

РАСПОЗНАВАНИЕ СИМВОЛОВ НА БАЗЕ НИЗКОУРОВНЕВЫХ МОДЕЛЕЙ ОПИСАНИЯ ГРАФИЧЕСКИХ ИЗОБРАЖЕНИЙ

Д. Васин¹, М. Ершов¹

¹ НИИ Прикладной математики и кибернетики
Нижегородского государственного университета им. Н.И. Лобачевского

Аннотация

В докладе рассмотрен алгоритм формирования эффективных признаков и решающих правил распознавания символов на низкоуровневой структурированной модели описания бинарного растрового изображения

Одной из основных проблем, возникающих при решении задачи распознавания образов, и существенно влияющих, как на эффективность распознавания, с одной стороны, так и на его качество – с другой, является выбор системы эффективных признаков классификации.

В работе рассматривается формирование решающих правил в задаче распознавания символов, находящихся в произвольной ориентации по отношению к горизонту, на низкоуровневой структурированной модели описания бинарного растрового изображения (БРИ) [1].

Низкоуровневая структурированная модель описания БРИ. Непроизводным элементом модели является штрих. Формально штрих – одномерный кластер связанных пикселей заданного цвета, т.е. представляется набором чисел $S=(t, N, K, C)$, где t – номер строки растра; N, K – координата начала и конца совокупности подряд идущих пикселей одного цвета вдоль линии растра; C – код цвета связанных пикселей, образующих штрих.

Для бинарных изображений штрих описывается тройкой типа $S=(t, N, K)$, где t – номер строки растра; N, K – координаты начала и конца штриха.

Совокупность штрихов для заданной строки растра t определяет строку штрихов $SH=\{S_i\}(i=1,2,\dots,m)$, где S_i – штрих.

Описание растра с помощью набора строк штрихов является однозначным. Отметим, что штриховой формат представления и хранения растровых данных является в 10 – 20 раз более компактным, по сравнению с пиксельным.

Топологические свойства штриха $S=(t, N, K)$ определяются через его связность со штрихами смежных строк [1]. Два штриха S_t и S_{t-1} смежных строк t и $t-1$ связаны, если $(N_{t-1} \geq N_t) \wedge (N_{t-1} \leq K_t) \vee (K_{t-1} \geq N_t) \wedge (K_{t-1} \leq K_t)$. Связный штрих – это набор $H=\{S, sw_p, sw_s\}$, sw_p, sw_s – количество связанных штрихов со штрихами смежных строк, а t – номер текущей строки штрихов растра. Тогда $sw_p=0$, если в предыдущей строке $p=t-1$ нет ни одного штриха, для которого выполняется условие $(N_p \geq N_t) \wedge (N_p \leq K_t) \vee (K_p \geq N_t) \wedge (K_p \leq K_t)$,
(1)

иначе, $sw_p = \mu_p$, μ_p – кратность выполнения условия (1), т.е. количество связанных штрихов предыдущей строки.

Аналогично, $sw_s=0$, если в следующей строке $s=t+1$ нет ни одного штриха, для которого выполняется условие: $N_s \geq N_t \wedge N_s \leq K_t \vee K_s \geq N_t \wedge K_s \leq K_t$,
(2)

иначе, $sw_s = \mu_s$, μ_s – кратность выполнения условия (2), т.е. количество связанных штрихов последующей строки.

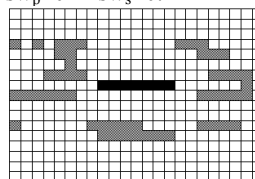
Здесь N, K – координаты начала и конца исследуемого на связность штриха, N_t, K_t – координаты начала и конца текущего штриха предыдущей смежной строки штрихов,

N_s, K_s – координаты начала и конца текущего штриха следующей смежной строки штрихов.

