1. Выбор документа.
2. Подготовка документа к проверке.
3. Проверка текста с использований образцов во внутренней базе образцов и источников.
4. Выдача отчет о проверке

Заключение. В данной работе рассмотрены некоторые алгоритмы поиска заимствования, а также возможность работы с базой данных для быстрой идентификации совпадающих значений. Несмотря на наличие как платных, так и бесплатных систем поиска плагиата, решение проблемы «нечеткого дубликата», который мы понимаем как частично или полностью заимствованный документ, является актуальной задачей на сегодняшней день.

В дальнейшем предполагается провести сравнение алгоритмов поиска заимствований, выбрать наиболее оптимальные из них, и провести вычислительные эксперименты с использованием автоматизированной системы.

ЛИТЕРАТУРА

1. Толково-энциклопедический словарь русского языка [Электронный ресурс] // Slovar.cc. 2012. URL: https://slovar.cc/rus/tolk-enc/1462480.html (дата обращения 23.12.2016).
2. Альгирдас Аушра. Научная электронная библиотека как средство борьбы с плагиатом // Образовательные технологии и общество. 2006. Том 9. №2. С. 270-276.
3. Романов, А.С.. Модификация метода накопительных сумм для проверки однородности текста и выявления плагиата // А.С. Романов, Электронные средства и системы управления. 2013. № 2. С. 30-38.
4. Алгоритм шинглов [Электронный ресурс] // Википедия. Свободная энциклопедия. URL: https://ru.wikipedia.org/wiki/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC_%D1%88%D0%B8%D0%BD%D0%B3%D0%BB%D0%BE%D0%B2 (дата обращения 23.12.2016).
5. Модифицированный алгоритм шинглов // OrionXL. URL: http://orionxl.ru/modificirovannyj-algoritm-shinglov.html (дата обращения 19.12.2016).
6. Загорулько, Ю. А. Выявление нечетких дубликатов при автоматическом формировании тематических коллекций документов на основе web-публикаций // Ю.А, Загорулько, Н. В. Саломатина, А. С. Серый Е. А. Сидорова, В. К. Шестаков. Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2013. Том 11. № 4. С. 59-70.
7. Береснева, Д.Д. Краткий обзор методов выявления искусственных текстов [Электронный ресурс] // URL: http://conf58.mipt.ru/static/reports_pdf/1020.pdf (дата обращения 24.12.2016).
8. Астапова, О.П. Исследование и разработка методов поиска плагиата в многоязычных корпусах текстов [Электронный ресурс] // URL: http://seminar.at.ispras.ru/wp-content/uploads/2012/07/Astapova-thesis1.pdf (дата обращения 24.12.2016).
9. Васин А.Д. Разработка системы анализа текста на наличие заимствований [Электронный ресурс] // А.Д. Васин, Д.С.Бургонский, URL: http://36.msiu.ru/files/203-xioamehtvkgs.pdf (дата обращения 23.12.2016).
10. Цхай, А.А. Обнаружение плагиата с использованием нереляционных баз данных // А.А.Цхай, С.В.Будаков, С.В.Мурзинцев, Л.С.Ким, Вестник алтайской науки. 2015. №1. С. 280-285.

УДК 615.035.4

Кудрина М.А., Дулимова И.Е.

ФГАОУ ВО «Самарский национальный исследовательский университет имени академика С.П.Королева», Самара, Россия

СКРЫТИЕ ИНФОРМАЦИИ В АУДИОФАЙЛАХ МЕТОДАМИ СТЕГАНОГРАФИИ

Разработана автоматизированная система встраивания данных, позволяющая осуществлять встраивание информации в файл-контейнер и извлечение встроенной информации. Исследована эффективность методов стеганографии при встраивании текстовой информации в аудиофайлы и проведен сравнительный анализ с помощью разработанной системы. Для сравнения эффективности методов при встраивании разных объемов данных в качестве характеристики использовалось количество информации и степень отклонения от исходного сигнала. В системе реализованы следующие методы стеганографии: метод наименее значащих битов, метод фазового кодирования. В системе реализовано встраивание текстовой информации в аудиофайлы, также визуализирован механизм извлечения встроенного сообщения. Определены оптимальный объем данных, встраиваемых в аудиофайл, и степень отклонения от исходного файла для решения задачи эффективного встраивания.

