

# 「定量化」時代の生物画像処理

## Biological Image Processing in the Quantitative Age

木 森 義 隆

Yoshitaka Kimori

自然科学研究機構新分野創成センターイメージングサイエンス研究分野

**要 旨** 近年のバイオイメーシング研究においては、生物画像情報の「定量化」がキーワードになっている。これまで目視で処理してきた事象に対し、数量的根拠を示すことがより重要視されるようになってきた。本稿では、定量化過程における計算機の内部画像処理のうち、特に重要なものに焦点をあて、その内容を概説する。とりわけ、形状計測過程で生じるデジタル誤差や測定誤差などについて取り上げ、それらが生じる原因を考察したい。

キーワード：定量化、画像処理、形状解析、セグメンテーション

### 1. はじめに

電子顕微鏡あるいは各種の光学顕微鏡を用いた、いわゆるバイオイメーシング研究においては、「定量化」がキーワードになっている。これまで目視で処理してきた事象に対し、数量的根拠を示すことが、より重要視されるようになってきた。解析対象を目視で観察し、その内容を漏らさず記述していくということよりも、むしろ、対象の性質を抽象化し、それを言語ではなく定量的に記述することに力点を置いた研究が展開されつつある。国内の動向としては、2011年から画像情報処理および生命科学分野の研究者が一堂に会する、「バイオイメーシ・インフォマティクスワークショップ」<sup>1)</sup>が開催されており、「定量生物学の会」<sup>2)</sup>を中心として形成されている研究者コミュニティも今後の発展が見込まれている。また、2014年より、総合研究大学院大学においても「定量生物学」の授業科目が開講され<sup>3)</sup>、当該研究領域における人材育成も行われるようになった。

定量化を目的とするバイオイメーシング研究においては計算機の利用が前提となる。計算機に実装された画像処理・解析過程は、基本的に全て開示できるため、その過程に対し客観性、再現性、反証可能性等の担保が期待される。これらは当該研究の大きなメリットといえる。ただ、注意すべきは、ブラックボックス化して使用した解析ソフトウェアからの出力データや、計算機上での手動操作による定量化結果には、これらの条件はあてはまらないケースもあるということである。処理内容を把握し、判断基準を明確に定義できてはじめて有用な結果になり得る。

本稿では、定量化過程における計算機の内部画像処理のうち、

特に重要なものに焦点をあて、その内容を概説する。とりわけ、形状計測に関するデジタル誤差や測定誤差など、これまで無視できるとされてきた事柄についても改めて取り上げ、それらが生じる原因を考察したい。

### 2. 「定量化」とは何か

図1aに生物画像の一例を示す。これは、カボチャ子葉細胞の電子顕微鏡写真である<sup>4)</sup>。この画像の構成関係をみると、まず、背景と細胞領域に分けることができ、つぎに細胞領域は核や葉緑体、ペルオキシソームなどのオルガネラの部分領域に分けることができる。この構造を図1bに示す。このような生物画像に対する定量化とは、画像中に存在する解析対象（細胞全体や各種オルガネラなど）の属性を計測し、数値化することである。通常、対象の属性としては、幾何学的形状（領域の面積や周囲長など）、および対象領域の輝度（各画素の保持する光強度）で表現される。

ただ、これらの属性は原画像中にそれとして認識できるような特異的な符号付きで存在しているわけではない。通常、解析対象は全画像中の部分領域となるが、その画素値は背景などの「その他」の領域と明確に区別できるような、特別な情報が付与されているものではない。たとえば、ある蛋白質を蛍光色素でラベルし、蛍光顕微鏡で観察したとする。ラベルが存在する領域は周囲に比べ輝度が高くなるが、そこに含まれる画素の値はその領域のみに特異的に出現するものではない。したがって、特定の領域を抽出（検索）するために、直接的に輝度値の情報を用いることはできない。この点が、生物学的データとしての画像が、その他のデータ、すなわち塩基配列や蛋白質の3次元結晶構造データなどと大きく異なる点である。

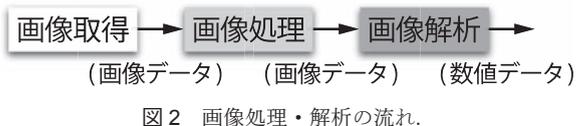
塩基配列は、塩基の情報をA, T, G, Cという記号で表現し、3次元結晶構造データは、特定原子の位置をxyz空間の

〒444-8787 愛知県岡崎市明大寺町東山5-1

E-mail: kimori@orion.ac.jp

2015年1月27日受付

座標として一意に表現できる。これらのデータは、人間による情報処理のため、特定の生物学的情報に一对一対応するかたちで、あらかじめ定められた符号である。したがって、ある塩基配列データの GC 含量を知りたい場合は、配列中の符



