Primary Symbol Recognition Using Scanline Structural Descriptions

Oleksii Gorokhovatskyi, Olena Peredrii

Department of Informatics and Computer Technics, Kharkiv National University of Economics, UKRAINE, Kharkiv,
E-mail: gorohovatsky@rambler.ru, elena_peredriy@ukr.net

Abstract - The method for the construction of scanline linear descriptions of an object on image was proposed. Description is based on interpretation of structural links of the object elements and is stable despite object scale and movement changes. Comparison procedures of proposed descriptions were investigated, efficiency of suggested approach was confirmed by computer modelling experiments.

Key words – symbol, block definition, linear description, state, description comparison, primary recognition, states of an image.

I. Introduction

The problem of text recognition remains one of the most actual in the area of computer vision systems because of its extraordinary complexity. A lot of existing systems have limitations concerning the language of recognition and font size. The most effective methods of optical character recognition are constrained usually with the use of neural networks at considerable temporal expenses on their training. Using of structural and numerical features [1, 2] that are characterized by simplicity of construction in the real tasks are not investigated enough as well as some humanlike recognition approaches e.g. by form [3-5].

The objective of this thesis is the development of primary recognition method based on scanline symbol representation (invariant-steady description of its structural properties), which allows to decrease the set of etalon images with further secondary recognition stage. During this we must also achieve high quality for primary processing in acceptable real time.

II. Scanline linear representation

Main idea of the proposed approach consists in the construction of some symbol description as a chain of its structural states [1]. We name these descriptions "linear" just because each of them is a set of symbols like line of char elements and is based on line analysis of the image. Construction of such descriptions and their further comparison procedure must provide their stability to scale and font changes inside a single font class.

Let's consider image $I(x,y)$ as a set of horizontal lines $L=\{l_1,l_2,...,l_n\}$ with some pixel step Δy. We will name the sequence of pixels with identical brightness as "block" b, alternation of blocks forms a line. Within each line m blocks can be separated as $l=\bigcup_{i=1}^{m} b_i$. We will imply as block in future only those of them, which carry a semantic sense in specific task, in our case – a set of black pixels. For example, Fig. 1 illustrates the line of image that consists of four blocks.

A value m is formed coming from practical considerations after the analysis of etalon pattern images. For example, for English alphabet the amount of such blocks in lines varies from 1 to 4. It is necessary to take into account that m can be different even for the same character of different font types.

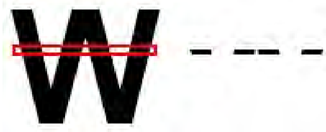

Fig. 1. Block definition

Let us denote by $S^i=\{s_1,s_2,...,s_{k_i}\}$, $i=\overline{1,m}$ a set of the states that can form combinations of blocks in every line, k_i is an amount of all possible states in S^i. Formation of the certain state by blocks can be obtained both on the basis of only current block analysis and taking previous blocks into account. For example, a single block in a line can be mapped to one of the marked states $S^1=\{s_1,s_2,s_3,s_4\}$, where s_1 corresponds to horizontal line, s_2 – vertical line, s_3 – diagonal line, s_4 – split of a single line. Possible set of the states can be generated and composed in groups after an automatic analysis of all etalon images.

Main complexity of states estimation for each of S^i set is connected with the necessity of providing the complete scope of all possible states in all patterns, and also with one-to-one correspondence of their sequences to etalon images.

Thus, each of image lines l is mapped to one of all possible states: $l \to s$, accordingly, the set of all lines forms the set of the states $L \to C=\bigcup s$, which finally forms its linear scanline description.

We will define a set of rules, which will form the states. Blocks and lines are characterized by parameters that describe their structure generally. Weight centers and block widths were chosen as such parameters $b=\{c,w\}$.

Actually, rules for reflection of set of blocks for a specific line to some state are represented by case-situations, coming from the analysis of etalon images. For example, here are decision mapping rules for S^1 set and line l_i:

$$\left[w_{1_i} <= \delta W\right] \Rightarrow l_i \to s_1,\ \left[w_{1i} > \delta W\right] \Rightarrow l_i \to s_2,$$

$$\left[\left|w_{1_i} - w_{1_{i-1}}\right| <= \delta_1,\ \left|c_{1_i} - c_{1_{i-1}}\right| <= \delta_2\right] \Rightarrow l_i \to s_3,$$

where $\delta \in [0;1]$ is predefined threshold value, W is the total width of image, $b_{1i}=\{c_{1i},w_{1_i}\}$, $l_i=\{b_{1_i}\}$, $b_{1_{i-1}}=\{c_{1i-1},w_{1i-1}\}$, $l_{i-1}=\{b_{1_{i-1}}\}$.

