

УДК 621.374

УСТРОЙСТВО, СЧИТЫВАЮЩЕЕ СИМВОЛЫ ПЕЧАТНЫХ ТЕКСТОВ

Пинт Э.М.,

к.т.н., профессор

Пензенский государственный университет архитектуры и строительства

Россия, г. Пенза

Еличев К.А.,

к.т.н., доцент

Пензенский государственный университет архитектуры и строительства

Россия, г. Пенза

Аннотация

Было разработано устройство для определения типов символов печатных текстов разных шрифтов. По разработанной программе определяются типы символов печатных текстов разных шрифтов, несмотря на то, что в изображениях символов возможны определенные помехи, дополнительные украшения, нехарактерные отклонения, разные толщины и размеры частей формы символов.

Ключевые слова: печатный символ, комбинация направления, обход, код направления, эталон, несовпадения, заполненные ячейки.

PRINTING TEXT SYMBOLS

Pint E.M.,

Ph.D., professor

Penza State University of Architecture and Construction

Russia, Penza

Elichev K.A.,

Ph.D., Associate Professor

Penza State University of Architecture and Construction

Russia, Penza

Annotation

A device was developed for determining the types of characters of printed texts of different fonts. According to the developed program, the types of characters of printed texts of different fonts are determined, despite the fact that in the images of characters certain interference, additional decorations, uncharacteristic deviations, different thicknesses and sizes of parts of the shape of the characters are possible.

Keywords: printed symbol, direction combination, bypass, direction code, standard, mismatches, filled cells.

Анализ изданий показывает, что существующие способы позволяют определять типы печатных символов при определенных вариациях их формы, но количество определяемых типов символов оказывается ограниченным и замыкается, в основном, в пределах одного шрифта [1, 2, 3, 4, 5, 6].

Проведенный авторами анализ формы символов печатных текстов дал возможность в качестве важных и необходимых признаков символов использовать определенную комбинацию направлений, в которую входят направления, образуемые частями контура символа в порядке его окантовки. В используемой комбинации направления отмечаются определенными номерами и идут по кругу под углом $2\pi/8$ (где $\pi = 180^\circ$) друг относительно друга (рис. 1).

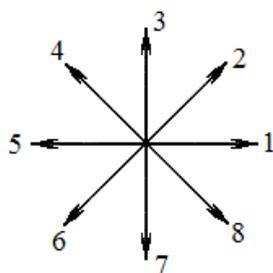


Рис. 1. Найденная комбинация направлений

ПК по разработанной программе осуществляет окантовку формы изображения символа, начиная с концевой точки. Каждый символ печатного текста представляется компьютером в виде кода, состоящего из номеров

основных направлений, входящих в контур символа и в найденную комбинацию направлений. Основное направление определялось количеством заполненных ячеек запоминающего устройства ПК, которое должно быть больше или равно количеству заполненных ячеек, составляющих максимальную толщину линий и других частей изображения символа.

Представление символа в виде кода, состоящего, в основном, из номеров основных направлений позволило однозначно определять тип каждого символа относительно других типов, независимо от масштаба элементов символа, от разных шрифтов, то есть выбранная комбинация направлений являлась оптимальной.

По разработанной программе, ПК составлял код каждого символа, убирая из изображения символа заполненные ячейки, составляющие лишнюю толщину элементов символа, убирая из полученного кода следующие направления: возникающие от возможных помех изображения, от ненужных украшений формы символа для некоторых шрифтов, от возможных изменений наклонов элементов изображения символа [7].

С целью определения типов печатных символов для них были составлены эталонные коды, учитывающие особенности формы символов.

Рассмотрим для примера определение сходных по форме букв «В» и «З» рассматриваемых шрифтов (рис. 2, а, б и рис. 3, а, б). На этих рисунках даются изображения в памяти ПК букв «В» и «З» рассматриваемых шрифтов. На этих рисунках приведены изображения в памяти персонального компьютера букв «В» и «З» рассматриваемых шрифтов, где сплошной линией показана первая окантовка формы букв, а пунктирной линией – вторая окантовка, начинающаяся с ячейки, где наблюдается развилка. Зачеркнутые ячейки иллюстрируют то, что заполнение в них исключается при ликвидации лишней толщины частей букв.

