения настроек

Методы увеличения эффективности ВВС, предложенные в Главе III, требуют предоставления настроек пороговых величин для реализации. Таблица 6 содержит рекомендуемые значения настроек и пороговых величин, полученные

опытным путём. Использование настроек в указанных границах позволяет добиться максимального среднего увеличения эффективности ВВС за счёт использования предложенных методов.

Таблица 6

Рекомендуемые значения настроек

Настройка	**Значение**
Длина кортежа n	$3-5$
Нормированный порог мощности кортежа $T, \%$	$0.1-0.4$
Минимальная длина градиентного кортежа P	6
Максимальная сила градиента k	1.1
Минимальная ширина градиента N	4
Допуск поиска в градиенте R	10
Вес совместного учёта кортежей w	0.6
Отношение схожести элементов матрицы кортежей S_T	0.88
Предельный порог схожести матриц кортежей S	0.72

4.1.3. Алгоритм прогноза значений пикселей в фоновой зоне

Предложенные алгоритмы увеличения эффективности анализа методом WS предполагают значительные изменения алгоритма проведения анализа методом WS. Рисунок 36 содержит блок-схему предлагаемого алгоритма анализа с учётом всех предложенных методом повышения эффективности.

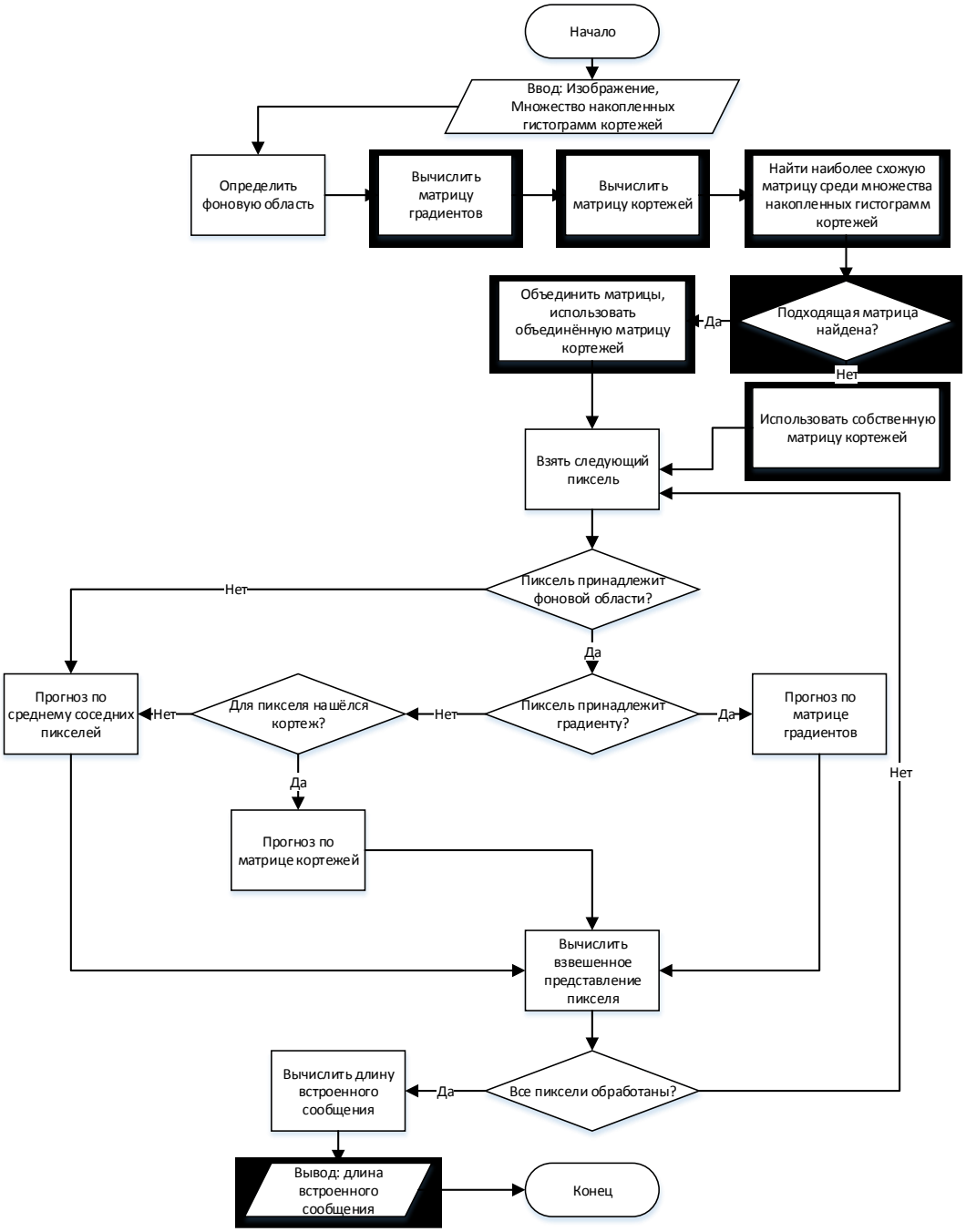

Рисунок 36 - Блок-схема алгоритма анализа методом WS с учётом предложенных методов повышения эффективности

4.2. Архитектура прототипа программной системы-реализации

Реализация разработанного метода представляет собой встраиваемый программный модуль на языке Java. Поскольку методы ВВС встроенных

сообщений, как правило, выступают составными частями более сложных систем защиты информации, модуль не располагает собственным интерфейсом пользователя, взамен определяя универсальный интерфейс данных для ввода тестируемых изображений и вывода результата. Формат поступающих данных стандартный для инфраструктуры программ на языке Java. Рисунок 37 показывает крупномасштабную архитектуру приложения, а рисунок 38 – архитектуру классов модуля со стереотипами их взаимодействия.

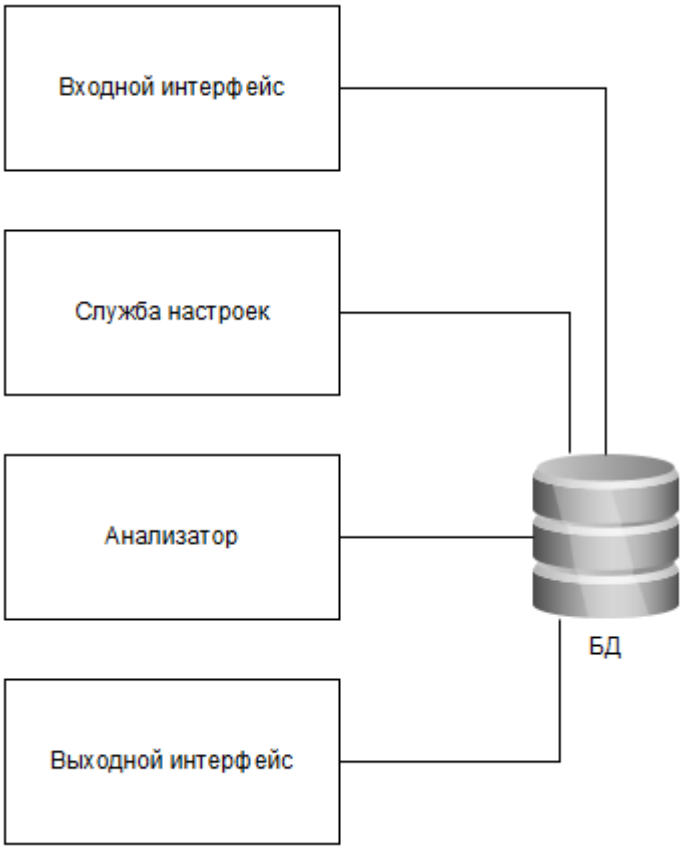

Рисунок 37 – Крупномасштабная архитектура программного модуля

Программный модуль рассчитан на асинхронное взаимодействие с внешней вызывающей системой.

