

УДК 004.93*12

А.В. Афонасенко, А.И. Елизаров

Обзор методов распознавания структурированных символов

Рассмотрены основные методы, используемые при распознавании печатных символов.

Дается подробное описание каждого метода в отдельности.

Также рассмотрены достоинства и недостатки, присущие каждому из методов.

Распознавание структурированных (печатных) символов различных изображений обеспечивает решение ряда научных и прикладных задач при идентификации объектов различной природы. Современные методы распознавания символов используются для решения как типовых задач, например распознавание текста, так и специализированных задач, ориентированных на распознавание символьной информации, нанесенной на поверхность различных объектов. В настоящее время существует достаточно большое количество программ, предназначенных для распознавания текста (например, FineReader, Readiris, ScanSoft OmniPage, CuneiForm и др.). Каждая из этих программ предлагает свою реализацию решения задачи обработки и распознавания изображений. В основном эти программы являются коммерческими, поэтому методы решения задач, заложенные в них, известны только их разработчикам, и практически невозможно определить для каких задач они подходят и какие задачи им не под силу. Кроме того, все эти программы поставляются в виде исполняемых модулей, что делает невозможными анализ работоспособности программ, качества их работы и модификацию используемых ими математических моделей и алгоритмов.

В данной статье рассмотрим наиболее известные и распространенные методы решения задачи распознавания символов. На данный момент выделяют три основных подхода для решения сформулированной задачи: структурный, признаковый и шаблонный. Каждому методу присущи свои достоинства и недостатки. Рассмотрим подробно каждый метод в отдельности.

Шаблонные методы распознавания символов. Первым этапом работы шаблонного метода является преобразование отсканированного изображения в растровое (поточечное). Далее производится его сравнение со всеми имеющимися в базе системы шаблонами. Наиболее подходящим шаблоном считается тот, у которого будет наименьшее количество точек, отличных от исследуемого изображения. Шаблон для каждого класса обычно получают, усредняя изображения символов обучающей выборки. У этих методов достаточно высокая точность распознавания дефектных символов (склеенных или разорванных). Недостаток данного метода — невозможность распознать шрифт, хоть немного отличающийся от заложенного в систему (размером, наклоном или начертанием). Алгоритм, основанный на шаблонном методе, должен заранее знать шрифт, который ему представляют для распознавания. При существующем богатстве печатной продукции в процессе обучения невозможно охватить все шрифты и их модификации. Другими словами, этот фактор ограничивает универсальность таких методов.

Рассмотрим методологию распознавания на конкретном примере шаблонного метода, который был разработан еще в 60-е годы прошлого века [1] и применялся при создании читающего устройства «РУТА 701», он также используется и в настоящее время.

Мерой подобия в данном методе выбран коэффициент подобия изображения символа с обобщенным образом S -го класса, выражаемый следующей формулой:

$$R_s = \sum_{j=1}^n \left(\ln \frac{P_{js}}{1 - P_{js}} \right) x_j + \sum_{j=1}^n \ln(1 - P_{js}),$$

где R_s — коэффициент подобия опознаваемого символа к эталонному изображению S -го класса символов; P_{js} — вероятность появления черного цвета в j -м элементе эталонного изображения S -го класса (выделяют три интервала вероятностей P_{js} : $0,00 \div 0,25$; $0,25 \div 0,75$; $0,75 \div 1,00$); x_j — значение интенсивности, соответствующей j -му элементу распознаваемого символа. Изображение символа отождествляется с эталонным классом, давшим максимальный коэффициент подобия R среди всех R_s .

Структурные, или топологические, методы распознавания хранят информацию не о поточечном написании символа, а о его топологии. Иными словами, эталон содержит информацию о взаимном расположении отдельных составных частей символа [2, 3]. Ясно, что при этом становится неважным размер распознаваемой буквы и даже шрифт, которым она напечатана. Но основной проблемой структурных методов распознавания остается идентификация знаков, имеющих дефекты (например, разрыв линии или слияние соседних линий). Рассмотрим методологию распознавания на конкретном примере [3].

