МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

**«КРЫМСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ имени В. И. ВЕРНАДСКОГО»**

(ФГАОУ ВО «КФУ им. В. И. Вернадского»)

**Таврическая академия (структурное подразделение)**

**Факультет математики и информатики**

**Кафедра информатики**

Иванча Николай Александрович

**Разработка элементов дополненной реальности для визуализации туристических объектов Крыма**

Выпускная квалификационная работа

Обучающегося 4 курса

Направления подготовки 01.03.02. Прикладная математика и информатика

Форма обучения очная

Научный руководитель

Доцент кафедры информатики,

Кандидат физико-математических наук М. Г. Козлова

К ЗАЩИТЕ ДОПУСКАЮ:

Заведующий кафедрой,

доктор физико-математических наук,

профессор В. И. Донской

Симферополь, 2019

**Аннотация**

**Иванча Н. А. Разработка объектов дополненной реальности для визуализации туристических объектов Крыма.** Выпускная квалификационная работа бакалавра по направлению 01.03.02. Прикладная математика и информатика. Таврическая академия Крымского федерального университета имени В. И. Вернадского.

Целью работы является создание приложения дополненной реальности, использую оптимальное программное обеспечение. Содержание работы состоит в исследовании, тестировании и разборе современных технологий, позволяющих работать с дополненной реальностью. Результатом работы является создание мобильного приложения.

Ключевые слова: дополненная реальность, AR, распознавание, текст, оптические системы распознавания, OCR, бинаризация, макет, маркер, target, объект, модель, моделирование, редактор, программное обеспечение, среда разработки, предобработка, шум, искажение, изображение, методы, классификация, шаг, сегментация, обучение, анализ, приложение, визуализация, город, Крым.

Страниц – 55, таблиц – 0, иллюстраций – 24, приложений – 1, библиографических источников – 37.

Содержание

[Введение 4](#_Toc10713846)

[1. Подходы к распознаванию текста 7](#_Toc10713847)

[1.1. Оптическое распознавание символов 7](#_Toc10713848)

[1.2. Метод распознавания текста №1 11](#_Toc10713849)

[1.3. Метод распознавания текста №2 13](#_Toc10713850)

[1.4. Метод распознавания текста №3 14](#_Toc10713851)

[1.5. Распознавание текста по структуре скелета букв 16](#_Toc10713852)

[1.6. Методы бинаризации изображений 19](#_Toc10713853)

[1.7. Tesseract OCR 21](#_Toc10713854)

[2. Подходы к работе с объектами дополненной реальности 29](#_Toc10713855)

[2.1. Vuforia 32](#_Toc10713856)

[2.2. ARToolKit 34](#_Toc10713857)

[2.3. Подходы к разработке объектов дополненной реальности 35](#_Toc10713858)

[2.3.1. Blender 35](#_Toc10713859)

[2.3.2. 3ds Max 36](#_Toc10713860)

[2.3.3. ZBrush 37](#_Toc10713861)

[3. Разработка комплекса визуализации объектов дополненной реальности при посещении городов Крыма 38](#_Toc10713862)

[3.1. Распознавание текста 38](#_Toc10713863)

[3.2. Создание объектов дополненной реальности 38](#_Toc10713864)

[3.3. Визуализация элементов дополненной реальности 38](#_Toc10713865)

[Приложение 48](#_Toc10713866)

[Список источников 53](#_Toc10713867)

# Введение

В последние годы несложно заметить стремительное развитие технологий, цель которых облегчить деятельность человека. Одной из таких технологий является augmented reality (AR) или дополненная реальность. Дополненная реальность – это технология, которая в реальном времени дополняет привычный для человека физический мир, элементами компьютерной графики (картинки, текст и др.), используя при этом специальное программное обеспечение и доступные каждому – смартфоны, планшеты или другие подобные устройства.

Очень важно отличать дополненную реальность от виртуальной (virtual reality, VR) и смешанной (mixed reality, MR). В отличие от дополненной реальности, где виртуальные объекты проецируются на реальное окружение, виртуальная представляет собой новый, созданный техническими средствами мир, который передаётся человеку через органы зрения, слуха и осязания. Смешанная реальность является наиболее сложной, так как объединяет оба подхода.

