

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

Распознавание рукописного текста

АВТОРЕФЕРАТ

дипломной работы

студента 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Русина Алексея Сергеевича

Научный руководитель

доцент, к.п.н

А. С. Гераськин

22.01.2022 г.

Заведующий кафедрой

д. ф.-м. н., доцент

М. Б. Абросимов

22.01.2022 г.

Саратов 2022

ВВЕДЕНИЕ

В настоящее время существуют устройства, способные перенести текст с твердого носителя в память компьютера. Обычно такой текст представлен в виде растрового изображения, что делает работу с ним весьма сложной. Для решения этих задач необходимо провести процесс распознавания текста на изображении с созданием файла в том или ином текстовом формате. Перевод изображений рукописного, машинописного или печатного текста в текстовые данные называется распознаванием текста.

Задача распознавания рукописного текста носит название HWR (handwriting recognition). Существуют два различных класса задач HWR:

- онлайн-распознавание — распознавание текста ведётся параллельно с вводом текста;
- оффлайн-распознавание — распознавание текста ведётся на уже синтезированном изображении.

При онлайн-распознавании процесс формирования изображения текста и процесс его ввода в систему распознавания совмещены, что позволяет системе отслеживать процесс начертания символов. Это даёт возможность получать помимо графической информации ещё и информацию о структуре входных изображений, например, о направлении и скорости движения пера или о его нажиме при написании символа. На данный момент онлайн-системы распознавания широко используются на планшетных ПК.

В задаче оффлайн-распознавания, в отличие от предыдущей, системе доступна только графическая информация. Уже это делает её значительно труднее онлайн-версией. Неполный список проблем, типичных при распознавании рукописного текста оффлайн в общем случае, включает в себя:

- высокая вариативность начертания символов — по размеру, наклону, набору составных частей, связям между ними и др.;
- орфографические ошибки в тексте;

- специфические особенности начертания, не позволяющие уверенно разделять символы;
 - пересечение элементов текста, наложение частей текста друг на друга;
 - помарки, кляксы, исправления, дефекты носителя (бумаги), а также артефакты, возникающие при сканировании;
 - непараллельность строк текста;
- и другие.

Точное распознавание латинских символов в печатном тексте в настоящее время возможно только если доступны чёткие изображения, такие как сканированные печатные документы. Точность при такой постановке задачи превышает 99%, абсолютная точность может быть достигнута только путем последующего редактирования человеком.

Данная работа посвящена оффлайн-распознаванию рукописного текста. В ней для сегментации текста используется сочетание метода диаграмм Вороного и анализ количества «черных» точек, расположенных на горизонтальной или вертикальной прямой. Также входная последовательность для скрытой марковской модели получается уникальным образом, что отличает реализованную программу распознавания рукописного текста от других.

Цель работы — изучение методов и разработка системы оффлайн-распознавания рукописного текста на основе скрытых марковских моделей.

Задачи:

- изучить и реализовать метод Отсу предварительной обработки изображения;
- изучить и реализовать метод сегментации изображения с помощью диаграмм Вороного;
- разработать программу распознавания рукописного текста на основе скрытых марковских моделей;
- проанализировать полученные результаты.

Дипломная работа состоит из введения, 3 разделов, заключения, списка использованных источников и 2 приложений. Общий объем работы – 43 страниц, из них 20 страниц – основное содержание, включая 11 рисунков и 1 таблицу, список использованных источников из 14 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

Во введении обосновывается актуальность выбранной темы дипломной работы, дается определение распознавания рукописного текста, определяются цели и задачи дипломной работы.