号“G”および“C”の数を計算することで目的が達成できる。これに対し、画像データはこのような処理目的の符号化がなされていない情報で構成されている。このため、原画像から、出力として要求される数値データに至るまでには隔りがある。この隔りを埋めるものが画像処理および画像解析過程である。

### 3. 生物画像における画像処理・解析

最初に画像処理過程と画像解析過程の違いをみていくことにする。もちろんこれらは、統合されたシステムとして機能しており、両者の区別は明瞭化できない部分もある。しかし狭義には、各過程に係る入力と出力データの観点から区別することができる。画像処理過程の入力データは画像であり、出力も画像である。一方、画像解析過程においては、入力データは画像であり、出力は数値（記述データ）となる。これらの過程を含んだ、画像取得から始まる画像データの処理・解析の流れは図2のようになる。この定義によれば、「定量化」は、画像解析過程が担うことになる。画像処理過程では解析対象の抽出を行い、その属性を計測できるような画像に変換することが求められる。3.2節で説明するが、対象の属性に対する解析方法は多数提案されており、とくに形状特徴に対する計測方法は一意性を持っている。これに対し、定量化のための画像処理はいまだ定まった方法（やその組み合わせ）は存在していない。提案されている種々の手法の適用範囲は非常に限定されており、汎用性や頑健性のあるアルゴリズムの構築が大きな課題となっている。その中でも定量化に必須でありながら、最も実現が困難な画像処理のひとつがセグメンテーション処理<sup>5-8)</sup>である。

#### 3.1 定量化のための画像処理：セグメンテーション

セグメンテーションは、背景領域と解析対象領域とを区別する処理である。上述したように原画像には、解析対象領域を示す特異的な情報は付与されていないため、計算機には両領域の区別ができない。セグメンテーション処理により、背景領域に対し、解析対象領域に「レベル差」を与えることにより、その領域を認識できるようにする。レベル差のある画像とは、背景領域の画素値を0、対象領域の画素値を1とするような2値化画像のことである。原画像を $f(x, y)$  ( $x, y$ は画素の座標)、2値化画像を $B(x, y)$ と表し、解析対象領域（部分集合）を $\Gamma$ とすると2値化処理は以下のように書くことができる。

$$B(x, y) = \begin{cases} 1 & \text{if } f(x, y) \in \Gamma, \\ 0 & \text{if } f(x, y) \notin \Gamma. \end{cases} \quad (1)$$

セグメンテーション処理においては、この部分集合 $\Gamma$ をいかに定めるかということが本質的な問題である。しかし、

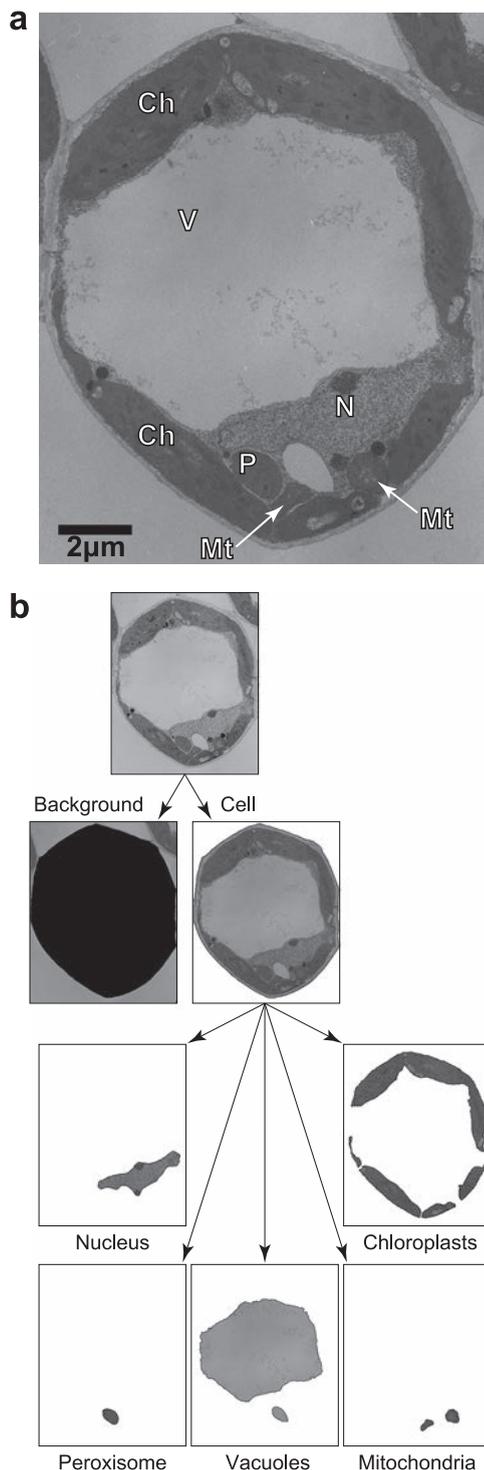


図1 (a) カボチャ子葉細胞の電子顕微鏡写真。画像は、PODB3: The Plant Organelles Database 3<sup>19)</sup>より取得（データ登録者：Kondo, M., National Institute for Basic Biology）。Ch: Chloroplast, Mt: Mitochondrion, N: Nucleus, P: Peroxisome, V: Vacuole. (b) 画像の構造。