Дополненная реальность уверенно занимает лидирующие позиции на рынке технологий. Это подтверждает столь частое упоминание AR в презентациях тех или иных IT-гигантов. Такой интерес к дополненной реальности вызван тем, что она предоставляет возможность взаимодействовать с окружением на принципиально новом уровне, а также имеет огромный спектр областей применения.

Основным двигателем AR технологий на данный момент является индустрия развлечений, которая видит в дополненной реальности возможность установления более тесных связей между аудиторией и своими персонажами. Так, например, в конце 2016 года компанией Niantic была выпущена мобильная игра про «карманных монстров» Pokemon GO, которая всего за несколько месяцев побила рекорды популярности среди мобильных приложений и положила начало кардинальным изменениям в индустрии развлечений [1].

Не обошли стороной новые технологии и рекламную индустрию. С появлением дополненной реальности различные компании открыли для себя новые способы донесения привлекательности своего продукта до своих потенциальных потребителей. Дополненную реальность используют для наполнения плоских поверхностей листовок и билбордов интерактивным контентом, который имеет огромное преимущество перед статичной рекламой. Например, известный журнал об автомобилях Top Gear предоставил своим читателям возможность просматривать видеоролики об автомобилях прям на страницах издания, используя для этого только лишь смартфон со специальным приложением. Такой же рекламный ход совершила и компания Faberlic, предоставив своим клиентам возможность посмотреть на косметические товары из своего печатного каталога в натуральную величину, прочитать состав на этикетке и сделать заказ [2].

Несмотря на огромную популярность дополненной реальности в сферах рекламы и развлечения, потенциал использования AR намного шире, что наглядно подтверждают современные примеры применения технологии в медицине. Первый самый очевидный пример использования AR в медицине – это организация максимально наглядного обучения студентов медицинских вузов. Наглядную анатомию в дополненной реальности способно продемонстрировать приложение HoloAnatomy, разработанное компанией Microsoft для очков смешанной реальности HoloLens [3].

Более серьёзными с точки зрения медицины являются приложения: MITK pille, которое позволяет, используя планшет, прямо во время операции «заглянуть» в тело пациента, и приложение для людей с ограниченными возможностями зрения и слуха Aira, в котором реализован нейросетевой помощник, распознающий и проговаривающий всё, что видит камера очков [4].

Своё применение дополненная реальность нашла и в сфере дизайна и проектирования. Так, например, компания Ikea интегрировала AR в своё приложение Ikea Place, с помощью которого позволяет пользователям проверить, как мебель может выглядеть в их домах [5].

Также дополненная реальность оказалась востребована в сфере туризма. Уже существуют приложения, помогающие людям, оказавшимся в незнакомом городе, успешно ориентироваться на местности и прокладывать маршруты к необходимым локациям. Примером такого приложения является разработка компании MapWay – приложение Bus Times London, которое, используя камеру смартфона, помогает туристам найти свою автобусную остановку в пределах города Лондон, рисуя средствами дополненной реальности путь до неё [9].

При посещении музеев, памятников архитектуры или других достопримечательностей средствами дополненной реальности могут быть реализованы такие опции как виртуальный экскурсовод или визуализация каких-либо событий, связанных с данной достопримечательностью.

Технология дополненной реальности востребована в столь многих сферах деятельности человека по следующим причинам:

1. Возможность получать информацию в реальном времени.
2. Необычный способ представления информации, благодаря которому возможно привлечь больше внимания окружающих и более точно донести до них необходимую информацию.
3. Усиливает запоминание деталей и понимание процесса за счёт наглядности. Актуально для сферы образования.
4. Инновационность. Несмотря на уверенное положение на рынке технологий, воспринимается как нечто новое и современное.
5. Большой, вечно расширяющийся, спектр сфер использования.