Распознаваемый символ подвергается процедуре скелетизации (утоношению). Процедура скелетизации и ее использование в системах распознавания текста давно изучается разными авторами, ей посвящена многочисленная литература [4–11].

Каждый полученный контур скелетного представления описывается в виде последовательного набора особых точек и так называемого цепного кода, состоящего из точки привязки, числа кодов и массива направлений из очередной точки на следующую точку. Особые точки – это концевые точки и точки ветвления (триоды) т.е. точки, соседи которых образуют не менее трех связных областей. В примере, представленном на рис. 1 образ, обладает двумя внутренними контурами, одной концевой точкой и тремя триодами.

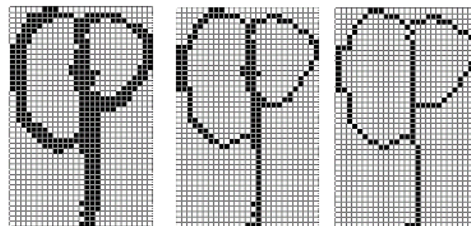


Рис. 1. Скелетизация образа, состоящего из одного внешнего и двух внутренних контуров

В полученном описании производится огрубляющая предобработка, состоящая в удалении коротких линий, объединении близких триодов, уничтожении малых внутренних контуров. Для внешнего контура находится его тип или топологический код. Для этого контур записывается в виде последовательного набора номеров особых точек, соответствующих обходу по часовой стрелке. Затем с помощью перенумерации особых точек и изменения начала контура делается попытка отождествления контура с одним из основных типов.

Кратко опишем признаки, используемые при распознавании по полученному скелетному представлению. Для каждой особой точки скелетного представления вычисляются следующие топологические признаки:

- нормированные координаты особой точки (вершины графа);
- длина ребра до следующей вершины в процентах от длины всего графа;
- нормированное направление из данной точки на следующую особую точку;
- нормированное направление входа в точку, выхода из точки (для триодов эти значения различаются, для точек индекса «1» совпадают с точностью до знака);
- кривизна дуги, точнее «левая» и «правая» кривизна дуги, соединяющей особую точку со следующей вершиной (кривизна слева и справа). Кривизна вычисляется как отношение максимального расстояния от точек дуги (находящихся соответственно слева/справа от прямой) до прямой, соединяющей вершины, к длине отрезка, соединяющего те же вершины.

На рис. 2 условно показаны некоторые из топологических признаков. Граф имеет 4 особые точки: a_0, a_1, a_2, a_3 . При обходе графа по маршруту $a_0 \Rightarrow a_1 \Rightarrow a_2 \dots$ в вершине a_1 условно показаны следующие признаки: вектор r_1 – направление входа в точку, вектор r_2 – направление выхода из точки, вектор r_3 – глобальное направление на следующую особую точку. Двухнаправленный вектор h показывает величину «левого» отклонения дуги (a_1, a_2) от прямой; «правое» отклонение равно нулю.

Как видно из приведенного описания, число признаков равняется восьмикратному числу вершин. Оно различается для разных топологических кодов, и признаки с одинаковым номером для разных топологических кодов могут иметь разный смысл.

Для некоторых кодов число особых точек и, соответственно, число топологических признаков слишком мало. Так, для кода, соответствующего символу «0», топологических признаков вообще нет, так как нет ни одной особой точки. Поэтому могут вычисляться и использоваться следующие дополнительные признаки:

- размеры и положение компонент и дыр;
- «черная» и «белая» ширина верхней половины символа;
- модифицированные прямые прогибы.

Прогибы вычисляются, как расстояния от точек скелетного представления до выпуклой оболочки построенного представления. Дополнительно запоминается положение точек максимального прогиба. Для некоторых топологических кодов число топологических признаков может быть достаточно велико, что может потребовать слишком большого набора эталонов для обучения, поэтому в ряде случаев в распознавании используется часть признаков.