Входной интерфейс принимает изображения для анализа и передаёт их анализатору, либо записывает в базу данных модуля в случае, если анализатор занят.

Анализатор содержит реализацию метода ВВС встроенных сообщений, принимает изображения на анализ от входного интерфейса либо самостоятельно забирает сообщения из базы данных по таймеру, производит анализ и записывает результат в базу данных.

Выходной интерфейс принимает запросы от внешнего приложения, обращается к базе данных, забирая оттуда результаты анализа, и возвращает их вызывающему приложению.

Служба настроек содержит интерфейс для задания настроек системы (пороговых значений и настроек метода ВВС, представленных в таблице 6) внешним приложением, записывает их в базу данных и предоставляет по запросу Анализатора.

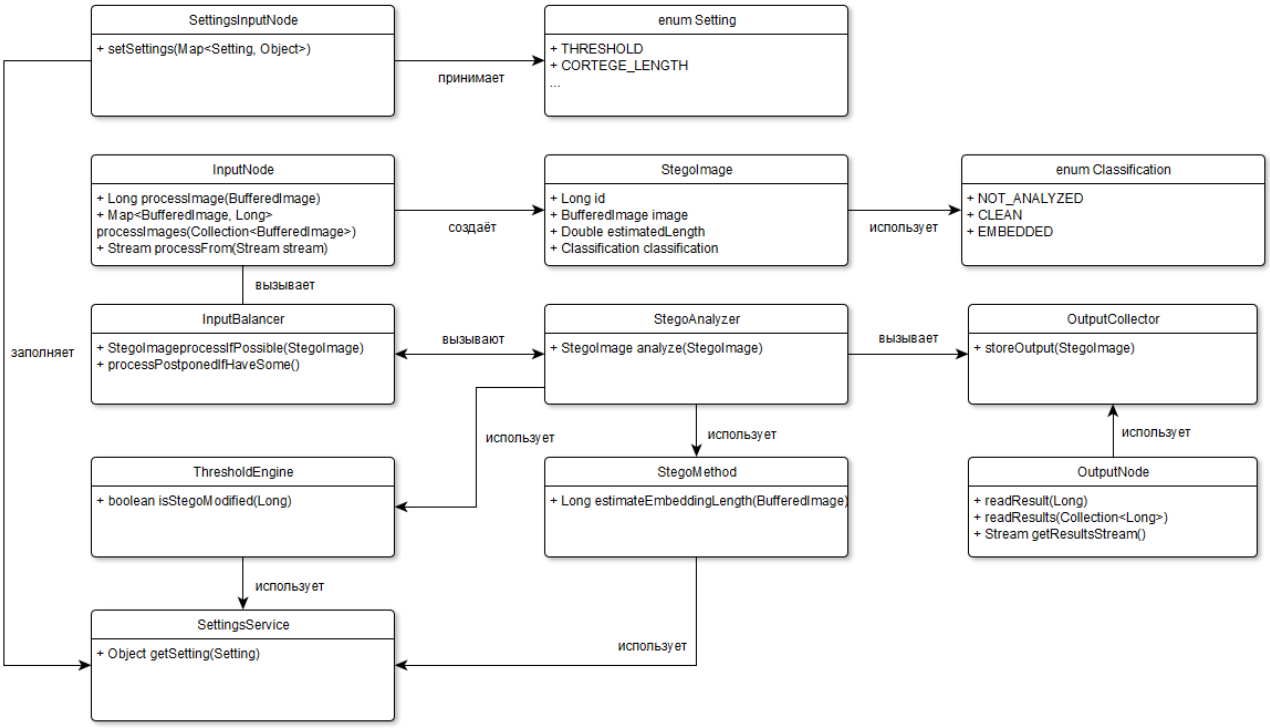

Рисунок 38 – Архитектура классов программного модуля

4.3. Экспериментальная оценка эффективности предложенного метода

4.3.1. Количественная оценка эффективности предложенного метода

В сложившейся практике оценки эффективности методов встраивания и ВВС, количественные показатели эффективности получаются путём анализа работы метода, лежащего в основе той или иной реализации, на репрезентативной выборке [83]. В условиях высокой доступности реализаций рассматриваемых методов ВВС в чистом виде (без программных оболочек, предоставляющих пользовательский интерфейс), а также в условиях частого использования реализаций методов стеганоанализа в качестве составных частей более крупных систем, оценка проводится непосредственно с реализациями методов, а не с программными системами, включающими реализации в качестве составных частей [84, 85, 86].

Экспериментальная оценка эффективности предложенного метода получена способом, описанным в главе I. Репрезентативная выборка описана также в главе I. Реализация метода предложена автором и приведена в [87]. Реализация разработанного метода соответствует архитектуре, приведённой в пункте 4.2.

Кривые доверительных интервалов построены на основе анализа исходного тестового множества, содержащего изображения с различной долей однородного фона.

На графике на рисунке 39 приведены кривые доверительных интервалов, иллюстрирующие повышение эффективности ВВС методом WS за счёт применения предложенных алгоритмов для различных значений полезной нагрузки в рассматриваемом диапазоне.

На графике пунктирные линии соответствуют эффективности анализа оригинальным методом WS, сплошные – улучшенным методом WS. Видно, что вне зависимости от значения полезной нагрузки, при применении предложенных методов эффективность анализа возрастает.

Таблица 7 содержит численные оценки прироста эффективности ВВС при использовании алгоритма предсказания по кортежам. Прирост эффективности оценен как падение доли ложной классификации при заданной доле корректной классификации (95%). Положительное число соответствует уменьшению вероятности некорректной классификации.