По суперпозиции значений sw_p и sw_s можно ввести следующую классификацию графических ситуаций. В текущей строке:

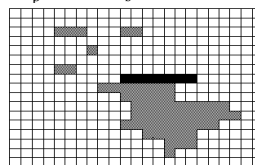
Штрих изолированный (ШИ):

$$sw_p=0 \wedge sw_s=0.$$



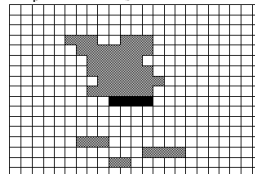
Штрих начала растрового объекта (ШН):

$$sw_p=0 \wedge sw_s=1$$



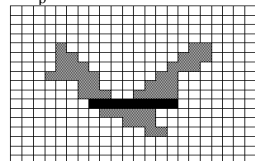
Штрих конца растрового объекта (ШК):

$$sw_p=1 \wedge sw_s=0$$

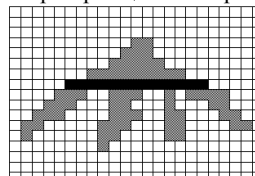


Штрих слияния растровых объектов (ШС):

$$sw_p > 1$$

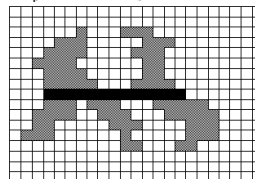


Штрих расщепления растровых объектов (Шр): $sw_s > 1$



Штрих слияния и расщепления растровых объектов (ШСр):

$$sw_p > 1 \wedge sw_s > 1.$$



Растровый простой объект РПО – кластер связанных штрихов, не содержащий графических ситуаций типа слияния и расщепления, то есть для любого штриха кластера выполняется условие

$$sw_p < 2 \wedge sw_s < 2. \quad (3)$$

Растровый составной объект РСО – кластер связанных штрихов, для каждого штриха которого выполняется условие

$$sw_p > 0 \vee sw_s > 0, \quad (4)$$

и для любых двух элементов (штрихов или пикселей) найдется хотя бы одна соединяющая их пиксельная траектория, состоящая из связанных смежных элементов.

Предложенная иерархия моделей описания РИГД на нижних уровнях позволила перейти к структурированному описанию РИГД. При этом РИ представляется совокупностью связанных растровых компонент (СРК) – РПО, РСО, узловых штрихов (Шр, Шс) и штрихов начала и конца объекта (Шн, Шк).

Алгоритм построения эффективных признаков и решающих правил классификации символов

В качестве рассматриваемых символов были выбраны заглавные буквы русского алфавита А – Я, цифры 0 – 9 и знаки препинания – точка, запятая, прямой и обратный слеш.

Примеры обучающих символов:

АБВГДЕЁЖЗИЙКЛ
МНОПРСТУФХЦЧ
ШЩЪЫЬЭЮЯ
1234567890,./\

Для каждого символа алфавита была сформирована матрица связанных штрихов и фиксировались следующие признаки:

1. Число внутренних областей (ВО).
2. Отношение аспекта (отношение ширины к высоте) (Е).
3. РПО, для которых вычисляются:
 - количество штрихов, входящих в РПО;
 - средняя длина штриха.
4. Процент заполненности – соотношение пиксельной площади символа к площади описывающего прямоугольника.
5. Число “вертикальных” РПО (ВРПО). РПО признается “вертикальным”, если средняя длина штриха более 70% высоты символа. Уровень 70% высоты символа как признак вертикальности выбран на основе эмпирических наблюдений.
6. Особые штрихи (Шн, Шк, Шр, Шс).

Приведем фрагмент автоматически сформированной таблицы значений вышеуказанных признаков (см. таблицу).

Символ	ВО	Е	Число ВРПО	% заполн.	Число Шс/Шр
7	0	0.52-0.65	0	21.5-26.6	1/0
3	0	0.59-0.64	0		2/1
Д	1	0.71-0.89	1		2/2
Ю	1	1.35-1.45	1		2/2
Ф	2	1-1.12	1		2/2
В	2	0.8-0.92	1		3/2

Алгоритм подсчета числа внутренних областей

Пусть имеется изображение с объектом, представленным единственной связной компонентой. В процессе работы на изображении находятся “особые” точки, хранящиеся в списке S.

