

## 論文

## 図形認識における特徴抽出の一方法\*

周 藤 安 造\*\*

## Abstract

In general, feature extraction in the picture recognition requires formidable computation. So, we developed a feature extraction method which incorporates local data operation and relatively simple algorithm. This method, named Geometric Factor Method, utilizes the geometry computation theory in Perceptrons.

The pictures are analyzed and feature values in the picture are extracted according to the geometric factors as defined in the method. In a simulation based on the method, feature values of randomly generated circles, triangles and rectangles are extracted and evaluated. Also, the method is tested on classification of human white cells for a medical picture processing application.

## 1. ま え が き

図形認識は最近さかんに研究されている分野で、普通の過程をとる。

- ① 図形入力
- ② 前処理
- ③ 特徴抽出
- ④ 図形判別

このうち、特徴抽出は図形認識において最も重要な部分であるにもかかわらず、普遍的な理論をうちたてることがむずかしく、これまでに非常に多くの方法が試みられてきた<sup>1)~4)</sup>。しかし、一般に図形認識における特徴抽出は、かなり高度な判断を必要とするため、実際の図形認識システムにおいて面倒な計算処理を加えねばならない。

このような点を考慮して、本研究では局所的なデータ操作と、比較的簡単なアルゴリズムにもとづく図形認識の一特徴抽出法を考案し、これを仮りに図形因子法 (Geometric Factor Method) と名づけた。図形因子法は、パーセプトロンの図形計算理論<sup>5)</sup>を応用して

図形因子 (geometric factor) という小正方形または小立方体の局所的な組合せパターンを定義し、これによって図形を分析して特徴抽出する方法であり、その機能はもとの図形の方向変換と位置変換に対して不変である。かかる特徴抽出法に関する理論を2値図形認識について確立し、その有効性をシミュレーション実験によって検証するが、濃淡図形認識への適用についても考察する。

また、実際の図形認識システムへの本法の応用として、医用画像の特徴抽出とその判別を試みる。

## 2. アルゴリズム

図形Sを2次元ユークリッド空間 (メッシュ空間) 内の2値情報からなる点の集まりとして表わし、簡単のために、Sを2つのパターンに判別するための特徴抽出を考える。そのために、空間要素  $(i, j)$  を2値図形関数

$$f(i, j) = 1 \quad (\text{黒メッシュ点})$$

$$f(i, j) = 0 \quad (\text{白メッシュ点})$$

で与え、図形空間Uと2値図形Sを次のように定義する。

$$U = \{(i, j) \mid f(i, j) = 1 \vee f(i, j) = 0\},$$

$$S = \{(i, j) \mid f(i, j) = 1\}.$$

図形因子法は、このような図形空間Uを次で定義する図形因子集合によって分析して、Uに含まれる図形

\* A Method for Feature Extraction in Picture Recognition by Yasuzou Sutou (Tokyo Shibaura Electric Co., Ltd, Computer Applications & System Engineering Sect., System Engineering Department, Electronic Computer Division).

\*\* 東京芝浦電気(株)電子計算機事業部・システム技術部応用ソフトウェア開発課

因子の個数を図形Sのもつ特徴値として抽出する方法である<sup>6)</sup>。

2.1 図形因子

図形因子は  $f(i, j)=1$  または  $f(i, j)=0$  となる空間要素  $(i, j)$  を系統的に組合せて生成した局所的な図形分析のための因子であり、図形情報を扱いやすい代数的数値に変換する機能がある。今後、図形因子を構成する要素  $(i, j)$  の個数をデータの局所性を示す目安として扱い、これを位数 (order)<sup>5)</sup> で表わし、図形因子を次の記号によって定義する。

図形因子  $\{P_{i,j}^{\nu\lambda}\}$ 。

ここで、

$\nu$ : 図形因子の位数

$\lambda$ : 図形因子番号

$i, j$ : 図形因子の代表座標点

とする。

このような図形因子は付録1に示す方法で生成するが、その一部を空間要素と2値図形関数で表わすと、次のようになる。

Geometric factors

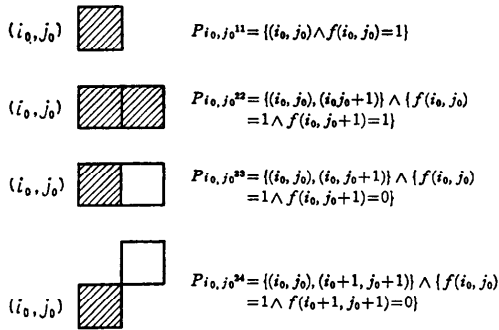


Fig. 1 Examples of geometric factor (square patterns).