Учитывая высокую популярность AR технологий, отличным решением будет внедрение элементов дополненной реальности в самую востребованную отрасль крымского полуострова – туризм. С помощью дополненной реальности можно визуализировать достопримечательности Крыма на физической карте, с целью привлечения большего числа туристов с разных уголков земного шара.

Целью работы является разработка программного обеспечения реализующего визуализацию туристических объектов Крыма в дополненной реальности.

Задачи, решаемые в работе:

1. Обзор научных источников по данному вопросу.
2. Изучение необходимых программных средств и библиотек.
3. Создание 3D-моделей достопримечательностей.
4. Реализация программного продукта для распознавания названий городов на изображении (карте).
5. Реализация программного продукта для визуализации 3D-моделей достопримечательностей.

# Распознавание текста

Задача распознавания текста особенно актуальна в наше время. Существует огромное количество печатных изданий, играющих важнейшие роли в развитии человечества, которые находятся далеко не в самом лучшем состоянии. И несмотря на то, что многим хотелось бы сохранить их первозданный вид, они всё же нуждаются в оцифровке во избежание потери информации. Для решения этой задачи придётся применить технологии распознавание текста.

Распознавание текста – это один из важнейших разделов распознавания образов с широчайшей сферой применения. Во-первых, это, упомянутая ранее, оцифровка. Помимо старинных печатных изданий, которые отцифровывают для сохранения культурного наследия, данная технология имеет большой спрос и для работы с современной печатной продукцией.

Программы распознавания текста невероятно востребованы среди представителей современного бизнеса. Главным образом они служат для автоматизации ввода и обработки данных из документов, за счёт чего помогают экономить время и деньги. Текст отсканированного документа, его фотографию или PDF-файл можно просматривать с экрана компьютера, но их содержимое нельзя копировать и изменять. Благодаря технологии распознавания текста исходное изображение может быть переведено в формат, доступный для редактирования. В качестве примера программного обеспечения, реализующего данную технологию, можно привести систему оптического распознавания текста ABBYY OCR, которая занимает лидирующие позиции в бизнес сфере [6].

Технология распознавания текста также используется для перевода текста с картинки с одного языка на другой. Основная область применения таких приложений – сфера туризма. Распознавание текста в данном случае используется для мгновенного перевода географических наименований, которые могут встретиться туристу в незнакомой для него стране. Примером такого использования технологии распознавания текста является приложение для мобильных устройств Google Переводчик. Данное приложение поддерживает мгновенный перевод любых надписей с 38 языков на русский, а также поддерживает распознавание рукописного ввода с 93 языков, что тоже является результатом работы программной части, отвечающей за распознавание текста [7].

Следует отметить, что процесс распознавания текста может быть реализован при помощи четырёх основных методов:

1. сравнение с подготовленными шаблонами;
2. распознавание, которое использует критерии распознаваемого объекта;
3. распознавание, основанное на структуре объекта;
4. распознавание при помощи нейронных сетей, самообучающихся алгоритмов.

## Оптическое распознавание символов

Оптическое распознавание символов (optical character recognition или OCR) – это электронное или механическое преобразование изображений машинописного, рукописного или печатного текста в машинно-кодированный текст, будь то отсканированный документ, фотография документа, фотография сцены (например, текст на вывесках и рекламных щитах в альбомной фотографии) или из текста субтитров, наложенного на изображение (например, из телевизионной трансляции).

Процесс оптического распознавания символов можно разделить на следующие этапы:

1. Предобработка

Оптические системы распознавания текста часто предварительно обрабатывают изображения, чтобы повысить шансы на успешное распознавание. К таким методам можно отнести следующие методы: выравнивание (de-skew) – если документ не был выровнен должным образом при сканировании, может возникнуть необходимость наклонить его на несколько градусов по часовой стрелке или против часовой стрелки, чтобы сделать линии текста идеально горизонтальными или вертикальными; удаление шумов (despeckle) – удаление положительных и отрицательных пятен, сглаживание краев; бинаризация (binarisation) – преобразование изображения из цветного в черно-белое с целью отделения текста (или любого другого желаемого компонента изображения) от фона; удаление линий (line removal) – удаление нерельефных частей изображения; анализ макета или «зонирование» – определение столбцов, абзацев, подписей и прочих элементов как отдельные блоки; обнаружение линий и слов (line and word detection) – устанавливает базовую линию для слов и символов, при необходимости разделяет слова; изоляция символов или «сегментация» (character isolation or «segmentation») – разделение изображения на сегменты с одним или несколькими символами на каждом, распознавание символов и дальнейшее соединение их; нормализация соотношения сторон и масштаба (normalize aspect ratio and scale).