この問題の解決は極めて難しい。なぜなら、この問題には、画像中の対象領域  $\Gamma$  を認識するためにはセグメンテーションが必要であるが、セグメンテーションするためには、 $\Gamma$  があらかじめ与えられて（すなわち認識されて）いなければならないという因果性のジレンマが存在するためである。これを解消するために考えられることのひとつは、対象領域を識別するための具体的なモデルを与えることである。対象が細胞であるなら、そのモデルを構築することになる。モデルを用い、画像中から類似する領域を検出し、その領域を  $\Gamma$  と定める。ただ、この場合、一般に複雑な形状の細胞をどのように表現するかという問題がでてくる。また、たとえば、いったんモデルが構築できたとしても、実際の細胞像とそれを合致させるためには、形状変形が必要となり、複数の変形モデルが必要となろう。しかし、変形の程度の妥当性の判断や、変形モデルと実際の細胞像との差を表すような評価法の設定も困難であり、結局は、認識対象モデルの導入も現実的な解決方法ではない。

現状では、実際に選択される2値化方法としては、輝度閾値処理となろう。本処理では、部分集合  $\Gamma$  は閾値  $T$  を与えることにより定められる。これは以下の条件式として表すことができる。

$$B(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq T, \\ 0 & \text{if } f(x,y) < T. \end{cases} \quad (2)$$

輝度閾値処理を用いる場合、セグメンテーション処理は画像  $f(x,y)$  を解析対象領域と背景領域とに区分するための閾値  $T$  を求めることと同値と言える。

図3に輝度閾値処理による2値化の実例を示す。図3a(左側)は原画像(8bit濃淡画像)であり、図3bはその輝度値ヒストグラムである。ヒストグラムはバイモーダルな分布となっており、ふたつのピークは、それぞれ背景の平均輝度付近と対象領域の平均輝度付近とに存在している。したがって、これらのピークの谷にあたる輝度値に閾値  $T$  を設定すれば、ふたつの領域を区分できることになる。自動的に閾値を選定する方法はこれまで数多く提案されているが<sup>9,10)</sup>、ここでは、Otsuの方法<sup>11)</sup>を用いて2値化処理を行った。Otsuの方法とは、輝度値ヒストグラムに判別分析を適用し、最適な  $T$  を選定する方法である。ヒストグラム中の背景領域および解析対象領域に対し、それらのクラス内分散とクラス間分散の比を分離度として求め、分離度が最大になる値を  $T$  とする。

この手法による2値化処理の結果を原画像の右側に示す。この画像の場合、 $T=129$ であった。原画像は2値化画像として、背景が黒、対象領域が白い画素で表現される。このように表現されてはじめて、計算機が解析対象の存在を認識できるようになる。

しかし、実際に処理すべき画像の多くでは、輝度値ヒストグラムがバイモーダルな分布を示すものは稀であるし、対象領域の輝度値がひとつのピークに割り当てられている保証もない。図4aには、蛍光顕微鏡で撮影された植物細胞の画像

