RECOGNITION OF HANDWRITTEN NUMERALS USING DECISION GRAPH

S. Katsuragi, H. Genchi, K. Mori and S. Watanabe

Central Research Laboratory
Tokyo Shibaura Electric Co., Ltd.
Kawasaki, Japan

Summary

An optical character reader capable of reading handwritten numerical characters has been developed and applied to mail automation.

The recognition of characters is performed by extracting geometrical features out of horizontally divided zones together with their mutual connective relation. The topological structure of character strokes is analyzed by a sequential machine with an input string of stroke segments combined with information on the connective relation.

The extracted features are sequentially matched with standard references. A new decision method called decision graph matching is fully utilized. The decision graph consists of a set of decision transition diagrams to be compared with the string of extracted features and a penalty count system to detect the optimum match.

High flexibility against unlimited variability of handwritten character shapes is achieved by means of a stored logic system using rewritable magnetic core memory.

1. Introduction

Postal code number reader-sorters have been developed by Tokyo Shibaura Electric Company, Ltd., under contract with the Ministry of Postal Services of Japan (Fig. 1). The machine sorts letters by recognizing three digit handwritten postal code numbers in the preprinted entry boxes on the envelope. The total system of the prototype model was described in a previous paper[1]. The present paper reports the recognition logic used in the latest model.

The major difficulty of recognizing handwritten characters is almost unlimited variability of character shapes, sizes, skew etc., which vary mainly with individual writers. The postal service is nation-wide and open to the public so that control over character shapes or writing instruments is practically impossible. Consequently, the machine has to read a tremendous amount of character variations. To cope with thi6 difficulty, hardware implementation has as much importance as logical design of the recognition logic.

2. Feature extraction by division and recombination of stroke segments

2-1 Approach to the feature detection

Feature detection is a technique whereby a two-dimensional pattern can be reduced into a set of combinational or logical parameters comprizing its geometrical shape. Many contributions have been made, among which the FIDAC (Ledley et al.)[2] is one of the most elaborate systems detecting geometrical features using a computer program. Another Ledley-[3] article is also concerned with the linguistic description of two-dimensional pictures. Greanias^ presented a stroke tracing technique by means of a CRT scanner showing a practical and economical application to the optical character reader.

The method described hereafter is also a trial to transform a two-dimensional stroke pattern to a string of definite symbols in strictly a mechanical way, none the less having a great flexibility to match character variations. A similar method is found in the classical articles of Grimsdale"' [7] using an electronic computer, but the following method is more simplified and machine oriented.

2-2 Scanning and preprocessing

Three postal code number entry boxes (6. 5 mm x 9 mm) with two additional small ones are printed in red near the upper-right corner of postal cards or envelopes. The code numbers are expected to be written within these entry boxes in black or blue. Three digits are recognized by the machine. Figure 2 shows the simplified block diagram of the recognition logic. Two vidicon cameras, one of them equipped with a red optical filter, are used to scan numerals. The video signal of both cameras is sampled at intervals of 0. 2 mm, quantized into binary level and fed to the segmentation circuit. Then the video image is segmented into individual numerals and tored in a magnetic core memory.

The stored pattern is thinned into a skeleton by removing boundary bits of wide strokes. The thinning method is similar to Unger's[5], except that a rectangular 3 x 4 bit binary combination is used to determine whether or not to remove a boundary bit. Then the scale of the pattern is roughly normalized by halving it if its horizontal height exceeds 40 bits. The pattern is thus represented on the 36 by 40 bit grid.

2-3 Stroke segment detection

As a primary stage in detecting the geometrical features of character shapes, stroke segments are extracted out of a pattern by scanning the whole pattern field with a 3 x 3 bit submatrix.

Direction of the stroke segment is determined according to the binary configuration of a 3 x 3 bit partial pattern in the matrix. The possible 512 (2^) partial patterns are classified into one of seven stroke segment directions, including "blank (BLANK)", "vertical (VERT)", etc., as is listed symbolically in Fig. 3. Figure 4 shows examples of bit combination which are classified as "incrementally slanted (INC)". Most stroke edge irregularities are deleted by this classification.

The 3 x 3 submatrix overlappingly scans the pattern every bit horizontally from left to right and every second bit vertically from top to bottom (see Fig. 5). The overlapped scanning extracts rather redundant information from a pattern, but it moderates the effect of the loss of information which otherwise occurs at the boundaries of the field partition.

