

005006147

КУРУШИН Даниил Сергеевич

МОДЕЛИРОВАНИЕ ПРОЦЕССА
ИДЕНТИФИКАЦИИ ГРАФИЧЕСКИХ ОБЪЕКТОВ

Специальность 05.13.18 — Математическое моделирование,
численные методы и комплексы программ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

15 ДЕК 2011

Работа выполнена на кафедре информационных технологий и автоматизированных систем
ФГБОУ ВПО «Пермский национальный исследовательский политехнический университет».

Научный руководитель:

доктор экономических наук, профессор
Долгова Елена Владимировна

Официальные оппоненты:

доктор физико-математических наук,
профессор Шатров Анатолий Викторович

кандидат технических наук
Скляренко Максим Сергеевич

Ведущая организация:

ФГБОУ ВПО «Пермский
государственный национальный
исследовательский университет»

Защита диссертации состоится 20 декабря 2011 г. в 14:00 на заседании диссертационного совета Д 212.188.08 при Пермском национальном исследовательском политехническом университете по адресу: 614990, г. Пермь, Комсомольский проспект, д. 29, ауд. 423.

С диссертацией можно ознакомиться в библиотеке Пермского национального исследовательского политехнического университета.

Автореферат разослан « 18 » ноября 2011 г.

Ученый секретарь диссертационного
совета, доктор физико-математических
наук, доцент



Л.Н. Кротов

1 Общая характеристика работы

Актуальность темы. Первые работы по распознаванию рукописного текста относятся к концу 1970-х — началу 1980-х годов (History of Pen and Gesture Computing: Annotated Bibliography in On-line Character Recognition, Pen Computing, Gesture User Interfaces and Tablet and Touch Computers). С тех пор был достигнут определенный прогресс в части распознавания отдельных символов, вводимых стилусом (реже мышью) в специальные поля. Также с приемлемым качеством решается задача распознавания т.н. рукопечатных символов, вводимых в поля анкет (т.н. блочеттеры). Существенно большую сложность представляет распознавание неограниченного слитного рукописного текста, причем ряд исследователей вообще полагает эту задачу неразрешимой на современном уровне развития технологии: «Решить задачу распознавания слитного текста с высоким результатом можно будет, только когда компьютер сможет понимать смысл предлагаемого текста» (А. Абраменко).

Задача распознавания оцифрованной рукописи может быть решена несколькими методами. Так, выделяют модели, основанные на идентификации отдельных объектов и их частично упорядоченных наборов. В 1998 г. Larry S. Yaeger, Brandyn J. Webb и Richard F. Lyon для компьютера Apple NEWTON разработали модель, комбинирующую нейронные сети и контекстно-зависимый поиск для распознавания символов, вводимых стилусом в специальное поле на экране.

А.В. Мисюрёв, исследуя возможности идентификации таких объектов, как символы кириллического алфавита, цифры и иные знаки, используемые при заполнении анкет, предлагает алгоритм распознавания, основанный на выделении из раstra с изображением буквы первичных признаков и использовании нейрореподобной модели для оценки близости входного изображения известным символам.

Принципиально другой подход рассмотрен в работах Котович Н. В., Славина О. А., Клейнберга Е. и др. Он основан на предварительной скелетизации изображения, преобразовании его в векторную форму и последующей идентификации векторного представления символов. Исследования Пламодон Р. и Шринари С. показывают меньшую чувствительность таких моделей к искажениям символов, возникающим при письме и сканировании.

Современные системы, поддерживающие распознавание рукописного текста, такие как:

- Chinese Handwriting for Linux — приложение для распознавания китайских символов (2001);
- Stylus Handwriting Input Panel — система ввода рукописного текста, основанная на распознавании штрихов для планшетных ПК (2008);
- HaRe — система рукописного ввода для иврита (2006);
- Tomoe — приложение для распознавания рукописного ввода на японском языке (2004);
- CellWriter — панель рукописного ввода, поддерживающая большинство современных языков (после обучения), (2011);
- Kadmos OCR/ICR — API поддержки рукописного ввода для C, C++, VB, .NET, Delphi и Java (2006), —

преимущественно ориентированы на работу с отдельными символами. Исключение составляет Stylus Handwriting Input Panel, однако эта система допускает только ввод текста «на лету» и не способна к распознаванию растровых изображений. Другим исключением является служба Evernote, предлагающая услугу индексирования рукописных документов на большинстве языков. Однако Evernote распознает лишь отдельные слова в тексте и тратит на индексирование одного документа до 24-х часов. Как можно видеть, в современных комплексных программ проблема распознавания слитного рукописного текста решена лишь частично.

Таким образом, создание математической модели идентификации графических объектов, работающей в условиях слитного рукописного текста, и разработка программного комплекса на ее основе являются актуальными задачами.

Целью работы является разработка и исследование модели идентификации графических объектов — примитивов, составляющих рукописный текст неограниченной формы и объема, обеспечивающей повышение точности распознавания слитного рукописного текста, создание

проблемно-ориентированного комплекса программ для выделения графических объектов из рукописи и их последующей идентификации.

