

**СИСТЕМА ON-LINE РАСПОЗНАВАНИЯ РУКОПИСНОГО ТЕКСТА**

С. А. Запрягаев, Э. Э. Киселёв

*Воронежский государственный университет***Поступила в редакцию 12.10.2012 г.**

**Аннотация.** Рассматривается проблема распознавания рукописного текста. В статье предложен алгоритм классификации рукописного символа, основанный на комбинации известных методов. Полученные результаты создают основу для создания программной оболочки, работающей по данному алгоритму.

**Ключевые слова:** распознавание образов, классификация объекта, символ, программа оболочка.

**Annotation.** The problem of handwriting recognition. An algorithm classification of handwritten characters based on a combination of known techniques. These results provide a basis for the creation of software, runs on the algorithm.

**Keywords:** pattern recognition, classification of the object, the symbol, the program envelope.

**ВВЕДЕНИЕ**

Среди традиционных задач распознавания образов, задача распознавания рукописного текста является чрезвычайно сложной в связи с широким диапазоном изменения свойств объектов подлежащих распознаванию. Поэтому набор методов решения такой задачи достаточно широк и не универсален [1].

Типичная задача автоматического распознавания образов формулируется как выделение и обработка конечного числа классов объектов или образов [2]. В процессе распознавания автоматическому устройству, снабженному механизмом восприятия образов, задается некий набор признаков, выделяемый из объекта и служащих для идентификации объекта. В связи с появлением большого класса устройств, для которых перо является основным средством ввода, проблема динамического распознавания рукописного текста является весьма актуальной. Цель настоящей работы состоит в разработке надежных алгоритмов распознавания рукописного ввода во время ввода данных.

**1. ОБЩИЕ ОПРЕДЕЛЕНИЯ**

Для решения поставленной задачи распознавания используем некоторые стандартные определения [3].

*Образ* – это формализованное описание любого объекта, полученное в рамках используемой математической модели объекта. В качестве образа может использоваться скаляр, вектор, матрица, реализация случайной функции, граф.

*Класс образов* – категория, определяющая совокупность объектов, обладающих рядом общих свойств. Образы, принадлежащие к одному классу, близки по своим свойствам, а объекты, принадлежащие к разным классам, имеют существенные различия в их свойствах.

*Признак* – это измеряемая характеристика объекта, которая может использоваться для составления описания образов и классов образов.

*Эталонное описание класса* – это обобщенные априорные сведения и характеристики класса в пространстве используемых признаков. У каждого класса имеется свое эталонное описание.

Задача распознавания (классификации) объекта ставится следующим образом. Имеется некоторый способ кодирования объектов (например, рукописные цифры), принадлежащих заранее известному конечному множеству классов  $C = \{C_1, \dots, C_q\}$ , и некоторое конечное множество объектов (обучающее множество), про каждый из которых известно, какому классу он принадлежит. Необходимо построить алгоритм, способный по любому входному объекту решить

какому классу этот объект принадлежит. Качество распознавания оценивается как вероятность ошибки классификации на другом множестве объектов с заранее известными ответами (тестовом множестве).

Распознавание рукописного ввода – это способность компьютера получать и интерпретировать рукописный ввод. Распознавание текста может производиться так называемым off-line методом для уже написанного на бумаге текста или в режиме on-line посредством считывания движений кончика пера по поверхности графического планшета (рис. 1).

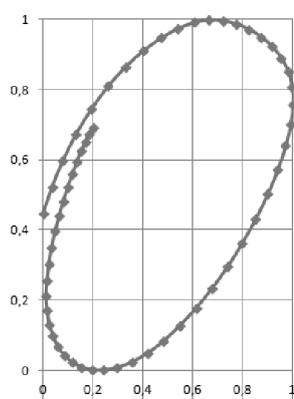


Рис. 1. Арабская цифра «0», введенная пером

Как правило система онлайн-распознавания обычно состоит из:

1. пера, которым пользователь осуществляет ввод;
2. поверхности, чувствительной к касаниям;
3. программного обеспечения, которое интерпретирует движения пера по пишущей поверхности.