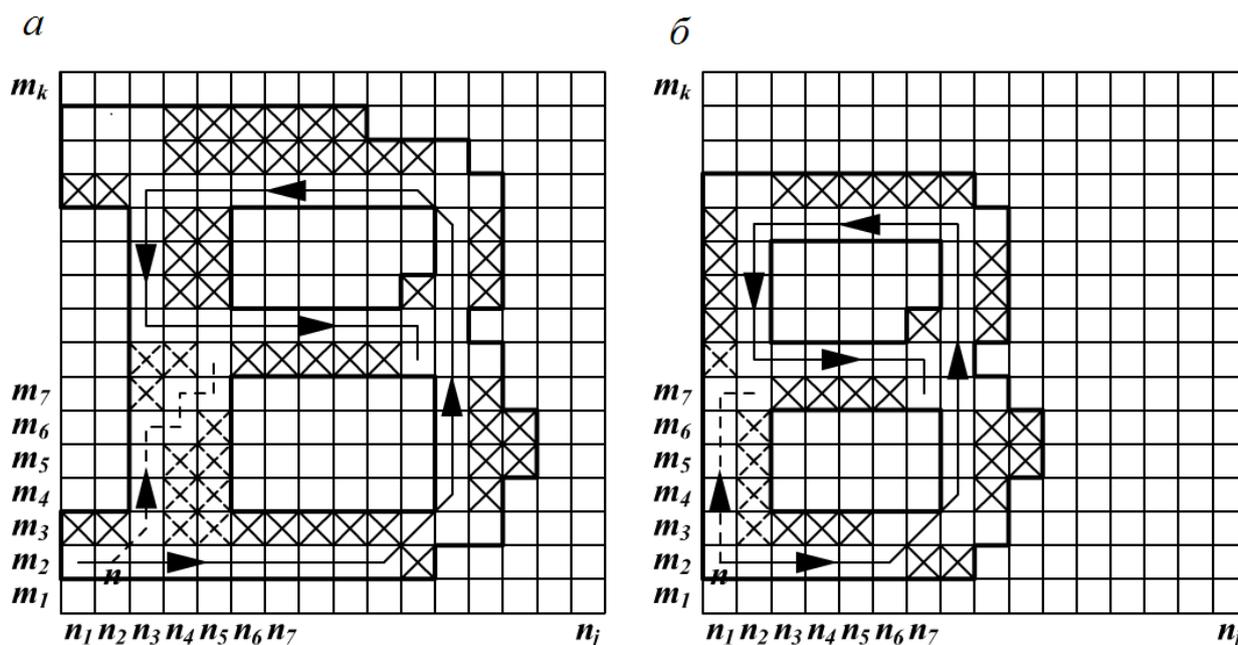


Рис. 2. Изображения букв «В»:

a – для шрифта с засечками; *б* – для шрифта без засечек

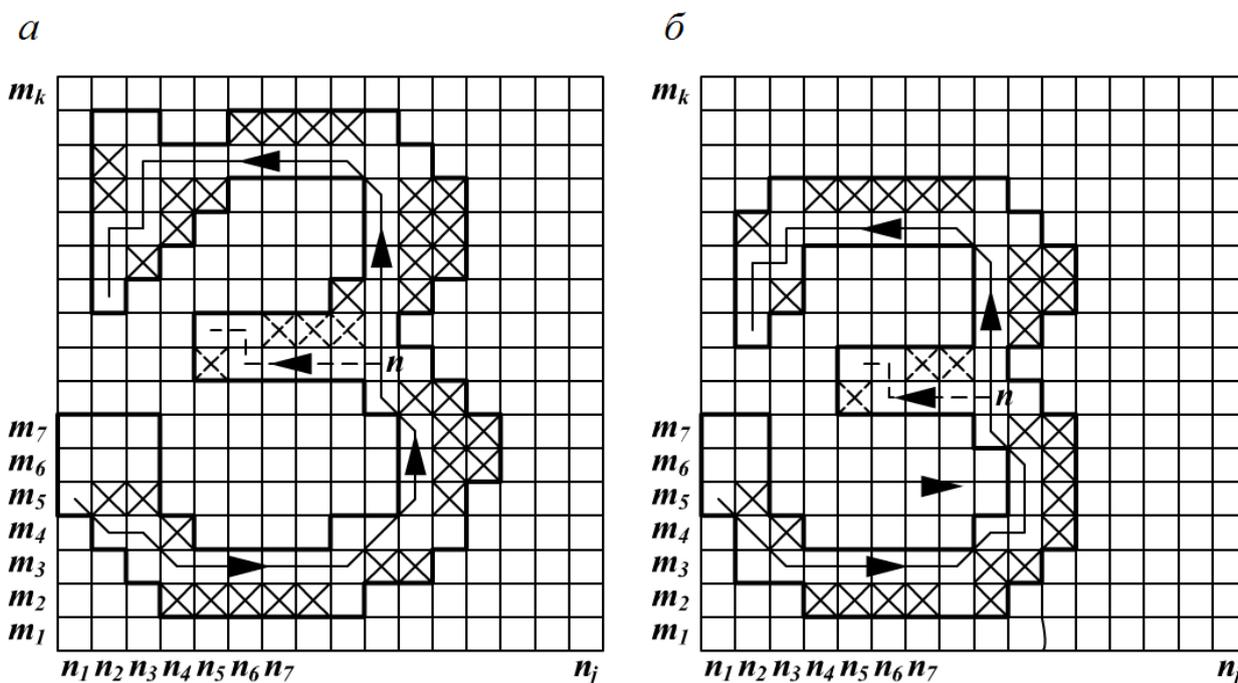


Рис. 3. Изображения букв «З»:

