

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ
компьютерной безопасности и
криптографии

**Технология улучшения изображений для их анализа методами машинного
обучения**

АВТОРЕФЕРАТ

дипломной работы

студента 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Шляхова Андрея Владимировича

Научный руководитель

профессор, д. ф.-м. н.

В. А. Молчанов

23.01.2021 г.

Заведующий кафедрой

д. ф.-м. н., доцент

М. Б. Абросимов

23.01.2021 г.

Саратов 2021

ВВЕДЕНИЕ

В настоящее время компьютерное зрение является важнейшей технологией, применяемой во многих сферах человеческой жизни. Если говорить про потребительскую электронику, компьютерное зрение применяется в мобильных телефонах, в системах управления беспилотными автомобилями, системах дополненной реальности и видеонаблюдения. Широкое применение системы компьютерного зрения находят в медицине и промышленности. Они используются в системах контроля качества, управления процессами, моделирования объектов, анализа медицинских изображений и так далее. Для создания таких систем необходимо применение различных математических методов анализа изображений или видео данных.

Целью данной работы является рассмотрение и программная реализация различных методов улучшения изображений для их анализа, распознавания и классификации.

Задачи работы:

1. Рассмотреть проблему распознавания образов и подходы к ее решению методами машинного обучения.

2. Изучить методы улучшения изображений, такие как:

- Методы сглаживания изображений для снижения влияния шумов на конечный результат. В частности, будут рассмотрены методы размытия Гаусса и медианной фильтрации.

- Морфологические операции и их применение для выделения необходимых участков изображения.

- Методы выделения границ изображения. В частности, будут рассмотрены операторы Собеля, Щарра, Робертса и Прюитт.

- Методы порогового преобразования для бинаризации изображений. В частности, будет рассмотрен алгоритм Оцу и Бернсена.

Главной практической задачей работы является реализация программного продукта, который использует все вышеперечисленные методы для распознавания и извлечения информации из машинно-читаемых зон документов, удостоверяющих личность. Данный программный продукт использует возможности библиотеки OpenCV для осуществления необходимых преобразований исходного изображения, и инструмент Google Tesseract OCR для распознавания текста.

Дипломная работа состоит из введения, 6-ти разделов, заключения, списка использованных источников и 1-го приложения. Общий объем работы – 64 страницы, из них 49 страниц – основное содержание, включая 46 рисунков, список использованных источников из 20 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

Первый раздел дипломной работы посвящен рассмотрению понятий теории распознавания образов и машинного обучения применительно к изображениям. Данный раздел содержит шесть подразделов.

Первый подраздел содержит общее описание задач классификации и распознавания образов. Во многих практических задачах необходимо принять некоторое решение относительно содержания изображения или классифицировать присутствующий на нем объект. Классификация объекта основывается на характерных признаках изображения объекта. Классификацию так же называют распознаванием образов. Система распознавания имеет память для хранения объектов распознавания. Представление объектов в памяти может быть встроенным, или сформировано по большому количеству эталонных образцов, или может быть запрограммировано в терминах специфических признаков изображения.

Второй подраздел содержит основные понятия из теории распознавания образов. Рассматриваются следующие основные определения:

- *Идеальный класс* – множество объектов, объединенных общими важными свойствами.
- *Классификация* – это процесс назначения меток объектам согласно некоторому описанию свойств объектов.
- *Классификатор* – это алгоритм, принимающий на вход описание объекта и выдающий в качестве результата метку класса.
- *Класс выбросов* – это класс объектов, которые не удалось отнести ни к одному известному классу в данной системе.
- *Распознавание образов* – это научная дисциплина, целью которой является классификация объектов. Объекты называются *образами*.
- *Прецедент* – образ, правильная классификация которого известна.

- *Признак* – количественное измерение объекта, используемое для классификации образов.

Третий подраздел содержит описание общей схемы системы классификации и ее характерные признаки.

Система классификации в общем виде состоит из следующих частей:

- Датчик/преобразователь – это устройство для восприятия реального физического объекта. Оно генерирует описание этого объекта, которое пригодно для машинной обработки. Например, таким устройством может быть цветная фотокамера.

- Экстрактор характерных признаков – некий программный модуль, который из данных, полученных с датчика, извлекает ключевую информацию, необходимую для классификации. Это достигается за счет применения различных методов фильтрации, удаления шума и преобразований входных данных. Применительно к обработке изображений к таким методам относится применение бинаризации, морфологических преобразований, методов выделения границ объектов на изображении. Другое название экстрактора признаков – *препроцессор*.

- Классификатор – правило отнесения объекта к одному из классов на основании анализа его характерных признаков.

Основными характеристиками системы классификации являются:

- *Коэффициент ошибок* – означает, насколько часто система классификации неправильно определяет класс объекта.

