Воронежский государственный университет

На правах рукописи

Митрофанова Елена Юрьевна

НЕЙРОСЕТЕВЫЕ СЖИМАЮЩИЕ ПРЕОБРАЗОВАНИЯ ДАННЫХ И АЛГОРИТМЫ СОЗДАНИЯ ЦИФРОВЫХ ВОДЯНЫХ ЗНАКОВ В ОБЪЕКТАХ МУЛЬТИМЕДИА ГРАФИЧЕСКИХ И ЗВУКОВЫХ ФОРМАТОВ

Специальность
05.13.17 – «Теоретические основы информатики»

Диссертация
на соискание ученой степени кандидата
технических наук

Научный руководитель
доктор технических наук,
профессор А. А. Сирота

Воронеж, 2014

Содержание

Введение 4
1. Анализ известных методов и алгоритмов создания цифровых водяных знаков 17
1.1. Методы компьютерной стеганографии 18
1.2. Применение компьютерной стеганографии для создания цифровых водяных знаков 30
2. Нейросетевые функциональные модели преобразования данных и их применение для разработки универсальных алгоритмов создания цифровых водяных знаков 53
2.1. Теоретические обоснования возможности построения сжимающих преобразований данных на основе нейронных сетей персептронного типа 54
2.2. Восстановление параметров векторной регрессии при обучении нейронной сети по прямым и косвенным наблюдениям. 56
2.3. Теорема о сходимости весов двухслойной линейной нейронной сети при построении оптимальных линейных оценок случайных векторов 62
2.4. Базовые алгоритмы создания цифровых водяных знаков при использовании в качестве контейнеров статистических эквивалентов сигналов и изображений 74
2.4.1. Нейросетевые функциональные модели преобразования данных и базовый алгоритм создания ЦВЗ при реализации отображений автоассоциативного типа 75
2.4.2. Нейросетевые функциональные модели преобразования данных и базовый алгоритм создания ЦВЗ при реализации отображений гетероассоциативного типа 87
3. Алгоритмы создания цифровых водяных знаков для объектов графических и звуковых форматов представления данных 99
3.1. Модели и алгоритмы создания цифровых водяных знаков для цветных изображений 99

3.2. Модификации алгоритма создания ЦВЗ с целью адаптации по уровню ошибки искажения контейнера и противодействия преобразованиям JPEG 109

3.3. Модели и алгоритмы создания цифровых водяных знаков для аудио-файлов 121

4. Исследование характеристик цифровых водяных знаков для реальных объектов. Описание программного комплекса для создания цифровых водяных знаков 130

4.1. Типовые методики оценки качества цифровых водяных знаков 130

4.2. Программный комплекс для создания цифровых водяных знаков с использованием нейросетевых функциональных моделей преобразования данных в интересах защиты авторских прав на объекты электронного контента 150

Заключение 166

Список литературы 171

Введение

На современном этапе развитие информационных систем и технологий, глобальных компьютерных сетей и средств мультимедиа стимулирует разработку новых методов анализа, хранения, воспроизведения и передачи данных по каналам информационных коммуникаций. К их числу относятся методы и средства обеспечения высоконадежной обработки данных в информационных структурах и системах, методы повышения надежности и безопасности использования информационных технологий. В настоящий момент одним из востребованных подходов в этой области является применение технологий, базирующихся на использовании методов компьютерной стеганографии, позволяющих скрытно встраивать необходимые данные в любые информационные массивы и объекты цифрового контента (ОЦК) (файлы аудио и видеоданных, файлы текстовых форматов, неподвижные изображения и пр.). Указанные технологии широко используются при решении задач создания защищенной связи и передачи данных, аутентификации пользователей, создания цифровых водяных знаков и камуфлирования программного обеспечения. Технологии цифровых водяных знаков (ЦВЗ) успешно используются для решения задач подтверждения авторского права по отношению к ОЦК, контролю их использования авторизованными пользователями и другими лицами, подтверждения подлинности и надежности используемых копий программных средств и мультимедиа объектов.

Методы компьютерной стеганографии основаны на том, что процесс встраивания последовательности данных в исходных объект цифрового контента носит скрытый характер, при этом не нарушается целостность и функциональность ОЦК. Компьютерная стеганография базируется на двух принципах:

- файлы, содержащие оцифрованное изображение или звук, могут быть до некоторой степени видоизменены без потери функциональности;
- человеческие органы восприятия неспособны различить незначительные изменения в цвете изображения или качестве звука.

Наибольшее распространение в компьютерной стеганографии получили методы, основанные на использовании избыточности современных мультимедиа файлов, позволяющие передавать достаточно большой объем скрываемой информации, сохраняя при этом высокий уровень качество ОЦК и возможность их использования по основному назначению.

В области компьютерной стеганографии известны работы отечественных ученых [1, 5, 6, 11, 19, 48] и, прежде всего, В. Г. Грибунина, А.В. Аграновского, А.В. Балакина, И. М. Окова, И. В. Туринцева, В. А. Хорошко, А. Д. Азарова, М. Е. Шелеста и др. Кроме того, вопросы стеганографического скрытия информации в свое время рассматривали зарубежные исследователи: Симмонс (G.J. Simmons), Д. Фридрич (J. Fridrich), Р. Андерсон (R. Anderson), В. Бендер (W. Bender), Н. Моримото (N. Morimoto), К. Качин (C. Cachin), И. Питас (I. Pitas), Х. Фарид (H. Farid), К. Салливан (K. Sullivan) и другие [49, 51, 52, 60-62]. Результаты исследования стеганографических алгоритмов на устойчивость приводят в своих работах Д. Фридрич (J. Fridrich). Р. Попа (R. Popa), Н. Джонсон (N. Johnson), С. Волошиновский (S. Voloshynovskiy) [54-56, 60, 80, 83]. Также необходимо отметить работы Б. Пфицманна (B. Pfitzmann), Б. Шнайера (B. Schneier) и С. Кравера (S. Craver) по вопросам согласования терминологии и формирования базовых стеганографических протоколов [57, 79, 81].

Как уже отмечалось, одно из важных направлений компьютерной стеганографии, получившее широкое распространение в последнее время, связано с применением технологий цифровых водяных знаков [5, 6, 7, 11, 21, 22, 48, 56, 82]. ЦВЗ представляют собой специальные метки, внедряемые в файл, в цифровое изображение или цифровой сигнал в целях контроля их правомерного использования. ЦВЗ делятся на два типа — видимые и невидимые. Видимые ЦВЗ довольно просто удалить или заменить. Для этого могут быть использованы графические или текстовые редакторы. Невидимые ЦВЗ представляют собой встраиваемые в компьютерные файлы метки, не воспринимаемые человеческим глазом или слухом. Для эффективного применения технологий ЦВЗ, необходимо выполнить ряд противоречивых требований, а именно: обеспечить аудио - и

визуальную незаметность сообщений, сохранить исходное качество исходного контейнера и, одновременно, обеспечить высокую достоверность извлечения сообщения с учетом возможных непреднамеренных и преднамеренных воздействий в канале передачи информации. Указанные противоречия не снимаются в полной мере в известных методах и алгоритмах создания ЦВЗ [6,8,11,19,48].

Проблемой разработки методов встраивания и декодирования цифровых водяных знаков занимались многие отечественные и зарубежные ученые [1, 5, 11, 23, 49, 60]: В.Г. Грибунин, А.В. Аграновский, А.В. Балакин, В.А. Митекин, И.В. Туринцев, А.Н. Фионов, К. Качин (C. Cachin), Р. Андерсон (R. Anderson), Х. Фарид (H. Farid), К. Салливан (K. Sullivan), Д. Фридрич (J. Fridrich), Н. Провос (N. Provos) и др. Наиболее популярные программы для реализации технологии внедрения невидимых меток (MyPictureMarc, EIKONAmark 4.5.1, StegMark, SandMark, StegSign, StegSafe и пр.) основаны на алгоритмах, полученных в рамках подобных научных исследований. Так как методы создания цифровых водяных знаков начали разрабатываться совершенно недавно, то здесь имеется много нерешенных проблем.

Одной из них является проблема сохранения качества маркируемых при внедрении ЦВЗ файлов при их использовании по основному назначению в сочетании с устойчивостью встраиваемых меток к возможным преобразованиями контейнера (зашумлению, аффинным трансформациям, обрезке, цифро-аналоговому и аналогово-цифровому преобразованиям, сохранению в другом формате и пр.) и обеспечении достоверности дальнейшего восстановления ЦВЗ.

Кроме того, большинство известных алгоритмов создания ЦВЗ имеет один общий недостаток: используемые в данных алгоритмах процедуры встраивания информации в объекты, представленные в виде цифрового контента (файлы аудио и видеоданных, неподвижные изображения и пр.), реализуют строго определенные последовательности операций преобразования данных, образующие конечное множество вариантов, которые могут быть идентифицированы и вскрыты. Также, к недостаткам большинства известных

алгоритмов встраивания ЦВЗ также можно отнести жесткую зависимость от формата файла, трудоемкость вычисления цифровых водяных знаков, возможное заметное ухудшение качества маркируемого файла (например, изображения) при встраивании метки, и кроме того, в ряде алгоритмов для извлечения цифрового водяного знака необходимо иметь исходный (немаркированный) файл.

Для преодоления вышеперечисленных недостатков известных алгоритмов перспективной задачей представляется дальнейшее развитие и разработка новых технологий создания ЦВЗ, обладающих невысокой сложностью внедрения и детектирования, визуальной незаметностью, адаптируемостью и универсальностью, хорошей устойчивостью к различного рода искажениям и трансформациям защищаемого цифрового объекта, возможностью обнаружения цифровой метки без исходного немаркированного файла. В этом плане перспективным подходом, который может быть использован для реализации новых алгоритмов создания ЦВЗ является применение нейросетевых технологий обработки информации [13,43,55,74].

В работах [54,76] аппарат нейронных сетей при формировании последовательности ЦВЗ не используется, а применяется только для восстановления или обнаружения скрытой последовательности ЦВЗ. В работе [34] разработана модель предсказателя значений среднечастотных коэффициентов матрицы дискретных косинусных преобразований (ДКП), отличающаяся использованием аппарата искусственных нейронных сетей, позволяющая оценить искажения изображения в процессе внедрения в него ЦВЗ с учетом нелинейности зависимости значений ДКП коэффициентов.

В работах [15,24-28,72] рассматривался принципиальный подход, основанный на построении нейросетевых функциональных моделей преобразования данных, в рамках которого специально обученные нейронные сети используются для реализации скрывающих и восстанавливающих преобразований. Показано, что при использовании предложенных в этих работах алгоритмов обработки информации процесс встраивания данных в файл-контейнер носит существенно менее прозрачный характер, чем в большинстве

известных алгоритмов стеганографии. Данный подход позволяет, в частности, реализовать создание ЦВЗ в файлах данных вещественных форматов. Однако предложенные в [25,36] алгоритмы встраивания не является достаточно эффективными для защиты объектов, имеющих целочисленные форматы представления данных.

В целом необходимо отметить, что применение нейросетевых технологий обработки информации принципиально позволяет решить две проблемы, связанные с созданием ЦВЗ, а именно: реализовать функциональный подход к построению алгоритмов встраивания (кодирования) и извлечения (декодирования) ЦВЗ и, одновременно, обеспечить универсальный характер выполняемых при этом преобразований информации. Как показывают предварительные исследования, именно функциональный характер выполняемых с помощью нейронных сетей преобразований (в отличие от классического алгоритмического подхода) и применение искусственных нейронных сетей различных типов для реализации технологий цифровых водяных знаков, позволяет обеспечить меньшую «прозрачность» процесса встраивания и хороший компромисс между визуальной незаметностью, устойчивостью внедрения цифровых меток и объемом используемых вычислительных ресурсов.

Таким образом, тема диссертации, посвященная разработке и исследованию нейросетевых функциональных преобразований, реализующих сжимающие отображения данных, и на их основе – алгоритмов создания цифровых водяных знаков для ОЦК графических и звуковых форматов, представляется актуальной.

Тема диссертации входит в план научно-исследовательских работ Воронежского государственного университета по кафедре технологий обработки и защиты информации и непосредственно связана с научным направлением Воронежского государственного университета «Математическое моделирование, программное и информационное обеспечение, методы вычислительной и прикладной математики и их применение к фундаментальным и прикладным исследованиям в естественных науках».

Тема диссертации соответствует паспорту специальности 05.13.17 – «Теоретические основы информатики» по следующим областям исследований: разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений (п.5 паспорта специальности); разработка методов обеспечения высоконадежной обработки информации и обеспечения помехоустойчивости информационных коммуникаций для целей передачи, хранения и защиты информации; разработка основ теории надежности и безопасности использования информационных технологий (п.11 паспорта специальности).

Целью работы является обоснование и исследование моделей и алгоритмов создания цифровых водяных знаков, основанных на построении нейросетевых сжимающих отображений, в интересах повышения скрытности и устойчивости, а также обеспечения универсальности алгоритмов создания ЦВЗ по отношению различным форматам объектов цифрового контента.

Объектом исследования являются информационные структуры и процессы, обеспечивающие создание и применение цифровых водяных знаков.

Предметом исследования являются нейросетевые модели и алгоритмы преобразования информации в интересах создания цифровых водяных знаков в объектах цифрового (электронного) контента.

Задачи исследования. Для достижения цели в работе рассматриваются и решаются следующие задачи.

1. Анализ известных методов и алгоритмов создания цифровых водяных знаков в объектах цифрового контента.

2. Проведение теоретических обоснований и исследований моделей преобразования данных, основанных на построении нейросетевых сжимающих отображений, обеспечивающих универсальный характер процесса создания цифровых водяных знаков в объекты-контейнеры различных классов.

3. Разработка и исследование нейросетевых функциональных моделей и алгоритмов преобразования данных, обеспечивающих реализацию процесса

встраивания ЦВЗ в исходные объекты-контейнеры и их последующее извлечение при минимальном уровне искажений исходного контейнера. Проведение экспериментальных исследований возможностей разработанных алгоритмов создания ЦВЗ для реальных контейнеров (файлов графических и звуковых форматов).

4. Разработка методик анализа статистической заметности и возможности восстановления встроенной последовательности ЦВЗ сторонним наблюдателем для оценки качества ЦВЗ в объектах графических форматов, а также обоснование рекомендаций относительно возможности использования разработанных класса алгоритмов для решения различного класса задач. Разработка программных средств для моделирования предложенных алгоритмов создания ЦВЗ.

Методы проведения исследования. При решении исследований использовались методы и подходы современной теории информационных процессов и систем, а именно: методы теории статистических решений; методы теории вероятностей и математической статистики; технологии объектно-ориентированного программирования; теория случайных процессов и полей методы и технологии статистического имитационного моделирования; методы цифровой обработки и распознавания сигналов и изображений; аппарат искусственных нейронных сетей; методы компьютерной стеганографии, а также технологии проведения прямого компьютерного эксперимента, выполняемого по отношению к реальным объектам цифрового контента.

Основные результаты, выносимые на защиту и их научная новизна.

На защиту выносятся следующие результаты, впервые достаточно подробно развитые или полученные в настоящей работе:

1. Теоретические обоснования возможности построения гетероассоциативных и автоассоциативных нейросетевых сжимающих преобразований данных универсального типа и исследования их свойств.

2. Базовые нейросетевые функциональные модели преобразований данных, обеспечивающих встраивание ЦВЗ в исходные объекты-контейнеры различных

классов и их последующее извлечение, а также закономерности, описывающие потенциальные характеристики качества ЦВЗ при использовании статистических эквивалентов объектов-контейнеров.

3. Алгоритмы создания цифровых водяных знаков для объектов, имеющих целочисленный формат представления данных, а также исследования возможности их применения для файлов графических и звуковых форматов.

4. Методики и результаты анализа статистической заметности и возможности восстановления встроенной последовательности ЦВЗ сторонним наблюдателем для оценки качества ЦВЗ в объектах графических форматов, а также рекомендации относительно по использованию разработанных алгоритмов для решения различного класса задач.

5. Структура и алгоритмы функционирования многофункционального программного комплекса для встраивания и декодирования цифровых водяных знаков с использованием обоснованных нейросетевых функциональных моделей преобразования данных.

Научная новизна полученных результатов работы определяется следующим.

1. Доказана теорема о сходимости весовых коэффициентов двухслойной линейной нейронной сети гетероассциативного типа с сокращенным числом нейронов в скрытом слое к компонентам собственных векторов, получаемых при решении обобщенной задачи на собственные значения при обучении по совокупности реализаций статистически связанных входного и выходного векторов. Показано, что при обучении такой нейронной сети на выходе получается приближенное представление оптимальной в классе линейных оценки выходного вектора относительно наблюдения входного вектора в виде разложения по первым собственным векторам выборочной матрицы ковариации оценки с минимальной дисперсией остаточной ошибки.

Получены необходимые и достаточные условия построения оптимальных в классе линейных оценок случайных процессов и полей на основе реализации процедур прямого и косвенного обучения двухслойных нейронных сетей по

совокупности реализаций статистически связанных входных и выходных случайных векторов.

2. Установлена целесообразность построения и использования универсальных сжимающих преобразований для стеганографического встраивания ЦВЗ в объекты-контейнеры различных типов с минимальным уровнем дисперсии вносимых искажений на основе искусственных двухслойных нейронных сетей прямого распространения. Обоснованы две принципиально новые схемы преобразования данных для создания ЦВЗ в объектах контейнерах различных типов, отличающиеся применением гетероассоциативного и автоассоциативного сжимающих отображений и специальной процедуры извлечения и модификации «высокочастотной» составляющей фрагмента контейнера при встраивании ЦВЗ. На основе обработки данных, представленных в виде реализаций случайных векторов, (как аналогов аудио файлов) или случайных полей (как аналогов изображений) получены зависимости для основных показателей качества контентно-независимых цифровых водяных знаков: дисперсии и максимальной абсолютной ошибки искажения контейнера, а также вероятности ошибки при восстановлении ЦВЗ.

3. Предложены и исследованы структуры алгоритмов встраивания ЦВЗ, представленных в виде двоичной последовательности, в файлы контейнеры, имеющие целочисленный формат представления данных (изображения, звук, видео), а также алгоритмов восстановления ЦВЗ, основанных на применении алгоритмов классификации фрагментов контейнера, реализуемых с использованием линейной или нелинейной нейронной сети прямого распространения. На основе анализа большого количества аналогичных экспериментальных данных установлены закономерности, описывающие потенциальные характеристики качества встраивания и восстановления ЦВЗ для объектов различных классов.

4. Предложенные методики оценки качества разработанных алгоритмов создания цифровых водяных знаков основаны на применении оригинальных схем преобразования контейнеров-изображений, реализующих воспроизведение

процесса встраивания и специальный анализ данных для оценки статистической заметности встроенных меток и возможности их восстановления сторонним наблюдателем. При их использовании получены впервые экспериментальные оценки вероятности восстановления сообщения ЦВЗ в зависимости от наличия априорной информации относительно конфигурации фрагментов контейнера и модифицируемых участков изображения для различных типов алгоритмов создания меток.

Достоверность результатов работы. Результаты исследований, сформулированные в диссертации, получены на основе корректного использования взаимно дополняющих друг друга теоретических и экспериментальных методов исследований, совпадением результатов, полученных различными методами, между собой, а, в ряде частных случаев, с известными, наглядной физической трактовкой установленных закономерностей и соотношений. Анализ разработанных нейросетевых моделей и алгоритмов создания ЦВЗ проводился как с использованием статистических моделей, так и путем экспериментирования с естественными файлами-контейнерами.

В целом полученные в диссертации выводы и рекомендации формировались на основе всестороннего анализа результатов исследований, допускают ясное физическое истолкование, имеют понятную содержательную трактовку и совпадают в ряде частных случаев с известными результатами, полученными другими авторами.

Теоретическая и практическая ценность результатов диссертации. Теоретическая значимость результатов диссертационной работы заключается в том, что полученные результаты отвечают потребностям важного направления в области обработки и анализа цифровых сигналов и изображений – создания цифровых водяных знаков для объектов цифрового контента, имеющих графические и аудио форматы в интересах подтверждения авторских прав на объекты электронного (цифрового) контента, а также скрытого маркирования ОЦК Полученные в диссертации зависимости для основных показателей качества ЦВЗ позволяют обоснованно выбрать необходимый алгоритм обработки

информации и сформировать рекомендации по его применению, а также определять эффективность эвристических алгоритмов на основе сравнения показываемых ими результатов с потенциально достижимыми.

Выполненные теоретические обоснования и доказанные утверждения относительно свойств нейросетевых функциональных моделей и алгоритмов преобразований данных имеют общее значение и могут быть использованы для построения универсальных сжимающих отображений в системах цифровой обработки сигналов и изображений, распознавания образов и статистического анализа данных.

Результаты диссертационной работы имеют практическое значение для разработки специального математического и программного обеспечения в интересах создания цифровых водяных знаков как средств скрытого маркирования объектов цифрового (электронного) контента, обеспечивающих эффективный контроль использования объектов интеллектуальной собственности, а также для диагностики аудио- и графических данных.

Реализация результатов работы. Полученные в диссертации результаты в части разработки нейросетевых моделей и алгоритмов создания цифровых водяных знаков для аудио- и графических объектов электронного (цифрового) контента реализованы: в ООО «Фото Технологии-В» при выполнении НИОКР «Программное обеспечение для реализации новой информационной технологии защиты авторских прав на основе создания цифровых водяных знаков с использованием нейросетевых функциональных моделей преобразования данных» государственный контракт №8501р/13581 от 04.10.2010; в Воронежском государственном университет при выполнении гранта Российского фонда фундаментальных исследований в рамках научного проекта № 13-01-97507 р_центр_ а «Модели и методы для реализации новых информационных технологий создания цифровых водяных знаков для защиты объектов цифрового контента», 2013 г., а также НИР «Разработка моделей, методов и алгоритмов обработки информации для создания информационных технологий и систем нового поколения» № 01201263910, 2011-2013 г.

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались: на XII, XIV Международных научно-технических конференциях «Кибернетика и высокие технологии XXI века» (Воронеж) в 2011, 2013 годах; на XI, XII, XIV Международных конференциях «Информатика: проблемы, методология, технологии» (Воронеж) в 2011, 2012 и 2014 годах.

Публикации. По теме диссертации опубликовано 10 работ, в том числе 3 статьи в изданиях, рекомендованных ВАК, а также свидетельство о регистрации программы. В совместных работах научному руководителю принадлежит постановка задачи и определение направления исследований. В опубликованных работах автору настоящей диссертации принадлежит доказательство теорем и проведение рассуждений, необходимых для решения поставленных задач, обоснование алгоритмов обработки информации, организация экспериментов для обоснования алгоритмов, полученных теоретическим путем, анализ и интерпретация полученных результатов. При разработке в соавторстве с М.А. Дрюченко программного комплекса для создания цифровых водяных знаков автору принадлежит обоснование структуры комплекса в части, а также нейросетевых функциональных моделей и алгоритмов преобразования данных.

Объем и структура диссертационной работы. Диссертация состоит из введения, четырех разделов, заключения и списка литературы из 83 наименований. Объем диссертации составляет 179 страниц, включая 170 страницы основного текста, содержащего 51 рисунков, 11 таблиц и 9 страниц списка литературы.

В первой главе диссертации определяются основные понятия и единая терминология в области компьютерной стеганографии и технологии создания цифровых водяных знаков для объектов цифрового (электронного) контента. Проводится анализ понятий и методов компьютерной стеганографии, а также анализируется специфика требований, выдвигаемых при применении технологий ЦВЗ. Особое внимание уделяется обзору известных алгоритмов и технологий создания ЦВЗ.

Во второй главе сформулирована и доказана теорема о сходимости весовых коэффициентов двухслойной линейной нейронной сети с сокращенным числом нейронов в скрытом слое к компонентам собственных векторов, получаемых при решении обобщенной задачи на собственные значения. Сформулирована постановка задачи и предложена общая модель встраивания последовательности ЦВЗ с учетом применения искусственных нейронных сетей различной архитектуры. Приводятся универсальные схемы преобразования данных для создания цифровых водяных знаков на основе автоассоциативных и гетероассоциативных сжимающих отображений.

В третьей главе диссертации рассмотрены нейросетевые функциональные модели и универсальные алгоритмы преобразования данных в интересах решения задачи создания ЦВЗ в контейнерах с целочисленным форматом представления данных. Представлены обобщенные схемы встраивания и извлечения информации при реализации предлагаемого нейросетевого подхода. Получены закономерности, описывающие потенциальные характеристики качества создаваемых ЦВЗ для объектов различных классов (цветные изображения в формате RGB и YCbCr, аудио-файлы в формате WAV). Рассмотрены модификации алгоритма создания ЦВЗ в файлах графических форматов, которые могут быть подвергнуты сжатию в стандарте JPEG.

В четвертой главе представлены разработанные и реализованные методики анализа возможности обнаружения встроенного ЦВЗ и анализа возможности восстановления встроенной последовательности ЦВЗ при определенных воздействиях стороннего наблюдателя. Приводится описание разработанного на языке C++ программного комплекса для создания цифровых водяных знаков, реализующего ранее рассмотренные модели и алгоритмы для работы с файлами графических форматов.

Автор выражает глубокую признательность научному руководителю проф. А.А. Сироте за постоянное внимание и руководство, а также коллективам кафедры информационных систем и кафедры технологий обработки и защиты информации ВГУ за оказанную поддержку.

1. Анализ известных методов и алгоритмов создания цифровых водяных знаков

В настоящее время в связи с интенсивным развитием и распространением технологий, которые позволяют с помощью компьютера интегрировать, обрабатывать и синхронно воспроизводить различные типы данных и информации (мультимедийные технологии), вопрос идентификации и подтверждения авторского права по отношению к объектам цифрового (электронного) контента (ОЦК), является чрезвычайно актуальным [19].

Для того чтобы компьютерный файл, представляющий собой объект авторского права, содержал всю необходимую информацию о его авторе и истории его создания, применяются специальные метки или, так называемые, цифровые водяные знаки (ЦВЗ). Применение ЦВЗ позволяет не только затруднить нарушение авторских прав, но проконтролировать его использование авторизованными пользователями и другими лицами. Наряду с информацией, представленной в форме цифровых изображений (фотографиями, рисунками, отсканированными бумажными документами и т.д.), ЦВЗ также широко используются и для маркирования видео - и аудиоданных. Получение доступа к информации в связи с появлением и распространением глобальных компьютерных сетей представляет достаточно простую процедуру. В то же время, легкость и скорость такого доступа предполагают активное использование ЦВЗ при размещении уникальных фотографий, видео, аудиотреков в электронном виде в глобальной сети Интернет [19].

Базовым подходом, используемым для реализации технологий ЦВЗ, является применение методов компьютерной стеганографии. Поэтому целью настоящей главы является анализ понятий и методов компьютерной стеганографии, что необходимо для последующего изложений материалов диссертации. При изложении основ компьютерной стеганографии в первом параграфе главы анализируется специфика требований, выдвигаемых при применении технологий ЦВЗ. Второй параграф главы посвящен обзору известных

алгоритмов и технологий создания ЦВЗ, а также анализу недостатков и ограниченных возможностей известных решений.

1.1. Методы компьютерной стеганографии

Компьютерная стеганография (КС) является сравнительно молодым и быстро развивающимся направлением в информатике. На международной конференции Information Hiding: First Information Workshop в 1996 году было предложено использовать единую терминологию и основные понятия КС [11], которые далее потребуются при изложении результатов диссертации и приводятся ниже.

Стеганография – это наука о передаче информации в скрытой форме, причем сам факт передачи может быть неизвестен внешнему наблюдателю.

Различают два вида стеганографии:

- скрытие информации от пассивного наблюдателя; в этом случае основная цель – не допустить обнаружения встроенной в ОЦК информации или отвлечения наблюдателя на эту информацию без необходимости.
- скрытие информации от активного наблюдателя, когда наличие встроенной в ОЦК информации в принципе известно, но получение этой информации должно быть существенно затруднено.

Термины которые приводятся ниже применятся во многих известных работах [7,11,19,48], посвященных компьютерной стеганографии.

Сообщение – информация (данные), которую необходимо встроить в ОЦК, а затем восстановить для передачи потребителю. Будем далее по тексту диссертации обозначать сообщение как вектор $d \in R^m$. В качестве такого сообщения в данной работе рассматривается ЦВЗ.

Контейнер или носитель – любой файл или поток данных, структура и размер которого позволяют встроить (скрыть) необходимые данные. Будем далее обозначать контейнер вектором $z \in R^n$, $n >> m$.

Как правило, объем стеганографически встраиваемой информации напрямую зависит от объема контейнера. Чем больше объем контейнера, тем больше информации можно встроить. К числу наиболее часто используемых в качестве контейнеров цифровых носителей относятся: неподвижные

изображения, текстовые, аудио- и видеофайлы, а также исполняемые файлы и некоторые протоколы (например, протокол транспортного уровня TCP) [22, 48].

Незаполненный (пустой) контейнер (НК) – контейнер, который не содержит встроенного сообщения, но является потенциально пригодным для скрытия в нем информации.

Заполненный контейнер (ЗК) – контейнер, содержащий встроенную информацию. Как правило, выдвигается требование, чтобы заполненный ЗК был «внешне» неотличим от не заполненного контейнера или чтобы встроенная информация не мешала воспринимать основную внешнюю информацию.

Стеганографический ключ (стегоключ) – закрытый или открытый ключ, необходимый для скрытия местонахождения данных в контейнере. Будем далее обозначать стегоключ вектором $k \in K \subset R^r$, $r << n$.

Чаще всего в качестве стегоключа могут выступать специально сформированные данные, а также определенная последовательность действий или алгоритм, задающий последовательность операций при встраивании/извлечении данных в элементы естественных контейнеров. При необходимости встроенные данные могут быть дополнительно зашифрованы обычными методами. Стегоключи могут быть как открытыми, так и закрытыми, все зависит от особенностей рассматриваемой способа стеганографического скрытия информации (ССИ) и его применения. При этом может использоваться один, несколько ключей. В некоторых случаях ключ может отсутствовать вообще. При знании ключа пользователь сможет извлечь ранее скрытое сообщение.

Стеганографическим преобразованием данных будем далее называть пару отображений вида

$$F_1 : Z \times D \times K \to \tilde{Z}, \qquad F_2 : \tilde{Z} \times K \to \tilde{D},$$

где Z – множество всех контейнеров; D – множество всех сообщений; K – множество ключей; $\tilde{Z}$ – множество заполненных контейнеров; оператор F_1 (прямой оператор) реализует встраивание информации, а оператор F_2 (обратный

оператор) ее восстановление. При выполнении соответствующих преобразований для элементов указанных множеств должны выполняться следующие требования:

$$\tilde{z} = F_1(z, d, k),\ z \in Z,\ d \in D,\ k \in K,\ \tilde{d} = F_2(\tilde{z}, k),\ \tilde{d} \in \tilde{D},$$

$$\|z - \tilde{z}\| \to \min, \quad \|d - \tilde{d}\| \to \min,$$

то есть при выполнении стеганографических преобразований свойства контейнера и восстанавливаемого сообщения должны искажаться в минимальной степени.

Стеганографической системой (стегосистемой) будем называть $\{Z, D, K, F_1, F_2\}$ – совокупность контейнеров, сообщений, ключей и связывающих их преобразований. Другими словами, стеганографическая система (СС) это совокупность средств и методов для формирования скрытого канала передачи информации (*стегоканала).*

Источник внешнего воздействия (ИВВ) – потенциальный или реальный субъект (нарушитель, сторонний пользователь и пр.) осуществляющий преднамеренное (атаку) или непреднамеренное воздействие на стегоситему с *целью* обнаружения стегоканала, извлечения передаваемой в нем информации или ее уничтожения и искажения. При этом обнаружение стегоканала является наиболее трудоемкой задачей.

Скрытая пропускная способность (СПС) стегоканала – максимальное количество информации, которое может быть встроено в один элемент (пиксель, отсчет) контейнера. Обязательным условием при этом является безошибочность передачи скрываемых данных получателю, а также их защищенность от таких атак нарушителя как попытка выявления факта наличия стегоканала, получения содержания скрытых сообщений, умышленное введение сфальсифицированных данных или же разрушение встроенной в контейнер информации.

Стеганографическая стойкость – способность стегосистемы скрывать от квалифицированного нарушителя факт передачи сообщений, а также способность противостоять попыткам нарушителя разрушить, исказить, удалить скрытно передаваемые сообщения, а также способность подтвердить или опровергнуть подлинность скрытно передаваемой информации. Считается, что если у

нарушителя имеется возможность выявить факт наличия скрытого канала передачи сообщений, то СС, как правило, уже считается нестойкой.

Стегоанализ – процедура обнаружения факта ССИ и, если возможно, определения стеганографического ключа и (или) выделения ранее скрытой информации.

Основными положениями современной компьютерной стеганографии являются следующие [6]:

- методы скрытия должны обеспечивать аутентичность и целостность файла - контейнера;
- предполагается, что источникам внешнего воздействия полностью или частично известны возможные стеганографические методы;
- безопасность информационных технологий основывается на сохранении стеганографическим преобразованием основных свойств открыто передаваемого файла при внесении в него сообщения и некоторой неизвестной информации — ключа;
- даже если факт скрытия сообщения стал известен ИВВ, извлечение самого сообщения представляет сложную вычислительную задачу.

Подавляющее большинство методов КС базируется на двух ключевых принципах [11]:

- файлы, которые не требуют абсолютной точности представления и воспроизведения данных (например, файлы с изображением, звуковой информацией и т.д.), могут быть видоизменены (конечно, до определенной степени) без потери своей функциональности.
- органы чувств человека неспособны надежно различать незначительные изменения в модифицированных таким образом файлах и/или отсутствует специальный инструментарий, который был бы способен выполнять данную задачу.

В основе базовых подходов к реализации методов КС в рамках той или иной информационной среды лежит выделение малозначительных фрагментов этой среды и замена существующей в них информации другой информацией, которую

необходимо скрыть. Поскольку в КС рассматриваются среды, поддерживаемые средствами вычислительной техники и компьютерными сетями, то вся информационная среда в результате может быть представлена в цифровом виде.

Таким образом, незначительные для кадра информационной среды фрагменты относительно того или иного алгоритма или методики анализа заменяются фрагментами со скрываемой информацией. Под кадром информационной среды в данном случае подразумевается определенная ее часть, выделенная по характерным признакам. Такими признаками зачастую являются семантические характеристики выделяемой части информационной среды. Например, в качестве такого кадра может рассматриваться какое-нибудь отдельное изображение, звуковой файл, Web-страница и т.д.

Для существующих методов компьютерной стеганографии вводят следующую классификацию (рисунок 1.1) [19] .

По способу выбора контейнера различают суррогатные (или так называемые эрзац-методы), селективные и конструирующие методы стеганографии.

В суррогатных (безальтернативных) методах стеганографии полностью отсутствует возможность выбора контейнера, и для скрытия сообщения избирается первый попавшийся контейнер – эрзац-контейнер, который в большинстве случаев не оптимален для скрытия сообщения заданного формата.

В селективных методах КС считается, что скрытое сообщение должно воспроизводить специальные статистические характеристики шума контейнера. Для этого генерируют большое количество альтернативных контейнеров с последующим выбором наиболее оптимального из них для конкретного сообщения. Особым случаем такого подхода является вычисление некоторой хэш-функции для каждого контейнера. При этом для скрытия сообщения выбирается тот контейнер, хэш-функция которого совпадает со значением кэш-функции сообщения (то есть стеганограммой является выбранный контейнер).

В конструирующих методах стеганографии контейнер генерируется самой стеганосистемой. При этом существует несколько вариантов реализации. Так,

например, шум контейнера может имитироваться скрытым сообщением. Это реализуется с помощью процедур, которые не только кодируют скрываемое сообщение под шум, но и сохраняют модель изначального шума. В предельном случае по модели шума может строиться целое сообщение.

Рисунок 1.1 – Классификация методов компьютерной стеганографии

По способу доступа к скрываемой информации различают методы для потоковых (беспрерывных) контейнеров и методы для фиксированных (ограниченной длины) контейнеров.

По способу организации контейнеры, подобно помехоустойчивым кодам, могут быть систематическими и несистематическими.

В первых можно указать конкретные места стеганограммы, где находятся информационные биты собственно контейнера, а где – шумовые биты, предназначенные для скрытия информации (как, например, в широко распространенном методе наименее значащего бита).

В случае несистематической организации контейнера такое разделение невозможно. В этом случае для выделения скрытой информации необходимо обрабатывать содержимое всей стеганограммы.

По используемому принципу скрытия методы компьютерной стеганографии делятся на два основных класса: методы непосредственной замены и спектральные методы. Если первые, используя избыток информационной среды в пространственной (для изображения) или временной (для звука) области, заключаются в замене малозначительной части контейнера битами секретного сообщения, то другие для скрытия данных используют спектральные представления элементов среды, в которую встраиваются скрываемые данные (например, в разные коэффициенты массивов дискретно-косинусных преобразований, преобразований Фурье, Карунена-Лоева, Адамара, Хаара и т.д.).

Основным направлением компьютерной стеганографии является использование свойств именно избыточности контейнера-оригинала, но при этом следует принимать во внимание то, что в результате скрытия информации происходит искажение некоторых статистических свойств контейнера или, же нарушение его структуры. Это необходимо учитывать для уменьшения демаскирующих признаков [19].

В особую группу можно выделить методы, которые используют специальные свойства форматов представления файлов:

- зарезервированные для расширения поля файлов, которые зачастую заполняются нулями и не учитываются программой;
- специальное форматирование данных (сдвиг слов, предложений, абзацев или выбор определенных позиций символов);
- использование незадействованных участков на магнитных и оптических носителях;
- удаление файловых заголовков-идентификаторов и т.д.

В основном для таких методов характерны низкая степень скрытности, низкая пропускная способность и слабая производительность.

По назначению различают методы собственно для скрытой передачи (или скрытого хранения) данных и методы для скрытия данных в цифровых объектах с целью подтверждения авторских прав на них.

По типам контейнера выделяют методы с контейнерами в виде текста, аудиофайла, изображения и видео. В настоящее время методы компьютерной стеганографии развиваются по двум основным направлениям [7]:

- методы, основанные на использовании специальных свойств компьютерных форматов;
- методы, основанные на избыточности аудио и визуальной информации.

Сравнительные характеристики существующих стеганографических методов приведены в таблице 1.1.

Таблица 1.1 - Сравнительные характеристики стеганографических методов

Стеганографические методы	Краткая характеристика и особенности методов	Недостатки	Преимущества
1. Методы использования специальных свойств компьютерных форматов данных			
1.1. Методы использования зарезервированных для расширения полей компьютерных форматов данных	Поля расширения имеются во многих мультимедийных форматах, эти поля заполняются нулевой информацией и не учитываются основной программой	Низкая степень скрытности, передача небольших ограниченных объемов информации	Простота использования

Продолжение таблицы 1.1

Методы специального форматирования текстовых файлов	1.3. Методы использования известного смещения слов, предложений, абзацев	Методы основаны на изменении положения строк и расстановки слов в предложении, что обеспечивается вставкой дополнительных пробелов между словами	1. Слабая производитель-ность метода, передача небольших объемов информации 2. Низкая степень скрытности	Простота использования. Имеется опубликованное программное обеспечение реализации данного метода
	1.4. Методы выбора определенных позиций букв (нулевой шифр)	Акростих - частный случай этого метода (например, начальные буквы каждой строки образуют сообщение)		
	1.5. Методы использования специальных свойств полей форматов, не отображаемых на экране	Методы основаны на использовании специальных "невидимых", скрытых полей для организации сносок и ссылок (например, использование черного шрифта на черном фоне)		
1.6. Методы скрытия в неиспользуемых местах гибких дисков		Информация записывается в обычно неиспользуемых местах ГМД (например, в нулевой дорожке)	1. Слабая производительно сть метода, передача небольших объемов информации 2. Низкая степень скрытности	Простота использования. Имеется опубликованное программное обеспечение реализации данного метода
1.7. Методы использования имитирующих функций (mimic-function)		Методы основаны на генерации текстов и является обобщением акростиха. Для тайного сообщения генерируется осмысленный текст, скрывающий само сообщение	1. Слабая производительно сть метода, передача небольших объемов информации 2. Низкая степень скрытности	Результирующий текст не является подозрительным для систем мониторинга сети

Продолжение таблицы 1.1

1.8. Методы удаления идентифицирующего файл заголовка	Скрываемое сообщение шифруется и у результата удаляется идентифицирующий заголовок, оставляя только шифрованные данные. Получатель заранее знает о передаче сообщения и имеет недостающий заголовок	Проблема скрытия решается только частично. Необходимо заранее передать часть информации получателю	Простота реализации. Многие средства (White Noise Storm, S-Tools), обеспечивают реализацию этого метода с PGP шифроалгоритмом
2. Методы использования избыточности аудио и визуальной информации			
2.1. Методы непосредственной модификации элементов аудио- и видеофайлов (пространственно-временные методы)	Младшие разряды цифровых отсчетов содержат очень мало полезной информации. Их заполнение дополнительной информацией практически не влияет на качество восприятия, что и дает возможность скрытия конфиденциальной информации	За счет введения дополнительной информации искажаются статистические характеристики цифровых потоков. Для снижения компрометирующ их признаков требуется коррекция статистических характеристик	Возможность скрытой передачи большого объема информации. Возможность подтверждения авторского права, скрытого изображения товарной марки, регистрационных номеров и т.п.
2.2. Методы непосредственной модификации спектральных представлений аудио- и видеофайлов (спектральные методы)	В реальных алгоритмах сжатия осуществляют обнуление не пикселей изображения, а спектральных коэффициентов. Преимущество такого подхода заключается в том, что близкие к нулю спектральные коэффициенты имеют тенденцию располагаться в заран	То же	То же

Как видно из таблицы 1.1, только первый класс методов основан на использовании специальных свойств компьютерных форматов представления

данных, а не на избыточности самих данных. Таким образом, можно сделать вывод, что основным направлением компьютерной стеганографии является использование избыточности аудио и визуальной информации.

Области практического применения методов современной компьютерной стеганографии достаточно широк. С точки зрения авторов [7, 11, 37] основные области использования компьютерной стеганографии:

- скрытная передача информации;
- преодоление систем мониторинга и управления сетевыми ресурсами;
- камуфлирования программного обеспечения;
- создание цифровых водяных знаков для различных приложений.

Остановимся подробнее на каждой из перечисленных задач [7].

Скрытная передача информации. Это область использования КС является наиболее эффективной. Так, например, только одна секунда оцифрованного звука с частотой дискретизации 44100 Гц и уровнем отсчета 8 бит в стерео режиме позволяет скрыть за счет замены наименее значимых младших разрядов на скрываемое сообщение около 10 Кбайт информации. При этом, изменение значений отсчетов составляет менее 1 %.

Преодоление систем мониторинга и управления сетевыми ресурсами. Стеганографические методы, направленные на противодействие системам мониторинга и управления сетевыми ресурсами, позволяют противостоять попыткам контроля над информационным пространством при прохождении информации через серверы управления локальных и глобальных вычислительных сетей.

Камуфлирование программного обеспечения (ПО). Другой важной задачей стеганографии является камуфлирование ПО. В тех случаях, когда использование ПО незарегистрированными пользователями является нежелательным, оно может быть закамуфлировано под стандартные универсальные программные продукты (например, текстовые редакторы) или скрыто в файлах мультимедиа (например, в звуковом сопровождении компьютерных игр).

Создание цифровых водяных знаков. В общем случае цифровые водяные знаки используются в следующих приложениях:

- скрытое или видимое маркирование объектов цифрового контента, представленных в виде графических, видео и аудио файлов для их идентификации, представления истории создания, а также внесения любой интересующей пользователя информации без существенных искажений, мешающих использовать ОЦК по основному назначению;
- подтверждение авторских прав на объекты цифрового контента, а также подлинности используемых копий объектов;
- создание RFID-меток повышенной скрытности и устойчивости.

Следует также отметить ряд принципиальных отличий постановки задачи создания ЦВЗ при реализации стеганографического метода и противодействию негативным воздействиям ИВВ:

- как правило, сам факт наличия ЦВЗ в файле-контейнере может быть известен потенциальному ИВВ с высокой степенью достоверности;
- при осуществлении внешнего воздействия ИВВ интересует не столько содержание встроенного сообщения, а, в большей степени, его уничтожение и модификация;
- файл-контейнер ОЦК с внесенным в него ЦВЗ может подвергаться трансформациям и различным преобразованиям формата представления данных, позволяющим при этом использовать объект по основному назначению, хотя и с худшим качеством.

Указанные обстоятельства выдвигают повышенные требования к скрытности самого способа создания ЦВЗ (скрытности стегоключа), сложности воспроизведения или повторения алгоритма встраивания ЦВЗ нарушителем или сторонним пользователем, устойчивости ЦВЗ к различного рода трансформациям и преобразованиям формата.

1.2. Применение компьютерной стеганографии для создания цифровых водяных знаков

Анализ целого ряда публикаций [6,7,11,48] по существующим алгоритмам и технологиям создания цифровых водяных знаков (ЦВЗ), позволяет сделать вывод, что использование ЦВЗ для подтверждения авторских прав в настоящее время приобретают все большую популярность.

ЦВЗ делятся на два типа – видимые и невидимые. Видимые ЦВЗ довольно просто удалить или заменить. Для этого могут быть использованы графические или текстовые редакторы. Невидимые ЦВЗ представляют собой встраиваемые в компьютерные файлы вставки, не воспринимаемые человеческим глазом или ухом. Обычно для технологий создания ЦВЗ выдвигают следующие требования:

- незаметность ЦВЗ для пользователей;
- индивидуальность алгоритма нанесения (достигается с помощью стеганографического алгоритма с использованием ключа);
- возможность для автора обнаружить недопустимое использование файла;
- невозможность удаления неуполномоченными лицами;
- устойчивость к изменениям носителя-контейнера (к изменению его формата и размеров, к масштабированию, сжатию, повороту, фильтрации, введению спецэффектов, монтажу, аналоговым и цифровым преобразованиям).

Представляет интерес работа [6], в которой рассматриваются следующие основные типы ЦВЗ, представленные в сводной таблице 1.2.

Таблица 1.2 – Основные особенности технологий цифровых водяных знаков

Технология ЦВЗ	Особенности ЦВЗ	Примечание
Робастные (РЦВЗ)	Обладают высокой устойчивостью к внешним воздействиям	Анализ литературы показывает, что РЦВЗ посвящено наибольшее количество работ
Хрупкие (ХЦВЗ)	Разрушаются при незначительной модификации заполненного контейнера. Применяются для аутентификации сигналов.	Отражают не только факт модификации контейнера, но также вид и местоположение этого изменения
Полухрупкие (ПЦВЗ)	Устойчивы по отношению к одним воздействиям и не устойчивы — к другим	Позволяют выполнять сжатие изображения, но запрещать вырезку из него или вставку в него фрагмента

Задачу встраивания и выделения сообщений из другой информации выполняет стегосистема [11].

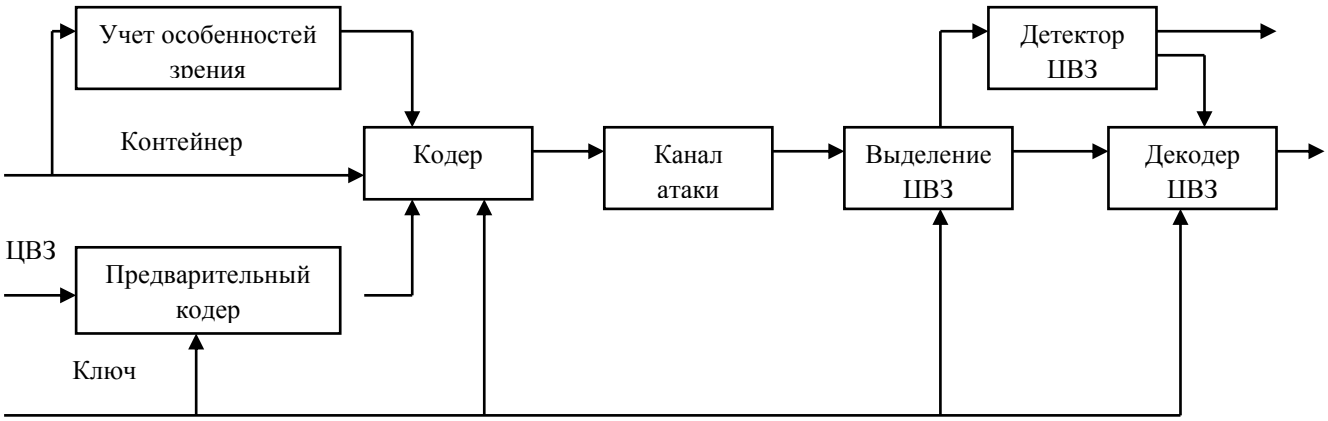


Рисунок 1.2 - Структурная схема типичной стегосистемы ЦВЗ

Стегосистема состоит из следующих основных элементов, представленных на рисунке 1.2:

- *прекодер* – устройство, предназначенное для преобразования скрываемого сообщения к виду, удобному для встраивания в сигнал-контейнер;

- *стегокодер* – устройство, предназначенное для осуществления вложения скрытого сообщения в другие данные с учетом их модели;

- устройство выделения встроенного сообщения;

- *стегодетектор* – устройство, предназначенное для определения наличия стегосообщения;

- *декодер* – устройство, восстанавливающее скрытое сообщение. Этот узел может отсутствовать, как будет пояснено далее.