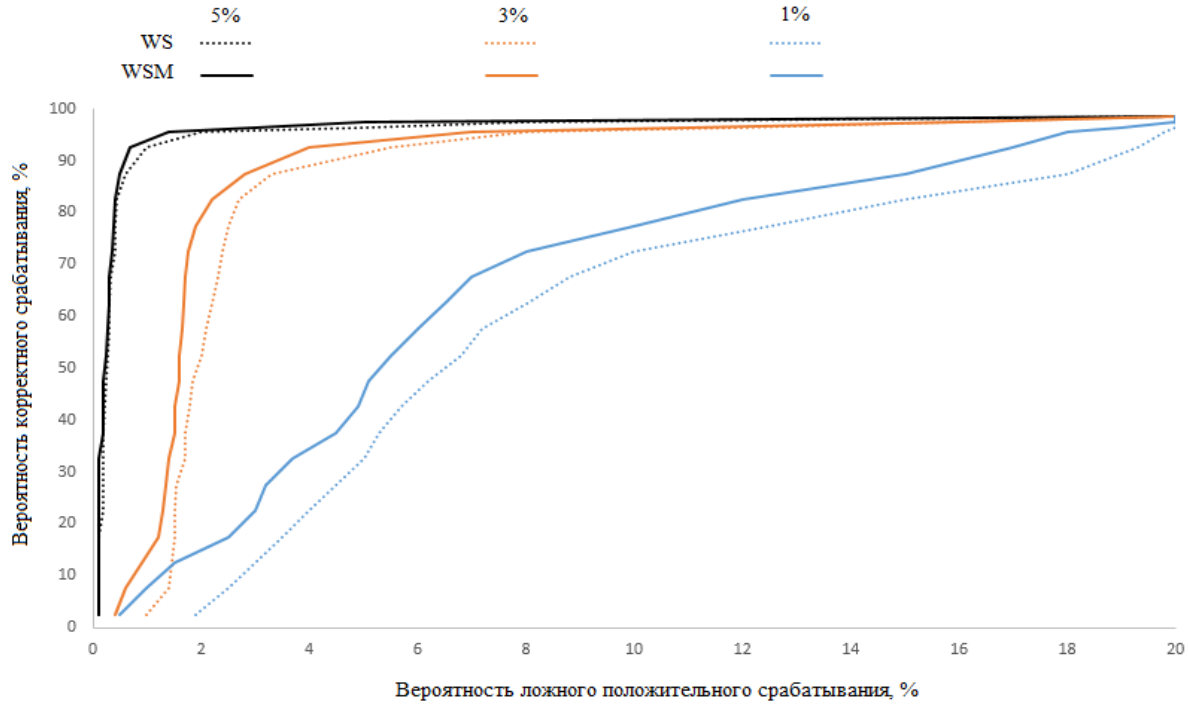

Рисунок 39 - Кривые доверительных интервалов метода WS и улучшенного метода

Таблица 7

Увеличение эффективности ВВС при использовании прогноза по кортежам

ОНК/Разрешение	**800x800**	**1500x1500**	**2400x2400**	**В среднем**
1%	9,75%	9,1%	6,95%	8,6%
2%	8,65%	7,6%	5,75%	7,3%
3%	5,6%	5,45%	5%	5,4%
4%	4,1%	4,05%	4,0%	4,03%
5%	3,8%	3,7%	3,3%	3,6%

Таблица 8 содержит численные оценки прироста эффективности ВВС при использовании метода адаптивного предсказания в градиентах.

Таблица 8

Увеличение эффективности ВВС при использовании адаптивного прогноза в градиентах

ОНК/Разрешение	**800x800**	**1500x1500**	**2400x2400**	**В среднем**
1%	5,1%	5,1%	5,2%	5,1%
2%	4,45%	4,6%	4,9%	4,65%
3%	4,0%	4,4%	4,9%	3,1%
4%	2,3%	2,8%	3,3%	2,8%
5%	2,2%	2,6%	3,0%	2,6%

Таблица 9 содержит численные оценки прироста эффективности ВВС при использовании предложенного метода. Поскольку алгоритм прогноза значения пикселя за счёт накопления статистики подразумевает проявление эффекта при последовательном анализе множества изображений, таблица содержит результат для различных значений количества проанализированных ранее изображений, а также результат усреднения.

Таблица 9

Увеличение эффективности ВВС при совместном использовании методов

ОНК/Размер выборки	**1**	**1000**	**10000**	**100000**	**В среднем**
1%	9,5%	9,6%	9,6%	9,7%	9,6%
2%	9,1%	9,1%	9,3%	9,3%	9,2%
3%	8,6%	8,6%	8,7%	8,7%	8,65%
4%	6,9%	7,0%	7,1%	7,1%	7,0%
5%	6,3%	6,3%	6,5%	6,7%	6,45%

Из таблицы 8 видно, что увеличение эффективности в зависимости от объёма встраивания и размера выборки составляет от 6,3% до 9,7%. Ни на одном значении полезной нагрузки из рассматриваемых не получено итоговое падение

эффективности. В среднем увеличение эффективности составляет 8,1% на рассматриваемом диапазоне значений полезной нагрузки [74, 78, 80].

4.3.2. Качественная оценка новизны функционала по сравнению с существующими решениями

Известной проблемой качественной оценки разработанного функционала по выявлению встроенных сообщений является отсутствие стандартов и рекомендаций по такому функционалу в нормативно-правовой базе РФ. Стеганография и стеганоанализ, к области которых обычно относят рассматриваемый функционал, не упоминаются в нормативно-правовом поле России [88]. Таким образом, невозможно однозначно сделать вывод о соответствии разработанного функционала тем или иным стандартным требованиям либо о возможности внедрения такого функционала в качестве стандартных составных частей систем, оформленных по подобным требованиям [89].

Единственным способом нормативного регулирования рассматриваемого функционала являются патенты. Патенты оформляются на средства выявления встроенных сообщений, которые выступают в качестве основы для реализаций программных комплексов защиты информации [90]. Поскольку именно методы выявления встроенных сообщений являются предметом исследования данной работы, и они же составляют основу программных реализаций для выявления встроенных сообщений, сравнительный анализ предложенного метода с существующими запатентованными решениями даёт объективную оценку качественной новизны предложенного функционала.

Патентный поиск произведён по следующим критериям:

— статистический стеганоанализ в частотной области представления изображения;
— коэффициентный стеганоанализ изображения;

— наличие строгого математического обоснования метода анализа и отбираемых для анализа характеристик изображения;

Следующие основные качественные характеристики разработанного метода используются для сравнения с существующими решениями:

— универсальность по формату представления анализируемого изображения для хранения;
— узкая ориентированность на выявление встроенных сообщений в НЗБ неподвижных цифровых изображений;
— применимость как к полноцветным изображениям, так и к изображениям в оттенках серого;
— использование статистического стеганоанализа;
— доступность выходных данных метода для непосредственной бинарной классификации.

В результате поиска отобраны 12 патентов не старше 10 лет относительно момента проведения поиска (2018 год). Таблица Б1 приложения Б содержит сводную таблицу патентов. Ниже приведены краткие характеристики исследованных патентов с акцентом на отличиях и сходствах с разработанным решением.

1. Патент CN 104021227 от 03.09.2014 «Digital forensics-oriented anomaly steganalysis method and system» [91].

Выделены следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— отсутствие узкой специализации на анализе цифровых изображений;
— использование гибридного алгоритма, алгоритм статистического стеганоанализа используется только в качестве вспомогательного к алгоритму элементного анализа;
— отсутствие однозначного вывода о классификации тестируемого изображения – генерируемый системой отчёт подлежит дополнительной интерпретации человеком или третьей системой.

2. Патент CN 104008521 от 27.08.2014 «LSB replacement steganalysis method based on grey co-occurrence matrix statistic features» [92].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— строгая привязанность к анализируемому методу встраивания, не позволяющая эффективно проводить стеганоанализ встраивания с использованием адаптивных методов анализа; возможность анализа только усреднённой цветовой плоскости изображения (либо изображений в оттенках серого);
— отсутствие математического обоснования выбираемых характеристик для анализа (в ассоциированной статье характеристики описаны как «очевидные»).

3. Патент CN 103971321 от 06.08.2014 «Method and system for steganalysis of JPEG compatibility» [93].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— строгая привязанность к формату анализируемого изображения: исходя из описания алгоритма, он оперирует коэффициентами таблицы квантизации формата JPEG, что ограничивает его принципиальную применимость этим форматом. Несмотря на то, что применяемые в описываемом решении методы в теории обобщаются на любые выборки коэффициентов, в том числе и на коэффициенты пространственной области, не удалось найти родственных патентов, применяющих подобное обобщение.