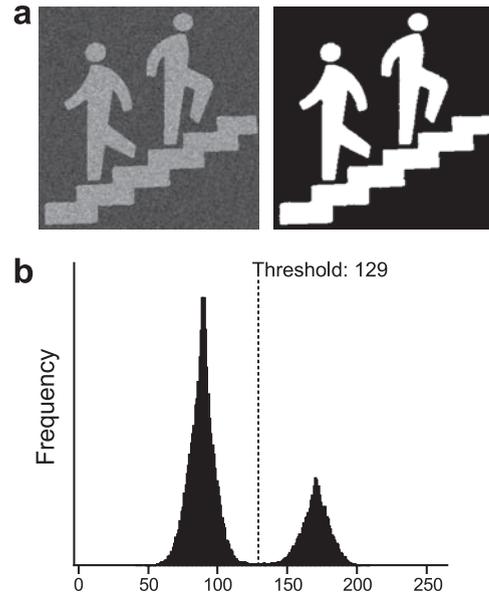


図3 閾値処理によるセグメンテーション。(a) 原画像(左)とその2値化画像(右)。(b) 原画像輝度値のヒストグラム。

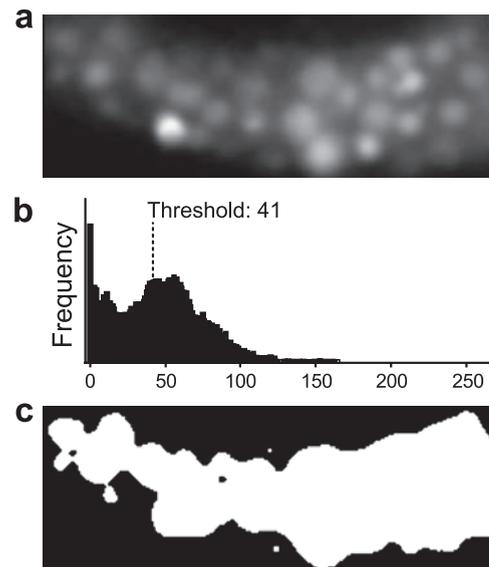


図4 (a) 原画像. PODB3<sup>19)</sup>より取得(データ登録者:Yoshimoto, K., National Institute for Basic Biology). (b) 原画像の輝度値ヒストグラム。(c) 自動閾値法による2値化結果。

を示す<sup>12)</sup>。GFP蛍光により、オートファゴソーム膜の蛋白質(AtATG8a)が可視化されており(粒子状にみえる)、この存在領域を解析対象領域とする。輝度値ヒストグラムを図4bに示すが、解析対象領域のピークは不明である。Otsuの方法による2値化画像( $T=41$ )を図4cに示す。細胞領域がおおまかにはセグメンテーションされているが、解析対象のセグメンテーションに失敗していることがわかる。再度、原画像(図4a)を見ると、解析対象領域の輝度レベルがその周囲とあまり差がないことがわかる。対象領域のコントラストが極めて弱い状態であり、このような場合は輝度閾値法単独でのセグメンテーションは困難である。そこで、その前

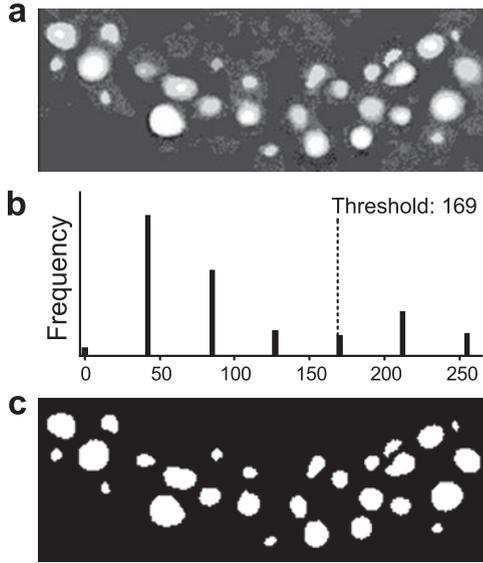


図5 (a) 対象領域の強調画像. (b) 輝度値ヒストグラム. (c) 自動閾値法による2値化結果.

処理として、コントラストの強調処理が有効に働く場合がある。ただ、この際、画像全体のコントラストを無選択に強調するのではなく、解析対象領域を選択的に強調する必要がある。このため、ここでは rotational morphological processing (RMP) に基づく white top-hat フィルタ<sup>13)</sup>を適用した。原画像にこのフィルタを適用した結果を図5aに示す。RMP に基づく white top-hat フィルタにより、対象物が選択的に抽出され、それらの領域は高い輝度値が割り当てられている。その輝度値ヒストグラムは図5bである。バイモーダルに近い分布形状になっていることがわかる。Otsuの方法による2値化の結果 ( $T=169$ ) は図5cである。解析対象領域が明瞭にセグメンテーションされていることがわかる。

通常、セグメンテーション処理はこのような複数の処理で構成されるが、どの対象にも対応できるような最良の構成は定まっていない。対象に応じて最適な構成になるよう工夫していくことが求められる。そして、この処理によってはじめて対象領域に対する解析(定量化)が可能になる。

なお、セグメンテーションの問題において、人間の「知識」の利用が重要とされる場合もある。その具体例のひとつが上述した細胞モデルの導入である。ただ、現段階では、私たちが観察している細胞の微細な形態やその動的構造の情報を過不足のないかたちで表現する(定式化する)ことは不可能のように思われる。このような知識利用型のセグメンテーションの体系化に関してもいまだ多くの課題が残っている。

### 3.2 画像解析：形状特徴の定量化

ここでは、画像解析過程における解析対象の形状特徴の計測手法について述べる。この計測を行うためには、解析対象はセグメンテーションされていなければならない。本過程への入力画像は2値化画像となる。領域の形状特徴のうち、もっとも基本的なものは空間モーメントである。画像  $f(x, y)$  の

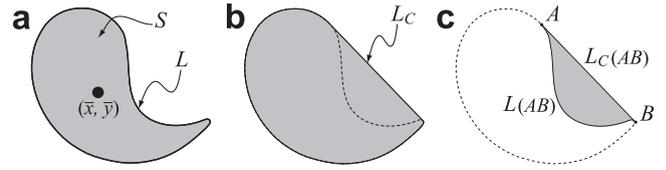


図6 領域の形状特徴量.