Задачи исследования. Для достижения поставленной цели были решены следующие задачи:

1. Анализ существующих методов распознавания рукописного текста, изучение применяемых моделей выделения и идентификации графических примитивов — элементов рукописи.
2. Разработка модели и алгоритма выделения примитивов из слитного рукописного текста, удовлетворяющих требованиям устойчивости к искажениям, возникающим при сканировании рукописи.
3. Разработка модели скелетизации векторных представлений графических объектов.
4. Разработка адаптированной нейросетевой модели идентификации примитивов, предназначенной для работы со структурными представлениями графических объектов.
5. Доказательство перцепционной представимости структурных представлений графических объектов.
6. Создание проблемно-ориентированного комплекса программ на основе численных методов решения вышеуказанных задач.
7. Проверка модели путем сравнения результатов идентификации с известными моделями и результатами, полученными экспертами.

Объект и предмет исследования. Объектом исследования является рукописный текст и графические примитивы, его составляющие. Предметом исследования являются характеристики графических примитивов, позволяющие выполнить их идентификацию.

Методы исследования. Для решения задач, сформулированных в работе, использованы методы нейросетевого анализа, обработки изображений, вычислительного эксперимента, искусственного интеллекта, технологии объектно-ориентированного программирования. При разработке проблемно-ориентированного программного комплекса использовались АЯП Python и Java, среда разработки Netbeans 6.9.

Достоверность и обоснованность полученных результатов подтверждается согласованностью результатов моделирования с расчетами, основанными на известных моделях а также с результатами идентификации графических объектов, выполненной экспертом.

Научная новизна работы состоит в следующем:

1. Создана математическая модель идентификации графических объектов, работающая с примитивами, составляющими слитный рукописный текст, и нечувствительная к искажениям, возникающим при оцифровке рукописей;
2. Впервые разработан адаптационный слой нейросетевой модели, что позволило понизить размерность нейронной сети на 1 – 2 порядка;
3. Впервые предложен однопроходный метод сегментации рукописного текста, слабо чувствительный к возникающим при оцифровке изображений градиентам яркости;
4. Впервые предложен метод геометрической скелетизации векторных представлений графических объектов.

Практическая ценность:

1. Предложенная модель и ее программная реализация позволяют снизить размерность нейронной сети, используемой для идентификации графических объектов, на 1 – 2 порядка.
2. Разработанное алгоритмическое и программное обеспечение может использоваться для исследования графической структуры рукописных текстов в приложениях распознавания текста, индексации оцифрованных рукописей, графологической экспертизы.

3. Разработанные модели и программное обеспечение может использоваться в таких сферах, как распознавание образов, робототехника (ориентирование в пространстве), идентификация объектов и т.п.

Внедрение результатов. Разработанная модель идентификации графических примитивов используется в учебном процессе Пермского национального исследовательского политехнического университета и Пермского государственного национального исследовательского университета при изучении дисциплин «Системы искусственного интеллекта», «Компьютерная графика», «Автоматическая обработка естественного языка».

Апробация работы. Результаты работы докладывались на следующих конференциях и семинарах:

1. Международная научно-практическая конференция «Перспективы развития информационных технологий», 2011, Новосибирск.
2. Научный семинар кафедры Математического моделирования систем и процессов, Пермского государственного технического университета, 2011, Пермь, рук. д.ф.-м.н., профессор Трусов П.В.
3. Научно-практический семинар кафедры информационных технологий и автоматизированных систем Пермского государственного технического университета, 2010, Пермь, рук. д.э.н., профессор Файзрахманов Р.А.
4. Международная интернет-конференция «Инновационные технологии: теория, инструменты, практика», 2010, Пермь.
5. Краевая дистанционная научно-практическая конференция «Молодежная наука Прикамья», 2009, Пермь.
6. Всероссийская конференция «Теория и практика речевых исследований», 2001, Москва.

Основные научные положения, выносимые на защиту:

1. Графические примитивы, составляющие слитный рукописный текст, могут быть формально описаны постоянным конечным количеством сегментов, причем для корректной идентификации примитива достаточно знать только траекторию пера, описываемую кортежем углов между сегментами.
2. Разработанная, с учетом приведенного выше утверждения, математическая модель процесса идентификации графических объектов позволяет сократить размерность нейронной сети на 1 – 2 порядка по сравнению с признаковыми моделями.
3. Полученный в ходе работы проблемно-ориентированный комплекс программ позволяет разбивать рукопись на графические примитивы и идентифицировать их с вероятностью 0.81-0.98.

Публикации. Соискатель имеет 11 опубликованных работ по теме диссертации в центральных (5 работ) и местных (6 работ) изданиях, в которых отражены основные положения диссертации. Список работ приводится в конце автореферата.

Объем и структура работы. Диссертация состоит из введения, трех глав, заключения, списка литературы; изложена на 87 страницах, содержит 15 рисунков; библиографический список включает 67 наименований, 4 приложения, 14 таблиц.

Личный вклад автора. Постановка задачи осуществлялась совместно с научным руководителем к.т.н., д.э.н., проф. Долговой Е.В. Основные результаты диссертационного исследования получены автором самостоятельно. Автором проведен анализ предметной области, предложена гипотеза о том, что графические примитивы, составляющие слитный рукописный текст, могут быть формально описаны постоянным конечным количеством сегментов, разработано структурное описание графического примитива, предложена и исследована модель идентификации.

