

白血球の自動分類

(特徴抽出 その2)

中村 雅文*・小川 秀夫*・谷口 慶治*

Automatic Classification of White Blood Cells
(Feature Extraction Part 2)

Masafumi Nakamura, Hideo Ogawa and Keiji Taniguchi

The present paper deals with the following problems for the automatic white blood cell classification.

(1) A technique for extracting a nucleus region of the white blood cell is described. It can be accomplished by using the histogram of the directional difference. From this histogram, an optimum threshold value for extracting the nucleus region is obtained.

(2) The texture parameters which express effectively the difference of the texture between white blood cells are proposed. That is, the shape of the chromatin in the nucleus and the amount of the granules are proposed to discriminate the difference between monocytes and large lymphocytes and the difference between eosinophils and basophils, respectively.

1. ま え が き

白血球の自動分類は近年その必要性がますます増大し、処理速度および識別精度においても極めて高度なものが要求されるに至っている。

筆者らは先にこの自動分類システムの構成にあたり、塗株標本中の白血球を5種類に分類するアルゴリズムを示した⁽¹⁾。ここでは、特徴パラメータとして容易に抽出できる濃度レベルおよび形状に関する量を提案し、その有効性を示した。しかし、文献(1)では白血球自動分類の前処理の問題について触れていなかった。ここでは、まずこの問題を解決する方法を述べている。つぎに、文献(1)で述べた特徴パラメータだけでは、単球と大リンパ球のように形状が非常に似かよっている白血球の場合、識別が困難であったが、これを容易にするため特徴パラメータとして“クロマチン量”を提案している。さらに、文献(1)では顆粒の検出法として、光電変換器の電圧レベルの変動を用いていたが、ここではデジタル処理による新しい方法を提案している。

* 電子工学科

2. 前 処 理

2-1 パターンの観測

対象とする白血球画像は、末梢血液をプレパラート上に塗抹し、メイギムザ染色を施したものである。この顕微鏡画像を、カラーテレビカメラにより、赤 (R)、青 (B)、緑 (G)、の3原色に対応する電気信号に変換し、それぞれをA-D変換することにより、デジタル画像、 $f_R(i, j)$ 、 $f_B(i, j)$ 、 $f_G(i, j)$ を得る。本稿では画像の特定の領域 (例えば核領域) を強調するために、3原色デジタル画像を適切な重み、 w_R, w_B, w_G 、により再合成した画像 $f(i, j)$ を用いる。すなわち

$$f(i, j) = w_R f_R(i, j) + w_B f_B(i, j) + w_G f_G(i, j) \quad (1)$$

である。この過程を図1に示す。

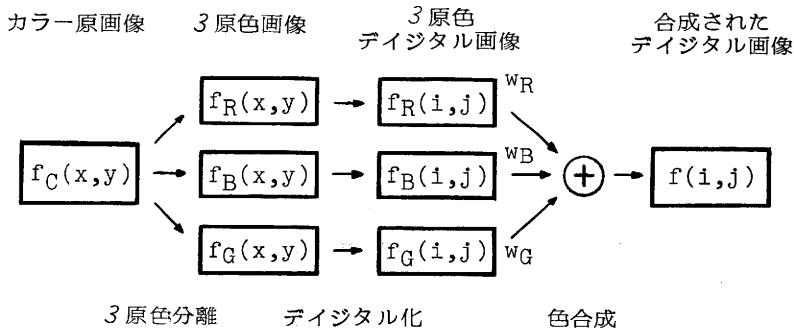


図1 白血球画像の観測過程

2-2 核領域の抽出 (2値化)

白血球を識別するための特徴として、核の形状、濃度、クロマチン網工などは極めて重要である。したがって、これらの特徴量を安定に抽出するためには、画質の悪い場合においても核領域が正しく抽出される必要がある。実際、白血球標本の染色むらおよび白血球の種類による濃度値のばらつきなどのため、核領域を自動的に正しく抽出することは必ずしも容易ではない。また従来この抽出