$(p+q)$  次のモーメントは次のように表される。

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y). \quad (3)$$

また、0次 ( $p=q=0$ ) のモーメント  $m_{00}$  は以下のようになる。

$$m_{00} = \sum_x \sum_y f(x, y). \quad (4)$$

これは、画素値  $f(x, y)$  の総和であり、2値化画像の場合は対象領域の面積(画素の総和)を表す。また、1次のモーメント  $m_{10}$  ( $p=1, q=0$ )、ならびに  $m_{01}$  ( $p=0, q=1$ ) を  $m_{00}$  で正規化すると、次式で表すように対象領域の重心 (centroid) の座標  $(\bar{x}, \bar{y})$  が与えられる。

$$\bar{x} = \frac{m_{10}}{m_{00}} = \frac{\sum_x \sum_y x f(x, y)}{\sum_x \sum_y f(x, y)}, \quad (5)$$

$$\bar{y} = \frac{m_{01}}{m_{00}} = \frac{\sum_x \sum_y y f(x, y)}{\sum_x \sum_y f(x, y)}. \quad (6)$$

以下にその他の代表的な形状特徴量を挙げる。まず根源的なものとしては、対象領域の面積  $S$  や周囲長  $L$  がある(図6a)。周囲長とは、対象領域の境界(背景と接する画素で構成される)における画素単位の距離である。これらを用いて次のような特徴量を定義できる。

$$\text{円形度} : \frac{4\pi S}{L^2}. \quad (7)$$

$$\text{複雑度} : \frac{L^2}{4\pi S}. \quad (8)$$

円形度は0から1の値をとり、領域が円形の場合に値が最大となる。対象のコンパクトさを測る特徴として多用される。また、この逆数は複雑度と呼ばれ、領域の輪郭が複数の凹凸で構成されるなど、複雑になるほど値が大きくなる。

また、領域を囲む最小の凸領域(凸包, convex hull)を求め、その周囲長を  $L_c$  とすると(図6b)、凸状度(convexity)を定義できる。

$$\text{凸状度} : \frac{L_c}{L}. \quad (9)$$

領域に凹部が多い形状ほど(すなわち、 $L$  の値が大きくなるほど)凸状度の値は小さくなる。また、領域全体に占める凹部の割合は、凹率として求められる。凹部は、凸包から対象領域を差分することで抽出できる(図6c)。

$$\text{凹率} : \frac{L(AB)}{L} \cdot \frac{L_c(AB)}{L_c(AB)} \quad (10)$$

ここで、 $L(AB)$  は、弧  $AB$  の長さ、 $L_c(AB)$  は弦  $AB$  の長さである。

なお、上記以外にも、Fourier 記述子、wavelet 記述子を用いた輪郭線の記述や Euler 数による領域の位相幾何学的記述、さらに pattern spectrum による幾何学パターンの分布表示や fractal 次元を用いた複雑さの計量など、多数の形状特徴の表現・定量化手法が考案されている。詳細は、文献 14)、15) を参照されたい。

さて、このような特徴量の適用に際し、実用上で問題になることがある。特徴量の多くは面積  $S$  と周囲長  $L$  に基づいているため、まずはこれらを計測する必要がある。しかし、面積は対象領域の画素を数えあげるにより容易に求まることに對し、周囲長の計測はそれほど容易ではない。デジタル画像特有の誤差（デジタル誤差）を考慮しなければならないためである。

周囲長計測の定義は複数存在するが、通常用いられている方法は以下のふたつである。

- 1) 対象領域の境界を構成する画素の総和を求める。

$$L_1 = N. \quad (N \text{ は境界画素の総和}) \quad (11)$$

- 2) 境界画素において、ある画素と水平・垂直方向に接する画素との距離を 1 (単位) とし、対角方向に接する画素との距離を  $\sqrt{2}$  (単位) とし、隣接する境界画素間の距離の総和を求める。

$$L_2 = N_0 + \sqrt{2}N_D. \quad (12)$$

( $N_0$  は水平・垂直方向の連結数であり、 $N_D$  は対角方向の連結数)