Процесс on-line распознавания рукописных символов можно рассматривать как процесс классификации параметрических плоских кривых, определяющих ортогональные координаты

пера на плоскости. Исходно эти кривые имеют вид последовательности точек, которые отбираются цифровым пером в режиме реального времени

$$s = f\{p_i\}, i = 1, \dots, n, \quad (1)$$

где  $p_i = (x_i, y_i)$  – декартовы координаты точек плоскости пера (рис. 2). На рис. 3 представлены абсолютные координаты  $y_i$ , выданные планшетом, в системе координат, в которой в качестве начальной точки выбрана исходная точка написания цифр от 0 до 9. Кроме значения координат точек в различных устройствах могут фиксироваться условные величины давления пера на плоскость, а также наклон пера. По совокупности этих параметров можно построить алгоритм распознавания, который будет эффективно применяться в процессе on-line ввода [4].

Типичный алгоритм распознавания состоит из следующих блоков:

1. Блок выделения отдельного объекта (сегментация).

Данный блок в системе распознавания рукописных символов не является тривиальным и может быть реализован с использованием диаграмм Вороного [5].

2. Блок предварительной обработки ввода данных.

Большинство методов классификации работают с данными, представленными в фиксированном формате и удовлетворяющими определенным требованиям к качеству, размерности, инвариантности и т.д. Однако, этих характеристик, как правило, сложно добиться при рукописном вводе текста. Низкое качество данных обусловлено сочетанием нескольких основных факторов. Одним из них является добавление шума во время оцифровки, как правило, порожденным плохой настройкой цифрового планшета. Другое нарушение качества является следствием индивидуальных особенностей почерка.

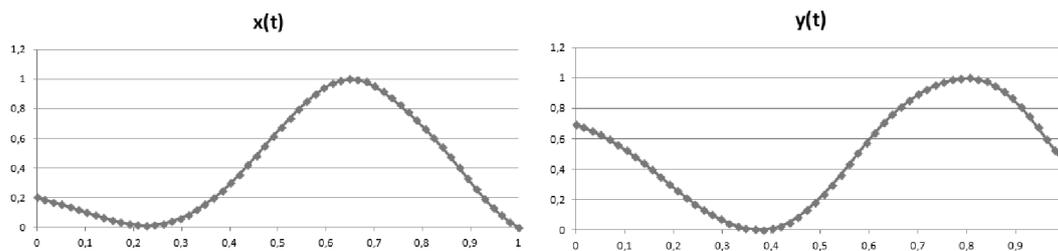


Рис. 2. Параметрические кривые  $x(t)$  и  $y(t)$  для цифры 0, нормированные на единичный интервал по осям

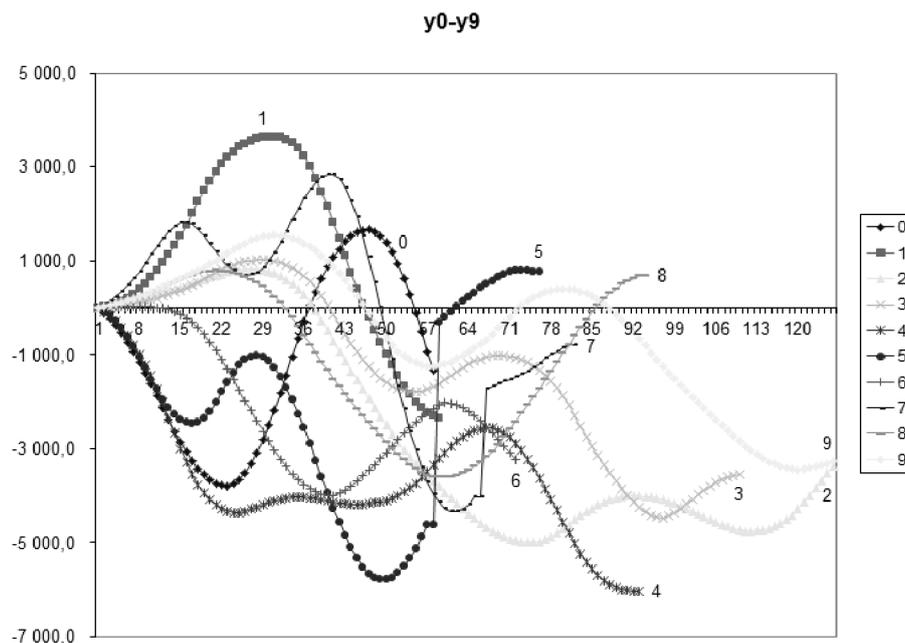


Рис. 3. Абсолютные координаты  $y_i$  в системе координат, в которой в качестве начальной точки выбрана исходная точка написания цифр от 0 до 9

