

平成 21 年 5 月 29 日現在

研究種目：若手研究（スタートアップ）

研究期間：2007～2008

課題番号：19800062

研究課題名（和文）

ライブセルイメージにおける特定オルガネラ領域の自動認識システムの開発

研究課題名（英文）

Development of the image segmentation system for intracellular organelles on living cell images

研究代表者

竹本 智子（TAKEMOTO SATOKO）

独立行政法人 理化学研究所・生物基盤構築チーム・研究員

研究者番号：00450403

研究成果の概要：

細胞生物・分子生物学にとって、画像解析によるオルガネラの定量分析は極めて重要だが、大量かつノイズが多い画像において、画像解析の前処理として欠かせない領域抽出を実現することは難しく、研究のボトルネックとなっている。本研究では、研究者が注目するオルガネラの画像特徴を自動解析し、その特徴に適した領域抽出アルゴリズムを自動選択するシステムを開発した。これによって、時間変化する大量画像から注目対象を精度よく自動抽出できるようになった。抽出結果によって、オルガネラ領域の3次元観察、体積・表面積の計測などが容易に可能となり、よりスムーズな研究発展に大きく寄与することが期待されている。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	1,360,000	0	1,360,000
2008年度	1,350,000	405,000	1,755,000
年度			
年度			
年度			
総計	2,710,000	405,000	3,115,000

研究分野：総合領域

科研費の分科・細目：情報学、生体生命情報学

キーワード：領域抽出、パターン認識、細胞内画像処理、特徴抽出

1. 研究開始当初の背景

2003年に終了宣言が成されたヒトゲノム解読プロジェクトを始め、各種のゲノム解明計画による生物の塩基配列の解読が終了し、最重要課題は各遺伝子から発現するタンパク質が細胞の中でどのような働きをするのかを解明することへ移ってきた。しかし、遺伝子やタンパク質は個々に機能を発現するのではなく、相互に複雑に絡み合って大規模な生命現象を実現していると考えられたため、個別の分子構造を一つずつ特定していくので

は生命現象の理解には到底たどりつけないことが分かってきた。そこで研究代表者の所属する研究グループでは、従来研究とは理解の軸を反対向き、すなわち生命現象の最小単位である細胞の形態こそがゲノムから発現する遺伝子、及びタンパク質の機能の表現系であるという考えに基づいて生命現象の理解を目指そうとした。これは、近年の関連研究の成果により、生物系内のタンパク質を網羅的に解析し数値・定量化することが可能となっていることから、その高次な表現系で

ある細胞の形態を数値化することによって、両者の数値的な比較が容易になると考えたものである。

また、近年の蛍光顕微鏡観察に代表される細胞のイメージング技術は飛躍的進歩を遂げている。特定のオルガネラを染色するための蛍光色素や蛍光タンパク質の開発に加え、顕微鏡装置の高精度化も進んだ。最近では、時間経過に伴う細胞内オルガネラの動態をも詳細に観察できるようになった。しかし一方で、染色や励起光照射に対する細胞ダメージの問題を無視することはできず、オルガネラの観察に十分な分解能を得られない問題もある。そのような場合でも細胞生物・分子生物学に関わる研究者らは、特定のオルガネラの形態を画像から認識することが可能であると言われているが、これは知識や経験に基づく人間の高次な認識機構であると言える。そのため、いつ、誰が行っても必ず同じ結果を生成するとは言えず、データの客観性を主張することが難しいのである。さらに、顕微鏡の高精度化によって莫大量となったデータに対し、研究者らが全て自力で形態認識を行うことは不可能である。したがって、計算機による特定オルガネラの自動認識による、客観性の獲得、省力化が求められている。