Application of rules to each image line $l \in L$ gives an opportunity to form a linear representation of image as a sequence of states $L \to C=\bigcup s$.

Simultaneously with the linear description for each of the objects, its profile is formed. Profile is a more general analogue of symbol description and is mapped as follows: $L \to P = \bigcup S$. Such profiles do not possess property of one-to-one correspondence to a single etalon class, and maps usually into a few classes. Such mapping can be used for auxiliary aims, for example to make recognition procedure more precise.

To provide profile and linear description stability in conditions of scale changing repetitive values merge into a single value.

Fig. 2 shows an example of etalon and test images.

Fig. 2. Examples of etalon images

Linear descriptions for images on Fig. 2 (letter "A" on the left side) are $C_1 = \{s_2 s_4 s_1 s_4 0 s_2 s_2 s_2 s_1\}$, $C_2 = \{s_2 s_4 s_1 s_4 0 s_2 s_3 s_4 s_6 s_1 s_3 s_2\}$ accordingly to the marks mentioned above. Values before "0" symbol characterize the states during horizontal scanning (from top to bottom), after "0" symbol – during scan with vertical lines from left to right. Differences in vertical profiles are related to the considerable changes of symbol structure on edges. The merged combined profiles of images coincide and equal to $P_1 = P_2 = \{1212121\}$. Images on the right of Fig. 2 ("B" letter) have distinctions on linear presentation $C_3 = \{s_1 s_7 s_1 s_7 s_1 0 s_1 s_{10} s_4\}, C_4 = \{s_1 s_4 s_7 s_6 s_1 s_4 s_7 s_1 0 s_1 s_{10} s_4\}$ as well as identical profiles $P_1 = P_2 = \{121213212\}$.

During comparison of linear descriptions, some factors should be taken into account. The main problem here is variable length of descriptions because of quality difference for two images in the same class. One of the possible approaches here is the weighted modification of Levenshtein's distance using [6].

III. Experiments

Etalon set during experimental investigations contained 442 images of Arial font bold upper and lower case symbols of different size, also with digits, forming in this way 62 etalon classes.

Test set contained 1590 images of Arial font symbols of different size, as well as another font without serifs, similar to Arial. Symbols were taken from synthetic and real-scanned images including documents with different scan quality.

Recognition procedure was based on the comparison of merged linear descriptions according to Levenshtein's distance minimization. Besides this, manual correction was performed in the cases of inability to recognize symbol without its context understanding, e.g. for letters "l" and "I", which are identical for Arial font.

Let us call the "primary recognition" the one, which gives the correct result of recognition in the first α outputs, which correspond to the recognized image with the best probability.

Table 1 shows the probabilities of primary recognition with different α value. As we can see, suggested approach based on linear descriptions gives correct number of etalon class in the first 10 outputs with the 0,95 probability.

TABLE 1

PRIMARY RECOGNITION PROBABILITIES

α value	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
Successful primary recognition rate	0.78	0.86	0.88
α value	$\alpha = 5$	$\alpha = 7$	$\alpha = 10$
Successful primary recognition rate	0.92	0.93	0.95

Conclusion

Investigations are devoted to the development of the method, which allows to perform primary recognition of a symbol on the image using its linear description. Mapping of an image into linear description is obtained with structural analysis of symbol during scanning with lines. Suggested representation combines quantitative parameters estimating simplicity with structural features analysis preserving invariance and stability to scale and movement disturbances.

Method of building linear descriptions and profiles is the main scientific novelty of this thesis. This approach allows to efficiently reduce etalon set after primary recognition. As a goal for further investigations we can consider ways for improving linear descriptions formation, e.g. not with the analysis of current line only, but with several previous ones too. The more common goal is connected with transforming suggested approach in a way to allow for high-quality direct recognition.

References

[1] C. L.Tan et al., "Imaged document text retrieval without OCR," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 6, pp. 838-844, June 2002.

[2] S. Lu, L. Li, C. L. Tan, "Document Image Retrieval through Word Shape Coding," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 11, pp. 1913-1918, Nov. 2008.

[3] S. Lu, C. L. Tan, "Retrieval of machine-printed Latin documents through Word Shape Coding," J. Pattern recognition, vol. 41, no.5, pp. 1799-1809, 2008.

[4] A. F. Smeaton, A.L. Spitz, "Using Character Shape Coding for Information Retrieval," in Proc. 4th International Conference on Document Analysis and Recognition, vol. 2, pp. 974-978, 1997.

[5] C.L. Tan et al., "Text Retrieval from Document Images Based on Word Shape Analysis," J. Applied Intelligence, vol. 18, no.3, pp. 257-270, 2003.

[6] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," Soviet Phys. Dokl., vol. 10, no. 8, pp. 707-710, February 1966.