Чтобы преодолеть эти проблемы, используется предварительная обработка, которая включает в себя замену, удаление, изменение порядка, и (или) извлечения данных. Предварительная обработка устраняет шум, нормализует почерк, и уменьшает количество избыточной информации. Это необходимо для того, чтобы исправить произвольные изменения почерка и облегчить кодирование необработанных данных в векторы признаков.

### 3. Блок выделения признаков.

Извлечение признаков – это преобразование входных объектов к единообразному и «компактному» виду, с сохранением значимой с точки зрения используемой классификации информации.

Удобным оказывается представление объекта точкой стандартного евклидова пространства  $R^d$ , принадлежащей некоторому фиксированному компактному (кубу, шару, сфере). Размерность  $d$  должна быть достаточно большой для успешного (в смысле качества) распознавания и достаточно малой для успешного (в смысле скорости) распознавания. Способ извлечения признаков зависит от природы и исходной кодировки объектов и подбирается вручную.

Извлечение признаков – одна из сложнейших задач, решаемых в процессе распознавания рукописного текста. Набор признаков должен

обладать информативностью, достаточностью, но не избыточностью [6].

### 4. Классификатор.

На основе выделенных и нормализованных признаков строится функция  $g(x) = l_0$ , которая является решающим правилом. В зависимости от значения  $l_0$  принимается решение о принадлежности исходного набора признаков к тому или иному классу [6].

Система распознавания рукописного текста не всегда должна содержать описанные блоки, но в среднем, они являются типичными для любого классифицирующего алгоритма [7].

Среди признаков обычно выделяются два основных класса: локальные и глобальные. К локальным относятся признаки, связанные с характеристикой точки  $p_i$  из набора on-line данных, полученные по координатам «соседних» точек из того же набора. Глобальные признаки описывают особенности всего набора данных.

Локальными признаками является, сами координаты точки  $p_i$ , аппроксимация первых и вторых производных в ней, кривизна сэмплизованной кривой  $s$ , вычисленная в точке  $p_i$ . Помимо этих признаков улучшение качества распознавания достигается при использовании таких локальных параметров как: угол  $\theta_i$  поворота в точке  $p_i$ , значение натурального параметра  $L_i$  кривой  $s$  в точке  $p_i$ , которые определены следующим образом:

$$\cos(\theta_i) = \frac{x_{i+1} - x_i}{d(p_i, p_{i+1})}, \sin(\theta_i) = \frac{y_{i+1} - y_i}{d(p_i, p_{i+1})} \quad (2)$$

$$L_i = \frac{1}{L} \sum_{k=1}^{i-1} d(p_k, p_{k+1}), L = \sum_{k=1}^{n-1} d(p_k, p_{k+1}). \quad (3)$$

Здесь  $d(p_i, p_{i+1})$  – евклидово расстояние между точками  $p_i$  и  $p_{i+1}$ ,  $n$  – число точек временной последовательности.

К глобальным параметрам обычно относится расстояние между начальной  $p_0$  и конечной  $p_{n-1}$  точками дискретной кривой  $s$ , позволяющее определить, является ли кривая замкнутой, общая длина  $L$  кривой, «суммарный» угол  $\theta_a$  :

$$\theta_a = \frac{1}{2\pi} \sum_{i=1}^{n-1} \theta_i. \quad (4)$$

При классификации признаков в on-line методах ввода для каждой из кривых  $s$ , входящих в состав некоторого символа или слова, формируется набор «характерных» признаков, который используется в дальнейшем для распознавания вводимой информации [8].