4. Патент CN 103745479 от 23.04.2014 «Digital steganography and steganalysis method for color image» [94].

Отличия решаемой патентом задачи от проверяемого решения исходят из особенностей алгоритма WS [4]:

— применимость только к цветным изображениям, основанная на необходимости совокупного анализа всех цветовых плоскостей изображения, иначе – значительная потеря в точности анализа;
— отсутствие непосредственно бинарной классификации как окончательного результата работы алгоритма: представляемая в качестве результата количественная оценка мощности встраивания подлежит дополнительной интерпретации с использованием допустимых порогов, которые нужно определять отдельно.

5. Патент CN 103310235 от 18.09.2013 «Steganalysis method based on parameter identification and estimation» [95].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— отсутствие узкой специализации на анализе цифровых изображений, неизбежно ведущее к потере преимуществ узкоспециализированных алгоритмов;
— необходимость подачи на вход алгоритму готовой базы знаний методов встраивания.

6. Патент US 2013208941 от 15.08.2013 «Steganalysis with neighborhood joint density» [96].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— несмотря на заявленное отсутствие строгой привязанности к формату анализируемого изображения, для выполнения процедуры анализа необходим его перевод в формат JPEG для получения матриц ДКП-коэффициентов. Это влечёт за собой неизбежное размывание оригинальных характеристик высшего порядка изначального изображения, что отрицательно сказывается на точности стеганоанализа. Данные о точности

стеганоанализа, приведённые в описании, получены при анализе только изображений, встраивание в которые изначально производилось в формат JPEG. Таким образом, алгоритм не полностью универсален относительно формата анализируемого изображения;

— алгоритм разработан для детектирования встраивания только с использованием методов семейства YASS. Несмотря на то, что методы данного семейства относятся к основным на сегодняшний день, описываемое решение не обладает универсальностью относительно метода встраивания;

7. Патент CN 103034853 от 10.04.2013 «Universal steganalysis method for JPEG images» [97].

Отличия от проверяемой задачи основываются на особенностях алгоритма SPAM и характере вносимых авторами метода вариаций [6]:

— практически полная неэффективность алгоритма анализа при применении адаптивных методов встраивания;
— отсутствие строгого математического обоснования процедуры выбора характеристик высшего порядка изображения, выступающих в роли элементов опорных векторов в процедуре машинного обучения;
— созданная строгая привязанность к формату JPEG, основанная на необходимости использования восстановленных коэффициентов пространственной области, полученных путём обратного ДКП коэффициентов изображения в формате JPEG.

8. Патент CN 102411771 от 11.04.2012 «Reversible image steganalysis method based on histogram peak value fluctuation quantity» [98].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— метод не нацелен непосредственно на анализ встраивания в пространственном домене, в основном работая с методами встраивания в частотную область, которые наибольшим образом влияют на итоговую гладкость гистограммы изображения;
— отсутствует бинарная классификация, получаемые значения предполагаемой мощности встраивания требуют дополнительной интерпретации.

9. Патент CN 102147913 от 10.08.2011 «Steganalysis method based on image smoothness variation characteristics» [99].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— привязанность к конкретным особенностям метода встраивания, описываемое решение не полностью универсально по методу встраивания;
— возможность анализа только изображений в оттенках серого.

10. Патент US 2011135146 от 09.06.2011 « Method and apparatus for steganalysis of texture images» [100].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— метод предназначен для анализа только текстурных, а не любых изображений. Несмотря на то, что авторами заявлена возможность применения решения в качестве универсального стеганоанализатора также для естественных изображений, представленное математическое обоснование эффективности справедливо только для текстурных изображений;
— наличие зависимости от формата изображения. Несмотря на то, что метод в теории подходит для анализа изображений в любом формате, строгая необходимость переведения изображения в формат JPEG создаёт

привязанность к преобразованным данным с затёртыми исходными характеристиками высшего порядка изображения. Таким образом, предлагаемое решение не полностью универсально по формату анализируемого изображения.

11. Патент US 2010091981 от 15.04.2010 «Steganalysis of suspect media» [101].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— отсутствие узкой специализации на анализе цифровых изображений, неизбежно ведущее к потере преимуществ узкоспециализированных алгоритмов;
— установленная неэффективность алгоритмов, основанных на применении функции квантизации к данным и анализе отклонений при использовании адаптивных методов встраивания делает описываемую систему не универсальной по методу встраивания.

12. Патент US 2003026447 от 22.6.2001 «Reliable detection of LSB steganography in color and grayscale images» [102].

Можно выделить следующие отличия решаемой патентом задачи от задачи проверяемого на патентоспособность решения:

— метод не является полностью универсальным по формату анализируемого изображения: несмотря на теоретическую возможность использования для стеганоанализа встраивания в LSB в любые данные изображения, наибольшую эффективность метод показывает только при анализе изображений, изначально образованных в формате с представлением данных в пространственной области.

Таблица 10 содержит сводные данные об отличиях существующих решений от разработанного.

Таблица 10

Отличия разработанного и существующего решений

Патент	Универсальность по формату анализируемого изображения	Ориентированность на область НЗБ естественных изображений	Доступность выходных данных для бинарной классификации	Применимость как к полноцветным, так и к серым изображениям	Наличие математического обоснования выбора характеристик
CN 104021227	Отличие	Отличие	Отличие		
CN 104008521				Отличие	Отличие
CN 103971321	Отличие				
CN 103745479			Отличие	Отличие	
CN 103310235	Отличие	Отличие			
US 2013208941	Отличие				
CN 103034853	Отличие				Отличие
CN 102411771	Отличие	Отличие	Отличие	Отличие	
CN 102147913				Отличие	
US 2011135146	Отличие	Отличие			
US 2010091981	Отличие	Отличие			
US 2003026447	Отличие				

Из анализа данных таблицы 10 видно, что ни одно представленное решение не обладает полной функциональной схожестью с предложенным. Кроме того, следует отдельно отметить, что ни в одном из рассмотренных решений не заявлен функционал BBC на малых значениях отношения нагрузка-контейнер, что является ключевым преимуществом представленного решения.

Таким образом, сделан вывод, что представленный метод обладает качественной новизной предлагаемого функционала по сравнению с существующими решениями.

4.4. Выводы

1. Метод выделения однородного фона, основанный на применении сегментационной нейронной сети, позволяет добиться большей эффективности выделения однородного фона изображения в задаче применения улучшенной модели прогноза пикселя в фоновой зоне.

2. Проведённые эксперименты подтверждают повышение эффективности ВВС предложенным методом по сравнению с методом WS за счёт применения предложенных алгоритмов прогноза пикселей. Минимальный прирост эффективности на тестовой выборке составил 6,4%, средний – 8,1%, максимальный – 9,7%.
3. Предложенные оптимальные значения настроек, используемых в методе, позволяют добиться максимального прироста эффективности при использовании предложенных методов.
4. Предложенный метод обладает качественной новизной функционала по сравнению с существующими решениями.

ЗАКЛЮЧЕНИЕ

Результаты диссертационной работы представляют собой решение актуальной научной задачи повышения эффективности методов выявления встроенных сообщений в неподвижных цифровых изображениях при малых значениях отношения нагрузка-контейнер.

Решённая задача актуальна по причине низкой эффективности существующих методов выявления встроенных сообщений при указанных условиях, а также за счёт постоянного роста объёмов контейнеров встраивания вкупе с известной зависимостью эффективности методов ВВС от значения отношения нагрузка-контейнер.