どの定義を用いるかにより、当然、結果が異なってくる。実際に図 7 に示す領域に対し、これらの定義を用いて測定してみると、定義 1 (式 (11)) では、 $L_1 = 22$  となり定義 2 (式 (12)) では  $L_2 = 6 + 16\sqrt{2} \approx 28.6$  となった。

さらに精度よく求めるために、様々な補正式が提案されている<sup>16)</sup>。たとえば、 $L_2$  に補正係数 0.948 をかけた定義式  $L_3$  は次のようになる。

$$L_3 = 0.948(N_0 + \sqrt{2}N_D). \quad (13)$$

また、誤差の起因である輪郭の凹凸形状の過剰評価を防ぐため、輪郭に平滑化処理を施した後、周囲長を計測するという方法もある。ただし、平滑化により、本来抽出すべき特徴を除去するという懸念もあるため、適用の際は平滑化のレベルに注意が必要である。

いずれの定義を用いるにせよ、解析対象の性質や分解能を勘案し、最適なものを選択しなければならない。

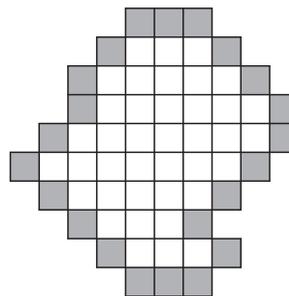


図 7 領域の周囲長. 灰色の画素を境界画素とする. 定義 1 (式 (11)) による周囲長の計測値は、 $L_1 = 22$  であり、定義 2 (式 (12)) による計測値は  $L_2 = 6 + 16\sqrt{2} \approx 28.6$  である。

#### 4. 手動計測による画像情報の定量化

画像中の解析対象の特性を定量化する場合は、まずその存在領域を切り出してくる必要がある、これを実現するものがセグメンテーション処理であった。しかし、上にも述べたが、自動セグメンテーション手法の適用対象は限られており、様々な画像データに対応できるような汎用性や頑健性に乏しい。そこで、現状では、(1) 解析者が対象を認識し、(2) 手動操作によって対象を切り出し、(3) 切り出された領域の特徴を計測する、という手動操作に基づく定量化が実施されることが多い。(2)の過程では、解析者がマウス等のポインティング・デバイスを用い、解析対象の輪郭をトレースし、閉じた曲線を作成する。そしてその内部を埋め、ひとつながりの領域を対象領域として切り出してくる。多くの場合、切り出された対象領域は 2 値化表現されているため、(3)の過程は計算機による自動処理が可能になる。

ただ、この処理の際、輪郭のトレースの正確さが定量化結果に直接影響することは自明である。しかし、これまで、得られた結果の精度やそれに含まれる誤差について、ほとんど考慮されることはなかった。本節では、手動トレースによる輪郭線抽出実験の実施結果に基づき、定量化操作に係る測定誤差の評価を試みる。

本実験では実際の画像計測を想定し、2 種類の図形の画像データ (Sample1, Sample2) とそれらにノイズを重畳した画像データ (Sample1\_noise, Sample2\_noise) を解析対象として用意した (図 8 上段)。これらの画像を計算機モニタで表示し、被験者 (15 人) が同一の計算機環境のもと、マウス操作による、図形の輪郭トレースを実行した。トレースした軌跡は一定幅の線として表示され、最終的に各図形の輪郭抽出画像 (2 値化画像) として保存された。なお、本実験ではあらかじめ式 (12) に基づき、Sample1 および Sample2 の画像の周囲長を自動計測し、その結果を「真値」(Ground Truth) として用いた。Sample1 に対する真値は 429.20 であり、Sample2 は 584.57 であった。