なお、本法は適用する図形因子の個数ができるだけ少なくなかつそれが局所的 (位数が低い) であるほど、実際の図形認識において有効となるが、これについては以降でのべる。

2.2 特徴抽出

まず、図形因子によって図形Sから特徴値を抽出するための関数として  $\varphi$  を定義する。次に、その特徴値を図形Sの位置変換および方向変換に対して不変に抽出するために、図形因子の同値類を定義する。

これらの定義にもとづいて、図形因子法による特徴抽出のアルゴリズムを明らかにする。

2.2.1 関数  $\varphi$

図形Sの存在する空間Uから特定な図形因子  $P_{i,j}^{\nu\lambda}$  を抽出するための関数として

$$\varphi(P_{i,j}^{\nu\lambda}) \tag{1}$$

を与える。ここで  $(P_{i,j}^{\nu\lambda})$  の機能は

$$\varphi(P_{i,j}^{\nu\lambda})=1 \quad (P_{i,j}^{\nu\lambda} \text{ が } U \text{ に含まれている})$$

$$\varphi(P_{i,j}^{\nu\lambda})=0 \quad (P_{i,j}^{\nu\lambda} \text{ が } U \text{ に含まれていない})$$

である。このような関数  $\varphi(P_{i,j}^{\nu\lambda})$  によって、図形空間Uを分析し、座標点  $(i, j)$  における特定な図形因子  $P_{i,j}^{\nu\lambda}$  を抽出する。

2.2.2 図形因子の同値類

2つの図形  $A_x, A_y$  および図形の変換群  $G$  が与えられ、もし

$$A_x = g A_y \tag{2}$$

となるような  $G$  の要素  $g$  が存在するなら図形  $A_x$  と  $A_y$  は  $G$ -同値であるといひ

$$A_x \equiv_G A_y \tag{3}$$

と書く<sup>5)</sup>。

本法では、 $G$ -同値の概念を図形因子に適用し、変換群  $G$  として方向変換群と位置変換群を考慮する。

1) 方向変換のもとで同値な図形因子

Fig. 2 において、 $g_{\theta} \in G$  を与え  $P_{i_0, j_0}^{\nu\lambda}$  と  $P_{i_0, j_0}^{\nu\lambda(\sigma)}$  ( $\sigma=1 \sim n(\sigma)$ ) が

$$P_{i_0, j_0}^{\nu\lambda(\sigma)} = g_{\theta} P_{i_0, j_0}^{\nu\lambda} \tag{4}$$

を満足するなら、 $P_{i_0, j_0}^{\nu\lambda}$  と  $P_{i_0, j_0}^{\nu\lambda(\sigma)}$  は方向変換のもとで同値 ( $G$ -同値) となり

$$P_{i_0, j_0}^{\nu\lambda} \equiv_G P_{i_0, j_0}^{\nu\lambda(\sigma)} \tag{5}$$

が成りたつ。

2) 位置変換のもとで同値な図形因子

Fig. 3 (次頁参照) において、 $g(m_{\tau}, n_{\tau}) \in G$  を与え  $P_{i_0, j_0}^{\nu\lambda}$  と  $P_{i_0, j_0}^{\nu\lambda(\tau)}$  ( $\tau=1 \sim n(\tau)$ ) が

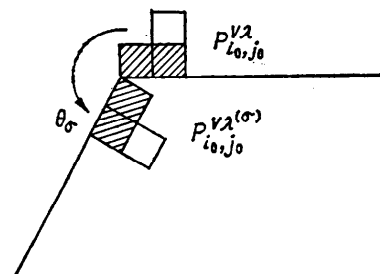


Fig. 2 Equivalent geometric factors under rotation.

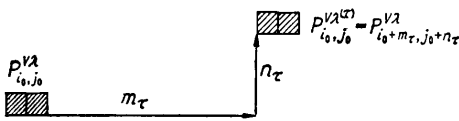


Fig. 3 Equivalent geometric factors under translation.

$$P_{i_0, j_0}^{V\lambda^{(n)}} = g(m_\tau, n_\tau) P_{i_0, j_0}^{V\lambda} = P_{i_0+m_\tau, j_0+n_\tau}^{V\lambda} \quad (6)$$

を満足するなら、 $P_{i_0, j_0}^{V\lambda}$  と  $P_{i_0, j_0}^{V\lambda^{(n)}}$  は位置交換のもとで同値 (G-同値) となり

$$P_{i_0, j_0}^{V\lambda} \equiv_G P_{i_0, j_0}^{V\lambda^{(n)}} \quad (7)$$

が成りたつ。

2.2.3 図形 S の特徴値

関数  $\varphi$  と図形因子の同値類の定義により、図形 S の存在する空間 U に含まれる特定な図形因子  $P_{i, j}^{V\lambda}$  の個数を

$$X_\lambda = \sum_{g(m_\tau, n_\tau) \in G} \sum_{g_\theta \in G} \varphi(g_\theta \cdot P_{i_0+m_\tau, j_0+n_\tau}^{V\lambda}) \quad (8)$$