В ходе выполнения диссертационной работы были решены поставленные частные задачи и получены следующие основные результаты:

1. Проведён анализ существующих методов ВВС в неподвижных цифровых изображениях в НЗБ. Отобраны методы, обладающие наибольшей эффективностью из доступных.
2. Проведён анализ эффективности отобранных методов ВВС, сделан вывод о недостаточной эффективности методов ВВС при малых значениях отношения нагрузка-контейнер. Сделан вывод о том, что метод Weighted Stego Image (WS) обладает наивысшей эффективностью в задаче ВВС в неподвижных цифровых изображениях из рассматриваемых.
3. Разработана и проанализирована модель ВВС методом WS в условиях малых значений полезной нагрузки. Сделан вывод о зависимости эффективности метода ВВС от доли однородного фона в анализируемом изображении.
4. Разработаны алгоритмы ВВС в НЗБ фоновых зон неподвижных изображений при малой полезной нагрузке.

5. Разработан метод ВВС в НЗБ фоновых зон неподвижных цифровых изображений с повышенной точностью при малой полезной нагрузке.

Таким образом, был разработан метод выявления встроенных сообщений в плоскость наименьшего значащего бита неподвижного цифрового изображения на базе метода ВВС WS, обладающий, по сравнению с аналогами, рассмотренными в главе I, повышенной эффективностью при применении в условиях малых значений отношения нагрузка-контейнер за счёт использования более сложных структур анализируемого изображения и накопления статистики анализатора в ходе работы.

Сформулированы **рекомендации** по применению результатов работы в различных прикладных областях. Представленные результаты дают инструмент для построения систем противодействия каналам скрытной передачи данных на основе встраивания. Разработанные алгоритмы и метод рекомендуется применять при противодействии встраиванию в НЗБ неподвижных изображений в ситуациях, когда:

— исходя из доступных потенциальному нарушителю данных, ожидается использование естественных изображений в качестве контейнеров для встраивания;
— объём доступных нарушителю данных позволяет организовать канал передачи данных на малых значениях отношения нагрузка-контейнер;

Разработанную модель выявления встроенных сообщений в фоновых зонах рекомендуется использовать в дальнейших исследованиях в области выявления встроенных сообщений в неподвижных изображениях с большой долей однородного фона.

Приведённые рекомендованные значения настроек разработанного метода ВВС позволяют организовать противодействие скрытному каналу передачи данных средством на основе метода наиболее эффективным способом.

В качестве **направлений дальнейших исследований** можно выделить исследования, связанные с:

- дальнейшим анализом модели выявления встроенных сообщений в фоновых зонах неподвижных изображений, с выделением более глубоких особенностей распределения значений пикселей в таких областях;
- анализом и выделением более сложных структур пикселей, прилегающих к анализируемому пикселю в фоновых зонах, позволяющих добиться большей точности прогноза пикселей;
- поиском, выделением и применением в задаче прогноза значений пикселей других крупномасштабных структур в фоновых зонах, кроме градиентов;
- анализом возможных средств адаптации методов встраивания в неподвижные изображения к разработанному методу, разработкой контрмер.

Положения, выносимые на защиту, **соотнесены с пунктами паспорта специальности** 05.13.19 – «Методы и системы защиты информации, информационная безопасность»: «5. Методы и средства (комплексы средств) информационного противодействия угрозам нарушения информационной безопасности в открытых компьютерных сетях, включая Интернет», «6. Модели и методы формирования комплексов средств противодействия угрозам хищения (разрушения, модификации) информации и нарушения информационной безопасности для различного вида объектов защиты вне зависимости от области их функционирования».

СПИСОК ЛИТЕРАТУРЫ

1. Steganography: A Powerful Tool for Terrorists and Corporate Spies // Stratfor [Электронный ресурс]. Режим доступа: https://www.stratfor.com/analysis/steganography-powerful-tool-terrorists-and-corporate-spies, свободный. Яз. англ. (дата обращения 22.08.2017).
2. Elżbieta Zielińska, Wojciech Mazurczyk, and Krzysztof Szczypiorski. 2014. Trends in steganography. Commun. ACM 57, 3 (March 2014), 86-95. DOI: https://doi.org/10.1145/2566590.2566610
3. Gayathri C., Kalpana V. Study on image steganography techniques // International Journal of Engineering and Technology (IJET). 2013. V. 5. P. 572–577.
4. Sharma V. K., Srivastava D. K., Mathur P. A Study of Steganography Based Data Hiding Techniques. – 2017.
5. Bachrach M., Shih F. Y. 11 Survey of Image Steganography and Steganalysis //Multimedia Security: Watermarking, Steganography, and Forensics. – 2017. – С. 201.
6. Patel A., Patel M. A Study of Different Steganalysis Methods. // International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 3, issue 2 – 2018. P. 118 – 123.
7. Hussain M. et al. Image steganography in spatial domain: a survey //Signal Processing: Image Communication. – 2018. – Т. 65. – С. 46-66.
8. Jindal S., Kaur N. Digital image steganography survey and analysis of current methods //International Journal of Computer Science and Information Technology & Security. – 2016. – Т. 6.
9. Prokhozhev N. et al. Passive Steganalysis Evaluation: Reliabilities of Modern Quantitative Steganalysis Algorithms //Proceedings of the First International Scientific Conference “Intelligent Information Technologies for Industry”(IITI’16). – Springer, Cham, 2016. – С. 89-94.

10. Буценко Ю. П., Савченко Ю. Г. Существует ли идеальное стеганографическое сокрытие информационного обмена?. – 2017.
11. Прохожев Н.Н., Михайличенко О.В., Башмаков Д.А., Сивачев А.В., Коробейников А.Г. Исследование эффективности применения статистических алгоритмов количественного стеганодетектирования в задаче детектирования скрытых каналов передачи информации // Программные системы и вычислительные методы. 2015. № 3. С. 281–292. doi: 10.7256/2305-6061.2015.3.17233
12. Шагрова Г. В., Садома Ю. В. Проблемы детектирования скрытой информации в цифровых изображениях и способы их решения //Инновационное развитие современной науки: проблемы, закономерности, перспективы. – 2018. – С. 62-64.
13. Волынкин П. А., Севостьянова А. С. Особенности стеганографии мультимедиа контента //European Scientific Conference. – 2018. – С. 99-101.
14. Patil A. et al. Survey on Recent Steganography Approaches. // International Journal of Innovative Research in Computer and Communication Engineering – 2017. P. 46 – 49.
15. Subhedar M. S., Mankar V. H. Current status and key issues in image steganography: A survey //Computer science review. – 2014. – Т. 13. – С. 95-113.
16. Laishram D., Tuithung T. A Survey on Digital Image Steganography: Current Trends and Challenges. // Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT) – 2018.
17. Roy C. Y., Goel M. K. Review on Image Steganography //Indian Journal of Science and Technology. – 2016. – Т. 9. – №. 47.
18. Kaur H., Kakkar A. Comparison of different image formats using LSB Steganography //Signal Processing, Computing and Control (ISPCC), 2017 4th International Conference on. – IEEE, 2017. – С. 97-101.