Ключевые слова:
стеганография, аудиофайл, метод наименее значащих битов, метод фазового кодирования.

В настоящее время как никогда существует острая необходимость защиты информации от несанкционированного доступа. Существует два основных решения этой задачи: криптография и стеганография. В криптографии основной целью является скрытие содержимого сообщений за счет шифрования данных. В то время как в стеганографии скрывается сам факт существования тайного сообщения.

Общей чертой различных способов стеганографии является то, что скрываемое сообщение встраивается в обычный, не вызывающий подозрение объект. После чего данный объект открыто доставляется адресату. В криптографии наличие шифрованного сообщения само по себе создает угрозу целостности информации, при стеганографии наличие скрытого сообщения остается незаметным.

Стеганография – это метод организации связи, который скрывает само наличие связи. Общей чертой различных способов стеганографии является то, что скрываемое сообщение встраивается в обычный, не вызывающий подозрение объект. После чего данный объект открыто доставляется адресату. В криптографии наличие шифрованного сообщения само по себе создает угрозу целостности информации, при стеганографии наличие скрытого сообщения остается незаметным [1].

Сообщение, факт передачи которого хотят скрыть, называют секретным сообщением. Файл, не содержащий секретного сообщения, называется пустым контейнером, а файл с включенным сообщением – заполненным контейнером. Стеганографический канал (стегоканал) – канал передачи стегоконтейнера. Ключ (стегоключ) – секретный ключ, нужный для сокрытия стегоконтейнера.

Предполагается, что при помощи стеганографического алгоритма секретное сообщение встраивается в контейнер так, чтобы не было заметных изменений этого файла. Контейнер пересылается по открытому каналу связи, не вызывая подозрений. Секретное сообщение извлекается получателем при помощи специального алгоритма. В качестве контейнеров могут быть использованы любые файлы, для которых придуманы такие алгоритмы. Наибольший интерес для использования в качестве контейнеров представляют файлы распространенного типа содержимого, например, фотографии или аудиофайлы.

Звуковые и видеофайлы, как правило, довольно избыточны, поэтому незначительное изменение потока данных не приводит к заметным искажениям. Так как необходимость передачи аудиофайлов через сеть в последнее время очень возросла, можно сделать вывод о том, что звуковые файлы можно успешно использовать в качестве стегоконтейнера.

Тема сокрытия информации методами стеганографии является актуальной, так как необходимость

защиты информации от несанкционированного доступа имеет важное научное и техническое значение, а также представляется интересным изучить возможность использования стеганографических методов применительно к аудиофайлам.

Для исследования эффективности методов стеганографии была разработана автоматизированная система для встраивания больших объемов данных, которые используются для формирования скрытого канала передачи информации, а также для решения задач помехоустойчивой аутентификации, защиты информации от несанкционированного копирования, отслеживания распространения информации по сетям связи, поиска информации в мультимедийных базах данных. Работа системы основана на двух стеганографических методах.

Одним из первых методов встраивания информации в аудиофайлы, затрагивающих область аудиоданных, является метод изменения малозначащих битов (LSB). Чаще всего он применяется для сокрытия в аудиофайлах формата WAV благодаря простоте осуществления вставки [2]. Метод заключается в использовании погрешности дискретизации, которая всегда существует в оцифрованных изображениях или аудио- и видеофайлах. Данная погрешность равна наименьшему значащему разряду числа, определяющего величину элемента файла. Поэтому модификация младших битов в большинстве случаев не вызывает значительной трансформации файла [3].

Используя звуковой сигнал, путем замены наименьших значащих бит каждой точки осуществления выборки, представленной двоичной последовательностью, можно зашифровать значительный объем информации. Сам процесс встраивания информации аналогичен тому, который используется для изображения-контейнера, то есть в каждом значении амплитуды наименьший значащий бит заменяется на бит сообщения. Схема работы метода показана на рис. 1.