2. 研究の目的

生命現象の理解を目標とする研究の多くは、1. ゲノム配列、2. 発現遺伝子、3. タンパク質、4. 機能の解明を主軸として行われているが、ゲノム配列には未だ不確定要素が多く、遺伝情報を正しく抽出できているかは不明である。また、1つの遺伝子からは異なる複数のタンパク質が生成される場合もあり、生命現象の機能解明までの道筋は未だ遠い。研究代表者が所属する研究グループでは、ゲノムの機能表現の高次な結果は、生物の形態であるという考えに基づき、形態の違いから遺伝子やゲノムを理解しようとする研究に関わっている。すなわち、従来研究の主軸方向とは逆向きのアプローチでゲノム配列と形態との関連を解明しようとするものであり、その第一歩として、生命を形態として捉えたときの基本単位である「細胞」の形態を情報科学の観点から明らかにしようとしている。

研究代表者は上記研究において、細胞や細胞内オルガネラの形態自動認識を目指している。形態の認識は、画像による定量解析の前処理として極めて重要な処理である。具体的には、生きた細胞の画像化技術である4次元(空間+時間)ライブセルイメージングによって取得した細胞のボリュームデータから、核やゴルジ体などの主要なオルガネラの形状、及び時間変化に対するダイナミクスを

自動認識技術により同定する。このとき取得したボリュームデータは大量で、さらに画像取得時のレーザー照射に対する細胞ダメージの問題から撮影時間を短縮せざるを得ないためにS/N比が極めて低いことがある。これらは人間が手作業でオルガネラの形態や位置を正しく認識することを妨げるのに十分であり、上記の研究目的達成の大きな妨げとなっている。そのため、細胞生物や分子生物学の専門家が自身の判断によって行う形態の認識の能力を、仮に画像データが大量で、S/N比が低くとも計算機が精度よくその能力を再現できるシステムが強く求められていた。

以上のことから本研究では、専門家(すなわち、細胞や細胞内オルガネラの観察者)が一部のボリュームデータに対して認識対象の領域の一部を指定することにより、計算機が動的に変化する対象の領域を自動認識し、可視化することを目的としたシステムの開発を行うものである。本システムの実現により、時々刻々と変化する細胞内のオルガネラを自由視点からの観察し、各オルガネラの体積や表面積、他のオルガネラとの位置関係を解析することが容易になり、ライブセルイメージングの新規性と相俟って細胞の4次元的な形態をはじめ定量的に捉えられることが期待される。

3. 研究の方法

本研究で開発するシステムは、ライブセルイメージングにより取得した大量のボリュームデータに対し、観察者が特定のオルガネラ領域の一部(例えば、1時相)を教師データとして指定することにより、残りの全領域を計算機が自動認識するものである。すなわち、システムに求められるものは観察者による認識の再現である。しかし人は、画像から何らかの特徴を見出し、それらと認識対象に関する知識を脳内で照合することによって、対象の存在領域を認識できるのであり、その機能は人間の脳内でも高次の処理である。そのプロセスは完全に解明されていないため計算機への実装は容易なことではない。

そこで本研究では人の画像認識に関する能力を、パターン認識による画像特徴量の識別問題と捉えて単純化することとした。この実現のために本研究で求められることは、

(A) 各オルガネラを表現する画像特徴量の検討、及び画像特徴量から生成される特徴空間において、オルガネラとそれ以外を識別する
(B) 識別関数の設計構築である。また、前述のように、同一のオルガネラでも位置により特徴が異なり、それらが時間変化することから、(C) 特徴量と識別関数の自動選択を実現する必要があると考えた。この(C)において特に問題となるのが、選択の評価規

準の定義であることから、研究期間の後半では画像特徴量の算出や識別に関わるアルゴリズムを自動選択できるプロトタイプシステムを検討し、どの選択基準がオルガネラの認識に最も適しているかを評価できるような枠組みを提案した。以上の検討に関して平成19～20年度に行った具体的な内容を以下に示す。

(1) 画像特徴量の検討

研究代表者は、人が画像認識において注目する特徴は、主に輝度、テクスチャ、境界の有無であると考えている。このうち研究代表者は、テクスチャ特徴の一種である高次局所自己相関特徴や、局所領域におけるモーメント特徴がオルガネラの特徴を記述するのに特に適しているのではないかと考えた。例えば高次局所自己相関特徴は、近傍画素の濃淡値の積を画像全体に対して足し合わせて得られる極めて局所的な特徴を数値化したものであることから、オルガネラのように微小領域での特徴変化に注目する必要がある場合に有効であると考えた。この他にも、輝度や境界の有無から生じる特徴量からなる高次元特徴空間を複数生成し、オルガネラの特徴表現として適したものを検討した。