Все предложенные математические модели алгоритмизированы с использованием АЯП Python и Java.

Автором разработан алгоритм сегментации слитного рукописного текста (на основе работ Wang и Suen). Предложен новый подход к структурному описанию рукописного текста, проведено доказательство перцептральной представимости графических примитивов, составляющих слитный рукописный текст.

Автором разработан проблемно-ориентированный комплекс программ предназначенный для исследования и уточнения параметров моделей.

2 СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы диссертационной работы, определены цель и задачи исследования, показана научная новизна и практическая ценность работы, изложены научные положения, выносимые на защиту.

В **первой главе** приводится анализ современного состояния исследований в области распознавания текста.

На рис. 1 представлена классификация методов распознавания, составленная на основании исследования «Основные методы, применяемые для распознавания рукописного текста» (Мерков А.В., 2005). Как можно видеть, существует два глобальных подхода: распознавание отдельных объектов (символов) и их упорядоченных наборов (кортежей). Для распознавания слитного письма методы, работающие с отдельными объектами или неприменимы вообще, или требуют предварительного разделения слитного текста на символы. Как показано в статье «Алгоритмы сегментации рукопечатных символов» (А.А. Михайлов, В.В. Постников, 2003), разделение слитного рукописного текста на отдельные символы сопряжено с рядом трудностей и зачастую не имеет однозначного решения.

Таким образом, для распознавания сплошного рукописного текста предпочтительным является использование методов, основанных на анализе упорядоченного набора объектов, однозначно выделяемых из текста.

Ряд исследователей (G.L. Martin, Y. Le Cun, O. Matan) для распознавания рукописных символов предлагают метод, основанный на выделении из растрового изображения первичных признаков (цвет, яркость, наличие границы) и использовании нейронной модели для оценки близости входного изображения известным символам. Отметим, что данная модель требует преобразования исходных данных к единому размеру 16×16 пикселей, т.е. к кортежу $\langle x_1, \dots, x_{256} \rangle$, $x \in [0, 256]$, что приводит к возникновению нелинейных искажений и ряда других проблем, понижающих качество распознавания текста.

Один из возможных подходов рассмотрен в работе «Распознавание скелетных образов» (Н.В. Котович, О.А. Славин, 2003). Этот подход основан на поиске концевых точек и точек ветвления в растровом скелете изображения символа. Л.М. Местецкий в работе «Непрерывный скелет бинарного изображения» предлагает метод и алгоритм выделения примитива из растрового изображения (в т.ч. применимый для анализа рукописи), основанный на получении обводящего контура. Также интерес представляет метод квантилей, предложенный W. Doyle в работе «Operations useful for similarity invariant pattern recognition» (1962) и алгоритмизированный в статье «Сегментация изображений методом квантилей» (2011). Исследования Пламон Р. и Шринари С. показывают меньшую чувствительность таких моделей к искажениям символов, возникающим при письме и сканировании.

По материалам первой главы формулируются следующие выводы:

1. Сравнительный анализ подходов к распознаванию слитного рукописного текста показывает преимущества методов, основанных на идентификации кортежей элементов (фрагментов символов).
2. Известные модели идентификации графических объектов чувствительны к искажениям, неизбежно возникающим в процессе письма.

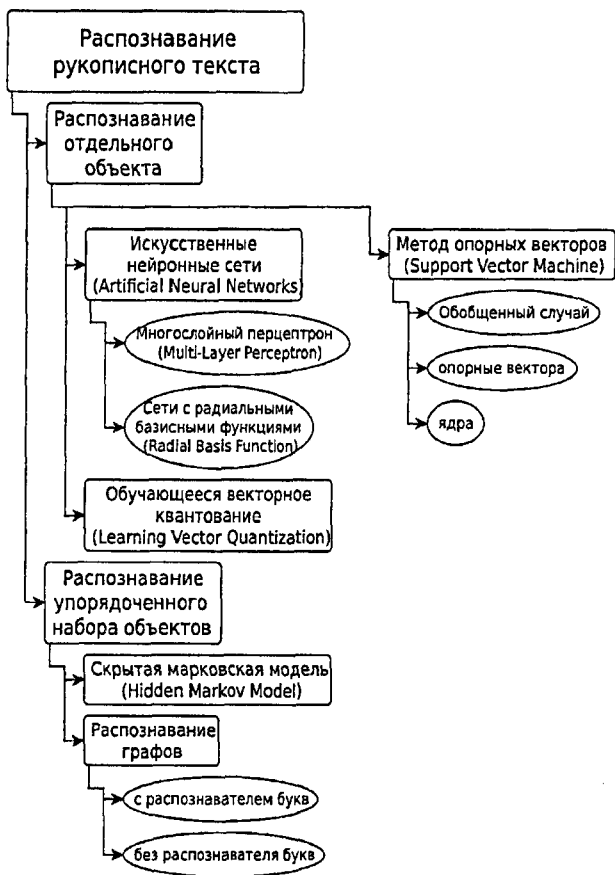


Рис. 1. Обзор методов распознавания рукописного текста

3. Широко используемые растровые методы скелетизации слитного рукописного текста приводят к существенным искажениям графических объектов, что затрудняет или делает невозможной их идентификацию.
4. Необходимы дальнейшие исследования, направленные на создание математической модели идентификации графических объектов, нечувствительной к искажениям, возникающим при оцифровке рукописей.
5. Актуальной задачей является сокращение размерности нейросетевых моделей, используемых для идентификации примитивов.