The classification of the stroke segment directions is partly influenced by this scaning mode. Since the pattern is scanned every bit horizontally, no distinction is necessary between left and right vertical lines while high and low horizontal lines are both distinguished. The stroke segment indicated by "X" in Fig. 3 is such that it does not have sufficient information to decide the direction, or has more than one direction within a pattern. Such undecided directions can be interpolated by adjacent stroke segments.

2-4 Horizontal feature detection by analysis of vertical connective relation.

The next stage of the feature detection procedure is to combine a string of stroke segments and determine a horizontal feature, which represents a partial figure of a character in a horizontal zone (Fig. 6). Figure 8 is an example of computer simulation of the feature detection process. The pattern on the left represents a thinned and normalized pattern; the middle represents stroke segments; and the right side shows a string of extracted features. The 40-bit height pattern field is scanned by the submatrix every two bits vertically, so that there are 20 horizontal zones from which the features are extracted. The feature is determined by a sequential logic with an input string of stroke segment directions.

The main problem of reducing a two-dimensional pattern into a combinational or linguistic string of symbols is not to destroy two-dimensional relationship involved in the pattern. Since the extraction of features is achieved by horizontally scanning the stroke segments, horizontal relation can be extracted without a loss. But the mutual relationship between two vertically adjacent stroke segments is lost by the horizontal zone division, even if they are out of the same continuous vertical stroke.

In order to represent a connective relation of stroke segments, a "tag" bit is affixed to every segment. The tag bit indicates whether this stroke segment connects to the ones in the upper horizontal zone. For the "blank" segment, this bit simply signifies the existence of a stroke in the upper adjacent 3 x 3 submatrix. The tag bit is indicated in Figs. 7 and 8 with an asterisk. Thus, an input binary pattern is converted to a rectangular array of stroke segment directions accompanied with tags as is illustrated in Fig. 8.

The horizontal sequence of these compound stroke segments has information on both the geometrical shapes of strokes within a horizontal zone and their connective relation to the upper strokes. The string is supplied to a sequential logic and classified to one of the features listed in Fig. 6.

Since topological properties are found among the relative and sequential relation of strokes, sequential analysis of the connective information (tag) plays a fundamental role. For example, a joint of two separate strokes is detected by a string sequentially ordered as "tagged segments, non-tagged segments, and tagged segments". Such a sequence signifies a horizontal stroke with both the right and left ends connected to the two separate strokes in the upper zone. Figure 7 shows examples of such sequences. Those are classified as "CUP" in Fig. 6. Similarly, a branching of a stroke is detected by a sequence of "tagged segments, blanks with tags, and tagged segments." These are classified as "PAI" in Fig. 6.

The connective tags are also utilized to detect a relative location of a stroke at the upper zone's strokes. Such information is especially useful to distinguish some types of "3" and "5", whose difference depends only upon whether the vertical stroke connects to the right or left of the upper horizontal bar. Roughly speaking, a sequential structure of tagged and non-tagged stroke segments determines an inter-zone characteristic of horizontal features and that of stroke segment directions (VERT, INC, etc.) traces an intra-zone characteristic, i. e. , upward or downward slope of stroke segments within a horizontal zone.

Horizontal features are classified into 19 categories (Fig. 6) each of which is modified by three kinds of connection representing information, "left", middle" or "right". These modifiers are indicated by suffices "-JL", "-M", or "-R" if necessary. The primary purpose of this information is to represent a relative location of the feature to the upper one. Thus, VRT-L following PAL-M represents a vertical stroke connecting to the left part of a pair of strokes in the upper horizontal zone, whereas a VRT-R following PAL.-M represents a stroke connecting the right part. Therefore, 57 distinct horizontal features are available for describing a partial pattern. The whole pattern is transformed to a string of horizontal features of length 20 (Fig. 8).

Topological structures are usually described in term of node, branch and connection matrix in the combinational topology or graph theory. In the above method, strokes and their connective relation are not separately treated, but expressed as a whole. Although a single horizontal feature may not posess complete information on the topological structure of strokes, any two successive features complete it.

Compared with the conventional graph theoretical description of features, the present method is reasonable for the recognition logic of handwritten characters in that: first, highly flexible logic to detect the feature is realized by a sequential machine with stored transition table, secondly, recognition of characters is easily performed by the sequential matching of the detected features as will be described in the following section.