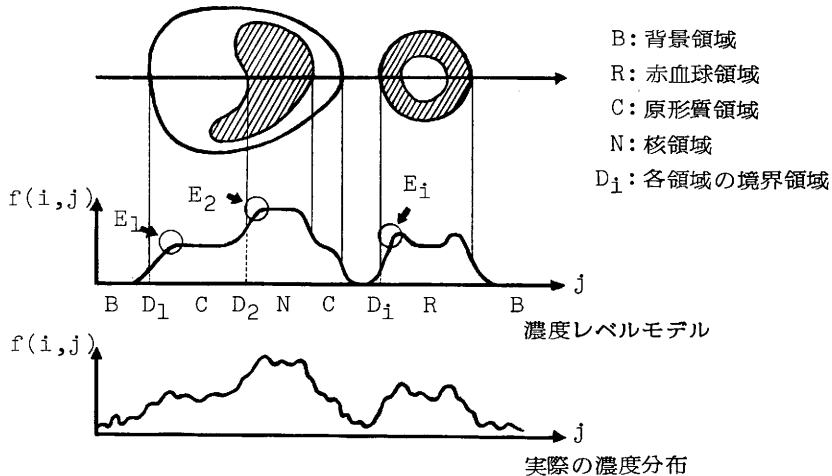


図2 白血球画像の濃度モデル

法（2値化法）として多くの手法⁽⁹⁾が提案されているが、それらは対象とする画像の性質に大きく依存するものである。そこで、ここではヒストグラム法に属する代表的な3つの手法を取り上げ、対象とする白血球画像の性質を基に各方式を比較し、最も有効な手法が差分ヒストグラム法によるものであることを示す。

白血球画像は、白血球の核、同原形質、赤血球、背景の各領域から構成され、その濃度レベルを図2の様にモデル化することができる。また実際の白血球画像を図3に示す。

(i) 濃度ヒストグラム法

一般に画像 $f(i, j)$ の濃度ヒストグラム $h_0(n)$ (2式) では、 $f(i, j)$ 上で同一濃度を持つ領域が存在する場合、それに対応して一つのクラスター（山）が形成される（図4(a)参照）。白血球画像では、核領域は他の領域に比べて濃度値が最も小さく（濃度レベルが高いほど数値は小さい）、したがって、 $h_0(n)$ 上で最も n の小さいクラスターが核領域に対応することになる。

$$h_0(n) = N\{(i, j) | f(i, j) = n\} \quad (2)$$

つぎに核領域に対応するクラスターを検出するアルゴリズムを示す。

[クラスター検出アルゴリズム]

$$\left. \begin{aligned} S_1(\ell | n) &= h_0(\ell) - h_0(n) \quad (\ell > n) \\ S_2(k | n) &= h_0(k) - h_0(n) \quad (k < n) \end{aligned} \right\} \quad (3)$$

とし、

$$\left. \begin{aligned} \bar{S}_1(n) &= \text{Max}_{\ell} \{S_1(\ell | n)\} \\ \bar{S}_2(n) &= \text{Max}_k \{S_2(k | n)\} \end{aligned} \right\} \quad (4)$$

とした場合

$$\varphi(n) = \bar{S}_1(n) \cdot \bar{S}_2(n)$$

をすべての n について求め、系列 $\{\varphi(n)\}$ を作る。

この $\varphi(n)$ より(5)式を満す n_T を求め、これをしきい値とする。

$$\varphi(n_T) = \text{Max}_n \left\{ \varphi(n) \mid T_D \leq \sum_{i=1}^n h_0(i) \leq T_U \right\} \quad (5)$$

ここで、 T_D, T_U は核面積から実験的に定められる定数である。

濃度ヒストグラム法では、全画素の濃度情報を用いるものであり、顕著な濃度差が存在する各領域を識別するためには簡単であり、有効な手法といえる。しかし、白血球画像では各領域の境界に連続的な濃度変化が見られるため、特に画質が悪い状態では $h_0(n)$ は一様分布に近い状態となり、一つのクラスターを検出することはほとんど困難となる。また、目的とする核領域が面積において小さい場合、 $h_0(n)$ では一つのクラスターとして識別することがさらに困難となり、これらの意味で実用的ではない。