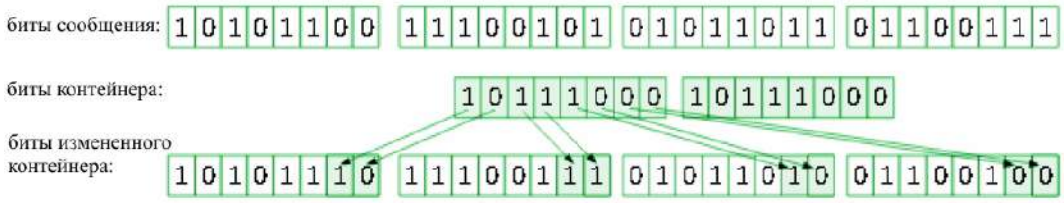

Рисунок 1 – Схема работы метода LSB

В рамках данной работы реализован метод изменения малозначащих битов, разработана автоматизированная система встраивания текстовых данных в контейнер аудиофайла, реализована обратная операция по извлечению встроенных данных из контейнера и произведено исследование возможностей использования данного метода. На рис. 2 представлен интерфейс системы.

Большое влияние на эффективность алгоритма оказывает отношение размера файла, используемого в качестве контейнера, к объему встраиваемых данных.

Популярность данного метода обусловлена тем, что он позволяет скрывать в относительно небольших файлах достаточно большие объемы информации. Основной недостаток метода – высокая чувствительность к малейшим искажениям контейнера, слабая устойчивость к посторонним воздействиям на сигнал (сжатие, воздействие шумов).

Также в системе реализован метод фазового кодирования, проведен сравнительный анализ и оценка эффективности методов.

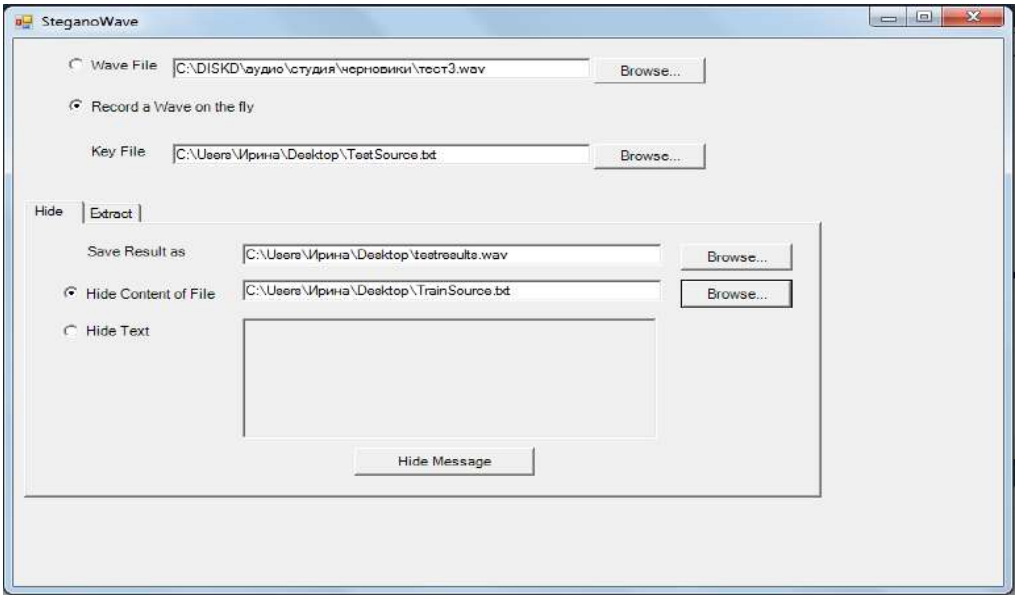

Рисунок 2 – Интерфейс системы

Метод фазового кодирования заключается в разбиении сигнала на сегменты и замещении фазы начального сегмента аудио-сигнала на фазу, которая характеризует данные (секретное сообщение). Фазы последующих сегментов регулируются в целях сохранения разности фаз между сегментами. Метод фазового кодирования является одним из методов, который устойчив к сжатию и воздействию шумов. Когда отношение фаз между частотами сильно меняется, происходит заметная фазовая дисперсия. Однако "неслышное" кодирование при этом все равно достигается, так как изменение фазы достаточно мало.

Фазовое кодирование осуществляется в соответствии с формулой:

$$S_m(t) = g(t)\cos[2\pi f_c t + \varphi_m(t)],$$

где g(t) определяет огибающую сигнала; $\varphi_m(t)$ является модулирующим сигналом; f_c – частота несущей; t – время.

Данный метод является более стойким к атакам. С другой стороны, он в некоторой степени снижает субъективное качество звука.