Вторая глава диссертационного исследования посвящена исследованию процесса распознавания рукописного текста и описанию модели идентификации графических объектов, его составляющих.

Процесс распознавания рукописного текста состоит из следующих фаз:

- сегментация текста из слитного рукописного текста выделяются примитивы, позволяющие сделать задачу их классификации относительно несложной;

- распознавание элементов рукописи — сегментированное изображение анализируется классификатором на основе нейронной сети и решаются две задачи:
 - определения класса элемента;
 - определения взаимного расположения элемента;
- генерация выходного документа, сохраняющего как информацию об оригинальном графическом представлении, так и текстовое представление с сохранением связи изображение ↔ текст.

Одной из ключевых стадий распознавания является сегментация растрового изображения. В настоящей работе предложен один из возможных подходов к сегментации рукописного текста на примитивы.

В общем виде этот процесс можно описать отображением:

$$I^s = S(I), \quad (1)$$

где I^s — сегментированное изображение, I — исходное изображение, а S — отображение исходного изображения в сегментированное.

Информация о цвете и яркости с точки зрения сегментации изображения не представляет интереса, поэтому первой операцией будет бинаризация — приведение изображения к черно-белому формату. Традиционно эта задача решается с использованием σ -функции, такой что:

$$\sigma = \begin{cases} 0, & x < x_0 \\ 1, & x \geq x_0 \end{cases}, \quad (2)$$

где x — входное значение, x_0 — пороговое значение функции. Однако такое решение не может быть оптимальным для нашей задачи, т.к. значение x_0 может отличаться в разных участках изображения, поэтому в нашем случае σ имеет вид:

$$\sigma = \begin{cases} 0, & x < x_0 \\ 1, & x \geq x_0 \end{cases}, \quad x_0 = \frac{\sum I_n(\vec{x})}{n} + b, \quad (3)$$

где I_n — последовательность пикселей изображения I длиной в n , \vec{x} — вектор координат текущей точки изображения (x, y) , b — смещение, подбираемое экспериментально.

После бинаризации изображения следует удалить остаточные пиксели фона, которые образовались вследствие выбора низкого порога яркости пикселя в предыдущем алгоритме. Удаление происходит следующим образом: из изображения выделяются четыре пикселя по следующей формуле:

$$f(n, \vec{x}) = \begin{cases} n = 0, & (x_1, x_2) \\ n = 1, & (x_1 + 1, x_2) \\ n = 2, & (x_1, x_2 + 1) \\ n = 3, & (x_1 + 1, x_2 + 1) \end{cases}, \quad (4)$$

где n — порядковый номер соседа обрабатываемого пикселя. Обрабатываемым является пиксель $n = 0$. Далее считается суммарная яркость \mathfrak{B} выделенных пикселей:

$$\mathfrak{B} = \sum_{i=0}^3 f(i, \vec{x}), \quad (5)$$

и, если она больше единицы, то пиксель помечается белым:

$$I'(\vec{x}) = \begin{cases} 1, & \mathfrak{B}(\vec{x}) > 1 \\ 0, & \mathfrak{B}(\vec{x}) \leq 1 \end{cases}, \quad (6)$$

где $\mathfrak{B}(\vec{x})$ — яркость в точке с координатами \vec{x} .

После проведения процедуры бинаризации (согласно (3)) и удаления остаточных фоновых пикселей (6) изображение достаточным образом подготовлено для процедуры скелетизации. Основная масса подобного рода алгоритмов основана на выделении средней линии фигуры, после чего лишние пиксели удаляются. Однако можно выделять и край линии, при таком подходе алгоритм получается заметно проще.

Из изображения выделяются четыре пикселя (проход слева-направо, сверху вниз). После чего формируется матрица размерностью 2×2 , в которую заносятся выделенные пиксели. Предлагаемый алгоритм скелетизации основан на следующих уравнениях:

$$N_w(\vec{x}_i) = \sum_{i=1}^4 \sigma_w(\mathcal{B}(\vec{x}_i)), \quad (7)$$

$$\sigma_w = \begin{cases} 1, & x = 1 \text{ (белый)} \\ 0, & x = 0 \text{ (черный)} \end{cases}, \quad (8)$$

$$N_{0 \rightarrow 1}(\vec{x}_i) = \sum_{i=1}^4 \nabla_{0 \rightarrow 1}^i \mathcal{B}(\vec{x}_i), \quad (9)$$

$$\nabla_{0 \rightarrow 1}^i = \begin{cases} 1, & B(\vec{x}_i) - B(\vec{x}_{i+1}) = -1, i \in [1, 3] \\ 0, & B(\vec{x}_i) - B(\vec{x}_{i+1}) \neq -1, i \in [1, 3] \\ 1, & B(\vec{x}_i) - B(\vec{x}_1) = -1, i = 4 \\ 0, & B(\vec{x}_i) - B(\vec{x}_1) \neq -1, i = 4 \end{cases}. \quad (10)$$

