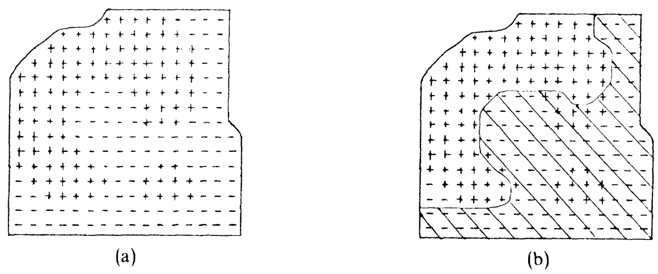

Fig. 1. Two-class problem in R^2. (a) Problem. (b) Decision boundary of TLU.

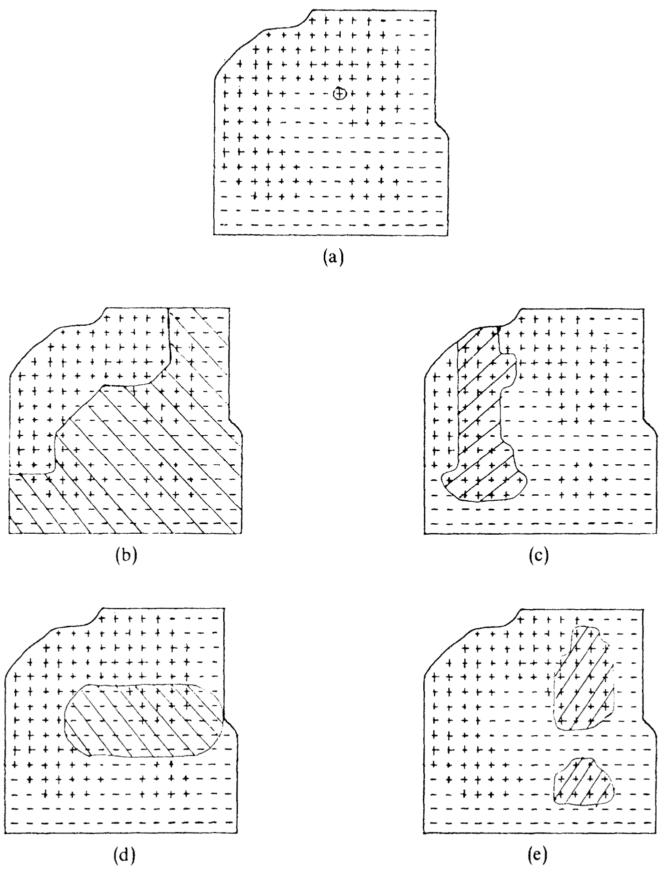

Fig. 2. Performance of seniority committee. (a) Committee's recognition rate. (b) Member one—class I or abstain. (c) Member two—class II or abstain. (d) Member three—class I or abstain. (e) Member four—class II or abstain.

real variables by placing 15 and 14 equally spaced thresholds along Fig. 1(a)'s horizontal and vertical axis, respectively.

A TLU cannot achieve a high recognition in the problem's system of real variables, but in the derived system of binary variables a TLU has a recognition rate of 88 percent (see Fig. 1(b)). A committee of four TLU's trained in the system of binary variables has a much higher recognition rate than a single TLU and can recognize all the patterns except the one circled in Fig. 2(a).

The committee, called a seniority committee, and its training algorithm are described in [3]. Figures 2(b)–(e) show the decision boundaries implemented by the committee members. Committee members one and three (two and four) classify patterns within their shaded region as belonging to class I (class II) and abstain from classifying the remaining patterns. The committee as a whole classifies a pattern according to the decision of the highest numbered nonabstaining committee member. If all the committee members abstain, the committee assigns a pattern to class II.

Finally, it seems reasonable to hope that mapping real variables to binary variables could be useful when a pattern recognition problem is multiclass and when networks of TLU's other than a committee are used.

References

[1] R. O. Duda and H. Fossum, "Pattern classification by iteratively determined linear and piecewise linear discriminant functions," *IEEE Trans. Elect. Comput.*, vol. EC-15, no. 2, pp. 220–232, Apr. 1966.

[2] A. N. Mucciardi and E. E. Gose, "A comparison of seven techniques for choosing subsets of pattern recognition properties," *IEEE Trans. Comput.*, vol. C-20, no. 9, pp. 1023–1031, Sept. 1971.

[3] M. L. Osborne, "A modification of veto logic for a committee of threshold logic units and the use of two-class classifiers for function estimation," Ph.D. dissertation, Oregon State University, 1974.