19. Voloshina N. et al. Effectiveness of LSB and MLSB information embedding for BMP images //Proceedings of the 18th Conference of Open Innovations Association FRUCT. – FRUCT Oy, 2016. – С. 378-384.
20. Singh A., Singh H. An improved LSB based image steganography technique for RGB images //Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on. – IEEE, 2015. – С. 1-4.
21. Li X. et al. A novel blind detector for additive noise steganography in JPEG decompressed images //Multimedia tools and applications. – 2014. – Т. 68. – №. 3. – С. 1051-1068.
22. Huang W., Zhao X. Novel cover selection criterion for spatial steganography using linear pixel prediction error //Science China Information Sciences. – 2016. – Т. 59. – №. 5. – С. 059103:1-059103:3.
23. Hu Y. C., Lee C. F., Liu Y. H. Reversible Image Steganography for Color Image Quantization Based on Lossless Index Coding //International Conference on Security with Intelligent Computing and Big-data Services. – Springer, Cham, 2017. – С. 185-195.
24. Никишова А. В., Македонский С. А. Модель оценки качества стеганографических систем //Промышленные АСУ и контроллеры. – 2017. – №. 7. – С. 37-43.
25. Desai M. B., Patel S. V. Performance analysis of image steganalysis against message size, message type and classification methods //Advances in Electronics, Communication and Computer Technology (ICAECCT), 2016 IEEE International Conference on. – IEEE, 2016. – С. 295-302.
26. Коржик В. И., Курбатов Е. В. Атака на систему цифровых водяных знаков с использованием методов статистического оценивания //Вопросы защиты информации. – 2005. – №. 2. – С. 14-20.
27. Amsaveni A., Vanathi P. T. A comprehensive study on image steganography and steganalysis techniques //International Journal of Information and Communication Technology. – 2015. – Т. 7. – №. 4-5. – С. 406-424.

28. A.D. Ker: A general framework for structural steganalysis of LSB Replacement, Proc. of the Information Hiding, pp.296-311, 2005.
29. М.А. Дрюченко. Алгоритмы выявления стеганографического скрытия информации в JPEG-файлах. // Вестник Воронежского государственного университета, №1, 2007. С. 21-30.
30. J.Fridrich, M.Goljan, R.Du Reliable Detection of LSB Steganography in Color and Grayscale Images, State Univ. of New York, Binghamton, NY, USA.
31. Lu, P., X. Luo et. al., An improved sample pairs method for detection of LSB embedding, Proc. of the 6th Information Hiding Workshop, Springer LNCS, vol.3200, pp.116-128, 2004
32. Mao Ye, Fenlin Liu, Chunfang Yang, Xiongfei He Steganalysis Based on Weighted Stego-Image for LSB Replacement Steganography. Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP '09. pp. 945-948
33. Xia Z. et al. Steganalysis of LSB matching using differences between nonadjacent pixels //Multimedia Tools and Applications. – 2016. – Т. 75. – №. 4. – С. 1947-1962.
34. Zhang H. et al. Steganalysis by subtractive pixel adjacency matrix and dimensionality reduction //Science China Information Sciences. – 2014. – Т. 57. – №. 4. – С. 1-7.
35. Zhang J., Cox I. J., Doerr G. Steganalysis for LSB matching in images with high-frequency noise //Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on. – IEEE, 2007. – С. 385-388.
36. He J., Huang J. Steganalysis of stochastic modulation steganography //Science in China Series F. – 2006. – Т. 49. – №. 3. – С. 273-285.
37. Niimi M. et al. Intensity histogram steganalysis in BPCS-steganography //Security and Watermarking of Multimedia Contents III. – International Society for Optics and Photonics, 2001. – Т. 4314. – С. 555-565.

38. Xia Z. et al. Steganalysis of least significant bit matching using multi-order differences //Security and Communication Networks. – 2014. – Т. 7. – №. 8. – С. 1283-1291.
39. Zhang, T. and X. Ping, Reliable detection of LSB steganography based on the difference image histogram, Proc. of the IEEE ICSAAP 2003, Part III, pp. 545-548, 2003.
40. Fridrich J. et al. Quantitative steganalysis of digital images: estimating the secret message length //Multimedia systems. – 2003. – Т. 9. – №. 3. – С. 288-302.
41. Nissar A., Mir A. H. Classification of steganalysis techniques: A study //Digital Signal Processing. – 2010. – Т. 20. – №. 6. – С. 1758-1770.
42. Евсютин О. О., Мещеряков Р. В., Шумская О. О. Стегоанализ цифровых изображений с использованием наивного байесовского классификатора //ДЕСЯТАЯ ВСЕРОССИЙСКАЯ МУЛЬТИКОНФЕРЕНЦИЯ ПО ПРОБЛЕМАМ УПРАВЛЕНИЯ МКПУ-2017. – 2017. – С. 56-58.
43. Kharrazi M., Sencar H. T., Memon N. Benchmarking steganographic and steganalysis techniques //Security, Steganography, and Watermarking of Multimedia Contents VII. – International Society for Optics and Photonics, 2005. – Т. 5681. – С. 252-264.
44. Chandramouli R., Memon N. D. Steganography capacity: A steganalysis perspective //Security and Watermarking of Multimedia Contents V. – International Society for Optics and Photonics, 2003. – Т. 5020. – С. 173-178.
45. Provos N., Honeyman P. Hide and seek: An introduction to steganography //IEEE security & privacy. – 2003. – Т. 99. – №. 3. – С. 32-44.
46. Schaathun H. G. Machine learning in image steganalysis. – Wiley, 2012.
47. Break Our Watermarking System (BOWS) image database. - http://bows2.ec-lille.fr
48. Break Our Steganographic System (BOSS) image database. - http://agents.fel.cvut.cz/boss/index.php?mode=VIEW&tmpl=materials
49. E-Trim Image database. - http://www.ipb.uni-bonn.de/projects/etrims_db/
50. Places image database. - http://places2.csail.mit.edu/

51. Gerald Schaefer, Michal Stich, "UCID: an uncompressed color image database", Proc. SPIE 5307, Storage and Retrieval Methods and Applications for Multimedia 2004, (18 December 2003); doi: 10.1117/12.525375; https://doi.org/10.1117/12.525375
52. Juarez-Sandoval O. et al. Compact image steganalysis for LSB-matching steganography //Biometrics and Forensics (IWBF), 2017 5th International Workshop on. – IEEE, 2017. – С. 1-6.
53. F. Korč, W. Förstner. eTRIMS Image Database for Interpreting Images of Man-Made Scenes. Technical report TR-IGG-P-2009-01, University of Bonn, Dept. of Photogrammetry, 2009.
54. Places: A 10 million Image Database for Scene Recognition. B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017
55. Степанов, Евгений Анатольевич, and И. К. Корнеев. "Информационная безопасность и защита информации." М.: ИНФРА-М (2001).
56. Сивачев А. В. и др. Эффективность стеганодетектирования на основе методов машинного обучения //Вопросы кибербезопасности. – 2017. – №. 2 (20).
57. Wang D. C. C., Vagnucci A. H., Li C. C. Digital image enhancement: a survey //Computer Vision, Graphics, and Image Processing. – 1983. – Т. 24. – №. 3. – С. 363-381.
58. Koenderink J. J. Color for the Sciences. – The MIT Press, 2010.
59. Кустов В. Н., Процко Д. К. Программная модель стеганографа на основе модификации метода замены наименее значащих битов //Вестник научных конференций. – ООО Консалтинговая компания Юком, 2017. – №. 2-3. – С. 54-61.
60. Казьмин Д. А., Цесарь А. Л. Исследование стеганографического метода замены младшего бита //Современные концепции развития науки. – 2017. – С. 53-55.