- Скорость – количество объектов, обрабатываемых за единицу времени.

- Стоимость – учет цены программного и аппаратного обеспечения и процесса разработки и тестирования.

Четвертый подраздел содержит формальную постановку задачи классификации. Задача классификации определяется через понятия пространства образов, пространства признаков, индикаторной функции и решающего правила.

Пятый подраздел содержит описание возможностей использования методов машинного обучения для решения задачи классификации.

Данные методы основаны на применении как отдельных, так и целых сетей *искусственных нейронов* – упрощенных моделей нейронов живых организмов.

Математическую модель искусственного нейрона, называемого *персептроном*, предложил Фрэнк Розенблатт в 1957 году. Зачастую отдельные нейроны объединяются в сети, которые называются *искусственными нейронными сетями*. Благодаря своей способности к обучению они широко применяются в системах распознавания образов.

Для решения задачи классификации так же может использоваться *нейронная сеть прямого распространения* – искусственная нейронная сеть, в которой каждый нейрон расположен на некотором уровне (слое).

В шестом подразделе описывается необходимость предварительной обработки изображений с целью повышения их качества, поскольку это напрямую влияет на точность системы распознавания образов. Среди методов улучшения изображений выделяются методы удаления шума, сглаживания, повышения контраста и выделения границ.

Во втором разделе работы рассматриваются методы сглаживания изображений. Цель применения данных методов состоит в удалении посторонних шумов с изображения, что приводит к повышению точности применения как последующих методов обработки изображения, так и методов распознавания образов. Данный раздел состоит из двух подразделов.

В первом подразделе содержится описание гауссова размытия. Данный метод основан на использовании двумерной функции Гаусса. Значения двумерного распределения используются для заполнения матрицы свертки. С помощью нее определяется среднее взвешенное значение у каждого пикселя. Значение текущего пикселя имеет наибольший вес (самое высокое значение распределения Гаусса), а соседние пиксели получают все меньший вес в зависимости от того, насколько далеко они находятся от текущего

оригинального пикселя. Это дает эффект размытости, которая сохраняет границы и края лучше, чем другие аналогичные фильтры размытия.

Во втором подразделе содержится описание медианного фильтра. Это распространенный метод фильтрации изображения, который вычисляет значение каждого пикселя как медиану значений пикселей из его окрестности.

Медианная фильтрация является очень эффективным методом обработки изображений, однако есть сложности с ее реализацией на аппаратном уровне.

В третьем разделе работы рассматриваются морфологические операции. Они позволяют изменять формы изображений объектов, например, устранять разрывы, отдельные точки и штрихи, либо наоборот, соединять разрозненные объекты. Данный раздел состоит из четырех подразделов.

В первом подразделе описываются базовые понятия бинарной морфологии, в частности, понятие структурирующих элементов – небольших бинарных изображений некоторой формы.

Во втором подразделе рассматриваются морфологические операции переноса, наращивания и эрозии. Это основные операции бинарной морфологии, которые применяются как самостоятельно, так и в составе остальных. Операция переноса множества пикселей позволяет перемещать пиксели на изображении. Операция наращивания позволяет увеличивать область объекта на изображении, а операция эрозии, наоборот, уменьшать область объекта на изображении.

В третьем подразделе рассматриваются морфологические операции замыкания и размыкания. Они представляют собой композицию операций наращивания и эрозии. Операция замыкания «заливает» узкие разрывы и щели малой ширины в объектах, удаляет малые дыры в объектах и заполняет промежутки в их контурах. При этом исходные размеры объекта сохраняются. В результате применения операции размыкания к бинарному изображению на нем сглаживаются контуры объектов, ликвидируются выступы небольшой ширины, обрываются узкие перешейки.

В четвертом подразделе рассматриваются морфологические операции top-hat и bothat. Операция top-hat (верх шляпы) позволяет выделять небольшие светлые детали на темном фоне, а операция bothat (низ шляпы) позволяет выделять небольшие темные детали на светлом фоне.

В четвертом разделе работы рассматриваются методы выделения границ объектов на изображении. Выделение границ объектов является одной из важнейших задач в цифровой обработке изображений. Данные методы позволяют извлекать из изображения нужные объекты и производить их дальнейшее распознавание. Раздел состоит из пяти подразделов.

В первом подразделе рассматривается общее описание методов, а именно рассматривается способ вычисления значения градиента яркости для каждого пикселя изображения.

Во втором подразделе рассматривается оператор Собеля; показывается матрица свертки, которая им используется; делается вывод о высокой эффективности применения данного оператора, хотя границы объектов получаются достаточно «толстыми».

В третьем подразделе рассматривается оператор Щарра; показывается матрица свертки, которая им используется; делается вывод о более высокой эффективности данного оператора по сравнению с оператором Собеля.