[4] J. E. Shoenfelt *et al.*, "Techniques for efficient encoding of features in pattern recognition," *IEEE Trans. Comput.*, vol. C-20, no. 9, pp. 1104–1106, Sept. 1971.

Handwriting Identification by Means of Run-Length Measurements

B. ARAZI

Abstract—**A handwritten text was scanned and digitized. The histogram of the run-lengths of the background intensity value was recorded. It is shown that the similarity between the histograms of two samples of the same person's handwriting is greater than that between the histograms of samples of two different person's handwriting. It is also shown how some characteristics of a handwriting can be measured quantitatively.**

Introduction

When a handwritten text is scanned and digitized, every picture element (pel) has one of two possible values. Let the background value and the written line value be denoted by a zero and a one, respectively. When the pel values obtained from one scan line are observed, it is noticed that there are short runs of value 1, which vary only slightly from the average length, and relatively longer runs of value 0 with a higher variation. The reason is that the length of runs of value 1 depends mostly on the thickness of the written line, which in turn depends on the pen used. The length of runs of value 0, on the other hand, depends on some characteristics of the handwriting such as the space between letters and between words, or the size of the empty area inside a letter. If a certain handwriting is scanned by a few hundred scan lines, and the histogram of the run-lengths of the zero value is recorded, then by observing the histogram it is possible to study some characteristics of the handwriting.

Specimens of the handwriting of 13 different persons, written with the same spacing between lines, were investigated in our experiment. Four samples of each handwriting were scanned by 300 scan lines, and the histogram of the run-lengths of zero values was recorded for each sample. Each sample was scanned horizontally (parallel to the written line) and vertically. Some of the specimens of handwriting used in the experiment are shown in Fig. 1. The scanning resolution is demonstrated on the handwriting at the bottom. Observing the vertical scanning resolution, it is obvious that a run of length greater than 120 occurs only if spaces between words are one below the other. Such runs cannot show any characteristic property of the handwriting, and the histogram of vertical scanning therefore stops at a run-length of 120. In horizontal scanning, the histogram for all possible run-lengths was recorded.

Manuscript received October 1, 1976; revised July 12, 1977.

The author was with the National Electrical Engineering Research Institute, Pretoria, South Africa. He is now with CSIR, Pretoria 0001, South Africa.