として求める。

ここで、 $\sum_{g_\theta \in G}$  によって特定な座標点  $(i_0+m_\tau, j_0+n_\tau)$  において  $P_{i_0+m_\tau, j_0+n_\tau}^{V\lambda}$  を方向変換して同値となる図形因子  $g_\theta \cdot P_{i_0+m_\tau, j_0+n_\tau}^{V\lambda}$  の個数が求まり、さらに  $\sum_{g(m_\tau, n_\tau) \in G}$  によって各  $g_\theta \cdot P_{i_0+m_\tau, j_0+n_\tau}^{V\lambda}$  を全空間 U に位置変換して同値となる図形因子の個数が求まる。なお、今後(8)式を簡単に

$$X_\lambda = \zeta_\lambda(P_{i, j}^{V\lambda}) \quad (9)$$

で表わし、 $X_\lambda$  を図形因子  $P_{i, j}^{V\lambda}$  によって求まる図形 S におけるひとつの特徴値または特徴パラメータとする。

ところで、特定な 2 値図形 S に対し任意の n 個の図形因子を適用すれば、特徴値は  $X_1, X_2, \dots, X_\lambda, \dots, X_n$  として求まる。このような n 次元ベクトルとしての特徴値 (特徴値ベクトル) を求めるのが本法の目的であり、これを図で示すと Fig. 4 のようになる。

2.2.4 特徴値の選択

一般に図形認識において、特徴抽出は面倒とされている。そこで、本法においてはできるだけ少ない図形因子の適用で有効な特徴値を抽出し、処理を効率化することを試みる。そのために、n 個の特徴値から最も図形判別能のよい  $n^0$  個 ( $n^0 \leq n$ ) を選び、実際の図形認識に適用する。

しかし、その際 ( $\binom{n}{n^0}$ ) の全組合せを調べて評価するのは大変であり、普通その方法として

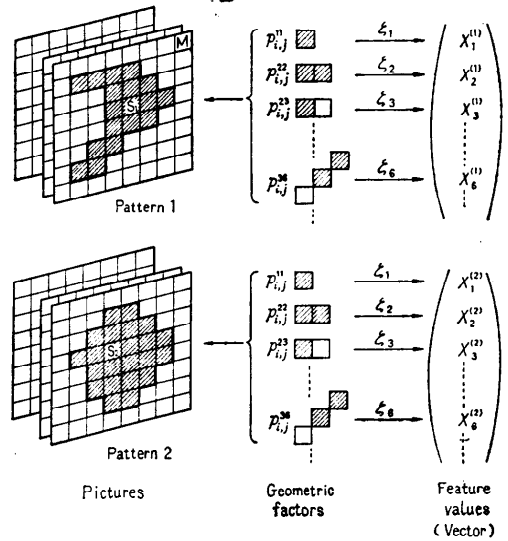


Fig. 4 Feature extraction using geometric factors.

- (1) 統計的方法
- (2) 情報論的方法
- (3) 逐次的方法

が考えられている<sup>7)</sup>。

このうち統計的方法のひとつによって、特定な図形  $S_1, S_2$  を判別する上で有効な (判別能のよい)  $n^0$  個の特徴値すなわち図形因子のパターンを(9)式の  $X_\lambda$  の値にもとづいて評価する。その際、各  $X_\lambda$  の値はサンプル・データが多ければ 2 つの図形  $S_1, S_2$  について、それぞれ平均値  $\bar{x}_{1\lambda}, \bar{x}_{2\lambda}$ 、標準偏差  $s_{1\lambda}, s_{2\lambda}$  の正規分布をすと仮定し、Fig. 5 の斜線領域  $e(S_1, S_2)$  の判別不能領域) を求め、できるだけ  $e$  を小さくする  $X_\lambda$  ほど、つまり平均値の差  $|\bar{x}_{1\lambda} - \bar{x}_{2\lambda}|$  の値を大きくし、かつ各標準偏差を小さくする  $X_\lambda$  ほど、 $S_1, S_2$  を判別する上で有効な特徴値すなわち図形因子のパターンであるとする。

特徴値  $X_\lambda$  の有効度は

$$|\bar{x}_{1\lambda} - \bar{x}_{2\lambda}| / (s_{1\lambda} + s_{2\lambda}) \quad (10)$$

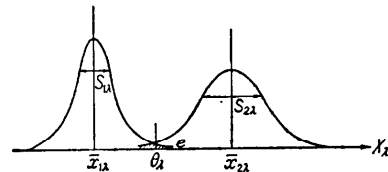


Fig. 5 Mean and standard deviation of  $X_\lambda$  in pictures  $S_1, S_2$ .

