

УДК 519.68:[681.5137+612.8.001.57+007.51/52]

Об одном подходе к различению элементов из больших совокупностей традиционных систем символов

Парфенов П. Г., Назарычев С. Л.  
Ярославский государственный университет  
150 000, Ярославль, Советская, 14

получена 30 марта 2006

Аннотация

В работе приведены результаты экспериментального характера, подтверждающие устойчивое различение элементов из больших совокупностей изолированных алфавитно-цифровых символов с помощью построения характеристического набора коэффициентов черно-белого изображения. В качестве такого рода совокупностей рассмотрены объединения как исторически сложившихся алфавитов, так и искусственно созданных алфавитно-подобных систем символов. Основным результатом работы представляется демонстрация мощности различения с помощью предложенного подхода.

Характеристический набор коэффициентов черно-белого цифрового изображения в том виде, в котором он будет использоваться в настоящей заметке, был введен в работе [1]. Этот набор дает существенную информацию о геометрических и топологических свойствах изображения. Следует отметить, что, по существу, хоть и в неявном виде, такого рода набор использовался в работе Дуды, на которого ссылается Прэтт [2]. Грэй [3] также, исходя фактически из этого набора, приводит формулы для вычисления эйлеровой характеристики и для оценки числа связных компонент изображения. В работе [4] частично были использованы свойства этого характеристического набора в виде линейных комбинаций его коэффициентов для различения элементов внутри отдельных алфавитов.

Под изображением  $A$  будем понимать матрицу  $A = (a_{ij})$  размера  $m \times n$  с элементами, принимающими значения либо 0, либо 1. Фрагментом  $s$  размера  $2 \times 2$  изображения  $A$  называется любая подматрица матрицы  $A$ , состоящая из элементов  $a_{ij}, a_{i+1j}, a_{ij+1}, a_{i+1j+1}$ . Символика  $s \subseteq A$  обозначает, что  $s$  является фрагментом  $A$ . Построим следующую таблицу:

| $S_0$      | $S_1$      | $S_2$      | $S_3$      | $S_4$      | $S_5$      | $S_6$      | $S_7$      | $S_8$      | $S_9$      | $S_{10}$   | $S_{11}$   | $S_{12}$   | $S_{13}$   | $S_{14}$   | $S_{15}$   |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
|            |            |            |            |            |            |            |            |            |            |            |            |            |            |            |            |
| 0 0<br>0 0 | 1 0<br>0 0 | 0 1<br>0 0 | 0 0<br>0 1 | 0 0<br>1 0 | 1 1<br>0 0 | 0 1<br>0 1 | 0 0<br>1 1 | 1 0<br>1 0 | 1 0<br>0 1 | 0 1<br>1 0 | 0 1<br>1 1 | 1 0<br>1 1 | 1 1<br>1 0 | 1 1<br>0 1 | 1 1<br>1 1 |
| $K_0$      | $K_1$      | $K_2$      | $K_3$      | $K_4$      | $K_5$      | $K_6$      | $K_7$      | $K_8$      | $K_9$      | $K_{10}$   | $K_{11}$   | $K_{12}$   | $K_{13}$   | $K_{14}$   | $K_{15}$   |

Рис. 1.

Очевидно, что всего существует 16 различных типов фрагментов, обозначенных символами  $s_l, l = \overline{0, 15}$ , и расположенных в первой строке таблицы. Во второй строке представлен геометрический аналог соответствующего типа, в третьей строке указан вид подматриц, представляющих данный тип. Четвертая строка указывает число фрагментов типа  $s_l, l = \overline{0, 15}$ , таких, что  $s_l \subseteq A$ . Этот набор неотрицательных целых чисел  $k_l, l = \overline{0, 15}$ , будем называть характеристическим набором коэффициентов изображения  $A$ , построенных по системе фрагментов размера  $2 \times 2$ .