В работе [7, 11] авторы вводится классификация и приводятся результаты анализа современных алгоритмов встраивания ЦВЗ в файлы - изображения При этом выделяются класс аддитивных алгоритмов, класс алгоритмов, на основе квантования.

Алгоритмы аддитивного внедрения ЦВЗ заключаются в линейной модификации исходного файла графического формата, а извлечение в декодере

производится на основе корреляционных методов. При этом ЦВЗ обычно складывается с исходным контейнером, либо «вплавляется» (fusion) в него.

Соответственно, аддитивные алгоритмы разделяются на алгоритмы на основе линейного встраивания данных и алгоритмы на основе слияния ЦВЗ и контейнера. В алгоритмах на основе линейного встраивания данных ЦВЗ представляет собой последовательность чисел длины N, которая внедряется в выбранное подмножество отсчетов исходного файла контейнера. Если вместо последовательности псевдослучайных чисел в изображение встраивается другое изображение (например, логотип фирмы), то соответствующие алгоритмы внедрения называются алгоритмами слияния. Размер внедряемого сообщения намного меньше размера исходного изображения. Перед встраиванием оно может быть зашифровано или преобразовано каким-нибудь иным образом.

Под квантованием понимается процесс сопоставления большого (бесконечного) множества значений с некоторым конечным множеством чисел. Понятно, что при этом происходит уменьшение объема информации за счет ее искажения. Квантование находит применение в алгоритмах сжатия с потерями.

Алгоритмы на основе квантования разделяются на алгоритмы встраивания ЦВЗ с использованием скалярного квантования (на вход квантователя подаются скалярные значения) и алгоритмы встраивания ЦВЗ с использованием векторного квантования (на вход квантователя подаются последовательности или блоки отсчетов).

Фрактальное кодирование может рассматриваться, как разновидность векторного квантования, причем в качестве кодовой книги выступают различные преобразования (поворот, сдвиг, масштабирование), позволяющие аппроксимировать блоки изображения малого размера (ранговые) блоками большего размера (доменами). Эта последовательность преобразований и передается декодеру.

Таблица 1.3 – Современные алгоритмы создания ЦВЗ для файлов – изображений

	Характеристика ЦВЗ	Принцип работы	Преимущества	Недостатки	Пример
1. Аддитивные алгоритмы создания ЦВЗ					
1.1. Алгоритмы на основе линейного встраивания ЦВЗ					
Последовательность псевдослучайных чисел, распределенных по закону Гаусса, длиной 1000 чисел		Модификация 1000 самых больших коэффициентов дискретного косинусного преобразования (ДКП)	Сильная робастность ЦВЗ при сжатии и других видах обработки сигнала	Трудоем-кость вычисления двумерного ДКП	Cox, [21, 48, 68]
Последовательность бинарных псевдослучайных чисел $w_{i\ I}$ {-1, 1}, длина которой определяется размерами исходного изображения		Модификация всех коэффициентов детальных поддиапазонов первого подуровня разложения при выполнении четырехуровневого вейвлет-преобразования	Возможность обнаружения ЦВЗ без исходного изображения. Сильная визуальная незаметность ЦВЗ	Для извлечения ЦВЗ необходимо иметь исходное изображение	Barni, [6]
Массив псевдослучайных чисел, распределенных по закону Гаусса, размером 1024 числа		Модификация всех коэффициентов LL поддиапазона вейвлет-преобразования изображения	Возможность модификации алгоритма для использования секретного ключа	Для извлечения ЦВЗ необходимо иметь исходное изображение	G. Nicchiotti [7,78]
Последовательность псевдослучайных действительных чисел, распределенных по закону Гаусса		Модификация наибольших коэффициентов детальных поддиапазонов трехуровневой декомпозиции изображения	Хорошая визуальная маскировка внедренных данных. Для обнаружения ЦВЗ не требуется наличие исходного изображения		R.Dugad [77]

Продолжение таблицы 1.3

Последовательность псевдослучайных действительных чисел, распределенных по закону Гаусса, длиной 1000 чисел	Модификация перцептуально значимых коэффициентов трехуровневой декомпозиции изображения с использованием биортогональных вейвлет-фильтров	Робастность ЦВЗ ко многим видам атак. Для обнаружения ЦВЗ не требуется наличие исходного изображения		J.Kim [69]
Последовательность псевдослучайных действительных чисел, распределенных по гауссовскому закону	Модификация наибольших коэффициентов каждого поддиапазона трехуровневой декомпозиции изображения	Для обнаружения ЦВЗ не требуется наличие исходного изображения		Y.Kim [67]
Массив биполярных псевдослучайных чисел	Модификация 1000 наибольших коэффициентов комплексного вейвлет-преобразования (ЦВЗ также подвергается преобразованию)	Для обнаружения ЦВЗ не требуется наличие исходного изображения		P.Loo [74]
Последовательность псевдослучайных действительных чисел, распределенных по гауссовскому закону	Модификация наибольших коэффициентов трехуровневого вейвлет-преобразования (коэффициенты отбираются в соответствии с заданным порогом)	Высокая робастность ЦВЗ к некоторым видам атак	Для извлечения ЦВЗ необходимо иметь исходное изображение	C.Lu [23, 67, 70, 75]

Продолжение таблицы 1.3

Последовательность псевдослучайных действительных чисел, длина которой зависит от пропускной способности изображения, вычисляемой на основе модели человеческого зрения	Модификация коэффициентов четырехуровнего вейвлет-преобразования, отобранных с учетом заданного порога	Высокая робастность внедряемого ЦВЗ	Для извлечения ЦВЗ необходимо иметь исходное изображение	C.Podilchuk [13,55]
Последовательность псевдослучайных действительных чисел, распределенных по гауссовскому закону	Модификация наибольших коэффициентов из высокочастотного и среднечастотного диапазонов преобразования Хаара	Высокая робастность к атакам с изменением масштаба.	Для извлечения ЦВЗ необходимо иметь исходное изображение	X-G.Xia[43]
Последовательность псевдослучайных действительных чисел, распределенных по гауссовскому закону (длина соответствует количеству модифицируемых коэффициентов)	Модификация значимых коэффициентов всех поддиапазонов пятиуровневого вейвлет-преобразования	Возможность модификации алгоритма для использования стегоключа	Для извлечения ЦВЗ необходимо иметь исходное изображение	H.-J. Wang [34,53,54, 63]
Последовательность псевдослучайных действительных чисел, распределенных по гауссовскому закону (длина соответствует количеству модифицируемых коэффициентов)	Алгоритм представляет собой модифицированный вариант предыдущего алгоритма, со слепым извлечением ЦВЗ	Для обнаружения ЦВЗ не требуется наличие исходного изображения	Сильно пониженная помехоустойчивость по сравнению с алгоритмом A27	H.-J. Wang [63]
1.2. **Алгоритмы на основе слияния ЦВЗ и контейнера**				
Черно-белый логотип, размером до 25% от исходного изображения	Модификация всех коэффициентов одноуровневой декомпозиции исходного изображения	Большой размер скрываемого ЦВЗ (до четверти размера исходного изображения)	Для извлечения ЦВЗ необходимо иметь исходное изображение	J.Chen [54]

Продолжение таблицы 1.3

Черно-белый логотип	Модификация всех коэффициентов детальных поддиапазонов вейвлет-преобразования исходного изображения	Для обнаружения ЦВЗ не требуется наличие исходного изображения		D.Kundur [71]
2. Алгоритмы создания ЦВЗ на основе квантования				
2.1. Алгоритмы с использованием скалярного квантования				
Последовательность ±1	Модификация высокочастотных коэффициентов изображения после целочисленного вейвлет-преобразования	Для обнаружения ЦВЗ не требуется наличие исходного изображения		C.-J. Chu [54]
Бинарное изображение, размером 1/2 от исходного изображения	Модификация ВЧ-НЧ и НЧ-ВЧ областей двухуровневого вейвлет-преобразования исходного изображения	Большой размер скрываемого ЦВЗ	Для извлечения ЦВЗ необходимо иметь исходное изображение; Низкая стойкость алгоритма по отношению к операциям обработки сигнала	Hsu [67]
2.2. Алгоритмы с использованием векторного квантования				
Последовательность символов, полученная из логотипа размером 25% от исходного изображения	Модификация n-мерного вектора коэффициентов дискретного вейвлет-преобразования исходного изображения	Большой размер скрываемого ЦВЗ Возможно контролировать робастность, уровень искажений и качество внедряемого изображения	Для извлечения ЦВЗ необходимо иметь исходное изображение	J.Chae, [54]

Продолжение таблицы 1.3

3. Алгоритмы создания ЦВЗ, использующие фрактальное преобразование				
Формируется из исходного изображения (до 15 различных ЦВЗ)	Модификация несколько «особых» точек с использованием известного из фрактального кодирования метода Стефана-Харриса. Для каждой особой точки выполняют изменение доменного блока в той же позиции так, чтобы он был более похож на ранговый блок.	Для обнаружения ЦВЗ не требуется наличие исходного изображения		Bas[63]
Строка бит	Модификация выбранного в соответствии с ключом рангового блока в доменном пуле. Если надо встроить 1, поиск выполняется в одной части пула, если 0 - в другой части.	Наличие секретного ключа; Устойчивость к сжатию JPEG		Puate [61]
Строка бит	Модификация вручную выбиранных двух неперекрывающиеся квадратных областей на изображении, называемые ранговой и доменной областью.	Наличие секретного ключа	Возможно заметное ухудшение качества изображения при встраивании ЦВЗ	Davern [50]

Рассмотрим более подробно известные алгоритмические решения для создания ЦВЗ для контейнеров-изображений. В работе [21] рассматриваются следующий ниже перечень базовых алгоритмов создания ЦВЗ для данных графических форматов, основанных на модификации файлов в исходном представлении (пространственно-временной подход, таблица 1.1.). Их преимуществом является то, что для внедрения ЦВЗ нет необходимости выполнять вычислительно громоздкие линейные преобразования изображений, а ЦВЗ внедряется за счет манипуляций яркостью $l(x,y) \in \{1,\ldots,L\}$ или цветовыми составляющими $(r(x,y), b(x,y), g(x,y))$.

Алгоритм Катера (Kutter) [68]. Пусть изображение имеет RGB-кодировку. Встраивание выполняется в канал синего цвета, так как к синему цвету система человеческого зрения наименее чувствительна. Рассмотрим алгоритм передачи одного бита секретной информации.

Пусть s_i - встраиваемый бит, $I = \{R, G, B\}$ - контейнер, $p = (x, y)$ - псевдослучайная позиция, в которой выполняется вложение. Секретный бит встраивается в канал синего цвета путем модификации яркости

$$l(p) = 0.299r(p) + 0.587g(p) + 0.114b(p)$$

$$b'(p) = \begin{cases} b(p) + ql(p), \text{ если } s_i = 0, \\ b(p) - ql(p), \text{ если } s_i = 1, \end{cases}$$

где q - константа, определяющая энергию встраиваемого сигнала. Ее величина зависит от предназначения схемы. Чем больше q, тем выше робастность вложения, но тем сильнее его заметность.

Извлечение бита получателем осуществляется без наличия у него исходного изображения, то есть вслепую. Для этого выполняется предсказание значения исходного, не модифицированного пикселя на основании значений его соседей. Предлагается для получения оценки пикселя использовать значения нескольких пикселей, расположенных в том же столбце и той же строке. Авторы использовали «крест» пикселей размером 7×7. Оценка $b''(p)$ получается в виде

$$b''(p)=\frac{1}{4c}\left(-2b''(p)\sum_{i=-c}^{+c}b''(x+i,y)+\sum_{k=-c}^{+c}b''(x,y+k)\right),$$

где c - число пикселей сверху (снизу, слева, справа) от оцениваемого пикселя. Так как в процессе встраивания ЦВЗ каждый бит был повторен cr раз, то мы получим cr оценок одного бита ЦВЗ. Секретный бит находится после усреднения разности оценки пикселя и его реального значения

$$\delta=\frac{1}{cr}\sum_{i=1}^{cr}\tilde{b}(p)-b_i(p).$$

Знак этой разности определяет значение встроенного бита. Алгоритм является робастным ко многим из известных атак: низкочастотной фильтрации изображения, его сжатию в соответствии с алгоритмом JPEG, обрезанию краев.

Алгоритм Брундокса (Bruyndonckx) [58]. ЦВЗ представляет собой строку бит. Для повышения помехоустойчивости применяется код БЧХ (код исправляющий кратные ошибки). Внедрение осуществляется за счет модификации яркости блока 8×8 пикселей.

Процесс встраивания осуществляется в три этапа.

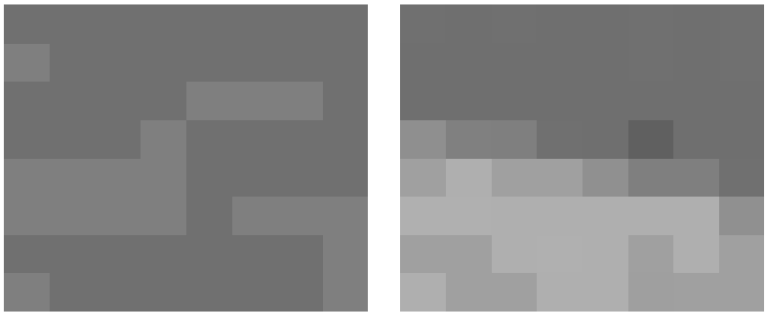
1. Классификация, или разделение пикселей внутри блока на две группы с примерно однородными яркостями.

2. Разбиение каждой группы на категории, определяемые данной сеткой.

3. Модификация средних значений яркости каждой категории в каждой группе.

Рассмотрим подробнее каждый из этих этапов.

1. При классификации авторы выделяют два типа блоков: блоки с «шумовым контрастом» (рисунок 1.4а) и блоки с резко выраженными перепадами яркости (рисунок 1.4б).

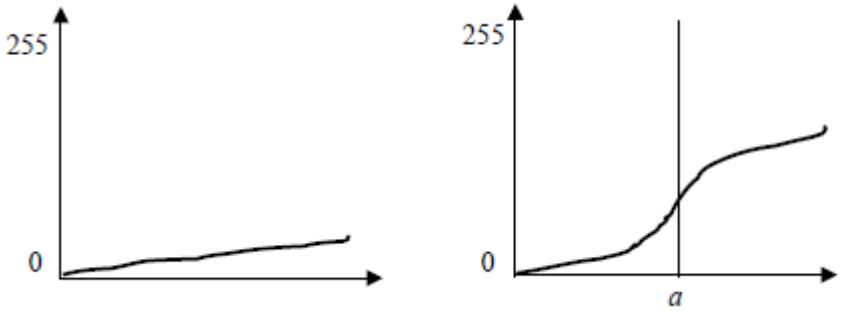


а) б)

Рисунок 1.4 – Два типа блоков: а) с нечетким контрастом и б) с резко выраженным контрастом

В блоках второго типа зоны с отличающейся яркостью не обязательно должны располагаться вплотную друг к другу, не обязательно должны содержать равное количество пикселей. Более того, некоторые пиксели вообще могут не принадлежать ни одной зоне. В блоках первого типа классификация особенно затруднена.

Для выполнения классификации значения яркости сортируются по возрастанию (рисунок 1.5а и б). Далее находится точка, в которой наклон касательной к получившейся кривой максимален (α). Эта точка является границей, разделяющей две зоны в том случае, если наклон больше некоторого порога. В противном случае пиксели делятся между зонами поровну.



а) б)

Рисунок 1.5 – Отсортированные значения яркостей блоков

2. Для сортировки пикселей по категориям на блоки накладываются маски, разные для каждой зоны и каждого блока. Назначение масок состоит в обеспечении секретности внедрения. Пример масок для двух зон приведен на рисунке 1.6а и б.

A	A	B	B	A	A	B	B
A	A	B	B	A	A	B	B
B	B	A	A	B	B	A	A
B	B	A	A	B	B	A	A
A	A	B	B	A	A	B	B
A	A	B	B	A	A	B	B
B	B	A	A	B	B	A	A
B	B	A	A	B	B	A	A

а)

B	B	B	B	A	A	A	A
B	B	B	B	A	A	A	A
B	B	B	B	A	A	A	A
B	B	B	B	A	A	A	A
A	A	A	A	B	B	B	B
A	A	A	A	B	B	B	B
A	A	A	A	B	B	B	B
A	A	A	A	B	B	B	B

б)

Рисунок 1.6 – Примеры используемых масок

3. Модификация. Множество пикселей оказалось разделенным на пять подмножеств: две зоны каждой из двух категорий и пиксели, не принадлежащие какой-либо зоне (для блоков первого типа). Обозначим среднее значение яркости для пикселей двух зон и категорий через $l_{1A}, l_{2A}, l_{1B}, l_{2B}$. Известно, что $l_{1A} < l_{2A}, l_{1B} < l_{2B}$. Встраивание бита ЦВЗ s осуществляется по следующему правилу:

$$s = \begin{cases} 1, \begin{cases} l'_{1A} > l'_{1B}, \\ l'_{2A} > l'_{2B}, \end{cases} \\ 0, \begin{cases} l'_{1A} < l'_{1B}, \\ l'_{2A} < l'_{2B}. \end{cases} \end{cases} \tag{1.1}$$

С другой стороны, необходимо обеспечить равенство значений яркости в каждой зоне

$$\frac{n_{1A} l'_{1A} + n_{1B} l'_{1B}}{n_{1A} + n_{1B}} = l_1 \text{ и } \frac{n_{2A} l'_{2A} + n_{2B} l'_{2B}}{n_{2A} + n_{2B}} = l_2 . \tag{1.2}$$

Для достижения этого яркость всех пикселей одной зоны меняется одинаково. Например, для зоны l, категории A это изменение составит $l'_{1A} - l_{1A}$.

Алгоритм извлечения ЦВЗ является обратным алгоритму внедрения. При этом вычисляются средние значения яркостей и находятся разности

$$s'' = \begin{cases} 0, \textit{если } l''_{1A} - l''_{1B} < 0 \textit{ и } l''_{2A} - l''_{2B} < 0, \\ 1, \textit{если } l''_{1A} - l''_{1B} > 0 \textit{ и } l''_{2A} - l''_{2B} > 0. \end{cases} \tag{1.3}$$

Алгоритм Ленгелаара (Langelaar) [69]. Данный алгоритм также работает с блоками 8×8. Вначале создается псевдослучайная маска нулей и единиц такого

же размера $pat(x,y) \in \{0,1\}$. Далее каждый блок B делится на два субблока B_0 и B_1, в зависимости от значения маски. Для каждого субблока вычисляется среднее значение яркости, l_0 и l_1. Далее выбирается некоторый порог α, и бит ЦВЗ встраивается следующим образом:

$$s = \begin{cases} 1, \text{если } l_0 - l_1 > +\alpha, \\ 0, \text{если } l_0 - l_1 < -\alpha. \end{cases} \tag{1.4}$$

Если условие (1.4) не выполняется, мы изменяем значение яркости пикселей субблока B_1. Для извлечения бита ЦВЗ вычисляются средние значения яркости субблоков - l_0'' и l_1''. Разница между ними позволяет определить искомый бит

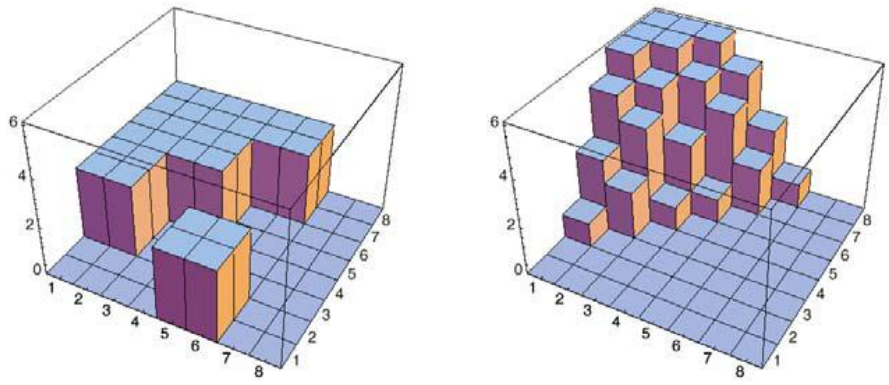
$$s = \begin{cases} 1, l_0 - l_1 > 0, \\ 0, \ l_0 - l_1 < 0. \end{cases}$$

Алгоритм Питаса (Pitas) [73]. ЦВЗ представляет собой двумерный массив бит размером с изображение, причем число единиц в нем равно числу нулей. Существует несколько версий алгоритма, предложенного Питасом. Вначале предлагалось встраивать бит ЦВЗ в каждый пиксель изображения, но позже благоразумно было решено использовать для этой цели блоки размером 2×2 или 3×3 пикселя, что делает алгоритм более робастным к сжатию или фильтрации. ЦВЗ складывается с изображением

$$l'(x,y) = l(x,y) + \alpha s(x,y).$$

В случае использования для внедрения блоков детектор ЦВЗ вычисляет среднее значение яркости этого блока. Отсюда появляется возможность неравномерного внедрения ЦВЗ в пиксели, то есть величина $\alpha \neq const$. Таким образом можно получить ЦВЗ, оптимизированный по критерию робастности к процедуре сжатия алгоритмом JPEG. Для этого в блоке 8×8 элементов заранее вычисляют «емкость» каждого пикселя. Затем ЦВЗ внедряют в соответствии с вычисленной емкостью. Эта оптимизация производится раз и навсегда, и

найденная маска применяется для любого изображения. На рисунке 1.7а и б показан ЦВЗ до и после оптимизации.



а) б)

Рисунок 1.7 – ЦВЗ а) до оптимизации и б) после оптимизации

Алгоритм Роджена (Rongen) [71]. Так же, как и в предыдущем алгоритме, ЦВЗ представляет собой двумерную матрицу единиц и нулей с примерно равным их количеством. Пиксели, в которые можно внедрять единицы (то есть робастные к искажениям), определяются на основе некоторой характеристической функции (характеристические пиксели). Эта функция вычисляется локально, на основе анализа соседних пикселей. Характеристические пиксели составляют примерно $1/100$ от общего числа, так что не все единицы ЦВЗ встраиваются именно в эти позиции. Для повышения количества характеристических пикселей в случае необходимости предлагается осуществлять небольшое искажение изображения.

Детектор находит значения характеристических пикселей и сравнивает с имеющимся у него ЦВЗ. Если в изображении ЦВЗ не содержится, то в характеристических пикселях количество единиц и нулей будет примерно поровну. Авторы рассчитали значение порога принятия решения, минимизирующего вероятность ложной тревоги.

Алгоритм PatchWork. В основе алгоритма Patchwork лежит статистический подход. Вначале псевдослучайным образом на основе ключа выбираются два пикселя изображения. Затем значение яркости одного из них увеличивается на некоторое значение (от 1 до 5), значение яркости другого –

уменьшается на то же значение. Далее этот процесс повторяется большое число раз (≈ 10000) и находится сумма значений всех разностей. По значению этой суммы судят о наличии или отсутствии ЦВЗ в изображении.

Для пояснения работы алгоритма введем ряд обозначений. Пусть значения выбираемых на каждом шаге пикселей a_i и b_i, величина приращения - δ. Тогда сумма разностей значений пикселей

$$S_n = \sum_{i=1}^{n}\left[\left(a_i + \delta\right) - \left(b_i - \delta\right)\right] = 2\delta n + \sum_{i=1}^{n}\left(a_i - b_i\right)$$

Математическое ожидание величины $\sum_{i=1}^{n}\left(a_i - b_i\right)$ стремится к нулю при достаточно большом n. Математическое ожидание величины S_n будет больше 2δ. Показано, что S_n имеет гауссовское распределение. Таким образом, в стегодетекторе в соответствии с ключом проверяется значение S_n и в том случае, если она значительно отличается от нуля, выносится решение о наличии ЦВЗ.

Алгоритм Patchwork является достаточно стойким к операциям сжатия изображения, его усечения, изменения контрастности. Основным недостатком алгоритма является его неустойчивость к аффинным преобразованиям, то есть поворотам, сдвигу, масштабированию. Другой недостаток заключается в малой пропускной способности. Так, в базовой версии алгоритма для передачи 1 бита скрытого сообщения требуется 20000 пикселей.

Алгоритм Бендера (Bender) [51]. Это алгоритм, основанный на копировании блоков из случайно выбранной текстурной области в другую, имеющую сходные статистические характеристики. Это приводит к появлению в изображении полностью одинаковых блоков. Эти блоки могут быть обнаружены следующим образом:

1. Анализ функции автокорреляции стегоизображения и нахождение ее пиков.

2. Сдвиг изображения в соответствии с этими пиками и вычитание изображения из его сдвинутой копии.

3. Разница в местоположениях копированных блоков должна быть близка к нулю. Поэтому можно выбрать некоторый порог и значения, меньшие этого порога по абсолютной величине, считать искомыми блоками.

Так как копии блоков идентичны, то они изменяются одинаково при преобразованиях всего изображения. Если сделать размер блоков достаточно большим, то алгоритм будет устойчивым по отношению к большинству из негеометрических искажений. В проведенных экспериментах показана робастность алгоритма к фильтрации, сжатию, поворотам изображения.

Основным недостатком алгоритма является исключительная сложность нахождения областей, блоки из которых могут быть заменены без заметного ухудшения качества изображения. Кроме того, в данном алгоритме в качестве контейнера могут использоваться только достаточно текстурные изображения.

Один из предложенных способов для проверки аутентичности изображений получил название метода проверочных сумм. Согласно этому методу отбирались семь старших бит восьми близлежащих пикселей. Получалось 56-битное слово. Выполнив эту операцию для всего изображения, имели $N \times N/8$ таких слов, где $N \times N$ - число пикселей в изображении. Затем они поразрядно складывались по модулю два, то есть вычислялась проверочная сумма длиной 56 бит. Эта сумма записывалась в младшие значащие биты, выбранных в соответствии с ключом, пикселей. В детекторе осуществлялась проверка этих бит, получившаяся проверочная сумма сравнивалась с эталонной, и выносилось решение о наличии или отсутствии модификации изображения. Таким образом, в данном алгоритме в качестве ключа использовались местоположение несущих проверочную сумму пикселей и само значение этой проверочной суммы.

Существующие на данный момент алгоритмы встраивания ЦВЗ в цифровые изображения обладают общим недостатком – большинство из них осуществляет встраивание в частотной области [74, 75], т.е. необходимыми этапами встраивания ЦВЗ в изображение являются прямое и обратное ДПФ всего изображения-контейнера. Данный подход, обеспечивая устойчивость встроенного ЦВЗ к повороту, масштабированию и кадрированию изображения-контейнера

(так называемый RST-invariant watermarking), в то же время значительно повышает вычислительную сложность процедуры встраивания и делает невозможным поблочное встраивание ЦВЗ [23].

В работе [23] авторами основное внимание уделяется алгоритмам создания стойких ЦВЗ для цифровых изображений и проводится следующая классификация:

- алгоритмы, основанные на встраивании ЦВЗ в области преобразования, инвариантного к заданным типам искажений, предполагающих встраивание ЦВЗ непосредственно в области преобразования (transform domain) [67, 70, 74, 75];
- алгоритмы встраивания ЦВЗ, использующие так называемые устойчивые к искажениям «метки синхронизации» (локальные экстремумы в заданных областях энергетического спектра изображения), для оценки и компенсации искажений изображения-контейнера [77, 78].

Основными недостатками первого типа алгоритмов являются вычислительная сложность и значительное искажение изображения-контейнера в процессе встраивания при вычислении прямого и обратного преобразования Фурье-Меллина. Кроме того, большинство алгоритмов данного класса используют так называемый «закрытый» детектор [23] и не позволяют встраивать в качестве ЦВЗ произвольную, не известную получателю, последовательность бит. Основным недостатком второго типа алгоритмов является их неустойчивость к атаке «watermark template attack» [23].

Кроме того, в работе [23] предложен новый алгоритм встраивания цифровых водяных знаков (ЦВЗ) в файлы графического формата, обеспечивающий устойчивость ЦВЗ к широкому набору преобразований изображения-контейнера и обладающий рядом преимуществ по сравнению с существующими алгоритмами встраивания стойких ЦВЗ. Представленный алгоритм позволяет встраивать ЦВЗ в файлы графического формата при помощи поблочной обработки исходного контейнера и не требующий выполнения спектральных преобразований при встраивании ЦВЗ. Данный алгоритм, как и

алгоритмы, основанные на спектральных преобразованиях данных контейнера, обеспечивает устойчивость ЦВЗ к повороту, масштабированию, кадрированию изображения-контейнера, а также к ряду других распространённых преобразований (сжатию с потерями, аддитивному зашумлению, линейной фильтрации и т.д.).

Как уже упоминалось, основной проблемой при реализации технологий создания ЦВЗ, является сохранение качества маркируемых при внедрении ЦВЗ файлов при их использовании по основному назначению в сочетании с достоверностью дальнейшего восстановления ЦВЗ. Единого стандарта и универсальной технологии создания ЦВЗ не существует. Компании - разработчики предлагают различные варианты, начиная от фактически видимых ЦВЗ, отвлекающих внимание и как следствие закрывающих смысловую часть данных, и заканчивая невидимыми цифровыми метками, для восстановления которых необходимы специальные программные средства. Одними из практических примеров использования технологии ЦВЗ являются продукты компании Digimarc, которая реализует их в виде плагинов к различным графическим редакторам (Adobe, HP, Philips и т.д.). ЦВЗ нанесённые продуктами Digimarc могут выдержать небольшую компрессию в формат JPEG [6, 37]. При этом модели и алгоритмы нанесения маркировок и алгоритмов генерации ЦВЗ Digimarc [11] неизвестны, и не подтверждены математическим аппаратом.

Основным недостатком известных алгоритмических решений является то, что в большинстве перечисленных алгоритмов создания ЦВЗ используются процедуры, реализующие строго определенные последовательности операций преобразования данных, в конечном счете, образующие конечное множество вариантов, которые могут быть идентифицированы и вскрыты. К недостаткам многих алгоритмов встраивания ЦВЗ также можно отнести жесткую зависимость от формата файла, трудоемкость вычисления цифровых водяных знаков, сравнительно низкую стойкость по отношению к операциям преобразования сигналов, возможное заметное ухудшение качества маркируемого файла (например, изображения) при встраивании метки, и, кроме того, в ряде

алгоритмов для извлечения цифрового водяного знака необходимо иметь исходный (немаркированный) файл.

В определенной степени преодолеть недостатки можно с использованием возможностей искусственных нейронных сетей (ИНС) [13, 43, 55, 75], используемых для создания универсальных функциональных моделей преобразования данных.

В работе [76] предложено использовать новый алгоритм создания цифровых водяных знаков на основе нейронных сетей. ЦВЗ незаметно встраивается в цветное изображение по аналогии с методом Катера. Затем с использованием нейронных сетей выборочно распознаются характеристики встроенных битов ЦВЗ на основе различий между непосредственно встроенных пикселей и этих же пикселей в симметричном крестообразном окне. Кроме того, представлены результаты проведенных экспериментов, иллюстрирующих стойкость алгоритма против различных типов атак.

В работе [54] рассматривается возможность применения нейронных сетей при встраивании и восстановлении ЦВЗ. Предлагаемая методика основывается на объединении двух подходов [53,63]. Изначально исходный ЦВЗ разделяется на блоки, для каждого из которых проводится дискретно-косинусное преобразование в частотной области. Перед встраиванием ЦВЗ случайным образом рассеивается, что делает процесс обнаружения ЦВЗ достаточно трудоемким, при этом коэффициенты ДКП также неупорядочены. Затем коэффициенты ДКП группируются в различные подмножества и кодируются в шести разрядные двоичные последовательности. После чего происходит непосредственно встраивание ЦВЗ в исходный файл графического формата. Для извлечения ЦВЗ используется специально обученная нейронная сеть, на выходе их которой получают шестиразрядные последовательности бит, которые преобразуются обратно в десятичное представление. Результаты проведенных экспериментов (различный уровень размытости, резкости, JPEG сжатие с потерями) показали, что извлеченный ЦВЗ имеет некоторые незначительные артефакты, что несколько влияет на надежность и незаметность предложенного метода встраивания ЦВЗ.

В работе [34] были разработаны методы обеспечения безопасности использования информационных технологий. Предложенные алгоритмы, базирующиеся на методах стеганографии, могут быть использованы, например, для анализа и фильтрации передаваемого трафика в сети, с целью пресечения утечки коммерческой информации предприятия. Первый разработанный метод стегоанализа базируется на выявлении статистической зависимости свойств заполненного и анализируемого контейнеров. Зависимость может быть выявлена обычным архиватором. На сегодняшний день данная схема стегоанализа является наиболее эффективной по сравнению с существующими аналогами. Второй метод обнаружения скрытой информации основан на статистической особенности встраиваемого сообщения, которое выглядит как случайная последовательность. Проверяя извлеченное из подозрительного контейнера сообщение на случайность, разработанный метод стегоанализа позволяет обнаруживать наличие внедрения с высокой точностью. Третий разработанный метод стегоанализа, также использует сжатие для выявления факта передачи секретного сообщения в исполняемых файлах. Предложенная схема обладает высокой эффективностью работы и, в отличие от имеющихся аналогов, не требует дизассемблировать программу. Четвертый разработанный метод, использующий внедрение стеганографических меток в исходные тексты программ, может быть использован для построения систем защиты авторских прав.

В работе [18] разработана модель предсказателя значений среднечастотных коэффициентов ДКП матрицы, отличающаяся использованием аппарата искусственных нейронных сетей, позволяющая оценить искажения изображения в процессе внедрения в него ЦВЗ с учетом нелинейности зависимости значений ДКП коэффициентов. Кроме того разработан алгоритм выбора способа изменения ДКП коэффициентов при внедрении цифровых водяных знаков в графические изображения па основе искусственных нейронных сетей и генетических алгоритмов, отличающийся изменением при внедрении бита ТДЗЗ всех ДКП коэффициентов, расположенных в пределах одной диагонали ДКП матрицы отобранного для этого блока изображения, и учетом при этом значений остальных

ДКП коэффициентов, позволяющий осуществить выбор субоптимального способа изменения ДКП коэффициентов.

Главная особенность рассматриваемого в данной работе подхода заключается в том, что комплекс специально обученных нейронных сетей используется для реализации как скрывающего, так и восстанавливающего ЦВЗ преобразований. Как показывают предварительные исследования, именно функциональный характер выполняемых с помощью нейронных сетей преобразований (в отличие от классического алгоритмического подхода) и применение искусственных нейронных сетей различных типов для реализации технологий цифровых водяных знаков, позволяет обеспечить меньшую «прозрачность» процесса встраивания и хороший компромисс между визуальной незаметностью, устойчивостью внедрения цифровых меток и объемом используемых вычислительных ресурсов. Предлагаемые математические модели и реализуемые на их основе принципы построения алгоритмов встраивания ЦВЗ отличаются от известных тем, что при их использовании обеспечивается гибкая и управляемая модификация как амплитудных характеристик файла-контейнера, так и статистических взаимосвязей между его элементами с минимальными результирующими искажениями. Аппарат искусственных нейронных сетей, как показывают выполненные исследования, позволяет реализовать алгоритмы, при использовании которых процесс встраивания данных в файл-контейнер носит существенно менее искажающий и более устойчивый характер. Для восстановления ЦВЗ и идентификационных номеров повышенной скрытности используется нейросетевые алгоритмы статистической классификации элементов ранее скрытой последовательности данных, которые обучаются с учетом особенностей реализованной процедуры встраивания. Функциональные возможности используемых нейросетевых алгоритмов обработки информации ориентированы, прежде всего, на минимизацию искажений файлов-контейнеров, представленных в форматах высокого качества, и обеспечение меньшей «прозрачности» процесса создания ЦВЗ, что позволяет разработать новую высоконадежную технологию создания ЦВЗ, не оказывающую влияния на

визуальное восприятие объекта.

Первоначально данный подход рассматривался в работах [13,43]. Показано, что при использовании предложенных в этих работах алгоритмов обработки информации процесс встраивания данных в файл-контейнер носит существенно менее прозрачный характер, чем в большинстве известных алгоритмов стеганографии. Данный подход позволил, в частности, реализовать создание ЦВЗ в файлах данных вещественных форматов. Однако предложенные в [13,43] алгоритмы встраивания не являются достаточно эффективными для защиты объектов, имеющих целочисленные форматы представления данных, и не позволяют реализовать универсальные алгоритмы и технологии создания ЦВЗ в файлах графических и звуковых форматов представления данных.

Выводы по главе

1. На основе известных публикаций отечественных и зарубежных авторов выполнено систематизированное изложение методов компьютерной стеганографии и их сравнительная характеристика. Определены основные понятия и единая терминология в области компьютерной стеганографии. Рассматриваются основные положения современной компьютерной стеганографии и классификация методов компьютерной стеганографии. Представлены основные области использования методов и алгоритмов компьютерной стеганографии.

2. Рассмотрены способы применения компьютерной стеганографии для создания ЦВЗ. Отмечено, что использование ЦВЗ для подтверждения авторских прав в настоящее время приобретают все большую популярность. Кратко рассмотрены и проанализированы основные особенности технологий цифровых водяных знаков и требования, предъявляемые к этим технологиям. Рассмотрена схема стегосистемы, предназначенная для встраивания ЦВЗ в цифровые изображения. Приводится подробная классификация и результаты анализа современных алгоритмов встраивания ЦВЗ в файлы–изображения при этом выделяются класс аддитивных алгоритмов, класс алгоритмов, на основе слияния ЦВЗ и контейнера, и класс алгоритмов на основе квантования.

3. Подробно рассматриваются основные алгоритмы создания ЦВЗ для данных графических форматов (изображений), основанных на модификации файлов в исходном представлении. Отмечено, что основной проблемой при реализации технологий создания ЦВЗ, является сохранение качества маркируемых при внедрении ЦВЗ файлов при их использовании по основному назначению в сочетании с достоверностью дальнейшего восстановления ЦВЗ. Основным недостатком известных решений является то, что в большинстве перечисленных алгоритмов создания ЦВЗ используются процедуры, реализующие строго определенные последовательности операций преобразования данных, в конечном счете, образующие конечное множество вариантов, которые могут быть идентифицированы и вскрыты.

4. Как показывает анализ, определенные перспективы для создания ЦВЗ повышенной скрытности и устойчивости, а также для реализации универсальных алгоритмов создания ЦВЗ, предоставляют возможности искусственных нейронных сетей. В этом плане рассмотрены результаты исследований, посвященных вопросам использования ИНС для создания ЦВЗ. При этом в качестве базового подхода предложено рассматривать подход, основанный на использовании нейросетевых функциональных моделей преобразования данных, главная особенность которого заключается в том, что комплекс специально обученных нейронных сетей используется для реализации как скрывающего и восстанавливающего ЦВЗ преобразований.

2. Нейросетевые функциональные модели преобразования данных и их применение для разработки универсальных алгоритмов создания цифровых водяных знаков

Как уже отмечалось в 1 главе, большинство известных и ранее рассмотренных алгоритмов ССИ имеет один общий недостаток: используемые в них процедуры встраивания информации реализуют строго определенные последовательности операций преобразования данных, образующие конечное множество вариантов, которые могут быть идентифицированы и воспроизведены сторонними лицами.

Представляет интерес рассмотрение возможности построения алгоритмов скрывающего и восстанавливающего преобразований для создания ЦВЗ на основе нейросетевых функциональных моделей, реализующих сжимающие отображения данных. Подобный подход, на наш взгляд может обеспечить ряд преимуществ перед известными, а именно:

- возможность создания универсальных алгоритмов, независящих от форматов обрабатываемых данных;
- обеспечение повышенной скрытности и устойчивости процедур встраивания и восстановления ЦВЗ, за счет сходимости во многих случаях нейросетевых алгоритмов обработки данных (оценивания и распознавания) к статистически оптимальным;
- реализацию менее прозрачных процедур ССИ, воспроизведение которых нарушителем, возможно, будет существенно затруднено.

В этом плане в настоящей главе рассмотрены теоретические обоснования возможности построения нейросетевых функциональных моделей и алгоритмов ССИ в интересах создания ЦВЗ при обработке данных, представленных в виде реализаций случайных векторов, отображающих фрагменты случайных процессов (как аналогов аудио файлов) или случайных полей (как аналогов видео файлов), а также исследования потенциальной эффективности выполняемой таким образом обработки.

2.1. Теоретические обоснования возможности построения сжимающих преобразований данных на основе нейронных сетей персептронного типа

Для построения алгоритмов обработки информации целесообразно рассмотреть теоретические обоснования возможности замены синтезируемых в рамках статистической теории решений алгоритмов их нейросетевыми эквивалентами. Подобный анализ применительно к задачам оценивания параметров случайных процессов и полей проведен в ряде известных работ [4, 16, 36, 41, 43, 44].

В качестве исходного положения рассмотрим следующий результат, представленный в работе [17]. Пусть x – ненаблюдаемый случайный вектор с априорной плотностью распределения $P(x)$, а $z = H(x) + v$ – наблюдаемый случайный вектор, для которого задана условная плотность распределения $P(z/x)$. Предположим, что на основе указанных плотностей распределения вероятностей может быть сгенерировано обучающее множество $\{x^{(p)}, z^{(p)} = H(x^{(p)}) + v^{(p)}, p = \overline{1,P}\}$, которое используется для построения оператора оценивания $x^{\circ}(z) = \Psi_W(z)$ с помощью многослойной нейронной сети. Здесь $\Psi_W(z)$– оператор, реализуемый нейронной сетью при произвольных значениях весовых коэффициентов W.

В ходе обучения нейронной сети осуществляются минимизация средней квадратичной ошибки, которую в данном случае запишем в виде

$$E(W) = \frac{1}{P}\sum_{p=1}^{P} e_p^T e_p = \frac{1}{P}\sum_{p=1}^{P}[x^{(p)} - \Psi_W(z^{(p)})]^T[x^{(p)} - \Psi_W(z^{(p)})] \underset{W}{\to} \min,$$

При достаточно общих предположениях существования математического ожидания величины $e^T e,\ e = x - \Psi_W(z)$ для произвольной совокупности весовых коэффициентов НС W величина $E(W)$ при $P \to \infty$ сходится по вероятности к своему математическому ожиданию

$$E(W) \underset{P\to\infty}{\to} M[e^T e],\ \ E(W) = M[e^T e] + O\left(\frac{1}{P}\right)$$

и, следовательно, к интегралу вида

$$\mathrm{M}[(x-\Psi_W(z))^T(x-\Psi_W(z))]=\int\int(x-\Psi_W(z))^T(x-\Psi_W(z))P(x,z)dxdz=$$

$$=\int J_W(z)dz=\int[\int(x-x^\circ(z))^T(x-x^\circ(z))P(x/z)dx]P(z)dz.$$

С другой стороны, как известно [45], необходимое условие минимума интеграла

$$E=\int\int(x-\tilde{x}(z))^T(x-\tilde{x}(z))P(x,z)dxdz=\int J(z)dz=$$

$$=\int[\int(x-\tilde{x}(z))^T(x-\tilde{x}(z))P(x/z)dx]P(z)dz,$$

где $\tilde{x}(z)$ – произвольная оценка x по наблюдениям z, удовлетворяет уравнениям

$$\frac{\partial J(z)}{\partial\tilde{x}}=0 \;\Rightarrow\; \tilde{x}(z)=\int xP(x/z)dx=\int x\frac{P(z/x)}{P(z)}dx.$$

Это означает, что при достижении минимума интеграла $J(z)$ вектор $\tilde{x}(z)$ является условным математическим ожиданием вектора x. Однако, поскольку, вид интегралов $J_W(z)$ и $J(z)$ совпадает, то при достижении минимума $J_W(z)$ и, соответственно, минимума предела $E(W)$ при $P\to\infty$ должно выполняться $x^\circ(z)=\tilde{x}(z)$. Вопрос существования многослойно нейронной сети, позволяющей воспроизвести любой заданный оператор $\tilde{x}(z)$ путем настройки весовых коэффициентов W, уже ранее обсуждался.

Представленные соображения позволяют сформулировать утверждение [17] о том, что при использовании нейронных сетей, обучаемых на основе данных, описываемых известными распределениями и формируемых путем выборки из генеральных совокупностей соответствующих случайных величин, обеспечивается сходимость по вероятности нейросетевых оценок к оценкам, получаемым на основе статистически оптимальных алгоритмов обработки информации.

Детальный сопоставительный анализ статистических и нейросетевых процедур при выполнении типовых преобразований данных может быть выполнен для нейронных сетей линейного типа. При отсутствии априорной информации построение преобразований данных для проведения процедур оценивания и сжатия данных удобно осуществлять в рамках нейросетевого

подхода. В [2, 42] показано, что использование аппарата НС позволяет определять пространственные связи между элементами сигналов и изображений в виде весов нейронной сети при обучении по эталонным фрагментам, т.е. фактически решать задачу восстановления регрессионной модели случайного процесса или случайного поля.

Общая методика получения указанных результатов основана на исследовании необходимых и достаточных условий минимума целевой функции – средней квадратичной ошибки представления данных на выходе НС. Отметим, что используемый целевой функционал для НС линейного типа является квадратичным относительно совокупности весов нейронных связей. Решение, минимизирующее этот функционал, может быть не единственным, но всегда существует. При этом гарантируется сходимость рекуррентных процедур поиска минимума, например градиентных алгоритмов обучения [15, 44, 46].

2.2. Восстановление параметров векторной регрессии при обучении нейронной сети по прямым и косвенным наблюдениям.

В работе [12] рассмотрена задача восстановления регрессионных моделей случайных процессов и полей с использованием нейронных сетей при их обучении по прямым наблюдениям.

Пусть $x=(x_1...x_M)^T$ – случайный вектор с математическим ожиданием $M[x]=m_x$ и матрицей ковариации $M[(x-m_x)(x-m_x)^T]=R_{xx}$, а $y=(y_1...y_N)^T$ – случайный вектор с математическим ожиданием $M[y]=m_y$ и матрицей ковариации $M[(y-m_y)(y-m_y)^T]=R_{yy}$. Пусть также связь между x и y определяется линейным оператором вида

$$y=Fx+c+v,$$

где $F=\|f_{nm}\|$ – матрица размера $N\times M$; c – векторная константа; $v=(v_1...v_N)^T$ – случайный, независимый от x и y, вектор возмущения ($M[v]=0$, $M[vv^T]=R_{vv}$). Фактически y и x подчиняются модели многомерной регрессии. При этом

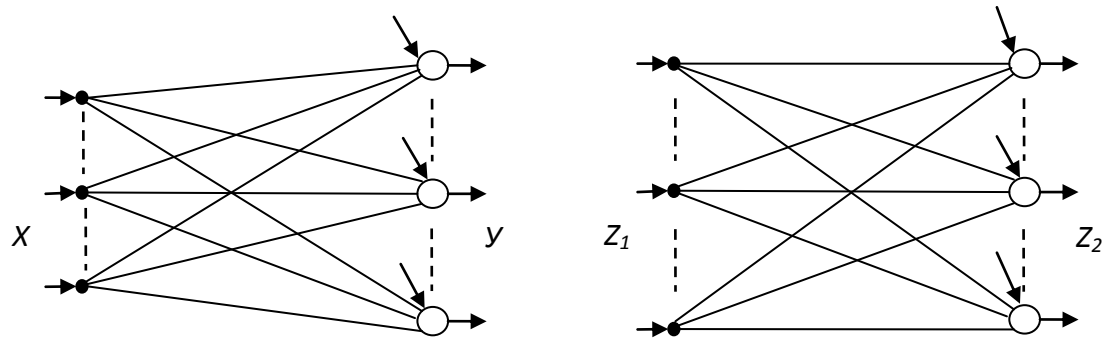
матрица F и вектор c с учетом независимости v и x, y связаны со статистическими характеристиками x, y и v следующим образом:

$$m_y = Fm_x + c,\ y - m_y = Fx + c - m_y - Fm_x + Fm_x + c = F(x - m_x) + v,$$

$$R_{yx} = M\left[(y - m_y)(x - m_x)^T\right] = FR_{xx}^{\ T},\ R_{yy} = M\left[(y - m_y)(y - m_y)^T\right] = FR_{xx}F^T + R_{vv}.$$

Если R_{xx} – невырожденная матрица (что, обычно, выполняется), то связь параметров модели векторной регрессии со статистическими характеристиками случайных векторов имеет вид

$$c = m_y - Fm_x,\ \ F = R_{yx}R_{xx}^{-1},\ R_{vv} = R_{yy} - FR_{xx}F^T. \qquad (2.1)$$



а) б)

Рисунок 2.1 - Однослойные линейные нейронные сети

В работе [12] было рассмотрено обучение однослойной нейронной сети (рисунок 2.1а) с линейными функциями активации и матрицей весов $W = \|w_{nm}\|$ размера $N \times M$ и вектором смещения $w_0 = (w_{10} \ldots w_{N0})^T$, для обучения сети используется выборка $\{x^{(p)}, y^{(p)}\}$, $p = \overline{1, P}$ и найдены необходимое и достаточное условия минимума целевой функции при обучении НС.