で求めるが、詳しくは付録2 (p.409)を参照されたい。

このようにして、すべての特徴値の有効度をランクづけしたのち、3.でのべるシミュレーションにおいてこれを順次、図形判別に適用していき、認識率が一定の水準に達したところで  $n^0$  を決定する。これによって、(11)式における  $n$  次元特徴値ベクトル  $X$  が  $n^0$  次元特徴値ベクトル  $X^0$  に変換され、実際に Fig. 4 で示す特徴抽出を効率的に行なうことができる。

$$X = \begin{pmatrix} \zeta_1(P_{i,j,k}^{r_1}) \\ \zeta_2(P_{i,j,k}^{r_2}) \\ \vdots \\ \zeta_n(P_{i,j,k}^{r_n}) \end{pmatrix}, \quad X^0 = \begin{pmatrix} \zeta_1(P_{i,j,k}^{r_1}) \\ \zeta_2(P_{i,j,k}^{r_2}) \\ \vdots \\ \zeta_n(P_{i,j,k}^{r_n}) \end{pmatrix} \quad (11)$$

2.3 濃淡図形の特徴抽出

本法による特徴抽出の機能を濃淡図形認識にまで一般化することを考える。

そのために、濃淡図形  $T$  を Fig. 6 で示す3次元空間、すなわちメッシュ平面を濃度レベルでスタックした空間  $V$  (濃度レベルが32であれば、32枚のメッシュ平面からなる) において表現し、 $V$  を構成する各要素  $(i, j, k)$  を小立方体 (cube) として扱い、これを2値図形関数

$$g(i, j, k) = 1 \\ g(i, j, k) = 0$$

で表わし、図形空間  $V$  と濃淡図形  $T$  を次のように小立方体の集合として定義する。

$$V = \{(i, j, k) | g(i, j, k) = 1 \vee g(i, j, k) = 0\}, \\ T = \{(i, j, k) | g(i, j, k) = 1\}.$$

次に、濃淡図形  $T$  から特徴値を抽出するための図形因子を  $g(i, j, k) = 1, g(i, j, k) = 0$  となる空間要素  $(i, j, k)$  の組合せパターン (小立方体の組合せパターン)

$$\{P_{i,j,k}^{r_n}\}$$

で定義し、その一部を Fig. 7 に示す。

このような  $P_{i,j,k}^{r_n}$  にもとづいて、関数  $\phi$  を与えれ

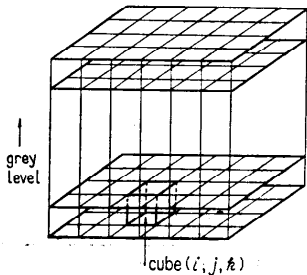


Fig. 6 3-dimensional picture space  $V$  and space element (cube).

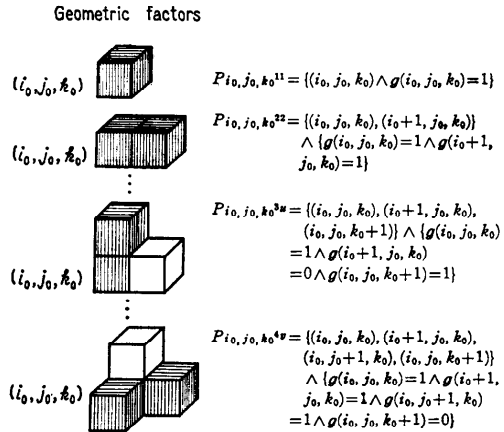


Fig. 7 Examples of geometric factor (cubic patterns).

ば

$$\phi(P_{i,j,k}^{r_n}) = 1 \quad (P_{i,j,k}^{r_n} \text{ が } V \text{ に含まれている}) \\ \phi(P_{i,j,k}^{r_n}) = 0 \quad (P_{i,j,k}^{r_n} \text{ が } V \text{ に含まれていない})$$

によって濃淡図形  $T$  の特徴値は(8)式に準じ

$$X_\lambda = \sum_{g(l,m,n) \in G} \sum_{g_{r_n} \in G} \phi(g_{r_n} P_{i_0+i, j_0+m, k_0+n}^{r_n}) \quad (12)$$

として求めることができる。

ここで、 $\sum_{g_{r_n} \in G}$  によって方向変換のもとで同値な図形因子の個数、 $\sum_{g(l,m,n) \in G}$  によって位置変換のもとで同値な図形因子の個数が求まる。

なお、 $g_{r_n}$  は  $g_{r_n}$  が平面的な方向変換に対する  $G$  の要素であるのに対して、3次元空間  $V$  における立体的な要素である。

2.4 図形判別

2.2.4 において個々の特徴値の図形判別能について評価したが、ここでは選択した特徴値を実際の図形判別に適用するための方法を示す。ところで、一般的な図形判別法としてはベイズ法、ミニマックス法、ネイマン-ピアソン法、線形判別関数法および非線形判別関数法などがあるが<sup>8)-10)</sup>、ここでは  $n$  次元ベクトル (実際には  $n^0$  次元に変換して判別処理するが、一般化するため  $n$  とする) として求まる特徴値を ( $n$  次元) ユークリッド空間に対応させ、この空間を超平面で分割する線形判別関数の設定を試みる。

そのためには、次の関数

$$f(X) = W_1 X_1 + W_2 X_2 + \dots + W_n X_n + W_{n+1} \quad (13)$$

において、特定な  $n$  個の特徴値  $X_1, X_1, \dots, X_n$  を与え、

$f(X)$  を計算し、 $f(X) > 0$  なら図形  $S_1$ ,  $f(X) \leq 0$  なら図形  $S_2$  となる確率が最も高くなるように重み係数  $W_1, W_2, \dots, W_n, W_{n+1}$  を決めてやればよい。本研究では、(13) 式の設定に  $n$  個の変数すなわち特徴値  $X_n$  の 2 グループ判別分析理論 (多変量解析法) を適用する。