Методы, основанные на изменении фазовой области, как правило, изменяют абсолютное значение фазы некоторых гармоник. Существуют вариации, в которых восстанавливаются значения разностей фаз

смежных фреймов. Делается это исходя из предположения, что человеческое восприятие звука чувствительно не к абсолютному значению фазы сигнала, а к разности фаз смежных фрагментов. В простейших случаях эти методы неустойчивы к модификациям данных.

Разработка логического проекта системы выполнена в бесплатных средах UML-моделирования Microsoft Office Visio 2007 и Draw.io. На рис. 3 приведена диаграмма вариантов использования разработанной системы.

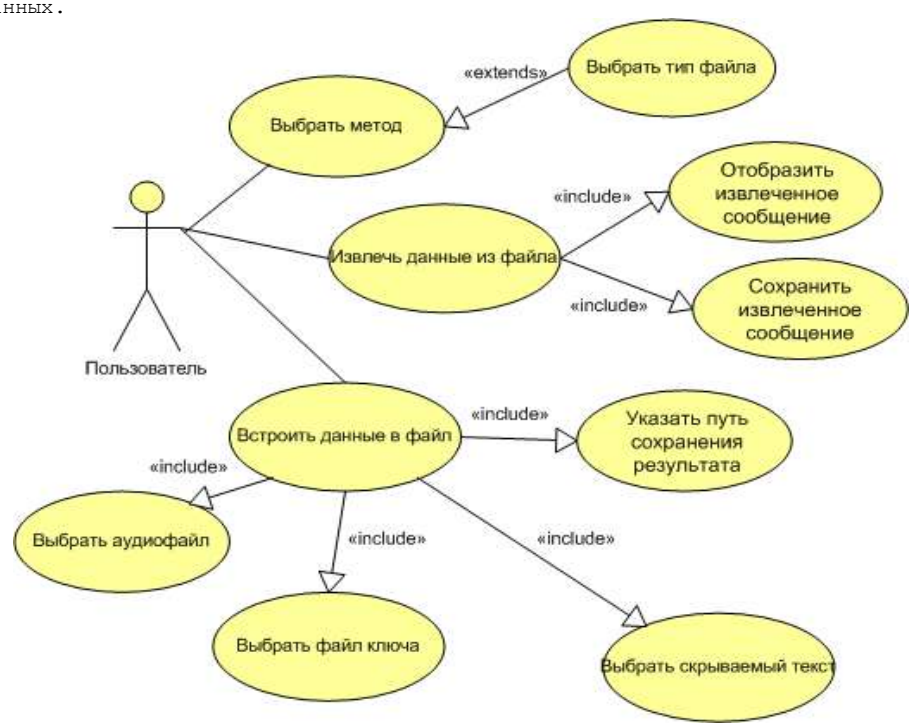

Рисунок 3 – Диаграмма вариантов использования системы

Для сравнения эффективности методов при встраивании разных объемов данных в качестве характеристики использовалось количество информации и степень отклонения от исходного сигнала.

В таблице 1 представлены результаты исследования зависимости степени отклонения от исходного сигнала от количества встраиваемой информации.

Зависимость степени отклонения от исходного сигнала от количества встраиваемой информации *Таблица 1*

Метод	Количество информации, I (байт)	Степень отклонения, %
LSB	50	10
LSB	500	27
LSB	1000	34
Фазового кодирования	50	7
Фазового кодирования	500	16
Фазового кодирования	1000	24

Полученные результаты позволили построить наглядное представление данной зависимости. На рис. 4 представлен график зависимости степени отклонения от исходного сигнала от количества встраиваемой информации для каждого из методов.