В результате было получено результирующее выражение для матрицы весов НС и вектора смещений

$$W = \tilde{R}_{yx}\tilde{R}_{xx}^{-1},\quad w_0 = \tilde{m}_y - W\tilde{m}_x, \qquad (2.2),$$

где оценки $\widetilde{R}_{yx}$, $\widetilde{R}_{xx}$ являются оценками максимального правдоподобия матриц ковариации R_{yx}, R_{xx}.

Таким образом, в [12] показано, что структура соотношений (2.1) для параметров F, c исходной модели и параметров W, w_0 (2.2) модели, восстанавливаемой с помощью НС, сохраняется, при этом w_0 и W является выборочными оценками. Следует отметить, что данный результат является известным [46] для оценок параметров многомерной линейной регрессии, получаемых в рамках традиционного статистического подхода.

Дополнительно, в [12] рассмотрена задача оценивания параметров векторной регрессии процесса по наблюдениям, получаемым в присутствии аддитивного шума.

Пусть $x_1 = (x_{1,1} \ldots x_{1,M})^T$ и $x_2 = (x_{2,1} \ldots x_{2,M})^T$ – случайные векторы. Пусть также $m_{x1} = M[(x_1]$, $M[(x_1 - m_{x1})(x_1 - m_{x1})^T] = R_{xx11}$, причем x_2 связан с x_1 матричным уравнением $x_2 = Fx_1 + Gu$, где F – матрица размера $M \times M$; G – матрица размера $M \times L$; u – вектор возмущения размерности L ($M[u] = m_u, M[uu^T] = R_{uu}$). Тогда

$$m_{x2} = Fm_{x1}, \quad R_{xx2} = FR_{xx11}F^T + GR_{uu}G^T.$$

Векторы x_1 и x_2 являются ненаблюдаемыми векторами состояний системы. Наблюдению доступны векторы $z_1 = (z_{1,1} \ldots z_{1,N})^T$ и $z_2 = (z_{2,1} \ldots z_{2,N})^T$

$$z_1 = Hx_1 + v_1, \quad z_2 = Hx_2 + v_2,$$

где H – матрица размера $N \times M$; v_1, v_2 – шумы наблюдений, $M[v_1] = M[v_2] = 0$, $M[v_1 v_1^T] = M[v_2 v_2^T] = R_{vv}$, $M[v_1 v_2^T] = 0$.

Из теоремы о нормальной корреляции следует, что оптимальная в среднеквадратичном оценка x_1 по наблюдению z_1 имеет вид

$$x_{1/1} = m_{x1} + R_{xz11}R_{zz11}^{-1}(z_1 - \widetilde{z}_{1/0})^T, \quad \widetilde{z}_{1/0} = m_{z1} = Hm_{x1},$$

$$R_{xz11} = M[(x_1 - m_{x1})(z_1 - m_{z1})^T] = R_{xx11}H^T,$$

$$P_{zz11} = M[(z_1 - m_{z1})(z_1 - m_{z1})^T] = HR_{xx11}H^T + R_{vv}.$$

а оптимальная оценка x_2 по наблюдению z_1 равна $\tilde{x}_{2/1} = F\tilde{x}_{1/1}$. Матрицы ошибок этих оценок имеют следующий вид:

$$P_{x1/1} = P_{xx11}{}^{T} - R_{xz11}R_{zz11}^{-1}R_{zx11}, \quad P_{x2/1} = FP_{xx11}F^{T} + GR_uG^{T}, \; R_{zx11} = R_{xz11}^{T}.$$

При этом оптимальная оценка z_2, то есть прогноз наблюдений на один шаг вперед, имеет вид

$$\tilde{z}_{2/1} = H\tilde{x}_{2/1} = HF\tilde{x}_{1/1} = HF(m_{x1} + R_{xx11}H^{T}(HP_{xx11}H^{T} + R_{vv})^{-1}(z_1 - Hm_{x1}). \quad (2.3)$$

Матрица ковариации ошибки такой оценки равна

$$P_{z2/1} = HR_{x2/1}H^{T} + R_{vv}.$$

Можно показать [12], что при обучении однослойной линейной НС по совокупности $\{z_1^{(p)}, z_2^{(p)}\}$, $p = \overline{1,P}$ получаемые оценки при $P \to \infty$ сходятся по вероятности к теоретическим оценкам (учитывая $m_{x1} = m_{x2} = 0$). Рассматривается однослойная сеть (рисунок 2.1б) с матрицей весов W размера $N \times N$ (смещение также для упрощения записи в явном виде не вводится). Целевая функция при обучении сети имеет вид

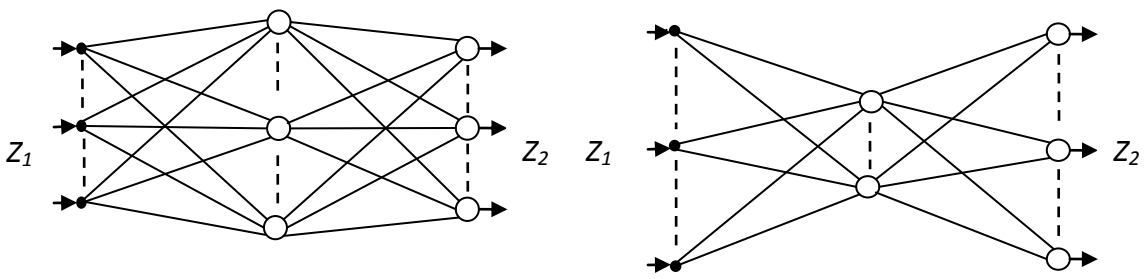
$$E = \frac{1}{2}\sum_{p=1}^{P}(z_2^{(p)} - Wz_1^{(p)})^{T}(z_2^{(p)} - Wz_1^{(p)}).$$

Из представленных результатов следует, что $W = \tilde{R}_{zz21}\tilde{R}_{zz11}^{-1}$, т.е. сеть реализует структуру оптимальной оценки. Получаемую при выполнении преобразования $\bar{z}_{2/1} = \tilde{R}_{zz21}\tilde{R}_{zz11}^{-1}z_1$ оценку z_2 по наблюдению z_1 называется выборочной оптимальной линейной оценкой. На основе неравенства Чебышева, а также выражений для первого и второго моментов слагаемых $\tilde{R}_{zz21}, \tilde{R}_{zz11}$, в [12] показано, что имеет место сходимость по вероятности

$$W \xrightarrow[P\to\infty]{} (HFP_{xx11}H^{T})(HP_{xx11}H^{T} + R_{vv})^{-1},$$

то есть преобразование, выполняемое после обучения нейронной сетью с весовой матрицей W, при $P \to \infty$ реализует близкую к оптимальной оценку (прогноз) вектора наблюдений z_2 по наблюдению z_1.

Представляет интерес решение более сложной задачи: получения алгоритма оценки ненаблюдаемого вектора x_2 путем косвенного обучения двухслойной нейронной сети только на основе выборки наблюдений $\{z_1^{(p)}, z_2^{(p)}\}$, $p = \overline{1,P}$. Данная задача решалась в работе [32] автором настоящей диссертации. При этом можно рассматривать двухслойную НС с одинаковым, расширенным или сокращенным количеством нейронов в скрытом слое (рисунок 2.2а и б) по отношению к выходному слою нейронов. Первый (скрытый) слой сети обучается, а второй фиксирован и определяется матрицей H известного вида. Матрицу весов первого слоя сети обозначим W, а общую матрицу выполняемого НС преобразования $W' = HW$.



а) б)

Рисунок 2.2 - Двухслойные линейные нейронные сети

Обозначим вектор сигналов в скрытом слое

$$\gamma = (\gamma_1 \ldots \gamma_M)^T = H\gamma\,,\ \gamma_k = \sum_{i=1}^{N} w_{ki} z_{1,i}\,,$$

а вектор сигналов на выходе НС

$$z_2 = H\gamma\,,\ \ z_{2,j} = \sum_{k=1}^{M} h_{jk}\gamma_k = \sum_{k=1}^{M}\sum_{i=1}^{N} h_{jk} w_{ki} z_{1,i}\,. \tag{2.4}$$

Тогда целевая функция, используемая для обучения НС, имеет вид

$$E = \frac{1}{2}\sum_{p=1}^{P}\sum_{j=1}^{N}\left(z_{2,j}^{(p)} - \sum_{k=1}^{M} h_{jk}\sum_{i=1}^{N} w_{ki} z_{1,i}^{(p)}\right)^2.$$

Рассмотрим необходимое условие минимума целевой функции:

$$\frac{\partial E}{\partial w_{mn}} = -\sum_{p=1}^{P}\sum_{j=1}^{N}\left(z_{2,j}^{(p)} - \sum_{k=1}^{M} h_{jk} \sum_{i=1}^{N} w_{ki} z_{1,i}^{(p)} \right) h_{jm} z_{1,n} =$$

$$\sum_{j=1}^{N} h_{jN}\left(\tilde{r}_{j,n}^{2,1} - \sum_{k=1}^{M} h_{jk} \sum_{i=1}^{N} w_{ki} \tilde{r}_{i,n}^{1,1} \right) = 0,\ m = \overline{1,M},\ n = \overline{1,N}.$$

В матричной форме получим эквивалентные уравнения

$$H^T H W \tilde{R}_{zz11} = H^T \tilde{R}_{zz21}.$$

Тогда, при условии, что $H^T H$ положительно определенная матрица, получим

$$W = (H^T H)^{-1} H^T \tilde{R}_{zz21} \tilde{R}_{zz11}^{-1}, \tag{2.5}$$

где $\tilde{R}_{zz21}$ и $\tilde{R}_{zz11}$ имеют вид (2.4).

Докажем теперь, что полученное решение является достаточным условием минимума целевой функции. Достаточное условие соответствует положительной определенности матрицы вторых частных производных по всем весовым коэффициентам обученной НС

$$P_W = \left\| \frac{\partial^2 E}{\partial w_{nm} \partial w_{nr}} \right\| > 0,$$

$$\frac{\partial^2 E}{\partial w_{nm} \partial w_{nr}} = -\frac{\partial}{\partial w_{nr}} \left(\sum_{p=1}^{P}\sum_{j=1}^{N}\left(z_{2,j}^{(p)} - \sum_{k=1}^{M} h_{jk} \sum_{i=1}^{N} w_{ki} z_{1,i}^{(p)} \right) h_{jm} z_{1,n}^{(p)} \right) =$$

$$= \sum_{j=1}^{N} h_{jm} \left(\sum_{k=1}^{M} h_{jk} \sum_{i=1}^{N} z_{1,i}^{(p)} z_{1,n}^{(p)} \right), \quad P_W = P H^T H \tilde{R}_{zz11}.$$

Таким образом, достаточное условие выполняется в силу положительной определенности $\tilde{R}_{zz1}$ и положительной определенности $H^T H$. Аналогично результатам [12], можно показать, что имеет место сходимость по вероятности

$$W \xrightarrow[P \to \infty]{} F P_{xx11} H^T (H P_{xx11} H^T + R_{vv})^{-1},$$

то есть сходимость к весовой матрице оптимальной оценки вектора x_2 по наблюдениям z_1. Это означает, что в скрытом слое при достаточно больших P формируется оптимальная в среднеквадратичном оценка x_2 по наблюдениям z_1. Данный результат свидетельствует о возможности косвенного обучения сети

указанной архитектуры по наблюдаемым параметрам для оценки ненаблюдаемых параметров x при известной матрице H, такой что $H^T H > 0$.

В работе [32] также рассмотрена более общая ситуация, когда совокупности реализаций $Z_1^P = \{z_1^{(1)},...,z_1^{(P)}\}$ и $Z_2^P = \{z_2^{(P)}...z_2^{(P)}\}$, как и ранее, используются в качестве обучающих данных при настройке весовых коэффициентов нейронной сети, имеющей архитектуру вида 2.2 а), при этом матрица весовых коэффициентов второго слоя не фиксируется, но число нейронов в скрытом слое в точности соответствует размерности векторов x_1 и x_2. Оказывается, что и в этом случае в условиях косвенного обучения на выходе нейронов скрытого слоя формируется оценка вектора y_2, связанного с z_2 соотношением

$$z_2 = HTy_2 + v_2,$$

где T – фиксированная матрица, которая может быть идентифицирована. Это означает, что экстраполированную оценку вектора x_2, обозначаемую $\tilde{x}_2$, по наблюдению z_1 можно найти из уравнения $\tilde{x}_2 = T\tilde{y}_2$, где $\tilde{y}_2 = W^{(1)} z_1$ – вектор, получаемый на выходе скрытого слоя нейронной сети рассматриваемой архитектуры.

Таким образом, исходя из представленных в п. 2.1.1 результатов, можно утверждать следующее: зная о существовании линейной зависимости, между наблюдаемыми и ненаблюдаемыми векторами, при настройке коэффициентов нейронной сети в условиях косвенного обучения по наблюдаемым данным при $P \to \infty$ имеет место сходимость по вероятности элементов матрицы преобразования для получения оценки x_2 по наблюдениям вектора z_1 к элементам матрицы, доставляющей оптимальную в классе линейных оценку.

2.3. Теорема о сходимости весов двухслойной линейной нейронной сети при построении оптимальных линейных оценок случайных векторов

Решение задач, связанных с обоснованием нейросетевых функциональных моделей и алгоритмов создания ЦВЗ, по нашему мнению, связано с возможностью реализации сжимающих отображений данных на основе линейных нейронных сетей прямого распространения, выполняемых по отношению к

областям определения функции произвольной конфигурации. Для этого в работе автора [32] выполнены теоретические обоснования, оформленные в виде соответствующей теоремы.

Рассмотрим двухслойную линейную сеть, архитектура которой представлена на рисунке 2.2б. Пусть в общем случае входной и выходной векторы имеют разную размерность: $z_1 = (z_{1,1},...,z_{1,N_1})^T \in R^{N_1}$ и $z_2 = (z_{2,1},...,z_{2,N_2})^T \in R^{N_2}$. Для обучения сети используются совокупности реализаций $Z_1^P = \{z_1^{(1)},...,z_1^{(P)}\}$ и $Z_2^P = \{z_2^{(P)}..z_2^{(P)}\}$. Особенностью архитектуры сети является использование сокращенного числа нейронов $M < N_1, M < N_2$ в скрытом слое по отношению к размерности входного и выходного векторов. Далее, для определенности, будем считать, что $m_{z2} = m_{z1} = 0$ и весовые коэффициенты смещения имеют нулевые значения. Общая матрица весов сети $W = \|w_{ij}\|$ теперь имеет следующий вид:

$$W = W^{(2)}W^{(1)},$$

$$W^{(1)} = \begin{pmatrix} w_{11}^{(1)} & \cdots & w_{1N_1}^{(1)} \\ \vdots & \ddots & \vdots \\ w_{M1}^{(1)} & \cdots & w_{MN_1}^{(1)} \end{pmatrix}, W^{(2)} = \begin{pmatrix} w_{11}^{(2)} & \cdots & w_{1M}^{(2)} \\ \vdots & \ddots & \vdots \\ w_{N_2 1}^{(2)} & \cdots & w_{N_2 M}^{(2)} \end{pmatrix}, \qquad (2.6)$$

где $W^{(1)}, W^{(2)}$ – матрицы весов первого и второго слоя НС. Обозначим вектор сигналов на выходе скрытого слоя

$$\gamma = (\gamma_1,...,\gamma_M)^T,\ \gamma = W^{(1)}z_1,\ \ \gamma_k = \sum_{i=1}^{N_1} w_{ki}^{(1)} z_{1,i},\ k = \overline{1,M},$$

тогда вектор сигналов на выходе сети определяется как

$$\tilde{z}_2 = W^{(2)}\gamma,\ \tilde{z}_{2,j} = \sum_{k=1}^{M} w_{jk}^{(2)}\gamma_k = \sum_{k=1}^{M} w_{jk}^{(2)} \sum_{i=1}^{N_1} w_{ki}^{(1)} z_{1,i},\ j = \overline{1,N_2}.$$

Целевая функция при обучении НС по совокупности реализаций на входе и выходе записывается в виде

$$E = \frac{1}{2}\sum_{p=1}^{P}\left(z_2^{(p)} - \tilde{z}_2^{(p)}\right)^T\left(z_2^{(p)} - \tilde{z}_2^{(p)}\right) = \frac{1}{2}\sum_{p=1}^{P}\sum_{j=1}^{N_2}\left(z_{2,j}^{(p)} - \sum_{k=1}^{M} w_{jk}^{(2)} \sum_{i=1}^{N_1} w_{ki}^{(1)} z_{1,i}^{(p)}\right)^2,$$

где $\tilde{z}_2^{(p)}$ - реакция НС на входное воздействие $z_1^{(p)}$.

Рассмотрим необходимое условие минимума целевой функции, которое определяется следующей системой уравнений:

$$\frac{\partial E}{\partial w_{mn}^{(1)}}=0,\ n=\overline{1,N_1},\ m=\overline{1,M},\ \frac{\partial E}{\partial w_{nm}^{(2)}}=0,\ n=\overline{1,N_2},\ m=\overline{1,Q}.$$

Сначала найдем значения компонентов матрицы $W^{(1)}$, исходя из следующих уравнений:

$$\frac{\partial E}{\partial w_{mn}^{(1)}}=\sum_{p=1}^{P}\sum_{j=1}^{N_2}\left(z_{2,j}^{(p)}-\sum_{k=1}^{M}w_{jk}^{(2)}\sum_{i=1}^{N_1}w_{ki}^{(1)}z_{1,i}^{(p)}\right)\frac{\partial}{\partial w_{mn}^{(1)}}\left(\sum_{k=1}^{M}w_{jk}^{(2)}\sum_{i=1}^{N_1}w_{ki}^{(1)}z_{1,i}^{(p)}\right)=$$

$$=\sum_{j=1}^{N_2}w_{jm}^{(2)}\left(\sum_{p=1}^{P}z_{2,j}^{(p)}z_{1,n}^{(p)}-\sum_{k=1}^{M}w_{jk}^{(2)}\sum_{i=1}^{N_1}w_{ki}^{(1)}\sum_{p=1}^{P}z_{1,i}^{(p)}z_{1,n}^{(p)}\right)=0,\ n=\overline{1,N_1},m=\overline{1,M}.$$

Обозначив $\frac{1}{P}\sum_{p=1}^{P}z_{2,j}^{(p)}z_{1,n}^{(p)}=r_{jn}^{(2,1)}$ и $\frac{1}{P}\sum_{p=1}^{P}z_{1,i}^{(p)}z_{1,n}^{(p)}=r_{in}^{(1,1)}$, перепишем последнее уравнение в виде

$$P\sum_{j=1}^{N_2}w_{jm}^{(2)}r_{jn}^{(2,1)}-P\sum_{j=1}^{N_2}w_{jm}^{(2)}\sum_{k=1}^{M}w_{jk}^{(2)}\sum_{i=1}^{N_1}w_{ki}^{(1)}r_{in}^{(1,1)}=0,\ n=\overline{1,N_1},\ m=\overline{1,M}.$$

В итоге в матричном виде получим следующие уравнения:

$$W^{(2),T}\tilde{R}_{z21}-W^{(2),T}W^{(2)}W^{(1)}\tilde{R}_{z11}=0 \text{ или } \tilde{R}_{z12}W^{(2)}-\tilde{R}_{z11}W^{(1),T}W^{(2).T}W^{(2)}=0, \quad (2.7)$$

где $\tilde{R}_{z21}=\left\|r_{jn}^{(2,1)}\right\|$, $\tilde{R}_{z11}=\left\|r_{in}^{(1,1)}\right\|$ – выборочные матрицы ковариации случайных векторов.

Найдем теперь соответствующие значения компонентов матрицы $W^{(2)}$, исходя из следующих уравнений:

$$\frac{\partial E}{\partial w_{nm}^{(2)}}=\sum_{p=1}^{P}\left(z_{2,n}^{(p)}-\sum_{k=1}^{M}w_{nk}^{(2)}\sum_{i=1}^{N_1}w_{ki}^{(1)}z_{1,i}^{(p)}\right)\frac{\partial}{\partial w_{nm}^{(2)}}\left(\sum_{k=1}^{M}w_{nk}^{(2)}\sum_{i=1}^{N_1}w_{ki}^{(1)}z_{1,i}^{(p)}\right)=$$

$$=\sum_{t=1}^{N_1}w_{mt}^{(1)}\left(\sum_{p=1}^{P}z_{1,t}^{(p)}z_{2,n}^{(p)}-\sum_{k=1}^{M}w_{nk}^{(2)}\sum_{i=1}^{N_1}w_{ki}^{(1)}\sum_{p=1}^{P}z_{1,t}^{(p)}z_{1,i}^{(p)}\right)=0,\ n=\overline{1,N_2},m=\overline{1,M}.$$

Обозначив, как и ранее, $\frac{1}{P}\sum_{p=1}^{P}z_{1,t}^{(p)}z_{2,n}^{(p)}=r_{tn}^{(1,2)}$ и $\frac{1}{P}\sum_{p=1}^{P}z_{1,t}^{(p)}z_{1,i}^{(p)}=r_{ti}^{(1,1)}$, перепишем последнее уравнение в виде

$$P\sum_{t=1}^{N_1} w_{mt}^{(1)} r_{tn}^{(1,2)} - P\sum_{t=1}^{N_1} w_{mt}^{(1)} \sum_{i=1}^{N} r_{ti}^{(1,1)} \sum_{k=1}^{M} w_{ki}^{(1)} w_{nk}^{(2)} = 0,\ n=\overline{1,N}_2,\ m=\overline{1,M}\ .$$

В матричной форме это уравнение записывается в виде

$$W^{(1)}\widetilde{R}_{z12} - W^{(1)}\widetilde{R}_{z11}W^{(1),T}W^{(2),T} = 0 \text{ или } \widetilde{R}_{z21}W^{(1),T} - W^{(2)}W^{(1)}\widetilde{R}_{z11}W^{(1),T} = 0, \quad (2.8)$$

где $\widetilde{R}_{z21} = \left\| r_{jn}^{(2,1)} \right\|$, $\widetilde{R}_{z11} = \left\| r_{in}^{(1,1)} \right\|$ – выборочные матрицы ковариации случайных векторов.

Будем далее предполагать, что объем обучающей выборки превышает N_1, N_2, что обеспечивает невырожденный характер матрицы $\widetilde{R}_{z11}$. Тогда для системы из двух матричных уравнений (2.7), (2.8)

$$\widetilde{R}_{z12}W^{(2)} - \widetilde{R}_{z11}W^{(1),T}W^{(2).T}W^{(2)} = 0,\quad W^{(2)} = \widetilde{R}_{z21}W^{(1),T}\left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1}, \quad (2.9)$$

получим

$$\widetilde{R}_{z12}\widetilde{R}_{z21}W^{(1),T} - \widetilde{R}_{z11}W^{(1),T}\left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1}W^{(1)}\widetilde{R}_{z12}\widetilde{R}_{z21}W^{(1),T} = 0.$$

Это уравнение эквивалентно также следующему:

$$R_{z11}^{-1}R_{z12}R_{z21}W^{(1),T} = W^{(1),T}L,\ L = \left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1}W^{(1)}R_{z12}R_{z21}W^{(1),T},$$

где L – матрица размера $M \times M$.

Проведем интерпретацию полученного решения. Для начала, представим величину ошибки, минимизируемой при обучении нейронной сети, в виде

$$E = \frac{1}{2}tr\left(\frac{1}{2}\sum_{p=1}^{P}(z_2{}^{(p)} - \widetilde{z}_2{}^{(p)})(z_2{}^{(p)} - \widetilde{z}_2{}^{(p)})^T\right) =$$

$$= \frac{1}{2}tr\left(\frac{1}{2}\sum_{p=1}^{P}(z_2{}^{(p)} - \bar{z}_{2/1}{}^{(p)} + \bar{z}_{2/1}{}^{(p)} - \widetilde{z}_2{}^{(p)})(z_2{}^{(p)} - \bar{z}_2{}^{(p)} + \bar{z}_2{}^{(p)} - \widetilde{z}_2{}^{(p)})^T\right) =$$

$$= \frac{1}{2}tr\left(\sum_{p=1}^{P}(z_2{}^{(p)} - \bar{z}_{2/1}{}^{(p)})(z_2{}^{(p)} - \bar{z}_{2/1}{}^{(p)})^T\right) + \frac{1}{2}tr\left(\sum_{p=1}^{P}(z_2{}^{(p)} - \widetilde{z}_2{}^{(p)})(z_2{}^{(p)} - \widetilde{z}_2{}^{(p)})^T\right) +$$

$$+\frac{1}{2}tr\left(\sum_{p=1}^{P}(z_2{}^{(p)}-\bar{z}_{2/1}{}^{(p)})(z_2{}^{(p)}-\tilde{z}_2{}^{(p)})^T\right)+\frac{1}{2}tr\left(\sum_{p=1}^{P}(z_2{}^{(p)}-\tilde{z}_2{}^{(p)})(z_2{}^{(p)}-\bar{z}_{2/1}{}^{(p)})^T\right),$$

где $\bar{z}_{2/1}^{(p)}=\tilde{R}_{z21}\tilde{R}_{z11}^{-1}z_1^{(p)}$ – выборочная оптимальная оценка $z_2^{(p)}$ по наблюдениям $z_1^{(p)}$. Поскольку

$$\sum_{p=1}^{P}z_2{}^{(p)}\bar{z}_{2/1}{}^{(p),T}=P\tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12},\ \sum_{p=1}^{P}\bar{z}_{2/1}{}^{(p)}\bar{z}_{2/1}{}^{(p),T}=P\tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}$$

$$\sum_{p=1}^{P}z_2{}^{(p)}\tilde{z}_2{}^{(p),T}=P\tilde{R}_{z21}W^{(1),T}W^{(2),T},\ \sum_{p=1}^{P}\bar{z}_{2/1}{}^{(p)}\tilde{z}_2{}^{(p),T}=P\tilde{R}_{z21}W^{(1),T}W^{(2),T},$$

то два последних слагаемых в выражении для E равны нулю. Это означает, что ошибка, минимизируемая относительно коэффициентов матриц $W^{(1)},W^{(2)}$, состоит из двух слагаемых, первое из которых не зависит от элементов $W^{(1)},W^{(2)}$, а второе – определяет среднюю квадратичную ошибку восстановления оптимальной линейной оценки при обучении НС рассматриваемой архитектуры.

$$E=\frac{1}{2}\left(\sum_{p=1}^{P}(z_2{}^{(p)}-\bar{z}_{2/1}{}^{(p)})^T(z_2{}^{(p)}-\bar{z}_{2/1}{}^{(p)})\right)+\frac{1}{2}\left(\sum_{p=1}^{P}(z_2{}^{(p)}-\tilde{z}_2{}^{(p)})^T(z_2{}^{(p)}-\tilde{z}_2{}^{(p)})\right).$$

Далее для наглядности рассмотрим случай, когда в скрытом слое нейронной сети один нейрон $M=1$, тогда L, как отношение квадратичных форм, является скалярной величиной. Обозначим $L=\lambda$. Отсюда следует, что вектор $W^{(1),T}$ удовлетворяет следующему соотношению:

$$\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}W^{(1),T}=\lambda W^{(1),T}. \qquad (2.10)$$

Это означает, что $W^{(1),T}$ является одним из собственных векторов $\{\varphi_i, i=\overline{1,N_1}\}$ матрицы $R_{z11}^{-1}R_{z12}R_{z21}$.

Рассмотренные выше условия минимизации среднеквадратичной ошибки являются необходимыми, но не достаточными условиями. Покажем, что как достаточное условие данный собственный вектор соответствует максимальному собственному числу полученного характеристического уравнения. Для этого перепишем (2.10) следующим образом:

$$\widetilde{R}_{z11}^{-1/2}\widetilde{R}_{z12}\widetilde{R}_{z21}\widetilde{R}_{z11}^{-1/2}\widetilde{R}_{z11}^{1/2}W^{(1),T} = \lambda\widetilde{R}_{z11}^{1/2}W^{(1),T}.$$

Обозначив вектор $x = R_{z11}^{1/2}W^{(1),T}$, получим новое характеристическое уравнение

$$\widetilde{R}_{z11}^{-1/2}\widetilde{R}_{z12}\widetilde{R}_{z21}\widetilde{R}_{z11}^{-1/2}x = \lambda x. \qquad (2.11)$$

Исходная матрица характеристического уравнения (2.10) $A = \widetilde{R}_{z11}^{-1}\widetilde{R}_{z12}\widetilde{R}_{z21}$ и матрица характеристического уравнения (2.11) $B = \widetilde{R}_{z11}^{-1/2}\widetilde{R}_{z12}\widetilde{R}_{z21}\widetilde{R}_{z11}^{-1/2}$ являются подобными и имеют одинаковый спектр собственных чисел. Если собственные значения для (2.10) и (2.11) совпадают, то собственные векторы связаны соотношением $x_i = R_{z11}^{1/2}\varphi_i$, $i = \overline{1, N_1}$.

Рассмотрим теперь матрицу $C = \widetilde{R}_{z21}\widetilde{R}_{z11}^{-1}\widetilde{R}_{z12}$, которая является матрицей ковариации оценки $\bar{z}_{2/1} = \widetilde{R}_{z21}\widetilde{R}_{z11}^{-1}z_1$ и соответствующее характеристическое уравнение для этой матрицы

$$\widetilde{R}_{z21}\widetilde{R}_{z11}^{-1}\widetilde{R}_{z12}y = \lambda y. \qquad (2.12)$$

Матрица $B = TT^T$ и матрица $C = T^TT$, где $T = \widetilde{R}_{z21}\widetilde{R}_{z11}^{-1/2}$, связанные подобным образом, имеют одинаковые ненулевые собственные числа [40]. Поэтому решения для (2.10), (2.11), (2.12) имеют одинаковый спектр собственных чисел.

Пусть вектор $W^{(1),T}$ является собственным вектором матрицы A, т.е. $W^{(1),T} = \varphi_i$, тогда $W^{(2)} = y_i$, выражаемый в соответствии с (2.9), есть собственный вектор матрицы C. Для доказательства подставим в (2.12) выражение (2.7) для $W^{(2)}$. Тогда с учетом (2.10) и (2.11) получим

$$\widetilde{R}_{z21}\widetilde{R}_{z11}^{-1}\widetilde{R}_{z12}W^{(2)} = \widetilde{R}_{z21}\widetilde{R}_{z11}^{-1}\widetilde{R}_{z12}\widetilde{R}_{z21}W^{(1),T}\left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1} =$$

$$= \lambda_i\widetilde{R}_{z21}W^{(1),T}\left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1} = \lambda_i W^{(2)} = \lambda_i y_i.$$

Напомним, что при минимизации ошибки E в ходе обучения нейронной сети осуществляется минимизация ошибки аппроксимации вектора выборочной оценки $\bar{z}_{2/1}^{(p)}$. Применение в нейронной сети весовой матрицы $W^{(2)} = y_i$ формирует выходной вектор $\tilde{z}_2$, пропорциональный данному собственному вектору матрицы C. Исходя из известных свойств разложения Карунена-Лоева [32, 38], указанное преобразование дает наименьшую остаточную дисперсию ошибки в случае, когда данный собственный вектор соответствует максимальному собственному значению $\lambda_{\max} = \lambda_1$. Это и является достаточным условием минимума E.

По аналогии подобное доказательство можно провести и для любого количества нейронов в скрытом слое. Без потери общности рассмотрим нейронную сеть с $M = 2$ нейронами в скрытом слое. Матрица $W^{(1)}$ теперь состоит из двух строк, а матрица $W^{(2)}$ из двух столбцов

$$W^{(1)} = \begin{pmatrix} W_1^{(1)} \\ W_2^{(1)} \end{pmatrix}, \quad W^{(2)} = \begin{pmatrix} W_1^{(2)} & W_2^{(2)} \end{pmatrix}.$$

При этом матрица $W^{(1)}$ удовлетворяет следующему соотношению:

$$\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}W^{(1),T} = W^{(1),T}L, \; L = \left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1}W^{(1)}R_{z12}R_{z21}W^{(1),T}, \qquad (2.13)$$

где L – несимметричная, в общем случае, матрица размера $M \times M$, являющаяся произведением двух симметричных матриц: положительно определенной матрицы $D = \left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1}$ и, как минимум, неотрицательно определенной матрицы $G = W^{(1)}R_{z12}R_{z21}W^{(1),T}$. Матрица L имеет вещественные собственные числа [9]. Выполним для неё диагонализирующее преобразование, задаваемое невырожденной матрицей U, состоящей из правых собственных векторов L [40], и перепишем (2.13) следующим образом:

$$\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}W^{(1),T}U = W^{(1),T}UU^{-1}LU = W^{(1),T}U\begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix}.$$

Последняя запись эквивалентна следующей системе уравнений:

$$\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}V = V\begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix}, \quad V = W^{(1),T}U = \begin{pmatrix} V_1 & V_2 \end{pmatrix}, \tag{2.14}$$

$$V_1 = W_1^{(1),T}u_{11} + W_2^{(1),T}u_{21},\ V_2 = W_1^{(1),T}u_{12} + W_2^{(1),T}u_{22},$$

$$\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}V_1 = v_1V_1,\ \tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}V_2 = v_2V_2,$$

где v_1, v_2, u_{11}, u_{21}, u_{12}, u_{22} – элементы матриц $U^{-1}LU$ и U соответственно. Векторы-столбцы V_1, V_2 являются решениями характеристического уравнения и, соответственно, собственными векторами $\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}$. Отсюда следует, что столбцы матрицы $W^{(1),T}$ являются линейной комбинацией двух собственных векторов матрицы $\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}$

$$W_1^{(1),T} = \alpha_{11}\varphi_1 + \alpha_{12}\varphi_2,\ W_2^{(1),T} = \alpha_{21}\varphi_1 + \alpha_{22}\varphi_2\ ,$$

где

$$\alpha_{11} = \frac{-u_{22}}{u_{12}u_{21} - u_{11}u_{22}}, \alpha_{12} = \frac{u_{21}}{u_{12}u_{21} - u_{11}u_{22}}, \alpha_{21} = \frac{u_{12}}{u_{12}u_{21} - u_{11}u_{22}}, \alpha_{22} = \frac{-u_{11}}{u_{12}u_{21} - u_{11}u_{22}},$$

а матрицу весов нейронной сети можно представить в виде

$$W^{(1),T} = \begin{pmatrix} \varphi_1 & \varphi_2 \end{pmatrix}U,\ W^{(1)} = U^T\begin{pmatrix} \varphi_1^T \\ \varphi_2^T \end{pmatrix}.$$

Это означает, что $W^{(1),T}$ является линейной комбинацией двух собственных векторов $\{\varphi_i, i = \overline{1, N_1}\}$ матрицы $\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}$.

Опять, рассмотрим матрицу $C = \tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}$, которая является выборочной матрицей ковариации оценки $\bar{z}_{2/1} = \tilde{R}_{z21}\tilde{R}_{z11}^{-1}z_1$ и соответствующее ей характеристическое уравнение

$$\tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}y = \lambda y. \tag{2.15}$$

Как и в случае наличия одного нейрона в скрытом слое матрица $B = \tilde{R}_{z11}^{-1/2}\tilde{R}_{z12}\tilde{R}_{z21}\tilde{R}_{z11}^{-1/2} = TT^T$ и матрица $C = T^TT$, где $T = \tilde{R}_{z21}\tilde{R}_{z11}^{-1/2}$ имеют одинаковые ненулевые собственные числа. Поэтому решения для (2.14), (2.15) имеют одинаковый спектр собственных чисел.

Покажем, что векторы φ_1, φ_2 соответствуют двум максимальным собственным значениям матрицы $\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}$. Для доказательства подставим в (2.15) выражение (2.9) для матрицы $W^{(2)}$. Тогда получим

$$\tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}W^{(2)} = \tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}W^{(1),T}\left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1} =$$

$$= \tilde{R}_{z21}W^{(1),T}\left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1}W^{(1)}R_{z12}R_{z21}W^{(1),T}\left(W^{(1)}R_{z11}W^{(1),T}\right)^{-1} = W^{(2)}L^T .$$

Выполним для матрицы L^T диагонализирующее преобразование, задаваемое невырожденной матрицей Ω, состоящей из левых собственных векторов L [40]

$$\tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}W^{(2)}\Omega = W^{(2)}\Omega\Omega^{-1}L^T\Omega = W^{(2)}\Omega\begin{pmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{pmatrix},$$

Если повторить анализ, проведенный ранее для уравнения (2.14), можно также показать, что каждый столбец $W^{(2)}$ является линейной комбинацией двух собственных векторов матрицы $C = \tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}$. Соответственно, применение весовой матрицы $W^{(2)}$ формирует вектор $\tilde{z}_2$ на выходе нейронной сети в виде разложения по этим двум собственным векторам матрицы C. При обеспечении минимальной средней квадратичной ошибки в процессе обучения нейронной сети эти векторы обязательно соответствуют максимальным собственным значениям C. Поэтому и векторы φ_1, φ_2 соответствуют двум максимальным собственным числам матрицы $\tilde{R}_{z11}^{-1}\tilde{R}_{z12}\tilde{R}_{z21}$.

Необходимо отметить, что существует бесконечное множество решений, доставляющих минимум используемому целевому функционалу. Действительно, пусть матрицы весовых коэффициентов $W^{(1)}$, $W^{(2)}$ определяют точку минимума E. Пусть T произвольная невырожденная матрица размера $M \times M$. Тогда матрицы весовых коэффициентов $W_*^{(1)} = TW^{(1)}$ и $W_*^{(2)} = W^{(2)}T^{-1}$ выполняют такое же преобразование данных в двухслойной нейронной сети рассматриваемой архитектуры и также являются одним из возможных решений задачи на нахождение экстремума.

Полученный результат может интерпретироваться в форме следующей теоремы.

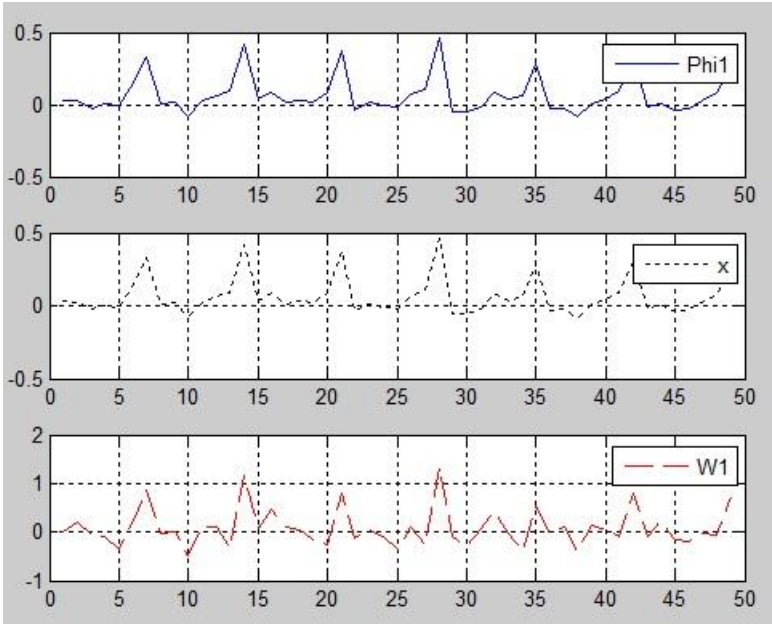
Теорема. В результате обучения с минимизацией средней квадратичной ошибки двухслойной линейной нейронной сети с архитектурой, представленной на рисунке 2.2б и имеющей $M < N_1, M < N_2$ нейронов в скрытом слое по совокупности реализаций векторов z_1, z_2 объемом $P > N_1, P > N_2$ на выходе получается приближенное представление оценки $\bar{z}_{2/1} = \tilde{R}_{z21}\tilde{R}_{z11}^{-1}z_1$ вектора z_2 относительно наблюдения z_1 в виде разложения по первым M собственным векторам выборочной матрицы ковариации $\tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}$ с минимальной дисперсией остаточной ошибки.

Следствие. Пусть векторы $z_1 = z_2 = z$ тождественны и имеют одинаковую размерность N, а при обучении сети на входе и на выходе используются одна и та же совокупность реализаций $Z^P = \{z^{(1)},\ldots,z^{(P)}\}$.

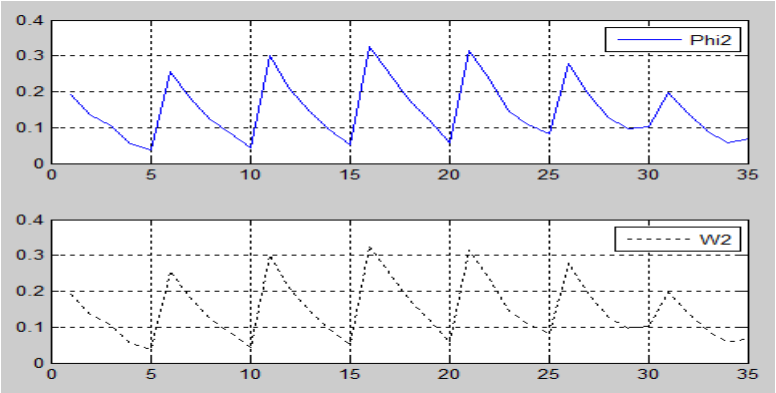
Тогда линейная автоассоциативная двухслойная нейронная сеть прямого распространения с архитектурой, представленной на рисунке 2.2б и имеющей $M < N$ нейронов в скрытом слое, после обучения по выборке объемом $P > N$ реализаций случайного вектора формирует выходные реакции, эквивалентные разложению по первым M функциям базиса Карунена-Лоева выборочной матрицы ковариации $\tilde{R}_{z11}$ случайного вектора z. При этом для весовых коэффициентов обученной нейронной сети имеет место соотношение $W_1W_2 = I$, где I – единичная матрица.

В данном случае нейронная сеть выполняет сжимающее преобразование данных, при котором процедура обучения сети организуется так, чтобы при поступлении на вход сети обучающей последовательности данных (реализации случайного вектора) целевые значения выхода сети имели такие же значения, что и значения данных на входе. После этого при подаче любого входного воздействия на выходе нейронов промежуточного слоя реализуется сжатие информации и формируются отклики, близкие главным компонентам входных данных.

Необходимо отметить, что данный результат, ранее полученный в работе В.Г. Попова, А.А. Сироты [38], является следствием из выше доказанной автором диссертации теоремы, которая носит существенно более общий характер. Важно также отметить, что в известной литературе, например в [47], данные свойства автоассоциативной нейронной сети с рассматриваемой архитектурой обсуждаются и экспериментально устанавливаются, однако более или менее строгие доказательства авторами не приводятся.



а)



б)

Рисунок 2.3 - Базисные векторы разложения Карунена-Лоева (а) и матрицы весов первого и второго слоя нейронной сети с заданной архитектурой (б)

Для наглядной иллюстрации доказанной теоремы на рисунке 2.3 приведены примеры моделирования процесса обучения нейронной сети с архитектурой, представленной на рисунке 2.2б, для одного нейрона в скрытом слое по набору векторов $Z_1^P = \{z_1^{(1)},...,z_1^{(P)}\}$ и $Z_2^P = \{z_2^{(P)},...,z_2^{(P)}\}$, полученных путем генерации и развертки фрагментов случайных полей при $P = 1000$.

На рисунке 2.3а и б представлены результаты экспериментов, полученные для фрагментов случайных полей z_1 размером 7x7 элементов и z_2 размером 7x5. На графиках используются следующие обозначения: Phi1 – теоретически рассчитанный собственный вектор матрицы $B = \tilde{R}_{z11}^{-1/2}\tilde{R}_{z12}\tilde{R}_{z21}\tilde{R}_{z11}^{-1/2}$; x – вектор $x = R_{z11}^{1/2}W^{(1),T}$; W1 – матрица весов первого слоя (при развертке в вектор) НС; Phi2 – теоретически рассчитанный собственный вектор матрицы $C = \tilde{R}_{z21}\tilde{R}_{z11}^{-1}\tilde{R}_{z12}$; W2 – матрица весовых коэффициентов выходного слоя НС (при развертке в вектор).

Таким образом, получены обоснования сходимости весовых коэффициентов двухслойной линейной нейронной сети с сокращенным числом нейронов в скрытом слое к компонентам собственных векторов, получаемых при решении обобщенной задачи на собственные значения. При этом на выходе сети формируется приближение оптимальной выборочной линейной оценки в виде разложения по первым собственным векторам матрицы ковариации оценки, число которых равно числу нейронов в скрытом слое.

В целом приведенные результаты показывают, что при настройке коэффициентов простейших линейных нейронных сетей в условиях прямого и косвенного обучения по представительным выборкам наблюдаемых данных имеется возможность сформировать оценки векторов наблюдаемых и ненаблюдаемых параметров, эквивалентные по своим свойствам статистически оптимальным.

2.4. Базовые алгоритмы создания цифровых водяных знаков при использовании в качестве контейнеров статистических эквивалентов сигналов и изображений

Определим в общем виде модель стеганографического скрытия информации. Пусть Z, D, K есть множества возможных контейнеров, скрываемых сообщений и ключей, тогда процедура встраивания и извлечения сообщения может быть представлена в виде отображений вида

$$F_{*}: Z \times D \times K \to Z,\ \bar{z} = F_{*}(z, d, k),\ z \in Z,\ \bar{z} \in Z, d \in D,\ k \in K,\ \|z - \bar{z}\| \to \min,$$

$$F_{**}: Z \times K \to D,\ \tilde{d} = F_{**}(\bar{z}, k),\ \tilde{d} \in D,\ \|d - \tilde{d}\| \to \min,$$

где $z, \bar{z}$ – исходный и заполненный контейнер; $d, \tilde{d}$ – исходное и восстановленное сообщение. Далее в качестве контейнера без ограничения общности будем рассматривать реализацию случайного поля, заданного на двумерной дискретной сетке (как аналога сигнала или изображения), а в качестве сообщения или ЦВЗ – двоичную последовательность $d^{(p)}$, $p = \overline{1, P}$, где $d^{(p)} \in \{-1, +1\}$ – скалярная величина, несущая один бит информации. При этом свойства контейнера должны модифицироваться так, чтобы внесенное при встраивании стегосообщения (ЦВЗ) изменение контейнера практически невозможно было бы выявить при визуальном и статистическом контроле, а само встраиваемое ЦВЗ должно быть максимально устойчиво к различного рода искажениям.

Для решения задачи ССИ целесообразно использовать нейронные сети различных типов и архитектур, которые обеспечивают принципиальные возможности воспроизведения функциональных моделей преобразования данных в соответствии с приведенными выше общими соотношениями.

Применение нейросетевых функциональных моделей и алгоритмов создания ЦВЗ, по нашему мнению, связано с возможностью построения универсальных сжимающих отображений данных на основе линейных НС прямого распространения. При этом в соответствии с теоремой, сформулированной в п. 2.1.2, могут быть рассмотрены два типа преобразований: преобразования, реализующие в терминологии [35] гетероассоциативное сжимающее

отображение, и преобразования, реализующие (как частный и более простой случай) автоассоциативное сжимающее отображение).