(ii) ラプラシアンヒストグラム法

先の手法で、各領域に対応するクラスターの共通部分、すなわち $f(i, j)$ で各領域の境界領域 (D_i) における濃度は一種の雑音とみなすことができる。したがって、 $h_0(n)$ を求める過程において、この領域の濃度を無視することにより、クラスターを顕著なものとするができる。これを基に、画像 $f(i, j)$ のラプラシアン (7式) を求め、この値の大きい領域 (図2の E_i 部分) を対象として濃度ヒストグラムを求めるものである。

$$h_1(n) = N \{(i, j) | \Delta(i, j) \geq T_1, f(i, j) = n\} \quad (6)$$

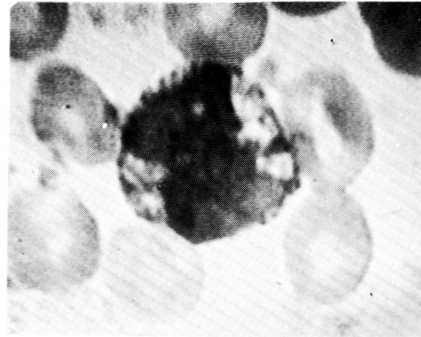
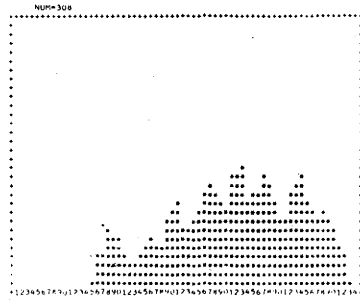


図3 白血球画像（好塩基球）



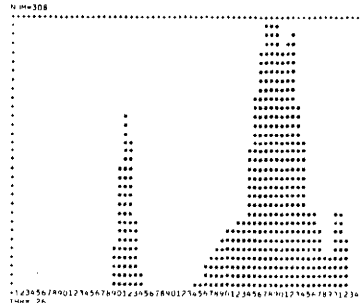
(a) 画像の濃度ヒストグラム



(b) 画像内でラプラシアン値の高い方20%の濃度ヒストグラム

$$\Delta(i, j) = |4 \cdot f(i, j) - \{f(i-1, j) + f(i+1, j) + f(i, j-1) + f(i, j+1)\}| \quad (7)$$

すなわち、ラプラシアンヒストグラム法は図2に示したモデルの様に各領域が平坦、あるいは単調な濃度変化を有する場合には、 E_1 領域のみの濃度を用いることになり、原理的に各領域の濃度に対応した極めて分散の小さいクラスターを形成することができる。また、各領域の境界付近のみの濃度情報を用いるため、領域の大きさ(面積)には大きく依存しない。したがって、小さな領域に対しても、対応するクラスターを検出することが容易となる。しかし、白血球画像では図2に実際の濃度分布として示



(c) 画像内で差分値の低い方40%の濃度ヒストグラム

したように、各領域あるいは境界領域の濃度は平坦あるいは単調な変化ではなく、2次の変化分が多く存在する。そのため、(6)式の条件を満たす画素は各領域に存在することになり、実際の $h_1(n)$ は図4(b)に示した様に原理的な予測とは大きくかけ離れたものとなり、この場合、有効な手法とは言えない。

図4 濃度ヒストグラム

iii) 差分ヒストグラム法

白血球の核を抽出するための2値化法として最も有効な差分ヒストグラム法を提案する。これは濃度ヒストグラム $h_2(n)$ を次の様に求めるものである。

Step 1 平均化：画像 $f(i, j)$ の高い周波数成分の雑音を除去するため、平均化を行なう。

$$\bar{f}(i, j) = \frac{1}{9} \sum_{\ell, m=-1, 0, 1} f(i+\ell, j+m) \quad (8)$$

Step 2 方向差分：画像 $\bar{f}(i, j)$ の平坦な領域と濃度変化を有する領域を区別するため、各画素の差分値 $\nabla(i, j)$ を、方向差分を基に次の様に求める。

$$\nabla(i, j) = \text{Max} \{ \nabla_V(i, j), \nabla_H(i, j), \nabla_L(i, j), \nabla_R(i, j) \} \quad (9)$$