なお、(13) 式による  $n$  次元特徴空間の超平面分割は Fig. 8 のようになる。

### 3. シミュレーション\*

2 値図形認識に対する本法の有効性を検証するために、円、三角形および矩形の各図形データを与え、これらの特徴抽出とその特徴値を評価するための図形判別シミュレーションを行なう。なお、シミュレーション結果に一般性をもたせるために、シミュレーションデータの作成において各図形の形状と大きさに統計的なバラツキを与えている。

#### 3.1 シミュレーションの手順

シミュレーションの手順を Fig. 9 に示す。シミュレーションは 2 つのフェーズからなり、フェーズ 1 で図形データとパターン化した図形因子を TOSBAC-

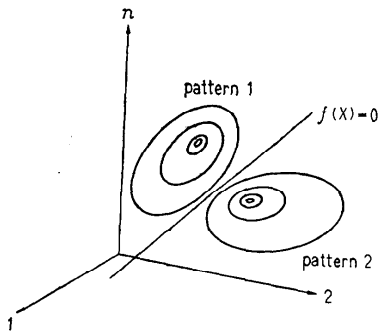


Fig. 8 Feature space and decomposition of hyperplane by linear discriminant function.

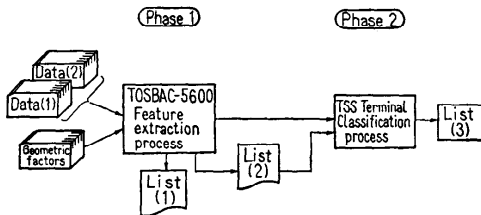


Fig. 9 Simulation processes.

\* 本シミュレーションはトレーニング・データのみによるテストである。

5600 に入力し、特徴抽出する。そのために、まずパターン化した図形因子で図形データを分析し、特徴値を  $n$  個のパラメータとしてリスト (1) に出力する。次に、各パラメータ  $X_n$  についてその有効度を (10) 式によって評価し、それをランクづけしてリスト (2) に出力する。

フェーズ 2 は、TOSBAC-5600 TSS を利用した会話型による図形判別シミュレーションで、フェーズ 1 で求めたリスト (2) にもとづいて有効度の高い  $X_n$  から順にひとつずつ追加しながら 2 グループ判別処理を行なう。その際、認識率が一定の限界に達したところでシミュレーションを終り、結果をリスト (3) に出力する。以上のシミュレーションを

- (1) 円と三角形
- (2) 円と矩形
- (3) 三角形と矩形

の各図形データについて行なう。

#### 3.2 データのつくり方

円、三角形、矩形の各図形データを Fig. 10 のように横軸  $x$ 、縦軸  $y$ 、を与えて、それぞれ 50 件ずつ作成する。

ここで  $x, y$  は、いずれも平均値 10 (メッシュ点の数)、標準偏差 2 の正規分布にもとづく乱数発生ルーチンによって生成した数値である。

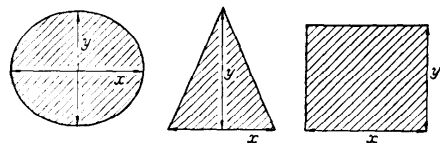
各図形の作成においては、このような  $x, y$  をそれぞれ 50 件ずつ生成順に組合せることによって、形状と大きさに十分なバラツキを与えているが、その結果は Table 1 (次頁参照) のようになった。なお、 $x$  軸、 $y$  軸に対応する各数値は、各図形データの件数を示している。

#### 3.3 図形因子の設定

位数 3 までの図形因子を付録 1 (p. 409 参照) に示す方法によって設定した。フェーズ 1 で適用した図形因子は 43 個で、そのうちフェーズ 2 で図形判別に適用したのは有効度の高い 5 個であった。

#### 3.4 シミュレーション結果

フェーズ 1、フェーズ 2 のシミュレーション結果を次に示す (次頁参照)。



circle triangle rectangle

Fig. 10 Simulation data patterns.

Table 1 Distribution of smulation data

		x 軸										
		6	7	8	9	10	11	12	13	14		
y 軸	6			1	2							3
	7				2	1						3
	8		2	1		2	3					8
	9		2	2	1	3	2					10
	10	1		2	3	2	3	1				12
	11		1	2		1	1					5
	12		2	3					1			6
	13		1			1						2
	14				1							1
			1	8	11	9	10	9	1	1	0	50

3.4.1 フェーズ I の結果

各図形群データを判別する上で有効な特徴値  $X_k$  が抽出できる図形因子パターンと各特徴値の平均値、標準偏差等は、それぞれ Table 2, Table 3 および Table 4 に示すとおりであった。なお、ここで有効度 (effective degree) は (10) 式にもとづいて

Table 2 Effective geometric factors for classification of circles and rectangles.