3. Recognition by sequential decision graph

The string of horizontal features has a great deal of variability reflecting variations of shapes, sizes, skew, etc. , of handwritten numerals. The object of the decision logic is to identify these variations as one of ten numerals.

In the articles by Grimsdale[6]» [7], Greanias[4] or Sakai[8], extracted features are combinationally matched with a standard list of features. In the present method, the features are matched sequentially. The horizontal zones are ordered from the top to the bottom, hence the sequence of the features extracted from them are arranged in the same order.

It must be noted that the sequential order of features is not much affected by character variations. For example, the character "8" has two loops vertically connected to each other. Divided into horizontal zones, it is transformed into one of the sequences like CAP-M, PAI-M, (PAL-M), CUP-M, CAP-M, PAI-M, (PAL-M), CUP-M. The number of PAL-M's (Parallel lines) may be different for each individual character, but the sequential order is not altered. In order to identify all these strings as the same, sequential logic is the most suitable.

The vertical size and location of characters has no essential influence upon the recognition if a sequential logic is used. And the feature extraction logic compensates for the horizontal deviation of characters. Therefore,the logic is independent of character displacement in both directions, despite the fixed mode of scanning and partition of the character field.

If one feature string corresponds to a single numeral, a sequential machine whose output is one of the ten numerals will make a sufficient realization of a decision logic. In practice, however, it is almost impossible to design such a logic because of an enormous amount of input string variations to be dealt with.

This difficulty is overcome by means of a combination of one-to-many recognition logic with an indirect output system.

The recognition i6 made by utilizing a set of sequential diagrams. Figure 9 is an example of a transition diagram of the sequential recognition logic which identifies one of the variations of the numeral "8". Each diagram identifies only one character variation. For every character, several transition diagrams are provided, each of which identifies a subclass of the character variations. This is also the case in the piecewise linear decision of pattern matching method.

The input sequence is compared with every diagram. In Fig. 9, circled numbers indicate internal states; arrows indicate the directions of transition; and the accompanying symbols specify input features meeting the transition condition.

The comparison with the transition diagram starts from the entry state ("0" in Fig. 9) and transits step by step to the next internal state corresponding to the input symbol. The destination of the transition is determined by the input feature. The final decision of recognition is not directly performed by the output of the logic, nor by the arrival to a certain final state. Instead, a penalty count system is employed. A penalty is assessed to that diagram every time when a feature improper to that diagram appears in following up that transition diagram. In Fig. 9, the symbol in parenthesis indicates such improper inputs ("others" mean other features than are specified). The total number of penalty represents a degree of mismatching of the input sequence with that diagram.

After comparison with all diagrams, the minimum value of the penalty is detected for every entry. If the minimum determines a unique numeral, it is selected as an answer, otherwise the recognition is rejected. But, if the minimum value equals or exceeds two, the answer is also rejected. Therefore, if the penalty count reaches two, further transition may be terminated in order to save recognition time.

The method described above is considered as an extension of a decision tree[9]. In the case of the decision tree, the transition is also sequentially executed. One of the problems of the tree decision logic is that the number of nodes (decision point8) increases exponentially when the steps of sequential decisions increase. Another problem, closely related to the one above, is that the initial or early stage decisions have too strong influence upon determining the tree structure of the logic. Once the decision takes a wrong branch, it could stray into completely irrelevant branches. To restore the erroneous decision, a great many nodes following that decision would be necessary.

The decision graph is a set of sequential transition diagrams considered as decision logic. This is an extension of the tree in the sense that it contains loops within it. In other words, there are more than one path which arrives at the same node. This means that, even if the decision takes an erroneous branch, it is possible to be recovered afterwards. Because of this structure, the defects of the decision tree are moderated, and a great deal of variations of feature sequences are identified by a rather simple diagram in a compact and economical way.

The design of recognition diagrams is very important in order to attain a high tolerance against character variations. The present systematic design technique cannot be applied to this problem, since the number of input sequences is too large to be dealt with. Therefore, the design is carried out by an intuitive cut-and-try method using computer simulation. In designing a sequential logic, assignment of internal states is the main problem. In case of intuitive design, it is desirable for internal states to have some intuitive meanings which are easily understood by a designer. This is especially important in not causing any trouble or confusion when a modification is to be made after the design is completed.