2.4.1. Нейросетевые функциональные модели преобразования данных и базовый алгоритм создания ЦВЗ при реализации отображений автоассоциативного типа

Постановка задачи ССИ с использованием аппарата искусственных нейронных сетей в данном случае может быть сформулирована следующим образом. Требуется с использованием функциональных возможностей нейронных сетей для любого фрагмента контейнера, представленного в виде вектора $z \in R^n$, и вектора встраиваемых данных $d \in R^1$ построить отображения

$$\tilde{z} = F_c(z),\ \bar{z} = F_m(\tilde{z}, d),\ \|\bar{z} - z\| \to \min,\ \tilde{d} = F_d(\bar{z}),\ \tilde{d} \in D,\ \left\|d - \tilde{d}\right\| \to \min,$$

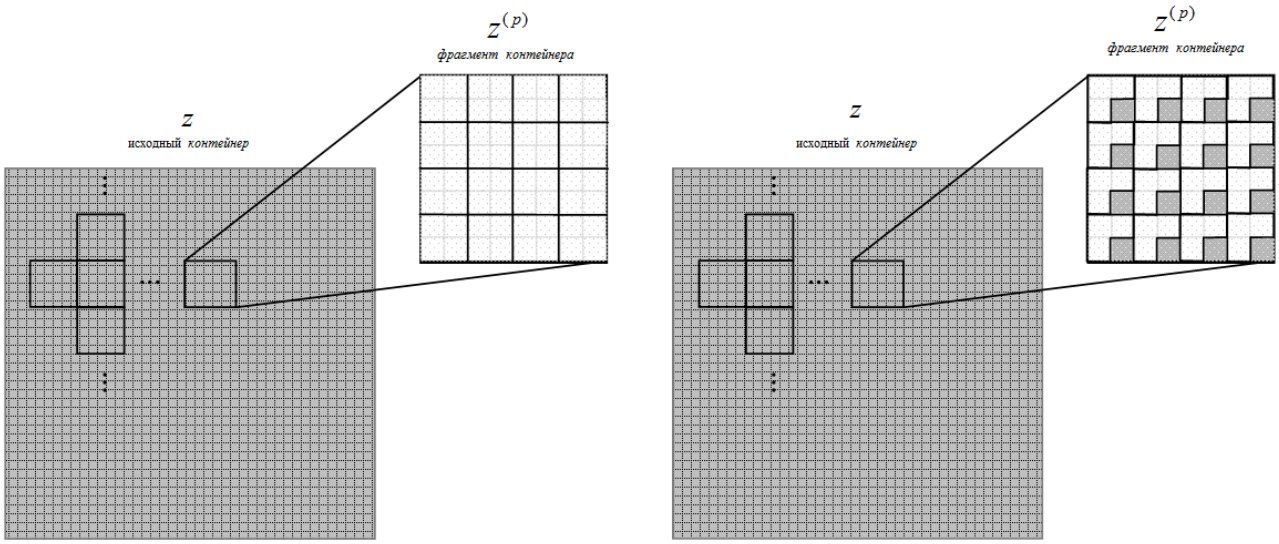
где первый оператор F_c реализует сжимающее преобразование входных данных, обеспечивающее подготовку вектора-контейнера к встраиванию данных, оператор F_m реализует собственно встраивание ЦВЗ, а оператор F_d – выполняет декодирование (восстановление) ранее скрытого сообщения на основе решения задачи классификации предъявляемого фрагмента контейнера для двух гипотез $H_1 : d^{(p)} = -1, H_2 : d^{(p)} = +1$.

Рассмотрим реализацию описанной модели встраивания данных для работы с данными вещественного формата. Для обоснования базового алгоритма создания ЦВЗ будем изначально использовать стохастическое представление данных. В качестве ЦВЗ без ограничения общности будем рассматривать двоичную последовательность $d^{(p)}$, $p = \overline{1,P}$, при этом $d^{(p)} \in \{-1,+1\}, p = \overline{1,P}$ – скалярная величина, которая несет в себе один бит информации в пределах встраиваемого сообщения.

Пусть $z = (z_1, \ldots z_n)^T$ случайный вектор, представляющий фрагмент файла контейнера, совокупность реализаций которого $z^{(p)}$, $p = \overline{1,P}$ используется для

встраивания элементов последовательности ЦВЗ, и, соответственно, в качестве обучающей выборки – для построения нейронной сети, обеспечивающей встраивание. Пусть для определенности математическое ожидание и матрица ковариации вектора z равны: $M[z]=0$, $M[zz^T]=R_z$.

Следует отметить, что при реализации предлагаемого подхода область определения (форма фрагмента контейнера) при выполнении сжимающих отображений может быть произвольной конфигурации. На рисунке 2.4 представлены различные примеры отображения конфигурации фрагментов, подвергаемых модификации в исходном контейнере: на рисунке 2.4а – это область квадратной формы, а на рисунке 2.4б – это решетка внутри области квадратной формы (белым цветом выделены пиксели исходного контейнера, в которые происходит встраивание последовательности ЦВЗ).



а) б)

Рисунок 2.4 - Примеры отображения конфигурации области модификации исходного контейнера для автоассоциативных преобразований

На рисунке 2.5 схематично представлен процесс встраивания последовательности ЦВЗ в исходный файл-контейнер и последующее восстановление ЦВЗ.

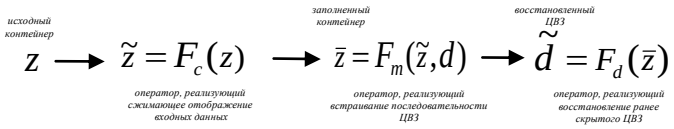
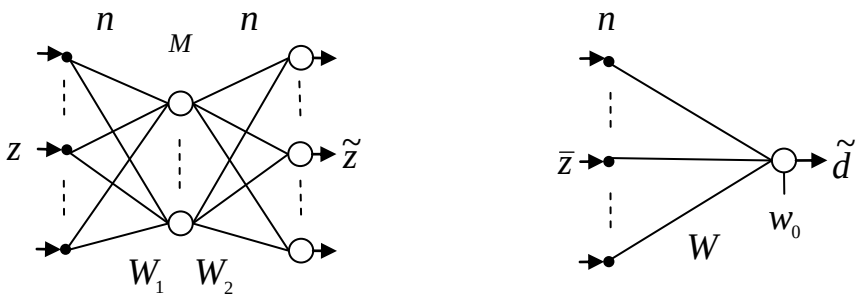


Рисунок 2.5 – Встраивание и восстановление последовательности ЦВЗ

Как следует из результатов п.2.1, структура автоассоциативной нейронной сети, реализующей оператор F_c, в общем случае может иметь вид, приведенный на рисунке 2.6а. При этом количество нейронов в скрытом слое нейронной сети равно $M \le n$. Матрицы весовых коэффициентов первого и второго слоя сети обозначены соответственно W_1, W_2. Для реализации алгоритма встраивания данных сначала проводится обучение сети, обеспечивающей первое (сжимающее отображение). Обучение проводится по совокупности реализаций входного и целевого вектора $\{z^{(p)}, y^{(p)} = z^{(p)}\}$, $p = \overline{1,P}$ так, чтобы минимизировать величину средней квадратичной ошибки представления входного вектора на выходе сети:

$$E = \frac{1}{2}\sum_{p=1}^{P}(y^{(p)} - W_2 W_1 z^{(p)})^T (y^{(p)} - W_2 W_1 z^{(p)}) \to \min .$$



а) б)

Рисунок 2.6 - Нейронная сеть, реализующая сжимающее преобразование для последующего встраивания (а); нейронная сеть, реализующая восстановление скрытой информации (б)

В результате обучения нейронной сети, представленной на рисунке 2.6а, получается индивидуальное для данного набора данных и конкретной конфигурации фрагментов сжимающее отображение с весьма незначительными

потерями, при котором реально получаемый на выходе сети вектор $\tilde{z} = (\tilde{z}_1, \ldots \tilde{z}_n)^T$ может быть представлен в виде разложения по первым M собственным векторам $\varphi_i, i = \overline{1, M}$ выборочной матрицы ковариации входного вектора

$$\hat{R}_z^{(P)} = \frac{1}{P-1} \sum_{p=1}^{P} z^{(p)} z^{(p),T},$$

которые являются одновременно функциями разложения Карунена-Лоэва [32, 38]. Тогда для любой реализации входного вектора $z^{(p)}$ получаемый на выходе вектор $\tilde{z}^{(p)}$ можно представить в виде

$$\tilde{z}^{(p)} = \sum_{i=1}^{M} \alpha_i^{(p)} \varphi_i, \quad z^{(p)} = \sum_{i=1}^{n} \alpha_i^{(p)} \varphi_i,$$

где $\alpha_i^{(p)}$, $i = \overline{1, M}$ – коэффициенты разложения по первым $M \le n$ собственным векторам $\varphi_i, i = \overline{1, M}$ матрицы $\hat{R}_z^{(P)}$. С целью минимизации ошибки искажения контейнера при выполнении подобного разложения будем использовать сеть, для которой $M = n - 1$, тогда $\tilde{z}^{(p)}$ будет отличаться от $z^{(p)}$ только «высокочастотной» составляющей с малой амплитудой и дисперсией соответствующей минимальному собственному числу выборочной матрицы ковариации.

После того, как сеть, реализующая оператор F_c, обучена, выполняется оператор F_m, реализующий окончательное встраивание ЦВЗ в файл контейнер. При этом вычисляется нормированный вектор

$$\varphi_n = r_{\min} \Big/ \sqrt{\left(r_{\min}{}^T r_{\min}\right)}, \; r_{\min} = \frac{1}{P} \sum_{p=1}^{P} (z^{(p)} - \tilde{z}^{(p)})$$

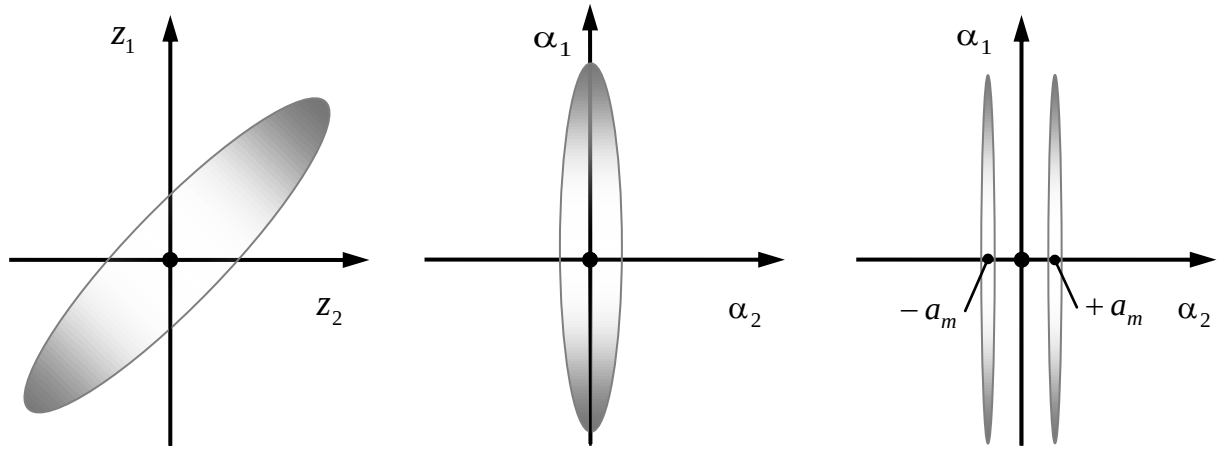
и каждый фрагмент файла-контейнера, описываемый теперь (после сжатия) вектором $\tilde{z}^{(p)}$, $p = \overline{1, P}$, модифицируется последовательностью ЦВЗ на основе следующего соотношения:

$$\bar{z}^{(p)} = \tilde{z}^{(p)} + a_m d^{(p)} \varphi_n, \; \tilde{z}^{(p)} = W_2 W_1 z^{(p)}, \; p = \overline{1, P}, \qquad (2.16)$$

где a_m – амплитуда вносимого искажения.

Соотношение (2.16), описывает результирующие преобразование, выполняемое при заполнении исходных фрагментов файла-контейнера, в

результате чего образуется последовательность $\bar{z}^{(p)}$, $p=\overline{1,P}$. На рисунке 2.7 дана наглядная иллюстрация описываемого информационного процесса на примере преобразования данных, представляемых в виде гауссовского коррелированного вектора. На рисунке 2.7а представлен исходная область локализация значений входного коррелированного случайного гауссовского вектора, которая имеет вид эллипсоида (гиперэллипсоида в многомерном пространстве). После выполнения декоррелирующего преобразования Карунена-Лоэва эллипсоид принимает вид, показанный на рисунке 2.7б. Рисунок 2.7в отражает вносимую модификацию входного вектора, когда вместо случайной величины $\alpha_n = \alpha_2$, распределенной по гауссовскому закону, используются два случайных точечных значения $a_2 = \pm a_m$, что соответствует использованию данных, имеющих две проекции гиперэллипсоида на ось a_1.



а) б) в)

Рисунок 2.7 - Область локализации значений исходного случайного вектора (а), после выполнения декоррелирующего преобразования (б), после встраивания ЦВЗ (в)

Для восстановления, скрытых таким образом данных и, соответственно, для реализации оператора F_d также может быть использована нейронная сеть,

реализующая необходимое функциональное преобразование данных заполненного контейнера. Для эффективного восстановления скрытых данных необходимо решить задачу классификации наблюдаемого вектора, в качестве которого может выступать вектор модифицированной выходной части фрагмента контейнера $\bar{z}$ по его принадлежности к одному из классов H_1, H_2, характеризующихся различными математическими ожиданиями $H_1 : m_+ = a_m \varphi_n, H_2 : m_- = -a_m \varphi_n$ и общей матрицей ковариации R_η. В рассматриваемом случае решающее правило получается путем обучения простейшей однослойной линейной нейронной сети, структура которой приведена на рисунке 2.6б. На рисунке 2.6б используются следующие обозначения: W – матрица весовых коэффициентов сети, w_0 – вес постоянного смещения (вес поляризации). Такая сеть обучается для решения задачи классификации входного для нее вектора с целью выделения значения ранее скрытого в нем элемента последовательности d. Волнистая линия в обозначениях элементов восстанавливаемой последовательности $\tilde{d}^{(p)}$, $p = \overline{1,P}$ означает наличие возможных ошибок при проведении классификации. При этом доказано, что для гауссовских случайных векторов в результате обучения однослойной линейной НС по совокупности $\{\bar{z}^{(p)}, d^{(p)}, p = \overline{1,P}\}$ при $P \to \infty$ формируется преобразование входных данных в выходные, реализующее структуру оптимального решающего правила.

Таким образом, для эффективного восстановления скрытых данных необходимо решить задачу классификации наблюдаемого вектора $\bar{z}$ по его принадлежности к одному из классов H_1 и H_2, характеризующихся различными математическими ожиданиями (см. также рисунок 2.7а). В [13, 43] показано, что в рассматриваемом случае указанное решающее правило получается путем обучения простейшей однослойной линейной нейронной сети, структура которой приведена на рисунке 2.6б. При этом сформулировано следующее утверждение: для гауссовских случайных векторов в результате обучения однослойной линейной НС (рисунок 2.6б) по совокупности $\{\bar{z}^{(p)}, d^{(p)}, p = \overline{1,P}\}$ при $P \to \infty$

формируется преобразование входных данных в выходные, реализующее структуру оптимального решающего правила.

При статистическом моделировании для проведения экспериментального анализа в качестве контейнера рассматривались реализации однородных гауссовских случайнных полей (ГСП) с корреляционными функциями вида

$$R(x,x',y,y') = \sigma^2 \exp(-\alpha r), \quad r = |x-x'| + |y-y'|,$$

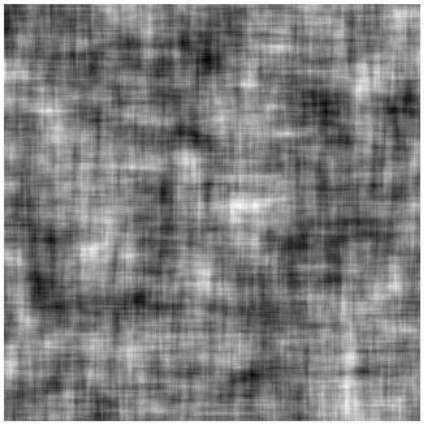
$$R(x,x',y,y') = \sigma^2 \exp(-\alpha r), \quad r = (x-x')^2 + (y-y')^2,$$

$$R(x,x',y,y') = \sigma^2 \exp(-\alpha r), \quad r = \sqrt{(x-x')^2 + (y-y')^2}.$$

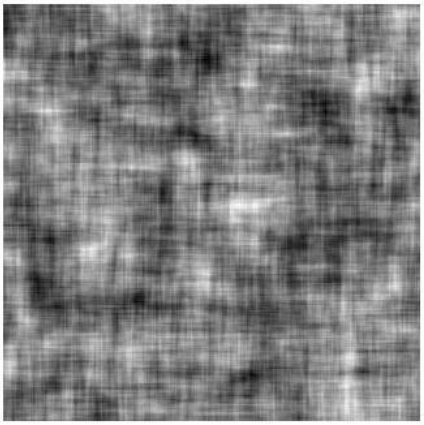
Фрагменты (кадры) случайного поля используются в качестве реализаций вектора z. При дискретном представлении с заданными интервалами дискретизации $\Delta x, \Delta y$ координат для описания такого случайного поля используется матрица ковариации $R = \|r_{ij}\|$, $r_{ij} = R(i\Delta x, j\Delta y)$. При моделировании задавались параметр корреляции $\alpha = 0.1$, дисперсия поля $\sigma^2 = 1$, и корреляционные функции вида:

$$r_{ij} = \sigma^2 \exp\left(-\alpha\left(|i| + |j|\right)\right), \; r_{ij} = \sigma^2 \exp\left(-\alpha\left(i^2 + j^2\right)\right), \; r_{ij} = \sigma^2 \exp\left(-\alpha\sqrt{\left(i^2 + j^2\right)}\right).$$

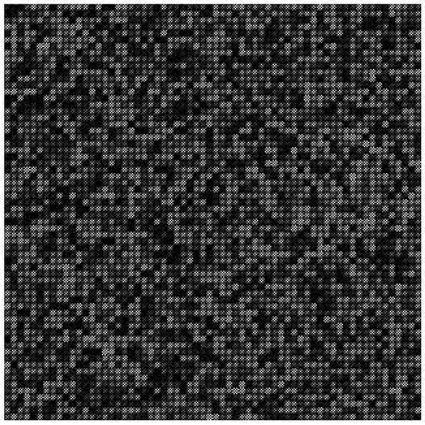
На рисунке 2.8а представлен пример реализации случайного поля с функцией корреляции первого типа, используемой в качестве контейнера.



а)



б)



в)

Рисунок 2.8 – Пример генерации реализации случайного поля на основе функции корреляции первого типа, исходный контейнер (а); заполненный контейнер (б); разница между исходным и заполненным контейнером (в)

На рисунке 2.8б представлен результирующий заполненный контейнер, а на рисунке 2.8в показана разность между исходным и заполненным контейнером с коэффициентом умножения 100 для обеспечения наглядности отображения.

На рисунке 2.9а представлены зависимости среднеквадратичной ошибки искажения контейнера от амплитуды a_m вносимого искажения. При моделировании задавались: размер фрагмента $n = 5 \times 5$, используемого для встраивания одного элемента последовательности ЦВЗ, объем обучающей выборки для формирования НС $P_1 = 4096$, объем тестирующей выборки $P_2 = 4096$, число эпох обучения НС $T = 5000$. Для восстановления ЦВЗ использовалась линейная нейронная сеть с архитектурой, представленной на рисунке 2.6б. На рисунке 2.9б приведены зависимости вероятности ошибки при восстановлении одного бита встраиваемой информации от амплитуды встраиваемой последовательности.

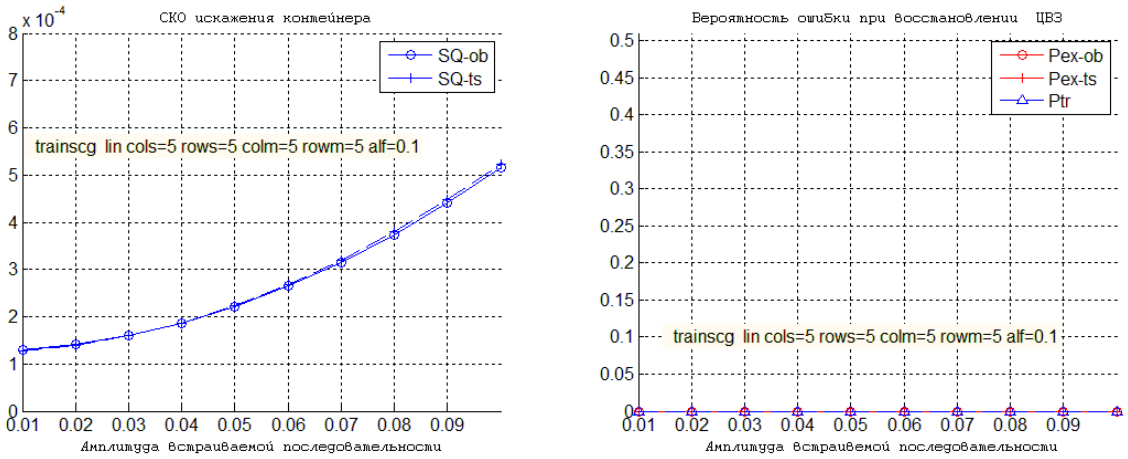
Здесь и далее по тексту диссертации на графиках используются следующие обозначения: SQ-ob – среднеквадратичная ошибка искажения контейнера при тестировании на данных, используемых для обучения нейронной сети; SQ-ts –

среднеквадратичная ошибка искажения контейнера при тестировании на данных, не участвовавших в обучении нейронной сети. Pex-ob – оценка вероятности ошибки при восстановлении ЦВЗ, при тестировании на данных, используемых для обучения нейронной сети; Pex-ts – оценка вероятности ошибки при восстановлении ЦВЗ при тестировании на данных, не участвовавших в обучении нейронной сети; Ptr – теоретически оцененная вероятность ошибки, которая в общем случае может быть рассчитана по формуле:

$$P_{tr} = 1 - \Phi(\alpha),\ \alpha = 0{,}5\sqrt{4\varphi_n^T R_\eta^{-1} \varphi_n}),\ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\ .$$

При использовании сжимающего преобразования автоассоциативного типа теоретическая ошибка в данном случае равна нулю, вследствие сингулярности матрицы ковариации (см.также наглядное отображение этой ситуации на рисунке 2.7), что и проявилось на графиках.

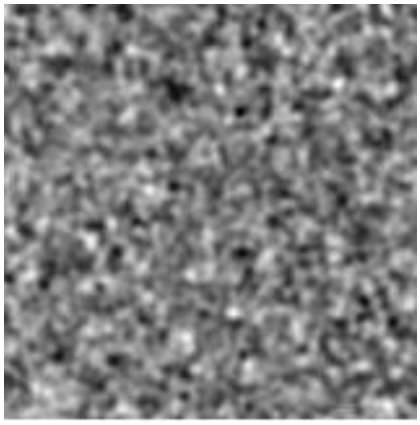
Кроме того, на графиках показаны условия и параметры, задаваемые при моделировании: trainscg - метод обучения сжимающей НС; lin - линейная НС, используемая для восстановления ЦВЗ; cols - размерность элементарного блока по оси OX; rows - размерность элементарного блока по оси OY; colm - размерность внутреннего блока по оси OX; rowm - размерность внутреннего блока по оси OY; alf – коэффициент, определяющий корреляцию поля.



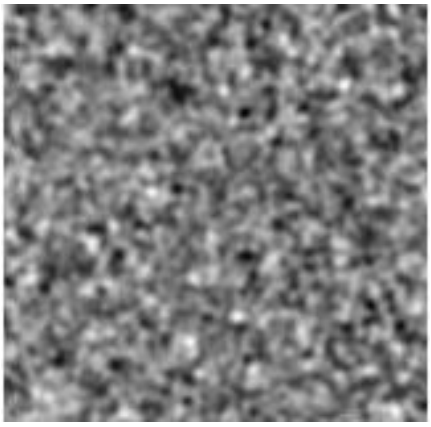
а) б)

Рисунок 2.9 - Зависимости среднеквадратичной ошибки искажения (а); вероятности ошибки восстановления ЦВЗ (б) от амплитуды встраиваемой последовательности ЦВЗ

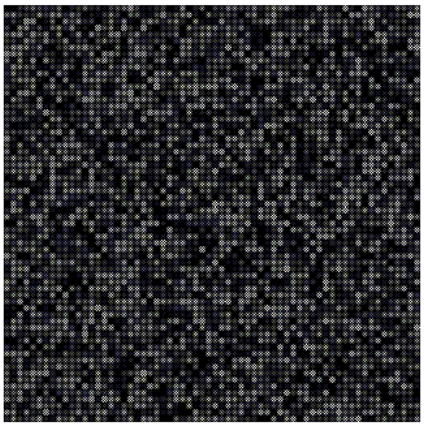
На рисунке 2.10 показаны аналогичные результаты при использовании в качестве контейнера реализации случайного поля с функцией корреляции второго типа, при этом параметр корреляции задается равным $\alpha = 0.3$, для реализации исходного фрагмента поля в виде реализации гауссовского вектора.



а)



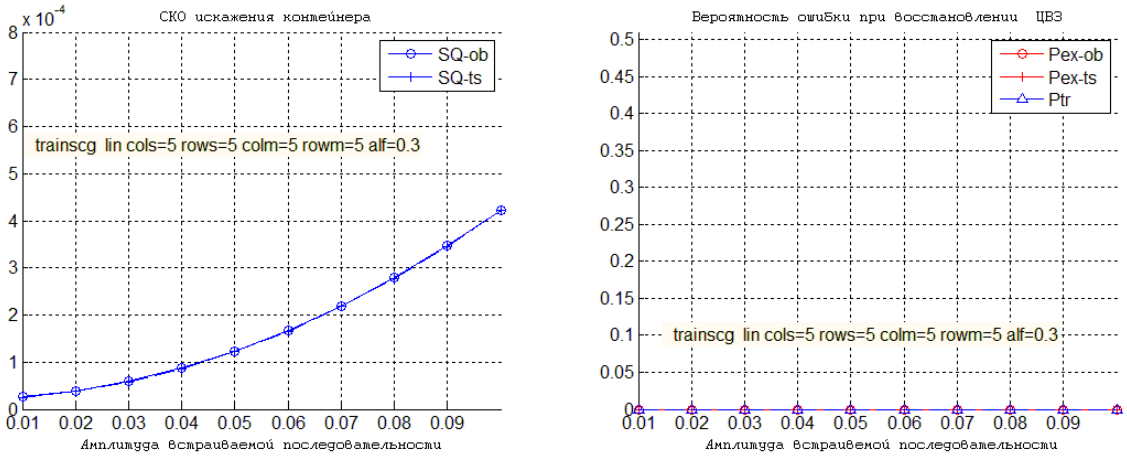
б)



в)

Рисунок 2.10 – Пример генерации реализации случайного поля на основе функции корреляции первого типа, исходный контейнер (а); заполненный контейнер (б); разница между исходным и заполненным контейнером (в)

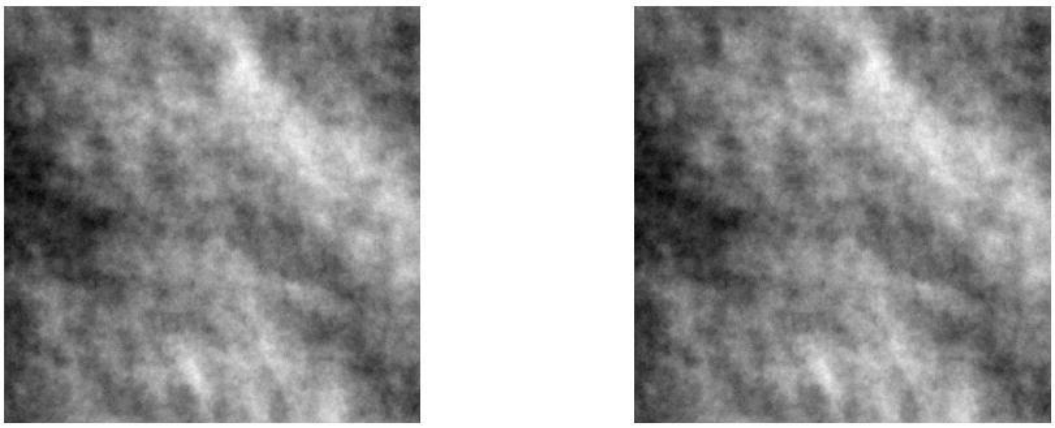
На рисунке 2.11 показаны зависимости среднеквадратичной ошибки искажения и вероятности ошибки восстановления ЦВЗ от амплитуды встраиваемой последовательности.



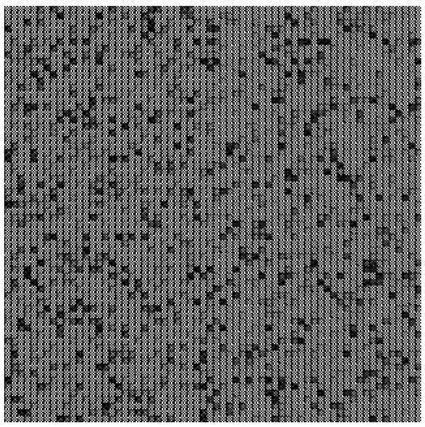
а) б)

Рисунок 2.11 - Зависимости среднеквадратичной ошибки искажения (а); вероятности ошибки восстановления ЦВЗ (б) от амплитуды встраиваемой последовательности ЦВЗ

На рисунке 2.12 показаны аналогичные результаты при использовании в качестве контейнера реализации случайного поля с функцией корреляции третьего типа, при этом параметр корреляции задается равным $\alpha = 0.01$, для реализации исходного фрагмента поля в виде реализации гауссовского вектора.



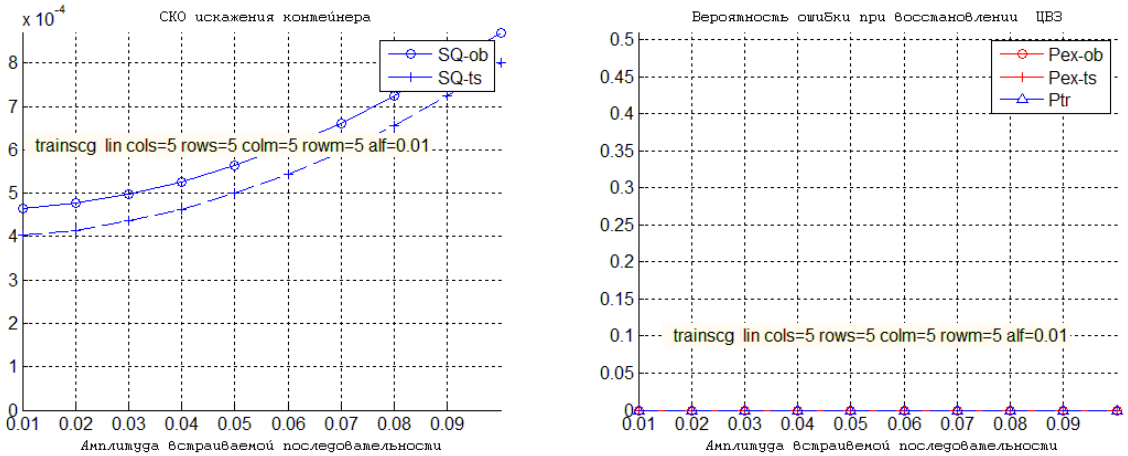
а) б)



в)

Рисунок 2.12 – Пример генерации реализации случайного поля на основе функции корреляции первого типа, исходный контейнер (а); заполненный контейнер (б); разница между исходным и заполненным контейнером (в)

На рисунке 2.13 показаны зависимости среднеквадратичной ошибки искажения и зависимость вероятности ошибки восстановления ЦВЗ.



а) б)

Рисунок 2.13 - Зависимости среднеквадратичной ошибки искажения (а); вероятности ошибки восстановления ЦВЗ (б) от амплитуды встраиваемой последовательности ЦВЗ

Необходимо отметить, что для всех вариантов формирования реализации гауссовского случайного поля с увеличением амплитуды встраиваемой

последовательности увеличиваются значения средней квадратичной ошибки искажения контейнера. В то же время, вероятность ошибки восстановления ЦВЗ при использовании автоассоциативного преобразования равна нулю, так как матрица ковариации R_η в этом случае является вырожденной.

2.4.2. Нейросетевые функциональные модели преобразования данных и базовый алгоритм создания ЦВЗ при реализации отображений гетероассоциативного типа

Рассмотрим вариант создания цифровых водяных знаков на основе сжимающих отображений гетероассоциативного типа [29]. Основное отличие от ранее описанного подхода заключается в процедуре реализации встраивания ЦВЗ. Здесь каждый фрагмент, используемый для встраивания элемента двоичной последовательности $\tilde{d}^{(p)}$ разбивается на две части: входную и выходную. Соответственно, на вход сети, выполняющей сжатие, подаются данные из одной части фрагмента контейнера; при этом на выходе формируется приближенная оценка данных другой части фрагмента, которые затем и подвергаются модификации. Это обеспечивает маскирование процесса встраивания и требует использования дополнительно обученной нейронной сети, определяющей корреляционные связи этих частей фрагмента для оценки (прогноза) значений данных выходной части фрагмента при известных значениях первой части.

Пусть $z=(z_1,\ldots z_n)^T$, $z\in R^n, n=N_1+N_2$ случайный вектор, представляющий некоторую область Ω изображения (фрагмент контейнера). Пусть для определенности математическое ожидание и матрица ковариации вектора z равны: $M[z]=0$, $M[zz^T]=R_z$.

Вектор z всегда может быть представлен как составной $z=(z_1^T,z_2^T)^T$, где $z_1\in R^{N_1}$ отображает некоторую подобласть фрагмента Ω_1, называемую входной частью, а $z_2\in R^{N_2}$ – подобласть фрагмента Ω_2, называемую выходной частью, $\Omega_1\cup\Omega_2=\Omega$. В качестве обучающей выборки для построения НС, реализующей сжимающее отображение, будем использовать данные $\{z_1^{(p)},z_2^{(p)}\}$, $p=\overline{1,P}$

входной и выходной частей совокупности фрагментов, рассматриваемые как реализации случайных векторов z_1, z_2.

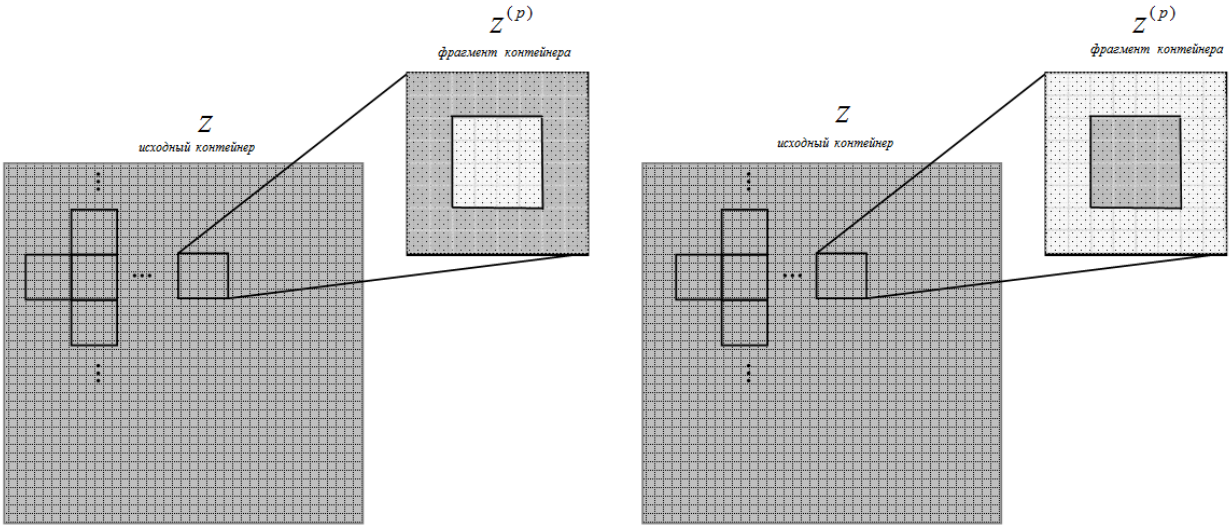
Необходимо отметить, что векторы z_1, z_2 связаны соотношением:

$$z_2 = \breve{z}_{2/1} + V = Hz_1 + V, \quad H = R_{z21}R_{z11}^{-1} \tag{2.17}$$

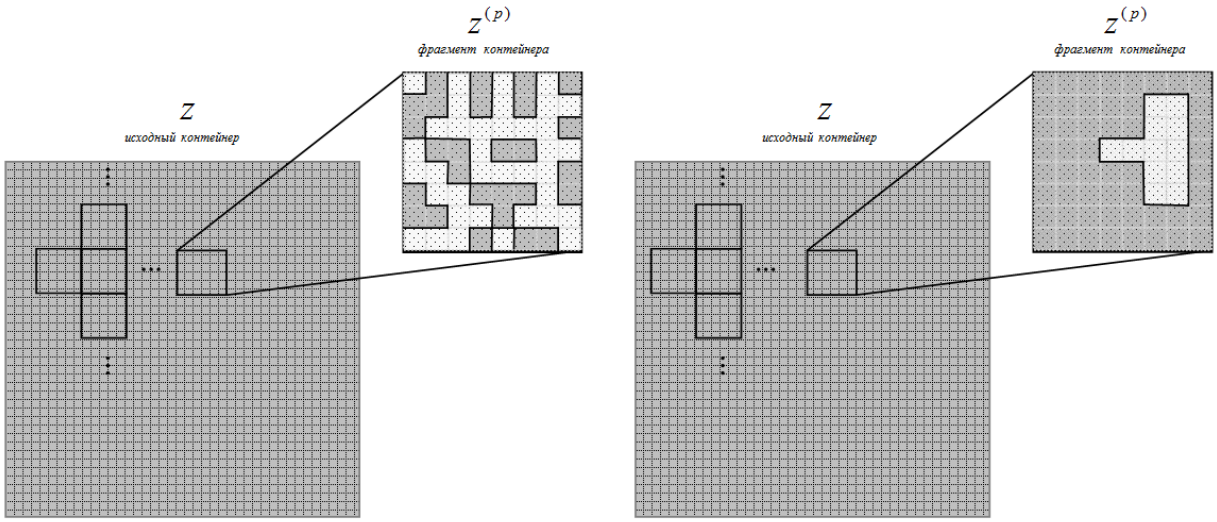
$$M[V] = 0, \quad M[VV^T] = R_{z22} - R_{z21}R_{z11}^{-1}R_{z12}$$

где $\breve{z}_{2/1}$ – имеет смысл теоретической оптимальной (в классе линейных) оценки z_2 относительно наблюдения z_1; V – стохастическая составляющая, некоррелированная с $\breve{z}_{2/1}$; $R_{z11} = M[z_1 z_1^T]$, $R_{z22} = M[z_2 z_2^T], R_{z21} = M[z_2 z_1^T]$.

Область определения Ω (форма фрагмента контейнера) при выполнении сжимающих отображений для изображений может быть произвольной конфигурации. На рисунке 2.14 представлены примеры отображения конфигурации фрагментов, подвергаемых модификации.

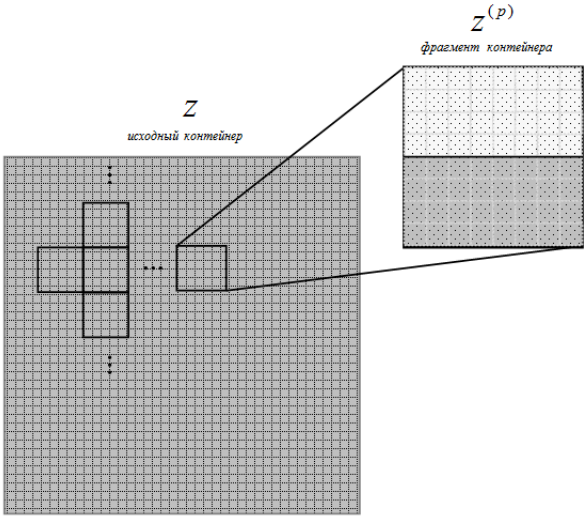


а) б)



в)

г)



д)

Рисунок 2.14 - Примеры отображения конфигурации области модификации исходного контейнера

Это области прямоугольной формы, решетки случайной конфигурации внутри области прямоугольной формы и т.п. Серым цветом выделены пиксели входной часть фрагмента Ω_1, белым цветом выделены пиксели выходной части фрагмента Ω_2, в которой происходит встраивание элементов ЦВЗ. При этом следует отметить, что рисунки 2.14а и г соответствуют случаю $N_2 < N_1$, рисунки 2.14б и в соответствуют случаю $N_2 > N_1$, рисунок 2.14д соответствует случаю $N_2 = N_1$.

В случае реализации алгоритма создания ЦВЗ на основе гетероассоциативных преобразований постановка задачи ССИ с использованием аппарата искусственных нейронных сетей и введенных обозначений может быть изменена следующим образом. Требуется с использованием функциональных возможностей нейронных сетей для любого фрагмента контейнера, представленного в виде вектора $z \in R^n$, и вектора встраиваемых данных $d \in R^1$, построить отображения

$$\hat{z}_{2/1} = F_e(z_1), \tilde{z}_{2/1} = F_c(z_1),\ \bar{z}_2 = F_m(\hat{z}_{2/1}, \tilde{z}_{2/1}, d) + (z_2 - \hat{z}_{2/1}),\ \left\|\bar{z}_2 - z_2\right\| \to \min,$$
$$\tilde{d} = F_d(\bar{z}_2),\ \tilde{d} \in D,\ \left\|d - \tilde{d}\right\| \to \min,$$

где оператор F_e реализует оптимальную оценку (прогноз) значений данных выходной части фрагмента при известных значениях входной части; оператор F_c реализует приближенную оценку (прогноз) значений данных выходной части фрагмента на основе сжимающего отображения (см. теорему п.2.1) при известных значениях входной части, обеспечивающего подготовку вектора-контейнера к встраиванию данных, оператор F_m реализует модификацию выходной части при встраивании элементов ЦВЗ, а оператор F_d – восстановление ранее скрытого сообщения.

На рисунке 2.15 схематично представлен процесс встраивания последовательности ЦВЗ в исходный файл-контейнер и последующее восстановление ЦВЗ.

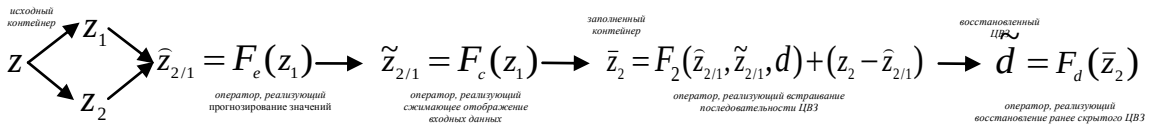


Рисунок 2.15 – Встраивание и восстановление последовательности ЦВЗ

Структура нейронной сети, реализующая оператор F_e, в общем виде приведена на рисунке 2.16а. Матрица весовых коэффициентов обозначена W_0.

Обучение нейронной сети проводится по совокупности реализаций входного и целевого вектора $\{z_1^{(p)}, z_2^{(p)}\}$, $p = \overline{1, P}$.

Необходимо также отметить, что для однослойной НС (рисунок 2.16а), имеющей N_1 входных контактов и N_2 нейронов в выходном слое и матрицу весовых коэффициентов W_0, после обучения по выборке объемом $P > N_1, P > N_2$ реализаций случайных векторов $z_1^{(p)}, z_2^{(p)}, p = \overline{1, P}$ на выходе формируется реакция, соответствующая структуре оптимальной оценке z_2 по наблюдению z_1 в виде $\hat{z}_{2/1} = \hat{R}_{z21}\hat{R}_{z11}^{-1}z_1$ $(W = \hat{R}_{z21}\hat{R}_{z11}^{-1})$ [2, 42].

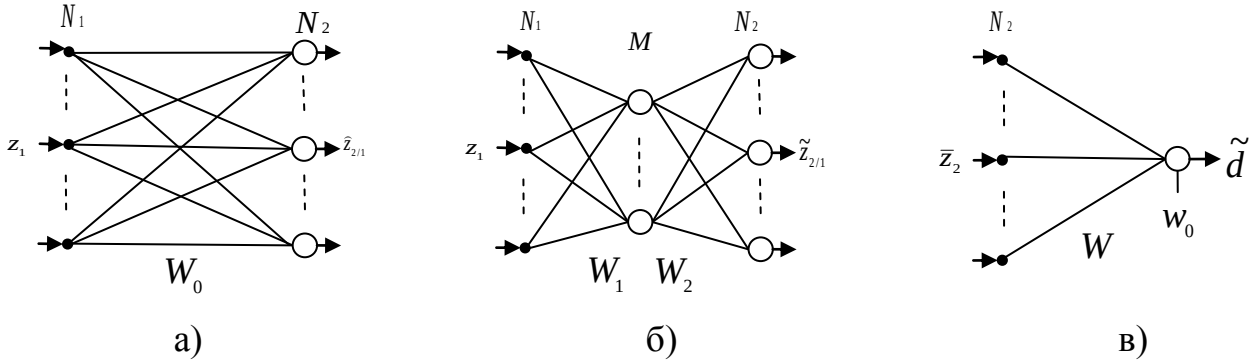


Рисунок 2.16 - Нейронная сеть, определяющая оператор связи соседних фрагментов (а); нейронная сеть, реализующая сжимающее преобразование для последующего встраивания (б); нейронная сеть, реализующая восстановление скрытой информации (в)

В результате обучения нейронной сети, представленной на рисунке 2.16а, получаемый на выходе сети вектор $\hat{z}_{2/1} = (\hat{z}_{2,1}, \ldots \hat{z}_{2,N_2})^T$ используется для прогнозирования фрагментов контейнера на основе соседних фрагментов контейнера и для определения стохастической составляющей прогноза

$$V^{(p)} = z_2^{(p)} - \hat{z}_{2/1}^{(p)}.$$

В результате обучения второй нейронной сети, представленной на рисунке 2.16б получается индивидуальное для данного набора данных сжимающее отображение с весьма незначительными потерями, при котором реально получаемый на выходе сети вектор $\tilde{z}_{2/1} = (\tilde{z}_{2,1}, \ldots \tilde{z}_{2,N_2})^T$ может быть представлен в

виде разложения по первым M собственным векторам $\varphi_i, i=\overline{1,M}$ выборочной матрицы ковариации оценки вектора z_2 относительно вектора z_1

$$\tilde{R}_{z_2}^{(P)} = \frac{1}{P-1}\sum_{p=1}^{P}\tilde{z}_{2/1}{}^{(p)}\tilde{z}_{2/1}{}^{(p),T},$$

которые являются одновременно функциями разложения Карунена-Лоэва [32, 38]. Тогда для любой реализации вектора $z_1{}^{(p)}$ получаемый на выходе вектор $\tilde{z}_{2/1}{}^{(p)}$ можно представить в виде

$$\tilde{z}_{2/1}{}^{(p)} = \sum_{i=1}^{M}\alpha_i^{(p)}\varphi_i,$$

где $\alpha_i^{(p)}$, $i=\overline{1,M}$ – коэффициенты разложения по первым $M \le N_1$ собственным векторам $\varphi_i, i=\overline{1,M}$ матрицы $\hat{R}_{z2}^{(p)}$. С целью минимизации ошибки искажения контейнера при выполнении подобного разложения будем использовать сеть, для которой $M = N_1 - 1$, тогда $\tilde{z}_{2/1}{}^{(p)}$ будет отличаться от $\hat{z}_{2/1}{}^{(p)}$ только «высокочастотной» составляющей с малой амплитудой и дисперсией соответствующей минимальному собственному числу выборочной матрицы ковариации.