$$\left. \begin{aligned} \text{但し } \nabla_V(i, j) &= \frac{1}{3} | \{ f(i-1, j-1) + f(i-1, j) + f(i-1, j+1) \} - \{ f(i+1, j-1) + f(i+1, j) + f(i+1, j+1) \} | \\ \nabla_H(i, j) &= \frac{1}{3} | \{ f(i-1, j-1) + f(i, j-1) + f(i+1, j-1) \} - \{ f(i-1, j+1) + f(i, j+1) + f(i+1, j+1) \} | \\ \nabla_L(i, j) &= \frac{1}{2} | \{ f(i-1, j) + f(i, j+1) \} - \{ f(i, j+1) + f(i+1, j) \} | \\ \nabla_R(i, j) &= \frac{1}{2} | \{ f(i-1, j) + f(i, j+1) \} - \{ f(i, j-1) + f(i+1, j) \} | \end{aligned} \right\} \quad (10)$$

Step 3 $h_2(n)$ を次のように求める。

$$h_2(n) = N \{ (i, j) \mid \nabla(i, j) \leq T_2, \bar{f}(i, j) = n \} \tag{11}$$

Step 4 この $h_2(n)$ より、先に示したクラスター検出アルゴリズムを用い、核に対応するクラスターを検出する。

この手法では、平均化により画像の濃度変化をなめらかなものとし、さらにその変化分(差分値)の小さい部分、すなわち各領域の内部に対応し、濃度の均一な部分のみに着目し、変化分の大きい部分、すなわち各領域の境界、あるいは領域内においても雑音などにより、大きな濃度変化が存在する部分を無視するものである。したがって、図2におけるモデルに近い状態で、かつ濃度が一定の部分のみを観測することになり、対応する $h_2(n)$ は極めて顕著なクラスターを構成することになる。また、核濃度に近い濃度を持つ赤血球は、図2のモデルに示されているように、中央部に濃度レベルの低い領域が存在する。したがって、赤血球全体として見た場合には、濃度が一定となる領域は比較的小さくなり、そのため、 $h_2(n)$ において赤血球領域に対応するクラスターが抑制されることになる。図4(c)に $h_2(n)$ を示す。同図からもこの手法の有効性は明らかである。

上記の手法を用いて、画像 $f(i, j)$ を2値化する。この一例を図5に示す。しかしこの2値パターンにはまだ目的とする核以外に雑音による孤立点、赤血球の濃度レベルの高い部分および他の白血球の核等が含まれている場合がある。そこでこの2値パターンに対して、まず孤立点を除去し、つぎに“1”の部分に対して輪郭追跡を行ない、グルーピングを行なう。さらにそれぞれのグループに対して、画面の中央に存在し、定められた範囲の大きさ(面積)をもつものを核領域と判定する。このようにして、核領域のみを有する2値パターンが得られる。

2-3 擬似原形質の指定

手引きの血液標本では、白血球のまわりに通常多くの赤血球が存在する。白血球の原形質の濃度レベルは種類により異なるが、一般には背景と赤血球の中間領域を占める。また原形質の色も白血球の種類により異なるので、色合成等により原形質を強調することも困難である。このため領域解析法等を用いて原形質を抽出する方法があるが、ここでは比較的簡単な方法で原形質を擬似的に定めた。まず白血球の原形質の多くが核を囲む類円形であることに着目し、抽出された核の周囲上の任意の2点間の距離を求め、この最大値を直径とする円を考える。つぎにこの円を外側に2画素分だけ広げ、この円内で核以外の領域を擬似原形質に割りあてる。