Обучение метода состоит в построении деревьев распознавания для каждого из определенных заранее (вручную или автоматически) топологических кодов. Рассмотрим простую процедуру построения деревьев распознавания, приносящую неплохие результаты.

Для каждого топологического кода в обучающем множестве проводится отбор всех имен символов, имеющих достаточно большое представительство. Для каждого имени проводится анализ имеющихся значений признаков $p[i]$, $i = 1, \dots, N$, где N – число признаков для текущего топологического кода. Обозначим A_i – множество имеющихся значений для признака i для символов с именем A . Тогда для каждого i , $0 < i \leq N$, A_i представляется в виде

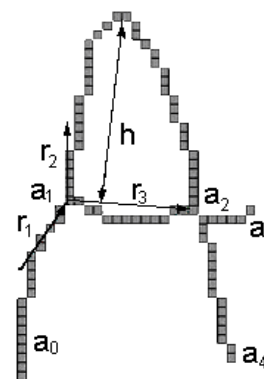


Рис. 2. Примеры скелетных признаков

$$A_i = A_{i1} \cup A_{i2} \cup \dots \cup A_{im}, \text{ где } A_{ij} = \{x \in R : s_{ij} \leq x \leq e_{ij}\},$$

im – различно для каждого i и для каждого A .

Далее производится поиск конфликтов. Если для некоторых символов A , B значения признаков пересекаются, т.е. $A_i \cap B_j \neq \emptyset$ для $\forall i, i = 1 \dots N$, тогда проводится попытка разрешить конфликт. Делается попытка найти некоторый наилучший для разбиения (наиболее дисперсионный) признак j , выбрать точку деления этого признака k и разбить множество A на два непересекающихся подмножества A' , A'' таким образом, что

$$\begin{aligned} A &= A' \cup A'', \\ A_j &= A_j' \cup A_j'', \end{aligned}$$

где $A_j' = A_{j1} \cup A_{j2} \cup \dots \cup A_{jk}$, $A_j'' = A_{jk+1} \cup A_{jk+2} \cup \dots \cup A_{jm_j}$, $0 < k < m_j$.

Затем процедура повторяется, т.е. для каждого A' , A'' проводится построение областей значений признаков и поиск конфликтов с разноименными символами с возможной дальнейшей разбивкой множеств A' , A'' и т.д. Конечно, все конфликты разрешить удастся не всегда, поэтому при распознавании в ряде случаев будет выдаваться не одна альтернатива, а несколько. Оценки результирующих альтернатив будут зависеть как от значений признаков (топологических и нетопологических), так и от представительности конфликтующих символов в обучающем множестве.

Таким образом, распознавание является древовидным, текущее дерево распознавания выбирается с помощью топологического кода.

Если символ после прохода по дереву распознавания остался нераспознанным, делается попытка улучшения изображения с помощью следующих операций:

- склеить концы линий по направлениям. Для этого рассмотреть направления всех концевых дуг скелетного представления, и если направления каких-либо линий сходны и указывают друг на друга, можно попытаться их соединить – возможно, это была сплошная линия;

- склеить точки скелета, находящиеся на минимальном расстоянии одна от другой;

- отбросить самую короткую линию (дугу графа). Излишние короткие дуги (линии) нередко возникают при рукописном написании;

- игнорировать малую компоненту. Если в результате анализа изображения оказалось, что имеется несколько компонент, не исключено, что просто добавились помехи;

- удалить внутренний контур. Если один или несколько из внутренних контуров малы, не исключено, что они появились в результате дефектов написания или сканирования, а не как характерные элементы символа.

Если модификации представления невозможны, то образ остается нераспознанным. Иначе выполняется модификация изображения с возможной повторной скелетизацией в зависимости от типа модификации.

Характеристики метода распознавания по скелетным признакам [9] на тестовой последовательности печатных символов таковы: точность $\approx 95,6\%$, быстродействие – 1000 образов/с.