1. Просматривая штриховой файл, находим штрихи расщепления.

2. Пусть штрих расщепления находится в *i*-столбце. Просматриваем все связанные с ним штрихи в столбце (*i*+1), для каждого $H = (t, N, K)$ добавляем в список S особые точки (*i*+1, *N*-1), (*i*+1, *K*+1).

3. Просматриваем все особые точки (*x*, *y*) из списка.

3.1. Если координаты точки выходят за рамки матрицы Р, то удаляем её из списка как несуществующую

3.2. Запускаем поиск в ширину со стартом в (*x*, *y*). При этом *цвет* 1 – белый, *цвет* 2 – любой дополнительный цвет (отличный от черного и белого).

3.3. Проверяем все особые точки из S, следующие по списку за точкой (*x*, *y*).

Если они окрашены в *цвет* 2, то удаляем их как лишние.

3.4. Просматриваем граничные точки всего изображения $[P_{ij}], [P_{Nj}], [P_{i1}], [P_{iM}]$ (*i* = 1,...,N; *j* = 1,..., M). Если хотя бы одна из них будет закрашена в *цвет* 2, то значит точка (*x*, *y*) не является точкой, принадлежащей внутренней области объекта (замкнутая область белого цвета). Удаляем её из списка S.

4. Количество элементов в списке S равно количеству внутренних областей в объекте.

Алгоритм автоматического распознавания символов

Разработан на основе иерархической модели, в узлах которой производится вычисление указанных признаков.

Приведем примеры решающих правил для некоторых символов шрифта «Times New Roman». В записи решающих правил выражение вида Ш₁ = ξ означает, что рассматриваемый символ содержит ξ штрихов типа Ш₁ ∈ {Шн, Шк, Шр, Шс}.

1) ВО = 2, это “В” ∨ “Ф”. Если Е < 0.95, то это “В”, иначе “Ф”.

2) ВО = 1 ∧ Е ∈ [0.0-0.66], это “0” ∨ “4” ∨ “6” ∨ “9”: если Шс = 1 ∧ Шр = 2 – это “6”, если Шс = 2, ∧ Шр = 1 – это “9”. Пусть площадь внутренней области S₀, а площадь описанного вокруг буквы прямоугольника S_n. Если (S₀/S_n) < 0.2, то это “4”, иначе “0”.

3) Если ВО = 1 ∧ Е ∈ [0.71-0.89] ∧ Шс=2 ∧ Шр=1, это “Б” ∨ “Р” ∨ “Ъ”. Для их распознавания поворачиваем символ на 90 градусов. Далее: если Шс = 2 ∧ Шр = 1, это “Б”; если Шс = 1 ∧ Шр = 1, это “Р” ∨ “Ъ”. Если Шс находится в первой трети повернутого символа, то это “Ъ”, если в последней трети – то это “Р”.

4) Если ВО = 0 ∧ Е ∈ [0.65-0.87] ∧ ВРПО=1, это “Г” ∨ “Е” ∨ “Т” ∨ “Ч”: если ВРПО находится во второй половине символа, то это “Ч”, если в первой половине символа, то поворачиваем символ на 90 градусов, и если (Шс+Шр)=1, это “Г”, если (Шс+Шр)>1, это “Е”, если ВРПО расположен в середине символа, то это “Т”.

В ходе анализа работы распознавания с помощью введенных выше решающих правил было отмечено следующее:

- исходные эталоны были расположены горизонтально с минимальным углом отклонения от горизонта, имели обычную ориентацию (не были повернуты на 90 или 270 градусов, не были перевернуты на 180 градусов);

- при малом угле наклона текста по отношению к горизонтальной линии (до 2°) значения признаков (особые штрихи, вертикальные ПРО, аспект и т.д.) не претерпевают значительных изменений. Измеряемые значения находятся в допустимых рамках, принимаемых решающими правилами;

- при большем наклоне текста (ближе к 5° и более) качество распознавания резко падает. Наиболее чувствительными к повороту являются признаки количества вертикальных ПРО, количества штрихов слияния/расщепления, взаимное расположение особых штрихов, отношение аспекта. Наиболее устойчивым, ожидаемо, является число внутренних областей у символа;
- при наклоне даже эталонных символов на $\pm 5^\circ$ правильно распознано было около трети эталонов.

