Information Processing in Japan Vol. 11, 1971

A Pattern Classification System using Equi-Variance Characteristic Parameters

YOSHIO NOGUCHI* AND TAIZO IIJIMA*

1. *Introduction*

We consider a rectangular area divided into $n_1 \times n_2$ meshes, where each mesh has a value either 0 (white) or 1 (black). In this case when the values of n_1 and n_2 are large, the dimensionality of pattern is very large and large amount of redundancy of pattern is included. As the process of pattern recognition can be considered as a sort of information compression, because its purpose is to ignore the information which is irrelevant to class features. We propose, in this paper, the pattern recognition system which optimally compresses the observed raw data to reduce the pattern dimensionality.

Regarding a pattern feature as a vector whose elements, in a certain sense, numerically asses properties of an aggregate of patterns, the processing which extracts pattern feature reducing the dimensionality of pattern is called feature extraction.

One of these feature extraction methods is known as factor analysis. Let the number of pattern categories be L, the number of meshes be K, and in the case of inequality $K>L$ be valid, eigenvalue equation must be solved in order to determine the feature space.

But it is not appropriate to apply this method directly to the observed pattern on retina. Because the feature extraction processer must calculate the scalar product between vectors with large dimension. Beside this method, we propose the sampling processing as one of the preprocessings of which aim is to reduce the dimensionality of the pattern. By this sampling processing, the observed pattern is transformed into the sampled pattern. By means of this processing, the similarity between any two patterns becomes large, but the distortion of pattern owing to the existance of noise becomes smaller. Moreover the canonicalization processing is applied in order to eliminate the average pattern from each individual objected pattern.

The feature extraction method defines the optimal coordinate system for the pattern representation space. In this coordinate system, the average square distance of the component of each representation vectors to the coordinate axis is sharply concentrated on a few coordinates. Each one of these coordinate

This paper first appeared in Japanese in Joho Shori (the Journal of the Information Processing Society of Japan), Vol. 12, No. 1 (1971), pp. 26-36.
* Electrotechnical Laboratory.

axis has different weight, therefore this coordinate system is not suitable to be used as the classification space. We propose, furthermore, the Affinian transformation of the feature space, in order to equalize the mean-square distance between a set of vectors and the coordinate axis, subjected to the constraint that the length of each axis is equal. This N-dimensional space which is spanned with these coordinate axis is name as *equi-variance space*, in this pattern.

For the convenience of classification procedure, we use the code distance as classification criteria.

The block diagram of the proposed system is shown in Fig. 1.

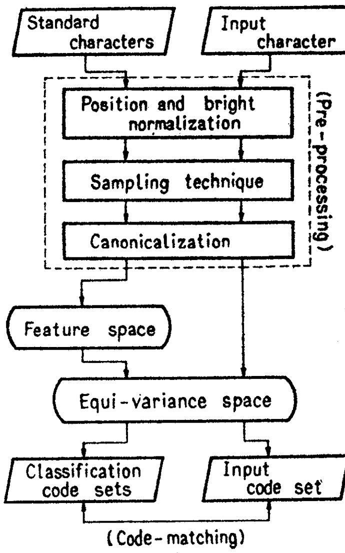

Fig. 1. Block diagram of this pattern recognition system (design process is left flow and classification process is right flow).

2. *Preprocessing*

Preprocessing means many kinds of processing which normalize the observed object pattern before the feature extraction.

By means of these processings various kinds of parameters of pattern can be adjusted to normaled point. Positional normalization and bright normalization of object pattern are used in this recognition system.

Let $e(x_i, y_j)$ be quantized character on retina, and let $f(x_i, y_j)$ be normalized character, then quantized character is transformed as follows;

$$f(x_i, y_j) = e(x_i + [x_a], y_j + [y_a]) - k_0, \tag{1}$$

where x_a, y_a, k_0 are normalizing parameters, and $[x]$ means greatest integer $\leq x$. It is necessary that these parameters satisfy following equations.

$$\sum_{j=1}^{43} \sum_{i=1}^{31} e(x_i + x_a, y_j) \sin\left(\frac{\pi x_i}{21_x}\right) = 0$$

アイウエオ
カキクケコ

Fig. 2. JEIDA-5 black- and-white Japanese Katakana (scale: 5/1).