В четвертом подразделе рассматривается оператор Робертса; показывается матрица свертки, которая им используется; делается вывод о высокой скорости работы данного оператора и о его чувствительности к шуму.

В пятом подразделе рассматривается оператор Прюитт; показывается матрица свертки, которая им используется; показывается возможность многократного применения данного оператора к изображению за счет вращения матрицы свертки.

В пятом разделе работы рассматривается пороговое преобразование изображения, которое позволяет преобразовывать полутоновое изображение в двухцветное. Методы порогового преобразования часто применяются совместно

с морфологическими операциями, что позволяет выделять нужные объекты на изображении и производить их дальнейшую классификацию.

Пороговую обработку можно рассматривать как задачу теории статистических решений, в которой цель состоит в минимизации средней ошибки разбиения пикселей на две или более группы (также называемые *классами*). Данный раздел состоит из двух подразделов.

В первом подразделе рассматривается метод Оцу. Этот метод является одним из наиболее эффективных способов вычисления порога бинаризации. Он оптимален в том смысле, что он максимизирует *межклассовую дисперсию*. Его основная идея заключается в том, что если классы хорошо разделяются по порогу, они должны отличаться значениями яркости своих пикселей, и наоборот, порог, хорошо разграничивающий классы по яркости, будет наилучшим. Все вычисления в методе Оцу выполняются над гистограммой изображения, что так же повышает эффективность, так как гистограмма является легко вычисляемым одномерным массивом.

Во втором подразделе рассматривается метод Бернсена. В данном методе изображение разделяется на квадраты фиксированного размера. Порог для каждого пикселя вычисляется как среднее арифметическое между наибольшим и наименьшим значениями яркости пикселей в пределах данного квадрата. К преимуществам применения данного метода можно отнести высокую скорость работы. Недостатком же является появление сильных паразитных помех при обработке однородных областей, что в некоторых случаях приводит к появлению ложных черных пятен.

В шестом разделе работы рассматривается программная реализация практической части работы. Раздел состоит из трех подразделов.

В первом подразделе содержится общее описание комплекса из двух программ для распознавания машинно-читаемых зон на изображениях документов, а именно: программы по распознаванию машинно-читаемой зоны и телеграм-бота, выступающего в качестве визуальной части.

Программа по распознаванию машинно-читаемой зоны написана на языке Python и представляет собой WEB-сервис, который принимает HTTP POST запросы с изображениями для анализа, и возвращает JSON сообщение с данными, которые удалось извлечь из машинно-читаемой зоны отправленного документа.

Для каждого полученного изображения программа выполняет следующие шаги:

1. Изображение подвергается предварительной обработке методами, описанными в данной работе, в ходе которой выделяется часть изображения, на которой расположена машинно-читаемая зона. Это выполняется с помощью API, предоставляемого библиотекой OpenCV.

2. Из выделенной части извлекается текстовая информация путем применения движка для оптического распознавания текстов Google Tesseract OCR. Для этого используется обученная модель, оптимизированная под распознавание моношириного шрифта машинно-читаемой зоны.

3. Из полученных строк машинно-читаемой зоны извлекается вся закодированная в них информация.

Во втором подразделе приводится описание шагов обработки изображения. Для тестового изображения приводятся результаты его обработки с помощью гауссова размытия, морфологических операций эрозии и bothat, оператора Собеля и метода Оцу.

В третьем подразделе приводится описание телеграм-бота, который представляет собой клиента для программы распознавания машинно-читаемых зон. Телеграм-бот написан на языке Java. Ему можно отправлять фотографии документов, а он в ответном сообщении в человекочитаемом виде присылает извлеченную информацию. Телеграм-бот доступен по имени *@MRZ_recognition_bot*.

ЗАКЛЮЧЕНИЕ

В ходе выполнения работы рассмотрена проблема распознавания образов и подходы к ее решению методами машинного обучения. Так же были рассмотрены различные методы улучшения изображений, такие как:

- Методы сглаживания изображений для снижения влияния шумов на конечный результат. В частности, рассмотрены методы размытия Гаусса и медианной фильтрации.

- Морфологические операции и их применение для выделения необходимых участков изображения.

- Методы выделения границ изображения. В частности, рассмотрены операторы Собеля, Щарра, Робертса и Прюитт.

- Методы порогового преобразования для бинаризации изображений. В частности, рассмотрены методы Оцу и Бернсена.

В практической части работы реализована программа, позволяющая с помощью изученных методов распознавать и извлекать информацию из машинно-читаемых зон документов, удостоверяющих личность, таких как паспорта и ID-карты.

Все поставленные в рамках данной работы задачи выполнены полностью.

Реализованная программа может быть использована для автоматизации деятельности сотрудников различных организаций, которые в рамках выполнения своих профессиональных обязанностей проверяют, анализируют и вносят в специальные формы идентификационные данные клиентов.