

УДК 621.374

***ОПРЕДЕЛЕНИЯ ПЕРСОНАЛЬНЫМ КОМПЬЮТЕРОМ ТИПОВ
ПЕЧАТНЫХ СИМВОЛОВ***

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

Разработанное устройство позволяет определять типы печатных символов разных шрифтов, несмотря на изменения, в определенных пределах размеров символов, положения элементов формы символов, на наличие помех и т.д. Данное устройство целесообразно применять в таких отраслях народного хозяйства как строительство, металлургия, химическая промышленность, автомобилестроение и в оборонной. Формируются коды знаков и их образов.

Ключевые слова: коды, обход, шрифт, ячейки, матрица, изображение символа, признаки печатных символов, персональный компьютер, программа.

***DETERMINATION BY A PERSONAL COMPUTER OF TYPES OF PRINTED
SYMBOLS***

Pint E.M.

Ph.D., professor

Penza State University of Architecture and Construction

Russia, Penza

Дневник науки | www.dnevniknauki.ru | СМЭЛ № ФС 77-68405 ISSN 2541-8327

Elichev K.A.,

Ph.D., Associate Professor

Penza State University of Architecture and Construction

Russia, Penza

Annotation

The developed device allows you to determine the types of printed characters in different fonts, despite the changes, within certain limits of the size of the characters, the position of the elements of the form of characters, the presence of interference, etc. This device is advisable to apply in such sectors of the economy as construction, metallurgy, chemical industry, automotive and defense. Codes of signs and their images are generated.

Keywords: codes, bypass, font, cells, matrix, symbol image, signs of printed symbols, personal computer, program.

Из существующих способов определения типов печатных символов наиболее прогрессивными [1, 2, 3, 4, 5] являются: способ определения по элементам формы символа, способ определения с использованием самообучающихся систем. К сожалению, эти способы позволяют определять типы печатных символов, в основном, в пределах одного шрифта.

Разработанное устройство состоит из воспринимающей части и персонального компьютера (ПК). Воспринимающая часть считывает изображение печатного символа с оригинала путем создания последовательности электрических импульсов, поступающих в определенном порядке в прямоугольную матрицу, состоящую из ячеек запоминающего устройства ПК. Заполненным ячейка матрицы соответствуют элементарные участки изображения символа.

После проведения исследований авторами были найдены существенные признаки печатных символов. Ими оказались направления элементов формы символа. Была найдена комбинация направлений (рис. 1), достаточная для определения типов печатных символов анализируемых шрифтов.

По созданной программе ПК в соответствии с выбранной комбинацией (рис. 1), начиная с концевой точки изображения символа, опрашивает соседние заполненные ячейки матрицы с целью определения основного направления. Основное направление определялось количеством заполненных ячеек, которое было больше, чем количество ячеек, составляющих максимальную толщину элементов изображения символа. В итоге образовывался код печатного символа, состоящий из последовательности номеров основных направлений. Этот код получался путем окантовки по основным направлениям формы изображения символа.