手動トレースによる輪郭線抽出画像は、程度の差はあるが被験者ごとにぶれや特定の方向への偏りなどの誤差を含んでいた。まずは、平均化によるこの誤差の低減の度合いを確認

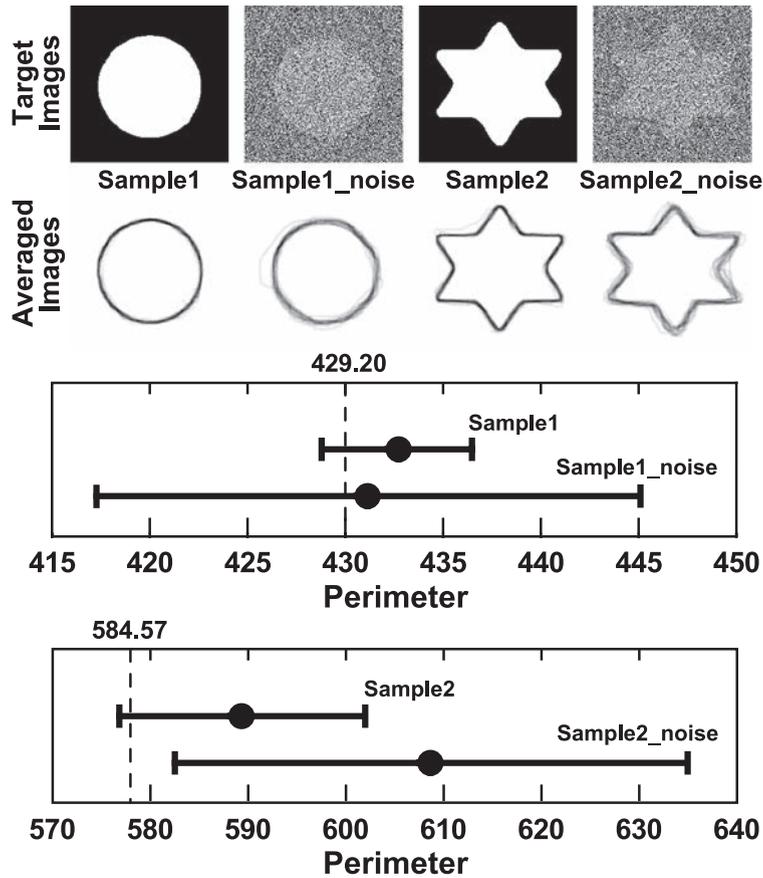


図8 手動トレースに基づく輪郭線抽出. 上段: 解析対象画像と輪郭線抽出結果の平均化画像. 下段: 周囲長計測結果. 平均値 ± 標準偏差の値を示している (Sample1:  $432.71 \pm 3.77$ , Sample1\_noise:  $431.16 \pm 13.89$ , Sample2:  $589.38 \pm 12.50$ , Sample2\_noise:  $608.71 \pm 26.09$ , それぞれ  $N=15$ ).

した. 図8の各対象画像の下に, それに対する被験者全員の輪郭線抽出結果の平均化画像を示す.

Sample1 および Sample2 をトレースした平均化画像は, その輪郭が対象画像のそれに近似されていることがわかる. これは, 偶然的に発生するゆらぎ等の誤差が平均化によりキャンセルされたことを表す. これに対し, Sample1\_noise および Sample2\_noise を対象とした結果は, 平均化されてもおお輪郭線に位置ずれがあり, 誤差がより大きかったことがわかる. これは, ノイズにより対象画像の輪郭線が不明瞭になっているため, その特定に対する被験者の認識に差異が生じているためと考えられる.

次に, このような個人誤差を含んだ輪郭線抽出画像の周囲長を計測した. 図8下段には式(12)を用いて計測した結果(平均値 ± 標準偏差)を示す. いずれの結果も周囲長の平均値(グラフ中の黒丸で示す)は, 真値(点線で示す)よりも大きい. これは, 手動トレースによる輪郭線は誤差として局所的に凹凸が生じているため, それらの距離を評価した周囲長としては, 真値よりも長くなる傾向があることを示す. さらに, 標準偏差はノイズのある画像でより大きくなっていることがわかる. 対象画像の輪郭線の不明瞭さによって, 被験者の輪郭線の決定に大きなばらつきがあることが評価されている.

表1 平均誤差率 [%]

|               |      |
|---------------|------|
| Sample1       | 0.99 |
| Sample1_noise | 2.03 |
| Sample2       | 1.44 |
| Sample2_noise | 4.67 |

さらに輪郭抽出画像の周囲長の測定値を  $v_M$ , 真値を  $v_{GT}$  と表し, 以下の式を用いてこれらの誤差率 ( $\epsilon$ ) を計算した.

$$\epsilon = \frac{|v_M - v_{GT}|}{v_{GT}} \times 100[\%]. \quad (14)$$

表1にそれぞれの画像に対する誤差率の平均値を示す. やはり, ノイズのない画像で誤差は小さくなり, ノイズが重畳すれば誤差は増大している. また, Sample1 に比べ Sample2 の方が誤差が大きい (Sample1\_noise に対する Sample2\_noise の誤差も同様). これは, Sample2 の方がトレースする輪郭の距離が長いため, 誤差が蓄積する傾向にあることを示している.