Для совокупности из одиннадцати экземпляров алфавитов, а именно русского, английского, греческого, угаритского [5], финикийского [6], иврита, алфавита друидов [7], футурама [8], валлийского коэлбрена [9], рунического письма [10] и символов египетской письменности [7], а также экземпляров знаков зодиака были построены соответствующие характеристические наборы. Для каждого элемента алфавитов выбрано достаточно случайное его представление в окне размером  $30 \times 40$  пикселей и полученная совокупность соответствующих векторов была подвергнута анализу на совпадение.

Результаты анализа позволяют говорить о высокой эффективности различения символов с помощью указанного подхода, при этом, естественно, символы из разных алфавитов, имеющие одинаковые представления в окне  $30 \times 40$  пикселей, имеют равные характеристические наборы коэффициентов.

В качестве иллюстрирующего примера на рис. 2, 3 и 4 приведены соответствующие экземпляры знаков греческого алфавита, алфавита друидов и финикийского алфавита и их соответствующие характеристические наборы коэффициентов, построенные по системе фрагментов размера 2×2.

|   |      |    |    |    |    |    |    |    |    |   |   |    |    |    |    |     |   |      |    |    |    |    |    |    |    |    |   |   |    |    |    |    |     |
|---|------|----|----|----|----|----|----|----|----|---|---|----|----|----|----|-----|---|------|----|----|----|----|----|----|----|----|---|---|----|----|----|----|-----|
| α | 856  | 14 | 11 | 13 | 12 | 20 | 22 | 20 | 20 | 0 | 0 | 13 | 12 | 14 | 11 | 233 | ν | 972  | 15 | 3  | 15 | 4  | 10 | 22 | 9  | 21 | 0 | 0 | 14 | 3  | 14 | 2  | 167 |
| β | 988  | 12 | 4  | 12 | 6  | 17 | 27 | 15 | 25 | 0 | 0 | 13 | 7  | 13 | 5  | 127 | ξ | 1031 | 6  | 5  | 7  | 5  | 27 | 20 | 26 | 21 | 0 | 0 | 6  | 4  | 5  | 4  | 104 |
| γ | 945  | 19 | 10 | 18 | 9  | 5  | 18 | 7  | 18 | 0 | 0 | 17 | 8  | 18 | 9  | 170 | ο | 945  | 16 | 8  | 16 | 6  | 10 | 18 | 10 | 18 | 0 | 0 | 16 | 8  | 16 | 8  | 174 |
| δ | 804  | 15 | 15 | 15 | 18 | 16 | 27 | 13 | 24 | 0 | 0 | 15 | 18 | 15 | 18 | 261 | π | 889  | 15 | 7  | 13 | 4  | 18 | 29 | 23 | 30 | 0 | 0 | 12 | 3  | 14 | 6  | 208 |
| ε | 926  | 13 | 11 | 11 | 14 | 26 | 12 | 25 | 7  | 0 | 0 | 10 | 13 | 12 | 10 | 181 | ρ | 985  | 12 | 4  | 11 | 5  | 11 | 22 | 11 | 20 | 0 | 0 | 11 | 5  | 12 | 4  | 158 |
| ζ | 874  | 9  | 10 | 9  | 12 | 12 | 18 | 10 | 16 | 0 | 0 | 8  | 11 | 8  | 9  | 165 | σ | 931  | 14 | 5  | 12 | 5  | 18 | 21 | 20 | 19 | 0 | 0 | 12 | 5  | 14 | 5  | 190 |
| η | 860  | 17 | 9  | 16 | 6  | 9  | 33 | 13 | 35 | 0 | 0 | 15 | 5  | 16 | 8  | 229 | τ | 966  | 14 | 4  | 12 | 4  | 15 | 14 | 17 | 12 | 0 | 0 | 11 | 3  | 13 | 3  | 183 |
| θ | 798  | 18 | 5  | 17 | 6  | 18 | 38 | 18 | 34 | 0 | 0 | 18 | 7  | 19 | 6  | 271 | υ | 992  | 11 | 8  | 11 | 6  | 6  | 23 | 8  | 25 | 0 | 0 | 10 | 5  | 10 | 7  | 149 |
| ι | 1071 | 7  | 7  | 8  | 7  | 8  | 13 | 7  | 14 | 0 | 0 | 7  | 6  | 6  | 6  | 104 | φ | 775  | 24 | 7  | 23 | 7  | 13 | 33 | 14 | 32 | 2 | 0 | 25 | 7  | 26 | 7  | 276 |
| κ | 974  | 17 | 10 | 13 | 12 | 4  | 18 | 6  | 12 | 1 | 0 | 13 | 11 | 17 | 9  | 154 | χ | 878  | 19 | 14 | 19 | 15 | 11 | 19 | 10 | 18 | 0 | 0 | 18 | 14 | 18 | 13 | 205 |
| λ | 1009 | 10 | 14 | 10 | 16 | 6  | 12 | 4  | 10 | 0 | 0 | 9  | 15 | 9  | 13 | 134 | ψ | 887  | 15 | 7  | 13 | 9  | 3  | 40 | 3  | 36 | 0 | 0 | 12 | 8  | 14 | 6  | 218 |
| μ | 968  | 13 | 4  | 16 | 6  | 10 | 26 | 5  | 27 | 0 | 0 | 15 | 5  | 12 | 3  | 161 | ω | 895  | 12 | 13 | 14 | 13 | 9  | 29 | 7  | 31 | 0 | 0 | 13 | 12 | 11 | 12 | 200 |