Таким образом, с учетом (7), (8), (9), (10) функция сегментации (для указанной точки) примет вид:

$$I'(\vec{x}) = S(I'(\vec{x})), \quad (11)$$

где:

$$S = \begin{cases} 0, & 0 \leq N_w \leq 2, N_{0 \rightarrow 1} = 1 \\ 1, & N_w > 2 \\ 1, & N_{0 \rightarrow 1} \neq 1 \end{cases} \quad (12)$$

Пример изображения до и после обработки по формулам ((11), (12)) показан на рис. 3. Как можно видеть, из символов успешно выделены наборы однотипных сегментов, которые можно легко отделить друг от друга и распознать.

Распознавание штрихов, представленных как набор точек (в растровом формате), может выполняться любым традиционным способом, например, могут использоваться нейронные сети различных конфигураций. Однако использование нейронных сетей в общем случае приводит к необходимости или нормализовывать размеры изображения, или увеличивать размерность сети.

Первый вариант приводит к возникновению нелинейных искажений, что ухудшает качество распознавания (или опять-таки требует сети большей размерности). Второй вариант, хоть и не приводит к искажениям входного сигнала, тоже не лишен недостатков. Любая нейронная сеть большой размерности требует пропорционально большого количества элементов в обучающей выборке. Подготовка таких массивов данных сложна технически и может порождать ошибки, обусловленные человеческим фактором.

Для устранения этой проблемы будем использовать в качестве входного сигнала нейронной сети не растровое изображение, а его векторное представление. Под векторным представлением растрового изображения будем понимать такой нагруженный граф, у которого нагрузка вершин — пары координат (x, y) соответствующих узловых точек изображения. За узловые точки будем принимать точки соединения изображения линий в растровом изображении. В зависимости от выбираемой стратегии изображение может рассматриваться как состоящее из отрезков прямых.

Векторное представление примитива — это направленный граф, построенный следующим образом: пусть p — точка с целочисленными координатами (соответствует 4-м пикселям изображения), такую точку будем считать вершиной (*vertex*) и обозначать как v (или w), если эти четыре пикселя имеют различия по цвету. Также будем считать, что между вершинами v и w существует грань (edge, e), если

$$E(v, w) = 1, \quad (13)$$

где E — Евклидово расстояние и если отрезок VW отделяет пиксель черного цвета от пикселя белого цвета таким образом, что черный пиксель остается слева от условного направления движения. Продолжая двигаться таким образом от вершины к вершине, мы получаем направленный граф G . Граф представляет собой путь (path, P) $\{v, \dots, v_n\}$, такой что:

$$\exists P, \exists e_{i \dots n}(v_i, v_{i+1}) = 1, \wedge e_i \neq e_j, \forall i, j \leq n. \quad (14)$$

Путь P называется замкнутым, если $v_n = v_0$. Учитывая (11), замкнутый путь P_i соответствует некоторому элементу, обозначим его W_j . Индексы отличаются, т.к. некоторые пути могут соответствовать незначимым элементам или оставаться после фильтрации шума.

Т.о. задача классификации путей P_i в элементы W_j может быть сформулирована следующим образом: найти такую классифицирующую функцию F , что:

$$F(P_i) = \begin{cases} W_j, & P_i \in W \\ \emptyset, & P_i \notin W \end{cases} \quad (15)$$

где W — множество известных рукописных элементов (задача определения такого множества является задачей лингвистического исследования и не включена в настоящую работу, хотя и выполнена автором). Будем считать, что F — многослойная нейронная сеть, тогда нам необходимо дополнительно преобразовать замкнутый путь P таким образом, чтобы каждый его элемент v_i принадлежал диапазону $[-1, 1]$.

Элементы пути P представляют собой вершины, т.е. вектора координат $\{x, y\}$. Эта форма представления информации удобна для, например, отображения на экране, но нерациональна с точки зрения распознавания формы пути (для отражения его в элемент множества W). Форму пути можно описать, разбив его на известное количество незамкнутых путей P' равной длины и рассмотрев угол поворота α_i при переходе $P'_i \rightarrow P'_{i+1}$. Для решения этой задачи будем использовать дополнительный (адапционный) слой нейросетевой модели, имеющий целью преобразование замкнутого пути с неопределенным (но конечным) количеством вершин в незамкнутый, состоящий из заранее определенного числа отрезков.

Чтобы отфильтровать возможные дефекты формы штриха, вызванные толщиной линий, и существенно упростить путь, перейдем от замкнутого пути P к незамкнутому P^m , представляющему собой среднюю линию.