The sequence of horizontal features can be naturally divided into several subsequences, each of which represents a partial shape of the character. The numeral "8" gives a good example, where the sequence is divided into two subsequences representing upper and lower loops. Each subsequence can be separately identified and then mutually recombined. Hence the internal states can also be separately assigned according to these subsequences. Such an intuitive decomposition and recombination method facilitated the design of the sequential transition diagrams.

4. Inplementation

Emphasis is put on the hardware feasibility and flexibility of the recognition logic during the system design. The recognition logic is apt to become complicated and large in size, since the variation of handwritten characters is very large. Besides, the logic must be capable of easy revision to match the true distribution of character variations. Even if the logic is initially designed using a great number of character samples, the whole set of characters to be read is much greater, and further improvement is inevitable after the machine is made up.

A solution to those problems is a stored logic system. Since the recognition logic is composed of sequential machines, this is realized by stored transition tables and table interpreting circuits. The tables are stored in a 36 bit, 8K word magnetic core memory with a cycle time of 1. 5 u.s. If the logic has to be revised, it is easily accomplished by changing the contents of the memory. The tables are supplied by a computer prepared perforated paper tape through a paper tape reader. By this, the computer simulation is directly linked to the actual reading machine.

The required recognition speed for the mail sorter is a little more than 18 characters per second (6 letters a second) so that a high speed core memory is sufficient for this purpose.

5. Conclusion

The recognition logic for handwritten numerals has been designed in order to overcome the almost unlimited possible variability of handwritten character variations. The logic consists of geometrical feature detection and feature matching with references. Both are performed by a sequential logic with stored transition table. The hardware implementation technique of the stored logic system using rewritable core memory enhanced system flexibility and shortened the turn-around time for improvements.

The recognition logic is applied to the handwritten postal code number reader-sorter. Three models have been installed in Tokyo and Osaka Central Post Offices and constantly field tested to achieve high performance levels.

Acknowledgement

The authors are greatly indebted to T. Suzuki, General Manager, and Dr. N. Sawasaki, Director of the Central Research Laboratory for their constant encouragement. They are also grateful to the members of the Mail Automation Group who developed the mechanism of the reader-sorter. Thanks are extended to the staff of the Bureau of Posts, the Ministry of Postal Service of Japan, for valuable advice during developing and field testing of those machines.

References

(1) H. Genchi et al., "Recognition of Handwritten Numerical Characters for Automatic Letter Sorting", Proc. IEEE, vol.56, No. 8, pp. 1292-1301 (Aug. 1968).

(2) R. S. Ledley et al., "Pattern Recognition Studies in the Biomedical Sciences," Proc. SJCC, 1966, pp. 411-427.

(3) R. S. Ledley, "Programming and Utilizing Digital Computers," McGraw-Hill. 1962, pp. 362-375.

(4) E. C. Greanias et al., "The Recognition of Handwritten Numerals by Contour Analysis," IBM Jour. pp. 14-21 (Jan. 1963).

(5) S. H. Unger, "Pattern Detection and Recognition," Proc. IRE, vol.47, pp. 1737-1752 (Oct. 1959).

(6) R. L. Grimsdale et al., "A System for the Automatic Recognition of Patterns-," Proc. IEE, vol. 106B, pp. 210-221 (Dec. 1958).

(7) R. Li. Grimsdale et al. "Character Recognition by Digital Computer using a Special Flying-Spot Scanner," Computer Jour., July 1961, pp. 129-136.

(8) T. Sakai, "Pattern Recognition," J. of the Institute of Electronics and Communication Engineers of Japan, vol 8, No. 914, pp. 10-19 (Nov. 1967). in Japanese

(9) L. S. Frishkoph and L. D. Harman, "Machine Reading of Cursive Script", Information Theory, Symposium on Information Theory, London, I960; C. Chery ed., Butterworths, 1961, pp. 300-316.