Рис. 2. Греческий алфавит

|   |      |    |    |    |    |    |    |    |    |   |   |    |    |    |    |     |   |      |    |    |    |    |    |    |    |    |   |   |    |    |    |    |     |
|---|------|----|----|----|----|----|----|----|----|---|---|----|----|----|----|-----|---|------|----|----|----|----|----|----|----|----|---|---|----|----|----|----|-----|
| Α | 914  | 11 | 11 | 13 | 8  | 13 | 25 | 14 | 30 | 0 | 0 | 12 | 7  | 10 | 10 | 193 | Ξ | 855  | 2  | 6  | 6  | 3  | 49 | 21 | 48 | 28 | 0 | 0 | 5  | 2  | 1  | 5  | 248 |
| Β | 1031 | 2  | 2  | 4  | 4  | 19 | 18 | 15 | 18 | 0 | 0 | 3  | 3  | 1  | 1  | 150 | Ι | 1026 | 2  | 2  | 5  | 3  | 18 | 22 | 14 | 24 | 0 | 0 | 4  | 2  | 1  | 1  | 147 |
| Χ | 1004 | 11 | 12 | 11 | 13 | 3  | 15 | 2  | 14 | 0 | 0 | 10 | 12 | 10 | 11 | 143 | Ζ | 989  | 8  | 3  | 8  | 5  | 24 | 19 | 22 | 17 | 0 | 0 | 7  | 4  | 7  | 2  | 156 |
| Υ | 1100 | 2  | 9  | 2  | 10 | 7  | 14 | 6  | 13 | 0 | 0 | 1  | 9  | 1  | 8  | 89  | Σ | 1005 | 13 | 7  | 14 | 7  | 11 | 14 | 10 | 5  | 0 | 0 | 13 | 6  | 12 | 6  | 138 |
| Η | 944  | 11 | 5  | 11 | 3  | 6  | 32 | 8  | 34 | 0 | 0 | 10 | 2  | 10 | 4  | 191 | Ψ | 1098 | 6  | 9  | 5  | 8  | 14 | 1  | 16 | 1  | 0 | 0 | 4  | 7  | 5  | 8  | 89  |
| Θ | 1007 | 7  | 6  | 8  | 8  | 16 | 21 | 13 | 20 | 0 | 0 | 7  | 7  | 6  | 5  | 140 | Ω | 1009 | 9  | 14 | 9  | 16 | 11 | 9  | 10 | 8  | 0 | 0 | 8  | 15 | 8  | 13 | 132 |
| Τ | 1022 | 15 | 15 | 14 | 6  | 10 | 19 | 10 | 17 | 0 | 0 | 13 | 5  | 14 | 4  | 117 | Κ | 1066 | 7  | 8  | 10 | 6  | 3  | 15 | 2  | 20 | 0 | 0 | 9  | 5  | 6  | 7  | 107 |
| Υ | 1046 | 3  | 14 | 6  | 11 | 5  | 13 | 5  | 19 | 0 | 0 | 5  | 10 | 2  | 13 | 119 | Ε | 993  | 5  | 5  | 4  | 6  | 37 | 18 | 37 | 6  | 0 | 0 | 3  | 5  | 4  | 4  | 154 |
| Σ | 992  | 8  | 13 | 11 | 16 | 25 | 12 | 19 | 12 | 0 | 0 | 10 | 15 | 7  | 12 | 200 | Ϸ | 989  | 4  | 10 | 4  | 10 | 29 | 10 | 29 | 10 | 0 | 0 | 3  | 9  | 3  | 9  | 152 |
| Κ | 1069 | 4  | 13 | 5  | 13 | 11 | 6  | 10 | 7  | 0 | 0 | 4  | 12 | 3  | 12 | 102 | Γ | 1091 | 2  | 4  | 2  | 2  | 10 | 20 | 12 | 22 | 0 | 0 | 1  | 1  | 1  | 3  | 100 |
| Ο | 905  | 11 | 14 | 11 | 13 | 21 | 18 | 22 | 19 | 0 | 0 | 10 | 12 | 10 | 13 | 192 | Υ | 859  | 14 | 14 | 15 | 12 | 7  | 33 | 8  | 36 | 0 | 0 | 14 | 11 | 13 | 13 | 222 |
| Φ | 854  | 19 | 8  | 19 | 8  | 3  | 34 | 3  | 34 | 0 | 0 | 19 | 8  | 19 | 8  | 235 | Ι | 1070 | 8  | 3  | 10 | 2  | 3  | 22 | 2  | 25 | 0 | 0 | 9  | 1  | 7  | 2  | 107 |
| Χ | 974  | 14 | 15 | 15 | 15 | 3  | 14 | 2  | 15 | 0 | 0 | 14 | 14 | 13 | 14 | 149 | Λ | 1042 | 10 | 12 | 8  | 13 | 1  | 17 | 2  | 14 | 0 | 0 | 7  | 12 | 9  | 11 | 113 |
| Υ | 1008 | 17 | 6  | 15 | 8  | 7  | 12 | 9  | 10 | 0 | 0 | 14 | 7  | 16 | 7  | 133 | Ρ | 1041 | 11 | 2  | 8  | 5  | 2  | 26 | 2  | 20 | 0 | 0 | 7  | 4  | 10 | 1  | 132 |