図形因子	位数	平均値		標準偏差		有効度
		円 $\bar{x}_{1k}$	矩形 $\bar{x}_{2k}$	円 $S_{1k}$	矩形 $S_{2k}$	
1	3	14	4	1.70	0	564
2	3	10	0	1.70	0	564
3	3	18	37	2.84	4.57	258
4	3	18	37	2.84	4.57	258
5	3	10	29	2.84	4.57	258
6	3	5	4	0.97	0	152
7	3	1	0	0.97	0	152
8	2	50	70	6.04	9.14	136
9	3	46	66	6.04	9.14	136
10	3	50	70	6.16	9.14	133

Table 3 Effective geometric factors for classification of circles and triangles.

図形因子	位数	平均値		標準偏差		有効度
		円 $\bar{x}_{1k}$	三角 $\bar{x}_{2k}$	円 $S_{1k}$	三角 $S_{2k}$	
1	3	0	8	0	1.66	465
2	3	4	12	0	1.66	465
3	3	0	3	0	0.94	322
4	3	4	10	0	2.00	302
5	3	0	3	0	1.09	276
6	3	4	10	0	2.20	253
7	3	0	5	0	2.22	243
8	3	29	15	4.57	3.26	176
9	3	37	24	4.57	4.08	147
10	3	269	111	74.02	38.10	141

Table 4 Effective geometric factors for classification of triangles and rectangles.

図形因子	位数	平均値		標準偏差		有効度
		三角 $\bar{x}_{1k}$	矩形 $\bar{x}_{2k}$	三角 $S_{1k}$	矩形 $S_{2k}$	
1	3	5	0	2.22	0	243
2	3	10	5	2.00	0.99	159
3	3	10	5	2.20	0.97	129
4	3	96	181	35.97	51.07	98
5	3	82	164	33.72	49.43	98
6	3	98	183	36.40	51.41	96
7	3	99	182	36.59	51.23	95
8	3	118	205	39.40	54.12	93
9	2	125	214	40.51	55.35	93
10	3	24	18	4.08	2.84	93

$$\{|\bar{x}_{1k} - \bar{x}_{2k}| / (s_{1k} + s_{2k})\} \times 100 \quad (14)$$

で与えている。

3.4.2 フェーズの結果

いずれの判別シミュレーションにおいても、最も有効度の高い特徴値だけで完全に2グループ100件の図形データが判別できた。ちなみに上位5個までの特徴値を順次、追加して判別処理したが、いずれのシミュレーションについても判別ミスは1件も生じなかった。これによって本法による特徴抽出の有効性を一般的に検証できた。

3.5 医用画像処用への応用

本法の医用画像処理への応用として、核型による白血球の分類を試みた。すなわち、白血球の核組織を2値図形として扱い、本法によって図形分析し、白血球を判別する上で有効な特徴値を抽出した。

各白血球の核図形データの一部を Fig. 11 に示すが、今回のシミュレーションでは23件のリンパ球と23件の好中球、42件のリンパ球と31の大単核白血球を判別するための特徴抽出を行なった。

シミュレーションは、まずフェーズ1において位数3までの43個の図形因子によって核図形を分析して特徴抽出を行ない、その中から判別上、特に有効な特徴値が抽出できる図形因子10個を選んだ。次に、フェーズ2ではこれらの10個の特徴値を有効度の高いものから順にひとつずつ追加しながら、判別処理を行ない認識率の変化を求めた。

その結果、フェーズ1においてリンパ球と好中球を判別するのに有効な図形因子パターンとその特徴値(平均値、標準偏差)は Table 5, 同じくリンパ球と大単核白血球については Table 6 に示すとおりであっ

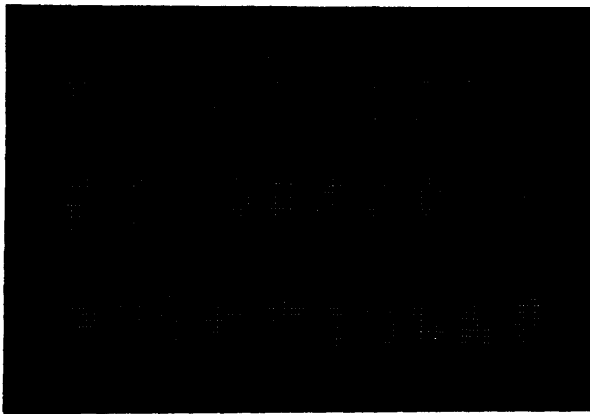


Fig. 11 Samples of white cell nuclear picture.

Table 5 Effective geometric factors for classification of lymphocytes and neutroplils.