После того, выполняется оператор F_m, реализующий окончательное встраивание ЦВЗ в файл контейнер. При этом вычисляется нормированный вектор

$$\varphi_{N_2} = r_{\min} \Big/ \sqrt{\left(r_{\min}{}^T r_{\min}\right)},\ r_{\min} = \frac{1}{P}\sum_{p=1}^{P}(\hat{z}_{2/1}{}^{(p)} - \tilde{z}_{2/1}{}^{(p)})$$

и каждый фрагмент, описываемый вектором $\tilde{z}_{2/1}{}^{(p)}$, $p=\overline{1,P}$, модифицируется последовательностью ЦВЗ на основе следующих соотношений с учетом стохастической составляющей прогноза:

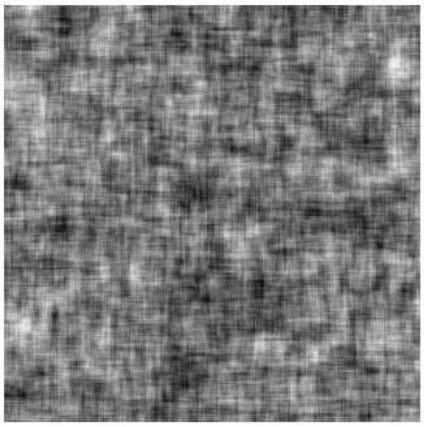
$$\bar{z}_2{}^{(p)} = \tilde{z}_{2/1}{}^{(p)} + a_m d^{(p)}\varphi_{N_2} + V^{(p)},\quad p=\overline{1,P}, \tag{2.17}$$

$$\tilde{z}_{2/1}{}^{(p)} = W_2 W_1 z_1{}^{(p)},\ \hat{z}_{2/1}{}^{(p)} = W_0 z_1{}^{(p)},\ V^{(p)} = z_2{}^{(p)} - \hat{z}_{2/1}{}^{(p)} \tag{2.18}$$

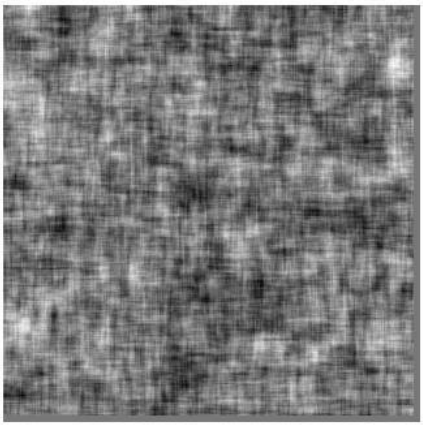
где a_m – относительная амплитуда вносимого искажения.

Для восстановления, таким образом скрытых данных и, соответственно, для реализации оператора F_d может быть использована нейронная сеть, архитектура которой показана на рисунке 2.16в. Такая сеть обучается для решения задачи классификации входного для нее вектора $\bar{z}_2$ с целью выделения значения ранее скрытого в нем элемента последовательности d. На рисунке 2.16в используются следующие обозначения: W – матрица весовых коэффициентов сети, w_0 – вектор постоянного смещения (вектор поляризации). Волнистая линия в обозначениях элементов восстанавливаемой последовательности означает наличие возможных ошибок при проведении подобной классификации последовательности входных сигналов $\bar{z}_2^{(p)}$, $p=\overline{1,P}$ в виде последовательности $\tilde{d}^{(p)}$, $p=\overline{1,P}$.

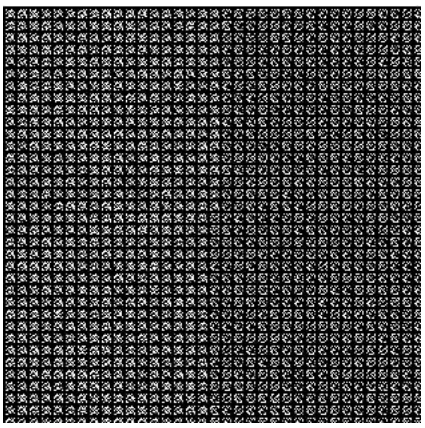
На рисунке 2.17а представлен пример реализации случайного поля с функцией корреляции первого типа, используемой в качестве контейнера.



а)



б)



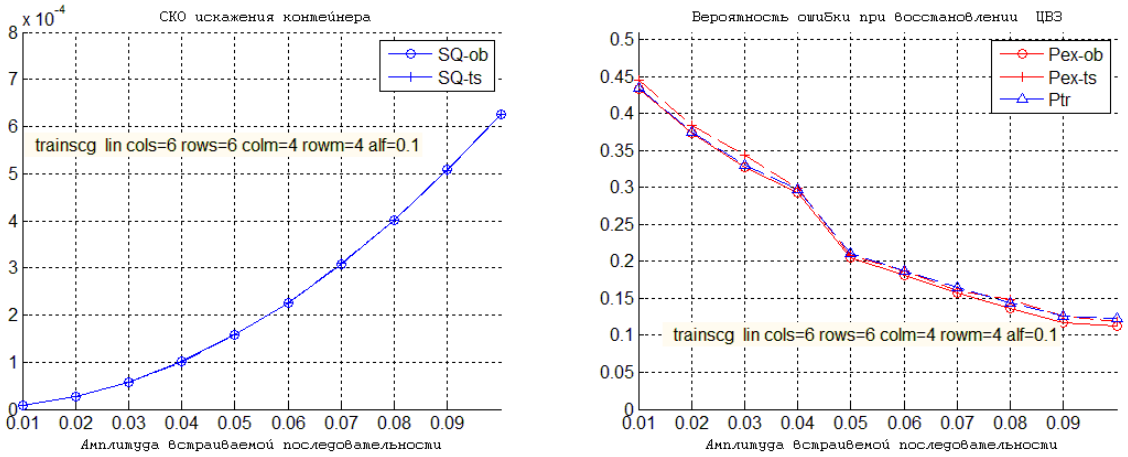
в)

Рисунок 2.17 – Пример генерации реализации случайного поля на основе функции корреляции первого типа, исходный контейнер (а); заполненный контейнер (б); разница между исходным и заполненным контейнером (в)

На рисунке 2.17б представлен результирующий заполненный контейнер, а на рисунке 2.17в показана разность между исходным и заполненным контейнером с коэффициентом умножения 1000 для обеспечения наглядности отображения.

На рисунке 2.18а представлены зависимости среднеквадратичной ошибки искажения контейнера от амплитуды a_m вносимого искажения. При моделировании задавались: исходный размер фрагмента контейнера $z^{(p)}$ $N_1 = 6\times 6$, размер фрагмента, используемого для встраивания одного элемента последовательности ЦВЗ $z_2^{(p)}$ $N_2 = 4\times 4$, объем обучающей выборки для формирования НС $P_1 = 2000$, объем тестирующей выборки $P_2 = 2000$, число эпох обучения НС $T = 5000$. На рисунке 2.18б и в приведены зависимости вероятности ошибки при восстановлении одного бита встраиваемой информации от амплитуды встраиваемой последовательности. При этом теоретически оцененная вероятность ошибки, может быть рассчитана по формуле:

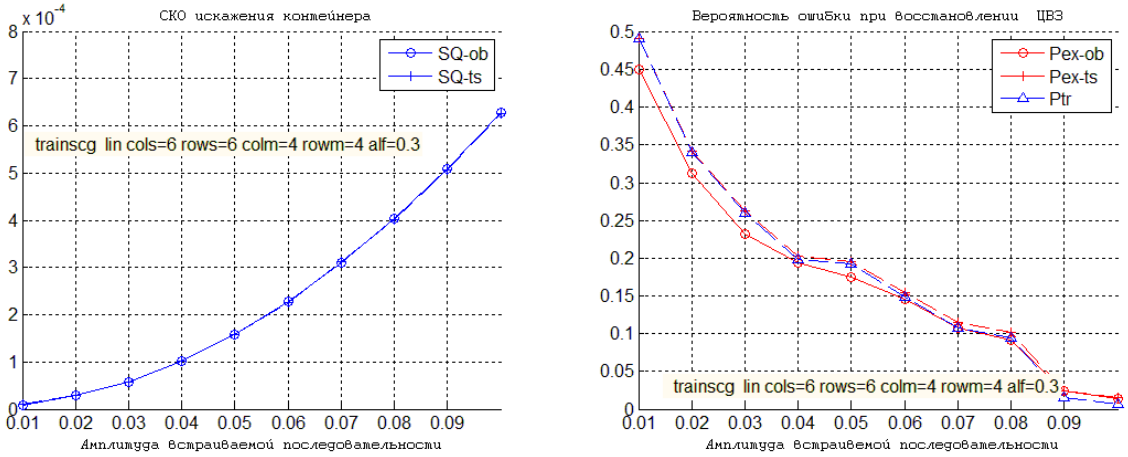
$$P_{tr} = 1 - \Phi(\beta), \beta = 0{,}5\sqrt{4\varphi_{N_2}^T R_\eta^{-1} \varphi_{N_2}}), \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt .$$



а) б)

Рисунок 2.18 - Зависимости среднеквадратичной ошибки искажения (а); вероятности ошибки восстановления ЦВЗ (б) от амплитуды встраиваемой последовательности ЦВЗ

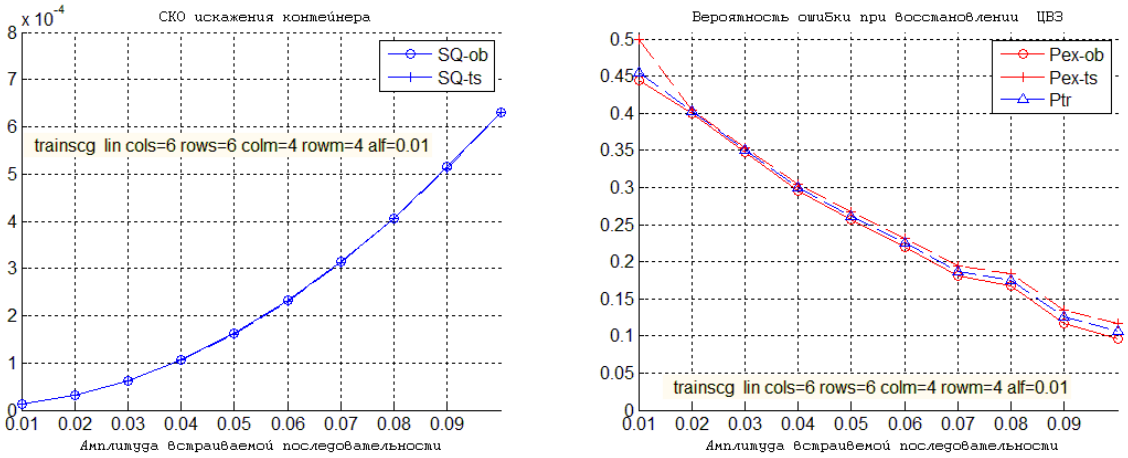
На рисунке 2.19 показаны зависимости среднеквадратичной ошибки искажения и вероятности ошибки восстановления ЦВЗ от амплитуды встраиваемой последовательности при использовании в качестве контейнера реализации случайного поля с функцией корреляции второго типа.



а) б)

Рисунок 2.19 - Зависимости среднеквадратичной ошибки искажения (а); вероятности ошибки восстановления ЦВЗ (б) от амплитуды встраиваемой последовательности ЦВЗ

На рисунке 2.20 показаны зависимости среднеквадратичной ошибки искажения и зависимость вероятности ошибки восстановления ЦВЗ от амплитуды встраиваемой последовательности при использовании в качестве контейнера реализации случайного поля с функцией корреляции третьего типа.



а) б)

Рисунок 2.20 - Зависимости среднеквадратичной ошибки искажения (а); вероятности ошибки восстановления ЦВЗ (б) от амплитуды встраиваемой последовательности ЦВЗ

Анализ полученных зависимостей показывает, что при увеличении амплитуды встраиваемой последовательности значение СКО возрастает, при этом следует отметить, что результаты СКО полученные для автоассоциативных и гетероассоциативных преобразований имеют ошибку порядка 10^{-4}. Для случая автоассоциативных преобразований СКО определяется не только амплитудой встраиваемой последовательности, но и величиной отбрасываемой высокочастотной компоненты. Для преобразований гетероассоциативного типа величина СКО монотонно возрастает из-за внесения модифицируемой компоненты с увеличивающейся амплитудой и наличия маскирующей стохастической составляющей. Вероятность восстановления информации при увеличении амплитуды стремится к нулю для преобразований гетероассоциативного типа, что обуславливает увеличение достоверности получаемых результатов при восстановлении элементов $\tilde{d}^{(p)}$ двоичной

последовательности ЦВЗ. Вероятность ошибки восстановления ЦВЗ при использовании автоассоциативного преобразования равна нулю, так как матрица ковариации R_η в этом случае является вырожденной.

Выводы по главе

1. Выполнены теоретические обоснования сходимости весовых коэффициентов двухслойной линейной нейронной сети с сокращенным числом нейронов в скрытом слое к компонентам собственных векторов, получаемых при решении обобщенной задачи на собственные значения. При этом на выходе сети формируется приближение оптимальной линейной оценки в виде разложения по первым собственным векторам матрицы ковариации оценки, число которых равно числу нейронов в скрытом слое. Показано также, что при настройке коэффициентов простейших линейных нейронных сетей в условиях прямого и косвенного обучения по представительным выборкам наблюдаемых данных имеется возможность сформировать оценки векторов наблюдаемых и ненаблюдаемых параметров, эквивалентные по своим свойствам статистически оптимальным.

2. Исходя из полученных теоретических обоснований целесообразно, в ходе дальнейших исследований, рассмотреть два базовых варианта нейросетевых функциональных моделей и алгоритмов преобразования данных в интересах создания цифровых водяных знаков: на основе нейронных сетей, реализующих сжимающие отображения автоассоциативного типа преобразований и на основе нейронных сетей, реализующих сжимающие отображения гетероассоциативного типа. Сформулирована постановка задачи и общая модель встраивания последовательности ЦВЗ в общем виде и с учетом применения искусственных нейронных сетей различной архитектуры.

3. Результаты проведенных экспериментальных исследований на основе имитационного статистического моделирования, когда в качестве контейнера рассматривались реализации однородного гауссовского случайного поля с разными корреляционными функциями, показывают, что при реализации рассмотренных вариантов моделей и алгоритмов создания ЦВЗ для

преобразований автоассоциативного и гетероассоциативного типа обеспечивается относительная средняя квадратичная ошибка искажения фрагментов контейнера порядка 10^{-4}. Вероятность ошибки восстановления элементов двоичной последовательности ЦВЗ для гетероассоциативных преобразований достигает порядка 10^{-1} при относительной амплитуде вносимого $a_m = 0.01$, порядка 10^{-2} при относительной амплитуде вносимого искажения $a_m = 0.09$, для автоассоциативных преобразований вероятность ошибки восстановления ЦВЗ равна нулю.

3. Алгоритмы создания цифровых водяных знаков для объектов графических и звуковых форматов представления данных

В главе 2 настоящей диссертации были проведены теоретические обоснования возможности построения нейросетевых функциональных моделей и алгоритмов ССИ в интересах создания ЦВЗ. Рассмотренные базовые алгоритмы создания ЦВЗ ориентированы на контейнеры с вещественной формой представления данных, статистическими эквивалентами которых являются реализации случайных процессов и полей. Применение их в непосредственном виде для контейнеров, содержащих данные, представленные в целочисленном формате (изображения, видео - и аудио файлы) не является достаточно эффективным. Поэтому представляет интерес развитие и исследование базовых алгоритмов для создания ЦВЗ в распространенных объектах цифрового контента имеющих графические и звуковые форматы представления данных.

Таким образом, в настоящей главе рассмотрены нейросетевые функциональные модели и универсальные алгоритмы преобразования данных в интересах решения задачи контентно-независимого встраивания и восстановления сообщений (создания ЦВЗ) в контейнерах с целочисленным форматом данных, отвечающих требованиям повышенной робастности, визуальной и статистической незаметности встраиваемой метки. При этом получены закономерности, описывающие потенциальные характеристики качества создаваемых ЦВЗ для объектов различных классов (цветные изображения в формате RGB и YCbCr, аудио-файлы в формате WAV).

3.1. Модели и алгоритмы создания цифровых водяных знаков для цветных изображений

Для работы с файлами-контейнерами, имеющими целочисленный формат представления данных, необходимо определенным образом модифицировать описываемые в п. 2.2 базовые алгоритмы создания цифровых водяных знаков при реализации отображений автоассоциативного и гетероассоциативного типа. Проблема состоит в том, что данные, используемые для обучения нейронной сети, а также параметры сети должны быть заданы в вещественном формате, тогда как

входные и выходные данные, а также данные, используемые для модификации контейнера, при встраивании имеют целочисленное представление. Поэтому требуется соблюдать определенный порядок преобразования данных в ходе выполнения процедур встраивания ЦВЗ.

Рассмотрим процесс обработки информации для модели первого типа, основанной на применении автоассоциативной нейронной сети и подробно описанной в п.2.2.1. Прежде всего, формируется соответствующая совокупность фрагментов контейнера $\{g^{(p)}\}, p=\overline{1,P}$ для встраивания, в рамках которой, определятся исходный вектор $g^{(p)}$, описывающий этот фрагмент. При этом для цветных изображений при формировании вектора проводится соответствующая развертка всех цветовых компонент.

На рисунке 3.1 показан исходный контейнер (цветное изображение в формате RGB) поделенный на блоки - фрагменты контейнера. Длина вектора n в данном случае определяется конфигурацией и размером фрагмента. Цветным RGB изображением называется, состоящий из цветных пикселей, массив, причем каждый такой пиксель является триплетом, элементы которого соответствуют трем цветовым компонентам: красному, зеленому, синему. Изображения в формате RGB составлены их трех 8-битных цветовых каналов.

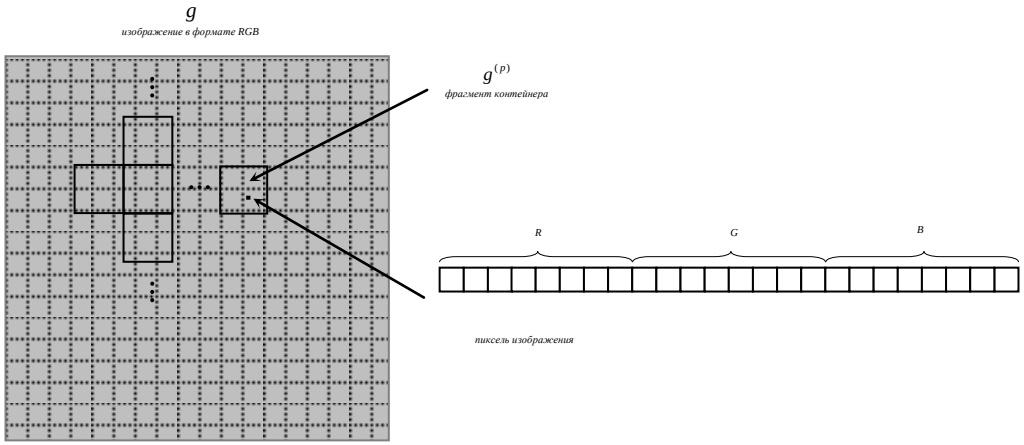


Рисунок 3.1 - Исходный контейнер (цветное изображение в формате RGB)

Каждый такой вектор, имеющий целочисленное представление с разрядностью представления данных, характерной для данного типа файла-

контейнера, преобразуется в вектор $z^{(p)}$, содержащий вещественные данные в диапазоне значений –0.5…+0.5 (с центрированием)

$$z^{(p)} = (z_1^{(p)}, \ldots, z_n^{(p)})^T = F_{double}(g_1^{(p)}, \ldots, g_n^{(p)}) - 0.5,$$

где $F_{double}(g_1^{(p)}, \ldots, g_n^{(p)})$ – функция преобразования целочисленных данных в данные вещественного формата в диапазоне 0,…,1. Далее в соответствии с приведенным выше алгоритмом проводится обучение нейронной сети F_c по выборке $z^{(p)}$, $p = \overline{1,P}$.

После завершения процесса обучения входная последовательность данных $z^{(p)}$, $p = \overline{1,P}$ пропускается через сеть и формирует результирующую последовательность $\tilde{z}^{(p)}$, $p = \overline{1,P}$. Для каждого $\tilde{z}^{(p)}$ выполняется преобразование, обеспечивающее подготовку вектора-контейнера к встраиванию данных

$$\tilde{z}_{-}^{(p)} = F_{double}[F_{\text{int}}(\tilde{z}^{(p)} + 0.5)] - 0.5, \; p = \overline{1,P},$$

где $F_{\text{int}}(\tilde{z}^{(p)} + 0.5)$ – функция, реализующая преобразование вещественных данных в диапазоне 0,…,1 к целочисленному формату представления с округлением в соответствии с используемой в контейнере разрядностью представления данных.

На основе полученной реализации $\tilde{z}_{-}^{(p)}$, $p = \overline{1,P}$ определяется нормированный собственный вектор, соответствующий минимальному собственному числу

$$\varphi_n = r_{\min} \Big/ \sqrt{\left(r_{\min}^{\;T} r_{\min}\right)}, \; r_{\min} = \frac{1}{P} \sum_{p=1}^{P} (z^{(p)} - \tilde{z}_{-}^{(p)}).$$

Для его использования при последующем встраивании данных в фрагменты контейнера получим вектор φ_n^* на основе следующего преобразования:

$$\varphi_n^* = F_{double}[F_{\text{int}}(a_m \varphi_n + 0.5)] - 0.5,$$

где a_m – относительная амплитуда вносимого искажения, определяемая по отношению к максимальному значению уровня яркости пикселей. В данном случае она задается после преобразования элементов исходного изображения,

находящихся в диапазоне значений 0,…,255, в вещественный формат представления данных в диапазоне 0,…,1.

Наконец, при встраивании сообщения, образующего скрываемую последовательность данных ЦВЗ $d^{(p)}, p=\overline{1,P}$, формируется последовательность заполненных фрагментов контейнера $\bar{z}^{(p)}$, $p=\overline{1,P}$, представленных в целочисленном формате, на основе следующего соотношения:

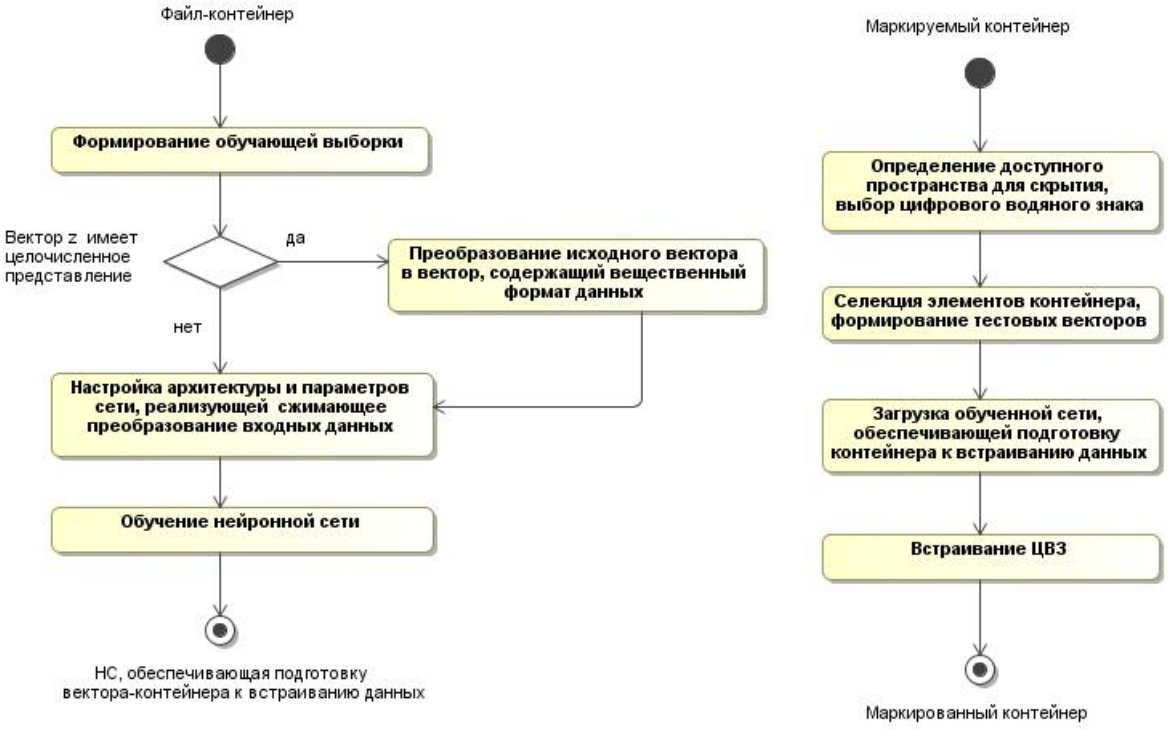
$$\bar{z}^{(p)} = F_{\text{int}}(\tilde{z}_{-}^{(p)} + d^{(p)}\varphi_n^* + 0.5), \quad p=\overline{1,P}. \tag{3.1}$$

Таким образом, в данном случае процесс встраивания состоит в добавлении или вычитании (в зависимости от знака $d^{(p)}$, $p=\overline{1,P}$) малого по амплитуде сигнала, который по форме соответствует округленному собственному вектору выборочной матрицы ковариации входной совокупности данных.

Получаемая на основе (3.1) последовательность используется для обучения нейронной сети (рисунок 2.6б), обеспечивающей восстановление ЦВЗ. Для этого выполняется преобразование, приводящее $\bar{z}^{(p)}$, $p=\overline{1,P}$ в вещественный формат

$$\bar{z}_d^{(p)} = F_{double}(\bar{z}^{(p)}) - 0.5, \quad p=\overline{1,P}.$$

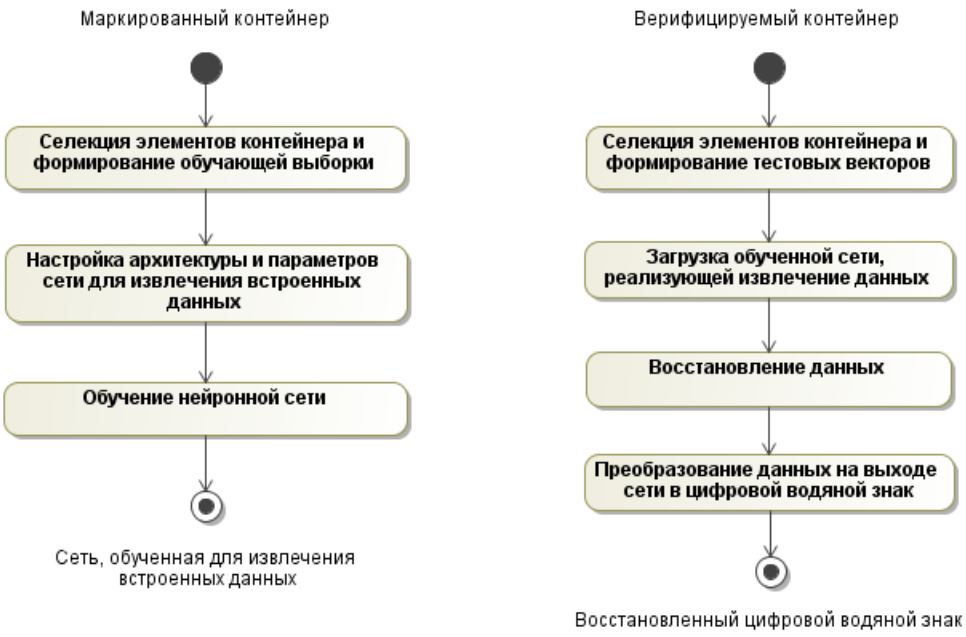
Обобщенную схему встраивания информации в рамках описанного подхода удобно представить в нотации языка UML, приведенной на рисунке 3.2. Схема ССИ включают два основных этапа. На первом этапе производится формирование обучающих множеств, настраиваются параметры ИНС, происходит преобразование исходного вектора в вещественный формат данных, происходит процесс ее обучения (рисунок 3.2а). На втором этапе выбирается подходящий контейнер и реализуется алгоритм встраивания (рисунок 3.2б).



а) б)

Рисунок 3.2 – Обобщенная схема встраивания информации при реализации предлагаемого нейросетевого метода

Схема извлечения информации дана на рисунке 3.3. Первоначально осуществляется обучение НС, предназначенной для восстановления скрытых данных (рисунок 3.3а). Дальнейшее извлечение данных реализуется путем подачи в необходимом порядке элементов контейнера, содержащих встроенный ЦВЗ, на «восстанавливающую» нейронную сеть (рисунок 3.3б).



а) б)

Рисунок 3.3 – Обобщенная схема извлечения информации при реализации предлагаемого нейросетевого метода

По схожей схеме реализуется процесс обработки информации для модели второго типа, основанной на применении гетероассоциативной нейронной сети и подробно описанной в п.2.2.2.

Оценка эффективности рассмотренных алгоритмов создания ЦВЗ, проводилась с использованием реальных объектов цифрового контента. В качестве показателей эффективности использовались среднеквадратичная ошибка искажения контейнера, вероятность ошибки при восстановлении встроенного ЦВЗ и максимальная абсолютная ошибка искажения контейнера.

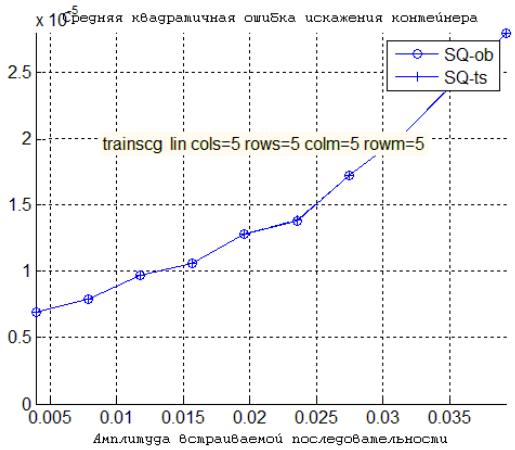
Пример исходного контейнера-изображения приведен на рисунке 3.4а.

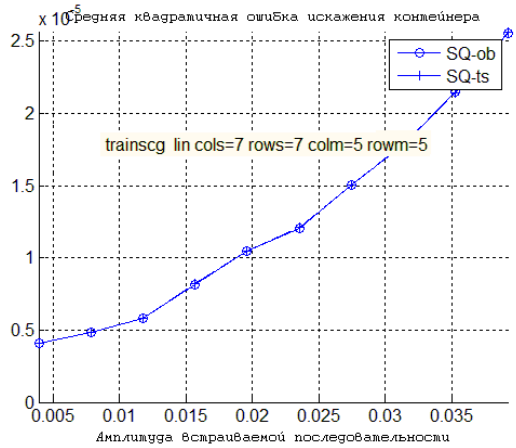
а) б) в)

Рисунок 3.4 –Исходный контейнер (а); заполненный контейнер (б); разница между исходным и заполненным контейнером (в)

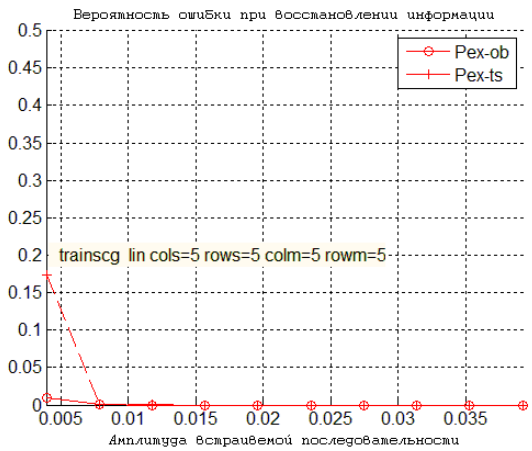
На рисунке 3.4б представлен результирующий заполненный контейнер, а на рисунке 3.4в показана разность между исходным и заполненным контейнером с коэффициентом умножения 1000 для обеспечения наглядности отображения.

Результаты работы нейросетевого алгоритма функционального встраивания в файлы графического формата для автоассоциативных преобразований представлены на рисунке 3.5а, в и д для гетероассоциативных преобразований представлены на рисунке 3.5б, г и е.

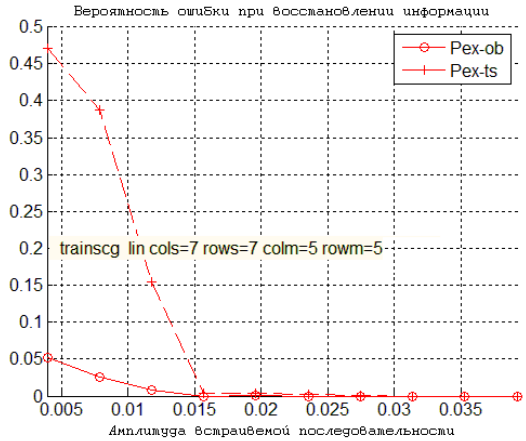




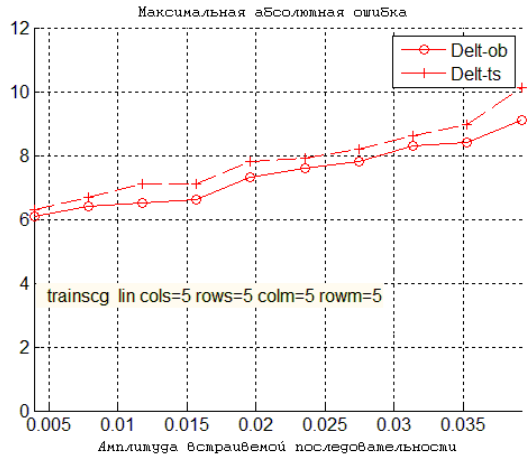
а) б)



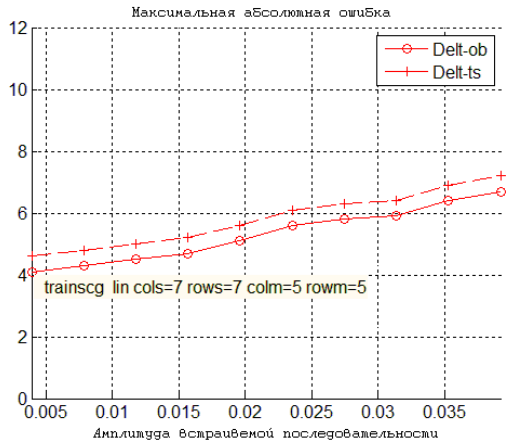
в)



г)



д)



е)

Рисунок 3.5 - Зависимость среднеквадратичной ошибки искажения (а, б), зависимость вероятности ошибки восстановления ЦВЗ (в,г), максимальная абсолютная ошибка искажения контейнера (д,е)

На графиках используются следующие обозначения: SQ-ob – среднеквадратичная ошибка искажения контейнера при тестировании на данных, используемых для обучения нейронной сети; SQ-ts – среднеквадратичная ошибка искажения контейнера при тестировании на данных, не участвовавших в обучении нейронной сети; Pex-ob – оценка вероятности ошибки при восстановлении ЦВЗ, при тестировании на данных, используемых для обучения нейронной сети; Pex-ts – оценка вероятности ошибки при восстановлении ЦВЗ при тестировании на данных, не участвовавших в обучении нейронной сети; Delt-ob – абсолютная ошибка искажения контейнера при тестировании на данных,

используемых для обучения нейронной сети; Delt-ts – абсолютная ошибка искажения контейнера при тестировании на данных, не участвовавших в обучении нейронной сети.

Анализ полученных зависимостей показывает, что используемый алгоритм успешно применяется для задач создания ЦВЗ. Следует отметить, что исследования проводились не только для цветовой модели RGB, но и для цветовой модели YCbCr. Пространство YCbCr является популярным форматом эффективного представления цветных изображений, используемым в цифровом видео [10]. При этом для перевода из цветовой модели RGB в YCbCr необходимо выполнить соответствующие преобразования.

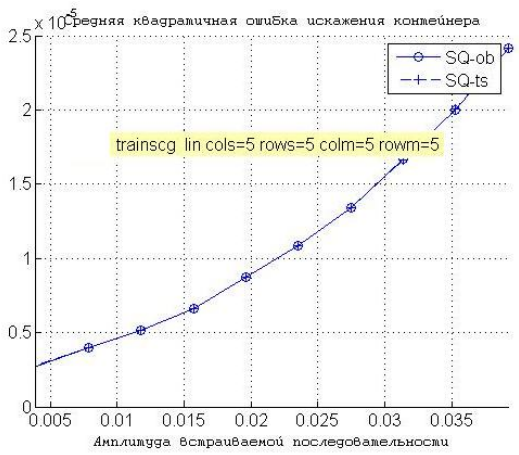
Составляющая Y содержит информацию о яркости изображения, а составляющие Cb и Cr (так называемые цветоразностные составляющие) - о его цветности.

Преобразование из цветовой модели RGB, где составляющие принадлежат диапазону 0…1, в модель YCbCr имеет вид:

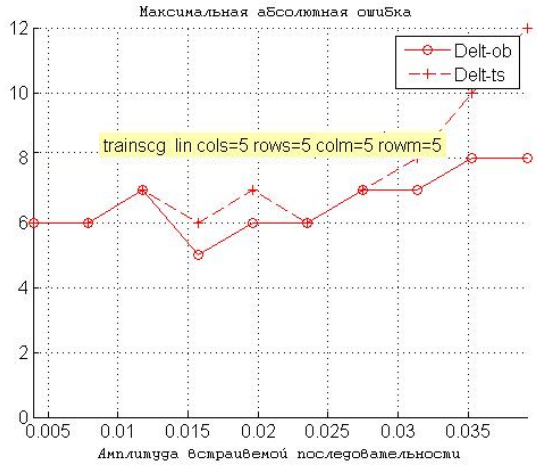
$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.0 \\ 112.0 & -93.784 & -18.214 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

В результате данного преобразования составляющая Y принадлежит диапазону значений 16…235, а составляющие Cb и Cr принадлежат диапазону значенпй 16…240. Оставшиеся допустимые значения вне указанных диапазонов (0…15 и 236…255 для Y, 0…15 и 241…255 для Cb, Cr) используются для внесения дополнительной информации (например, звука), которая передается вместе с видеопотоком данных.

В данном случае исследовался только алгоритм первого типа (основанный на преобразованиях автоассоциативного типа). На рисунке 3.6 приведены зависимости среднеквадратичной ошибки искажения контейнера и вероятность ошибки при восстановлении информации от амплитуды встраиваемой последовательности, полученные для аналогичных условий.



а) б)



в)

Рисунок 3.6 - Зависимость среднеквадратичной ошибки искажения(а), зависимость вероятности ошибки восстановления ЦВЗ (б), максимальная абсолютная ошибка искажения контейнера (в)

На рисунке 3.6 графики SQ-ob и SQ-ts практически идентичны и совпадают. Величина вероятности ошибки при восстановлении ЦВЗ в данном случае значительно меньше, чем для цветовой модели RGB, что свидетельствует и том, что реализованный алгоритм может еще более успешно использоваться для цветовой модели YCbCr.

Качество алгоритма встраивания ЦВЗ в первую очередь обуславливается визуальной незаметностью искажающих изменений результирующего контейнера по сравнению с исходным контейнером. Надежность алгоритма восстановления ЦВЗ определяется его устойчивостью по отношению к различным

трансформациям (случайным или преднамеренным), выполняемым над заполненным контейнером. В главе 4 подробно будут рассмотрены типовые методики оценки качества цифровых водяных знаков.

Как видно из рисунка 3.4в результате применении алгоритма были получены результирующие изображения очень высокого качества. При сравнении маркированных контейнеров с их оригиналами отыскать визуальные отличия практически невозможно.

3.2. Модификации алгоритма создания ЦВЗ с целью адаптации по уровню ошибки искажения контейнера и противодействия преобразованиям JPEG

Ранее в п.2.2.1 было показано, что с целью минимизации ошибки искажения контейнера будет использоваться сеть, для которой $M = n - 1$, тогда $\tilde{z}^{(p)}$ будет отличаться от $z^{(p)}$ только «высокочастотной» составляющей с малой амплитудой и дисперсией соответствующей минимальному собственному числу выборочной матрицы ковариации, которая определяется следующим соотношением:

$$\Delta_z^{(p)} = z^{(p)} - \tilde{z}^{(p)}.$$

С учетом требования адаптации по возникающим искажениям необходимо выполнить следующие операции. Рассчитывается относительная энергия вносимых искажений, возникающих при сжатии фрагмента при выполнении оператора F_c

$$E^{(p)} = \Delta_z^{(p)T} \Delta_z^{(p)} \Big/ \tilde{z}^{(p)T} \tilde{z}^{(p)}.$$

Затем величина $E^{(p)}$ сравнивается с некоторым выбранным пороговым значением $0 \le r < 1$. Случай $r = 0$ соответствует отсутствию адаптации.

В случае если $E^{(p)}$ не превышает значение r, то в данный фрагмент контейнера встраивается один бит информации $d^{(p)} \in \{-1, +1\}$, $p = \overline{1, P}$, в противном случае встраивание не происходит. При этом возможны два варианта:

- производится запоминание номеров фрагментов, в которые осуществлялось встраивание элементов бинарной последовательности;

- номера фрагментов не фиксируются, но при восстановлении ЦВЗ производится идентификация фрагментов, в которых производилось встраивание.

Встраивание последовательно ЦВЗ производится рассмотренным ранее способом.

В случае если ранее номера фрагментов запоминались, то НС имеет один выход и реализует стандартное двухальтернативное решающее правило, определяющее с некоторой вероятностью значения элемента встроенной последовательности $\tilde{d}^{(p)} \in \{-1,+1\}$ (структура НС для восстановления скрытых данных приведена на рисунке 3.7а).

В случае, когда запоминание номеров фрагментов ранее не проводилось, НС имеет три выхода и обеспечивает формирование решений, отражающие три альтернативы, $w^{(p)} = 0$, идентифицирующее отсутствие встраивания элемента сообщения в данный фрагмент, $w^{(p)} = 1$, что соответствует идентификации значения $\tilde{d}^{(p)} = 1$, $w^{(p)} = -1$, что соответствует идентификации значения $\tilde{d}^{(p)} = -1$ (структура НС для восстановления скрытых данных приведена на рисунке 3.7б.

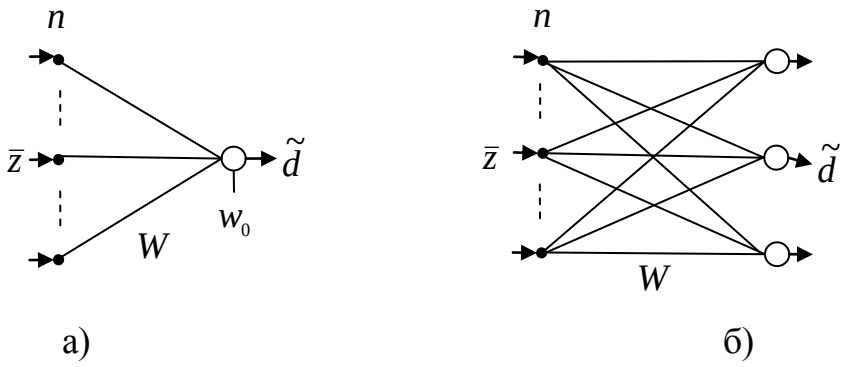
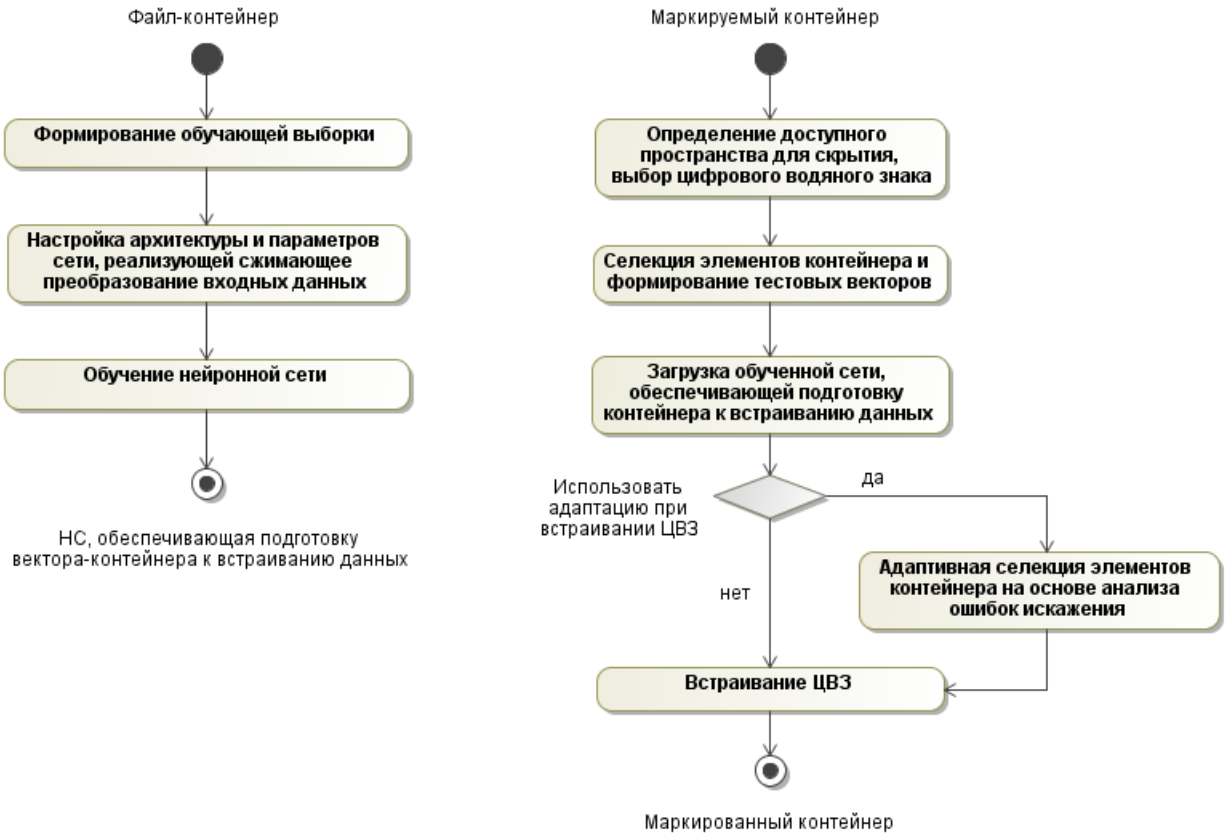


Рисунок 3.7 – Нейронная сеть, реализующая восстановление скрытой информации для случая стандартного двухальтернативного решающее правила (а); для случая трехальтернативного решающего правила (б)

Обобщенную схему встраивания информации в рамках описанного подхода удобно представить в нотации языка UML, приведенной на рисунке 3.8. Схема ССИ включают два основных этапа. На первом этапе производится формирование обучающих множеств, настраиваются параметры ИНС и происходит процесс ее обучения (рисунок 3.8а). На втором этапе выбирается подходящий контейнер,

формируются векторы входных воздействий, и реализуется алгоритм встраивания (рисунок 3.8б).

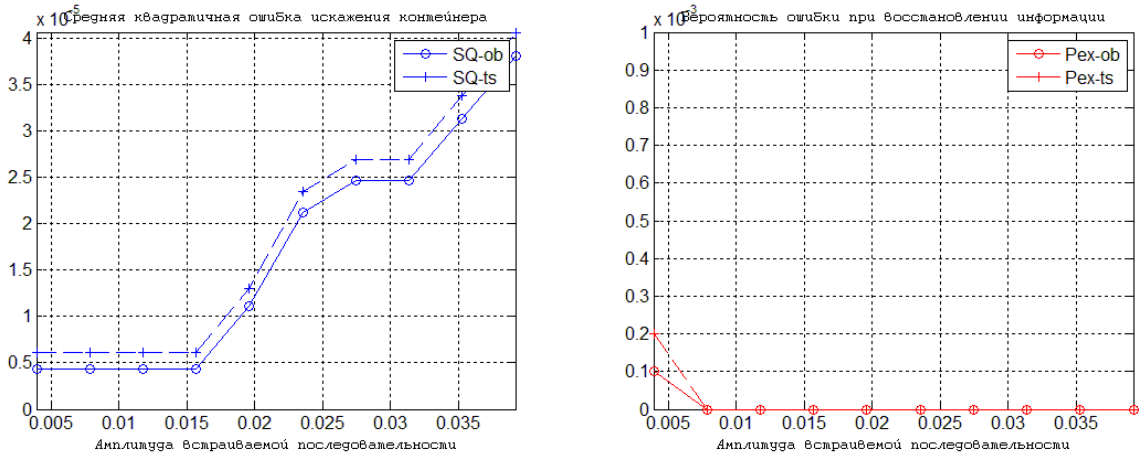
Схема извлечения информации представлена выше на рисунке 3.2 и остается без изменений.



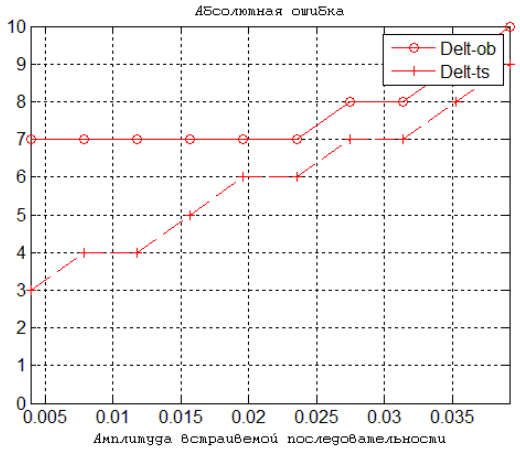
а) б)

Рисунок 3.8 - Обобщенная схема встраивания информации при реализации адаптивного нейросетевого метода

Экспериментальный анализ разработанного алгоритма проводился для реальных изображений формата BMP.