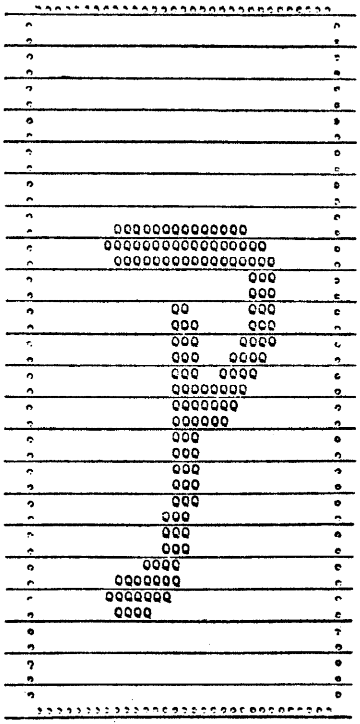

Fig. 3. Standard quantized chracter with 31×43 meshes.

$$\left.\begin{aligned}&\sum_{i=1}^{31}\sum_{j=1}^{43} e(x_i,\ y_j+y_a)\sin\left(\frac{\pi y_j}{21_y}\right)=0\\ &k_0=\frac{1}{41_x 1_y}\sum_{j=1}^{43}\sum_{i=1}^{31} e(x_i,\ y_j)\end{aligned}\right\} \qquad (2)$$

where $21_x=3.10$ and $21_y=4.30$.

Let $g(x_i, y_j)$ be a sampled pattern transformed from $f(x_i, y_j)$, then

$$g(x_i, y_j)=\sum_{t=-3}^{3}\sum_{s=-3}^{3}\left[e^{-\frac{(x_i-u_{i+s})^2+(y_j-v_{j+t})^2}{2\sigma^2}}\times f(u_{i+s},\ v_{j+t})\times\frac{c}{\sigma}\right], \qquad (3)$$

where c is constant, $1\leq i\leq 11$, $1\leq j\leq 15$, and σ is parameter of avoid and it's value is 0.177.

Gaussian weights in the above formula is shown in Fig. 4. Sampling interval of this sampling process is 0.30 mm on the rectangular area of retina.

As the value of component of sampled pattern is a real value, then sampled character can be assumed as a vector. Let $\boldsymbol{g}_r$ be the sampled standard character

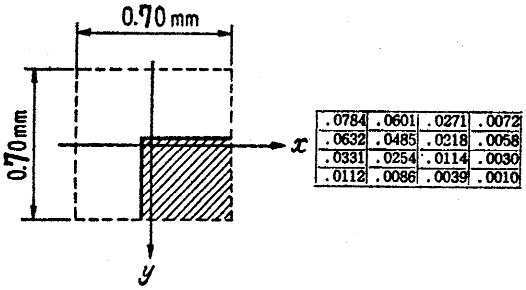

Fig. 4. Partial list of weighting fun.

which belong to the r-th category among the set of L categories.

Let $\boldsymbol{\tau}$ be the average character among the set of sampled standard character, then

$$\boldsymbol{\tau}=\frac{\sum_{r=1}^{L}\boldsymbol{g}_r}{\|\sum_{r=1}^{L}\boldsymbol{g}_r\|}. \tag{4}$$

If we subtract the average character from each representation vector, the canonicalized character can be obtained as follows

$$\boldsymbol{h}_r=\boldsymbol{g}_r-\alpha_r\boldsymbol{\tau}\;;\;\alpha_r=(\boldsymbol{g}_r,\boldsymbol{\tau}) \tag{5}$$

The norms of $\boldsymbol{g}_r$, $\boldsymbol{h}_r$ and the value of α_r are shown in Table 1 as example.

Table 1. Norm of sampled standard Characters, norm of canonicalized ones, and factor size of the average character.