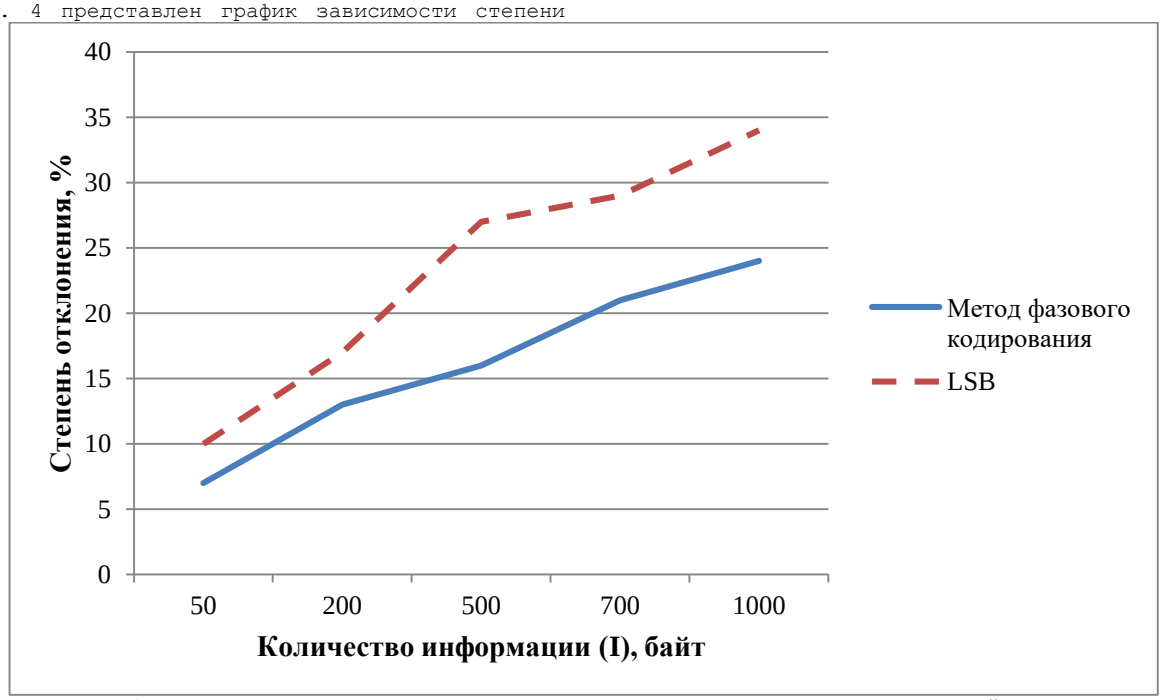

Рисунок 4 – График зависимости степени отклонения от количества встраиваемой информации

Также была исследована зависимость размера встраиваемого сообщения от размера контейнера при выполнении условия сохранения минимального показателя степени отклонения от исходного сигнала. На рис. 5 представлены результаты исследования этой зависимости.