1. Распознавание текста

Существует два основных типа алгоритма распознавания текста. Матричное сопоставление включает в себя сравнение изображения с сохраненным шаблоном, сравнивая пиксели; оно также известно, как «сопоставление шаблонов», «распознавание образов» или «корреляция изображений». Это зависит от того, насколько входной символ правильно изолирован от остальной части изображения, а сохранённый шаблон имеет такой же шрифт и масштаб. Этот метод лучше всего работает с машинописным текстом и плохо работает, когда встречаются новые шрифты.

Извлечение объектов разлагает символы на «объекты», такие как линии, замкнутые контуры и «перекрёстки». Выделение признаков уменьшает размерность представления и делает процесс распознавания вычислительно эффективным. Эти характеристики сравниваются с абстрактным векторным представлением символа, которое может быть сведено к одному или нескольким прототипам шаблона. Общие методы обнаружения признаков в компьютерном зрении применимы к данному типу распознавания, который обычно наблюдается в «интеллектуальном» распознавании почерка и в большинстве других современных программах распознавания.

Такие программы, как Cuneiform и Tesseract, используют двухпроходный подход к распознаванию символов. Второй проход известен как «адаптивное распознавание» и использует формы букв, распознанные с высокой степенью уверенности на первом проходе, чтобы лучше распознавать оставшиеся буквы на втором проходе. Это выгодно для необычных шрифтов или низкокачественных сканов, где шрифт искажён (например, размытый или выцветший).

Современное программное обеспечение для распознавания символов, например OCRopus или Tesseract, использует нейронные сети, которые были обучены распознавать целые строки текста вместо того, чтобы фокусироваться на отдельных символах.

1. Постобработка

Точность распознавания может быть увеличена, если выходные данные ограничены лексиконом – списком слов, которые могут встречаться в документе. Это могут быть, например, все слова в русском языке или более технический лексикон для конкретной области. Этот приём может быть проблематичным, если документ содержит слова, отсутствующие в лексиконе, например имена собственные. Tesseract использует свой словарь, чтобы повлиять на шаг сегментации символов, для повышения точности. Выходной поток может представлять собой обычный текстовый поток или файл символов, но более сложные системы распознавания текста могут сохранять исходный макет страницы и создавать, например, аннотированный PDF-файл, включающий как исходное изображение страницы, так и текстовое представление с возможностью поиска.

«Анализ ближайшего соседа» может использовать частоты встречаемости для исправления ошибок, отмечая, что некоторые слова часто встречаются вместе. Например, «Вашингтон, округ Колумбия» обычно гораздо чаще встречается в английском языке, чем «Вашингтон DOC». Знание грамматики сканируемого языка также может помочь определить, является ли слово глаголом или существительным, например, что позволяет повысить точность.

1. Оптимизация для конкретных приложений

Всё больше и больше поставщиков технологий распознавания начали настраивать системы распознавания таким образом, чтобы более эффективно работать с конкретными типами входных данных. Помимо специфичного для конкретного приложения лексикона, более высокая производительность может быть достигнута за счёт учёта бизнес-правил, стандартных выражений или богатой информации, содержащейся в цветных изображениях. Эта стратегия называется «Ориентированность на приложение распознавания» или «Индивидуально-настраиваемая система распознавания» и применяется для распознавания номерных знаков, счетов, скриншотов, удостоверений личности, водительских удостоверений и производства автомобилей.

The New York Times адаптировала технологию OCR в собственный инструмент, который они называют Document Helper, что позволяет их интерактивной команде новостей ускорить обработку документов, которые необходимо просмотреть. Они отмечают, что это позволяет им обрабатывать до 5400 страниц в час, обрабатывая материал так, чтобы репортёры могли ознакомиться с его кратким содержанием [8].