В первом разделе описывается задача предварительной обработки изображения, а именно устранение шумов и других дефектов изображения. Также описывается метод Отсу и его реализация. Метод Отсу является одним из наиболее эффективных методов глобальной бинаризации. Данный метод имеет ряд достоинств, таких как простота реализации метода, хорошая адаптивность метода, быстрое время выполнения (требуется $O(N)$ операций, где N – количество пикселей в изображении). Недостатком метода Отсу является чувствительность к неравномерной яркости изображения. Решением данного недостатка может быть введение локальных порогов бинаризации. После данного этапа изображение текста представляется в виде матрицы состоящей из нулей и единиц.

Во втором разделе описывается проблема сегментации изображения текста. На этом этапе производится разделение текста на строки, слова и, если возможно, на отдельные символы. Для хорошо структурированных текстов (как правило, машинописных) используется метод горизонтального или вертикального сканирования изображения. Суть данного метода в подсчете количества «черных» точек, расположенных на горизонтальной или вертикальной прямой. Для сегментации рукописного текста и в реализованной программе используется метод, основанный на построении диаграммы Вороного. Диаграмма Вороного представляет собой разбиение области на сегменты, которые называют «ячейками Вороного», а точки, с помощью которых осуществляется разбиение на сегменты называются «генераторами». Построение диаграммы будет производиться в 2 шага:

- 1) нахождение «генераторов»;
- 2) построение диаграммы.

Построение диаграммы Вороного осуществляется с помощью алгоритма Форчуна за $O(n * \log n)$. Суть алгоритма заключается в перемещении прямой от «генератора» с наибольшей ординатой к «генератору» с меньшей (или наоборот).

После второго этапа изображение рукописного текста представляется в виде отдельных символов входного текста, представленных как матрицы состоящие из нулей и единиц.

В третьем разделе описывается скрытая марковская модель, определяются три основные задачи, которые необходимо решить для использования скрытой марковской модели в распознавании рукописного текста, описываются алгоритмы решения трех основных задач и их реализация, а также производится анализ полученных результатов тестирования программы.

В первом подразделе третьего раздела дается определение марковской модели и скрытой марковской модели. Скрытые марковские модели описывают двухэтапный случайный процесс. Первый этап состоит из дискретного стохастического процесса, который является стационарным, причинным и простым. Рассматриваемое пространство состояний конечно. Таким образом, процесс вероятностно описывает переходы состояний в дискретном пространстве конечных состояний. Его можно визуализировать как конечный автомат с ребрами между любой парой состояний, которые помечены вероятностями перехода. Поведение процесса в данный момент времени t зависит только от состояния непосредственного предшественника.

На втором этапе для каждого момента времени t дополнительно, путем вывода или выходных данных, генерируется O_t . Распространение ассоциативной вероятности зависит только от текущего состояния S_t , а не от каких-либо предыдущих состояний или выводных данных. Эта последовательность выводных данных единственное, что можно наблюдать в поведении модели. С другой стороны, состояние последовательности принятое во время генерации данных не может быть исследована. Это и есть, так

называемая, «скрытость», из которой выводится определение скрытых моделей Маркова.

Во втором подразделе третьего раздела определяются три основные задачи, которые необходимо решить для использования скрытой марковской модели в распознавании рукописного текста.

Три основные задачи:

- задача оценки;
- задача декодирования;
- задача обучения.

В первом подразделе второго подраздела третьего раздела описывается задача оценки и алгоритм прямого-обратного хода, применяющийся для эффективного решения данной задачи. Задача оценки является задачей вычисления вероятности появления последовательности наблюдений, учитывая параметры модели. Суть алгоритма прямого-обратного хода заключается в вычислении, рекурсивно оцениваемых прямых переменных, а затем суммирования всех прямых переменных в момент времени T . Количество задействованных вычислений составляет порядка TN^2 , где T – длина последовательности наблюдений, N – число возможных состояний.