	$\|g^{(r)}\|$	$\alpha^{(r)}$	$\|h^{(r)}\|$
ア	0. 407977	0. 320105	0. 252943
イ	0. 364605	0. 228557	0. 284075
ウ	0. 415422	0. 289514	0. 297924
エ	0. 420333	0. 291749	0. 302594
オ	0. 448944	0. 340923	0. 292101

3. *Patterns in Feature Space*

Pattern which passed through the preprocessing is represented as a point in the K-dimensional vector. This dimension is very large compared with the number of categories. N.B.: $K=11\times15$

In order to reduce the dimensionality of measurement vectors and at the same time to eliminate the less important variables, we apply the feature extraction process. If we make the assumption that each standard pattern has equal probability density, then optimal coordinate system can be obtained solving the eigen value equation:

$$\lambda_m\varphi_m=\boldsymbol{A}\varphi_m\;;\;a_{ij}=\frac{1}{L}(\boldsymbol{h}_i,\boldsymbol{h}_j), \tag{6}$$

where a_{ij} is an element of matrix $\boldsymbol{A}$ and L is the total number of categories.

We agree to label the normalized eigen vectors in the descending order of the corresponding eigen values.

Then the Euclidian space with normalized eigen vectors as coordinate axes is named as the feature space.

In this space, each character may be expressed as

$$\boldsymbol{h}_r = \sum_{m=1}^{L-1} \beta_{rm}\varphi_m\,;\ \beta_{rm} = (\boldsymbol{h}_r, \varphi_m), \tag{7}$$

where the expansion coefficients $\{\beta_{rm}\,;\ m=1, 2, \cdots, L-1\}$ are named as *feature coefficients* in this paper. Let $\boldsymbol{\Phi}$ be a $(L-1)\times K$ matrix with ${}^t\varphi_i$ as the i-th row vector. Then a pattern in the K-dimensional sampled space is transformed into a point in the $L-1$ dimensional feature space as follows;

$$\boldsymbol{\beta}_r = \boldsymbol{\Phi}\boldsymbol{h}_r, \tag{8}$$

where $\boldsymbol{\beta}_r$ is a column vector with β_{rm} as the m-th row element.

Examining pattern with various kinds of noise in the feature space, it became clear that the amount of noise which was generated by the parallel shift of the pattern over the original retina, is larger than any other kinds of noise.

Let $\boldsymbol{m}_r$ be the translation-noise vector of r-th object of pattern in the K-dimensional sampled space, then translation-noise in the feature space can be defined as

$$\boldsymbol{\varepsilon}_r = \boldsymbol{\Phi}\boldsymbol{n}_r.$$

Fig. 5. shows the level of translation-noise.

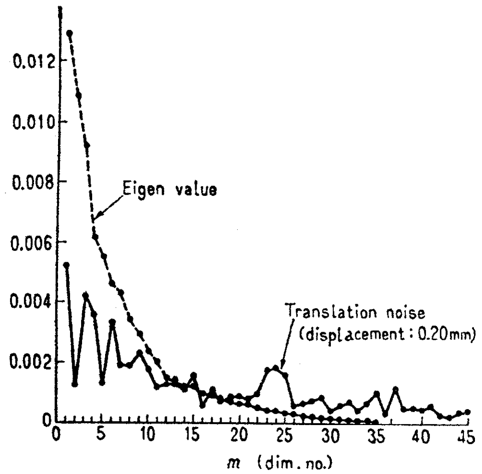

Fig. 5. Comparison of average power-spectra between the standard characters and the translation-noises.