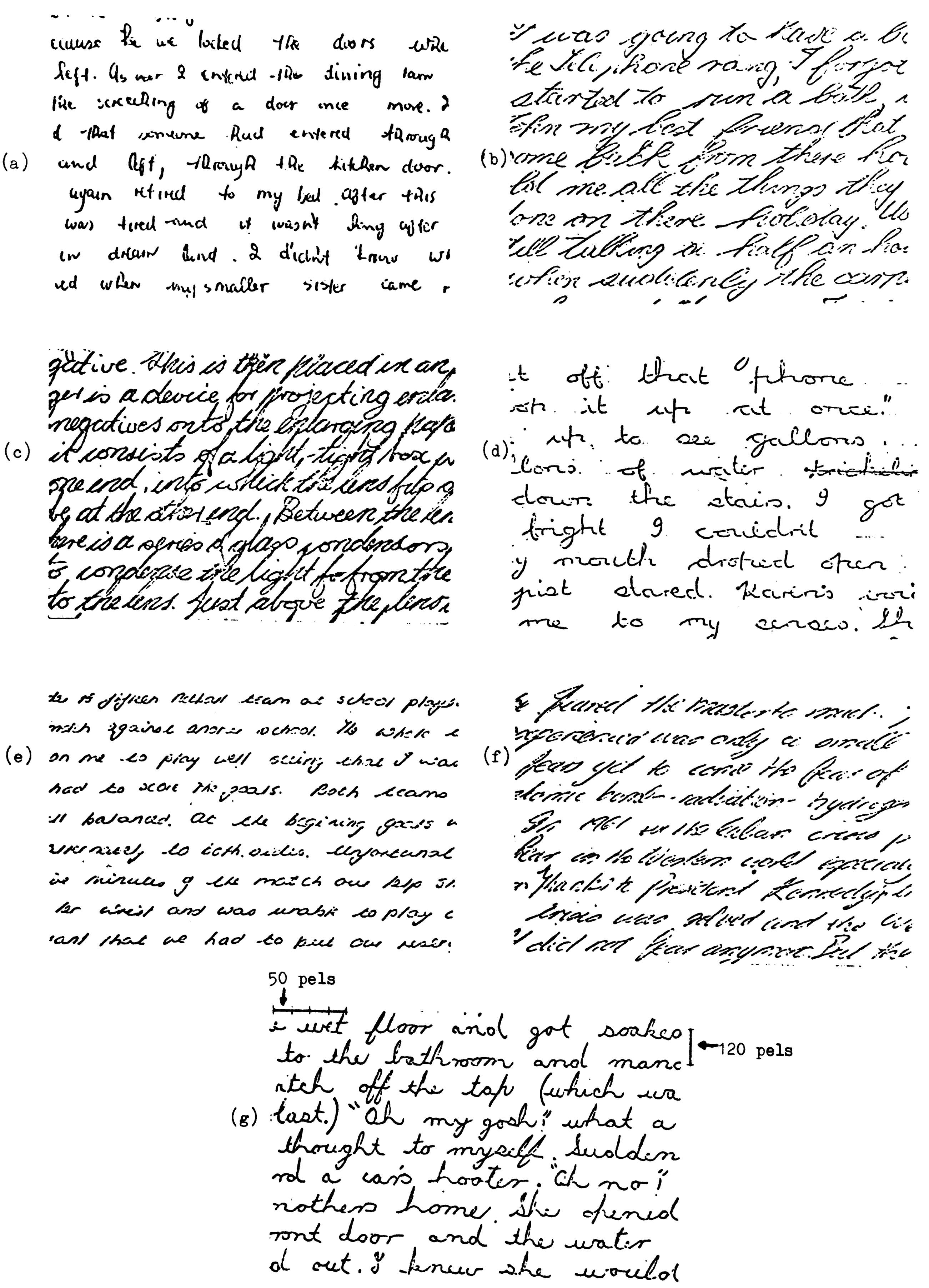

Fig. 1. Some specimens of handwriting used in experiment.

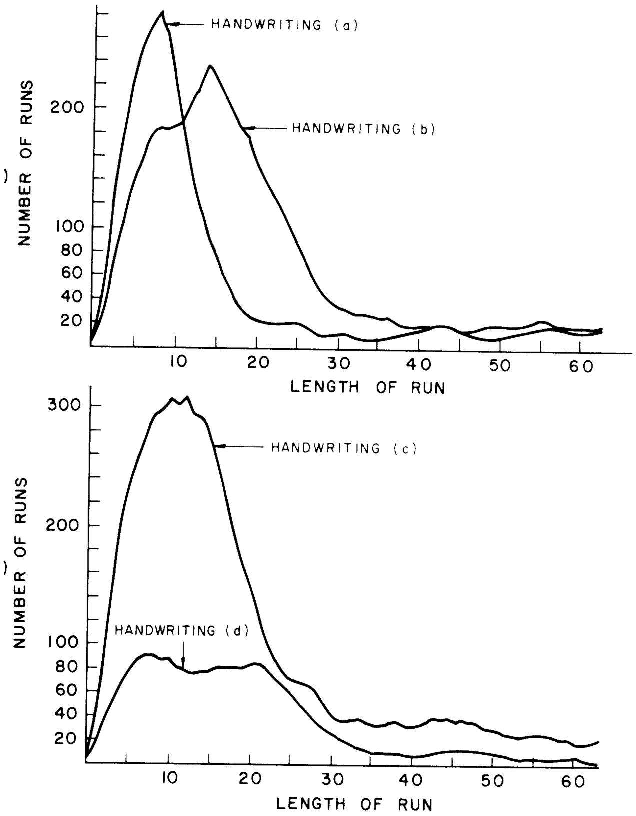

Fig. 2. Some run-length histograms obtained from horizontal scanning.

Before the histograms which were obtained were investigated in detail, a "smoothing" operation was performed by replacing the actual value of each point on the histogram by the average of that value and the values of some adjacent points. This was done since only the general shape of the histogram contains the necessary information on the handwriting, small variations of values being due to the specific sample taken.

Horizontal Scanning

Most of the horizontally scanned runs were short. The histograms obtained from some specimens of handwriting shown in Fig. 1 are given in Fig. 2 (only the first 63 lengths are shown). The location of the peak gives information on the average width of the letters. The wider the letters, the longer are the runs, and the further the peak will be shifted to the right. It is evident from Fig. 2(a) that the peak value for handwriting (a) corresponds to run-length 8, while the peak value for handwriting (b) corresponds to run of length 15. Inspection of these two specimens of handwriting in Fig. 1 shows that the letters of handwriting (b) are wider.

Another property of the handwriting can be studied by considering the width of the peak near the origin of the histogram. The wider this peak, the greater is the deviation of the run-length from the average. This property is also demonstrated in Fig. 2(a). Handwriting (b) is more irregular than (a).

