МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»

ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

Кафедра теоретической информатики и компьютерных технологий

РЕФЕРАТ

Дисциплина: «Алгоритмы компьютерной графики»

Тема: «Распознавание рукописного текста»

Выполнил студент

2 курса, группы ИУ9-41Б

Ионов Т. Р.

Проверил

старший преподаватель

Вишняков И.Э.

Дата \_\_\_\_\_\_\_\_\_\_\_\_\_

Подпись \_\_\_\_\_\_\_\_\_\_\_\_\_

Москва 2021

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 3](#_Toc70537218)

[1. Основные понятия 4](#_Toc70537219)

[2. Стадии распознания 5](#_Toc70537220)

[2.1 Предобработка изображения 5](#_Toc70537221)

[2.1.1 Устранение шума 5](#_Toc70537222)

[2.1.2 Выравнивание строк 6](#_Toc70537223)

[2.2 Сегментация 7](#_Toc70537224)

[2.3 Нормализация 8](#_Toc70537225)

[2.4 Выделение признаков 8](#_Toc70537226)

[2.5 Классификация 9](#_Toc70537227)

[2.6 Постобработка 10](#_Toc70537228)

[ЗАКЛЮЧЕНИЕ 11](#_Toc70537229)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 12](#_Toc70537230)

# ВВЕДЕНИЕ

С каждым днем все больше достижений в области информатики и, особенно, в машинном обучении и анализе изображений находят свое применение в реальной жизни. Компьютерное зрение, зарекомендовав себя как один из самых эффективных способов анализа видео и изображений, давно интегрировано во множество областей, облегчая или заменяя работу человека. Распознавание текста – одна из ключевых задач компьютерного зрения. Это способность компьютера получать и интерпретировать записи из таких источников, как бумажные документы, фотографии, сенсорные экраны и другие устройства.

Данную задачу можно отнести к разделу распознавания образов. В последние десятилетия, благодаря использованию современных достижений компьютерных технологий, были разработаны новые методы обработки изображений и распознавания образов, благодаря чему стало возможным создание удачных промышленных систем распознавания рукописного текста, как, например, Nebo by MyScript или PDFelement Pro by Iskysoft. Тем не менее, создание приложения в данной области по-прежнему остается творческой задачей и требует дополнительных исследований в связи со специфическими требованиями по разрешению, быстродействию, надежности распознавания и объему памяти, которыми характеризуется каждая конкретная задача. Проблема действительно актуальна, ведь оцифрованный вид рукописных записей позволил бы автоматизировать бизнес-процессы множества компаний, упростив работу человека.

# Основные понятия

Цель компьютерного зрения заключается в формировании полезных выводов относительно объектов и сцен реального мира на основе анализа изображений, полученных с помощью датчиков.

Следует разделить распознавание рукописного текста на два класса: офлайн-распознавание с листа бумаги или другой поверхности с помощью оптического сканирования и онлайн - с помощью взаимодействия с поверхностью экрана компьютера. Представление текста в офлайн-распознавании делится на две подгруппы: печатанный и рукописный.

При онлайн-распознавании текста слова и символы распознаются во время выполнения, и информация о процессе написания знаков хранится в буфере памяти. Самым распространенным способ онлайн-записи является использование пера и планшета. Используя знание о количестве, наборе и скорости штрихов, программа может эффективно и точно распознавать написанный пером текст.

При офлайн-распознавании обрабатывается статическое представление имеющих смысл записей. В обеих подгруппах объектом обработки является изображение, полученное при помощи сканера или камеры. В виду вариативности стиля письма и, зачастую, высокой степенью зашумления, данный тип распознавания является сложной задачей и требует нетривиальных методов, самые успешные из которых относятся к классу алгоритмов машинного обучения. Определение автомобильных номеров и интерпретация типизированных документов с помощью алгоритмов искусственного интеллекта стало самым успешным применением технологий распознавания текста.

Работа в области распознавании рукописного текста (РРТ) представляет из себя самую сложную и интересную задачу на данный момент и вносит огромный вклад в улучшении взаимодействия человека и компьютера. Здесь и далее будем рассматривать офлайн-распознавание рукописного текста, как наиболее интересный объект изучения.

# Стадии распознания

## 2.1 Предобработка изображения

Изображения, подлежащие распознаванию, могут быть зашумлены. Причиной тому могут быть искажения, вносимые в изображение поверхностями, активно отражающими свет (Например: металл, стекло, вода), качество используемой аппаратуры и др. Поэтому необходима *предварительная обработка* — последовательность алгоритмов, обрабатывающая исходное изображения, подготавливающая изображение к следующим этапам.