4. *Patterns in Equi-Variance Space*

If these eigen vectors are put into devices as the classifiers of the pattern, it can be assumed that accuracy of device is uniform at each device. However the average amount of measurements of the pattern set which are extracted by eigen vectors is not equal. Then the lossness of pattern information calculated by means of this device is different at each classifier. On the other hand, the subspace of the feature space is enough to extract the greater part of pattern. From these points of view, we constract the new space for the purpose of classification using the subspace of the feature space to equalize the information measurements which are extracted at each axis. We name this new space as *equi-variance space.*

Let χ_k be a new coordinate axis, then the axis is defined as follow,

$$\begin{cases} \chi_k = \sum_{m=1}^{N} u_{km} \frac{\varphi_m}{\sqrt{\lambda_m}} \\ \sum_{k=1}^{N} u_{km} u_{kn} = \delta_{mn} \end{cases} . \tag{10}$$

Let $\{\chi^l;\ l=1, 2, \cdots, N\}$ be the reciprocal system of $\{\chi_k\}$, then reciprocal system satisfies

$$(\chi_k,\ \chi^l) = \delta_k{}^l. \tag{11}$$

In this transformation, $\{u_{km}\}$ is defined by the linear combination of $\{\chi_k\}$. The column vector χ^l must satisfy the condition

$$\|\chi^l\| \fallingdotseq \text{const}. \tag{12}$$

As the space whose coordinate axes $\{\chi_k\}$ is not an orthogonal space, representative pattern is expanded with reciprocal system. Let $\boldsymbol{h}_{Nr}$ be the pattern vector within N-dimensional-subspace of the feature space, then $\boldsymbol{h}_{Nr}$ can be expanded as follows;

$$\begin{cases} \boldsymbol{h}_{Nr} = \Sigma \gamma_r{}^R \chi_k \\ \gamma_r{}^k = (\boldsymbol{h}_r,\ \chi^k) \end{cases} . \tag{13}$$

Let χ be $N \times K$ matrix with ${}^t\chi^i$ as i-th row vector, then

$$\chi^t \chi \fallingdotseq \begin{pmatrix} \mu_0 & \mu_1 & \mu_2 \cdots & \vdots \\ \mu_1 & \mu_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mu_1 \\ \cdots & \cdots & \mu_1 & \mu_0 \end{pmatrix} \tag{14}$$

where $\mu_0 = \max \mu_i$.

The expansion coefficients in reciprocal system is named as *equi-variance characteristic parameters.* With respect to these parameters, let $\boldsymbol{C}$ be the $N \times L$ matrix with $\boldsymbol{\gamma}_r$ as r-th column vector, then

$$\boldsymbol{C}^t \boldsymbol{C} = \begin{pmatrix} 1 & & 0 \\ & 1 & \\ & & \ddots \\ 0 & & & 1 \end{pmatrix} \tag{15}$$

5. *Clustering*

Using these characteristic parameters, many kinds of classification techniques can be applied. For simplicity, the code distance can be used as an appropriate criterion of decision in classification. From this point of view, the set of N characteristic parameters is encoded as a N-bits code system corresponding to the pattern with a suitable threshold value. In the equi-variance space, the components of patterns are distributed evenly over each coordinate, then each characteristic parameter calculated at and axis can be treated on a footing of equality. Therefore each bit in the code system obtained by means of encoding the characteristic parameter has an equal weight, thus each bit can also be treated with equality.

In order to obtain the reference classification code system, the characteristic parameter is encoded into 3-level symbol. Let $[\gamma_r{}^k]$ be an encoded symbol, then it can be obtained as: $[\gamma_r{}^k]=1$, if $\gamma_r{}^k \geq d$; $[\gamma_r{}^k]=0$, if $-d \leq \gamma_r{}^k < d$; $[\gamma_r{}^k]=-1$, if $\gamma_r{}^k \leq -d$, where d is the threshold value. Let $\{\gamma_r{}^k\}$ be distributed according to N(0, 1) and let the probability of occurance of 1, 0, -1 be equal with each other, then threshold value becomes $d=0.4308$. In addition to this, let $D([\gamma_r{}^k], [\gamma_s{}^k])$ be the code distance at k-th bit in code systems corresponding to $\boldsymbol{h}_r$ and $\boldsymbol{h}_s$, then it can be defined as follows,

$$D([\gamma_r{}^k], [\gamma_s{}^k]) = \begin{cases} 1 ; & [\gamma_r{}^k][\gamma_s{}^k] = -1 \\ 0 ; & [\gamma_r{}^k][\gamma_s{}^k] = 0 \text{ or } 1 . \end{cases} \tag{16}$$