Рис. 3. Алфавит друидов

|   |      |    |    |    |    |    |    |    |    |   |   |    |    |    |    |     |   |      |    |    |    |    |    |    |    |    |   |   |    |    |    |    |     |
|---|------|----|----|----|----|----|----|----|----|---|---|----|----|----|----|-----|---|------|----|----|----|----|----|----|----|----|---|---|----|----|----|----|-----|
| Ⲁ | 964  | 12 | 8  | 16 | 9  | 20 | 16 | 15 | 19 | 0 | 0 | 16 | 9  | 12 | 8  | 147 | Ⲉ | 1056 | 9  | 2  | 9  | 4  | 7  | 28 | 5  | 26 | 0 | 0 | 8  | 3  | 8  | 1  | 106 |
| ⲁ | 1014 | 9  | 6  | 13 | 10 | 18 | 15 | 10 | 15 | 0 | 0 | 13 | 10 | 9  | 6  | 123 | ⲉ | 1039 | 17 | 7  | 15 | 8  | 7  | 19 | 8  | 16 | 0 | 0 | 14 | 7  | 16 | 6  | 92  |
| Ⲃ | 1070 | 8  | 12 | 10 | 13 | 7  | 7  | 4  | 8  | 0 | 0 | 9  | 12 | 7  | 11 | 93  | Ⲋ | 1061 | 16 | 3  | 15 | 3  | 8  | 16 | 9  | 15 | 0 | 0 | 14 | 2  | 15 | 2  | 92  |
| ⲃ | 955  | 9  | 10 | 11 | 15 | 19 | 16 | 12 | 13 | 0 | 0 | 11 | 15 | 9  | 10 | 166 | ⲋ | 968  | 16 | 8  | 18 | 11 | 13 | 21 | 8  | 20 | 0 | 0 | 17 | 10 | 15 | 7  | 139 |
| Ⲅ | 978  | 13 | 10 | 22 | 8  | 18 | 10 | 11 | 21 | 0 | 0 | 21 | 7  | 12 | 9  | 131 | Ⲍ | 1098 | 7  | 6  | 7  | 6  | 6  | 15 | 6  | 15 | 0 | 0 | 6  | 5  | 6  | 5  | 83  |
| ⲅ | 1068 | 2  | 10 | 5  | 10 | 8  | 20 | 5  | 23 | 0 | 0 | 4  | 9  | 1  | 9  | 97  | ⲍ | 1065 | 12 | 6  | 11 | 5  | 4  | 21 | 6  | 21 | 0 | 0 | 10 | 4  | 11 | 5  | 90  |
| Ⲇ | 1024 | 16 | 2  | 14 | 4  | 16 | 16 | 16 | 12 | 0 | 0 | 18 | 3  | 15 | 1  | 119 | Ⲏ | 939  | 19 | 6  | 16 | 8  | 17 | 25 | 18 | 20 | 1 | 0 | 16 | 7  | 19 | 5  | 155 |