図形因子	位数	平均個数		標準偏差		有効度
		リンパ $\bar{x}_{1k}$	好中 $\bar{x}_{2k}$	リンパ $S_{1k}$	好中 $S_{2k}$	
1	3	5	14	0.85	4.18	176
2	3	0	5	0.42	2.27	174
3	3	73	25	17.41	9.92	173
4	3	85	34	18.42	13.33	162
5	3	86	36	19.08	12.27	162
6	3	86	36	18.86	12.80	159
7	3	15	26	2.23	4.95	156
8	3	0	7	0.58	3.95	154
9	3	22	31	2.15	4.06	148
10	2	30	42	2.83	5.59	146

Table 6 Effective geometric factors for classification of lymphocytes and monocytes.

図形因子	位数	平均個数		標準偏差		有効度
		リンパ $\bar{x}_{1k}$	単球 $\bar{x}_{2k}$	リンパ $S_{1k}$	単球 $S_{2k}$	
1	3	23	31	2.78	3.93	113
2	2	29	38	3.16	5.37	111
3	3	24	31	2.86	3.65	107
4	3	27	34	2.78	4.30	104
5	3	24	32	3.21	4.71	101
6	3	38	47	4.06	5.59	98
7	3	29	37	3.16	5.04	98
8	3	18	25	2.29	4.25	96
9	2	41	53	4.57	8.32	95
10	3	35	46	4.16	6.78	95

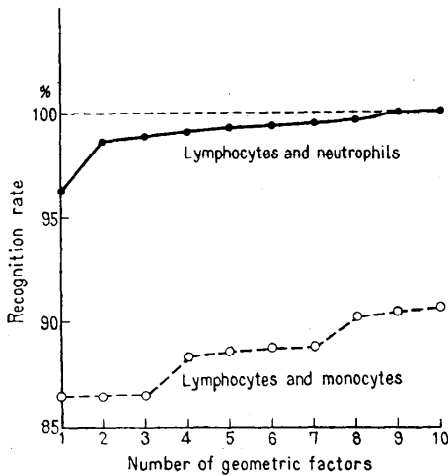


Fig. 12 Picture recognition rate in white cells classification.

た。また、Fig. 12 はフェーズ2の結果で、この図において判別に適用した図形因子の個数とそれに対応する認識率の関係をリンパ球と好中球の判別については実線で、リンパ球と大単核白血球の判別については点線でそれぞれ示す。

これによって、今回の白血球の分類においては、核図形に含まれる特徴値の有効度を評価すれば、それに対応したせいぜい10個の図形因子の適用でかなり高い認識率を得ることができるといえる。なお、本シミュレーションの結果は、大型電子計算機 TOSBAC-5600 モデル 140 を使って求めたが、どのシミュレーションにおいても、フェーズ1、フェーズ2に要したピロセス・タイムは十数秒で、処理を十分効率化することができた。

#### 4. 考 察

本法による特徴抽出は図形因子を定義し、それに位数の概念を導入した点に特長があるが、実際の図形認識においては、位数の設定が重要となる。すなわち、特徴抽出を効率的に行なうためには、データ処理を局所的に行なうこと、つまりどの程度図形因子の位数が低く設定できるかということがひとつの大きな課題となる。この点を明確にするために図形因子の機能と位数の関係を2値図形認識について分析することを考え、

- ① 大きさに関する特徴値を抽出する図形因子群を  $Q_1$
- ② 周囲長に関する特徴値を抽出する図形因子群を

Table 7 Order and feature extraction function of geometric factor groups.

図形因子群	図形因子	最高位数	機能
$Q_1$		1	図形の大きさが求まる
$Q_2$		2	図形の周囲長が求まる
$Q_3$		3	輪郭部の形状が求まる
$Q_4$		不定 (図形の大きさによって異なる)	内部の形状が求まる

$Q_2$

- ③ 輪郭情報に依存した形状に関する特徴値を抽出する図形因子群を  $Q_3$
- ④ 内部情報に依存した形状に関する特徴値を抽出する図形因子群を  $Q_4$

とし、Table 7 に示すように各群の最高位数を評価した。

この結果から、3位までの図形因子によって図形の大きさ、周囲長および輪郭情報に依存した形状に関する特徴値が抽出できることがわかる。これに対して  $Q_4$  の位数は一概に決めることはできない。

結局、図形因子の位数は、図形の大きさにあまり依存しない（その場合、最高位数は3）と思えるが、さらにシミュレーションを重ね明確にする考えである。

#### 5. あとがき

図形の特徴値が局所的なデータ操作と比較的簡単な演算によって抽出できることを明らかにした。なお、今後濃淡図形の特徴抽出に関するシミュレーションを行ない、その有効性と実用性について検証していく考えである。最後に、本研究において、いろいろご指導を受けた慶応義塾大学教授林喜男先生、東京大学付属病院関原成允先生に深く感謝申し上げる。

#### 参 考 文 献

- 1) Martin D. Levine: Feature Extraction; A Survey, Proc. IEEE, 57, 8, 1391~1402(1969).
- 2) Ernest L. Hall et al: A Survey of Preprocessing and Feature Extraction Techniques for Radiographic Images, IEEE, Trans, C-20, 9, 1032~1044 (1971).
- 3) 坂井利之, 長尾真: 文字・図形の認識機械 (情報科学講座 E. 19. 2), 共立出版 (1967).
- 4) 志村正道: パターン認識と学習機械, 昭晃堂 (1972).