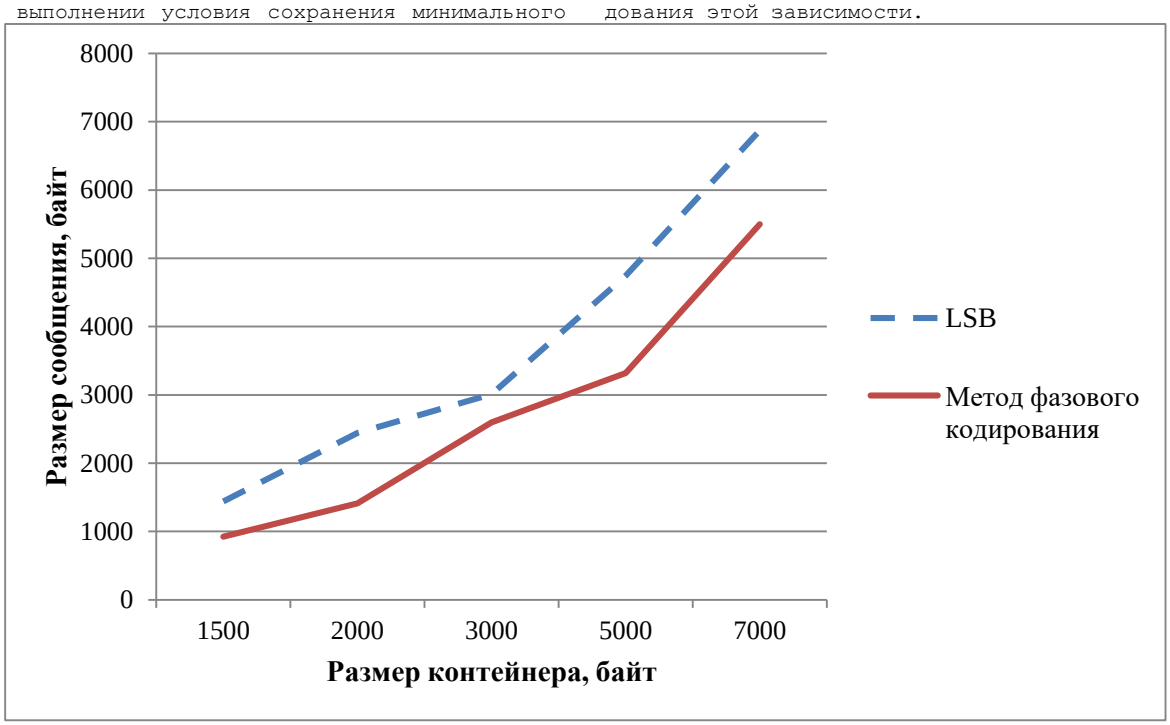

Рисунок 5 – График зависимости размера встраиваемого сообщения от размера контейнера

Исследования показывают, что, если объем встраиваемых данных при использовании метода LSB составляет не менее 10% от объема файла контейнера, то при первичном анализе можно обнаружить факт встраивания. При наиболее оптимальных параметрах (менее 10% от объема файла контейнера) искажения исходного аудифайла незначительны и вероятность обнаружения факта встраивания ниже. Метод фазового кодирования в меньшей степени зависит от объема встраиваемых данных и, в отличие от метода наименее значащих битов, является более стойким к попыткам обнаружения факта встраивания информации в контейнер.

Выводы и результаты

Разработана автоматизированная система, в которой реализованы следующие методы стеганографии: метод наименее значащих битов, метод фазового кодирования. В системе реализовано встраивание текстовой информации в аудиофайлы. Также визуализирован механизм извлечения встроенного сообщения.

Определены оптимальный объем данных, встраиваемых в аудиофайл, и степень отклонения от исходного файла для решения задачи эффективного встраивания.

Произведен сравнительный анализ результатов решения задачи встраивания при использовании двух различных методов.

ЛИТЕРАТУРА

1. Генне О. В. Основные положения стеганографии // журнал "Защита информации. Конфидент", №3, 2000.

2. Нечта И. В. Стеганография в файлах формата Portable Executable // Вестник СибГУТИ. 2009. № 1. С. 85 – 89.

3. Freeware program of steganography bmp, wav, voc. [Электронный ресурс] http://www.hein z-repp.onlinehome.de/Hide4PGP.htm.

УДК 004.932.2

Мишенев В.С., Кудрина М.А.

ФГАОУ ВО «Самарский национальный исследовательский университет имени академика С.П. Королева», Самара, Россия

ИСПОЛЬЗОВАНИЕ ВОЛНОВОГО АЛГОРИТМА ДЛЯ ПОСТРОЕНИЯ СКЕЛЕТА РАСТРОВОГО ИЗОБРАЖЕНИЯ

Ряд задач компьютерной графики требует построения скелета растрового изображения. Для этого существуют различные методы. В данной статье рассмотрен волновой метод скелетизации. В основе алгоритма лежит идея запуска сферической волны внутри объекта и дальнейшего отслеживания центральных пикселей каждой генерации волны. В результате получается векторное представление изображения в виде графа. Конечным этапом алгоритма является оптимизация полученного графа. Преимуществами алгоритма являются высокая скорость работы и малые затраты памяти. Разработана программа, реализующая пороговую бинаризацию изображения, построение скелетного графа по волновому алгоритму и последующую его оптимизацию методами последовательных приближений и ортогональной регрессии

Ключевые слова:

скелетизация растровых изображений, волновой алгоритм, алгоритм утончения линий, векторизация бинарных изображений

Введение

Для таких задач, как структурное распознавание рукописного текста, сравнение отпечатков пальцев, обработка медицинских изображений, картографических изображений и технических чертежей, необходимо строить скелет растрового изображения. Например, используя скелетное представление изображения, содержащего рукописный текст, можно на основе топологического кода построить нейронную сеть для распознавания символов [1]. Такая комбинация различных методов приводит к хорошим итоговым результатам распознавания.

На непрерывной плоскости скелетом называется множество точек, равноудаленных от границ изображений. Этапу построения скелета фигуры как правило предшествуют несколько вспомогательных этапов. Это предварительная обработка (устранение мелких шумов и пр.) и бинаризация изображения. Бинаризацией называется процесс преобразо-