| ⲇ | 909  | 9  | 18 | 9  | 16 | 17 | 17 | 19 | 19 | 0 | 0 | 10 | 17 | 10 | 19 | 182 | ⲏ | 925  | 10 | 10 | 13 | 8  | 15 | 29 | 14 | 34 | 0 | 0 | 4  | 9  | 11 | 11 | 168 |
| Ⲉ | 916  | 10 | 11 | 10 | 9  | 20 | 21 | 22 | 23 | 0 | 0 | 13 | 12 | 13 | 14 | 177 | Ⲑ | 1072 | 5  | 9  | 8  | 11 | 8  | 15 | 3  | 16 | 1 | 0 | 9  | 11 | 6  | 9  | 88  |
| ⲉ | 1022 | 18 | 5  | 18 | 5  | 10 | 11 | 10 | 11 | 0 | 0 | 17 | 4  | 17 | 4  | 119 | ⲑ | 994  | 14 | 11 | 13 | 9  | 7  | 18 | 10 | 19 | 0 | 0 | 12 | 8  | 13 | 10 | 133 |
| Ⲋ | 979  | 15 | 6  | 16 | 7  | 10 | 22 | 8  | 22 | 0 | 0 | 16 | 7  | 15 | 6  | 142 | Ⲓ | 1103 | 6  | 3  | 8  | 3  | 12 | 13 | 12 | 13 | 0 | 0 | 5  | 2  | 5  | 2  | 86  |

Рис. 4. Финикийский алфавит

## Список литературы

1. Парфенов П.Г. О некоторых свойствах характеристического набора коэффициентов черно-белого цифрового изображения // Модел. и анализ информ. систем. Ярославль. 2005. Т. 12, №1. С. 52-54.
2. Прэтт У. Цифровая обработка изображений. Кн.2. М.:Мир, 1982.
3. Gray S.B. Local properties of binary images in two dimensions // IEEE Trans.Computers. 1971. May, C-20, 5. P. 551-561.
4. Парфенов П.Г. Некоторые числовые характеристики изображений, инвариантные относительно сдвигов и поворотов // Модел. и анализ информ. систем. Ярославль. 1998. Вып.5. С. 58-61.
5. Угаритский алфавит <http://navosvet.narod.ru/d40/chmg12.HTML>
6. Википедия. Свободная энциклопедия. <http://ru.wikipedia.org/wiki/>
7. Алфавиты <http://general-1.narod.ru/files/alph.htm>
8. Инопланетный язык Futuramer.ru <http://www.futuramer.ru/info/alien/>
9. Валлийский коэлбрен <http://florell-page.narod.ru/Druids/Ogam/ogambukva.htm>
10. Руническое письмо. Терра-Лексикон: Иллюстрированный энциклопедический словарь. М.: ТЕРРА, 1998.