### 2.1.1 Устранение шума

С этой задачей отлично справляется пороговая бинаризация — процесс отделения фона от объекта, в данном случае — текста (рисунок 1).

Простейшие методы пороговой обработки заменяют каждый пиксель с координатами i j в изображении черным пикселем, если интенсивность изображения меньше некоторой фиксированной постоянной T или белый пиксель, если интенсивность изображения больше этой постоянной.

﻿Самой сложной частью в алгоритмах пороговой обработки является собственно поиск порога: при выборе слишком высокого порога, мы можем потерять значимую информацию, при выборе слишком низкого – получаем некорректные данные. Одним из лучших является метод Отсу.

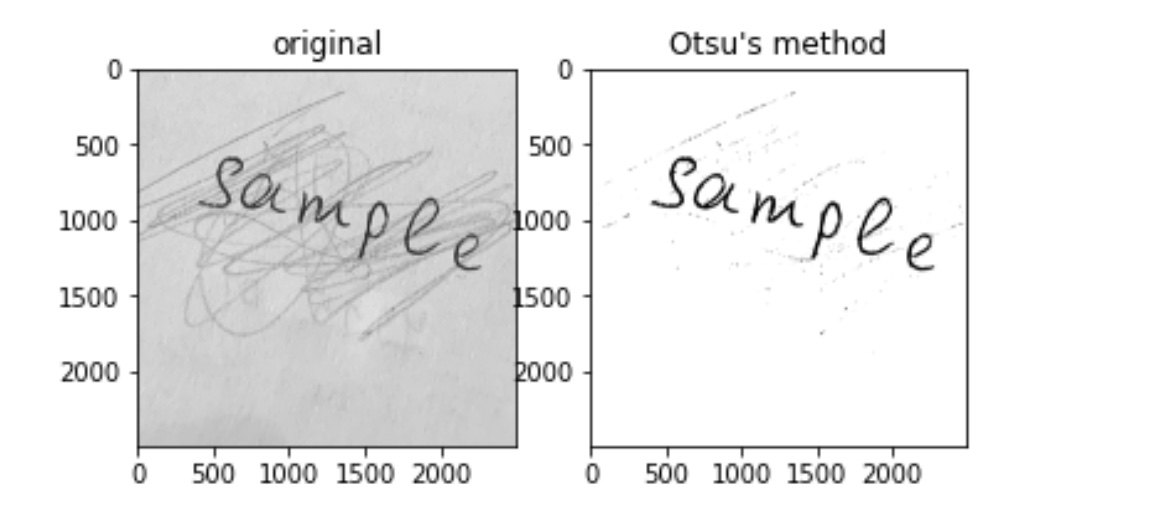
Метод Отсу ищет порог, уменьшающий дисперсию внутри класса, которая определяется как взвешенная сумма дисперсий двух классов:

(1)

где веса – это вероятности двух классов, разделенных порогом t,

.

Отсу показал, что минимизация дисперсии внутри класса приводит к максимизации дисперсии между классами:

которая выражается в терминах вероятности и среднего арифметического класса , которое может обновляться итеративно.

(2)

Рисунок 1

### 2.1.2 Выравнивание строк

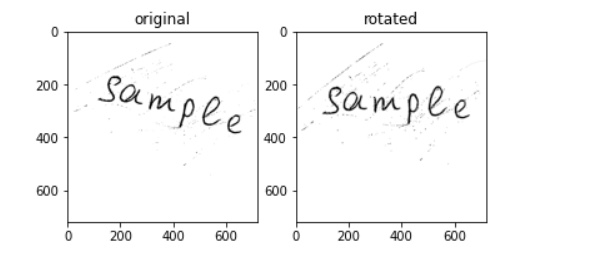
Выравнивание используется для исправления текстовых строк, если они находятся под углом к горизонтали вследствие неточного сканирования или иных факторов (рисунок 2). Широко известен алгоритм выравнивания с помощью преобразования Хафа. Сперва строки размываются, образовывая линии, а затем осуществляется поиск прямых линий с помощью преобразования Хафа. Далее, нужно вычислить угол, на который необходимо повернуть строку, чтобы привести ее к горизонтальному виду.

Рисунок 2

## 2.2 Сегментация

Следующим этапом является сегментация изображения. Сегментация изображения означает присвоение каждому пикселю определенной метки.

На этом этапе сначала с предварительно обработанного изображения извлекаются строки, затем строки разбиваются на слова и, в итоге, каждое слово разбивается на отдельные символы, которые предстоит классифицировать. Текстовый блок разбивается на строки по средствам горизонтального просвета между ними, а слова – вертикального между словами. Эти задачи не представляют трудности.