- 5) M. Minsky, S. Papert: Perceptrons—An introduction to computation geometry, MIT press, London (1969).
- 6) 開原成允, 周藤安造, 柴崎公子: パーセプトロンを適用した画像処理法による ME 画像解析, 第13 回日本 ME 学会大会資料集, 182(1974).
- 7) 山本真司: パターン認識における特徴の抽出と選択—IEEE Trans., C-20 特集号報告, 電子通信学会誌, 55, 12, 1637~1640 (1972).
- 8) 野口正一: パターン認識と処理の理論, 電気学会誌, 93, 11, 1007~1012 (1973).
- 9) 上坂吉則: パターン認識と学習の理論, 総合図書 (1971).
- 10) J. N. Nilsson: Learning machines, Mc-Graw-Hill Book Company, New York (1965).

付 録

1. 図形因子の生成

図形因子の生成過程は次のとおりである。

- ① Fig. 13 に示すメッシュ座標から 1~k 位までの図形因子を生成する。
  - ② 方向変換, 位置変換のもとで同値な図形因子を削除して黒メッシュ点だけからなる素な図形因子に変換する。なお, 方向変換に関しては, 90°, 180°, 270° に限定した。
  - ③ 同じ方法で, 白黒メッシュ点を系統的に組合せた図形因子を生成する。
- ②と③が求める図形因子である。

2. 特徴値  $X_\lambda$  の有効度評価

特徴値  $X_\lambda$  の有効度を次の方法で評価する。

各  $X_\lambda$  は Fig. 5 で示すような正規分布をすとし, さらに Fig. 5 における  $\theta_\lambda$  が斜線領域を等分する点 (threshold) であるとすれば, 次式が成りたつ。

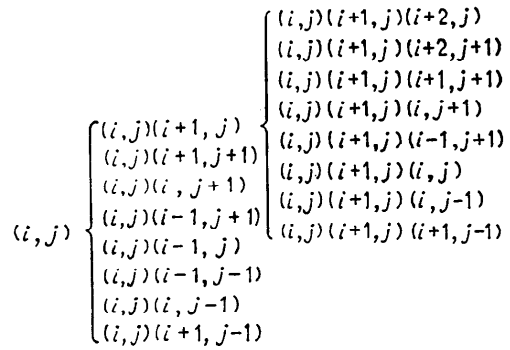
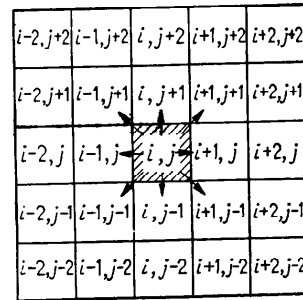
$$1/(\sqrt{2\pi}s_{1\lambda}) \int_{\theta_\lambda}^{\infty} \exp\{-(x_\lambda - \bar{x}_{1\lambda})^2/2s_{1\lambda}^2\} dx_\lambda$$

$$= 1/(\sqrt{2\pi}s_{2\lambda}) \int_{-\infty}^{\theta_\lambda} \exp\{-(x_\lambda - \bar{x}_{2\lambda})^2/2s_{2\lambda}^2\} dx_\lambda$$

(付 1)

(付 1) を変換して

$$1/\sqrt{2\pi} \int_{(\theta_\lambda - \bar{x}_{1\lambda})/s_{1\lambda}}^{\infty} \exp(-t^2/2) dt$$



Order 1      Order 2      Order 3.  
**Fig. 13** Generation of geometric factors.

$$= 1/\sqrt{2\pi} \int_{-\infty}^{(\theta_\lambda - \bar{x}_{2\lambda})/s_{2\lambda}} \exp(-t^2/2) dt \quad (付 2)$$

を導き, (付 2) から

$$(\theta_\lambda - \bar{x}_{1\lambda})/s_{1\lambda} = -(\theta_\lambda - \bar{x}_{2\lambda})/s_{2\lambda} \quad (付 3)$$

ゆえに

$$\theta_\lambda = (s_{1\lambda}\bar{x}_{2\lambda} + s_{2\lambda}\bar{x}_{1\lambda}) / (s_{1\lambda} + s_{2\lambda}) \quad (付 4)$$

が求まる。ところで, (付 2) において, (付 3) の値を大きくする特徴値  $X_\lambda$  ほど Fig. 5 の斜線部の領域を小さくするといえる。これは, (付 4) を (付 3) に代入して求まる。

$$|\bar{x}_{1\lambda} - \bar{x}_{2\lambda}| / (s_{1\lambda} + s_{2\lambda}) \quad (付 5)$$

の値の大小に変換できるから, 結局, (付 5) の値で特徴値  $X_\lambda$  の有効度が決まる。

(昭和 49 年 8 月 24 日受付)