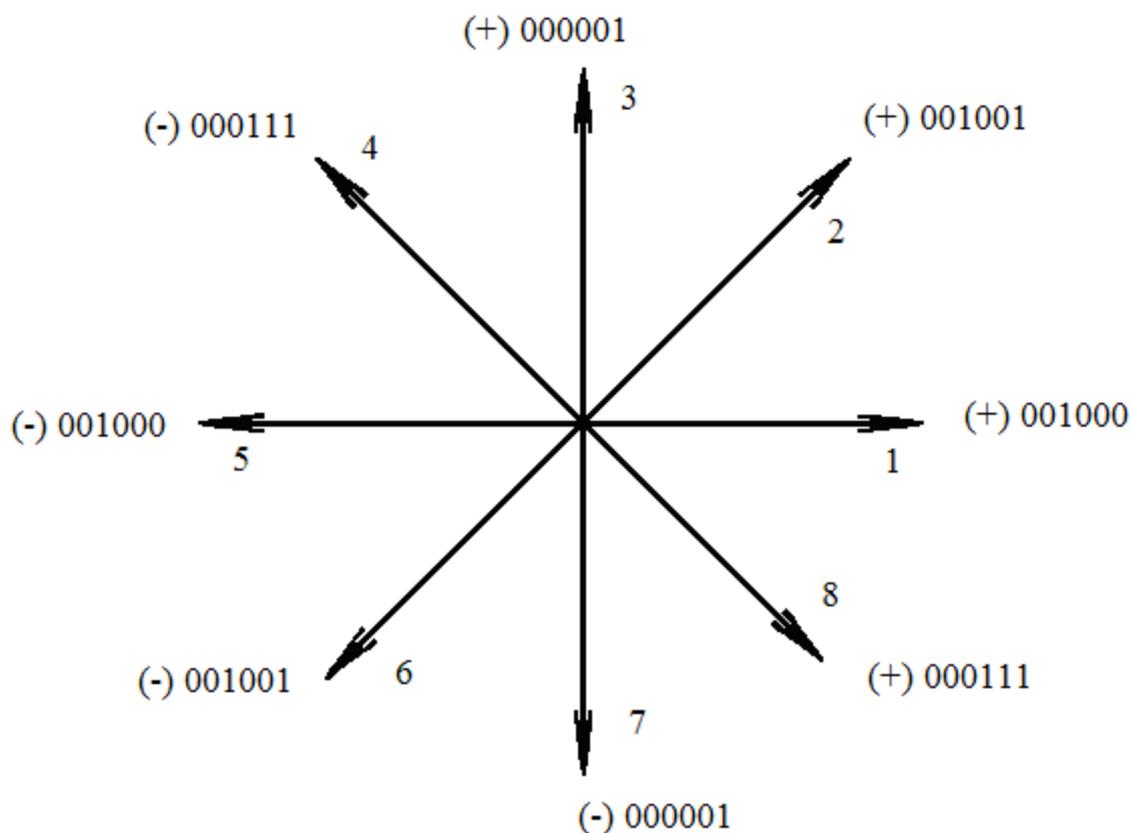


Рис. 1. Выбранная комбинация направлений, с числами для перехода на соседние ячейки по направлениям

Возникали случаи, когда при переходе ПК от одного элемента изображения к другому выявлялось неосновное направление, так называемая сумма, где количество заполненных ячеек было меньше, чем у основного направления. Тогда, чтобы окантовка не прерывалась, ПК переходил на соседнюю ячейку туда, где сумма заполненных ячеек больше, а при одинаковых суммах – на соседнюю заполненную ячейку в направлении с наименьшим номером.

Полученный код символа состоял, главным образом, из номеров основных направлений, а также из номеров определенных неосновных направлений.

Чтобы толщина элементов символа не оказывала влияния на определение типа символа, по определенным правилам освобождались от заполнения ячейки, составляющие лишнюю толщину элементов символа (так называемое стирание по толщине).

Полученный код символа по определенным правилам сокращался, то есть ПК убирал неосновные направления из кода, которые возникали из-за изменения положения элементов формы символа, из-за наличия помех, украшений формы и т.д..