## 2. ТРЕХЭТАПНЫЙ АЛГОРИТМ РАСПОЗНАВАНИЯ

Как отмечено выше, для увеличения эффективности классификации целесообразно использовать не одну методику распознавания, а их комбинацию. Предложенный в настоящей работе алгоритм распознавания рукописных цифр состоит из трех этапов, каждый из кото-

рых, последовательно, сокращает область решений и, в конечном итоге, приводит к единственному результату.

I этап. Подсчет числа экстремумов параметрических кривых.

На первом этапе предложенный алгоритм интерпретирует координаты точек, введенных пером, в виде двух параметрических функций  $x(t), y(t)$  декартовых координат точек ввода данных и анализирует каждую из этих функций отдельно.

Для сужения области решения на первом этапе выделяются локальные признаки в виде количества точек функции  $h(t) \in \{x(t), y(t)\}$ , в которых производная равна нулю. Каждый минимум обозначается «0», а каждый максимум – «1». Таким образом, для каждого введенного образца получим вектор

$$v = [u_{x1}, \dots, u_{xk}, u_{y1}, \dots, u_{yk}], \quad (5)$$

где  $xk$  – число экстремумов функции  $x(t)$ ,  $yk$  – число экстремумов функции  $y(t)$ ,  $u_\alpha \in 0, 1$ . Полученные векторы сравниваются с соответствующими векторами образцов каждой из цифр.

Эксперименты показывают, что один и тот же человек одну и ту же цифру не может написать по одной и той же траектории (реализовать одно и то же состояние) с точностью до количества экстремумов. Поэтому в ходе работы была построена табл. 1, в которой собраны практически все возможные состояния векторов-при-

Таблица 1

Экспериментальные данные о количестве экстремумов для каждой из цифр

0	1	2	3	4	5	6	7	8	9
-2 -2	1 1	2 4	3 4	-2 -4	-3 -3	-2 -2	4 4	4 2	-2 3
	0 0	2 3	3 3	-3 -4	-3 -4	-2 3	4 5	-4 2	3 4
	3 3	2 2	3 -3	-2 -3	-3 -2	3 3	5 5	4 -1	-2 -2
	2 1	2 -3			-3 -1		2 4	4 -2	2 2
		-3 4			-4 -4		3 5		4 3
							2 2		3 3
							-3 4		-2 4
							3 4		-2 5
							2 5		-4 3
							2 6		3 5
							3 6		

знаков, полученные вышеописанным способом. Здесь для краткости первое число в ячейке обозначает количество экстремумов по оси  $x$ , а второе – по оси  $y$  соответственно. Знак «минус» перед числом говорит о том, что первым экстремумом функции является точка минимума. Далее тип экстремума последовательно чередуется. Например, в столбце для символа «0» есть единственная запись  $[-2; -2]$ . Это означает, что цифру «0» можно записать только единственным способом, сигнатура которого будет выглядеть таким образом:  $[-2; -2]_0 \equiv \{0, 1; 0, 1\}$ .

Представленные выше данные практически полностью инвариантны к особенностям почерка человека. Из табл. 1 следует, что часть образцов может быть классифицирована на первом этапе, но другая часть имеет схожие сигнатуры, которые не дают однозначного ответа.

При регистрации величины давления осуществляется фиксация отрыва пера от бумаги при введении символа. В рассматриваемых примерах отрыв фиксируется на цифрах «5» и «7». Как видно из табл. 1, сигнатуры этих чисел не совпадают, поэтому, зафиксировав отрыв на данном этапе распознавания и зная сигнатуру введенного объекта, можно однозначно идентифицировать искомый символ. Таким образом, если отрыва пера не происходит, то цифры «5» и «7» следует исключить из области решений (при условии полного, стандартного написания этих символов). В общем случае эти цифры можно написать и без отрыва пера, что означает ограниченность формируемой системы распознавания.

II этап. Угол поворота геометрического вектора «начало-конец».

Следующий этап распознавания основан на использовании признака, относящегося к классу глобальных. Для его определения нужно иметь информацию обо всей траектории движения пера, а не только об отдельных ее точках. Алгоритм сравнивает координаты точек в начальный  $p_0 = \{x_0 = 0, y_0 = 0\}$  и в конечный  $p_{N-1} = \{x_{N-1}, y_{N-1}\}$  моменты времени ввода информации и строит вектор, соединяющий начало и конец ввода. В зависимости от того, в какой координатной четверти располагается этот вектор (первой, второй, третьей или четвертой), возможен отбор несоответствующих образцов.