Для этого (для каждого пути P_i) выполняем:

$$P_i^m = f^m(P_i), \quad (16)$$

где f^m — функция скелетизации контура. Определим ее:

$$f^m(P) = f^m(\{v_0, \dots, v_n\}) = \{v_0^m, \dots, v_n^m\}, \quad \text{где } v_k^m = \left(\frac{x_i + x_{i+1}}{2}, \frac{y_i + y_{i+1}}{2} \right) \forall i, i < n, \forall k, k = \frac{i}{2}. \quad (17)$$

Очевидно, что в зависимости от сложности исходных путей и их геометрических размеров количество вершин в них может отличаться. Соответственно, оно будет отличаться и в путях P^m .

Известно, что размерность входного вектора (обычно обозначаемого x) нейронной сети постоянна для всей выборки, поэтому к входным данным применяют процедуру нормализации (по размерности и значению). Угол поворота α_i пути P^m в каждой точке v_k лежит в диапазоне $(-\pi, \pi)$. Приведение диапазона $(-\pi, \pi)$ к $(-1, 1)$ может быть выполнено или σ -функцией, принятой в теории нейронных сетей, или обычным линейным преобразованием. Использования

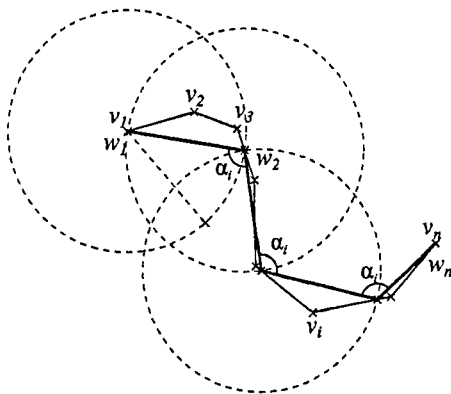


Рис. 2. Выделение сегментов и определение углов поворота

σ -функции предпочтительнее, т.к. она «растягивает» небольшие углы поворота и «сжимает» предельные, что соответствует физике процесса письма. Т.о. мы будем использовать:

$$\sigma(\alpha_i) = \frac{1}{1 + e^{\frac{10}{\pi} \alpha_i}}. \quad (18)$$

Осталось рассчитать значения α_i .

Как сказано выше, путь P^m необходимо разбить на конечное постоянное число сегментов p_j^m . Для этого определяем суммарную длину пути P^m как:

$$L = |P^m| = \sum_{i=0}^{n-1} \left| \overrightarrow{\{v_i, v_{i+1}\}} \right|, \quad (19)$$

определяем длину сегмента как:

$$l = \frac{L}{N}, \quad (20)$$

где N — необходимое число сегментов и одновременно размерность входного вектора нейронной сети.

Для разбиения пути P^m на сегменты воспользуемся следующим способом: начиная с произвольно выбранного конца пути P^m строится ряд окружностей радиуса l , так что центром первой является вершина v_1 , далее w_1 . Точка w_2 определяется методом поиска точки пересечения окружности и отрезка, с центром в этой точке строится новая окружность того же радиуса, определяющая вершину w_3 , и так далее, пока не будет достигнуто такое состояние, что все необработанные вершины лежат внутри очередной окружности. Тогда угол α_i есть угол между отрезками $w_i w_{i+1}$.

Таким образом, входной вектор нейронной сети \vec{x} рассчитывается:

$$\begin{aligned}\vec{x} &= f^n(P^m) = f^n(\{v_1, \dots, v_n\}) = \{x_1, \dots, x_N\}, \\ x_j &= \sigma(\alpha_j), \\ \alpha_j &= f^a(w_{j-1}, w_j, w_{j+1}), j \in [2, N-1].\end{aligned}\quad (21)$$

Проделанные выше преобразования позволили выполнить процедуру нормализации изображения без потери содержательно-значимой информации, а также существенно сократить размерность сети, абстрагировавшись одновременно от начального угла поворота изображения (обычно представляющего сложность для сетей, распознающих растровое изображение).

Настройка модели на работу с текстами определенного содержания и графического качества ведется путем подбора следующих параметров:

1. Смещения b — выражение (3);
2. Количество входов нейронной сети N — (20).

По материалам второй главы формулируются следующие выводы:

1. Из рукописного текста можно выделить составляющие его примитивы таким образом, что сохраняется визуальная читаемость текста.
2. Полученные примитивы можно преобразовать к векторной форме и скелетизировать.
3. Скелетное представление примитива может использоваться в качестве входной информации нейросетевой модели при условии ее адаптации.
4. В качестве математической модели распознавания элементов рукописи предлагается использовать нейронную сеть, дополненную описанным выше адаптационным слоем.

В третьей главе диссертационного исследования описывается структура проблемно-ориентированного комплекса программ, а также методика и результаты экспериментальных исследований, проведенных с его помощью.

В проблемно-ориентированный комплекс программ входят реализации следующих моделей:

1. Модели сегментации рукописного текста;
2. Модели скелетизации векторных представлений рукописных элементов;
3. Модели нейронной сети (на базе сети Кохонена), настроенной на классификацию рукописных элементов;
4. Словаря рукописных элементов для массива текстов, полученных из ответов студентов специальности АСУ;
5. Модели нейронной сети, настроенной на распознавание рукописных элементов;
6. Модели оптимального представления набора рукописных элементов текстом на естественном языке.

Алгоритм сегментации растрового изображения реализован на АЯП Java и протестирован на массиве оцифрованных рукописных текстов. По результатам экспериментов подготовлена и опубликована работа [1].