Во втором подразделе второго подраздела третьего раздела описывается задача декодирования и алгоритм Витерби для её решения. Задача декодирования включает в себя поиск оптимальной последовательности состояний, заданной последовательностью наблюдений и параметр модели. В алгоритме Витерби вычисляется максимальная вероятность оптимальной последовательности частичных состояний с соблюдением последовательности частичных наблюдений, учитывая модель. С помощью максимальной вероятности оптимальной последовательности частичных состояний можно вычислить совместную вероятность оптимальной последовательности состояний и последовательности наблюдений, заданной моделью.

В третьем подразделе второго подраздела третьего раздела описывается задача обучения и алгоритм Баума-Уэлча, который гарантирует локально оптимальное решение данной задачи. Задача обучения вычисляет оптимальный параметр модели, учитывая последовательность наблюдений. Суть алгоритма Баума-Уэлча в вычислении прямых и обратных переменных, чтобы с их помощью определить то, насколько хорошо модель соответствует последовательности наблюдений.

В третьем подразделе третьего раздела описывается реализация программы распознавания рукописного текста. Распознавание происходит следующим образом:

- матричное представление символа кодируется последовательностью;
- с помощью алгоритма Баума-Уэлча происходит обучение модели на обучающих последовательностях;
- с помощью алгоритма прямого-обратного хода оценивается вероятность принадлежности входной последовательности обучающим последовательностям;
- основываясь на оценке, полученной на предыдущем шаге, происходит распознавание символа.

Для кодирования символа необходимо разделить столбцы его матричного представления, полученного после этапа сегментации, на 3 равные части, затем каждой части столбца ставится в соответствие 0 или 1 и столбец кодируется числом от 0 до 7 включительно. Таким образом, изображение представляется в виде последовательности цифр от 0 до 7 включительно. Данное представление отличает реализованную программу от других программ распознавания рукописного текста, использующих скрытые марковские модели.

В четвертом подразделе третьего раздела производится анализ полученных результатов. Анализ проводился на изображениях английских слов. Анализ показал, что программа часто путает буквы «e» и «l». Это можно объяснить тем, что при написании эти буквы действительно похожи и даже

человек не всегда способен различить их. Для увеличения точности распознавания можно анализировать полученные последовательности букв при помощи словаря.

В заключении подведены итоги дипломной работы, изложены её основные выводы и обобщающие результаты.

ЗАКЛЮЧЕНИЕ

В настоящей работе были изучены и реализованы: метод Отсу обработки изображения, алгоритм Форчуна построения диаграммы Вороного, алгоритм прямого-обратного хода и алгоритм Баума-Уэлча. Так же была построена скрытая марковская модель, на основе которой была написана программа распознавания рукописного текста.

Метод Отсу, использующийся в реализованной программе распознавания рукописного текста, имеет ряд достоинств, таких как простота реализации метода, хорошая адаптивность метода, быстрое время выполнения (требуется $O(N)$ операций, где N – количество пикселей в изображении). Недостатком метода Отсу является чувствительность к неравномерной яркости изображения. Решением данного недостатка может быть введение локальных порогов бинаризации.

Сегментация рукописного текста с использованием диаграммы Вороного используется во многих современных системах распознавания рукописного текста. Алгоритм Форчуна является одним из самых эффективных алгоритмов построения диаграммы Вороного и строит диаграмму за $O(n * \log n)$.

Скрытая марковская модель, с помощью которой производится распознавание рукописного образа, часто используется в системах распознавания рукописного текста, так как в точности распознавания с ней может поспорить только система, использующая нейронные сети. Недостатками использования скрытой марковской модели являются: большая сложность вычислений, требование, чтобы каждой строке соответствовало малое количество скрытых состояний.

Реализованная программа имеет точность распознавания на 10-20% меньше чем у лучших известных программ распознавания рукописного текста, таких как ABBYY FineReader, SimpleOCR, PDFelement Pro, OCR Desktop (Free Online OCR), TopOCR.

На данный момент точность распознавания не является достаточной, но использование словарей, ограничение возможных входных символов, дополнительный смысловой анализ текста, могут сильно повысить точность распознавания текста.