Fig. 2(b) shows the histograms of another two specimens of handwriting. The main difference between them lies in the height of the peak. Examining these two specimens in Fig. 1 it is seen that the density of writing in (c) is higher than that in (d). Considering the fact that both samples were scanned by the same number of lines, it is clear that handwriting (c) had to produce a greater number of runs, which is apparent from Fig. 2(b). In addition to these handwriting characteristics there are others which are also revealed by the shape of the histograms, but which are less obvious than the ones mentioned. It was therefore conjectured that handwriting can be identified by the histogram of run-lengths of value zero obtained from it.

TABLE I
SIMILARITY BETWEEN HANDWRITING OF DIFFERENT PERSONS (HORIZONTAL SCANNING)

	a	b	c	d	e	f	[illegible]
a	100	190	150	262	[illegible]	[illegible]	[illegible]
b	462	100	345	[illegible]	352	[illegible]	[illegible]
c	410	388	100	[illegible]	518	[illegible]	[illegible]
d	513	320	465	100	186	[illegible]	[illegible]
e	358	246	321	[illegible]	100	222	[illegible]
f	418	231	342	238	204	100	[illegible]
g	399	209	284	303	[illegible]	181	100

Experimentally, the similarity between two histograms was measured as follows. For each run-length, the absolute value was found of the difference between the number of runs in the two histograms, and the sum of these values for all possible run-lengths was recorded. This sum can be expected to be low if the two histograms were obtained from different samples of the same handwriting.

The run-length histograms of two samples taken from the same handwriting were recorded for specimens of handwriting from 13 different persons. Let A_1^j and A_2^j denote the two histograms obtained from the jth handwriting. Another two run-length histograms were recorded of another two samples of the ith handwriting. Let H_1^i and H_2^i denote these last two histograms. The similarity between H_1^i and A_1^j and between H_2^i and A_2^j for $1 \leq j \leq 13$ was measured, and the average of each such pair of measurements taken. The results of these measured averages were then expressed as a percentage of the smallest value obtained. Table I lists the results obtained when comparing the specimens given in Fig. 1.

It is noticed that an element in the diagonal of the array has the lowest value in its row, which means that the minimum result is obtained when histograms of two samples of the same handwriting are compared. When all 13 specimens were compared, the value nearest to the minimum obtained was higher by at least 20 percent.

Vertical Scanning

When handwriting is scanned vertically, the histogram of the run-lengths usually has two peaks. One of these occurs in the range of short runs. This peak, like the one obtained for horizontal scanning, gives information on the structure of the letters. A second peak is the result of many runs which correspond to the distance between the bottom of a line of writing and the top of the letters in the line below. The location of this peak gives information on the average size of the letters. The larger the letters, the shorter is the mentioned distance, and the closer to the origin this second peak will be. This is demonstrated in Fig. 3(a). The second peak of handwriting (c) is closer to the origin than that of (e). This is due to the fact that the letters of handwriting (c) are larger, as can be seen in Fig. 1.

The width of the second peak shows how the size of the letters deviates from the average. Fig. 1 shows that handwriting (c) is more regular than (b). This difference can be noticed in Fig. 3(b). The histogram of (b) hardly has a second peak.

Another characteristic of handwriting is its slope. If a handwritten text is scanned vertically, the number of short runs increases with the slope. Comparing specimens (f) and (g) in Fig. 1, the first slopes more than the second. Their histograms in Fig. 3(c) show that handwriting (f) has a greater number of short runs. (It must