Разработанный метод предварительной обработки изображения можно условно разбить на 3 этапа: бинаризация, сегментация и скелетизация. На этапе бинаризации происходит разделение пикселей изображения на две группы: фоновые пиксели и пиксели объектов (см. 3). На этапе сегментации происходит окончательное уточнение границ объектов, которые затем уточняются до толщины в один пиксель. На рисунках 3 и 4 представлены исходные изображения и результаты сегментации. Видно, что слитно написанные слова разделены на символы, однако читаемость текста сохранилась.

Как можно видеть, из символов успешно выделены наборы однотипных сегментов, которые можно легко отделить друг от друга. Результаты сегментации были обработаны алгоритмом *potrace*.

Стихотворение

Стихотворение

Рис. 3. Текст до и после сегментации

Таблица 1. Изменение количества сегментов и их средней длины после преобразования

Пример	К-во сегментов	Длина сегмента (дуг)
1	6	175
2	64	20
3	57	36
4	128	16

В таблице 1 примеры 1 и 2 взяты с рис. 3 (1 — до сегментации, 2 — после), примеры 3 и 4 взяты с рис. 4 (1 — до сегментации, 2 — после, анализировался фрагмент).

Алгоритм скелетизации векторного представления графического примитива реализован на АЯП Python и протестирован на данных, полученных после векторизации результатов работы сегментатора, а также штрихов, подготовленных вручную. По результатам экспериментов подготовлена и опубликована работа [2], в которой доказано, что предложенная модель позволяет сократить размерность нейронных сетей, используемых при распознавании рукописного текста на 1 – 2 порядка (в зависимости от подхода).

Таблица 2. Сравнение результатов распознавания разными нейросетевыми моделями

Тип сети	Качество распознавания	Время обучения (сек)	Циклов обучения
Сеть на базе персептрона	0.81	15	1500
Нейронная сеть Хопфилда	0.98	5	150
Каре Кохонена	n/a	7	n/a

С помощью АЯП Python и модуля NumPy реализованы:

- Каре Кохонена,
- Многослойная сеть на основе персептрона,
- Нейронная сеть Хопфилда,
- Адаптационный слой, совместимый (после настройки) с любой нейросетевой моделью.

Разработанные алгоритмы протестированы на данных, полученных после скелетизации, а также штрихов, подготовленных вручную, по результатам экспериментов подготовлена и опубликована работа [3]. Сравнение результатов распознавания разными нейросетевыми моделями приведено в табл. 2.

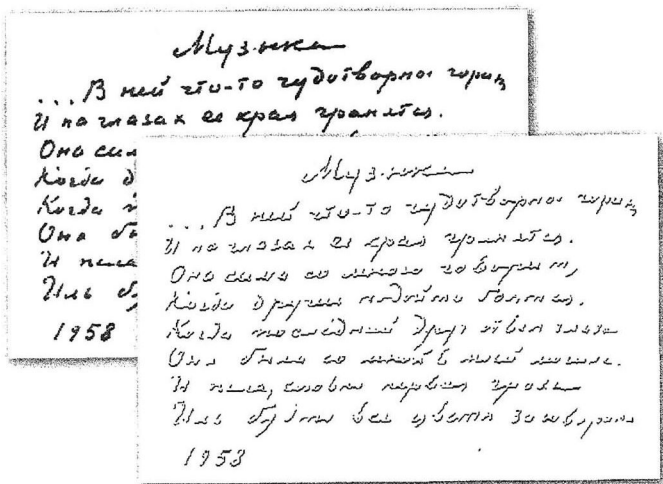


Рис. 4. Текст до и после сегментации

Модель скелетизации рукописных элементов будет верна тогда, когда она сможет привести множество рукописных примитивов к такому виду, что оно будет представимо в терминах нейронной сети и разделимо кусочно-линейной (или линейной) функцией, получаемой при обучении.

Проверку представимости и разделимости множества выполняют либо создавая некоторую обучающую выборку и экспериментируя с топологиями сетей, либо с использованием самоорганизующихся карт.

Для верификации данной модели была создана реализация самоорганизующейся карты Кохонена размерностью $32 \times 32 \times 3$, что позволяет представить 1024 рукописных элемента длиной в 4 сегмента (3 угла). В качестве входных данных для карты Кохонена были использованы рукописные элементы, полученные после сегментации (см. рис. 3).

Карта Кохонена выделила 4 основных класса примитивов. Повышение точности представления может быть достигнуто увеличением количества сегментов, однако для нас было важно установить, является ли модель, предложенная в главе 2, перцептрально представимой и будет ли множество путей, выделенное такой моделью, разделимо для нейронной сети, что и было доказано в результате проведенных экспериментов.

Анализ результатов классификации показал, что выделяемое моделью множество примитивов разделимо кусочно-линейной функцией и, следовательно, может быть разделено многослойной нейронной сетью (при условии создания обучающей выборки). Для реализации самоорганизующейся карты Кохонена использован пакет математического моделирования SciPy (с пакетом точных вычислений NumPy).