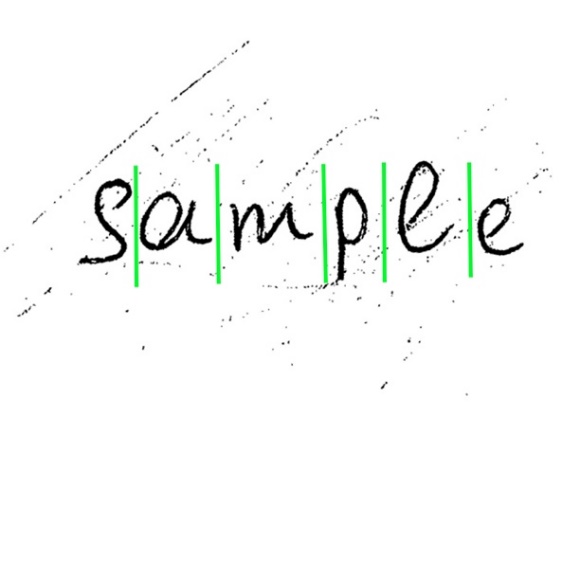
Самой сложный этап возникает при сегментации слова на символы. Существует множество вариантов решения данной проблемы, но наиболее распространенные – нейросетевой модели. Один из возможных вариантов – найти все возможные точки сегментации в слове, а после этого с помощью обученной нейронной сети исключаются некорректные точки сегментации. В последнее время большей популярностью пользуются сверточные нейронные сети (рисунок 3).

Рисунок 3

Рисунок 4.1

## 2.3 Нормализация

После определения символов, полезно их привести к одному стандарту. Основная цель этапа нормализации состоит в устранении различий, которые могли бы усложнить классификацию и снизить скорость распознавания одного и того же символа или слова у разных авторов. В основном параметрами, которые необходимо нормализовать, являются наклон, толщина и размер. Применяя преобразование сдвига, можно привести наклон символа ближе к вертикальному положению (рисунок 4).

Рисунок 4

## 2.4 Выделение признаков

Выделение признаков — это процесс снижения размерности, в котором исходный набор исходных данных сокращается до более управляемых групп для дальнейшей обработки, оставаясь при этом достаточным набором для точного и полного описания исходного набора данных. Признаки играют важную роль, их главная цель – повышение скорости распознавания за счет оптимизированного представления информации об объекте классификации. Возможны различные алгоритмы выделения признаков, часто они влияют на построение и обучение предсказывающей модели. Например, можно считать число переходов от пикселей фона к пикселям символа по вертикальным и горизонтальным линиям (рисунок 5 — crossings) и расстояния от границ до первого пикселя символа (рисунок 5 — distances).

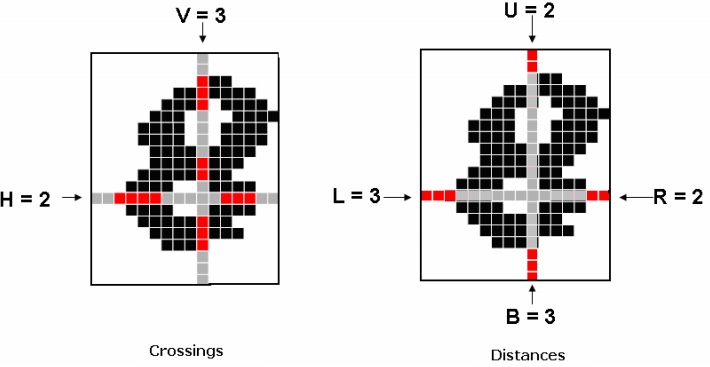


Рисунок 5

С другой стороны, для решающего дерева или сверточной нейронной сети значимыми могут являться признаки, отвечающие за количество вертикалей, горизонталей, отрезков, окружностей и т.п. В современных системах распознавания важные признаки выделяет сама нейронная сеть, обучаясь на тренировочной выборке.

## Классификация

Классифицировать объект — значит, указать номер (или наименование класса), к которому относится данный объект.

﻿Классификатор использует признаки, найденные среди исходных данных датчика, для присвоения объекту метки одного из *m* известных классов (формула (3)).

(3)

*C1, C2, …, Cm-1, Cm = Cr*

где *Cr* – класс выбросов,

*m* – размер алфавита.