Let $[\boldsymbol{h}_{Nr}]$ be the code system corresponding to $\boldsymbol{h}_{Nr}$, and let $D([\boldsymbol{h}_{Nr}], [\boldsymbol{h}_{Ns}])$ be the distance of code system between $\boldsymbol{h}_{Nr}$ and $\boldsymbol{h}_{Ns}$, then the distance is defined as follows;

$$D([\boldsymbol{h}_{Nr}], [\boldsymbol{h}_{Ns}]) = \sum_{k=1}^{N} D([\gamma_r{}^k], [\gamma_s{}^k]). \tag{17}$$

On the other hand, in the case of input pattern, the characteristic parameter is encoded into 2-level symbol. Let γ^k be the k-th equi-variance characteristic parameter, and let $[\gamma^k]$ be the code symbol derived from γ^k, then $[\gamma^k]$ is defined as $[\gamma^k]=1$, if $\gamma^k \geq 0$; $[\gamma^k]=-1$, if $\gamma^k < 0$. As a recognition criterion, it is expedient to use the shortest distance of code system from the unknown code system corresponding to the input pattern to each reference-classification code system.

6. *Computer Simulation Results*

The effectiveness of this recognition system has been tested experimentally for the case of the recognition of the full set of typed style Japanese Katakana. In case of putting the standard pattern as an input vector, each input pattern has been assigned exactly to one of the category among the set of 46 categories. But for the input pattern which includes noise element, this decision rule does not assign the input pattern to the correct category, as noise disturbs the distribution of pattern in the feature space. In the case of deviation of pattern in the feature space, the noisy pattern and the similar standard pattern are distributed within the neighborhood, then we can cluster the set of patterns

Table 2. Character groups which are invariant for translation of input character.

r		θ	categories included in the classified cluster
1	ア	0	ア
2	エ	0	エ
3	カ	0	カ
4	シ	0	シ
5	ス	0	ス
6	ナ	0	ナ
7	ハ	0	ハ
8	ラ	1	ラ
9	ウ	0	ウ, ワ
10	ヘ	1	ヘ, ノ
11	ム	1	ム, ネ
12	ミ	1	ミ, メ
13	ロ	1	ロ, ツ, ホ
14	ホ	1	ホ, キ, ヨ
15	ケ	2	ケ, キ, ヒ, リ
16	イ	1	イ, ソ, キ, メ, リ
17	オ	1	オ, キ, サ, ノ, メ
18	メ	1	メ, イ, オ, ノ, ミ
19	ニ	1	ニ, ク, フ, ユ
20	ト	2	ト, ヤ, キ, ツ, ヲ
21	ノ	1	ノ, キ, ク, サ, メ
22	ツ	1	ツ, タ, フ, リ, ヲ
23	リ	1	リ, イ, キ, タ, ツ, テ
24	ユ	1	ユ, マ, キ, ケ, ニ, フ
25	サ	1	サ, オ, ケ, ノ, ヒ, メ
26	キ	1	キ, ネ, テ, ホ, ヨ, リ, ル
27	タ	1	タ, ヌ, ク, ツ, フ, リ
28	モ	2	モ, イ, テ, ヒ, ホ, ミ
29	フ	1	フ, コ, ク, タ, ニ, ユ, ヨ, ヲ
30	レ	2	レ, キ, タ, ツ, テ, ニ
31	ヲ	2	ヲ, キ, ツ, テ, フ, ミ, ヨ, リ
32	ヨ	2	ヨ, ン, コ, キ, ツ, フ, ホ, ユ, リ, ヲ
33	ル	2	ル, キ, サ, テ, ホ, ミ, ム, ユ, リ, レ, ロ
34	ヒ	2	ヒ, セ, ク, サ, リ, ミ, モ, リ, ロ
35	ク	2	ク, タ, テ, ニ, ノ, ヒ, フ, ミ, ユ, ロ
36	テ	2	テ, チ, キ, ク, ニ, フ, ミ, モ, ユ, リ, ル, レ, ロ, ヲ

(θ; threshold value)

into a several groups. By means of computer simulation, character groups which are in variant against the translation noise are shown in Table 2. The range of permissible translation for these character groups is 0.10 mm for any direction. N.B.N := 15.