## Факторы влияющие на точность распознавания

На точность распознавания символов влияет большое количество факторов, многие из которых могут быть устранены в процессе предобработки. Следующие факторы оказывают наибольшее влияние на результат распознавания:

1. Лингвистическое разнообразие. Многие языки обладают одинаковым буквами: «O», «А», «Р» и др. или похожими: «N» и «И», «R» и «Я» и др. В каждом языке также существуют буквы похожие между собой: «М», «Н», «И» – в русском, «I», «L», «J» – в английском. Некоторые буквы имеют сходства с цифрами «О» и «0», «I» и «1» и т.д.
2. Искажение изображения. К возможным искажениям изображения можно отнести описанные ранее: нарушение перспективы, соотношения сторон и т.д.
3. Защита источника изображения. Часто встречается на документах с персональными данными. В случае паспортов – это невидимые символы.

Разнообразные вариации размеров, шрифтов и масштабов символов. Каждый отдельный символ может быть написан с использованием различных шрифтов, как стандартных (Times, Orator, Arial), а также множеством шрифтов, используемых, например, в пределах одной страны, для оформления одних и тех же документов.

1. Эффекты освещения (тени, блики и т. п.) при съёмке камерой. Многие системы распознавания символов показывают наилучшие результаты при обработке чёрного текста на светлом (в идеале – белом) фоне. В случае текстовых документов – это не критично. Но если применять такую систему для обработки нестандартных изображений: сканы паспортов, водительских удостоверений, номерных знаков автомобилей, то мы столкнёмся с невозможностью получить корректные данные без предварительной обработки изображения [10].

## Методы распознавания символов

Технология распознавание различных печатных символов с изображений находит своё применение среди множества научных и прикладных задач, которые нацелены на идентификацию всевозможных объектов в реальном мире. К числу задач, решаемых с помощью OCR, можно отнести не только класс задач по распознаванию текста, но и ряд других специализированных задач, ориентированных на распознавание символической информации, нанесённой на поверхность различных объектов. В зависимости от класса задачи применяют один из следующих методов распознавания символов.

### Шаблонные методы

Шаблонные или растровые методы распознавания текста основаны на сравнении исходного изображения с заранее подготовленными шаблонами, используя некую метрику. Метрика –  некоторое условное значение функции, определяющее положение объекта в пространстве. В случае, если два объекта расположены близко друг к другу, то метрики для таких объектов будут совпадать или быть предельно похожими, это будет означать, что два объекта представляют, например, одну и ту же букву, написанную разными шрифтами. В качестве метрики можно выбрать расстояние Хэмминга.

Данную метрику часто используют при кодировании информации и передаче данных, она показывает, как сильно объекты не похожи между собой. Например, после сеанса передачи на выходе имеется следующая последовательность бит (1000001), также нам известно, что должна прийти другая последовательность бит (1000101). Мы вычисляем метрику путём сравнения частей последовательности с соответствующими местами из другой последовательности. Таким образом метрика Хэмминга в нашем случае равна 3. Так как объекты отличаются в трёх позициях. 3- это степень непохожести, чем больше, тем хуже в нашем случае.

Следовательно, чтобы определить какая буква изображена нужно найти её метрику со всеми готовыми шаблонами. И тот шаблон, чья метрика окажется наименьшей будет ответом.

Но реализовав систему распознавания используя расстоянии Хэмминга можно прийти к выводу, что подсчёт одной лишь метрики не даёт положительного результата, так как многие буквы похожи между собой. Например «j» и «i» в английском алфавите, их метрика близка к нулю, что не гарантирует однозначности распознавания и приводит к ошибочному результату.

Решением данной проблемы является отличная от расстояния Хэмминга метрика, которая позволяет разграничить некоторое множество букв в отдельный класс. Примером такой метрики является метрика, основанная на горизонтальном и вертикальном отражении объектов и преобладании одного из отображений.