3. 特 徴 量

白血球の核及び原形質から得られる特徴量として、次のものがあげられる。

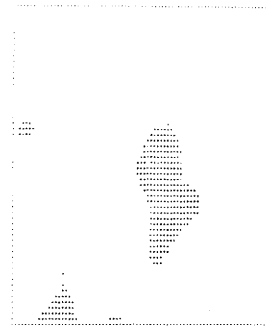


図5 差分ヒストグラム法による2値化パターン



図6 擬似原形質

C : 原形質領域
N : 核領域

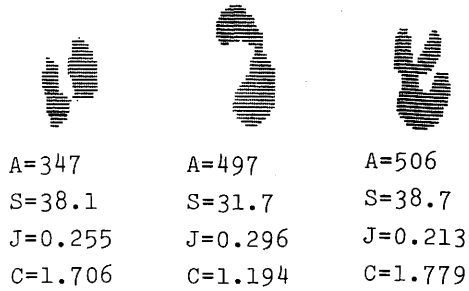
(1) 核面積：A，核領域を示す2値画像を

$P_N(i, j)$ とすると

$$A = \sum P_N(i, j) \quad (12)$$

但し

$$P_N(i, j) = \begin{cases} 1 : \text{核領域} \\ 0 : \text{その他の領域} \end{cases}$$



(a) 好中球

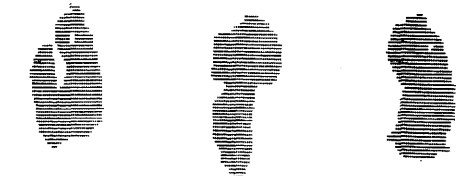
(2) 核周囲長：L，輪郭追跡法により，核周囲を構成する画素列が得られるが，この系列で，水平，垂直に連結する画素数をn，斜め方向に連結する画素数をmとした場合

$$L = n + \sqrt{2}m \quad (13)$$

(3) 核の形状：S

$$S = L^2 / A \quad (14)$$

Sの値は，一般に複雑な形状をもつ核ほど大きい。



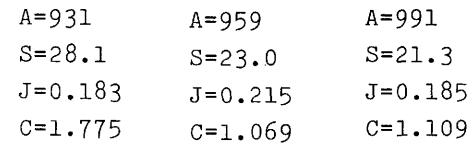
(b) 単球

(4) 核の平均濃度： D_N

$$D_N = \sum f(i, j) \cdot P_N(i, j) / \sum P_N(i, j) \quad (15)$$

但し

$f(i, j)$ ：原画像



(c) リンパ球

(5) 原形質の平均濃度： D_C

$$D_C = \sum f(i, j) \cdot P_C(i, j) / \sum P_C(i, j) \quad (16)$$

但し

$$P_C(i, j) = \begin{cases} 1 : \text{原形質領域} \\ 0 : \text{その他の領域} \end{cases}$$

(6) 核の極断面2次モーメント：J

$$J = \int_A r^2 dA / (\int_A dA)^2 \quad (17)$$

但し

r：重心から微小面積dAまでの距離

Jの値は，一般に円形に近い核ほど小さい。

(7) くぼみ量：C

単球とリンパ球の識別において，核の形

状から得られる特徴として，特徴量Jが有効であることを先に示したが，ここではさらに単球の核が一般に馬蹄形を呈することから，その“くぼみ”に着目し，“くぼみ”の定量化を行なう。まず図8のように核に外接する多角形を考える。このとき，同図の A_i, B_i 間のように互いに連結していない2点間が，多角形の辺となっている場合，その2点間に“くぼみ”があると判定し，くぼみ量Cを18式のように定める。

図7 核の形状パターン

A:核面積
S:核の形状
J:核の極断面2次モーメント
C:くぼみ量

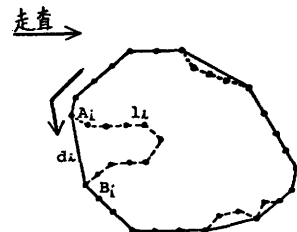


図8 くぼみの検出

$$C = \sum_i \frac{\theta_i}{L} \cdot \frac{\theta_i}{d_i} = \frac{1}{L} \sum_i \frac{\theta_i^2}{d_i} \tag{18}$$

但し L : 核周囲長, θ_i : A_i, B_i 間の凹部の長さ, d_i : 多角形の辺 A_i, B_i 間の長さ

(8) クロマチン量 : Q_N

単球と大リンパ球において特に核の形状からは識別困難なものが存在する。このような場合、検査技士は通常原形質の色調、アズール顆粒、^{*}クロマチン網工の相違点に着目して両者を識別している。そこで本稿ではクロマチン網工に着目し、これが単球では比較的繊細緻密であり、リンパ球では粗大粗荒であることから、局所的な変化をとらえることができ、処理が比較的簡単であるラブラシアン演算を用い、その値の平均と分散により、両者の相違を定量化する。

$$Q_N = \frac{1}{A} \sum \Delta(i, j) \tag{19}$$

$$\sigma^2 = \frac{1}{A} \sum (\Delta(i, j) - Q_N)^2 \tag{20}$$

但し

A : 核面積

$$\Delta(i, j) = |4f(i, j) - \{f(i-1, j) + f(i+1, j) + f(i, j-1) + f(i, j+1)\}| \tag{21}$$