In order to examine the topological property of the full set of standard patterns and that of reference-classification code cystems, two kinds of distance are investigated. One of these is the average Euclidian distance between the point $\boldsymbol{h}_r$ and the L members of the set $\{\boldsymbol{h}_s\}$, and the other is the mean code distance between reference classification code system $[\boldsymbol{h}_r]$ and the L members of the set $\{[\boldsymbol{h}_s]\}$. The results of these are shown in Table 3 and Table 4.

From the point of investigation of this recognition system, we also examine the distance between each standard pattern and the nearest neighbor standard

Table 3. Mean distance between the standard characters in the 15-dimensional feature space.

Category	distance	Category	distance	Category	distance
ア	.34600	チ	.33064	ム	.41463
イ	.36495	ツ	.35482	メ	.38825
ウ	.37648	テ	.34226	モ	.37420
エ	.38223	ト	.35804	ヤ	.37232
オ	.36561	ナ	.35652	ユ	.35731
カ	.38569	ニ	.35260	ヨ	.32902
キ	.36557	ヌ	.37967	ラ	.35479
ク	.33806	ネ	.34315	リ	.35055
ケ	.35440	ノ	.35645	ル	.36123
コ	.35335	ハ	.39232	レ	.38588
サ	.36525	ヒ	.37692	ロ	.39855
シ	.37657	フ	.34379	ワ	.39064
ス	.38084	ヘ	.38195	ヲ	.33152
セ	.38090	ホ	.36131	ン	.35811
ソ	.35444	マ	.35167	Mean of means =0.36365	
タ	.33973	ミ	.34852		

Table 4. Mean distance between the classification code sets.

ア	3.0652	チ	2.7391	ム	4.0434
イ	3.0869	ツ	2.8260	メ	3.1521
ウ	3.5652	テ	3.3260	モ	4.0217
エ	3.1086	ト	4.3695	ヤ	4.0652
オ	3.2826	ナ	3.3478	ユ	3.1739
カ	4.0000	ニ	2.9130	ヨ	3.7391
キ	2.0000	ヌ	3.4130	ラ	3.3478
ク	3.2173	ネ	2.2391	リ	2.8043
ケ	4.3695	ノ	3.5217	ル	3.3043
コ	4.1304	ハ	3.5434	レ	4.0434
サ	3.3695	ヒ	3.4347	ロ	3.1304
シ	4.1956	フ	2.8913	ワ	3.9565
ス	4.3260	ヘ	4.9565	ヲ	3.3913
セ	4.0869	ホ	2.9565	ン	4.4347
ソ	3.2608	マ	2.2391	Mean of means =3.452	
タ	3.3043	ミ	3.0869		

Table 5. Distance to the nearest neighbor character and norm of translation-noise.