По материалам третьей главы формулируются следующие выводы:

1. Разработанные на основании модели, описанной во второй главе исследования, алгоритмы удовлетворяют предъявляемым к ним требованиям по устойчивости к искажениям.
2. Применение адаптационного слоя позволяет разбить множество кривых, выделяемых в процессе сегментации, на конечное количество классов, таким образом можно считать доказанной перцептральную представимость структурных представлений графических примитивов.
3. Качество распознавания текста с использованием данной модели (в зависимости от аккуратности почерка и настройки модели) составляет от 0.81 до 0.98.

В заключении сформулированы основные выводы и результаты работы:

1. Проведен анализ существующих методов распознавания рукописного текста, изучение применяемых моделей выделения и идентификации графических примитивов — элементов рукописи.
2. Разработаны модель и алгоритм выделения примитивов из слитного рукописного текста. Результаты тестирования алгоритма показали его устойчивость к искажениям, возникающим при сканировании рукописи.
3. Разработана и алгоритмизирована модель скелетизации векторных представлений графических объектов.
4. Выполнена адаптация нейросетевой модели для идентификации графических объектов — элементов рукописного текста. Показана работоспособность такой модели.
5. Персеитронная представимость структурных представлений графических объектов доказана с использованием самоорганизующегося каре Коопена.
6. Создан проблемно-ориентированный комплекс программ, решающий задачи сегментации текста, скелетизации и идентификации элементов рукописи.
7. Качество распознавания текста с использованием данной модели (в зависимости от аккуратности почерка и настройки модели) составляет от 0.81 до 0.98 и более.

Основные публикации по теме диссертации

В журналах, рецензируемых ВАК:

1. Долгова Е.В., Курушин Д.С. Концепция математического моделирования сегментации слитного рукописного текста // Вестник МГОУ. 2011. №1. Серия «Физика и математика». С. 53-57.
2. Долгова Е.В., Курушин Д.С. Адаптация математической модели нейронной сети для системы распознавания слитного рукописного текста // Вестник МГОУ. 2011. №1. Серия «Физика и математика». С. 91-96.
3. Долгова Е.В., Курушин Д.С. Исследование математической модели штрихов рукописного текста. // Вестник МГОУ. 2011. №1. Серия «Физика и математика». С. 97-105.

Другие публикации:

4. Курушин Д.С., Нестерова Н.М., Низамутдинов О.Б. Технология «адаптивной эквивалентности» как способ автоматического анализа текста // Формирование гуманитарной среды и внеучебная работа в вузе, техникуме, школе: Материалы III Всерос. науч. — практ. конф. (22-23 апреля 1999г.): Т. IV / ПГТУ. — Пермь, 1999. — С. 699-700.
5. Курушин Д.С., Нестерова Н.М. О проблеме моделирования понимания в автоматизированных обучающих системах // Визуальная культура XX века и проблемы современного образования: Материалы междунар. молодеж. науч.-практ. конф. Пермь, 10-11 дек. 1999 г. / ПГТУ — Пермь, 1999. — С. 76-78.
6. Курушин Д.С., Нестерова Н.М., Низамутдинов О.Б. Проект обучающей системы с использованием технологий искусственного интеллекта // Теоретические и прикладные аспекты информационных технологий: Сб. науч. тр. — вып. 47 / РосНИИУМС. — Пермь, 1998. — С. 50-54
7. Курушин Д.С., Санникова А.Г., Левченко А.М. О возможности применения эвристических алгоритмов для автоматического распознавания значений счетчиков посещаемости интернет-ресурсов // Вестник ПГТУ. Электротехника, информационные технологии, системы управления. №3/Федер. агентство по образованию, ПГТУ — Пермь: Издательство ПГТУ, 2009. — С. 279-288
8. Курушин Д.С., Баранова И.А., Маркарян А.М., Нестерова Н.М., Серова Т.С. и др. Моделирование процессов понимания. «Семантика текста»: от формализованного представления содержания текста к его автоматическому выделению // С любовью к тексту: [монография]; Восточ. ин-т экономики, гуманитар. наук, управления и права. — Уфа, 2006. — С. 69-74.

9. Курушин Д.С., Нестерова Н.М., Новиков А.И. О возможности автоматического распознавания смысловой внутренней формы текста / Новосиб. гос. ун-т. — Новосибирск, 1999. — С. 114-115.
10. Курушин Д.С., Нестерова Н.М., Новиков А.И. Использование XML-технологий для моделирования процесса понимания текста // 2-ая Всероссийская конференция «Теория и практика речевых исследований» / РАН Ин-т языкознания, Москва, 6-7 дек. 2001 г. — М., 2001. — С. 76-78.
11. Курушин Д.С. О возможности использования нечетких алгоритмов проверки открытых вопросов в автоматизированных системах контроля знаний // Электротехника, ИТ, системы управления. 2009. №3. — С. 84-88.

Подписано в печать 18.10.2011. Формат 60×90/16.
Усл. печ. л. 1,0. Тираж 100 экз. Заказ № 1967/2011.

Издательство

Пермского национального исследовательского
политехнического университета.
Адрес: 614990, г. Пермь, Комсомольский пр., 29, к. 113.
Тел. (342) 219-80-33.