В связи с этим возникает необходимость корректировки наклона текста.

Определим основные требования к алгоритму определения угла наклона текста к горизонту:

- быстродействие (время работы алгоритма не должно зависеть от контента изображения);
- точность (малая величина систематической ошибки);
- корректная работа с полутоновыми и, в большинстве случаев, с цветными графическими документами;
- угол наклона может быть определён, как локально – для некоторой части изображения, так и для всего документа в целом;
- вместе с углом наклона, метод должен выдавать степень доверия к своему ответу, либо оценивать погрешности измерения.

В обзоре [2] предлагается несколько различных методов определения угла наклона документа. Разработанный алгоритм комбинирует некоторые из них и адаптирован к штриховому формату, что уменьшает количество необходимых вычислений. Для оценки угла наклона текста используем вертикальные проекционные гистограммы [3], представляющие собой одномерный массив размерности, равной числу строк раstra N. Каждая ячейка гистограммы хранит количество чёрных пикселей в соответствующей строке изображения.

Пусть имеется бинарное растровое изображение P, размером N * M пикселей.

$$P[x, y] = \begin{cases} 1, & \text{если пиксель } (x, y) - \text{чёрный} \\ 0, & \text{если пиксель } (x, y) - \text{белый} \end{cases}$$

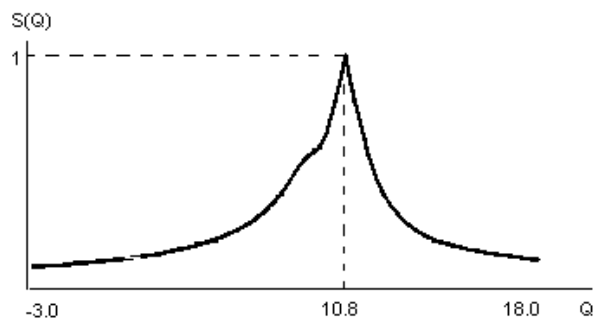
Пусть $S(i, Q)$ – суммарное количество чёрных пикселей в строке раstra с номером i при повороте исходного

$$\text{изображения на угол } Q: S(i, Q) = \sum_{j=0}^{j=M(Q)} P(i, j).$$

Вычисляем суммарную функцию угла наклона

$$Q: S(Q) = \sum_{i=0}^{i=N(Q)} S^2(i, Q)$$

Пример нормированной по максимальному значению функции S(Q), полученной на реальном графическом документе, повернутом на угол около 10° .



На базе разработанных алгоритмов создано оригинальное программное обеспечение.

В результате практической апробации созданного программного обеспечения было выявлено, что при горизонтальном расположении текста правильно распознается 96 – 98% символов, однако по мере увеличения угла поворота снижается до 90 – 93%.

Работа выполнена при финансовой поддержке РФФИ – проекты №13-07-00521 и № 13-07-12211

ССЫЛКИ

1. Васин Д.Ю., Васин Ю.Г., Громов В.П. Структурное описание растровых данных // Методы и средства обработки сложной графической информации: 6 Всероссийская конференция с участием стран СНГ, 25 – 27 сентября 2001 г. – Н. Новгород – С. 21 – 23.
2. Hull J. Document Image Skew Detection: Survey and Annotated Bibliography, Document Analysis Systems II, World Scientific, 1998. – P. 40 – 64.
3. Dan S. Bloomberg, Gary E. Kopec and Lakshmi Dasari. Measuring document image skew and orientation. // Xerox Palo Alto Research Center. – Режим доступа: <http://www.leptonica.com/papers/skew-measurement.pdf>, свободный.