Следует отметить, что в выбранной области применения этот этап практически не дает сокращения области решений, а лишь подтверждает первый этап.

III этап. Средние значения координат.

На этом этапе происходит сужение области решений до одного объекта. Абсолютные значения координат точки начала ввода данных  $x_0, y_0$  принимаются за начало отсчета декартовой системы координат. В результате получаем последовательность координат вида:

$$p'_0 = \{0, 0\}, \dots, p'_i = \{x'_i = x_i - x_0, y'_i = y_i - y_0\}, \dots, p'_{N-1} = \{x'_N, y'_N\}. \quad (6)$$

Алгоритм вычисляет среднее значение полученной последовательности координат  $p, q \in \mathbb{R}$  по осям:

$$p = \frac{1}{N} \sum_{i=1}^N x'_i, \quad q = \frac{1}{N} \sum_{i=1}^N y'_i, \quad (7)$$

Действительные числа  $p$  и  $q$  определяют точку  $b = (p, q)$ , на плоскости  $(x, y)$  в выбранной системе координат. Алгоритм измеряет евклидово расстояние между точкой введенного образца и каждой из моделей локализованной области решений:

$$d_i = \sqrt{(p_i - p)^2 + (q_i - q)^2}, \quad (8)$$

где  $i$  принадлежит области допустимых решений, образованной после II этапа. Введенный образ относится к  $i$ -му классу, если вычисленное евклидово расстояние минимально, т.е.  $d_i < d_j$  справедливо для всех  $i \neq j$  [9].

Таким образом, после данного этапа распознавания в области допустимых решений остается один класс, к которому алгоритм относит введенный образ.

Для проверки эффективности решающего правила III этапа был проведен эксперимент. В систему было введено по 5 образцов каждой из цифр. Для каждого из экземпляров по формуле (7) были получены числа  $p$  и  $q$ , и построены точки  $b$ . Результаты представлены на рис. 4.

Набор точек, относящихся к одному и тому же классу, составляют геометрическое множество точек, характерное для написания траектории образца этого класса. Как видно из диаграммы на рис. 4, некоторые из множеств пересекаются, что может создавать проблемы при идентификации введенного символа. На самом деле ухудшения качества распознавания не происходит, так как пересекающиеся множества имеют разные сигнатуры и «расходятся» еще на I и II этапах.

Следует отметить, что подход, реализованный на III этапе распознавания, предполагает, что все точки  $b$  в признаковом пространстве