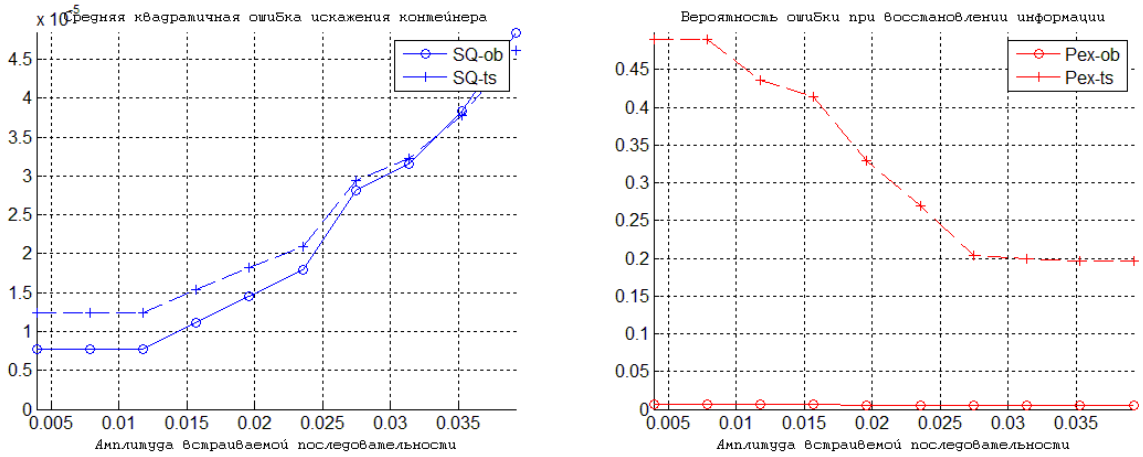
а) б)



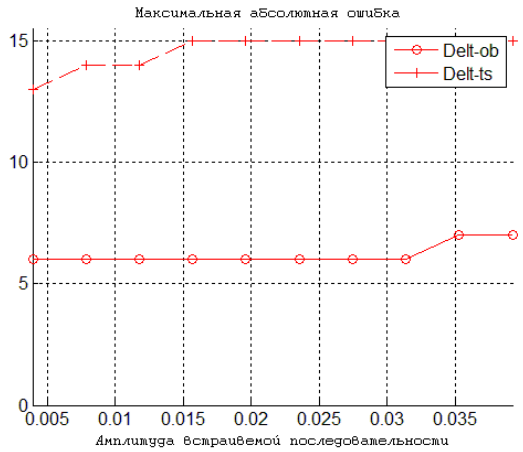
в)

Рисунок 3.9 - Зависимости для среднеквадратичной ошибки искажения контейнера (а), вероятности ошибки восстановления ЦВЗ (б), максимальной абсолютной ошибки искажения контейнера (в)

На рисунке 3.9 приведены зависимости средней квадратичной и абсолютной ошибок искажения контейнера и вероятности ошибки восстановления контейнера от амплитуды встраиваемой последовательности, полученные для следующих условий, полученные для следующих условий: амплитуда встраиваемого процесса при обучении $a_m = 1/255$, число эпох обучения $T = 5000$. Для восстановления скрытых данных используется НС, имеющая один выход и реализует стандартное двухальтернативное решающее правило (рисунок 3.7а).



а) б)



в)

Рисунок 3.10 - Зависимости для среднеквадратичной ошибки искажения контейнера (а), вероятности ошибки восстановления ЦВЗ (б), максимальной абсолютной ошибки искажения контейнера (в)

На рисунке 3.10 приведены зависимости средней квадратичной и абсолютной ошибок искажения контейнера и вероятности ошибки восстановления контейнера от амплитуды встраиваемой последовательности для случая запоминания номера фрагментов где реализовано встраивание, полученные для следующих условий: амплитуда встраиваемого процесса при обучении $a_m = 1/255$, число эпох обучения $T = 5000$. Для восстановления скрытых данных используется НС, имеющая три выхода и обеспечивает формирование решений, отражающие три альтернативы (рисунок 3.7б).

Приведенные результаты проведенных исследований в случае использования, для восстановления встроенного ЦВЗ, НС, имеющую три выхода показывают худшие результаты, чем при использовании стандартного двухальтернативного решающего правила, что обусловлено возникающими трудностями при классификации полученных решений на выходе НС.

Представляет интерес и развитие предложенных алгоритмов для реализации процедуры создания и сохранения ЦВЗ в файлах графических форматов, которые могут быть подвергнуты сжатию в стандарте JPEG. При сжатии изображений по стандарту JPEG отмечается появление на восстановленных изображениях при высоких степенях сжатия характерных артефактов: изображение рассыпается на блоки размером 8×8 пикселей (этот эффект особенно заметен на областях изображения с плавными изменениями яркости), в областях с высокой пространственной частотой (например, на контрастных контурах и границах изображения) возникают артефакты в виде шумовых ореолов.

В таблице 3.1 представлены примеры изображений в формате JPEG с различными уровнями сжатия. (1%, 5%, 10%, 50%, 75%, 100%).

Таблица 3.1 - Примеры искажений при сжатии изображений в формате JPEG

1%, 2 Кб	5%, 3 Кб
10%, 4 Кб	50%, 5 Кб
75%,7 Кб	100%, 43 Кб

С целью повышения устойчивости алгоритмов создания ЦВЗ к преобразованиям типа JPEG-компрессии с различными коэффициентами сжатия предложена их следующая модификация. Она основана на использовании в качестве единичных элементов фрагментов файла-контейнера, для построения обучающих и целевых выборок, значений средней яркости в блоках пикселей

размерностью $q \times q$ ($q = 8, q = 16$). При $q = 1$ получаем, как частный случай, ранее рассмотренный в п.3.1 алгоритм.

Рассмотрим реализацию данного подхода при использовании алгоритма, реализующего сжимающее отображение автоассоциативного типа. Исходный файл-контейнер (цветное изображение) g, имеющий целочисленное представление, с учетом трех цветовых компонент (красной, зеленой, синей), преобразуется в файл-контейнер s, содержащий вещественные данные в диапазоне значений –0.5…+0.5 (с центрированием)

$$s^{(r,g,b)} = F_{double}(g^{(r,g,b)}) - 0.5.$$

Затем полученный файл-контенейнер последовательно разбивается на блоки $b_{ij}^{(r,g,b)}$ размерностью $q \times q$, с учетом трех цветовых компонент, и таким образом получаем $B^{(r,g,b)} = \left\| b_{ij}^{(r,g,b)} \right\|$, $i = \overline{1, I}$, $j = \overline{1, J}$, $B^{(r,g,b)}$ – матрицы блоков размерностью $q \times q$ для каждой из цветовых компонент, где I – общее количество полученных блоков по оси OX, J – общее количество полученных блоков по оси OY . Отметим, что если подобное разбиение не охватывает каждый пиксель исходного контейнера, то последующее встраивание не будет реализовываться для данных пикселей.

Далее для элементов, соответствующих каждому блоку $b_{ij}^{(r,g,b)}$ размерностью $q \times q$, вычисляется среднее арифметического значений яркости для каждой цветовой компоненты в отдельности

$$b_{m_{ij}}^{(r,g,b)} = G[b_{ij}^{(r,g,b)}],\ B_m^{(r,g,b)} = \left\| b_{m_{ij}}^{(r,g,b)} \right\|$$

где $G[...]$ – символ операции вычисления среднего арифметического; $B_m^{(r,g,b)}$ – матрица средних значений яркости блоков для каждой из цветовых компонент.

Затем на основе матрицы средних значений блоков (которая теперь играет роль исходного изображения-контейнера) формируется соответствующая совокупность фрагментов контейнера $\{z_m^{(p)}\}, p = \overline{1, P}$ для встраивания, в рамках которой, определятся исходный вектор $z_m^{(p)} = (z_{m_1}^{(p)}, \ldots, z_{m_n}^{(p)})^T$, описывающий этот

фрагмент. Длина вектора n в данном случае определяется конфигурацией и размером фрагмента. При этом для цветных изображений при формировании вектора проводится соответствующая развертка всех цветовых компонент в единый вектор.

Далее в соответствии с приведенным ранее в п.3.1 алгоритмом проводится обучение нейронной сети F_c по выборке $z_m^{(p)}, p=\overline{1,P}$.

После завершения процесса обучения входная последовательность данных $z_m^{(p)}, p=\overline{1,P}$ пропускается через сеть и формирует результирующую последовательность $\tilde{z}_m^{(p)}, p=\overline{1,P}$.

На основе полученной реализации $\tilde{z}_m^{(p)}$, $p=\overline{1,P}$ определяется нормированный собственный вектор, соответствующий минимальному собственному числу

$$\varphi_n = r_{\min} \Big/ \sqrt{\left(r_{\min}^T r_{\min}\right)},\ r_{\min} = \frac{1}{P}\sum_{p=1}^{P}(z_m^{(p)} - \tilde{z}_m^{(p)}).$$

Для его использования при последующем встраивании данных в фрагменты контейнера получим вектор φ_n^* на основе следующего преобразования:

$$\varphi_n^* = a_m \varphi_n,$$

где a_m – амплитуда вносимого искажения.

Наконец, при встраивании сообщения (3.2), образующего скрываемую последовательность данных ЦВЗ $d^{(p)}, p=\overline{1,P}$, формируется последовательность $\bar{z}_m^{(p)}$, $p=\overline{1,P}$ новых усредненных значений для каждого блока пикселей размерностью $q \times q$

$$\bar{z}_m^{(p)} = \tilde{z}_m^{(p)} + d^{(p)}\varphi_n^*, \quad p=\overline{1,P}. \tag{3.2}$$

Для результирующего преобразования необходимо выполнить следующую последовательность действий: вычесть старое усредненное значение яркости $z_m^{(p)}$ пикселей в каждом блоке, и добавить новое значение $\bar{z}_m^{(p)}$ к элементам соответствующего фрагмента (обозначим как $s^{(p)}$):

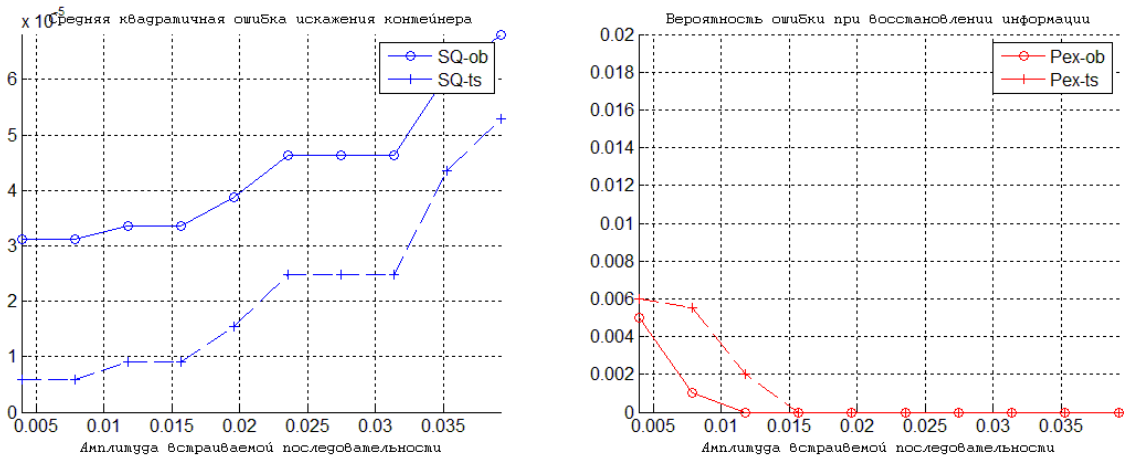
$$\Delta^{(p)} = s^{(p)} - z_m^{(p)};$$
$$\bar{s}^{(p)} = \Delta^{(p)} + \bar{z}_m^{(p)}.$$

Затем формируется последовательность заполненных фрагментов контейнера, представленных в целочисленном формате

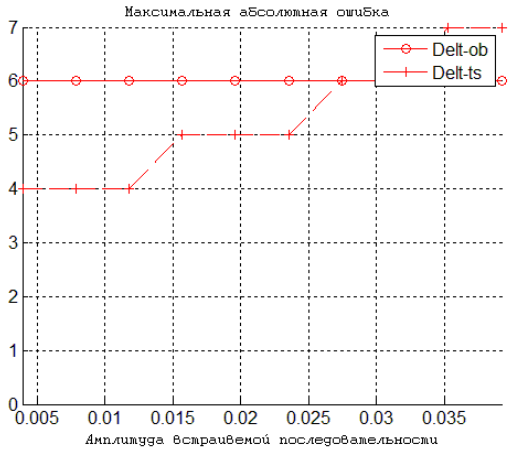
$$\bar{g}^{(p)} = F_{\text{int}}(\bar{s}^{(p)} + 0.5), \quad p = \overline{1, P}.$$

Восстановление скрытого ЦВЗ реализуется аналогичным рассмотренному ранее в п.3.1. алгоритмом.

На рисунке 3.11 приведены зависимости средней квадратичной и абсолютной ошибок искажения контейнера и вероятности ошибки восстановления контейнера от амплитуды встраиваемой последовательности, полученные для следующих условий: амплитуда встраиваемого процесса при обучении $a_m = 1/255$, число эпох обучения $T = 5000$, размер блоков пикселей 8×8.



а) б)



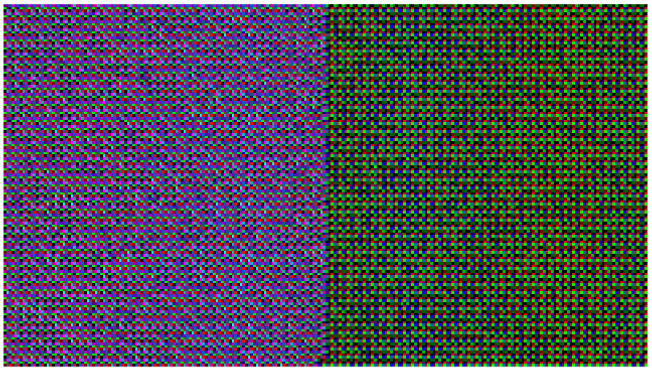
в)

Рисунок 3.11 - Зависимости для среднеквадратичной ошибки искажения контейнера (а), зависимости для вероятности ошибки восстановления ЦВЗ (б), зависимости для абсолютной ошибки искажения контейнера (в)

Пример фрагмента исходного и маркированного изображения, сформированного в результате реализации описанного блочного варианта встраивания ЦВЗ и сохраненного в формате JPEG с высоким качеством сжатия, приведены на рисунок 3.12а и б. При создании ЦВЗ выбирались блоки пикселей размером 8×8, $n=3\times3$, амплитуда встраиваемого сигнала $am=0.03$. При рассмотрении маркированного изображения в увеличенном масштабе на участках плавного изменения яркости становятся заметными артефакты в виде блочности. В результате применения JPEG сжатия бинарное изображение ЦВЗ также восстанавливается с артефактами, однако общий вид встраиваемой ранее пиктограммы остается различимым. Если маркированное изображение сохраняется в графические форматы не предусматривающие сжатия, последующее восстановление ЦВЗ происходит с минимальным количеством искажений.

а)

б)



в)

Рисунок 3.12 - Фрагмент исходного изображения (а); фрагмент изображения, содержащего ЦВЗ, встроенного с использованием блочного алгоритма (размер блоков 8×8, $a_m = 0.03$) и сохраненного в JPEG формате (б); разница между исходным и заполненным контейнером (в).

На рисунке 3.12в показана разность между исходным и заполненным контейнером с коэффициентом умножения 1000 для обеспечения наглядности отображения.

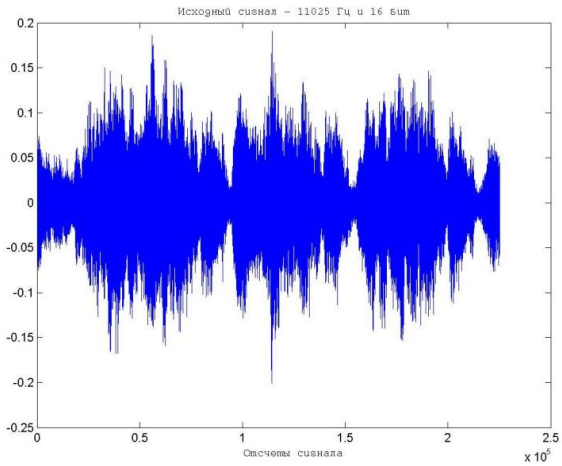
Таким образом, исходный ЦВЗ восстанавливается безошибочно для форматов контейнеров не предусматривающих сжатие, и с определенными искажениями в случае JPEG-сжатия с потерями. Увеличение амплитуды вносимых искажений a_m позволяет повысить робастность встроенных меток, однако это не всегда желательно, поскольку может привести к потере качества защищаемого файла.

3.3. Модели и алгоритмы создания цифровых водяных знаков для аудио-файлов

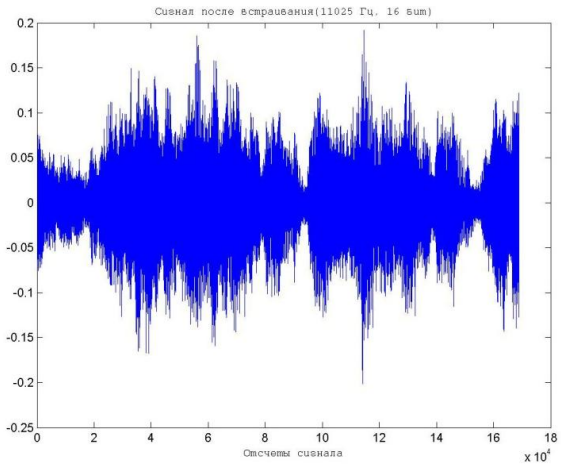
Особый интерес представляет развитие предложенного базового алгоритма создания ЦВЗ при реализации отображений автоассоциативного типа для аудио-файлов. Проведенные исследования показывают, что многие характеристики звука являются для каждого человека субъективными и индивидуальными. Частотная характеристика звука является для большинства людей одинаковой. Исследование восприятия звука показало, что порог слышимости на разных частотах неодинаков, поэтому встраивание ЦВЗ в высокочастотную область аудио - файла будет существенно менее заметно для слушателя. Соответственно, были проведены исследования вопросов создания ЦВЗ для аудио-файлов при вариации различных параметров (частота дискретизации сигнала, битность амплитуды сигнала и количество каналов воспроизведения).

Для реализации процесса создания ЦВЗ для аудио-файлов wav-формата используется алгоритм первого типа, описанный в п.3.1. При моделирования минимальное значение амплитуды задавалась равным $am = 0.5/65535$, исходя из типовой разрядности (16 бит) представления аудио-файлов.

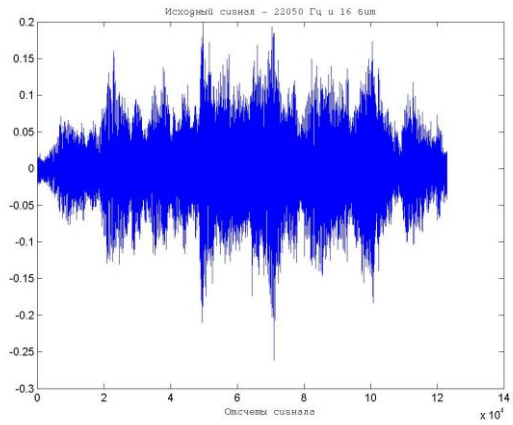
Проведем исследование качества алгоритма в зависимости от частоты дискретизации сигнала. На рисунке 3.13а представлен исходный моно сигнал с параметрами: частота дискретизации сигнала 11025 Гц, разрядность исходного сигнала 16 бит, на рисунке 3.13б представлен результирующий сигнал, содержащий встроенную последовательность ЦВЗ. На рисунке 3.13в представлен исходный моно сигнал с параметрами: частота дискретизации сигнала 22050 Гц, разрядность исходного сигнала 16 бит, на рисунке 3.13г представлен результирующий сигнал, содержащий встроенную последовательность ЦВЗ. На рисунке 3.13д представлен исходный моно сигнал с параметрами: частота дискретизации сигнала 44100 Гц, разрядность исходного сигнала 16 бит, на рисунке 3.13е представлен результирующий сигнал, содержащий встроенную последовательность ЦВЗ



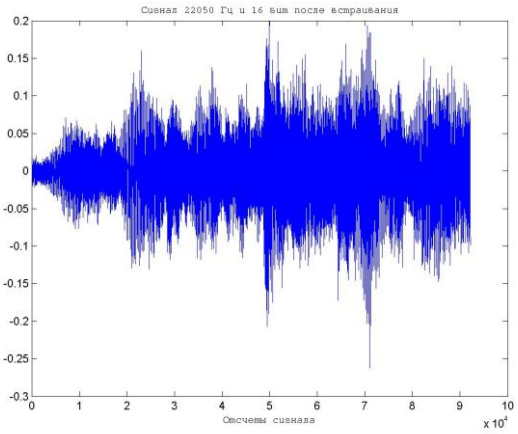
а)



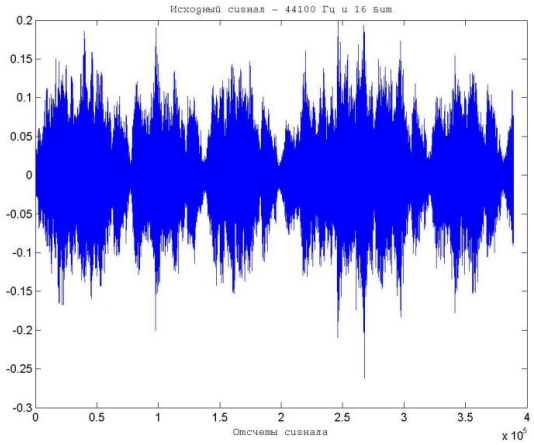
б)



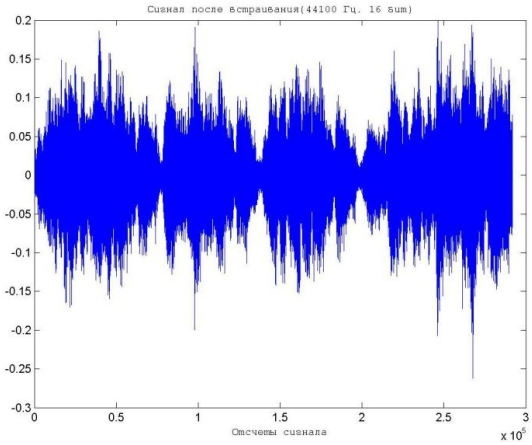
в)



г)



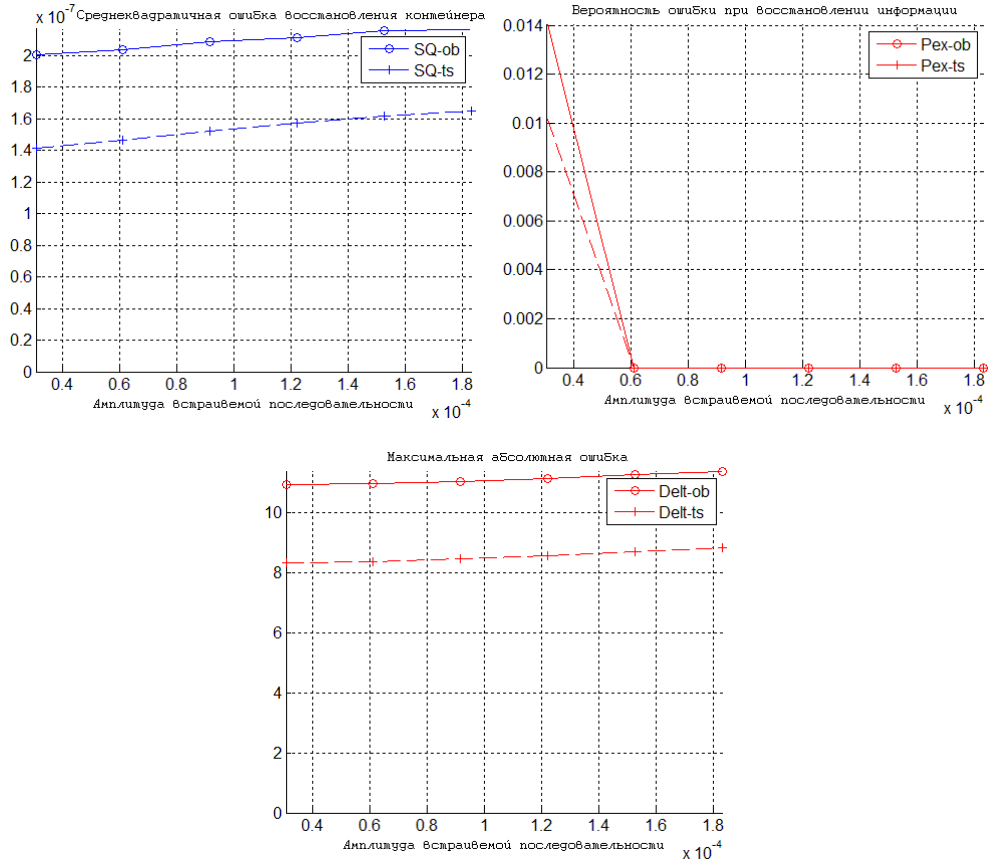
д)



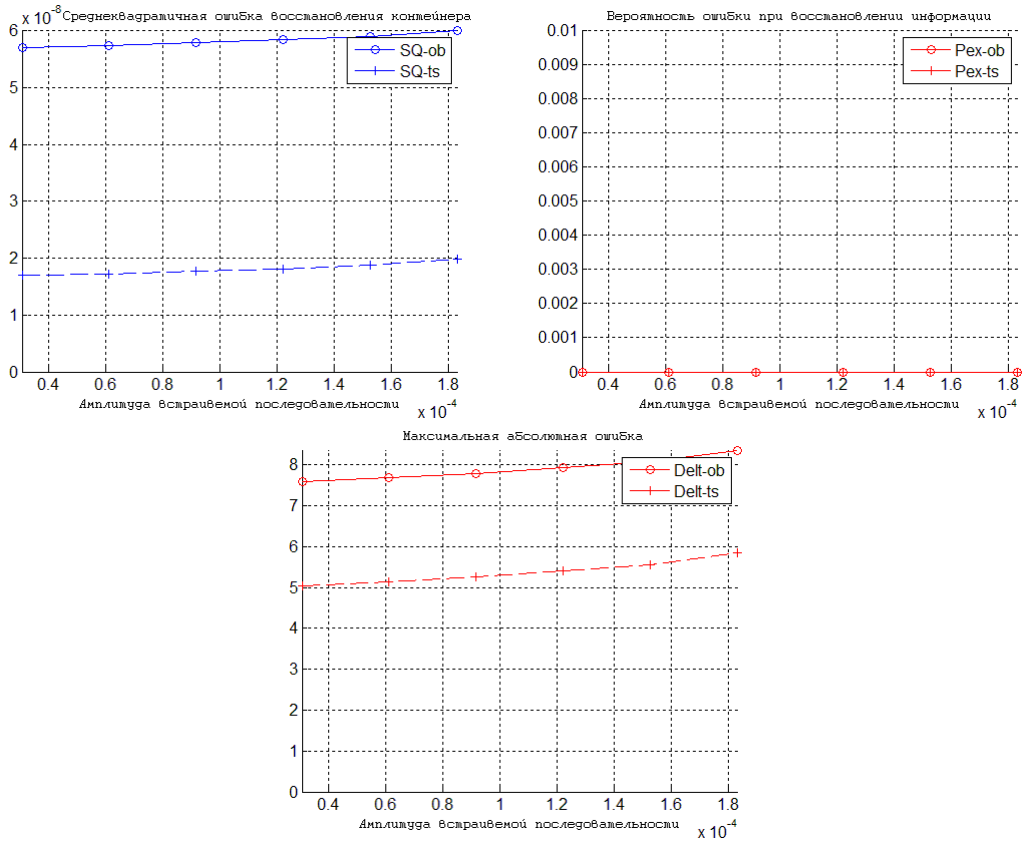
е)

Рисунок 3.13 - Исходный сигнал и результирующий сигнал 11025 Гц (а,б); исходный аудио-сигнал и результирующий сигнал 22050 Гц (в,г) исходный аудио-сигнал и результирующий сигнал 44100 Гц (д,е)

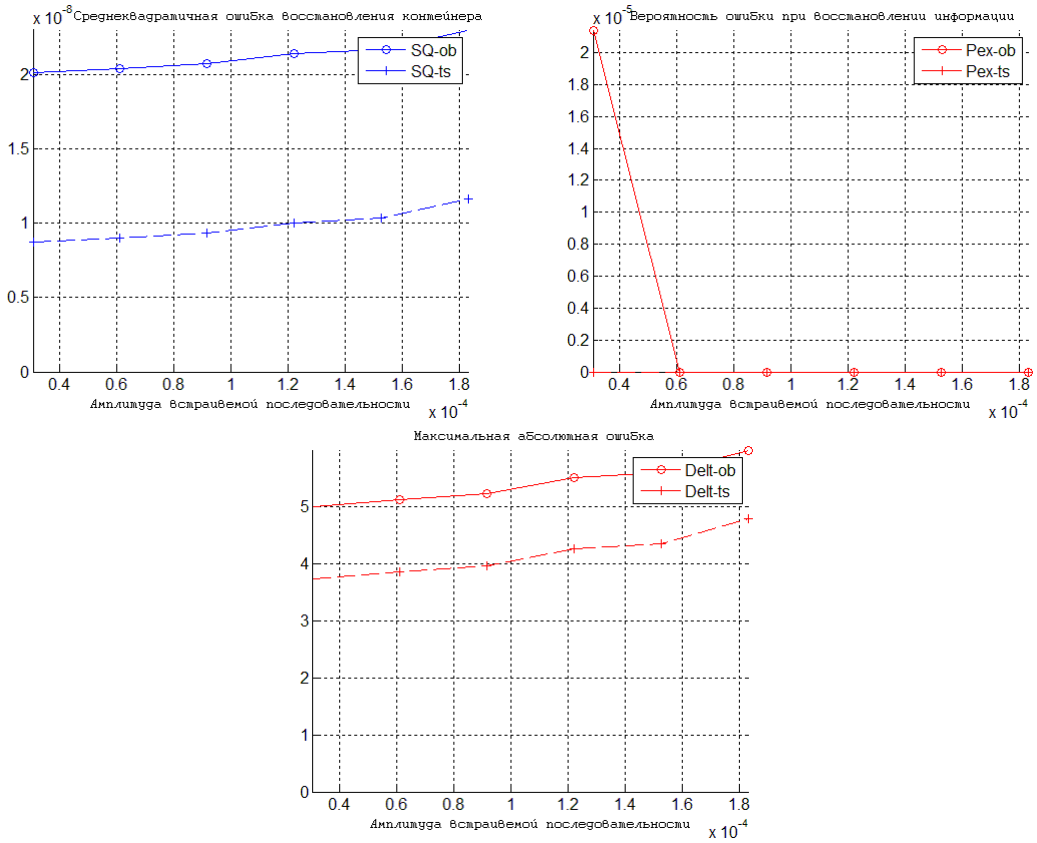
При визуальном сравнении видно, что сигналы по форме практически не отличаются, следовательно, не будут отличаться и по звучанию.



а)



б)



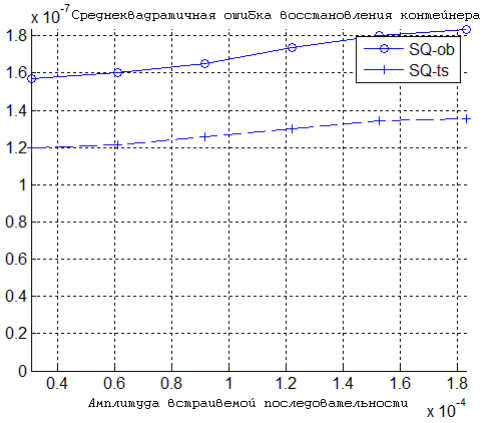
в)

Рисунок 3.14 - Зависимости среднеквадратичной ошибки искажения контейнера, вероятности ошибки восстановления ЦВЗ, абсолютной ошибки искажения контейнера для аудио-файлов с различными частотами дискретизации сигнала 11025 Гц (а); 22050 Гц (б); 44100 Гц (в)

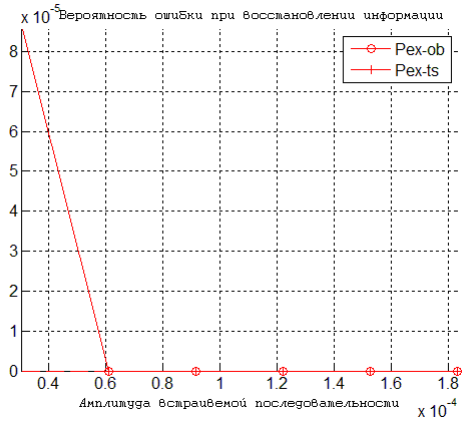
На рисунке 3.14 приведены зависимости средней квадратичной и абсолютной ошибок искажения контейнера и вероятности ошибки восстановления контейнера от амплитуды встраиваемой последовательности, полученные для исходных аудио-файлов с различными частотами дискретизации сигнала.

Следует отметить, что качество звучания информации зависит не только от частоты дискретизации сигнала, но и от битности (разрядности) сигнала. Аудио-файлы имеют различную битность сигнала от которой зависит какое количество бит могут занимать значения амплитуды.

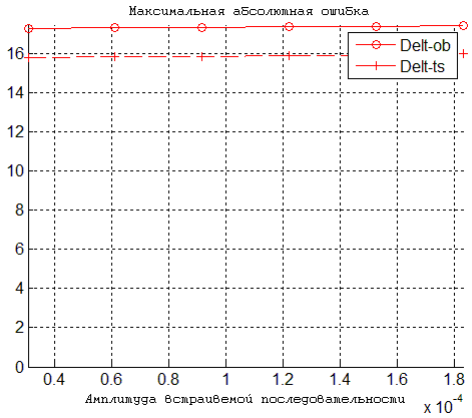
На рисунке 3.15 (а,б,в) приведены зависимости средней квадратичной и абсолютной ошибок искажения контейнера и вероятности ошибки восстановления контейнера от амплитуды встраиваемой последовательности, полученные для исходного моносигнала с параметрами: 22050 Гц, 8 бит, на рисунке 3.15 (г,д,е) приведены зависимости средней квадратичной и абсолютной ошибок искажения контейнера и вероятности ошибки восстановления контейнера от амплитуды встраиваемой последовательности, полученные для исходного моносигнала с параметрами 22050 Гц, 32 бит.



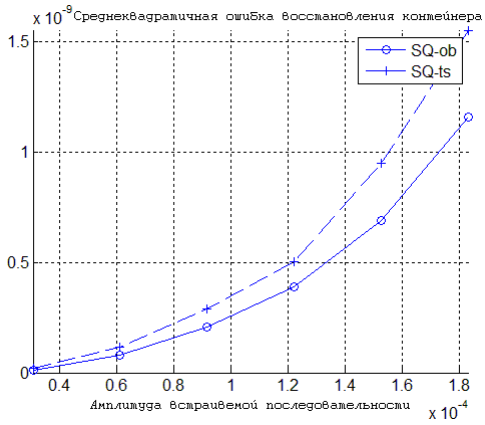
а)



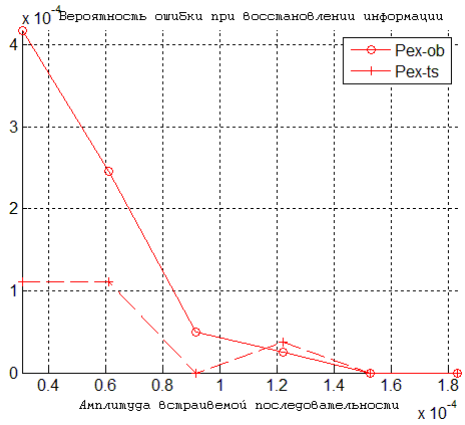
б)



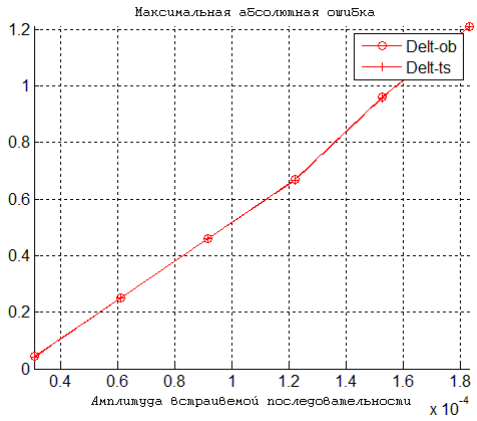
в)



г)



д)

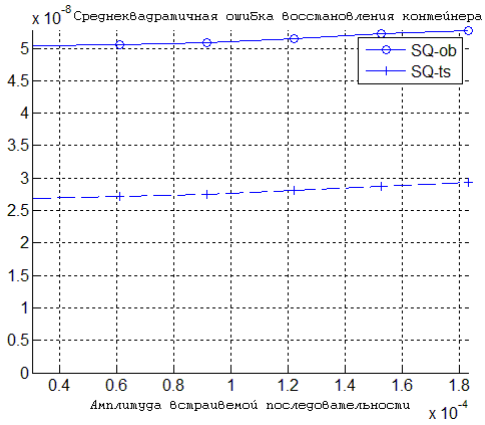


е)

Рисунок 3.15 - Зависимости среднеквадратичной ошибки искажения контейнера, вероятности ошибки восстановления ЦВЗ, абсолютной ошибки искажения контейнера для аудио-файлов с различными значениями битности 8 бит (а,б,в); 32 бит (г,д,е)

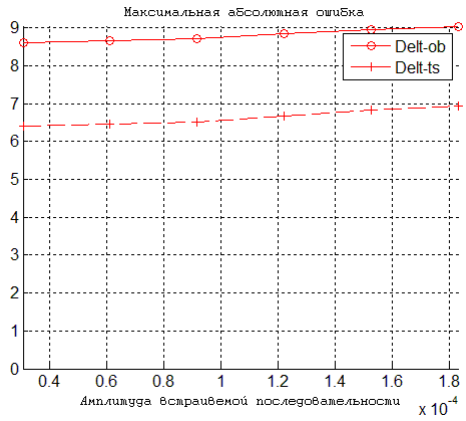
При работе с аудио-контентом часто встречается не только моносигнал, передающий 1 канал звучания, но и более каналов. Среди часто используемых вариаций количества каналов передачи аудио сигнала используется стерео сигнал, который включает в себя 2 канала.

На рисунке 3.16 приведены зависимости средней квадратичной и абсолютной ошибок искажения контейнера и вероятности ошибки восстановления контейнера от амплитуды встраиваемой последовательности для исходного сигнала с параметрами 22050 Гц, 16 бит, стерео сигнал.



а)

б)



в)

Рисунок 3.16 - Зависимости среднеквадратичной ошибки искажения контейнера (а); вероятности ошибки восстановления ЦВЗ (б); абсолютной ошибки искажения контейнера (в).

Как показали предварительные результаты экспериментов, разработанные алгоритмы для различных классов исходных данных (цветных изображений (RGB, YCbBr), аудио-файлы), а так же для различных форматов данных (JPEG) продемонстрировали хорошие результаты по минимизации искажений при создании маркированных контейнеров (алгоритм функционального встраивания и нейросетевой алгоритм модификации усредненного значения блоков), а также при восстановлении ЦВЗ.

Выводы по главе

1. Исходя из полученных в главе 2 настоящей диссертации теоретических обоснований, в ходе дальнейших исследований рассмотрены

реализации описываемых базовых алгоритмов создания цифровых водяных знаков для цветных изображений в форматах RGB, YCbCr, имеющих целочисленное представление данных. При реализации предложенных вариантов моделей и алгоритмов создания ЦВЗ для цветных изображений в формате RGB обеспечивается относительная средняя квадратичная ошибка искажения фрагментов контейнера порядка $10^{-4}...10^{-5}$. При этом для изображений в формате RGB обеспечивается вероятность ошибки восстановления элементов двоичной последовательности ЦВЗ порядка 10^{-1} при относительной амплитуде вносимого $a_m = 0.005$, порядка 10^{-2} при амплитуде $a_m = 0.035$.Для цветных изображений в формате YCbCr обеспечивается вероятность ошибки восстановления элементов двоичной последовательности ЦВЗ порядка 10^{-2} при амплитуде $a_m = 0.005$, порядка 10^{-3} при амплитуде $a_m = 0.035$.

2. Результаты проведенных экспериментальных исследований показывают целесообразность модификации алгоритма создания ЦВЗ с целью адаптации по уровню ошибки искажения контейнера и противодействия преобразованиям JPEG для алгоритма с реализацией сжимающих отображений автоассоциативного типа. При экспериментальных исследованиях можно ожидать достижения уровня относительной средней квадратичной ошибки искажения исходного файла-контейнера порядка $10^{-4}\ldots10^{-6}$, а также вероятности ошибки при восстановления элементов встраиваемого водяного знака не более 10^{-2}.

3. Результаты проведенных экспериментальных исследований для аудио-файлов (WAV формата) показывают возможность скрытия больших объемы информации, за счет того, что выбранные форматы не предусматривают предварительной декомпрессии и последующего после скрытия сжатия (например, MPEG), при этом проведены исследования для сигналов разной битности и частоты дискретизации сигнала. При реализации предложенных вариантов моделей и алгоритмов создания ЦВЗ для аудио-файлов обеспечивается относительная средняя квадратичная ошибка искажения фрагментов контейнера

порядка $10^{-7}...10^{-8}$ при вероятности ошибки восстановления элементов двоичной последовательности ЦВЗ порядка $10^{-1}...10^{-2}$ в зависимости от частоты дискретизации сигнала.

4. Исследование характеристик цифровых водяных знаков для реальных объектов. Описание программного комплекса для создания цифровых водяных знаков

В главе 2 и 3 настоящей диссертации были рассмотрены нейросетевые функциональные модели и универсальные алгоритмы преобразования данных в интересах решения задачи контентно-независимого встраивания и восстановления сообщений (создания ЦВЗ) в контейнерах, как с вещественной формой представления данных, так и с целочисленным форматом данных. При этом в главе 3 получены закономерности, описывающие потенциальные характеристики качества создаваемых ЦВЗ для объектов различных классов (цветные изображения в формате RGB и YCbCr, аудио-файлы в формате WAV).

Представляет интерес реализация и исследование типовых методик оценки качества созданных с помощью нейросетевых алгоритмов цифровых водяных знаков для изучения возможности обнаружения встроенного ЦВЗ или его частичного или полного уничтожения сторонним наблюдателем.

Кроме того, в настоящей главе приводится описание разработанного с участием автора настоящей диссертации программного комплекса для создания цифровых водяных знаков, реализующего ранее рассмотренные модели и алгоритмы.

4.1. Типовые методики оценки качества цифровых водяных знаков

Для оценки качества цифровых водяных знаков на основе нейросетевых функциональных моделей преобразования данных и базовых алгоритмов целесообразно разработать и реализовать методику анализа возможности обнаружения встроенного ЦВЗ и методику анализа возможности восстановления встроенной последовательности ЦВЗ при определенных воздействиях стороннего наблюдателя.

Методика анализа возможности обнаружения встроенного ЦВЗ. Рассмотрим методику анализа возможности обнаружения встроенного ЦВЗ для алгоритма создания ЦВЗ при реализации отображений автоассоциативного типа.

Предполагается, что стороннему наблюдателю известен алгоритм встраивания ЦВЗ, т.е. архитектура нейронной сети, реализующей сжимающее преобразование того или иного вида для последующего встраивания (рисунок 2.6а), кроме того известна конфигурация фрагмента исходного файла-контейнера, используемого для встраивания элемента двоичной последовательности ЦВЗ. Требуется оценить его возможности по выявлению факта наличия или отсутствия ЦВЗ в анализируемом файле-контейнере. Для этого, как следует из структуры алгоритма встраивания ЦВЗ, необходимо выполнить следующую последовательность действий.

1. Для анализируемого контейнера формируется обучающая выборка путем деления анализируемого файла-контейнера на блоки – фрагменты контейнера и формируется соответствующая совокупность фрагментов контейнера $\{g^{(p)}\}, p=\overline{1,P}$ для встраивания, в рамках которой, определятся исходный вектор $g^{(p)}$, описывающий каждый фрагмент. При этом для цветных изображений при формировании вектора проводится соответствующая развертка всех цветовых компонент.

2. Каждый такой вектор, имеющий целочисленное представление с разрядностью представления данных, характерной для данного типа файла-контейнера, преобразуется в вектор $z^{(p)}$, содержащий вещественные данные в диапазоне значений –0.5…+0.5 (с центрированием)

$$z^{(p)} = (z_1^{(p)},\ldots,z_n^{(p)})^T = F_{double}(g_1^{(p)},\ldots,g_n^{(p)}) - 0.5,$$

где $F_{double}(g_1^{(p)},\ldots,g_n^{(p)})$ – функция преобразования целочисленных данных в данные вещественного формата в диапазоне 0,…,1.

3. Проводится обучение нейронной сети по выборке $z^{(p)}, p=\overline{1,P}$, с архитектурой, представленной на рисунке 2.6а и известной стороннему наблюдателю.

4. После завершения процесса обучения входная последовательность данных $z^{(p)}$, $p=\overline{1,P}$ пропускается через сеть и формирует результирующую последовательность $\tilde{z}^{(p)}$, $p=\overline{1,P}$, тогда $\tilde{z}^{(p)}$ будет отличаться от $z^{(p)}$ только

«высокочастотной» составляющей с малой амплитудой и дисперсией соответствующей минимальному собственному числу выборочной матрицы ковариации, которая определяется следующим соотношением:

$$\Delta_z^{(p)} = z^{(p)} - \tilde{z}^{(p)}.$$

5. По полученной реализации $\Delta_z^{(p)}$, $p = \overline{1,P}$ определяется нормированный собственный вектор, соответствующий минимальному собственному числу

$$\varphi_n = r_{\min} \Big/ \sqrt{\left(r_{\min}^T r_{\min}\right)},\ r_{\min} = \frac{1}{P} \sum_{p=1}^{P} \Delta_z^{(p)},\ p = \overline{1,P},$$

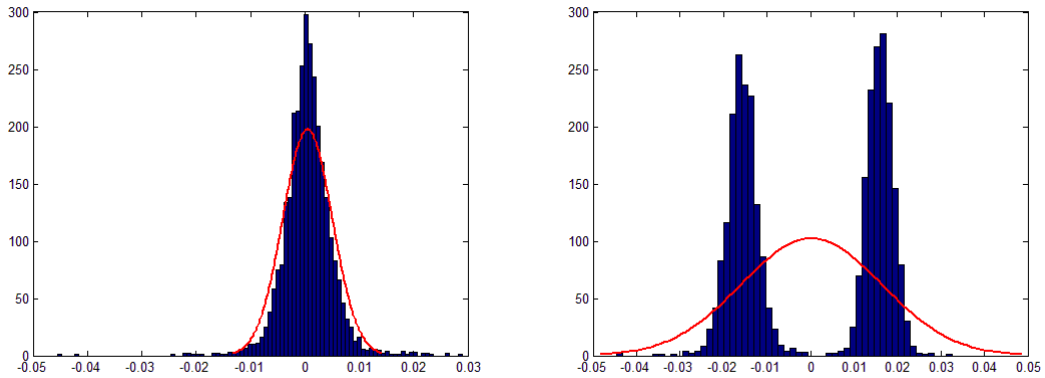
на основе которого вычисляются значения амплитуды «высокочастотной» составляющей

$$\alpha_i^{(p)} = \varphi_n^T \Delta_z^{(p)},\ p = \overline{1,P}.$$

6. Для совокупности полученных значений амплитуды «высокочастотной» составляющей строится гистограмма в заданном диапазоне значений (для графических форматов представления данных обычно можно выбрать -0.02…0.02 (рисунок 4.1а и б)).

7. Проводится визуальный и статистический анализ полученных данных с использованием стандартных критериев согласия.

Рассмотрим более детально обоснования к последнему пункту методики.



а) б)

Рисунок 4.1 – Гистограмма амплитуды «высокочастотной» составляющей для файла-контейнера, не содержащего встроенную последовательность ЦВЗ (а); для файла-контейнера, содержащего встроенную последовательность ЦВЗ (б)

На рисунке 4.1а красным цветом обозначен график функции плотности вероятности гауссовского закона распределения, полученные для следующих условий: амплитуда встраиваемого процесса при обучении $a_m = 1/255$, размер фрагмента файла-контейнера, используемого для встраивания последовательности ЦВЗ 9×9.