(9) 顆粒量 : G_N

顆粒球である好酸球と好塩基球を他の白血球から識別するために、次の様に顆粒量を定義する。まず核領域の各画素 (i, j) に対して、その濃度 $f(i, j)$ が式(22)を満す画素 (i, j) を突起点とする。

$$f(i, j) \geq f(k, \ell) \tag{22}$$

但し

$$(k, \ell) \in W_n(i, j) = \left\{ (p, q) \mid |i-p| \leq \frac{n-1}{2}, |j-q| \leq \frac{n-1}{2} \right\} \tag{23}$$

このようにして求められた突起点を基に、顆粒量 G_N を式(24)で定義する。

$$G_N = \sum_{m=1}^M g_m \tag{24}$$

但し

$$g_m = \frac{1}{n^2 - 1} \sum_{(k, \ell) \in W_n(i, j)} \{f(i, j) - f(k, \ell)\} \tag{25}$$

M : 突起点の総数

G_N の値は、核内の局所的な濃度のピークをとらえたものであり、突起点の数が多く、近傍の画素との濃度差が大きいほど大きくなる。

(10) 核の輪郭部分の濃度勾配 : D_E

好酸球と好塩基球を識別する場合、好酸球は原形質に顆粒が充満しており、核濃度と原形質濃度との差が小さく、また核の輪郭部分での濃度勾配が、好塩基球にくらべゆるやかであることに着目し、核の輪郭部分の濃度勾配 D_E を次の様に定義する。まず核内の各画素に対し、隣接する八つの画素のうち少なくとも一つが原形質の画素である場合、

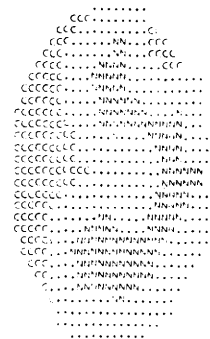


図9 核の輪郭部分

* 核内で糸状のクロマチン(核色素)がもつれあって網の目状等の構造をなしているが、そのもつれあって染色された様姿を核網工あるいはクロマチン網工と呼んでいる。

それらのすべての画素を核の輪郭部分とする(図9参照)。つぎにこの領域内の画素(i,j)に対し四方向について方向差分をとり(10式参照), それらの最大値をその画素(i,j)における差分値とし, 領域内で平均したものを D_E とする。

4. 特徴量の評価

図12は原形質濃度 D_C と核顆粒量 G_N の2次元分布を示したものである。好塩基球は, 顆粒が顆粒球中最大であり細胞全体を覆っているため, G_N の値の大きい方に分布している。また, 好酸球は原形質に顆粒が充満しているため, D_C の小さい方に分布している。(D_C の値は, 濃度レベルが高いほど数値は小さい。) 同図より顆粒球である好酸球と好塩基球が他の白血球, すなわち好中球, 単球, リンパ球から識別可能であることがわかる。さらに図13より核の輪郭部分の濃度勾配 D_E により, 好酸球と好塩基球が識別できることが明らかである。

図7は, 好中球, 単球, リンパ球の代表的なものについて, 核のパターンと面積 A , 形状 S , 極断面2次モーメント J , くぼみ量 C の各特徴量の数値を示したものであり, 図10は, 核形状-くぼみ量の2次元分布を示したものである。同図より好中球は一般に核の形状が複雑であるので, 単球, リンパ球から容易に識別できることがわかる。