Category	The nearest neighbor character and its distance	The largest norm of noise character	Category	The nearest neighbor character and its distance	The largest norm of noise character
ア	0.2370 (テ)	0.1405	ネ	0.2173 (オ)	0.1492
イ	0.1960 (チ)	0.1548	ノ	0.1935 (メ)	0.1289
ウ	0.1574 (ワ)	0.1371	ハ	0.2793 (ロ)	0.1745
エ	0.2159 (ソ)	0.1330	ヒ	0.1769 (セ)	0.1625
オ	0.2173 (ネ)	0.1625	フ	0.1983 (ク)	0.1344
カ	0.2515 (ヒ)	0.1815	ヘ	0.2970 (ケ)	0.1332
キ	0.2177 (オ)	0.1331	ホ	0.2347 (キ)	0.1590
ク	0.1392 (タ)	0.1429	マ	0.2290 (ネ)	0.1299
ケ	0.1723 (サ)	0.1421	ミ	0.2402 (ニ)	0.1594
コ	0.1562 (ユ)	0.1449	ム	0.3155 (ル)	0.1414
サ	0.1723 (ケ)	0.1772	メ	0.1935 (ノ)	0.1297
シ	0.2091 (レ)	0.1368	モ	0.2362 (ト)	0.1472
ス	0.1751 (ヌ)	0.1444	ヤ	0.2143 (ト)	0.1435
セ	0.1769 (ヒ)	0.1359	ユ	0.1562 (コ)	0.1296
ソ	0.2101 (ク)	0.1469	ヨ	0.1863 (ヲ)	0.1501
タ	0.1392 (ク)	0.1435	ラ	0.1771 (ヲ)	0.1436
チ	0.1960 (テ)	0.1320	リ	0.2191 (ソ)	0.1786
ツ	0.1943 (ウ)	0.1837	ル	0.2545 (ホ)	0.2275
テ	0.1900 (ヲ)	0.1508	レ	0.2091 (シ)	0.1876
ト	0.2143 (ヤ)	0.1664	ロ	0.2793 (ハ)	0.1705
ナ	0.2175 (チ)	0.1466	ワ	0.1574 (ウ)	0.1478
ニ	0.1912 (ユ)	0.1573	ヲ	0.1771 (ラ)	0.1350
ヌ	0.1751 (ス)	0.1455	ン	0.2599 (コ)	0.1323

pattern, and the largest norm of translation-noise character caused by parallel translation of standard pattern within the displacement 0.1 mm. Results are shown in Table 5.

7. *Conclusion*

The character groups are constracted on the basis of translation noise, because translation noise is the largest noise among the various kinds of distortion.

Threshold value used in this paper is not proper, it has been determined considering the norm of noise pattern. The method of feature extraction described here is being investigated very actively, and the relation of this method with the concept of entropy has been discussed. But consideration for this equi-variance space is not discussed. It is possible to consider the other useful decision procedure using the equi-variance characteristic parameters. Moreover it is important to investigate quantitatively the pattern distribution and the range of various kinds of noise elements. For the first step of this investigation, we have done this study. The results of this study can be applied for the arbitrary kind of character sets.

Acknowledgement

It is a pleasure to acknowledge the hospitality and helpful discussion of the members of Iijima Section and Mathematical Engineering Section. We also would like to thank Hisako Sasaki for her considerable assistance and Dr. Katsuhiko Noda for his encouragement.

References

[1] Iijima, T.: Pattern Recognition, Nikkan Kogyo Shinbunsya (1969).
[2] Iijima, T.: "Basic theory of feature extraction for visual pattern." *Jour. I.E.C.E.*, Japan 46, 11 (Nov. 1963).
[3] Iijima, T.: "Extraction theory of equi-variance characteristic parameters.", Technical Report on Automata & Automatic Control (Oct. 1966).
[4] Iijima, T. and Y. Noguchi: "Theory of preprocessing.", Technical Report on Automata, A 68-63 (March. 1969).
[5] Iijima, T. and Y. Noguchi: "Analysis of pattern." Joint Conv. Record of Electrical & Electronics Engrs. of Japan, 49 (1966).
[6] Watanabe, S.: "A method of self-featuring information compression." "Proc. Bionics. Symp. (May. 1966).
[7] Tou, J. T. Ed.: Computer and Information Sciences-II, pp. 57-89, Academic Press (1967).
[8] Sebestyen, G. S.: Decision-Making Processes in Pattern Recognition, Macmillan (1962).
[9] Watanabe, S.: Methodologies of Pattern Recognition, Academic Press (1969).
[10] Funakubo, Iwamatsu, Suzuki, Mori and Iijima, "ETL-OCR pilot model.", Jour. I.E.C.E., Japan, 52-C, 11, pp. 712-719 (Nov. 1969).
[11] Yoshimura, M. and T. Iijima: "A method for the disign of the font", Information Processing in Japan, 10 (1970).