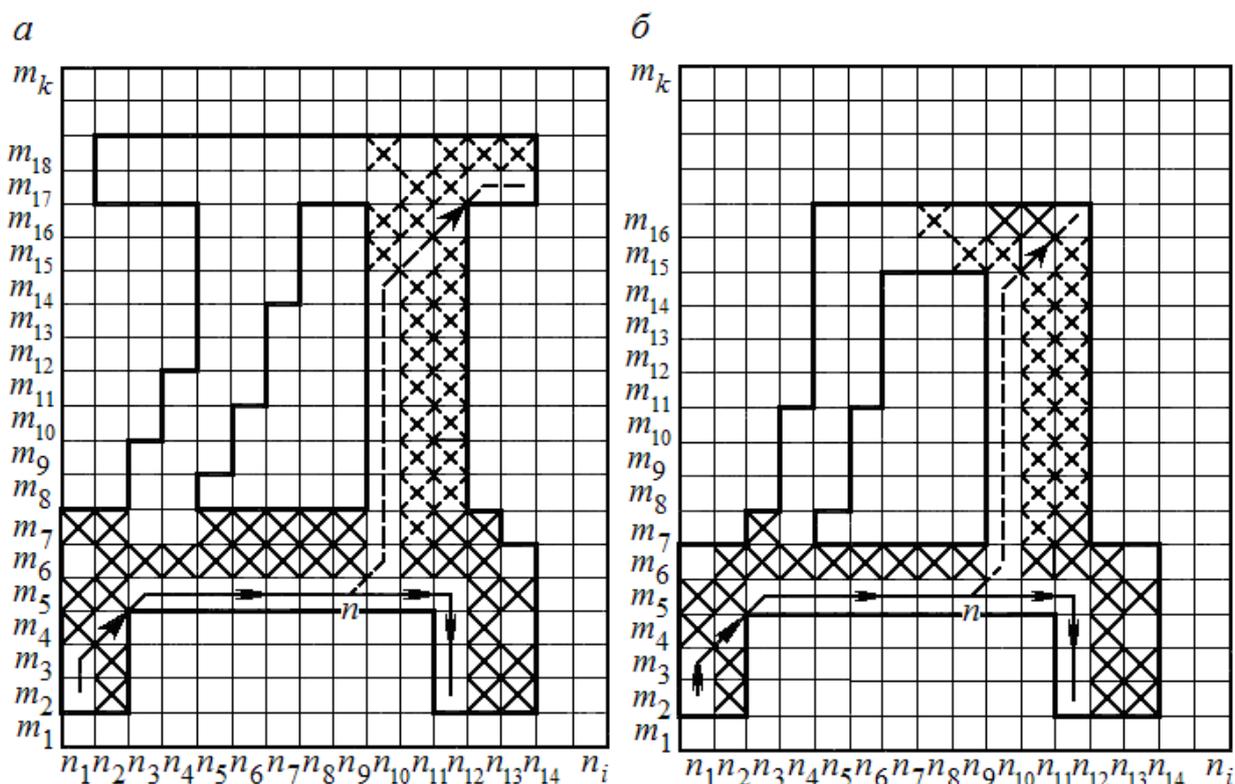
Затем сокращенный код печатного символа сопоставлялся с эталонными кодами. Для каждого типа печатного символа был составлен эталонный код, то есть последовательность направлений, получаемая путем окантовки по направлениям элементов идеально напечатанного контура символа. При таком сопоставлении подсчитывалось количество несовпавших номеров для каждого эталонного кода, и по меньшему количеству несовпадений определялся тип печатного символа [6].

Вначале предполагалось при перемещении от одной ячейки к другой по основному направлению стирать содержимое соседних заполненных ячеек, расположенных по неосновным направлениям. Однако, это правило не подошло для некоторых символов. Например, (рис. 2, а) в процессе окантовки

Дневник науки | www.dnevniknauki.ru | СМЭЛ № ФС 77-68405 ISSN 2541-8327

буквы «Д» от развилки (пунктирная линия) при переходе из ячейки $m_{16}n_{12}$ в направлении «2» (номера направлений указаны на рис. 1) на ячейку $m_{17}n_{13}$ стирается содержимое заполненных ячеек $m_{17}n_{11}$ и $m_{18}n_{10}$ в направлении «4» и содержимое ячеек $m_{17}n_{12}$ и $m_{18}n_{12}$ в направлении «3».

Это, как не трудно заметить, приводит к тому, что левая часть буквы «Д» остается не обойденной, то есть не все признаки буквы «Д» фиксируются. Аналогично для буквы «Д» другого шрифта (рис. 2, б) переход из ячейки $m_{14}n_{10}$ в направлении «2» на ячейку $m_{15}n_{11}$ вызывает стирание содержимого ячеек в направлениях «3» и «4», и левая часть буквы «Д» не обходится.



n – перекрестие; \times – ячейки, стираемые в процессе обхода от ячейки с наименьшим номером; \times – ячейки, стираемые в процессе обхода от перекрестия

Рис. 2. Окантовка буквы «Д» разных шрифтов

В связи с вышеуказанным было найдено следующее решение:

- при переходе по направлениям «2», «4», «6», «8» осуществляется стирание содержимого заполненных ячеек по направлениям «1» и «5», что исключает лишнюю толщину наклонных линий символа;

- при переходе по направлениям «3» и «7» осуществляется стирание содержимого заполненных ячеек по направлениям «1» и «5», что исключает лишнюю толщину вертикальных линий символа;

- при переходе по направлениям «1» и «5» осуществляется стирание содержимого заполненных ячеек по направлениям «3» и «7», что ликвидирует лишнюю толщину горизонтальных линий символа.