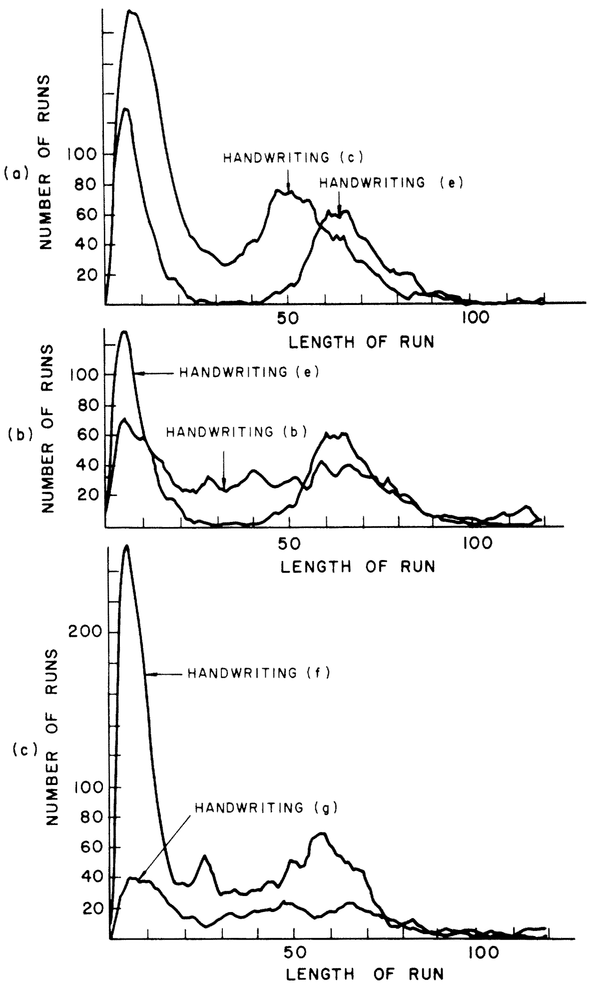

Fig. 3. Some run-length histograms obtained from vertical scanning.

be emphasized that another reason for this difference in the number of short runs is the difference in the densities of the two specimens).

The similarity between different specimens of handwriting was measured by comparing the run-length histograms obtained from vertical scanning. In contrast to what resulted when the data obtained from horizontal scanning were used, some measurements failed in that they led to faulty identification. Nine out of the 13 specimens of handwriting were found suitable for similarity measurements. For these it was found that if the squares of the values of each histogram are taken and similarity measurement performed, then the identification results are improved. (In the sense that if a table similar to Table I is compiled, the difference between the diagonal elements and the off-diagonal ones is increased.)

This result suggests that when scanning vertically, the information on the individual handwriting is more concentrated in and around the peaks. This assumption is supported by observing the histograms of those specimens whose similarity check failed. It was found that in all of them the second peak was either very flat or could not be recognized at all (e.g., specimens (b) and (g) shown in Fig. 3). This means that if the height of the written letters is too irregular, there is not enough information to enable the handwriting to be identified. It was found experimentally that handwriting can be considered as regular enough for purposes of identification only if the second peak has a width less than a certain value. Owing to this limitation, it can be said that the similarity check using run-length histograms obtained from vertical scanning is less powerful than that using horizontal scanning. On the other hand, the information which can be gathered on a certain handwriting observing the histogram obtained from vertical scanning is an important feature in itself.

TABLE II
SIMILARITY BETWEEN HANDWRITING OF DIFFERENT PERSONS (VERTICAL SCANNING)

	a	c	d	e	f
a	100	566	604	651	662
c	961	100	359	390	398
d	766	269	100	180	181
e	1263	446	275	100	141
f	1660	589	357	182	100

TABLE III
SIMILARITY OF HANDWRITING OF DIFFERENT PERSONS (AVERAGE OF HORIZONTAL AND VERTICAL SCANNING)

	a	c	d	e	f
a	100	358	443	431	465
c	685	100	529	454	500
d	639	367	100	183	230
e	810	383	224	100	181
f	1039	465	297	193	100

Five out of the seven specimens of Fig. 1 were found suitable for measurement. Their comparisons are listed in Table II. The values were calculated in exactly the same way as those for Table I, except that the measurements were performed on the squares of the values in the histograms.

A further test was to compare the values obtained by taking the average of horizontal and vertical scanning. The following table shows the average of horizontal and vertical scanning of the seven specimens of Table II.

When listing a table similar to Table III for all the nine specimens found suitable for vertical scanning analysis, there was a difference of at least 45 percent between the lowest value and the value nearest to it (i.e., the lowest value in any row of the table beside 100 percent was 145 percent). This is due to the fact that specimens which had close values when horizontal scanning analysis was performed had larger differences when scanned vertically, and vice-versa. By taking the described average, the handwriting is therefore identified with a relatively high degree of accuracy.

The Turing Machine Constructed by Trainable Threshold Elements

EIICHI TANAKA

Abstract—**According to recent research, threshold nets have some interesting functions, i.e., they can recognize both spacial and temporal patterns and memorize those patterns associatively. However, there is no report on constructing the Turing machine by trainable threshold elements. The realization of a finite automata and the Turing machine by a threshold net is described. The transformation between a finite automaton and an autonomous automaton is first discussed. The Turing machine is constructed by threshold elements, and its instructions can be changed by learning.**

Manuscript received February 21, 1975; revised July 6, 1977.
The author is with the Department of Information Science, Faculty of Engineering, Utsunomiya University, Utsunomiya 321-31, Japan.