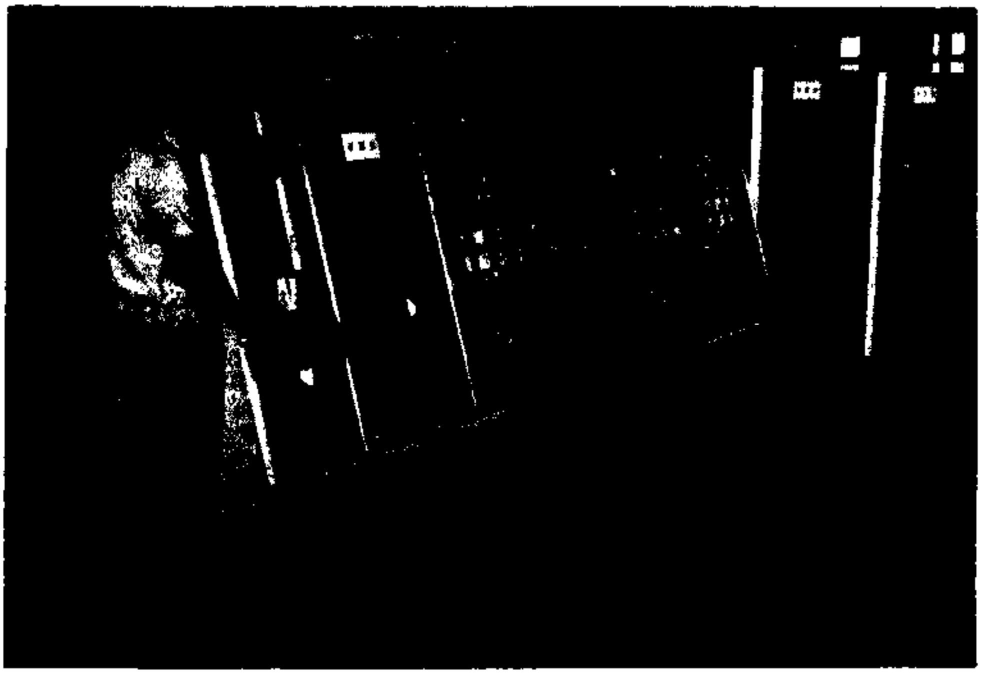

Fig. 1. Latest model handwritten postal code number reader-sorter under field test at the Central Post Office, Tokyo.

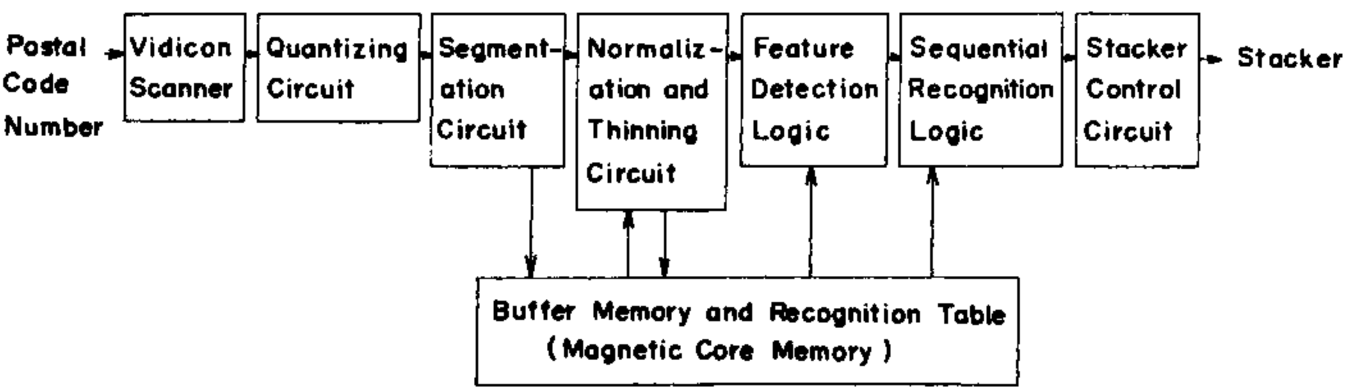

Fig. 2. Block diagram of the handwritten postal code number reader-sorter.

Fig. 3. Symbolical representation of stroke segment directions.

0	1	1
0	1	1
1	1	0

0	0	1
0	1	1
1	1	0

0	0	1
0	1	0
1	1	0

0	0	1
0	1	0
1	0	0

0	0	1
1	1	1
1	1	0

Fig. 4. Binary configurations within 3 x 3 matrix classified as "INC".