Необходимо отметить, что в настоящее время программа распознавания на основе скелетизации не используется в системе сама по себе [3], а комбинируется с иными методами распознавания, основанными на совершенно иных принципах, в первую очередь в комбинации с нейронной сетью, построенной на других признаках символов. Такая комбинация различных методов распознавания рукописных символов приводит к хорошим итоговым результатам распознавания.

Признаковые методы базируются на том, что изображению ставится в соответствие N -мерный вектор признаков. Распознавание заключается в сравнении его с набором эталонных векторов той же размерности. Задача распознавания, принятия решения о принадлежности образа тому или иному классу, на основании анализа вычисленных признаков, имеет целый ряд строгих математических решений в рамках детерминистического и вероятностного подходов [12, 14]. В системах распознавания символов чаще всего используется классификация, основанная на подсчете евклидова расстояния между вектором признаков распознаваемого символа и векторами признаков эталонного описания. Тип и количество признаков в немалой степени определяют качество распознавания. Формирование вектора производится во время анализа предварительно подготовленного изображения. Данный процесс называют извлечением признаков. Эталон для каждого класса получают путем аналогичной обработки символов обучающей выборки.

Основные достоинства признаковых методов – простота реализации, хорошая обобщающая способность, хорошая устойчивость к изменениям формы символов, низкое число отказов от распознавания, высокое быстродействие. Наиболее серьезный недостаток этих методов – неустойчивость к различным дефектам изображения. Кроме того, признаковые методы обладают другим серьезным недостатком – на этапе извлечения признаков происходит необратимая потеря части информации о символе. Извлечение признаков ведется независимо, поэтому информация о взаимном расположении элементов символа утрачивается.

Четких правил отбора признаков не существует, поэтому методы от разных разработчиков систем распознавания оперируют различными наборами признаков. Рассмотрим несколько методов создания наборов признаков, используемых в настоящее время.

В качестве признаков изображений, например, можно использовать последовательности m_k , предложенные Е.С. Абрамовым [14], которые строятся по описанному ниже алгоритму.

Пусть на изображении A имеется объект O . Рассмотрим N радиус векторов r_k с началом в центре изображения C и концом на границе изображения, отстоящих друг от друга на угловое смещение $2\pi/N$. Вдоль каждого вектора существуют точки изображения, принадлежащие объекту O , и точки, не принадлежащие этому объекту. Эти точки формируют отрезки B_{kl} , $1 \leq l \leq N_k$, которые представляют собой «куски» объекта O вдоль вектора r_k . Для каждого из этих отрезков мы можем определить длину L_{kl} и его расстояние от центра изображения d_{kl} . Величину, определяемую выра-

жением $m_k = \sum_{l=1}^{N_k} d_{kl} L_{kl}$, будем называть моментом изображения вдоль вектора r_k . Проведя нормировку $\widehat{m}_k = \frac{m_k}{\sum_k m_k}$, можно обеспечить инвариантность момента изображения к масштабированию. Это

позволяет обеспечить распознавание объектов на изображениях независимо от их масштаба.

Рассмотрим основную идею формирования признаков системы распознавания на основе функции инвариантных к сдвигу, вращению, предложенную Д.В. Горловым [15]. Для каждого символа вычисляются параметры (высота, ширина, количество точек) и признаки.

Процедура вычисления признаков состоит из пяти этапов.

1. Для контура текущего символа вычисляется значение центра тяжести.
2. Для текущей точки окружности u с центром в центре тяжести и радиуса $r = \text{const}$ вычисляется значение $\Phi_Q = P_Q(\theta^2(x, u))$, где x – текущая точка контура символа; Q – количество признаков; P_Q – полином Лежандра; $\theta(x, u) = \rho(x, u)/r$; $\eta = \rho^2(x, u)$; η – евклидово расстояние на плоскости.

3. Для всех точек контура символа вычисляется значение $\Phi'_Q = \sum \Phi_Q$.

4. Для всех точек окружности находится характерная точка, в которой значение Φ'_Q максимально.

5. Для этой точки окружности вычисляются значения $\Phi'_{Q-1} \dots \Phi'_1$.