61. Johnson N. F., Jajodia S. Exploring steganography: Seeing the unseen //Computer. – 1998. – Т. 31. – №. 2.
62. Буханцов А. Д. и др. Исследование алгоритмов скрытного внедрения информации в пространственные компоненты монохромного изображения. – 2017.
63. Neeta D., Snehal K., Jacobs D. Implementation of LSB steganography and its evaluation for various bits //Digital Information Management, 2006 1st International Conference on. – IEEE, 2006. – С. 173-178.
64. Белим С. В., Вильховский Д. Э. Выявление стеганографических вставок типа LSB-замещения в растоpовых изображениях //ББК 22.18 я43+ 32.973. 26–018.1 я43 М340. – 2017. – С. 183.
65. Torralba A., Oliva A. Statistics of natural image categories //Network: computation in neural systems. – 2003. – Т. 14. – №. 3. – С. 391-412.
66. Yudin O., Veselska O. Metods of digital filtration and their impacts on the quality of images of different classes //Наукоємні технології. – 2017. – Т. 36. – №. 4.
67. Chandramouli R., Kharrazi M., Memon N. Image steganography and steganalysis: Concepts and practice //International Workshop on Digital Watermarking. – Springer, Berlin, Heidelberg, 2003. – С. 35-49.
68. Вахаб А., Романенко Д. М. Методы цифровой стеганографии на основе модификации цветовых параметров изображения. – 2018.
69. Тарнополов Р. В. Метод повышения доступности видеопотока для информационных технологий закрытия информации. – 2017.
70. Fridrich J. et al. Steganalysis of content-adaptive steganography in spatial domain //International Workshop on Information Hiding. – Springer, Berlin, Heidelberg, 2011. – С. 102-117.
71. Zhang T., Ping X. A new approach to reliable detection of LSB steganography in natural images //Signal processing. – 2003. – Т. 83. – №. 10. – С. 2085-2093.
72. Ker A. D. A weighted stego image detector for sequential LSB replacement //Information Assurance and Security, 2007. IAS 2007. Third International Symposium on. – IEEE, 2007. – С. 453-456.

73. Chandramouli R., Memon N. Analysis of LSB based image steganography techniques //Image Processing, 2001. Proceedings. 2001 International Conference on. – IEEE, 2001. – Т. 3. – С. 1019-1022.
74. Башмаков Д.А. Точность предсказания пикселей фоновых зон цифровых изображений в задаче стеганодетектирования методом Weighted Stego // Кибернетика и программирование. — 2018. - № 2. - С.38-47. DOI: 10.25136/2306-4196.2018.2.25706. URL: http://e-notabene.ru/kp/article_25706.html
75. Башмаков Д.А., Сивачев А.В. Влияние параметров маски на практическую точность RS-анализа. Сборник трудов IV Всероссийского конгресса молодых ученых (Санкт-Петербург, 7-10 апреля 2015 г.). 2015. С. 49-53
76. Islam M. R. et al. An efficient filtering based approach improving LSB image steganography using status bit along with AES cryptography //Informatics, Electronics & Vision (ICIEV), 2014 International Conference on. – IEEE, 2014. – С. 1-6.
77. Сивачев А.В., Башмаков Д.А. Влияние предварительной обработки изображения - контейнера фильтрами на точность статистического стеганодетектирования. Сборник трудов IV Всероссийского конгресса молодых ученых (Санкт-Петербург, 7-10 апреля 2015 г.). 2015. С. 361-365.
78. Башмаков Д.А. Адаптивное предсказание пикселей пикселей в градиентных областях для улучшения точности стеганодетектирования в неподвижных цифровых изображениях // Кибернетика и программирование. — 2018. - № 2. - С.83-93. DOI: 10.25136/2306-4196.2018.2.25514. URL: http://e-notabene.ru/kp/article_25514.html
79. Andrew D. Ker, Rainer Böhme, "Revisiting weighted stego-image steganalysis", Proc. SPIE 6819, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, 681905 (18 March 2008); doi: 10.1117/12.766820; https://doi.org/10.1117/12.766820
80. Башмаков Д.А., Прохожев Н.Н., Михайличенко О.В., Сивачев А.В. Применение матриц соседства пикселей для улучшения точности стеганодетектирования неподвижных цифровых изображений с

однородным фоном // Кибернетика и программирование. — 2018. - № 1. - С.64-72. DOI: 10.25136/2306-4196.2018.1.24919. URL: http://e-notabene.ru/kp/article_24919.html

81. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – С. 3431-3440.
82. Прохожев Н.Н., Сивачев А.В., Михайличенко О.В., Башмаков Д.А. Повышение точности стеганодетектирования в области ДВП путем использования взаимосвязи между областями двумерного и одномерного разложений. Кибернетика и программирование. 2017. № 2. С. 78-87
83. Ker A. D. The ultimate steganalysis benchmark? //Proceedings of the 9th workshop on Multimedia & security. – ACM, 2007. – С. 141-148.
84. Ma X., Lin J. Research on Efficiency Evaluation for Steganalysis //Web Information Systems and Mining, 2009. WISM 2009. International Conference on. – IEEE, 2009. – С. 543-547.
85. Hernandez-Chamorro A. et al. A methodology of steganalysis for images //| 2009 International Conference on Electrical, Communications, and Computers. – IEEE, 2009. – С. 102-106.
86. Malik H., Chandramouli R., Subbalakshmi K. P. Steganalysis: Trends and Challenges //Multimedia Forensics and Security. – IGI Global, 2009. – С. 245-265.
87. Fridrich J. Structural LSB Detectors [Электронный ресурс] // URL: http://dde.binghamton.edu/download/structural_lsb_detectors/
88. Чичварин Н. В. Особенности изучения методов стеганографии в МГТУ им. НЭ Баумана //Инженерный вестник. – 2015. – №. 8. – С. 15-15.
89. Грибунин В., Оков И., Туринцев И. Цифровая стеганография. – Litres, 2017.
90. Абазина Е. С., Ерунов А. А. Цифровая стеганография: состояние и перспективы //Системы управления, связи и безопасности. – 2016. – №. 2.
91. Patent. Mai Yonghao. (2014). Digital forensics-oriented anomaly steganalysis method and system. CN 104021227.

92. Patent. Univ Xi An. (2014). LSB replacement steganalysis method based on grey co-occurrence matrix statistic features. CN 104008521.
93. Patent. Univ Huazhong. (2014). Method and system for steganalysis of JPEG compatibility. CN 103971321.
94. Patent. Fujian Shitong. (2014). Digital steganography and steganalysis method for color image. CN 103745479.
95. Patent. Inst Inf End CAS. (2013). Steganalysis method based on parameter identification and estimation. CN 103310235.
96. Patent. Liu Quingzhong. (2013). Steganalysis with neighborhood joint density. US 2013208941.
97. Patent. Univ Wuhan. (2013). Universal steganalysis method for JPEG images. CN 103034853.
98. Patent. Univ Beihang. (2011). Reversible image steganalysis method based on histogram peak value fluctuation quantity. CN 102411771.
99. Patent. Univ Beihang. (2011). Steganalysis method based on image smoothness variation characteristics. CN 102147913.
100. Patent. New Jersey Tech Inst. (2010). Method and apparatus for steganalysis of texture images. US 2011135146.
101. Patent. Shin Yun-Qing; Li Bin. Steganalysis of suspect media. US 2010091981.
102. Patent. Fridrich Jessica; Goljan Miroslav. (2009). Reliable detection of LSB steganography in color and grayscale images. US 2003026447.
103. Костырка О. В. Анализ преимуществ пространственной области цифрового изображения-контейнера для стеганопреобразования //Інформатика та математичні методи в моделюванні. – 2013. – №. 3,№ 3. – С. 275-282.
104. Пескова О. Ю., Халабурда Г. Ю. Применение сетевой стеганографии для скрытия данных, передаваемых по каналам связи //Известия Южного федерального университета. Технические науки. – 2012. – Т. 137. – №. 12 (137).