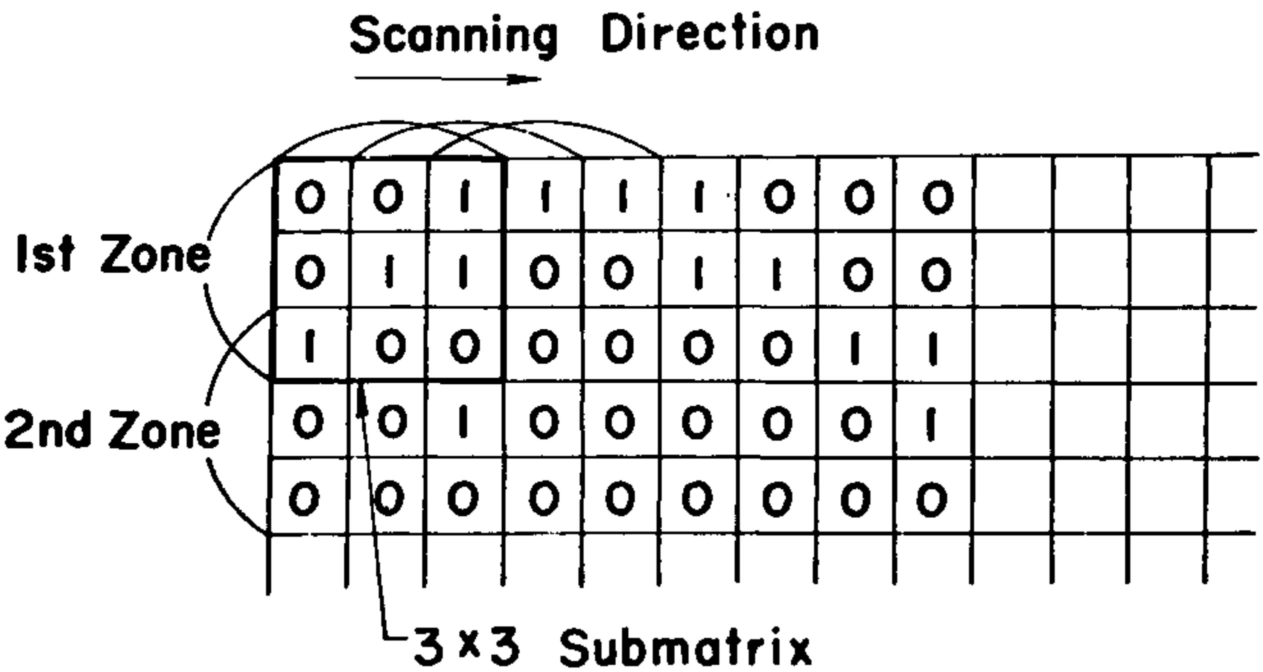

Fig. 5. Overlapped scanning of horizontal zones.