Такие буквы как «H» «I» «i» «O» «o» «X» «x» «l» обладают суперсимметрией (полностью совпадают со своими отражениями и значимые пиксели распределены равномерно по всему изображению), они выносятся в отдельный класс, что сокращает перебор всех метрик примерно в 6 раз. При проведении аналогичных действий в отношении других букв можно достигнуть уменьшения перебора примерно в 3 раза. Особой уникальностью обладает буква «J», которая находится в своём классе одна, и, следовательно, идентифицируются однозначно.

Следующем шагом может быть использована метрика Хэмминга, которая на данном этапе даёт лучшие показатели чем при прямом применении [11].

Реализовав систему распознавания текста на основе данного метода мы пришли к тому, что из-за огромного количества различных печатных изданий и используемых в них разнообразных модификаций шрифтов рассмотреть все случаи на этапе обучения невозможно.

К достоинствам данного метода можно отнести: простоту программной реализации, стабильную работу в условиях отсутствия помех, а также высокую точность распознавания символов, имеющих различные дефекты.

Но также данный метод имеет существенные недостатки, в числе которых: сильная «шрифтозависимость», сильная зависимость от качества предварительно подобранных шаблонов, медленная работа при большом количестве помех, а также резкое ухудшение точности распознавания при работе с повёрнутыми, наклонными и другими изображениями, имеющими различные искажения.

### Признаковые методы

Признаковые методы являются наиболее распространёнными [12], признак в них определяется как функция от значений, содержащихся в одном или более пикселях, и вычисляется так, что численно выражает некоторую значимую характеристику объекта.

Все признаки присущие символам можно классифицировать следующим образом:

* признаки пиксельного уровня: признаки, вычисляемые в каждом пикселе, такие как цвет, положение;
* локальные признаки: признаки, вычисляемые в некотором окне или ограниченной области изображения;
* глобальные признаки: признаки, вычисляемые по всему изображению. Обычно, это статистические свойства изображений, например, гистограмма, среднее значение, дисперсия и другие статистические моменты.

С другой стороны, все признаки могут быть условно разделены на низкоуровневые и высокоуровневые признаки. Низкоуровневые признаки могут быть извлечены непосредственно из исходного изображения, тогда как высокоуровневые признаки базируются на низкоуровневых признаках.

Множество подлежащих обработке символов разбивается на конечное число классов, называемых образами. Объекты каждого класса (образы) характеризуются некоторой совокупностью признаков, каждый из которых принимает одно из определённого множества значений. Пространство признаков разбивается на непересекающиеся области, каждая из которых соответствует определённому классу символов. Для предъявленного к распознаванию изображения символа вычисляются все признаки, значения которых задают положение некоторой точки в пространстве признаков. После этого определяют, в какую из областей пространства признаков попала данная точка, и исходный символ относят к классу символов, связанному с данной областью [13, 16].

Исходя из этого, к достоинствам данного метода можно отнести высокую обобщающую способность, а также устойчивость к изменению формы символов, что позволит распознавать с помощью данного метода различные шрифты и начертания.

В качестве недостатков можно выделить неустойчивость к различным дефектам изображения и потерю информации о символе на этапе извлечения признаков

### Структурные методы

Структурные методы переводят символ в его топологическое представление, отражающее информацию о взаимном расположении структурных элементов. В качестве основных структурных представлений изображений используются детекторы линий, полос, пятен или углов, находящихся в соединениях отрезков вида «Г», «T», «Y» и «X» [15, 16].

Также для построения описания изображений символов используется контурное представление. Построение структурных элементов на основе контуров может выполняться путём сегментации контуров, их аппроксимации или обнаружения точек максимальной кривизны. К контурным структурным элементам относятся: отрезки прямых линий, дуги окружностей или эллипсы, углы и т.д. [14, 16].

При проведении структурного сравнения изображений производится поиск соответствий между структурными элементами с учётом их взаимного расположения, типа, размеров, взаимной ориентации и т.д. Таким образом, в структурных методах вместо поиска в множестве параметров пространственного преобразования производится поиск в множестве вариантов соответствий структурных элементов [16].