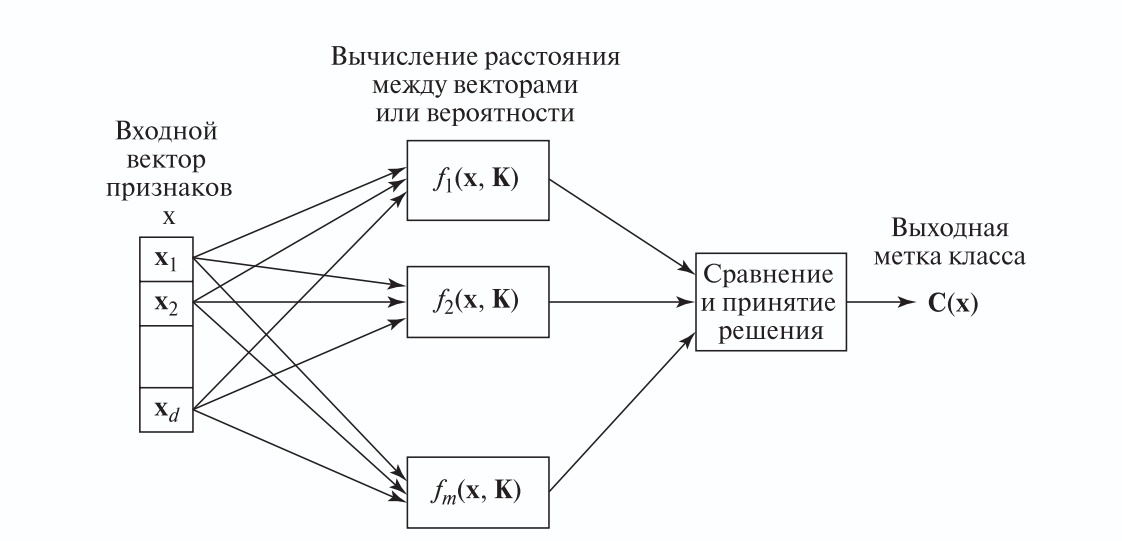
В качестве датчика используются методы определения позиции пикселя. Входными данными, представляющими объект для классификации, служит *d*-мерный вектор признаков *x*. ﻿В системе для каждого возможного класса имеется по одному блоку, в котором хранятся некоторые сведения *K* об этом классе и некоторые характеристики обработки. Результаты *m* вычислительных операций передаются на окончательный этап классификации, на котором принимается решение о классе объекта (рисунок 6). Эта схема достаточна общая и подходит под определение нескольких типов классификаций. В силу объемности рассуждений о подходах обучения, будем представлять модель как черный ящик, классифицирующий с высокой точностью входящий символ.

Рисунок 6

## Постобработка

На этом этапе обработки можно увеличить точность идентификации путем привязки словаря к выходному тексту для синтаксического и семантического анализа, однако, этот этап не является обязательным в РРТ.

# ЗАКЛЮЧЕНИЕ

В данной работе мы рассмотрели современные подходы к распознаванию рукописного текста. Самым важным этапом является правильный алгоритм предобработки данных, поскольку это значительно повышает точность классификации. Несмотря на выдающиеся продвижения в этой области, все еще требуется много исследований, чтобы получить универсальное программное обеспечение. Имея важную практическую ценность для человечества, сфера распознавания текста притягивает внимание огромного числа разработчиков и исследователей по всему миру.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Николенко, С. И., Кадурин, А. А., Архангельская, Е. В. Глубокое обучение. Погружение в мир нейронных сетей / С. И. Николенко, А. А. Кадурин, Е. В. Архангельская. – СПб.: Питер, – 2018. – 480 c.
2. Bezerra, B., Zanchettin, C. Handwriting: Recognition, Development and Analysis / B. Bezerra, C. Zanchettin. – New York: Nova Science Publishers, – 2017. – 402 p.
3. Impedovo, S., Downton, A. Progress in Handwriting Recognition / S. Impedovo, A. Downton. – Singapore: World Scientific Publishing, – 1997. – 622 p.
4. Impedovo, S. Fundamentals in Handwriting Recognition / S. Impedovo. – Berlin: Springer, – 1994. – 402 p.
5. Garcia, G., Suarez, O. Learning Image Processing with OpenCV / G. Garcia, O. Suarez. – Birmingham: Packt Publishing, – 2015. – 232 p.
6. Shapiro, L., Stockman, G. Computer Vision / L. Shapiro, G. Stockman. – Hoboken: Prentice Hall, – 2001. – 580 p.
7. Solem, J. Programming Computer Vision with Python: Tools and algorithms for analyzing images / J. Solem. – Newton: O'Reilly Media, – 2012. – 264 p.