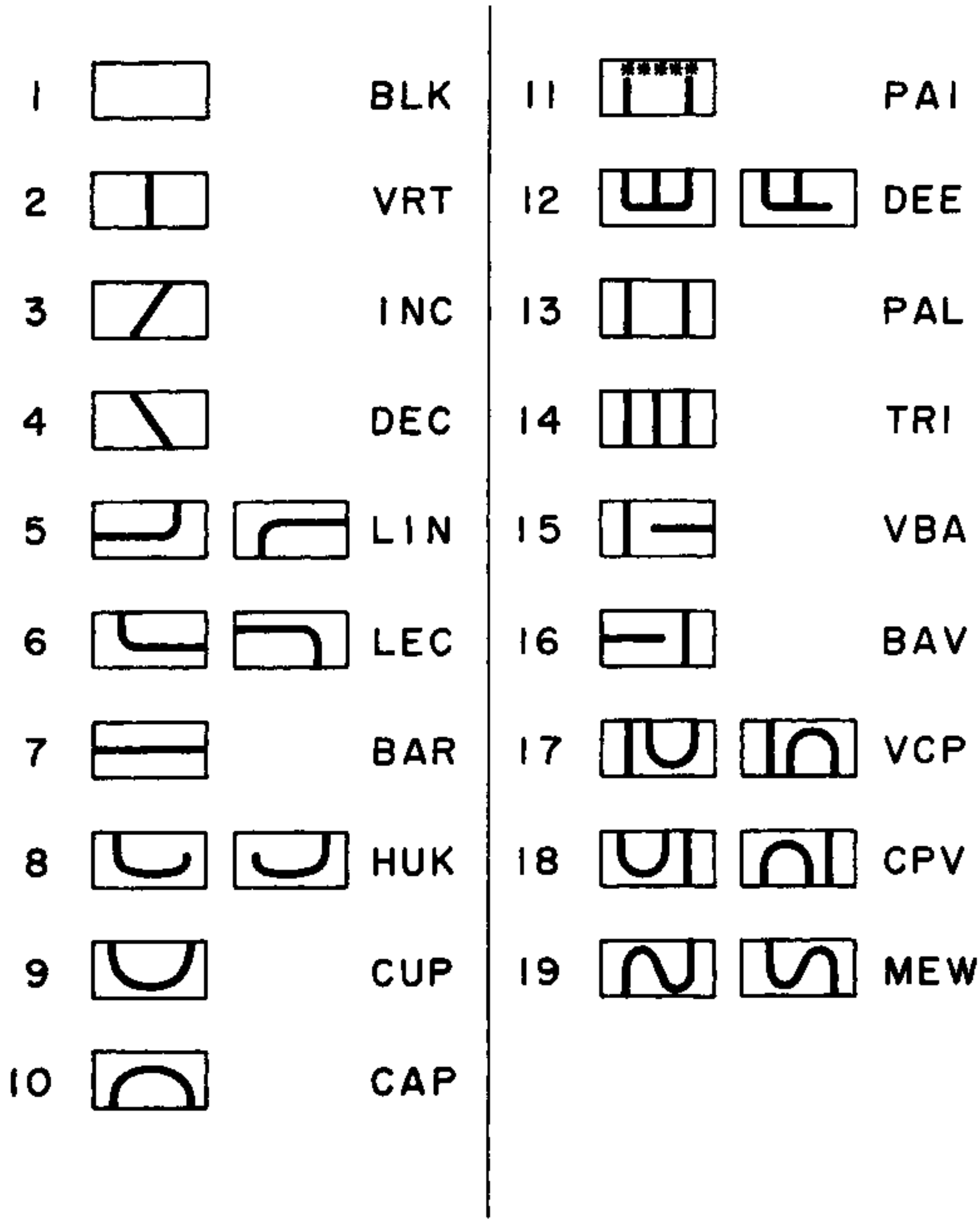

Fig. 6. Horizontal features and their symbolical illustrations.

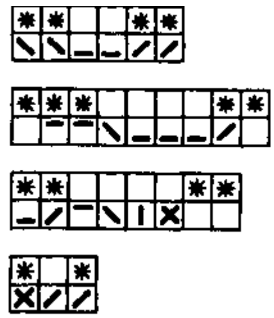

Fig. 7. Examples of strings of stroke segments classified as "CUP".

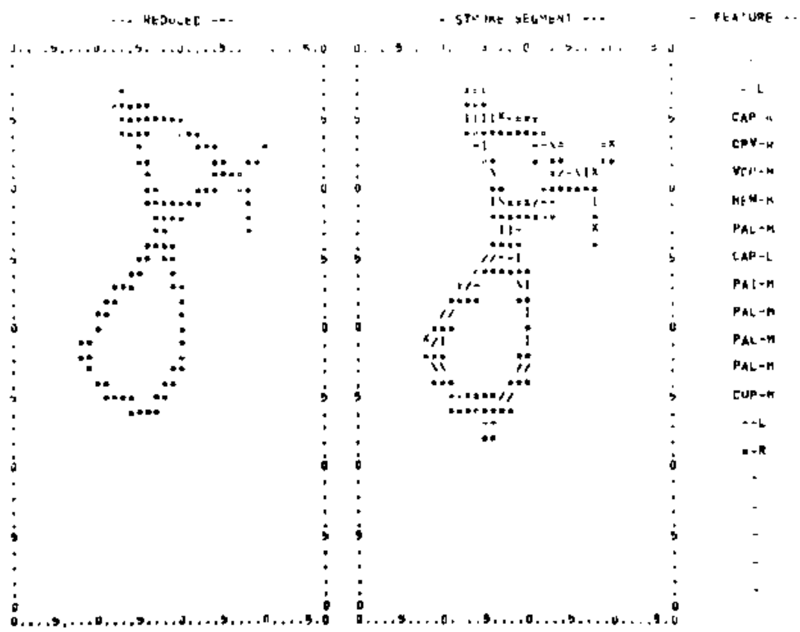

Fig. 8. Example of feature detection by computer simulation. The binary pattern at the left is scanned and converted to an array of stroke segment directions together with connective tags as is illustrated at the middle ("-" stands for "HIGH", "=" for "LOW", "x" for "indecided"). Horizontal features are detected by a string of extracted stroke segments in the horizontal zones as shown at the right.