Среди достоинств данного метода можно выделить способность распознавания искажённых символов и хорошие показатели точности при работе с различными типами шрифтов.

К недостаткам же стоит отнести неустойчивость к различным дефектам изображения, таким как «слипание» символов или «разрыв» структурных элементов одного символа [10].

### Нейросетевые методы

Нейросетевые методы основаны на применении различных типов искусственных нейронных сетей. Идея этих методов – моделирование работы мозга человека. Нейрон, как структурная единица нейронной сети, получает множество входных сигналов. В нашем случае входные сигналы описывают значение пикселя изображения, то есть, если имеется изображение 16х16, входных сигналов у нейрона должно быть 256. Каждый входной сигнал воспринимается с определенным коэффициентом и в результате, по окончанию распознавания на каждом нейроне скапливается определённый заряд, у какого нейрона заряд будет больше тот и испустит импульс. А так как все нейроны поименованы значениями букв, следовательно, среагировавший нейрон и несёт ответ распознавания.

Обучение нейронных сетей происходит на множестве обучающих примеров. Из вариантов обучения можно использовать обучение с учителем (персептрон) или самоорганизацию (сеть Кохонена). Процесс обучения можно описать следующим образом. Берётся очередное изображение из обучающей выборки и отправляется для обработки в построенную нейронную сеть. Сеть анализирует все позиции чёрных пикселей (цвет символов) и выравнивает коэффициенты минимизируя ошибку совпадения методом градиента, после чего определённому нейрону сопоставляется данное изображение.

Основными преимуществами данного метода являются способность к обобщению и высокая скорость работы.

Но также этому методу присущи и существенные недостатки: чувствительность к вращению и искажению символов, высокая зависимость от обучающей выборки и алгоритма обучения [10, 11].

## Методы бинаризации изображений

Большинство OCR использует бинаризацию изображений (выделение текста из фона) как первый шаг в процессе распознавания. Выполнив бинаризацию изображения на этапе предобработки и передав результат к готовому модулю распознавания текста, можно значительно повысить точность относительно прямого распознавания.

Процесс бинаризации представляет собой перевод цветного изображения или изображения в градациях серого в двухцветное – черно-белое. Основной параметр данного преобразования – порог, со значением которого затем сравнивается яркость всех пикселей. После сравнения, пикселю присваивается одно из двух возможных значений: 0 – «граница объекта» или 1 – «остальная область» [18].

Общая схема бинаризации представлена на рисунке 1.4.1.

Рис. 1.4.1. Общая схема бинаризации изображения

В зависимости от порога методы бинаризации можно условно разделить на две группы: первая группа использует фиксированный порог для заданного изображения (глобальная бинаризация), тогда как вторая группа использует локальные пороги (локальная бинаризация).

Глобальная бинаризация хорошо применима в случаях, когда текст занимает большую часть изображения и хорошо контрастирует с фоном. При реализации такого метода находится порог бинаризации , с помощью которого происходит деление на чёрное и белое, причём величина порога  остаётся неизменной в течение всего процесса бинаризации. К глобальным методам бинаризации относятся:

* бинаризация с нижним порогом;
* бинаризации с верхним порогом;
* бинаризация с двойным ограничением;
* неполная пороговая обработка;
* многоуровневое пороговое преобразование [17].

Метод бинаризации изображения с нижним порогом является одним из самых простых, так как в нём рассматривается только одно значение порога. Пусть  – это пиксель исходного изображения, тогда  – яркость пикселя , а  – функция переводящая пиксель в бинарное представление. В случае для бинаризации с нижним порогом справедлива следующая формула:

1.4.1

Если в приведённой выше формуле для точки изображения выполняется первое условие, то такая точка является точкой объекта, если же выполняется второе условие, то точка будет точкой фона.