Стирается содержимое заполненных ячеек неосновных направлений.

Как видно из рис. 3 при таком стирании совершается окантовка всей буквы «Д», включая ее левую часть.

Кроме того, при выбранном правиле стирания по толщине некоторые изображения одного и того же по смыслу символа окантовывались по разному. Например, в процессе окантовки буквы «Р» (рис. 4, а) для ячейки $m_{13}n_6$ выявляются сразу три одинаковые суммы заполненных ячеек по неосновным направлениям «1», «3», «4». Согласно правилу, изложенному выше, переход на соседнюю заполненную ячейку должен происходить по направлению «1».

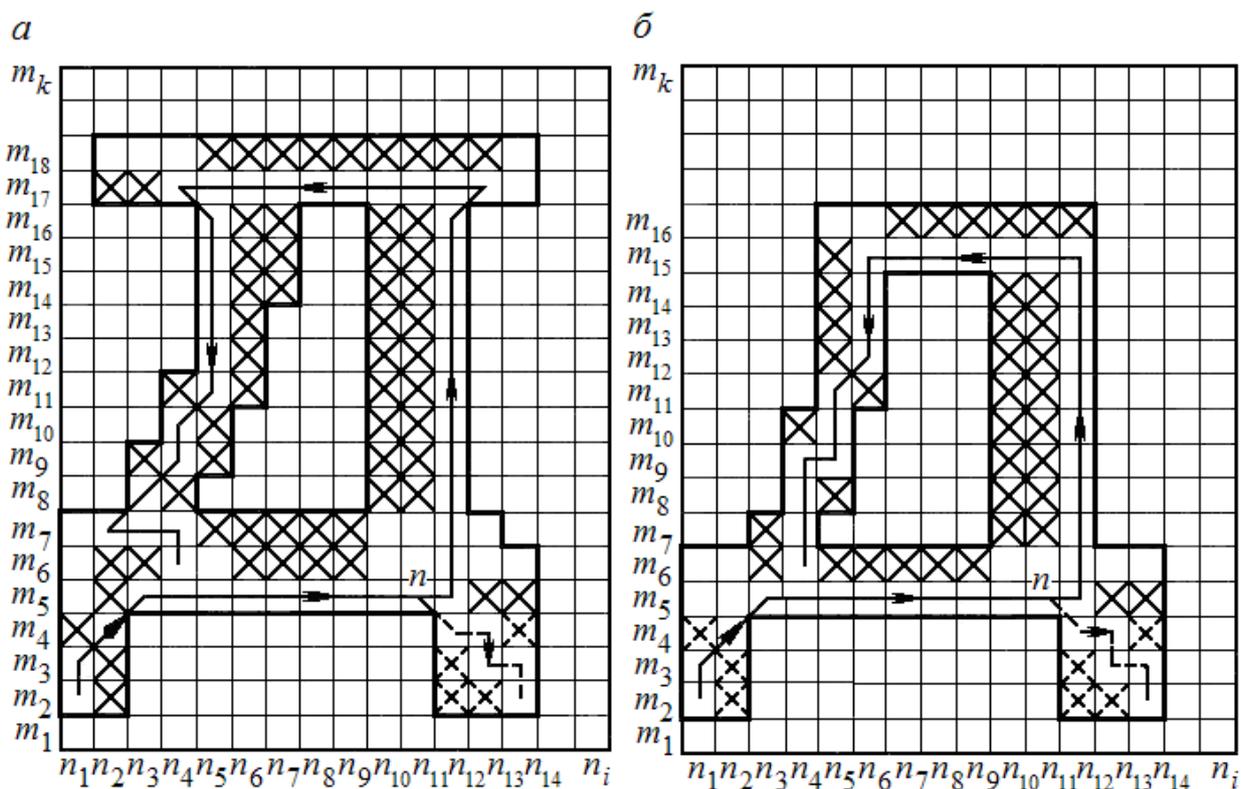


Рис. 3. Окантовка букв «Д» разных шрифтов

При этом стирается содержимое заполненной ячейки $m_{14}n_6$, и дальнейшая окантовка будет происходить от развилки m_8n_2 . Как видно из рис. 4, б, окантовка несколько отличного изображения буквы «Р» будет происходить без развилки по всем элементам изображения буквы. Подобное явление могло также наблюдаться в знаках: «3», «0», «Э», «2», «9». Во избежание подобных явлений, было найдено следующее дополнительное правило: если встречаются одинаковые суммы заполненных ячеек по четвертому и первому направлениям, переход ПК должен совершать на соседнюю заполненную ячейку по четвертому направлению, а во всех остальных случаях ранее принятое правило соблюдается. Как видно из рис. 4, в с учетом дополнительного правила окантовка несколько отличных по форме букв «Р» совпадает.