Визуальный анализ гистограмм амплитуды «высокочастотной» составляющей показывает, что гистограмма амплитуды «высокочастотной» составляющей, представленной на рисунке 4.1а близка с графиком функции плотности вероятности гауссовского закона распределения, что позволяет сделать вывод о том, что в анализируемом файле-контейнере не содержится встроенной последовательности ЦВЗ. В тоже время, гистограмма амплитуды «высокочастотной» составляющей, представленной на рисунке 4.1б визуально не схожа с графиком функции плотности вероятности гауссовского закона распределения, имеет бимодальный характер и отражает вносимую модификацию входного вектора, когда вместо случайной величины, распределенной по гауссовскому закону, используются два случайных точечных значения, что позволяет сделать вывод о том, что анализируемый файл-контейнер содержит встроенную последовательность ЦВЗ.

Статистический анализ полученных гистограмм основывается на применении критерия χ^2 («хи-квадрат», критерий согласия Пирсона).

Критерий χ^2 контролирует согласованность гипотетических вероятностей $p_k = P\{E_k\}$ случайных событий $E_1, E_2 \ldots E_r$ с их относительными частотами $h_k = h\{E_k\} = n_k / n$ в выборке из n независимых наблюдений [20]. В данном случае каждое событие E_k состоит в том, что некоторая случайная величина x попадает в определенный классовый интервал, так что критерий χ^2 позволяет сравнивать гипотетическое теоретическое распределение величины x с ее эмпирическим распределением. Используя критерия согласия Пирсона, требуется проверить нулевую гипотезу H_0 о подчинении наблюдаемого вектора $\alpha_i^{(p)}$ нормальному закону распределения (то есть об отсутствии встроенной последовательности ЦВЗ в анализируемом файле-контейнере).

Для проверки вводится следующая статистика [20]

$$\chi_n^2 = n\sum_{i=1}^{r}\frac{(h_i - p_i)^2}{p_i} = \sum_{i=1}^{r}\frac{(n_i - np_i)^2}{np_i}, \tag{4.1}$$

распределение которой при $n \to \infty$ стремится к распределению χ^2 с $m = r - s - 1$ степенями свободы, где r – число групп (частичных интервалов) выборки, s – число параметров предполагаемого распределения, p_i - предполагаемая вероятность попадания в i интервал, $h_i = n_i / n$ –соответствующее эмпирическое значение, n_i - число элементов выборки из i интервала, n - полный объем выборки. В частности, если предполагаемое распределение – нормальное, то оценивают два параметра (математическое ожидание и среднеквадратическое отклонение), поэтому число степеней свободы $m = r - 3$.

Для расчета введенной статистики, в качестве предполагаемой вероятности попадания в i интервал p_i используется плотность вероятности нормального закона распределения, для вычисления соответствующего эмпирического значения $h_i = n_i / n$ используется полученная гистограмма распределения амплитуды «высокочастотной» составляющей.

Для проверки нулевой гипотезы H_0 о соответствии гистограммы амплитуды «высокочастотной» составляющей гауссовскому закону распределения необходимо сравнить значение статистики χ_n^2 в соответствии с выражением 4.1, и значение статистики с критическим значением $\chi_{кр}^2(m,\alpha)$ для уровня значимости α и числа степеней свободы $m = r - 3$, в соответствии со значением квантили $\chi_{кр}^2(m,\alpha)$ в зависимости от уровня вероятности и числа степеней свободы. Область принятия предположительной гипотезы H_0 определяется условием $\chi_n^2 \le \chi_{кр}^2(m,\alpha)$. При несоблюдении указанного неравенства принимается альтернативная гипотеза H_1 о принадлежности выборки некоторому неизвестному распределению (отличного от распределения Гаусса).

Используя выражение (4.1) для гистограммы амплитуды «высокочастотной» составляющей, представленной на рисунке 4.1а (размер фрагмента 9×9, используемого для встраивания одного элемента последовательности ЦВЗ) при уровне значимости $\alpha = 0.05$ показано, что нулевая гипотеза H_0 не отвергается, и, следовательно, в файле не содержится встроенная последовательность ЦВЗ. Также, используя выражение (4.1), для гистограммы амплитуды «высокочастотной» составляющей, представленной на рисунке 4.1б, при аналогичных условиях при уровне значимости $\alpha = 0.05$ показано, что нулевая гипотеза H_0 отвергается и, следовательно, в файле содержится встроенная последовательность ЦВЗ.

Рассмотрим методику возможности обнаружения встроенного ЦВЗ для алгоритма создания ЦВЗ при реализации отображений гетероассоциативного типа. Предположим, что стороннему наблюдателю известна архитектура нейронной сети, определяющей оператор связи соседних фрагментов (рисунок 2.16а), нейронной сети, реализующей сжимающее преобразование того или иного вида для последующего встраивания (рисунок 2.16б), кроме того известна конфигурация фрагмента исходного файла-контейнера, используемого для встраивания элемента двоичной последовательности ЦВЗ. Требуется

определить факт наличия или отсутствия ЦВЗ в анализируемом файле-контейнере. Для этого необходимо выполнить следующую последовательность действий.

1. Для анализируемого контейнера формируется обучающая выборка путем деления анализируемого файла-контейнера на блоки – фрагменты контейнера и формируется соответствующая совокупность фрагментов контейнера $\{g^{(p)}\}, p=\overline{1,P}$ для встраивания, в рамках которой, определятся исходный вектор $g^{(p)}$, описывающий каждый фрагмент. При этом для цветных изображений при формировании вектора проводится соответствующая развертка всех цветовых компонент.

2. Каждый такой вектор, имеющий целочисленное представление с разрядностью представления данных, характерной для данного типа файла-контейнера, преобразуется в вектор $z^{(p)}$, содержащий вещественные данные в диапазоне значений –0.5…+0.5 (с центрированием)

$$z^{(p)} = (z_1^{(p)},\ldots,z_n^{(p)})^T = F_{double}(g_1^{(p)},\ldots,g_n^{(p)}) - 0.5,$$

где $F_{double}(g_1^{(p)},\ldots,g_n^{(p)})$ – функция преобразования целочисленных данных в данные вещественного формата в диапазоне 0,…,1.

3. Вектор $z=(z_1,\ldots z_n)^T$ представляющий фрагмент файла контейнера, совокупность реализаций которого $z^{(p)}$, $p=\overline{1,P}$ используется для встраивания элементов двоичной последовательности $\tilde{d}^{(p)}$ разбивается на две части: входную $z_1^{(p)} = (z_{1,1}^{(p)},\ldots,z_{1,h}^{(p)}), p=\overline{1,P}$ и выходную $z_2^{(p)} = (z_{2,1}^{(p)},\ldots,z_{2,s}^{(p)}), p=\overline{1,P}$.

4. Проводится обучение нейронной сети по совокупности реализаций входного и целевого вектора $\{z_1^{(p)}, z_2^{(p)}\}$, $p=\overline{1,P}$, с архитектурой, представленной на рисунке 2.16а и известной стороннему наблюдателю.

5. После завершения процесса обучения входная последовательность $z_1^{(p)}$, $p=\overline{1,P}$ пропускается через сеть и формирует результирующую последовательность $\hat{z}_2 = (\hat{z}_{2,1},\ldots\hat{z}_{2,s})^T$, используемую для прогнозирования

фрагментов контейнера на основе соседних фрагментов контейнера и для определения стохастической составляющей прогноза

$$V^{(p)} = z_2^{(p)} - \hat{z}_2^{(p)}.$$

6. Проводится обучение нейронной сети по выборке $z_1^{(p)}$, $p = \overline{1,P}$, с архитектурой, представленной на рисунке 2.16б и известной стороннему наблюдателю.

7. После завершения процесса обучения входная последовательность данных $z_1^{(p)}$, $p = \overline{1,P}$ пропускается через сеть и формирует результирующую последовательность $\tilde{z}_2^{(p)}$, $p = \overline{1,P}$, тогда $\tilde{z}_2^{(p)}$ будет отличаться от $\hat{z}_2^{(p)}$ только «высокочастотной» составляющей с малой амплитудой и дисперсией соответствующей минимальному собственному числу выборочной матрицы ковариации, которая определяется следующим соотношением:

$$\Delta_{z_2}^{(p)} = \hat{z}_2^{(p)} - \tilde{z}_2^{(p)}.$$

8. По полученной реализации $\Delta_{z_2}^{(p)}$, $p = \overline{1,P}$ определяется нормированный собственный вектор, соответствующий минимальному собственному числу

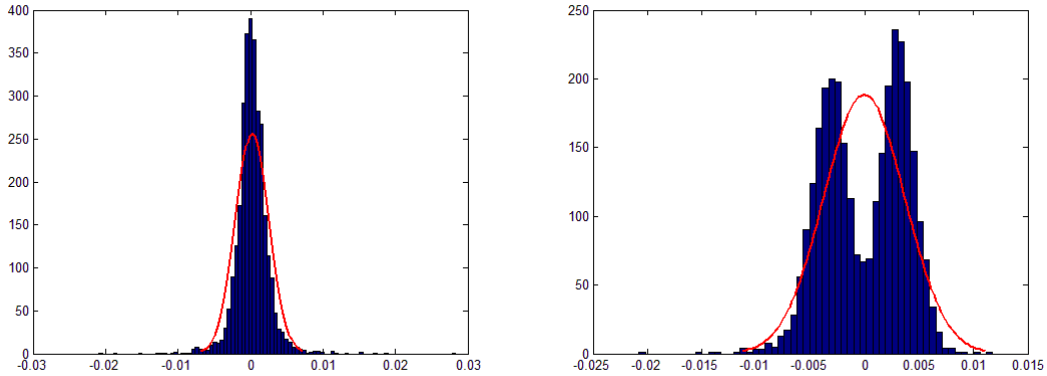
$$\varphi_n = r_{\min} \Big/ \sqrt{\left(r_{\min}^T r_{\min}\right)},\ r_{\min} = \frac{1}{P} \sum_{p=1}^{P} \Delta_{z_2}^{(p)},$$

на основе которого вычисляются значения амплитуды «высокочастотной» составляющей

$$\alpha_i^{(p)} = \varphi_n^T \Delta_{z_2}^{(p)}.$$

9. Для совокупности полученных значений амплитуды «высокочастотной» строится гистограмма в заданном диапазоне значений (для графических форматов представления данных обычно можно выбрать -0.02…0.02 (рисунок 4.2а и б)).

10. Проводится визуальный и статистический анализ полученных данных с использованием стандартных критериев согласия.



а) б)

Рисунок 4.2 – Гистограмма амплитуды «высокочастотной» составляющей, для файла-контейнера, не содержащего встроенную последовательность ЦВЗ (а); для файла-контейнера, содержащего встроенную последовательность ЦВЗ (б)

На рисунке 4.2а красным цветом обозначен график функции плотности вероятности гауссовского закона распределения, полученные для следующих условий: относительная амплитуда встраиваемого процесса при обучении $a_m = 1/255$, исходный размер фрагмента контейнера 7×7, размер фрагмента, используемого для встраивания одного элемента последовательности ЦВЗ 5×5.

Визуальный анализ гистограмм амплитуды «высокочастотной» составляющей показывает, что гистограмма амплитуды «высокочастотной» составляющей, представленной на рисунке 4.2а близка с графиком функции плотности вероятности гауссовского закона распределения, что позволяет сделать вывод о том, что в анализируемом файле-контейнере не содержится встроенной последовательности ЦВЗ. В тоже время, гистограмма амплитуды «высокочастотной» составляющей, представленной на рисунке 4.2б визуально не схожа с графиком функции плотности вероятности гауссовского закона распределения, имеет бимодальный характер и отражает вносимую модификацию входного вектора, когда вместо случайной величины, распределенной по гауссовскому закону, используются два случайных точечных значения, что позволяет сделать вывод о том, что анализируемый файл-контейнер содержит встроенную последовательность ЦВЗ.

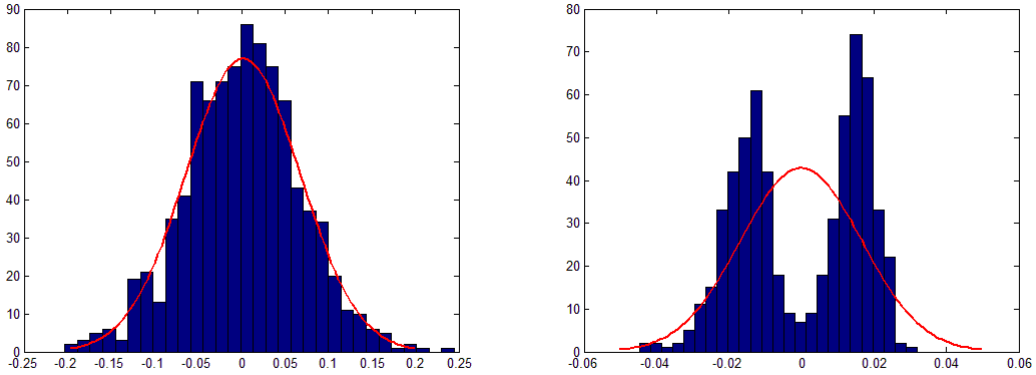
Статистический анализ полученных гистограмм амплитуды «высокочастотной» составляющей также может основываться на применении критерия χ^2 («хи-квадрат», критерий согласия Пирсона).

Используя выражение (4.1) для гистограммы амплитуды «высокочастотной» составляющей, представленной на рисунке 4.2а (исходный размер фрагмента контейнера $z^{(p)}$ 7×7, размер фрагмента, используемого для встраивания одного элемента последовательности ЦВЗ $z_2^{(p)}$ 5×5) при уровне значимости $\alpha=0.05$ показано, что нулевая гипотеза H_0 не отвергается и, следовательно, в файле не содержится встроенная последовательность ЦВЗ. Для гистограммы амплитуды «высокочастотной» составляющей, представленной на рисунке 4.2б при аналогичных условиях при уровне значимости $\alpha=0.05$ получено, что нулевая гипотеза H_0 отвергается и, следовательно, в файле содержится встроенная последовательность ЦВЗ.

Анализ полученных результатов на основе разработанной методики анализа возможности обнаружения встроенного ЦВЗ при реализации автоассоциативных и гетероассоциативных показывает, что при известной архитектуре нейронной сети, реализующей сжимающее преобразование того или иного вида для последующего встраивания, известной конфигурация фрагмента исходного файла-контейнера, используемого для встраивания элемента двоичной последовательности ЦВЗ возможно определить факт наличия или отсутствия ЦВЗ в анализируемом файле-контейнере.

Кроме того, следует отметить, что при реализации преобразований автоассоциативного типа не всегда представляется возможным достоверно провести визуальный и статистический анализ. На рисунке 4.3. представлены гистограммы амплитуды «высокочастотной» составляющей полученные для следующий условий исходный размер фрагмента контейнера 9×9, размер фрагмента, используемого для встраивания одного элемента последовательности ЦВЗ 9×9, относительная амплитуда встраиваемого процесса при обучении $a_m=1/255$ (рисунок 4.3а), исходный размер фрагмента контейнера 9×9, размер

фрагмента, используемого для встраивания одного элемента последовательности ЦВЗ 9×9, относительная амплитуда встраиваемого процесса при обучении $a_m = 5/255$ (рисунок 4.3б).



а) б)

Рисунок 4.3 – Гистограмма амплитуды «высокочастотной» составляющей для файла-контейнера, содержащего встроенную последовательность ЦВЗ относительная амплитуда встраиваемого процесса при обучении $a_m = 1/255$ (а); относительная амплитуда встраиваемого процесса при обучении $a_m = 5/255$ (б)

Методика анализа возможности восстановления встроенной последовательности ЦВЗ при негативных воздействиях стороннего наблюдателя. Рассмотрим методику анализа возможности восстановления встроенной последовательности ЦВЗ после воздействий стороннего наблюдателя, направленных на частичное или полное уничтожение встроенной последовательности ЦВЗ при реализации отображений автоассоциативного типа. Предположим, что стороннему наблюдателю известна архитектура нейронной сети, реализующей сжимающее преобразование того или иного вида для последующего встраивания (рисунок 2.6а), кроме того не известна или известна частично конфигурация фрагмента исходного файла-контейнера, в который происходит встраивание последовательности ЦВЗ.

Следует отметить, что в случае если не известна конфигурация фрагмента файла-контейнера, то количество всех возможных вариантов конфигурации

фрагмента исходного файла-контейнера, в общем случае определяется следующим соотношением:

$$count = \sum_{s=1}^{n} C_n^s = \sum_{s=1}^{n} \frac{A_n^s}{P_s} = \sum_{s=1}^{n} \frac{n!}{(n-s)!\,s!},$$

где s - размер области модификации фрагмента контейнера, в которой происходит встраивание элементов последовательности ЦВЗ, n - размер области модификации фрагмента контейнера, C_n^s - количество сочетаний без повторений из n по s элементов, A_n^s - количество размещений из n по s элементов, P_s - количество перестановок из s элементов.

Требуется определить возможность восстановления ранее встроенной последовательности ЦВЗ в зависимости от степени перекрытия области конфигурации фрагмента исходного контейнера, в который производилось встраивание последовательности ЦВЗ, и предполагаемой сторонним наблюдателем области конфигурации фрагмента анализируемого файла-контейнера. Для этого необходимо выполнить последовательность действий по отношению к анализируемому контейнеру, содержащему ранее встроенный ЦВЗ, которые фактически повторяют действия пользователя при встраивании, с тем только отличием, что амплитуда выделяемой высокочастотной составляющей здесь зануляется.

1. Для анализируемого контейнера формируется обучающая выборка путем деления анализируемого файла-контейнера на блоки – фрагменты контейнера и формируется соответствующая совокупность фрагментов контейнера $\{g^{(p)}\}, p = \overline{1,P}$ для встраивания, в рамках которой определятся исходный вектор $g^{(p)}$, описывающий каждый фрагмент. При этом для цветных изображений при формировании вектора проводится соответствующая развертка всех цветовых компонент.

2. Каждый такой вектор, имеющий целочисленное представление с разрядностью представления данных, характерной для данного типа файла-

контейнера, преобразуется в вектор $z^{(p)}$, содержащий вещественные данные в диапазоне значений –0.5…+0.5 (с центрированием)

$$z^{(p)} = (z_1^{(p)}, \ldots, z_n^{(p)})^T = F_{double}(g_1^{(p)}, \ldots, g_n^{(p)}) - 0.5,$$

где $F_{double}(g_1^{(p)}, \ldots, g_n^{(p)})$ – функция преобразования целочисленных данных в данные вещественного формата в диапазоне 0,…,1.

3. Проводится обучение нейронной сети по выборке $z^{(p)}, p = \overline{1,P}$, с архитектурой, представленной на рисунке 2.6а и известной стороннему наблюдателю.

4. После завершения процесса обучения входная последовательность данных $z^{(p)}$, $p = \overline{1,P}$ пропускается через сеть и формирует результирующую последовательность $\tilde{z}^{(p)}$, $p = \overline{1,P}$, тогда $\tilde{z}^{(p)}$ будет отличаться от $z^{(p)}$ только «высокочастотной» составляющей с малой амплитудой и дисперсией соответствующей минимальному собственному числу выборочной матрицы ковариации, которая определяется следующим соотношением:

$$\Delta_z^{(p)} = z^{(p)} - \tilde{z}^{(p)}.$$

5. На основе полученной реализации $\Delta_z^{(p)}$, $p = \overline{1,P}$ определяется нормированный собственный вектор, соответствующий минимальному собственному числу

$$\varphi_n = r_{\min} \Big/ \sqrt{\left(r_{\min}{}^T r_{\min}\right)},\ r_{\min} = \frac{1}{P} \sum_{p=1}^{P} \Delta_z^{(p)}.$$

6. Каждый фрагмент файла-контейнера, описываемый теперь (после сжатия) вектором $\tilde{z}^{(p)}$, $p = \overline{1,P}$, модифицируется на основе следующего соотношения с учетом амплитуды вносимого искажения $a_m = 0$, тем самым уничтожается встроенная в анализируемый файл-контейнер последовательность ЦВЗ:

$$\bar{z}^{(p)} = \tilde{z}^{(p)} + a_m \varphi_n,\ \ \tilde{z}^{(p)} = W_2 W_1 z^{(p)},\ \ p = \overline{1,P}.$$

7. Полученный вектор $\bar{z}^{(p)}$ подается на вход заранее обученной сети для восстановления встроенной последовательности ЦВЗ и формируется результирующая последовательность $\tilde{d}^{(p)}$, $p=\overline{1,P}$.

8. Для анализируемого файла-контейнера определяется вероятность ошибки восстановления встроенной последовательности ЦВЗ, с учетом полученной результирующей последовательность $\tilde{d}^{(p)}$, $p=\overline{1,P}$.

В таблице 4.1 представлены результаты оценки вероятности ошибки восстановления ЦВЗ без вмешательства стороннего наблюдателя ($P_{ош1}$), результаты оценки вероятности ошибки восстановления ЦВЗ ($P_{ош2}$) при вмешательстве стороннего наблюдателя в зависимости от степени перекрытия предполагаемой и реальной конфигурации фрагмента контейнера. Оценки получены для следующих условий: относительная амплитуда вносимого искажения $a_m = 1/255$, объем обучающей выборки $P = 4160$, размер исходного фрагмента файла-контейнера 9×9, размер фрагмента файла-контейнера, предполагаемого сторонним наблюдателем, равен 9×9 $(100\%), 7\times7\,(60\%), 5\times5\,(30\%), 3\times3\,(10\%)$.

Таблица 4.1 – Вероятности ошибки восстановления ЦВЗ

Степень перекрытия предполагаемой и реальной конфигурации фрагмента контейнера, %	100%	60%	30%	10%
Вероятность ошибки восстановления ЦВЗ, $P_{ош1}$	0.07	0.07	0.07	0.07
Вероятность ошибки восстановления ЦВЗ, $P_{ош2}$	0.49	0.28	0.12	0.03

В таблице 4.2 представлены результаты, полученные при аналогичных условиях при относительной амплитуде вносимого искажения $a_m = 5/255$.

Таблица 4.2 – Вероятности ошибки восстановления ЦВЗ

Степень перекрытия предполагаемой и реальной конфигурации фрагмента контейнера, %	100%	60%	30%	10%
Вероятность ошибки восстановления ЦВЗ, $P_{ош1}$	0.02	0.02	0.02	0.02
Вероятность ошибки восстановления ЦВЗ, $P_{ош2}$	0.47	0.18	0.05	0.01

В таблице 4.3 представлены результаты, полученные при аналогичных условиях при относительной амплитуде вносимого искажения $a_m = 10/255$.

Таблица 4.3 – Вероятности ошибки восстановления ЦВЗ

Степень перекрытия предполагаемой и реальной конфигурации фрагмента контейнера, %	100%	60%	30%	10%
Вероятность ошибки восстановления ЦВЗ, $P_{ош1}$	<0.01	<0.01	<0.01	<0.01
Вероятность ошибки восстановления ЦВЗ, $P_{ош2}$	0.36	0.11	0.03	0.01

Полученные результаты свидетельствуют о том, что для преобразований автоассоциативного типа негативные воздействия стороннего наблюдателя частично или полностью уничтожают встроенную последовательность ЦВЗ в зависимости от степени перекрытия предполагаемой и реальной конфигурации фрагмента контейнера (при известных стороннему наблюдателю архитектуры нейронной сети, реализующей сжимающее преобразование того или иного вида для последующего встраивания (рисунок 2.6а)).

Рассмотрим методику возможности обнаружения встроенного ЦВЗ для алгоритма создания ЦВЗ при реализации отображений гетероассоциативного типа. Предположим, что стороннему наблюдателю известна архитектура

нейронной сети, определяющей оператор связи соседних фрагментов (рисунок 2.16а), архитектура нейронной сети, реализующей сжимающее преобразование того или иного вида для последующего встраивания (рисунок 2.16б), кроме того не известна или известна частично конфигурация фрагмента исходного файла-контейнера, в который происходит встраивание последовательности ЦВЗ. Требуется определить факт наличия или отсутствия ЦВЗ в анализируемом файле-контейнере. Для этого необходимо выполнить следующую последовательность действий.

1. Для анализируемого контейнера формируется обучающая выборка путем деления анализируемого файла-контейнера на блоки – фрагменты контейнера и формируется соответствующая совокупность фрагментов контейнера $\{g^{(p)}\}, p=\overline{1,P}$ для встраивания, в рамках которой определятся исходный вектор $g^{(p)}$, описывающий каждый фрагмент. При этом для цветных изображений при формировании вектора проводится соответствующая развертка всех цветовых компонент.

2. Каждый такой вектор, имеющий целочисленное представление с разрядностью представления данных, характерной для данного типа файла-контейнера, преобразуется в вектор $z^{(p)}$, содержащий вещественные данные в диапазоне значений –0.5…+0.5 (с центрированием)

$$z^{(p)} = (z_1^{(p)},\ldots,z_n^{(p)})^T = F_{double}(g_1^{(p)},\ldots,g_n^{(p)}) - 0.5,$$

где $F_{double}(g_1^{(p)},\ldots,g_n^{(p)})$ – функция преобразования целочисленных данных в данные вещественного формата в диапазоне 0,…,1.

3. Вектор $z=(z_1,\ldots z_n)^T$, представляющий фрагмент файла контейнера, совокупность реализаций которого $z^{(p)}$, $p=\overline{1,P}$ используется для встраивания элементов двоичной последовательности $\tilde{d}^{(p)}$ разбивается на две части: входную $z_1^{(p)} = (z_{1,1}^{(p)},\ldots,z_{1,h}^{(p)}), p=\overline{1,P}$ и выходную $z_2^{(p)} = (z_{2,1}^{(p)},\ldots,z_{2,s}^{(p)}), p=\overline{1,P}$.

4. Проводится обучение нейронной сети по совокупности реализаций входного и целевого вектора $\{z_1^{(p)}, z_2^{(p)}\}$, $p=\overline{1,P}$, с архитектурой, представленной на рисунке 2.16, а, и известной стороннему наблюдателю.

5. После завершения процесса обучения входная последовательность $z_1^{(p)}$, $p=\overline{1,P}$ пропускается через сеть и формирует результирующую последовательность $\widehat{z}_2=(\widehat{z}_{2,1},\ldots\widehat{z}_{2,s})^T$, используемую для прогнозирования фрагментов контейнера на основе соседних фрагментов контейнера и для определения стохастической составляющей прогноза

$$V^{(p)}=z_2^{(p)}-\widehat{z}_2^{(p)}.$$

6. Проводится обучение нейронной сети по выборке $z_1^{(p)}$, $p=\overline{1,P}$, с архитектурой, представленной на рисунке 2.16, б, и известной стороннему наблюдателю.

7. После завершения процесса обучения входная последовательность данных $z_1^{(p)}$, $p=\overline{1,P}$ пропускается через сеть и формирует результирующую последовательность $\widetilde{z}_2^{(p)}$, $p=\overline{1,P}$, тогда $\widetilde{z}_2^{(p)}$ будет отличаться от $\widehat{z}_2^{(p)}$ только «высокочастотной» составляющей с малой амплитудой и дисперсией соответствующей минимальному собственному числу выборочной матрицы ковариации, которая определяется следующим соотношением:

$$\Delta_{z_2}^{(p)}=\widehat{z}_2^{(p)}-\widetilde{z}_2^{(p)}.$$

8. На основе полученной реализации $\Delta_{z_2}^{(p)}$, $p=\overline{1,P}$ определяется нормированный собственный вектор, соответствующий минимальному собственному числу

$$\varphi_n=r_{\min}\Big/\sqrt{\left(r_{\min}^{\ T}r_{\min}\right)},\ r_{\min}=\frac{1}{P}\sum_{p=1}^{P}\Delta_{z_2}^{(p)}.$$

9. Каждый фрагмент файла-контейнера, описываемый теперь (после сжатия) вектором $\widetilde{z}_2^{(p)}$, $p=\overline{1,P}$, модифицируется на основе следующего

соотношения с учетом амплитуды вносимого искажения $a_m = 0$, тем самым удаляется встроенная в анализируемый файл-контейнер последовательность ЦВЗ:

$$\bar{z}_2^{(p)} = \tilde{z}_2^{(p)} + a_m \varphi_n, \quad \tilde{z}_2^{(p)} = W_2 W_1 z_1^{(p)}, \quad p = \overline{1, P},$$

$$\bar{z}_2^{(p)} = \bar{z}_2^{(p)} + V^{(p)}$$

10. Полученный вектор $\bar{z}_2^{(p)}$ подается на вход заранее обученной сети для восстановления встроенной последовательности ЦВЗ и формируется результирующая последовательность $\tilde{d}^{(p)}$, $p = \overline{1, P}$.

11. Для анализируемого файла-контейнера определяется вероятность ошибки восстановления встроенной последовательности ЦВЗ, с учетом полученной результирующей последовательность $\tilde{d}^{(p)}$, $p = \overline{1, P}$.

В таблице 4.4 представлены результаты вероятности ошибки восстановления ЦВЗ без вмешательства стороннего наблюдателя ($P_{ош1}$), результаты вероятности ошибки восстановления ЦВЗ ($P_{ош2}$) в зависимости от степени перекрытия предполагаемой и реальной конфигурации фрагмента контейнера исходного файла контейнера, полученные для следующих начальных условий: относительная амплитуда вносимого искажения $a_m = 1/255$, объем обучающей выборки $P = 4160$, размер исходного фрагмента файла-контейнера 11×11, размер фрагмента, используемого для встраивания одного элемента последовательности ЦВЗ 9×9, размер фрагмента файла-контейнера, предполагаемого сторонним наблюдателем, равен 9×9 $(100\%), 7 \times 7\,(60\%), 5 \times 5\,(30\%), 3 \times 3\,(10\%)$.

Таблица 4.4 – Вероятности ошибки восстановления ЦВЗ

Степень перекрытия предполагаемой и реальной конфигурации фрагмента контейнера, %	100%	60%	30%	10%
Вероятность ошибки восстановления ЦВЗ, $P_{ош1}$	0.25	0.25	0.25	0.25
Вероятность ошибки восстановления ЦВЗ, $P_{ош2}$	0.42	0.40	0.38	0.31

В таблице 4.5 представлены результаты, полученные при аналогичных условиях при относительной амплитуде вносимого искажения $a_m = 5/255$.

Таблица 4.5 – Вероятности ошибки восстановления ЦВЗ

Степень перекрытия предполагаемой и реальной конфигурации фрагмента контейнера, %	100%	60%	30%	10%
Вероятность ошибки восстановления ЦВЗ, $P_{ош1}$	0.01	0.01	0.01	0.01
Вероятность ошибки восстановления ЦВЗ, $P_{ош2}$	0.09	0.08	0.03	0.01

В таблице 4.6 представлены результаты, полученные при аналогичных условиях при относительной амплитуде вносимого искажения $a_m = 10/255$.

Таблица 4.6 – Вероятности ошибки восстановления ЦВЗ

Степень перекрытия предполагаемой и реальной конфигурации фрагмента контейнера, %	100%	60%	30%	10%
Вероятность ошибки восстановления ЦВЗ, $P_{ош1}$	<0.01	<0.01	<0.01	<0.01
Вероятность ошибки восстановления ЦВЗ, $P_{ош2}$	0.07	0.06	0.02	0.01

Полученные результаты свидетельствуют о том, что для преобразований гетероассоциативного типа негативные воздействия стороннего наблюдателя оказывают весьма незначительное воздействие на встроенную последовательность ЦВЗ в зависимости от степени перекрытия предполагаемой и реальной конфигурации фрагмента контейнера даже при известных стороннему наблюдателю архитектуры нейронных сетей, определяющих оператор связи соседних фрагментов (рисунок 2.16, а) и реализующей сжимающее преобразование для последующего встраивания (рисунок 2.16, б)).

Следует отметить, что в случае даже в случае полного совпадения реальной и предполагаемой конфигурация вероятность ошибки восстановления ЦВЗ после деструктивных действий стороннего наблюдателя существенно не увеличивается. Это связано с тем, что процесс встраивания последовательности ЦВЗ для гетероассоциативного типа основан на использовании дополнительной прогнозирующей сети с помощью, которой определяется стохастическая составляющая прогноза, а затем реализуется непосредственно процедура встраивания ЦВЗ. При негативных воздействиях со стороны стороннего наблюдателя входная часть фрагмента контейнера, на основе которого реализуется прогнозирование выходной части, остается неизменным, тем самым встраивание последовательности ЦВЗ с амплитудой $a_m = 0$ в целом не разрушает встроенную последовательность ЦВЗ. Иначе говоря, происходит маскирование ЦВЗ ранее выделенной стохастической составляющей.

Кроме того, в экспериментах получены результаты оценки вероятности ошибки восстановления ЦВЗ без вмешательства стороннего наблюдателя $P_{ош1} = 0.0230$ и вероятности ошибки восстановления ЦВЗ при вмешательстве стороннего наблюдателя $P_{ош2} = 0.0411$ для следующих условий: относительная амплитуда вносимого искажения $a_m = 1/255$, размер исходного фрагмента файла-контейнера 11×11, размер фрагмента, используемого для встраивания одного элемента последовательности ЦВЗ 9×1, размер фрагмента файла-контейнера, предполагаемого сторонним наблюдателем, равен 1×9.

Также в экспериментах получены результаты оценки вероятности ошибки восстановления ЦВЗ без вмешательства стороннего наблюдателя $P_{ош1} = 0.0710$, вероятности ошибки восстановления ЦВЗ $P_{ош2} = 0.0520$ для следующих условий: относительная амплитуда вносимого искажения $a_m = 1/255$, размер исходного фрагмента файла-контейнера 11×11, размер фрагмента, используемого для встраивания одного элемента последовательности ЦВЗ определяется случайным образом.

В целом анализ этих результатов, полученных с использованием разработанной методики, показывает, что если стороннему наблюдателю известны архитектуры нейронных сетей, реализующих встраивание ЦВЗ и его последующее восстановление, при известной или известной частично конфигурации фрагмента исходного файла-контейнера, в котором происходит встраивание, при реализации сжимающих преобразований автоассоциативного типа возможно полное или частичное разрушение последовательности ЦВЗ, тогда как для преобразований гетероассоциативного типа уничтожение ЦВЗ существенно затруднено.

4.2. Программный комплекс для создания цифровых водяных знаков с использованием нейросетевых функциональных моделей преобразования данных в интересах защиты авторских прав на объекты электронного контента

Программный комплекс для создания цифровых водяных знаков с использованием нейросетевых функциональных моделей преобразования данных в интересах защиты авторских прав на объекты электронного контента реализован в среде Qt-4.6.0 и предназначен для работы в операционной системе Windows 7/Windows XP/Linux. В качестве языка разработки был выбран язык C++ как один из наиболее универсальных и легко переносимых на другие платформы.

Программный комплекс разработан автором диссертации в соавторстве [3,30,31]. В рамках программного комплекса реализовано несколько вариантов алгоритмов создания ЦВЗ. Лично автор принял участие в обосновании структуры

программного комплекса. Кроме того, им в полном объеме реализованы нейросетевые алгоритмы создания ЦВЗ, описание которых дано в главе 3.

Функциональные возможности разработанного программного комплекса позволяют решать следующие задачи:

– осуществлять надежное встраивание/извлечение цифровых водяных знаков в виде черно-белых пиктограмм в ряд файлов мультимедийных форматов (bmp, png, jpg, avi и пр.)

– осуществлять детальную настройку предложенных алгоритмов встраивания ЦВЗ с тем, чтобы минимизировать результирующие искажения заполненных контейнеров;

– создавать пользовательские ЦВЗ (черно-белые изображения);

– обеспечить удобство и простоту графического интерфейса, наглядность представления результатов работы алгоритмов встраивания и извлечения ЦВЗ.

Укрупненная структурная схема программного комплекса представлена на рисунке 4.4. На ней приведены наиболее важные программные модули, сгруппированные по своему функциональному назначению: реализующие совокупность алгоритмов создания и встраивания ЦВЗ, а также совокупность алгоритмов для анализа контейнеров и извлечения встроенных меток. В рамках основной части программного комплекса на рисунке 4.5 представлены:

– модули интерфейсной части, обеспечивающие возможность загрузки, просмотра, настройки параметров, характеризующих качество файлов-контейнеров, детальной настройки различных параметров алгоритмов встраивания ЦВЗ, удобство отображения результатов применения реализованных алгоритмов встраивания ЦВЗ;

– модули для кодирования/декодирования файлов различных форматов, используемых в качестве контейнеров для внедрения в них ЦВЗ;

– модули реализации алгоритмов встраивания/извлечения ЦВЗ;

– модули для генерации и настройки параметров ЦВЗ;

– модули для моделирования, обучения и тестирования нейронных сетей;

– вспомогательные математические модули.

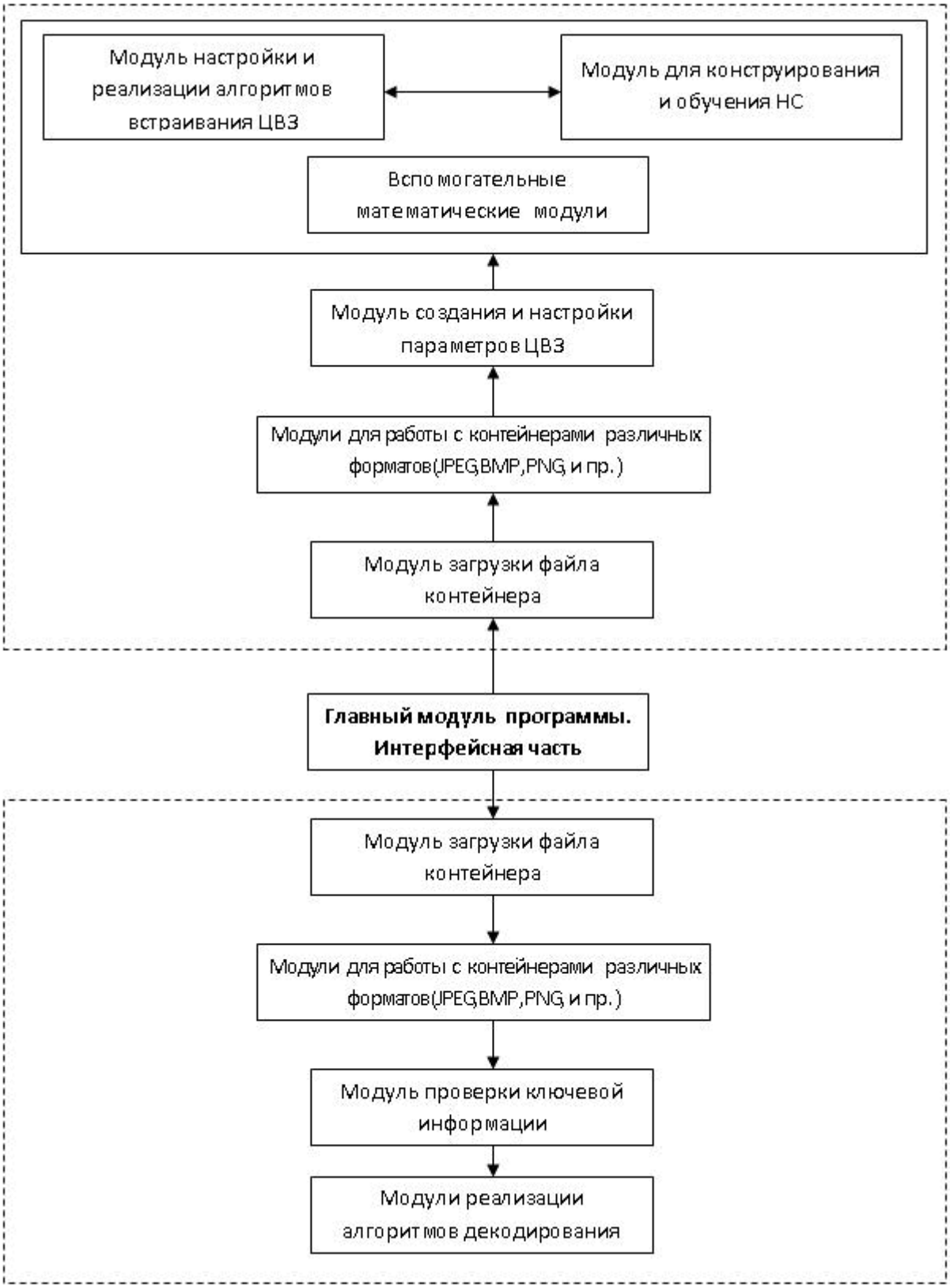


Рисунок 4.4 – Укрупненная структурная схема программного комплекса

При проектировании программного обеспечения, работающего с файлами различных форматов возникает необходимость отдельной программной реализации или использования специализированных библиотек для кодирования и декодирования используемых форматов. В разрабатываемом программном комплексе для кодирования и декодирования контейнеров графических форматов использовались встроенные классы библиотеки Qt, позволяющие получить доступ к элементам контейнера и модифицировать их необходимым образом. Для работы с контейнерами видео-форматов использовалась библиотека ffmpeg [59]. Для работы с нейронными сетями использовались модули разработанной библиотеки классов для моделирования и исследования нейронных сетей [14].

На рисунке 4.5 представлена упрощенная схема (в виде диаграммы классов UML), демонстрирующая основную логику реализованных алгоритмов встраивания и извлечения ЦВЗ.

Автором настоящей диссертации разработаны следующие классы в рамках реализованного программного комплекса (WmDrawEditorWidget, ImageResizeDlg, AlgFunctionalEmbeddingWidget, AlgFunctionalEmbedding), обеспечивающие реализацию методов пользовательского редактирования графических ЦВЗ и включают объекты, реализующие алгоритмы встраивания/восстановления ЦВЗ, а также объекты классов описания элементов управления для настройки параметров соответствующих алгоритмов.

Описание элементов управления главного окна приложения содержится в классе **MainWindow**.

Классы «Исходный файл-контейнер» **EmptyContainerPage** и «Маркированный файл-контейнер» **FilledContainerPage** описывают элементы графического интерфейса для соответствующих страниц загрузки и просмотра исходных незаполненных файлов и файлов, предположительно содержащих ЦВЗ.

Класс описания интерфейса «настройки ЦВЗ» **WatermarkConfPage** предназначен для задания способа создания цифровой метки (загрузки из файла, формирования из текстовой строки или использования пиктограммы созданной пользователем). В классах **WmDrawEditorWidget**, **ImageResizeDlg** реализованы

методы пользовательского редактирования графических ЦВЗ – рисование с помощью указателя мыши, инвертирование цветов, изменение размера пиктограммы. По окончании редактирования пиктограмма ЦВЗ должна быть сохранена в файл.

Поля классов «Алгоритмы встраивания ЦВЗ» **EmbAlgorithmConfPage** и «Алгоритмы извлечения ЦВЗ» **ExtractAlgorithmConfPage** включают объекты, реализующие алгоритмы встраивания/восстановления ЦВЗ, а также объекты классов описания элементов управления для настройки параметров соответствующих алгоритмов (**AlgFunctionalEmbeddingWidget** – настройки алгоритма функционального встраивания, **AlgLevelModificationWidget** – настройки алгоритма модификации уровней (блочный), **AlgLevelModificationNnWidget** – настройки алгоритма модификации уровней (НС)).

Объекты класса **AlgFunctionalEmbedding** реализуют логику работы алгоритма функционального встраивания ЦВЗ. Среди полей данного класса отметим значения высоты и ширины ЦВЗ (**unsigned int** wm_rows, wm_cols), количество встраиваемых экземпляров ЦВЗ (**unsigned int** num_wm_units), ширину и высоту кадра (**unsigned int** block_width, block_height), размерность входных и целевых данных НС (**unsigned int** ninputs, ntargets), объем обучающей выборки (**unsigned int** npatterns), вариант выбора элементов контейнера для встраивания (**int** elemChooseType), параметр для генерации ПСЧП (**int** seed), амплитуду встраиваемого процесса (**double** ampl), матрицу ПСЧП для хранения координат элементов контейнера используемых при формировании обучающей(тестовой) выборки (Matrix<**int**> psnr_matr), объект многослойный перцептрон (MultilayerPerceptron mp).

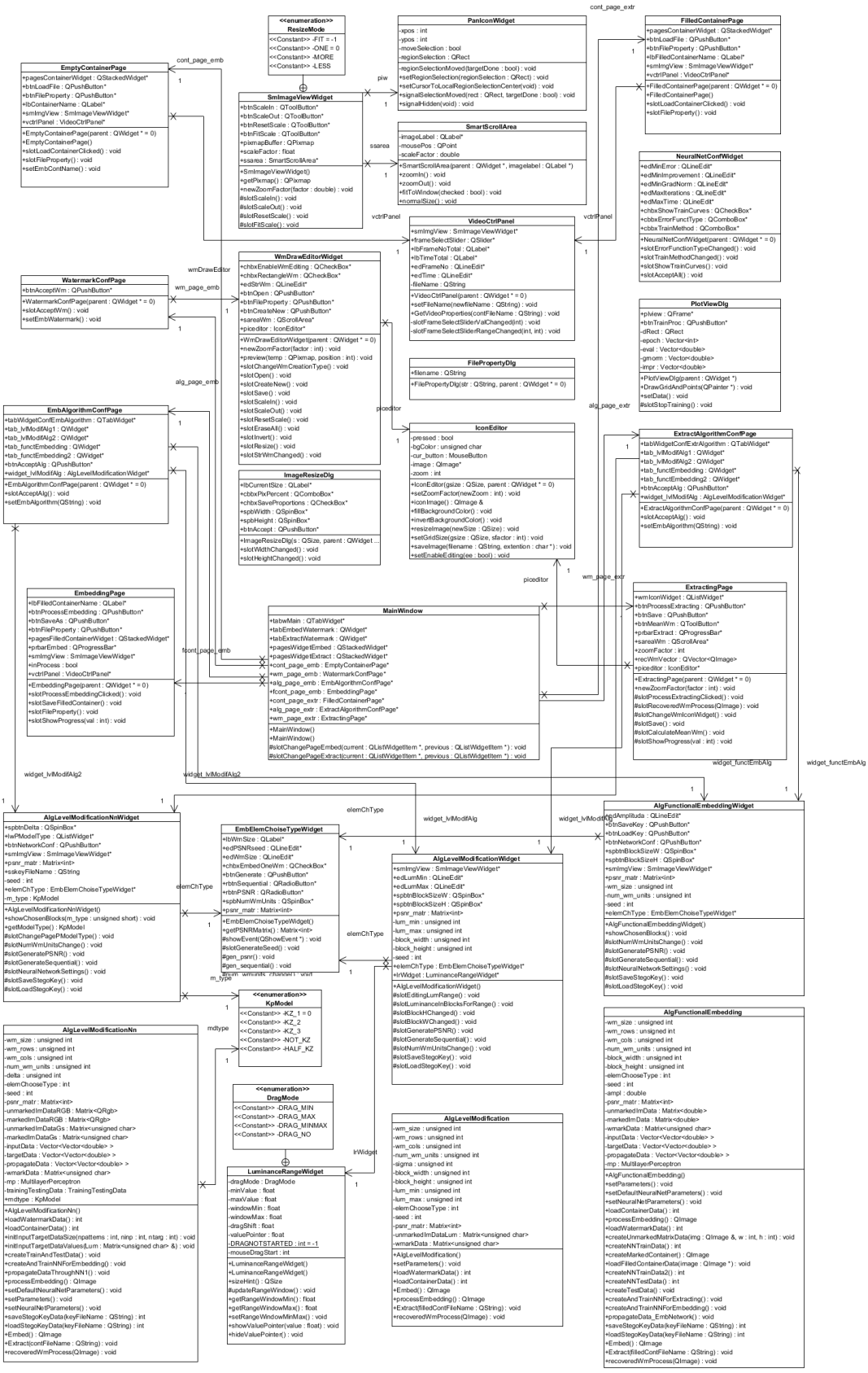


Рисунок 4.5 – Диаграмма классов UML, описывающая структуру основных классов программы встраивания и извлечения ЦВЗ

В классе **AlgFunctionalEmbedding** выполняется:

- задание общих параметров алгоритма встраивания/извлечения ЦВЗ;

- задание параметров для обучения нейронной сети;

- загрузка и заполнение матрицы ЦВЗ;

- формирование матрицы значений цветовых составляющих исходного и «заполненного» файла по кадрам, развернутым в строку;

- создание, настройка, обучение и тестирование НС 1-го и 2-го типа для встраивания и восстановления бит ЦВЗ;

- основные процедуры встраивания и восстановления ЦВЗ;

- формирование файла ключевых данных с его последующим сохранением на диск;

- загрузка файла ключевых данных, с разбором и корректировкой необходимых полей класса.