Метод бинаризации с верхним порогом можно считать частным случаем бинаризации с нижним порогом, при котором получается негатив исходного изображения:

Если необходимо выделить определённые области, значения яркости пикселей в которых могут изменяться в некотором диапазоне, то применяется метод бинаризации с двойным ограничением, который является объединением двух предыдущих методов:

С другой стороны, методы локальной бинаризации

может лучше справляться с неравномерным освещением и изменением цвета текста,

все же они более чувствительны к выбору параметров

# Приложение

# Список использованных источников

* + - 1. Мир после Pokemon Go: как технология дополненной реальности изменит нашу жизнь / А. Сошников. – URL: https://www.bbc.com/russian/features-36898771 (дата обращения: 13.04.2020). [1]
      2. Дополненная реальность в маркетинге: 15+ примеров использования / Е. Самородских ; TexTerra – агентство комплексного интернет- маркетига. – URL:

<https://texterra.ru/blog/dopolnennaya-realnost-v-marketinge-primery-ispolzovaniya.html> (дата обращения: 13.04.2020). [2]

* + - 1. HoloAnatomy : программный продукт / Microsoft [сайт]. – URL:

<https://www.microsoft.com/ru-ru/p/holoanatomy/9nblggh4ntd3/> (дата обращения: 13.04.2020). [3]

* + - 1. VR и AR в медицине / Яндекс Дзен [сайт]. – URL:

<https://zen.yandex.ru/media/id/5ad5fbd76104934709d6f592/vr-i-ar-v-medicine-5b17aa5d83090530bc94c7c4/> (дата обращения: 13.04.2020).

* + - 1. Say hej to IKEA Place / Ikea [сайт] – URL: <https://www.ikea.com/au/en/customer-service/mobile-apps/say-hej-to-ikea-place-pub1f8af050> (дата обращения: 13.04.2020).
      2. Распознавание текста с помощью решений ABBYY – все гениальное просто для бизнеса / Комсомольская правда [сайт].– URL: <https://www.kp.ru/guide/raspoznavanie-teksta.html> (дата обращения: 13.04.2020).
      3. Google Переводчик / Google Play [сайт]. – URL:

<https://play.google.com/store/apps/details?id=com.google.android.apps.translate> (дата обращения: 13.04.2020).

* + - 1. Optical character recognition / Wikipedia [сайт]. – URL: [https://en.wikipedia.org/wiki/Optical\_character\_recognition/](https://en.wikipedia.org/wiki/Optical_character_recognition#Techniques) (дата обращения: 13.04.2020).
      2. Bus Times London – TfL timetable and travel info : программный продукт / Google Play [сайт]. – URL:

<https://play.google.com/store/apps/details?id=com.mxdata.buslondon&hl=en/> (дата обращения: 13.04.2020).

* + - 1. Оптическое распознавание символов / Studbooks [сайт]. – URL: <https://studbooks.net/2016242/informatika/osnovnye_trudnosti_raspoznavaniya_simvolov/> (дата обращения: 13.04.2020).
      2. Методы распознавания текста / Habr [сайт]. – URL: <https://habr.com/ru/post/220077/> (дата обращения: 14.04.2020).
      3. Логический подход к искусственному интеллекту / под ред. Г. П. Гаврилова – Москва : Мир, 1998 – 493 с.
      4. Фор, А. Восприятие и распознавание образов. / А. Фор. – Москва : Машиностроение, 1989. – 272 с.
      5. Эндрю, А. Искусственный интеллект / Под ред. Поспелова Д.А. – Москва : Мир, 1985. – 265 с.
      6. Фу, К. Структурные методы в распознавании образов / К. Фу. – Москва : Мир, 1977. – 320 с.
      7. Тропченко, А. Ю. Методы вторичной обработки и распознавания изображений : учебное пособие / А. Ю. Тропченко, А. А. Тропченко. – Санкт-Петербург : Университет ИТМО, 2015. – С. 60–89.
      8. Ntirogiannis, K. An objective evaluation methodology for document image binarization techniques in DAS / K. Ntirogiannis, B. Gatos, I. Pratikakis. – 2008. P. 217–224.
      9. Исрафилов, Х. С. Исследование методов бинаризации изображений : статья / Х. С. Исрафилов – Москва : МГТУ им. Н.Э. Баумана, 2017. – 7 с.