В качестве признаков используются значения вида $\Phi_Q = \frac{1}{k} \sum_{i=1}^k P_Q(\theta^2(x_i, u))$, где k – количество

точек контура символа. Множитель $1/k$ введен с целью уменьшения влияния дискретности изображения. В качестве признаков выбираются ортогональные полиномы Лежандра:

$$P_1(\eta) = \frac{2}{b-a}(\eta - a) - 1, \quad P_2(\eta) = \frac{1}{2}(3P_1^2 - 1), \quad P_3(\eta) = \frac{1}{2}(5P_1^3 - 3P_1).$$

Из выкладок в [15] следует, что $a = 0$, $b = 2$.

Распознавание осуществляется следующим образом. Для контура символа вычисляются параметры (высота, ширина, количество точек) и признаки.

Последовательно контур символа сравнивается с объектами, полученными при обучении. При этом вычисляется мера близости в виде

$$L = \sum_{i=1}^Q |\Phi_i^* - \Phi_i|,$$

где Φ_i^* – признак объекта, полученный при обучении; Φ_i – признак объекта, полученный при распознавании; Q – количество признаков.

В заключение рассмотрим «метод зон», использующийся для распознавания символов автоматически считываемых с паспортов и виз [16]. На этапе обучения используется полный эталонный набор бинарных изображений символов, имеющих размеры 100×135 пикселей. При этом для каждого эталонного изображения в наборе выполняются следующие действия:

- 1) определение минимального прямоугольника, содержащего все «черные» пиксели изображения («рамки»);
- 2) равномерное разбиение рамки на $N \times M$ прямоугольных зон;
- 3) подсчет числа пикселей, принадлежащих каждой зоне;
- 4) формирование вектора зонного описания, состоящего из значений числа «черных» пикселей для каждой зоны, нормированных путем деления на суммарное число «черных» пикселей всего изображения.

При анализе каждого распознаваемого символа производятся следующие операции:

- 1) формирование вектора зонного описания;
- 2) формирование вектора расстояний;
- 3) классификация символа на основе вектора расстояний.

Формирование вектора расстояний предполагает последовательное вычисление евклидовых расстояний между вектором зонного описания тестируемого символа и векторами зонных описаний каждого эталона в списке.

Проведенные исследования показали, что для размера изображений символов порядка 10×15 пикселей оптимальным числом зон будет $N \times M$, где $N = 5$ и $M = 5$. При таком выборе зонного разрешения использованный способ нормирования вектора зонного описания позволил обеспечить устойчивость данного метода распознавания к изменению толщины штрихов символа вследствие неустойчивости яркостных характеристик изображения.

Использование шаблонного метода распознавания символов подходит в том случае, если исходный документ типографского качества (достаточно крупный шрифт, отсутствие плохо напечатанных символов или исправлений). Использование данного метода в этом случае позволяет получить достаточно быстрые алгоритмы распознавания.

При заполнении налоговых деклараций, при проведении переписей населения и т.д. используются различного вида бланки с полями. В данном случае сложность состоит в том, что необходимо распознавать символы, написанные от руки, довольно сильно различающиеся у разных людей. Поэтому наиболее оптимальным с точки зрения точности распознавания является структурированный метод распознавания, так как в данном методе используется распознавание символов по наличию в них определенных структурных элементов (отрезков, колец, дуг и др.).

В случаях распознавания изображений маркировки на поверхностях различных объектов, полученных с помощью систем автоматизированного ввода информации через различные типы цифровых фото- и видеокамер, например задача распознавания автомобильных номеров, символы подвергаются как искажениям случайными помехами, так и таким сложным, как аффинные и проективные. В данном случае будет наиболее оптимальным сочетание признакового и структурного методов распознавания.

Для быстрого порождения предварительного списка гипотез используется признаковый метод классификации. Типичная процедура классификации состоит в вычислении степени близости между входным изображением и известными системе классами изображений. В качестве ответа выдается список классов, упорядоченный по степени близости, т.е. фактически выдвигается ряд гипотез о принадлежности объекта тому или иному классу. Далее для целенаправленной проверки выдвинутых гипотез используется структурный метод распознавания.