a – для шрифта с засечками; *б* – для шрифта без засечек

Таблица 1 показывает первоначальные коды букв «В» и «З» и коды, полученные после упрощения первоначальных кодов. Черта указывает на запись кода после развилки.

Таблица 1

В рис. 2, а	Первоначальная запись по напр.	1111111122333333334555555577771111111723331313
	Упрощенная запись по напр.	11111111223333333345555555777711111117233333
В рис. 2, б	Первоначальная запись по напр.	111122333333335555557777111117333331
	Упрощенная запись по напр.	111122333333335555557777111117333331
З рис. 3, а	Первоначальная запись по напр.	81811111223343333345555577577555535
	Упрощенная запись по напр.	88111112233333334555557755555
З рис. 3, б	Первоначальная запись по напр.	8811111213343333345555757755535
	Упрощенная запись по напр.	88111111333333345555775555

Для определения какой-либо буквы, например «В» относительно других, персональный компьютер по разработанной программе номера упрощенного кода буквы «В» сравнивает с номерами эталонных кодов всех букв. При этом, при сравнении упрощенного кода буквы «В» с каким-либо эталонным кодом подсчитывается количество несовпадений номеров в этом эталонном коде и в упрощенном коде буквы «В» [7].

При сравнении, например, упрощенного кода буквы «В» со своим эталонным кодом количество несовпадений номеров будет равно нулю, либо минимальным по сравнению с количествами несовпадений, полученными после сравнения упрощенного кода буквы «В» с другими эталонными кодами. Таким образом, тип буквы относительно других типов букв определяется по нулевому или минимальному количеству несовпадений, получаемому после сравнения упрощенного кода определенной буквы со своим эталонным кодом.

Таблица 2 дает эталонные коды букв «В» и «З» и количество несовпадений от сравнения упрощенных записей букв по направлениям с приведенными стандартными видами.

Таблица 2

Эталонные коды букв		Количество несовпадений			
		В		З	
		рис. 2, а	рис. 2, б	рис. 3, а	рис. 3, б
В	1 2 3 5 7 1 3	0	0	3	4
З	2 1 3 4 5 7 5	4	5	0	0

Как нетрудно заметить из таблицы 2, сходные по форме буквы «В» и «З» в достаточной степени распознаются друг относительно друга; в то же время одноименные буквы разных шрифтов распознаются одинаково.

Было создано устройство для определения типов символов печатных текстов разных шрифтов.

В результате проведенных исследований были выбраны важные и достаточные признаки символов рассматриваемых шрифтов, позволившие персональному компьютеру надежно по разработанной программе определять типы символов печатных текстов разных шрифтов, несмотря на то, что в изображениях символов возможны определенные помехи, дополнительные украшения, нехарактерные отклонения, разные толщины и размеры частей формы символов.

Библиографический список

1. Lear I. «A machine that reads written, words». The New Scientist, 1959, v. 6, № 154.
2. «Direct Reading for Data Processing». Process Control and Automation, 1960, v. 7, № 3.
3. Переверзев-Орлов В.С., Поляков В.Г. Универсальный автомат для чтения печатного текста: Доклад на КОИМПАЧТ, 1961, г. Москва.
4. Letter segmenting apparatus for OCR comprising multi-level segmentor operable when binary segmenting fails by Toshio Miyazaki et al, NEC, published June 3,

1980. A detailed description of how an OCR system can identify individual letters in a written sample.

5. Браверман Э.М. Попыты по обучению машины распознаванию зрительных образов // Автоматика и телемеханика. – 1962. - № 3.

6. Chris Woodford. Last updated: January 2, 2018.
<http://www.explainthatstuff.com/how-ocr-works.html>.

7. Пинт Э.М., Петровнина И.Н., Романенко И.И., Еличев К.А. Заключительный алгоритм рационального метода распознавания компьютером печатных знаков разных шрифтов и распространение метода на образы, связанные с автоматизацией работы дорожных машин и автомобилей // Материалы IV международной научно-практической конференции «Перспективные направления развития автотранспортного комплекса». - Пенза: Изд. ПГУАС, 2011. – С. 165.

Оригинальность 80%