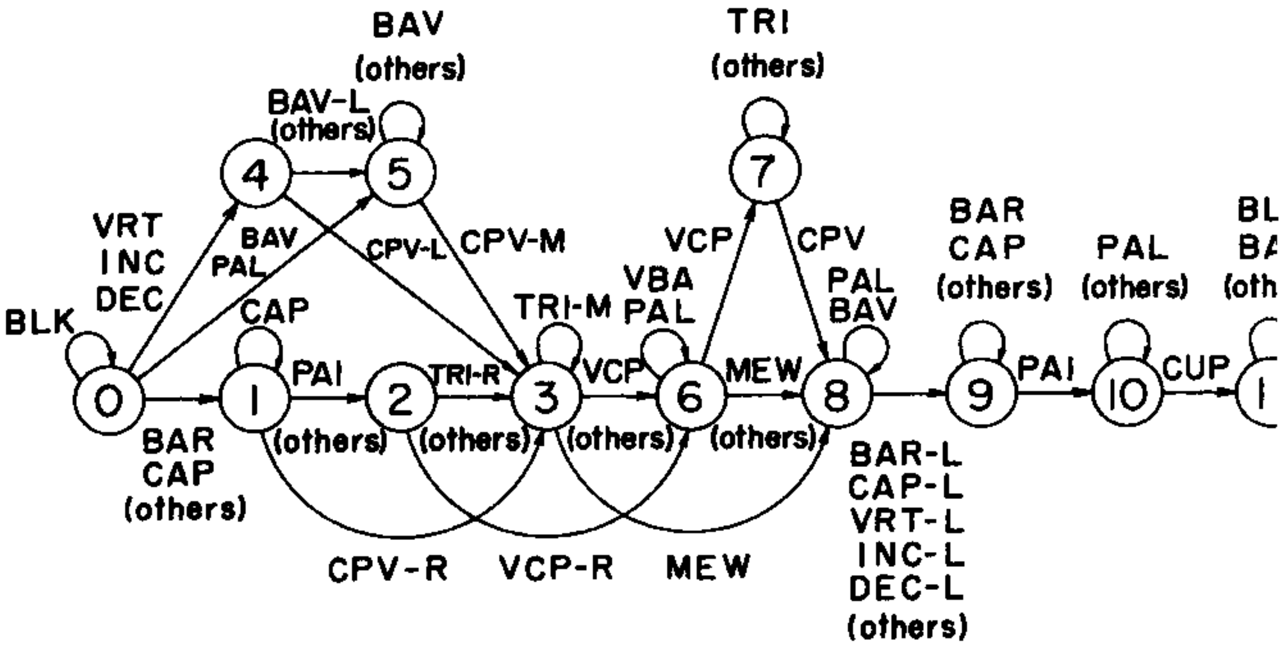

Fig. 9. Example of a sequential transition diagram which identifies one of the variations of the numeral "8". Circled numbers indicate internal state. Arrows indicate the direction of transition. Symbols over the arrows indicate input feature meeting the transition condition. For example, for the input sequence "BLK-M, BLK-L, CAP-R, CPV-R, VCP-M, MEW-M, PAL-M, CAP-L, PAI-M, PAL-M, PAL-M, PAL-M, CUP-M, BLK-L, BLK-R, BLK-M, ... " the internal state transits from the initial state 0 to 0, 0, 1, 3, 6, 8, 8, 9, 10, 10, 10, 10, 11, 11, 11, 11, . . . successively, and no penalty is encountered. But for the sequence "BLK-M, BLK-L, CAP-M, PAI-M, VRT-L, VBA-R, MEW-M, PAL-M, PAL-M, CUP-M, BAR-M, BLK-M, . . . " the internal state transits as 0, 0, 1, 2, 3, 6, 8, 8, 8, 9, 9, 9, . . . and a penalty condition is encountered more than once. The former sequence is accepted by this diagram but the latter is rejected.