Приложение А

Таблица А1

Численные оценки практической эффективности ВВС, метод RS-analysis

Номер коллекции	Цветовой слой	ОНК				
		1%	2%	3%	4%	5%
1	RED	78,20	44,07	23,78	12,63	6,46
	GREEN	68,75	39,00	18,70	7,46	3,98
	BLUE	74,32	54,02	35,62	24,77	15,42
2	RED	33,50	10,58	5,08	3,83	2,58
	GREEN	29,41	6,66	3,33	2,25	1,66
	BLUE	38,08	12,08	7,24	5,25	3,66
3	RED	26,91	15,58	10,66	7,75	5,83
	GREEN	23,75	8,50	5,16	4,16	3,33
	BLUE	30,25	14,33	10,08	7,83	5,91
4	RED	16,41	4,75	2,58	1,58	1,00
	GREEN	13,50	4,83	2,08	1,33	0,66
	BLUE	27,83	16,50	10,50	6,83	4,33

Таблица А2

Численные оценки практической эффективности ВВС, метод Difference Image Histogram

Номер коллекции	Цветовой слой	ОНК				
		1%	2%	3%	4%	5%
1	RED	68,85	35,92	18,70	8,75	4,27
	GREEN	63,38	29,45	12,73	6,26	2,98
	BLUE	73,03	51,94	34,32	22,38	14,52
2	RED	29,25	9,50	4,33	2,91	2,33
	GREEN	25,00	6,00	3,50	2,66	2,00
	BLUE	34,00	12,16	6,58	5,25	3,83
3	RED	29,91	14,33	9,25	6,83	6,08
	GREEN	26,08	10,08	6,33	5,00	3,91
	BLUE	36,16	16,66	10,75	8,41	7,08
4	RED	17,83	5,50	3,58	2,58	2,00
	GREEN	16,50	4,75	3,00	1,75	1,25
	BLUE	26,16	14,75	9,58	6,83	4,58

Таблица А3

Численные оценки практической эффективности ВВС, метод Sample Pairs analysis

Номер коллекции	Цветовой слой	ОНК				
		1%	2%	3%	4%	5%
1	RED	66,86	37,11	19,10	10,14	4,07
	GREEN	59,70	27,66	12,83	5,57	2,78
	BLUE	72,43	51,14	34,82	21,59	13,93
2	RED	21,08	8,33	4,91	3,16	2,58
	GREEN	15,25	4,41	2,33	1,66	1,41
	BLUE	24,75	10,00	5,75	3,83	2,91
3	RED	23,25	14,41	9,91	7,00	5,58
	GREEN	15,41	7,66	5,08	4,00	2,83
	BLUE	26,25	12,83	10,00	7,91	5,58
4	RED	12,41	4,50	2,58	2,00	1,00
	GREEN	11,83	4,50	2,25	1,08	0,83
	BLUE	25,25	15,00	10,00	6,16	3,75

Таблица А4

Численные оценки практической эффективности ВВС, метод Sample Pairs analysis

Номер коллекции	Цветовой слой	ОНК				
		1%	2%	3%	4%	5%
1	RED	73,83	39,00	17,31	9,25	4,05
	GREEN	-	39,10	16,51	6,36	3,88
	BLUE	81,29	57,51	34,22	21,19	13,33
2	RED	21,00	0,91	0,08	0,00	0,00
	GREEN	12,66	0,33	0,08	0,00	0,00
	BLUE	16,33	0,75	0,00	0,00	0,00
3	RED	14,24	2,66	0,58	0,08	0,00
	GREEN	6,25	0,75	0,16	0,08	0,00
	BLUE	14,16	2,08	0,58	0,08	0,00
4	RED	2,58	1,00	0,33	0,08	0,08
	GREEN	1,58	0,41	0,16	0,00	0,00
	BLUE	6,00	1,75	0,33	0,08	0,08

Приложение Б

Таблица Б1

Сводная таблица рассматриваемых патентов

Предмет поиска (объект исследования, его составные части)	Страна выдачи, вид и номер охранного документа. Классификационный индекс	Заявитель (патентообладатель), страна. Номер заявки, дата публикации	Название изобретения (полной модели, образца)
1	2	3	4
Алгоритмы и системы статистического стеганоанализа в неподвижных изображениях	Китай, патент, CN 104021227. G06F17/30; G06F21/60	Mai Yonghao, Китай. CN20141291747, 2014.06.26. 2014-09-03	Digital forensics-oriented anomaly steganalysis method and system
	Китай, патент, CN 104008521. G06T1/00	Univ Xi An Technology, Китай. CN20141234722, 2014.05.29. 2014-08-27	LSB replacement steganalysis method based on grey co-occurrence matrix statistic features
	Китай, патент, CN 103971321. G06T1/00	Univ Huazhong Science Tech, Китай. CN20141195302, 2014.05.09. 2014-08-06	Method and system for steganalysis of JPEG compatibility
	Китай, патент, CN 103745479. G06T7/00	Fujian Shitong Photoelectric Network Co Ltd, Китай. CN2014134750, 2014.01.24. 2014-04-23	Digital steganography and steganalysis method for color image
	Китай, патент, CN 103310235. G06K9/66; G06T1/00.	Inst Inf End CAS, Китай. CN20131214534 2013.05.31. 2013-09-18	Steganalysis method based on parameter identification and estimation
	США, патент, US 2013208941. G06T1/00	Liu Quingzhong, США. US201313757399, 2013.02.01. 2013-08-15	Steganalysis with neighborhood joint density
	Китай, патент, CN 103034853. G06K9/00	Univ Wuhan, Китай. CN2013106086, 2013.01.08. 2013-04-10	Universal steganalysis method for JPEG images
	Китай, патент, CN 102411771. G06T1/00	Univ Beihang, Китай. CN20111220482, 2011.08.03. 2012-04-11	Reversible image steganalysis method based on histogram peak value fluctuation quantity
	Китай, патент, CN 102147913. G06T1/00	Univ Beihang, Китай. CN2011189230, 2011.04.11. 2011-08-10	Steganalysis method based on image smoothness variation characteristics
	США, патент, US 2011135146. G06K9/36	New Jersey Tech Inst, США. US20100964052, 2010.12.09. 2011-06-09	Method and apparatus for steganalysis of texture images
	США, патент, US 2010091981. G06K9/00; H04K1/00	Shin Yun-Qing; Li Bin; New Jersey Institute of Technology, США. US20090422677, 2009.04.13. 2010-04-15	Steganalysis of suspect media
	США, патент, US 2003026447. G06K9/00; G06T1/00	Fridrich Jessica; Goljan Miroslav; The research foundation of Suny State Univercity Plaza, US20010887805, 2001.06.22	Reliable detection of LSB steganography in color and grayscale images