Классы **AlgLevelModification** и **AlgLevelModificationNn** инкапсулируют логику работы блочного и нейросетевого алгоритмов модификации уровней.

Настройка параметров нейронной сети осуществляется в классе **NeuralNetConfWidget**. С целью упрощения использования разработанного приложения ряд функциональных возможностей библиотеки для работы с НС (п.2.3) сделан недоступным в графическом окне настроек НС.

В классе **EmbeddingPage** реализован интерфейс страницы «встраивания ЦВЗ». В классе **ExtractingPage** реализован интерфейс страницы «восстановления ЦВЗ». Среди элементов управления данного класса можно отметить QListWidget *wmIconWidget, который содержит список пиктограмм восстановленных ЦВЗ, а также кнопку QToolButton *btnMeanWm, функциональное назначение которой заключается в усреднении набора восстановленных пиктограмм ЦВЗ.

Главное окно приложения, содержит две вкладки «Встраивание ЦВЗ» и «Извлечение ЦВЗ». На вкладке «Встраивание ЦВЗ» в левой части расположены кнопки для перехода на страницы выбора и загрузки контейнера, создания ЦВЗ, выбора и настройки параметров алгоритмов, встраивания и визуализации результатов.

На рисунке 4.6 представлено главное окно приложения с активной страницей загрузки и просмотра контейнеров. В данном примере в качестве контейнера выступает изображение формата IPEG.

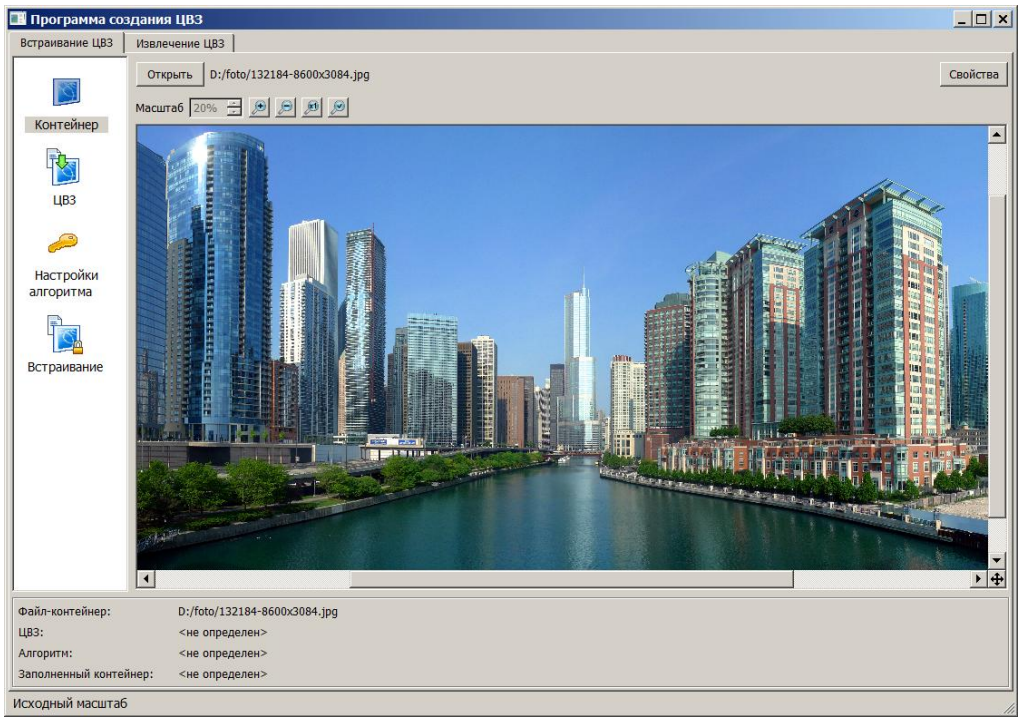


Рисунок 4.6 – Окно загрузки и просмотра исходного файла-контейнера

После загрузки защищаемого файла необходимо перейти на страницу создания ЦВЗ (рисунок 4.7). В качестве цифровых меток в разработанном приложении используются черно-белые изображения - пиктограммы. В меню «Варианты создания ЦВЗ» предусмотрено несколько методов создания подобных графических ЦВЗ.

В первом случае цифровой водяной знак будет загружаться из графического файла, при этом его глубина цвета уменьшается до 1 бита/пиксель. При необходимости загруженное ЦВЗ можно скорректировать, дорисовав или убрав отдельные пиксели, инвертировав цвета или изменив размер изображения-ЦВЗ.

Во втором случае при выборе переключателя «Создать новый ЦВЗ» задается размер пиктограммы, и дальнейшее рисование производится с помощью указателя мыши. Полученное в результате бинарное изображение должно быть сохранено в файл.

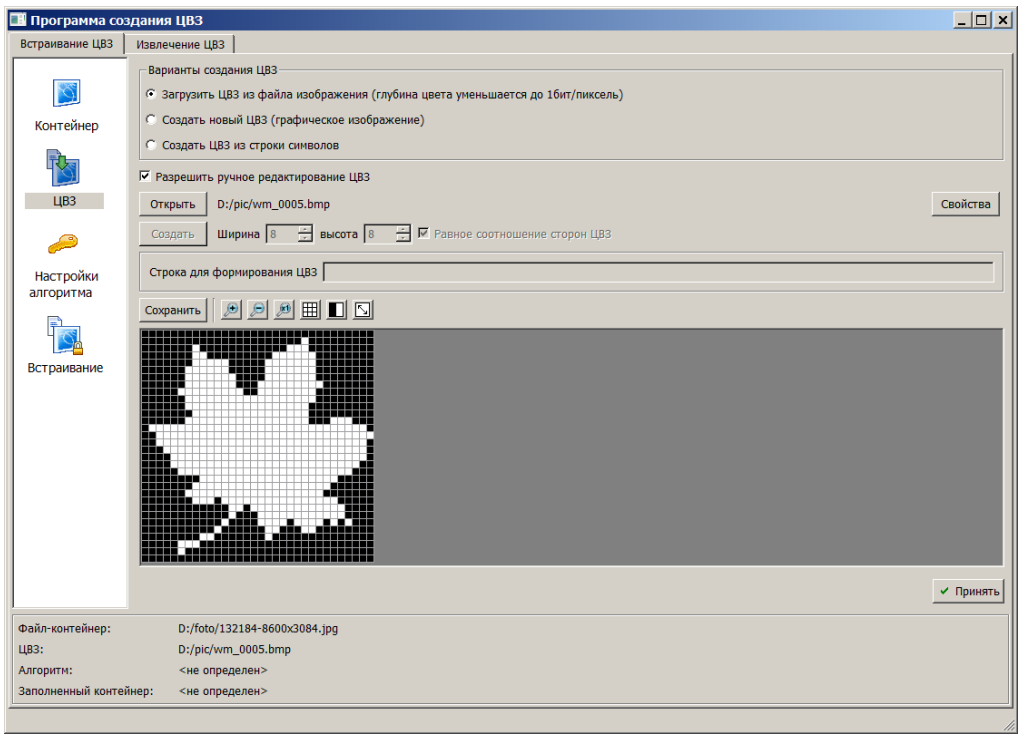


Рисунок 4.7 – Окно создания графического ЦВЗ

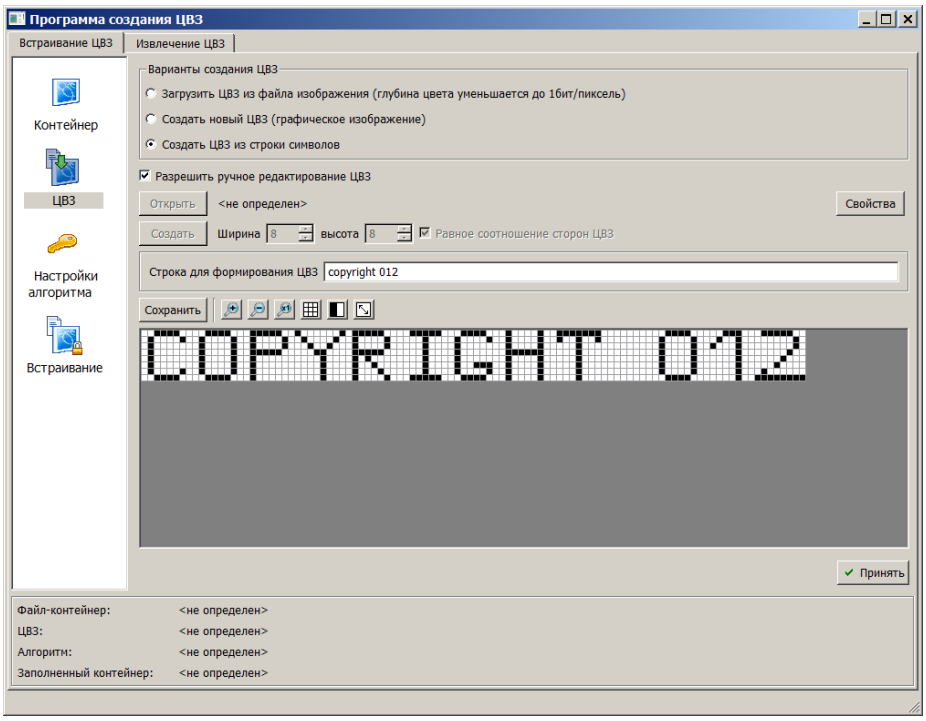


Рисунок 4.8 – Окно создания графического ЦВЗ на основе введенной строки символов

В программе также реализован вариант формирования цифровой метки на основе введенной пользователем строки символов. При этом каждый символ

строки преобразуется в графическое представление – матрицу размером 8x8 пикселей. Несмотря на то, что подобное преобразование увеличивает избыточность встраиваемых данных, оно призвано повысить устойчивость ЦВЗ по отношению к различного рода искажениям, а также увеличить надежности восстановления встроенной метки. В случае возможных искажений ЦВЗ вероятность сохранения разборчивости отдельного символа или строки символов при их графическом представлении, оказывается существенно выше по сравнению с ЦВЗ, сформированному из бит исходной строки. На рисунке 4.8 приведен пример формирования графического отображения на основе введенной пользователем строки символов.

Для восстановления ранее встроенной метки необходимо перейти на вкладку «Извлечение ЦВЗ». На данной вкладке в левой части окна расположены кнопки для перехода на страницы выбора и загрузки контейнера, выбора и настройки параметров алгоритмов, извлечения и визуализации результатов.

На рисунке 4.9 представлено главное окно приложения с активной страницей загрузки и просмотра контейнеров.

Рисунок 4.9 – Окно загрузки и просмотра «маркированного» файла-контейнера

Процедура восстановления ЦВЗ запускается нажатием кнопки «Извлечь». По окончании работы процедуры формируется список пиктограмм восстановленных ЦВЗ. На рисунке 4.10 представлено окно программы со списком из 40 извлеченных меток, каждая из которых имеет различный уровень искажений. Цифровая метка считается корректно восстановленной, если характер ее изображения схож с исходной пиктограммой. В данном примере все метки можно считать корректно восстановленными. Выбор метки для ее увеличенного просмотра осуществляется из списка эскизов ЦВЗ в верхней части окна.

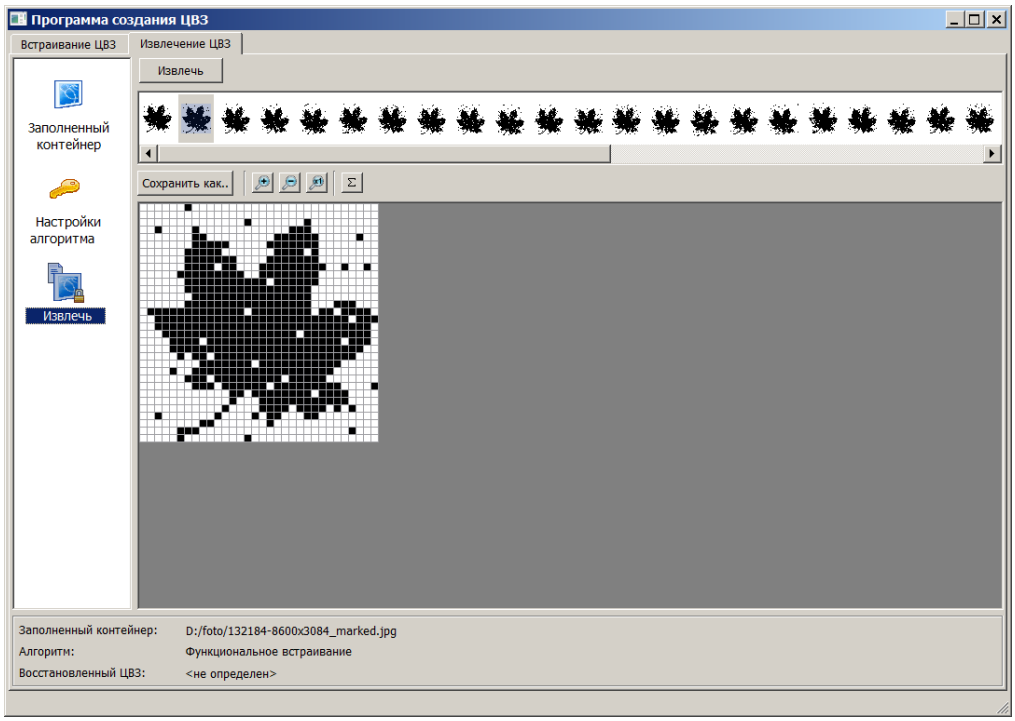
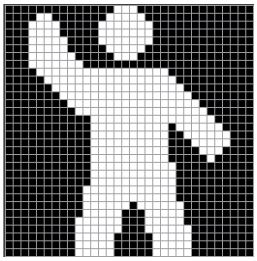


Рисунок 4.10 – Окно извлечения ЦВЗ и визуализации результата

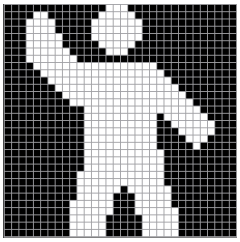
Далее приводятся результаты тестирования программного продукта на примере встраивания цифровых меток в графические файлы различных форматов с использованием всех реализованных алгоритмов. Оценить результаты тестирования, можно сравнив файлы, полученные на выходе алгоритмов встраивания и восстановления цифровой метки, с исходным контейнером и пиктограммой ЦВЗ. Исходный контейнер и ЦВЗ приведены на рисунке 4.11.



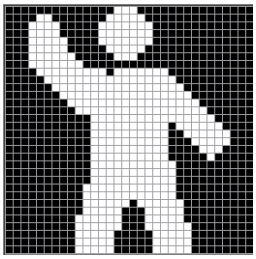
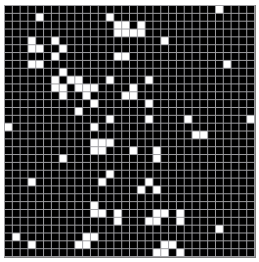
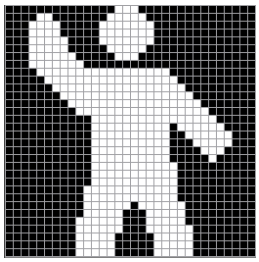
а б

Рисунок 4.11 –Исходный контейнер «aircraft.jpg» (75.4 Кб), размеры 900x588, глубина цвета 24 бит/пиксель (а); изображение ЦВЗ «wm_0001.bmp» (190 байт), размеры 32x32, глубина цвета 1бит/пиксель (б)

Таблица 4.7 – Результаты работы алгоритма функционального встраивания и извлечения ЦВЗ

Заполненный контейнер aircraft_m_afe.bmp (1.51 Мб),размеры 900x588, $a_m = 0.05$	 Восстановленная пиктограмма ЦВЗ

Продолжение таблицы 4.7

 Заполненный контейнер aircraft_m_afe.png (525 Кб), размеры 900x588, $a_m = 0.05$	 Восстановленная пиктограмма ЦВЗ
 Заполненный контейнер aircraft_m_afe.jpg (116.1 Кб), размеры 900x588, $a_m = 0.09$	 Восстановленная пиктограмма ЦВЗ
 Заполненный контейнер aircraft_m_afe.tiff (2.01 Мб), размеры 900x588, $a_m = 0.05$	 Восстановленная пиктограмма ЦВЗ

Как видно из таблицы 4.7 алгоритм нейросетевого функционального встраивания позволяет получить результирующие изображения очень высокого качества. При сравнении маркированных контейнеров с их оригиналами отыскать визуальные отличия практически невозможно. Надежное восстановление ЦВЗ происходит при сохранении изображений в форматы не предусматривающие сжатия или сжимающие данные без потерь. В случае JPEG-сжатия восстановление ЦВЗ происходит значительно хуже. Увеличение значения амплитуды встраиваемого процесса позволяет несколько улучшить характеристики восстанавливаемых ЦВЗ, однако данный шаг нежелателен, поскольку может привести к искажениям в защищаемом файле.

Основываясь на приведенных в результатах тестирования предложенных алгоритмов создания ЦВЗ можно сформулировать следующие критерии (или правила) выбора типа контейнеров [39]:

– использование контейнеров достаточно большого размера (достаточного для формирования обучающих множеств и обучения сетей, реализующих скрывающее и восстанавливающее преобразования);

– наличие определенного уровня вариативности (пространственной корреляции) контейнера (при изначально высоком уровне пространственной корреляции контейнера шум, вносимый встраивании ЦВЗ, будет демаскирующим признаком);

– наличие для контейнеров-изображений возможно большего числа пикселей, оттенки цветов которых плохо различаются глазом человека (выбор контейнера, содержащего наибольшие области зеленого, желтого и их смесей с белым цветов обеспечит наилучшую скрытность данных с точки зрения визуального анализа);

– использование контейнеров со «смешанным» характером изображения (монотонные изображения с плавными переходами яркости будут лучше восстанавливаться, однако возможные ошибки в таких контейнерах визуально более заметны по сравнению с «пестрыми» изображениями с большим числом перепадов яркости);

– получение контейнеров-изображений при помощи фотоаппарата или сканера, а не при помощи графических редакторов;

– отсутствие полезной составляющей в младших битовых плоскостях контейнера (при встраивании ЦВЗ высока вероятность потери ранее скрытой информации).

Кроме того, при использовании графических форматов данных для дополнительного снижения искажающих изменений в результате внедрения ЦВЗ с использованием, как функционального алгоритма скрытия, так и алгоритмов модификации уровней, следует учитывать особенности цветового восприятия человека. Темные цвета зрительно воспринимаются как черные, а оперируя в светлых тонах, глаз человека различает самые тонкие переходы как от цвета к цвету, так и от света к тени. Поэтому логично формировать обучающие и тестовые выборки для нейронных сетей из значений пикселей, выбранных в определенном цветовом (яркостном) диапазоне со смещением в сторону темных тонов.

Выводы по главе

1. Исходя из полученных в главе 2 настоящей диссертации, теоретических обоснований для нейросетевых функциональных моделей и универсальных алгоритмов преобразования данных, и, исходя из исследований, описывающих потенциальные характеристики качества создаваемых ЦВЗ для объектов различных классов (цветные изображения в формате RGB и YCbCr, аудио-файлы в формате WAV) выполненных в главе 3, разработаны типовые методики (при известном стороннему наблюдателю процедуры встраивания последовательности ЦВЗ (создания ЦВЗ)):

– методика анализа возможности обнаружения встроенного ЦВЗ;

– методика анализа возможности восстановления встроенной последовательности ЦВЗ при негативных воздействиях стороннего наблюдателя

2. Результаты проведенных исследований по оценки качества цифровых водяных знаков показывают, что при известной конфигурации фрагмента исходного файла-контейнера и при известной архитектуре нейронной сетей

возможно подтвердить или опровергнуть факт наличия встроенной последовательности ЦВЗ в файле-контейнере. Кроме того, результаты проведенных экспериментальных исследований на основе разработанных методик показывают, что при не известной или известной частично конфигурации фрагментов исходного файла контейнера возможно полное уничтожение скрытой последовательности ЦВЗ для преобразований автоассоциативного типа, для преобразований гетероассоциативного типа возможность частичного или полного разрушения последовательности ЦВЗ зависит от степени перекрытия реальной и предполагаемой конфигурации фрагмента контейнера.

3. Разработан программный комплекс для создания цифровых водяных знаков с использованием нейросетевых функциональных моделей преобразования данных в интересах задачи защиты авторских прав на объекты электронного (цифрового) контента. Комплекс реализован в среде Qt-4.6.0 и предназначен для работы в операционной системе Windows 7/Windows XP/Linux. В качестве языка разработки выбран язык C++ как один из наиболее универсальных и легко переносимых на другие платформы. В составе комплекса разработаны программные модули, сгруппированные по своему функциональному назначению и реализующие совокупность разработанных алгоритмов создания и встраивания ЦВЗ, включая алгоритмы, основанные на применении нейросетевых функциональных моделей преобразования данных, а также совокупность алгоритмов для анализа контейнеров и извлечения встроенных меток.

Заключение

В ходе выполнения диссертационной работы поставлены и решены следующие научные задачи:

- рассмотрены и проанализированы результаты ранее выполненных исследований, посвященных вопросам создания цифровых водяных знаков в объектах цифрового (электронного) контента;
- проведены теоретические обоснования, разработка и исследования нейросетевых функциональных моделей и алгоритмов преобразования данных, позволяющих реализовать процесс создания ЦВЗ в объектах цифрового (электронного) контента (файлы графических и аудио форматов) с минимальным уровнем искажения исходного контейнера;
- получены зависимости среднеквадратичной ошибки искажения контейнера, вероятности ошибки восстановления ЦВЗ, а также максимальной абсолютной ошибки искажения контейнера, описывающие потенциальные характеристики качества встраивания и восстановления ЦВЗ для реальных объектов различных классов;
- разработаны типовые методики для оценки качества алгоритмов встраивания последовательности ЦВЗ: методика анализа возможности обнаружения встроенного ЦВЗ и методика анализа возможности восстановления встроенной последовательности ЦВЗ при негативных воздействиях стороннего наблюдателя;
- разработаны программные средства для моделирования различных алгоритмов создания ЦВЗ с использованием нейросетевых функциональных моделей преобразования данных в интересах задачи подтверждения авторских прав на объекты электронного (цифрового) контента.

В ходе проведенных исследований получены следующие новые результаты и выводы.

1. Получены теоретические обоснования сходимости весовых коэффициентов двухслойной линейной нейронной сети с сокращенным числом нейронов в скрытом слое к компонентам собственных векторов, получаемых при

решении обобщенной задачи на собственные значения. При этом на выходе сети формируется приближение оптимальной линейной оценки в виде разложения по первым собственным векторам матрицы ковариации оценки, число которых равно числу нейронов в скрытом слое. Показано, что при настройке коэффициентов простейших линейных нейронных сетей в условиях прямого и косвенного обучения по представительным выборкам наблюдаемых данных имеется возможность сформировать оценки векторов наблюдаемых и ненаблюдаемых параметров, эквивалентные по своим свойствам статистически оптимальным. Полученные теоретические обоснования определили облик нейросетевых функциональных моделей преобразования информации с использованием сжимающих отображений и позволили реализовать два базовых варианта построения алгоритмов создания цифровых водяных знаков при реализации отображений автоассоциативного типа и гетероассоциативного типа.

2. Установлено, что при реализации алгоритма создания ЦВЗ на основе гетероассоциативных сжимающих отображений требуется с использованием функциональных возможностей нейронных сетей для любого произвольного фрагмента контейнера (прямоугольной формы, решетки случайной конфигурации внутри области прямоугольной формы и т.п.), представленного в виде вектора соответствующей размерности, построить отображения, реализующие в соответствующей последовательности следующую схему преобразования данных:

- оптимальную оценку (прогноз) значений данных выходной части фрагмента при известных значениях входной части;
- приближенную оценку значений данных выходной части фрагмента на основе сжимающего отображения при известных значениях входной части, обеспечивающего подготовку вектора-контейнера к встраиванию данных;
- модификацию выходной части на основе статистической обработки данных, предполагающей оценку нормированного собственного вектора выборочной матрицы ковариации выходной части контейнера и модификацию соответствующей высокочастотной составляющей каждого фрагмента контейнера на основе элементов двоичной последовательности ЦВЗ;

- восстановление ранее скрытого сообщения основе двоичного классификатора фрагментов контейнера.

Показано, что операторы, реализующие указанные преобразования, в общем случае могут быть получены на основе обучения нейронных сетей, имеющих специальную архитектур, согласованную с параметрами размерности и конфигурации входной и выходной части фрагментов контейнера.

3. Проведены экспериментальные исследования на основе имитационного статистического моделирования, когда в качестве статистических эквивалентов контейнера рассматривались реализации однородного гауссовского случайного поля с разными корреляционными функциями. Результаты выполненных исследований показывают, что при реализации рассмотренных вариантов моделей и алгоритмов создания ЦВЗ обеспечивается относительная средняя квадратичная ошибка искажения фрагментов контейнера порядка 10^{-4}. Вероятность ошибки восстановления элементов двоичной последовательности ЦВЗ для гетероассоциативных преобразований достигает порядка 10^{-1} при относительной амплитуде вносимого $a_m = 0.01$, порядка 10^{-2} при относительной амплитуде (по отношению к максимальному значению яркости пикселя) вносимого искажения $a_m = 0.09$, для автоассоциативных преобразований вероятность ошибки восстановления ЦВЗ равна нулю. В целом достижимый уровень средней квадратичной ошибки искажения контейнера при встраивании ЦВЗ зависит от характеристик временной или пространственной корреляции сигналов и изображений, на основе которых получены файлы-контейнеры; так для участков изображений, имеющих высокую степень пространственной корреляции, уровень ошибки существенно ниже; при этом визуально искажения контейнера, связанные с встраиванием ЦВЗ, могут быть более заметны по сравнению с участками, имеющими более низкую пространственную корреляцию. Для обеспечения одновременно низкого уровня средней квадратичной ошибки искажения контейнеров-изображений и их визуальной незаметности целесообразно в дальнейшем перейти к адаптивным схемам встраивания ЦВЗ, при которых встраивание элементов двоичной последовательности ЦВЗ

осуществляется для фрагментов, выбираемых на основе анализа характеристик пространственного градиента или других характеристик, определяющих степень «гладкости» используемого участка.

4. Проведены экспериментальные исследования для реальных изображений в формате RGB, полученные результаты показывают, что относительная средняя квадратичная ошибка искажения фрагментов контейнера достигает порядка $10^{-4}...10^{-5}$. При этом для изображений в формате RGB обеспечивается вероятность ошибки восстановления элементов двоичной последовательности ЦВЗ порядка 10^{-1} при относительной амплитуде вносимой двоичной последовательности $a_m = 0.005$ и порядка 10^{-2} при амплитуде $a_m = 0.035$; для изображений в формате YCbCr достигается вероятность ошибки восстановления элементов двоичной последовательности ЦВЗ порядка 10^{-2} при амплитуде $a_m = 0.005$, порядка 10^{-3} при амплитуде $a_m = 0.035$; для аудио-файлов (WAV формата) достигается относительная средняя квадратичная ошибка искажения фрагментов контейнера порядка $10^{-7}...10^{-8}$ при вероятности ошибки восстановления элементов двоичной последовательности ЦВЗ порядка $10^{-1}...10^{-2}$ в зависимости от частоты дискретизации сигнала.

5. Для исследования вопросов практического применения и оценки эффективности реализованных алгоритмов разработаны типовые методики анализа возможности обнаружения встроенного ЦВЗ и анализа возможности восстановления встроенной последовательности ЦВЗ при определенных воздействиях стороннего наблюдателя для оценки качества цифровых водяных знаков на основе нейросетевых функциональных моделей преобразования данных и базовых алгоритмов для изучения возможности обнаружения встроенного ЦВЗ или его частичного или полного уничтожения сторонним наблюдателем. При этом проведенные исследования по оценки качества цифровых водяных знаков показывают, что при известной конфигурации фрагмента исходного файла-контейнера и при известной архитектуре нейронной сетей возможно подтвердить или опровергнуть факт наличия встроенной последовательности ЦВЗ в файле-

контейнере. Кроме того, результаты проведенных экспериментальных исследований на основе разработанных методик показывают, что при не известной или известной частично конфигурации фрагментов исходного файла контейнера возможно полное уничтожение скрытой последовательности ЦВЗ для преобразований автоассоциативного типа, для преобразований гетероассоциативного типа возможность частичного или полного разрушения последовательности ЦВЗ зависит от степени перекрытия реальной и предполагаемой конфигурации фрагмента контейнера.

6. Для создания цифровых водяных знаков с использованием нейросетевых функциональных моделей преобразования данных в интересах задачи подтверждения авторских прав на объекты электронного (цифрового) контента в среде Qt-4.6.0 реализован программный комплекс, состоящий из программных модулей, сгруппированных по своему функциональному назначению и реализующих совокупность разработанных алгоритмов создания и встраивания ЦВЗ, включая алгоритмы, основанные на применении нейросетевых функциональных моделей преобразования данных, а также совокупность алгоритмов для анализа контейнеров и извлечения встроенных меток.

Таким образом, в диссертации решены все поставленные задачи научного исследования, и цель работы достигнута.

Список литературы

1. Аграновский, А.В. Стеганография, цифровые водяные знаки и стеганоанализ / А.В. Аграновский, А.В. Балакин, В.Г. Грибунин, С.А.Сапожников - М.: Вузовская книга,2009. – 220 с.

2. Алгазинов, Э.К. Анализ и компьютерное моделирование информационных процессов и систем / Э.К. Алагазинов, А.А. Сирота – М.: Диалог-МИФИ, 2009. – 416 с.

3. Алгазинов, Э.К. Программный модуль для создания цифровых водяных знаков с использованием искусственных нейронных сетей / Э.К. Алгазинов, М.А. Дрюченко, Е.Ю. Митрофанова, А.А. Сирота // Свидетельство о государственной регистрации программы для ЭВМ № 2012613096 от 30.03.2012 г. (Федеральная служба по интеллектуальной собственности, патентам и товарным знакам).

4. Афанасьев, В.И. Нейросетевые и статистически оптимальные алгоритмы оценки координат источников радиоизлучения в многопозиционных радиосистемах / В.И. Афанасьев, Э.А. Кирсанов, А.А. Сирота // Радиотехника. - 2003, №10. - С. 87-91.

5. Балакин, А.В. Разработка архитектуры программного комплекса и методов информационной защиты мультимедиа-информации с использованием цифровых водяных знаков : Дис. канд. техн. наук: 05.13.11, 05.13.19 / Балакин Александр Владимирович. – Ростов н/Д, 2005. -195 с.

6. Барсуков, В.С. Еще раз о стенографии – самой современной из древнейших наук / В.С. Барсуков, А.В. Шувалов // Специальная техника. – 2004. - №2. - С. 51-65.

7. Барсуков, В.С. Компьютерная стеганография вчера, сегодня, завтра / В.С. Барсуков, А.П. Романцов // Специальная техника. - 1998. - № 4,5. – С. 19-26.

8. Бахрушин, А.П. Спектральный анализ видеокадров на основе системы импульсных функций с целью синхронизации процессов внедрения и поиск цифровых водяных знаков / А.П. Бахрушин // Вестник Тихоокеанского государственного университета. – 2008. - №4. - С. 225-238.

9. Беллман, Р. Введение в теорию матриц / Р.Беллман, - М. :Наука, 1976. – 368 с.

10. Гонсалес, Р. Цифровая обработка изображений // Р. Гонсалес, Р. Вудс. – М.: Техносфера, 2005. - 1072 с.

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

12. Дрюченко, М.А. Восстановление регрессионных моделей случайных процессов и полей с использованием нейронных сетей / М.А. Дрюченко, Е.В. Воронова, А.А Сирота // Вестник Воронежского государственного университета. Серия: «Системный анализ и информационные технологии». – 2010.- №1. – С. 109-119.

13. Дрюченко, М.А. Нейросетевые модели и алгоритмы стеганографического скрытия информации / А.А. Сирота, М.А. Дрюченко // Труды Российского научно-технического общества радиотехники, электроники и связи имени А.С. Попова. – Москва, 2010. –Т.2. – С. 335-338.

14. Дрюченко, М.А. Статистические и нейросетевые алгоритмы синтеза и анализа стеганографически скрытой информации в аудио- и графических данных: дис. канд. техн. наук : 05.13.17 / Дрюченко Михаил Анатольевич – Воронеж, 2010. – 192 с.

15. Дуда, Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт. - М.: Мир, 1976. - 512 с.

16. Ефименко, В. С. Применение нейронных сетей в задачах оптимальной фильтрации / В. С. Ефименко, В.Н. Харисов, Е. Г. Стребков // Радиотехника. – 2000. - №7. - С. 56-61.

17. Кирсанов, Э.А. Обработка информации в пространственно-распределенных системах радио мониторинга: статистический и нейросетевой подход / Э.А. Кирсанов, А.А. Сирота. – М.: Физматлит, 2012. – 344 с.

18. Коваленко, М.П. Внедрение цифровых водяных знаков в графические изображения на основе искусственных нейронных сетей и генетических

алгоритмов: автореф. дис. канд. техн. наук: 05.13.19 / Коваленко Михаил Павлович. – Серпухов, 2012. -19 с.

19. Конахович, Г.Ф. Компьютерная стеганография. Теория и практика / Г.Ф. Конахович, А.Ю. Пузыренко. – М.:МК-Пресс, 2006. – 288 с.

20. Корн, Г. Справочник по математике (для научных работников и инженеров) / Г. Корн, Т. Корн. – Наука, 1974. – 832 с.

21. Левиев, Д.А. Описание алгоритмов встраивания ЦВЗ [Электронный ресурс] / Д.А. Левиев. - Режим доступа: http://re.mipt.ru/infsec/2004/essay/2004__Watermark_algorithms_and stability__LevievD.pdf

22. Мельников, Ю.П. Цифровые водяные знаки – новые методы защиты информации [Электронный ресурс] / Ю.П. Мельников, А.В. Теренин, В.Г. Погуляев // Компьютерная неделя. – 2007. - №48. – Режим доступа: http://www.pcweek.ru/security/article/detail.php?ID=105054.

23. Митекин, В.А. Алгоритм поблочного встраивания стойких ЦВЗ в крупноформатные изображения / Митекин В.А., Глумов Н.И. // Компьютерная оптика. – 2011. – Т. 35. - №3. – С. 368-372.

24. Митрофанова, Е.Ю. Адаптивные алгоритмы создания цифровых водяных знаков на основе сжимающих отображений / Е.Ю. Митрофанова, А.А. Сирота, М.А. Дрюченко // Сборник трудов Международной научно-технической конференции «Кибернетика и технологии XXI века (C&T 2013г.). – Воронеж, 2013. –Т.1. – С. 44-50.

25. Митрофанова, Е.Ю. Нейросетевые модели и алгоритмы стеганографического скрытия информации / Е.Ю. Митрофанова, А.А. Сирота, М.А. Дрюченко // Сборник трудов XI Международной научно-технической конференции «Информатика: проблемы, методология, технологии». – Воронеж, 2011. –Т.2. – С. 275-279.

26. Митрофанова, Е.Ю. Нейросетевые модели и алгоритмы стеганографического скрытия информации в BMP-файлах / Е.Ю. Митрофанова, А.А. Сирота, М.А. Дрюченко // Сборник трудов Международной научно-

технической конференции «Кибернетика и технологии XXI века (C&T 2011г.). – Воронеж, 2011. –Т.1. – С. 48-54.

27. Митрофанова, Е.Ю. Нейросетевые технологии создания скрытых цифровых водяных знаков / Е.Ю. Митрофанова, А.А. Сирота, М.А. Дрюченко// Нейрокомпьютеры: разработка и применение. – 2012. – №10. – С. 13–20.

28. Митрофанова, Е.Ю. Нейросетевые технологии создания цифровых водяных знаков в аудио-файлах / Е.Ю. Митрофанова, А.А. Сирота, М.А. Дрюченко // Сборник трудов XII Международной научно-технической конференции «Информатика: проблемы, методология, технологии». – Воронеж, 2012. –Т.1. – С. 361-362.

29. Митрофанова, Е.Ю. Нейросетевые функциональные модели преобразования данных и алгоритмы создания ЦВЗ при реализации сжимающих отображений гетероассоциативного типа / Е.Ю. Митрофанова, А.А. Сирота, М.А. Дрюченко // Сборник трудов XIV Международной научно-технической конференции «Информатика: проблемы, методология, технологии». – Воронеж, 2014. –Т.2. – С. 165-170.

30. Митрофанова, Е.Ю. Программный комплекс для создания цифровых водяных знаков как средства защиты объектов электронного контента / Э.К. Алгазинов, М.А. Дрюченко, Е.Ю. Митрофанова, А.А. Сирота // Информационные технологии. – 2012. – №9. – С. 60–66.

31. Митрофанова, Е.Ю. Программный комплекс для создания цифровых водяных знаков с использованием искусственных нейронных сетей / Э.К. Алгазинов, М.А. Дрюченко, Е.Ю. Митрофанова, А.А. Сирота // Свидетельство о государственной регистрации программы для ЭВМ № 2012613096 от 30.03.2012 г. (Федеральная служба по интеллектуальной собственности, патентам и товарным знакам).

32. Митрофанова, Е.Ю. Сходимость весов двухслойной линейной нейронной сети при построении оптимальных оценок случайных векторов / Е.Ю. Митрофанова, А.А. Сирота// Нейрокомпьютеры: разработка и применение. – 2011. – №7. – С. 39–48.

33. Назаров, Л.Е. Применение искусственных нейронных сетей для сжатия РСА - и сканерных изображений земной поверхности / Л.Е. Назаров - Исследование Земли из космоса. – 1999. – № 5. – С.44-50.

34. Нечта, И.В. Разработка методов обеспечения безопасности использования информационных технологий, базирующихся на идеях стеганографии: автореф. дис. канд. техн. наук : 05.13.17 / Нечта Иван Васильевич. – Новосибирск, 2012. - 21 с.

35. Осовский, С. Нейронные сети для обработки информации / пер. с польского И.Д. Рудинского. - М.: Финансы и статистика, 2002. - 344 с.

36. Перов, А.И. Особенности синтеза устройств обнаружения и оценки параметров сигнала нейросетевыми методами / А.И. Перов, Г. Г. Соколов // Радиотехника. - 2001, №7. - С. 22-29.

37. Петраков, А.В. Защита абонентского телетрафика / Петраков А.В., Лагутин В.С.. – М.: Радио и связь, 2001. – 504 с.

38. Попов, В.Г. Свойства сходимости весов автоассоциативной двухслойной линейной нейронной сети при построении сжимающих отображений случайных векторов / В.Г. Попов, А.А. Сирота // Нейрокомпьютеры. – 2009. – № 5. – С.3-11.

39. Программное обеспечение для реализации новой информационной технологии защиты авторских прав на основе создания цифровых водяных знаков с использованием нейросетевых функциональных моделей преобразования данных: отчет о НИОКР / А.А. Сирота. – Воронеж: федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Воронежский государственный университет», 2011. – 163 с.

40. Рао, С.Р. Линейные статистические методы и их применения / С.Р. Рао. – М.: Наука, 1968 - 548 с.

41. Сирота, А.А. Нейросетевые и статистически оптимальные алгоритмы принятия решений в задачах высокоточной координатометрии и обработки изображений / А.А. Сирота, Э.А. Кирсанов, О. В. Маслов // Сборник трудов III Международной конференции «Идентификация систем и задачи управления». -

М.: Институт проблем управления им. В.А. Трапезникова РАН, 2004. - С. 1629-1652.

42. Сирота, А.А. Нейросетевые модели и алгоритмы имитации текстур цветных изображений земной поверхности / А.А. Сирота, О. В. Маслов // Нейрокомпьютеры: разработка и применение. - 2005. - №4. - С. 28-32.

43. Сирота, А.А. Нейросетевые модели и алгоритмы стеганографического скрытия информации / А.А. Сирота, М.А. Дрюченко // Информационные технологии. – 2011 . – №3. – С.41-49.

44. Татузов, А.Л. Нейронные сети в задачах радиолокации / А.Л. Татузов. - М.: Радиотехника, 2009. - 432 с.

45. Тихонов, В.И. Статистический анализ и синтез радиотехнических устройств и систем / В.И. Тихонов, В.Н. Харисов. - М.: Радио и связь,1991. - 608 с.

46. Фукунага, К. Введение в статистическую теорию распознавания образов: Пер. с англ. / К. Фукунага. – М.:Наука, 1979. – 368 с.

47. Хайкин, С. Нейронные сети: полный курс / С. Хайкин. – М.: Вильямс, 2005. – 1104 с.

48. Хорошко, В.А. Методы и средства защиты информации / В.А. Хорошко, А.А. Чекатков. - К.: Юниор, 2003. – 504 с.

49. Anderson, R. J. Stretching the Limits of Steganography / R. J. Anderson // Information Hiding Springer Lecture Notes in Computer Science. - 1996. - P. 39-48.

50. Bassia, P. Robust Audio Watermarking In The Time Domain / P. Bassia, I. Pitas, N. Nikolaidis // II Department of Informatics. University of Thessalonica. – 2001. – Vol.3. – Issue. 2. – P. 232-241.

51. Bender, W. Techniques for Data Hiding / W. Bender, D. Gruhl, N. Morimoto, A. Lu // IBM Systems Journal – 1996. – Vol. 35. - P. 313-336.

52. Cachin, C. An Information-Theoretic Model for Steganography / C. Cachin // Information Hiding — 2nd International Workshop, Springer as Lecture Notes in Computing Science. – 1998. - Vol.1525. -P. 306-318.

53. Chang, C.-C. A technique for computing watermarks from digital images. / C.-C. Chang, C.-S. Tsai // Informatica. – 2000. – Vol. 24. – P. 391-396.

54. Chen, J. Digital Watermarking Based on Neural Network Technology for Grayscale Images / J. Chen, Tung-Shou Chen, Keh-Jian Ma, Pin-Hsin Wang // Encyclopedia of Multimedia Technology and Networking . – 2005. – Vol. 29. - P. 204-212.

55. Chuan-Yu, C. Using Counter-propagation Neural Network for Robust Digital Audio Watermarking in DWT Domain / Chuan-Yu Chang, Wen-Chih Shen, Hung-Jen Wang // Systems, Man and Cybernetics (SMC). IEEE International Conference on . – 2006. – Vol. 2. – P. 1214-1219.

56. Cox, J. Watermarking as communications with side information / J. Cox, M. Miller, A. McKellips // Proceedings of the IEEE. - 1999. - Vol. 87. - № 7. - P. 1127-1141.

57. Craver, S. On Public-Key Steganography in the Presence of an Active Warden / S. Craver // Information Hiding. Lecture Notes in Computer Science. – 1998. - Vol. 1525. – P. 355-368.

58. Darmstaedter, V. Low cost spatial watermarking / V. Darmstaedter, J.-F. Delaigle, J. Quisquater, B. Macq // Computers and Graphics. - 1998. – Vol. 22. – №. 4. – P. 417-424.

59. FFmpeg [Электронный ресурс]. – Режим доступа: http://ffmpeg.org/

60. Fridrich, J. Steganalysis of LSB Encoding in Color Images / J. Fridrich, R. Du, M. Long // Proceedings of ICME 2000. - 2000. – Vol. 3. - P. 1279-1282.

61. Gustavus, J. Simmons. The History of Subliminal Channels / J. Simmons Gustavus // II IEEE Journal on Selected Areas of Communications. - 1998. - Vol. 16. - №4. - P. 452-461.

62. Gustavus, J. Simmons. The Prisoner's Problem And The Subliminal Channel / J. Simmons Gustavus // Advances in Cryptology: Proceedings of Workshop on Communications Security (Crypto'83. David Chaum, ed.). - 1984. - P. 51-67.

63. Hwang, M.S. A time-stamping protocol for digital watermarking / M.S. Hwang, K.F. Hwang, C.C. Chang // Applied Mathematics and Computation. – 2005. – Vol. 169. – Issue 2. – P. 1276-1284.

64. Johnson, N. F. Information Hiding: Steganography and Watermarking - Attacks and Countermeasures / F. Johnson, Zoran Duric, Sushil Jajodia. - Springer, 2001. – 137 pp.

65. Johnson, N. F. Steganalysis of Images Created using Current Steganography Software / N. F. Johnson, S. Jajodia // Workshop on Information Hiding Proceedings. Portland Oregon. Springer as Lecture Notes in Computer Science. – 1998. - Vol. 1525. – P. 273-289.

66. Johnson, N.F. Steganalysis: The Investigation of Hidden Information / N.F. Johnson, S. Jajodia // IEEE Information Technology Conference, Syracuse, New York, USA. - 1998. – P. 113-116.

67. Kim, B.S. Robust digital image watermarking method against geometrical attacks / B.S. Kim, J.G. Choi, C.H. Park, J.U. Won, D.M. Kwak, S.K. Oh, C.R. Koh, K.H. Park // Real-Time Imaging. – 2003. – Vol. 9. – P. 139-149.

68. Kutter, M. Digital signature of color images using amplitude modulation / M. Kutter, F. Jordan, F. Bossen // Proceedings of the SPIE (Storage and Retrieval for Image and Video Databases V). - 1997. - Vol. 3022. – P. 518-526.

69. Langelaar, G. Robust labeling methods for copy protection of images / G. Langelaar // Proceedings of the SPIE (Storage and Retrieval for Image and Video Databases V). - 1997. - Vol. 3022. - P. 298-309.

70. Lin, C.Y. Rotation, scale, and translation resilient watermarking for images / C.Y. Lin, M. Wu, J.A. Bloom, I.J. Cox, M.L. Miller, Y.M. Lui // IEEE Trans on Image Processing. – 2001. – №10. – P. 767-782.

71. Maes, M. Digital image watermarking by salient point modification practical results / M. Maes, P. Rongen, van C. Overveld // SPIE Conference on Security and Watermarking of Multimedia Contents. - 1999. - Vol. 3657. - P. 273-282.

72. Mitrofanova, E.Yu. Digital Watermarking Based on Neural Network Technologies / M.A. Dryuchenko, E.Yu. Mitrofanova, A.A. Sirota // JCIT: Journal of Convergence Information Technology. – 2013. - Vol. 8. - Issue 11. - PP. 616 - 624.

73. Nikolaidis, N. Robust image watermarking in the spatial domain / N. Nikolaidis, I. Pitas // Signal Processing. – 1998. – Vol. 66. – № 3. - P. 385-403.

74. O'Ruanaidh, J.K. Rotation, scale, and translation invariant digital image watermarking / J.K. O'Ruanaidh, T. Pun // Proceedings of ICIP. –1997. – Vol. 1. – P. 536-539.

75. O'Ruanaidh, J.K. Rotation, scale, and translation invariant spread spectrum digital image watermarking / J.K. O'Ruanaidh, T. Pun // Signal Processing. –1998. – Vol. 66. – P. 303-317.

76. Pao-Ta, Yu Digital watermarking based on neural networks for color images / Pao-Ta Yu, Hung-Hsu Tsai, Jyh-Shyan Lin // Signal Processing. – 2001. – Vol. 81. – Issue 3. – P. 663-671.

77. Pereira, S. Robust template matching for affine resistant image watermarks / S. Pereira, T. Pun // IEEE Trans. on Image Processing. – 2000. – Vol. 9. – P. 1123-1129.

78. Pereira, S. Template based recovery of Fourier-based watermarks using log-polar and log-log map / S. Pereira, J.K. O'Ruanaidh, F. Deguillaume, G. Csurka, T. Pun //Proceedings of ICMCS. – 1999. – P. 870-874.

79. Pfitzmann, B. Information Hiding Terminology / B. Pfitzmann // Information Hiding, Springer Lecture Notes in Computer Science. – 1996. - P. 347-350.

80. Popa, R. An Analysis of Steganographic Techniques / R. Popa. - The Polytechnic University of Timisoara. Faculty of Automatics and Computers. Department of Computer Science and Software Engineering, 1998. – 59 pp.

81. Schneier, B. Applied Cryptography: Protocols. Algorithms, and Source Code in C / B. Schneier. - John Wiley and Sons, 1996. – 784 pp.

82. Van Schyndel, R. G. A Digital Watermark / R. G. van Schyndel, C. Osborne, A. Tirkel // IEEE In-tern. Conf. on Image Processing. - 1994. - P. 86-90.

83. Voloshynovskiy, S. Attack Modelling: Towards a Second Generation Watermarking Benchmark / S. Voloshynovskiy, S. Pereira, V. Iquise, T. Pun. // Signal Processing. – 2001. – Vol. 81. – P. 1177-1214.