

## Symbols Qualitative Description for Their Recognition

Khelashvili Davit<sup>1</sup>, Tsiskaridze Dimitri<sup>2</sup>

<sup>1,2</sup> Iv. Javakhishvili Tbilisi State University, Microprocessors and Microprocessor Systems Faculty

### Abstract

For symbol identification there exist a lot of algorithms and programs. In most cases symbol is described by means of point and vector methods [1,2]. In the submitted work it is used, symbols qualitative description which does not depend on symbol's size, width or on its outline drawing. The research of the input symbols graphics face is provides according to the classification of symbols, and other properties. One symbol description by this method needs 11 bits. In this work this method is described on an example of Georgian writing alphabetic letters, but if we study the every symbols of any language, we'd see, that for each character, one can find the group properties, that describe it uniquely, should be mentioned, that above mentioned work is the result of psychoevristical research [3].

**Keywords:** Qualitative Description, Identification, Recognition, Comparative analysis, Georgian writing alphabetic letters.

### Comparative analysis

For your information, Georgian writing alphabetic letters are:

ა ბ გ დ ე ვ მ თ ი კ ლ  
 მ ნ ო პ ჟ რ ს ტ უ ფ ქ  
 ლ ყ შ ჩ ც ძ წ ჭ ხ ჯ ჰ

The first parameter of symbols qualitative description, finding out of which is the most simple and its check does not require a difficult algorithm, is a difference between the writing levels of Georgian language symbols [4]. For example fig. 1.



Fig. 1

On the given example the letter „ო“ is written only at the second level, the letter „მ“ at the first and the second levels, the letter „უ“ is written at the second and the third levels and the letter „ქ“ at three levels. Division in this way gives good results, during the establishment of this sign; making mistake probability is low, if it was not caused by objective reasons, if all letters are not written at one level. Division of Georgian language alphabet in this way is shown in Table 1.

R <sub>1</sub>	symbols group	levels
1	ა ი ო ო	II
2	ბ გ დ ე ვ კ ლ ქ უ ფ ლ ყ ც ჯ	II-III
3	ბ მ მ ნ პ რ ს შ ჩ ძ ხ ჰ	I-II
4	ტ ქ წ ჭ	I-II-III

Table 1

### Separation of symbols according to levels.

This parameter has four values and consequently symbols are divided into four groups that is very small amount for its identification.

Let us consider the second property according to which the letters will be divided into more groups. Georgian language symbols mostly consist of arches and smaller part from pieces and reserved circle, therefore the majority of symbols have one side open. For example, in a fig. 2 is shown

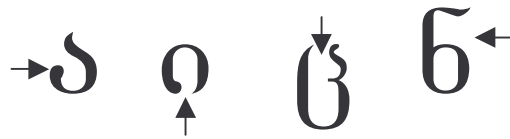


Fig. 2

The symbols which have one side open; for „ა“ the left side, „ო“ the bottom side, „ც“ the top side, „ბ“ the right side, but sometimes it is possible to have two sides open; for „ლ“ the right side and the top side, for „ქ“ the left side and the top side, for „რ“ the bottom side and the left side. For example, fig.3.

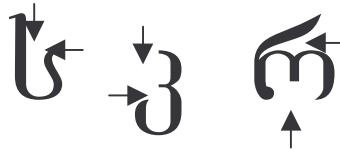


Fig. 3

If we give one bit to each side we will have  $2^4=16$  case. Let us mark these bits in this way.



Numerical value of the described property in mathematical aspect is not complex and has the following view

$$R_2 = \begin{cases} 4 & \text{if } S_l = n \\ 0 & \text{if } S_l < n \end{cases} \quad S_i = \sum_{k=n}^1 P_{ik}$$

$$n=0.4a_y \quad 0.3a_x < i < 0.7a_x \quad (i=0.3a_x, 0.4a_x, 0.5a_x, 0.6a_x, 0.7a_x)$$

$P_{ik}$  is matrix (0-1) of symbol.  $a_x, a_y$  is size of  $P_{ik}$  matrix.

In this way it is calculated the value of this parameter from above for open symbols.

$$R_2 = \begin{cases} 1 & \text{if } S_l = a_x - n + 1 \\ 0 & \text{if } S_l < a_x - n + 1 \end{cases}$$

$$n=0.6a_y \quad 0.3a_x < i < 0.7a_x \quad (i=0.3a_x, 0.4a_x, 0.5a_x, 0.6a_x, 0.7a_x)$$

$R_2=1$  is the value of the parameter from below for open symbols;  $R_2=4$  and  $R_2=1$  parameters are symmetrical and one  $S_i$  is enough.

The other values of  $R_2$  are calculated accordingly.

Georgian alphabet according to the value of this parameter is divided as follows

$R_2$	symbols	$R_2$	symbols
0	ტ	8	ა ბ ე ვ უ ფ მ შ ჰ ჯ
1	ო თ ლ	9	
2	ღ ჳ ლ ნ წ	10	ბ
3	რ ჩ	11	
4	ც	12	კ ქ ყ ძ ე
5		13	
6	ხ ს	14	
7		15	ჯ ჳ მ

Table 2



$$R_4 = \sum_{i=1}^{a_x} \begin{cases} R_4 + 1 & \text{if } (P_{ik} = 0) \& (P_{i-1,k} = 1) \\ R_4 & \end{cases} \quad K=0.3a_y$$

$$R_5 = \sum_{k=1}^{a_y} \begin{cases} R_5 + 1 & \text{if } (P_{ik} = 0) \& (P_{i,k-1} = 1) \\ R_5 & \end{cases} \quad K=0.3a_x$$

Taking into consideration the values of these parameters we shall receive the following tables

R <sub>4</sub>	symbols group
0	ა კ პ რ ს ჰ ბ ნ თ ხ ჳ ქ
1	ო ე ვ ყ ც გ ძ მ წ ბ ტ ქ
2	ლ ზ ღ თ შ უ წ ჯ
3	ჲ უ

Table 4

R <sub>5</sub>	symbols group
0	ს ი ო თ
1	ღ ლ ა რ ბ ხ ქ ე ყ გ მ წ ბ ტ ქ უ
2	ღ კ პ ნ თ ვ ც ძ შ უ წ ჯ
3	ჰ ჳ

Table 5

Taking into consideration the values of all parameters if we divide Georgian alphabet into groups, in each group there will be found one symbol [3]. In Table 6 is shown how many bits are necessary for each parameter.

parameters	bits
R1	2
R2	4
R3	1
R4	2
R5	2
sum	11

Table 6

Therefore one symbol description by this method needs 11 bits.

## References

1. J. Pacht, I. Ramm. Recognition of Hand-written Characters Using the DAP. MIT Press. London (1989).p 257.
2. Z. Bosikashvili. Optical Recognition System for Georgian Printed and Handwritten Texts. Acad. Sci. Georgia. The Georgian Technical University, Scientific-Technological Conference, Applied Program Technologies SOFTEC'99. 1999, E7.
3. V. Chavchanidze. Problems of Management of Intellectual Activity Psychoevristical Programming. Acad. Sci. Georgia. 1974.
4. D. Khelashvili, D. Tsiskaridze, D. Maskharashvili. Task for Computerized Recognition of the Typed and Printed Characters in Accordance with Language Peculiarities. Bull. Acad. Sci. Georgia 153, 1, 1996, p 70-72.

Received: 05.12.2003

Received after processing 16.04.2004