Работа выполнена при поддержке РФФИ, проект № 06–08–00751.

Литература

1. Багдонас А. Читающее устройство «РУТА 701» / А. Багдонас, Р. Жлабис, А. Кведаравичус, Р. Нашлюнас, П. Швагждис, П. Яшинкас // Автоматизация ввода письменных знаков в электронно-вычислительные машины: доклады науч.-техн. совещания. – Вильнюс, 1968. – С. 96–121.
2. Фу К. Структурные методы в распознавании образов: Пер. с англ.; под ред. М.А. Азейрмана. – М.: Мир, 1977.
3. Котович Н.В. Распознавание скелетных образов / Н.В. Котович, О.А. Славин. [Электронный ресурс]. Режим доступа: <http://ocr.ai.narod.ru/skeletrecognize.html>
4. Pavlidis T. Algorithms for Graphics and Image Processing. Computer Science Press, Rockville, MD, 1982.
5. Lam L., Lee S.W., Suen C.Y. Thinning Methodologies: A Comprehensive Survey. IEEE Trans. Pattern Analysis and Machine Intelligence. – 1992. – Vol. 14. – P. 869–885.
6. Plamondon R., Suen C.Y., Bourdeau M., Barriere C. Methodologies for Evaluating Thinning Algorithms for Character Recognition. // J. Pattern Recognition and Artificial Intelligence, special issue thinning algorithms. – 1993. – Vol. 7. № 5. P. 1247–1270.
7. Smith S.J., Bourgoin M.O., Sims K., Voorhees H.L. Handwritten character classification using nearest neighbor in large databases // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1994. – Vol. 22, № 9. – Sept. P. 915–919.
8. Wakahara T. Shape machine using LAT and its application to hand-written character recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1994. – Vol. 16, № 6. – June. P. 618–629.
9. Lam L., Suen C.Y. An Evaluation of Parallel Thinning Algorithms for Character Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1995. – Vol. 17, № 9. P. 914–919.
10. Plamondon R., Srinari S. On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2000. – Vol. 22, № 1. – January.
11. Щепин Е.В., Непомнящий Г.М. К топологическому подходу в анализе изображений // Геометрия, топология и приложения: Межвуз. Сб. науч. трудов. М.: Мин. высшего и средн. спец. образ. РСФСР, Московский институт приборостроения, 1990. – С. 13–25.
12. Дуда Р., Харг П. Распознавание образов и анализ сцен: / Пер. с англ.; под ред. Р.Л. Стефанюка. – М.: Мир, 1976.
13. Фукунага К. Введение в статистическую теорию распознавания образов: Пер. с англ. – М.: Наука, 1979.
14. Абрамов Е.С. Моделирование систем распознавания изображений (на примере печатных текстов): Дис. ... канд. техн. наук: 05.13.01: – М., 2006. – 234 с.

15. Горлов Д.В. Распознавание изображений на основе признаков, инвариантных к сдвигу, вращению, масштабированию: Автореф. дис. канд. техн. наук: – Красноярск, 2002. – 20 с.

16. Андреев С.В. Алгоритмическое обеспечение прототипа устройства считывания паспортов и виз / С.В. Андреев, А.В. Бондаренко, В.И. Горемычкин, А.В. Ермаков, С.Ю. Желтов. [Электронный ресурс]: [научная электронная библиотека ИПМ РАН].

Режим доступа: http://www.keldysh.ru/papers/2003/prep46/prepr2003_46.html.

Афонасенко Анна Владимировна

Томский государственный университет, аспирант радиофизического факультета

Эл. почта: anuta8181@mail.ru

Елизаров А.И.

ГОУ ВПО Томский государственный университет систем управления и радиоэлектроники

A.A. Afonassenko, A.I. Elizarov

Structure symbol recognitions method review

Basic method review using by print symbol recognition. Each methods analyzing and show advantage and shortage.