Таким образом, в результате проведенных экспериментов было получено следующее. Были найдены существенные и необходимые признаки печатных символов для их определения друг относительно друга.

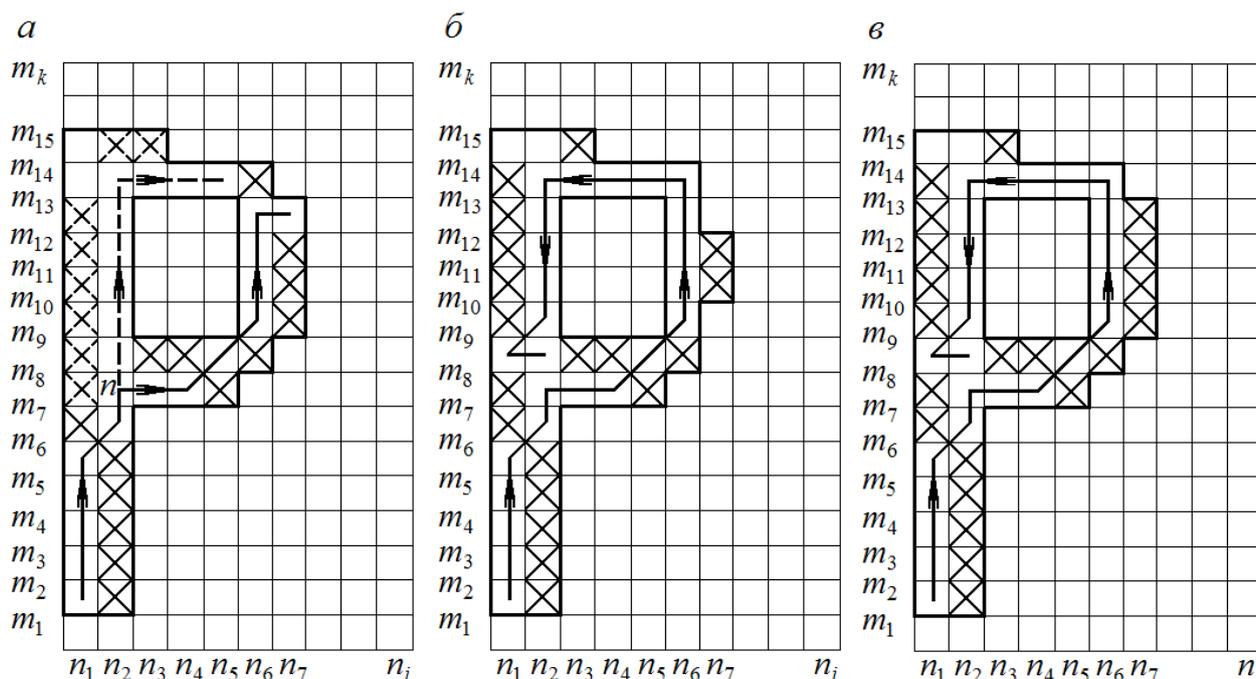


Рис. 4. Окантовка букв «Р»

Согласно разработанной программе были получены коды печатных символов после окантовки ПК их форм. Эти коды после преобразований позволили с высокой степенью надежности определять типы печатных символов, несмотря на разные шрифты, на неодинаковые размеры элементов символов, на определенные изменения положений этих элементов на возможные определенные помехи.

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. Chris Woodford. Last updated: January 2, 2018.
<http://www.explainthatstuff.com/how-ocr-works.html>
6. Пинт Э.М. Результаты исследования читающего устройства /Пинт Э.М., Романенко И.И., Еличев К.А. Р // Научно-теоретический журнал «Вестник» № 1. – 2014. - №1.– с. 182.

Оригинальность 82%