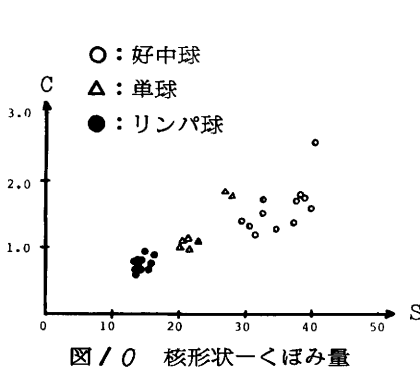


図10 核形状-くぼみ量

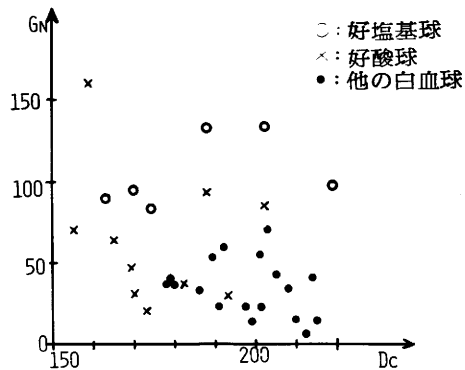


図12 原形質濃度-核顆粒量

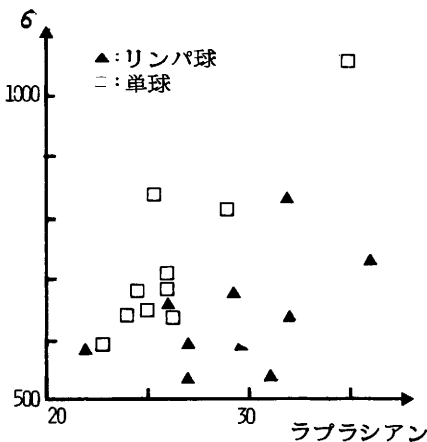


図11 ラプラシアン の 平均値-標準偏差

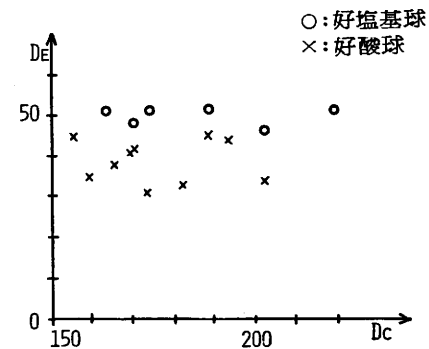


図13 原形質濃度-核の輪郭部分の濃度勾配

小，中リンパ球は，核の濃度レベルも比較的高く，核の形状もほぼ円形であり，S，J，Cなどの特徴量を用いれば，容易に識別できる。しかし，大リンパ球には核のいびつなものもあり，濃度レベルも低く，特に単球との識別が困難なものがある。しかし，単球には核に切れ込みがあるものが多く，くぼみ量Cが大きな値を示すため，特徴量Cは，両者の識別の際に有効な特徴量であると考えられる。また単球の核にこのような顕著な特徴がなく，核の形状からは識別が困難な場合には，クロマチン量 Q_N が有効なパラメータとなることが図11から確かめられる。

5 む す び

白血球の自動分類において，核領域を正しく抽出することは，識別精度を向上させる上で極めて重要である。本稿では，このためのしきい値を，画像の方向差分による濃度ヒストグラムから求める手法を示し，これにより従来，赤血球の存在などのため困難であった画像に対しても，安定に核を抽出できることを示した。

また新しい特徴量として，好酸球と好塩基球を他の白血球から識別するために“顆粒量”を定め，その有効性を示した。さらに核濃度から“クロマチン量”を定義し，これが形状からは識別困難であるような単球と大リンパ球の識別の際に，有効なパラメータとなることを示した。

謝 辞

本研究に御協力いただいた，山口明氏，金盛繁夫氏，また本論文を書くにあたり，御協力いただいた研究室の各氏に感謝する。

参 考 文 献

- (1) 谷口，小川，酒井，“白血球の自動分類（特徴抽出と識別法その1）”福井大工研 Vol.24 No.1（1976）
- (2) 森，大津，“認識問題としての2値化と各方法の検討”情報処理学会，イメージプロセッシング研資15-1（1977）
- (3) 山口，谷口，小川，“白血球自動分類における特徴抽出”信学会，PRL研資，77-48（1977）
- (4) 中村，山口，金盛，小川，谷口，“白血球自動分類における特徴抽出（その2）”信学会 PRL研資，78-11（1978）

<付 記>

本研究の一部は昭和49，50，51年度科学研究費の補助金を受けて行ったものである。