このような計測誤差を低減するためには, 複数回の計測結果や複数計測者の結果の平均が必要になる. また, 作業の標

準化も重要になってくる。計測の手順や判断基準を明確にし、再現性の担保された手動計測による定量化を実施することが求められる。

## 5. おわりに

定量化を目的としたバイオイメージング研究においては、その分野のひろがりに呼応し、ImageJ<sup>17)</sup> やその関連ソフトウェア<sup>18)</sup>をはじめとして、様々な画像処理・解析ツールが提供されている。多様な選択肢が存在し、従来できなかった解析手法の構築も可能になってきた。しかしその一方で、いかにそれらを便利に使いこなすかという、“how-to”ばかりが強調されている側面もある。繰り返しになるが、ソフトウェアの内部処理をブラックボックスとして出力された定量化データには、それに本来期待される客観性等の条件は付与されていない。それは本質的に目視に基づく主観評価結果と変わるところはない。定量化を指向するバイオイメージング研究では、結果に至る全ての過程をクリアにしておくことが要求される。このためには、デジタル誤差等を考慮したうえで、使用するアルゴリズムの詳細や計測手法の定義等を正確に把握しておく必要がある。

他方、現段階の画像処理・解析ツールを用いても、実現が困難な処理があることについて認識しておくことも重要である。画像処理として、最も重要になるのが、画像中から解析対象を切り分けるセグメンテーション処理であったが、その適用範囲は広いとはいえず、汎用性・頑健性のある処理の実現はまだ先になるように思える。したがって現状としては、この処理には、解析者の関与が不可欠となり、手動操作による定量化が実施されることになる。この場合においても、解析に係る様々な評価基準を明確にするとともに、計測誤差を適切に評価し、結果に反映させることが必要である。

これまで、画像データは最終的な研究結果となり得た。しかし、定量化時代のバイオイメージング研究においては画像自体が研究対象であり、それを科学計測の観点から処理することが要求される。本稿では、その基礎となる画像処理・解析技術とその過程で生じるいくつかの誤差の問題について概説した。

## 謝 辞

本稿の執筆にあたり、植物オルガネラデータベース (PODB3: The Plant Organelles Database 3) の画像を使用させていただきました。画像の使用についてご快諾いただいた、当該データベース構築代表者の真野昌二博士 (基礎生物学研究所) に感謝いたします。また、4章の「手動計測による画像情報の定量化」実験は、2014年12月10-12日の日程で開催された、自然科学研究機構新分野創成センター・基礎生物学研究所主催「生物画像データ解析トレーニングコース2014」(<https://is.cnsi.jp/biatc2014/>) 内で実施しました。被験者としてご協力いただいた受講者の方々に感謝いたします。

## 文 献

- 1) <http://www.bioimageinformatics.jp>
- 2) <http://www.q-bio.jp>
- 3) <http://ibep.ims.ac.jp/>
- 4) <http://podb.nibb.ac.jp/Organellome/podb3/result.php?id=71>
- 5) Haralick, R.M. and Shapiro, L.G.: *Comput. Vis. Graph. Image Proc.*, 29, 100-132 (1985)
- 6) Pal, N.R. and Pal, S.K.: *Pattern Recogn.*, 26, 1277-1294 (1993)
- 7) Zhang, H., Fritts, J.E. and Goldman, S.A.: *Comput. Vis. Image Und.*, 110, 260-280 (2008)
- 8) Ilea, D.E. and Whelan, P.F.: *Pattern Recogn.*, 44, 2479-2501 (2011)
- 9) Sahoo, P.K., Soltani, S. and Wong, A.K.C.: *Comput. Vis. Graph. Image Proc.*, 41, 233-260 (1988)
- 10) Sezgin, M. and Sankur, B.: *J. Electronic Imaging*, 13, 146-168 (2004)
- 11) Otsu, N.: *IEEE Transactions on System Man Cybernetics, SMC-9*, 62-66 (1979)
- 12) [http://podb.nibb.ac.jp/Organellome/bin/browseImage.php?ID=Image-k-yoshi\\_nibb.ac.jp-20060731091544](http://podb.nibb.ac.jp/Organellome/bin/browseImage.php?ID=Image-k-yoshi_nibb.ac.jp-20060731091544)
- 13) Kimori, Y., Baba, N. and Morone, N.: *BMC Bioinformatics*, 11, 373 (2010)
- 14) Loncaric, S.: *Pattern Recogn.*, 31, 983-1001 (1998)
- 15) Zhang, D. and Lu, G.: *Pattern Recogn.*, 37, 1-19 (2004)
- 16) Dorst, L. and Smeulders, A.W.M.: *Comput. Vis. Graph. Image Proc.*, 40, 311-333 (1987)
- 17) <http://imagej.nih.gov/ij/>
- 18) Schneider, C.A., Rasband, W.S. and Eliceiri, K.W.: *Nat. Methods*, 9, 671-675 (2012)
- 19) Mano, S., Nakamura, T., Kondo, M., Miwa, T., Nishikawa, S., Mimura, T., Nagatani, A. and Nishimura, M.: *Plant Cell Physiol.*, 55, e1 (2014)