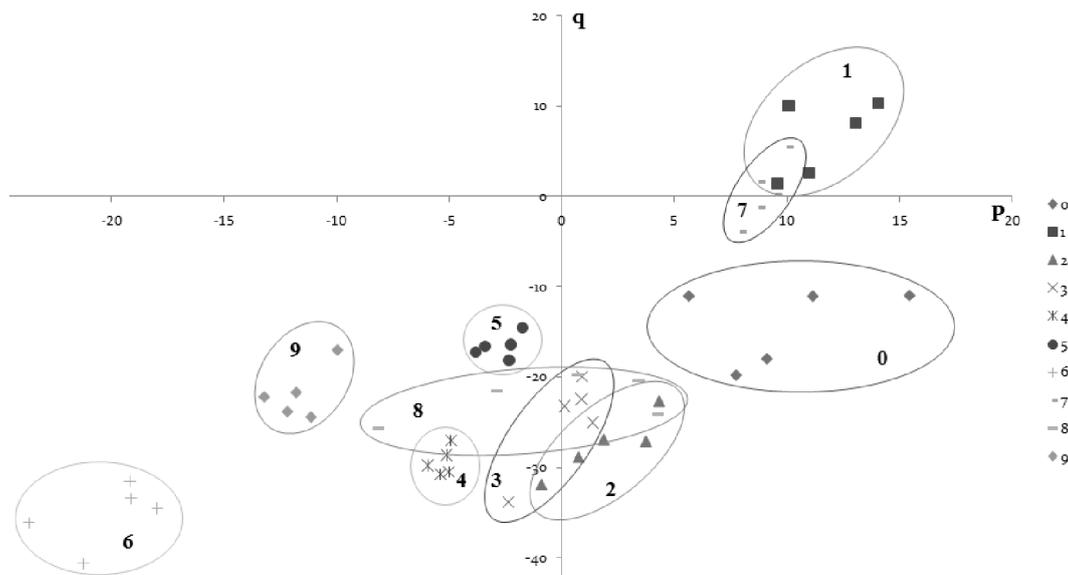


Рис. 4. Эксперимент проверки правильности теории III этапа

располагаются по равномерному закону распределения. Если считать, что точки распределены по иному закону (например, нормальному), то следует учитывать вероятности влияния каждого из классов на искомую точку. Также, необходимо отметить, что III этап, в отличие от I, не инвариантен к особенностям почерка человека. Это значит, что перед работой алгоритма нужно подать обучающую выборку из символов, для которых известна их принадлежность к классу.

### ЗАКЛЮЧЕНИЕ

В настоящей работе создана программная оболочка, работающая в соответствии с предложенным алгоритмом распознавания. В настоящее время областью ее применения являются рукописные арабские цифры.

Средняя эффективность распознавания изолированных рукописных цифр составляет 97%. Точность может снижаться в зависимости от

особенностей почерка пользователя. Подробные результаты экспериментов, проведенных отдельными символами представлены в табл. 2.

Как видно из представленных результатов уровень распознавания достаточный для реального применения.

### СПИСОК ЛИТЕРАТУРЫ

1. Дуда Р. Распознавание образов и анализ сцен. Р. Дуда, П. Харг – М.: «Мир», 1976. – 507 с.
2. Бутаков Б. А. Обработка изображений на ЭВМ / Б. А. Бутаков, В. Н. Островский, И. Л. Фадеев – М.: Радио и связь, 1987. – 240 с.
3. Местецкий Л. М. Математические методы распознавания образов. Курс лекций. – М.: МГУ, 2002–2004 – 85 с.
4. Golubitsky O. and Stephen M. Watt “Improved Character Recognition through Subclassing and Runoff Elections”, Ontario Research Center for Computer Algebra Technical Report TR-09-01.
5. Сегментация рукописных и машинописных текстов методом диаграмм Вороного / С. А. Запря-

Таблица 2

### Эффективность распознавания

Символ	Вероятность распознавания	Символ	Вероятность распознавания
0	98%	5	98%
1	96%	6	100%
2	98%	7	98%
3	96%	8	98%
4	96%	9	96%

гаев, А. И. Сорокин // Вестник Воронежского государственного университета. Серия: системный анализ и информационные технологии – Воронеж, 2010. №1, – С. 160–165.

6. Новикова Н. М. Структурное распознавание образов. Учебно-методическое пособие для вузов. Воронеж: Издательско-полиграфический центр ВГУ, 2008. – 30 с.

7. Суясов Д. И. Магистерская диссертация: Разработка алгоритмов распознавания текста на основе

клеточных автоматов: – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, Кафедра «Компьютерные технологии», 2007.

8. Tapia E. Understanding Mathematics: A System for the Recognition of On-Line Handwritten Mathematical Expressions: Dissertation / E. Tapia. – Berlin, 2004. – 109 p.

9. Ту Дж. Принципы распознавания образов. Дж. Ту, Р. Гонсалес – М.: «Мир», 1978 – 413 с.

**Запрягаев Сергей Александрович** – доктор физико-математических наук, профессор кафедры цифровых технологий Воронежского государственного университета. Тел. (473) 220-82-57. E-mail: zsa@main.vsu.ru

**Zapryagaev S. A.** – Doctor of Physics-math. Sciences, Professor of the dept. of digital technologies Voronezh State University. Tel. (473) 220-82-57. E-mail: zsa@main.vsu.ru

**Киселёв Эдуард Эдуардович** – магистрант кафедры цифровых технологий Воронежского государственного университета. Тел. 8 950 772 58 92. E-mail: vaivolley@yandex.ru

**Kiselyov E. E.** – Graduate student of the dept. of digital technologies Voronezh State University. Tel. 8 950 772 58 92. E-mail: vaivolley@yandex.ru