本研究ではまず、細胞生物・分子生物の専門家にライブセルイメージからの各オルガネラの領域指定をしてもらい、認識の教師データとした。次にこの教師データを用いて、観察対象とするオルガネラとそれ以外の画像特徴を算出し、上記の画像特徴のうちオルガネラとそれ以外を最も良く識別できる特徴について、バイズ誤り確率を用いて評価した。

(2) 識別関数の設計構築

一般に識別関数を設計するとき、特徴空間における特徴量の確率的分布構造が事前に既知の場合は、バイズ識別関数を用いることにより識別問題は解決する。しかし、生物を対象とした場合、事前に分布構造が分かっていることはなく、分布構造を推定する必要が生じる。分布構造の推定は通常、確率密度分布の推定により議論され、その方法は大別してノンパラメトリック、パラメトリック、セミパラメトリック法に分けられる。それぞれ一長一短であることが知られていることから、対象に応じて各方法の識別性能を比較する必要がある。本研究ではそれぞれの代表的なものとして、線形判別分析 (Linear Discriminant Analysis)、k-最近傍決定則 (k-Nearest Neighbor)、一般化線形モデルの1つである多項ロジットモデル (Multinomial Logit Model) を実装し、オルガネラ領域の特徴量を識別することとした。

(3) 特徴量と識別関数の選択

時間変化するオルガネラの認識に有効と考えられる画像特徴量は、一意に定義できるものではない。また、算出した特徴量を識別する規則にも多くの解決法が存在し、与えられた特徴量分布の特性や計算コストなどを考慮して最適なものを決定する必要がある。すなわち、(1)～(2)で検討した方法がすべての場合に有効であるとは限らないのである。

この問題に対し、本研究では入力として教師データが与えられたとき、対象によって特徴量と識別関数の最適組み合わせを自動的に選択するシステムの構築を行った。このときに何をもち「最適」と見なすかを定義することは容易ではない。そこで本研究では、各組み合わせによるオルガネラ領域の自動認識結果と、教師データとの領域の類似性を、「一致度」及び「輪郭形状間距離」の2つの尺度を用いて評価することにより「最適」な方法を決定することを目指した。輪郭形状間距離には、輪郭形状を複素自己回帰モデルによる離散的な点列で記述し、点列間のユークリッド距離等で数値化したものを用いた。これは回転、相似変換に対して不変な量であることが保障されており、動的に位置が変化するオルガネラ領域の評価にも有効であると考えた。

以上(1)～(3)で検討した内容のシステム開発を行った。なお、特徴量と識別関数を選択には大量の計算コストが必要であることから、2048CPUのLinuxクラスタシステムからなるRSCC (RIKEN Super Combined Cluster System) を使用し、並列計算等により計算コストの問題を回避した。

(4) 認識結果の可視化

自動認識システムを用いて、大量のボリュームデータから各オルガネラの認識を行った後、それらを3次元的に可視化した。これにより、観察者はオルガネラを自由視点から観察し、細胞内における位置関係等を詳細に把握することができる。また、理化学研究所においてCADの次世代システムとして開発されているVCADシステムを用いることにより、各オルガネラの表面積や体積の計測、及び4次元データのリアルタイムレンダリング (ここでは、ボリュームレンダリング) を行った。

4. 研究成果

(1) ライブセルイメージを用いたオルガネラの形態及び動態の可視化
ライブセルイメージング技術により取得した細胞の4次元ボリュームデータから、核、Golgi、mitochondria、Early endosome、Late

endosome、Lipid Droplet の計6つのオルガネラと、tubulin、actinの2つの細胞骨格系の自動認識を実現した。認識結果の一例を図1～図2に示す。また、図3は細胞内におけるLipid Dropletの時間変化を示しており、このように3次元空間内でオルガネラが動的に変化する様子が明確に示されたのは初めてである。

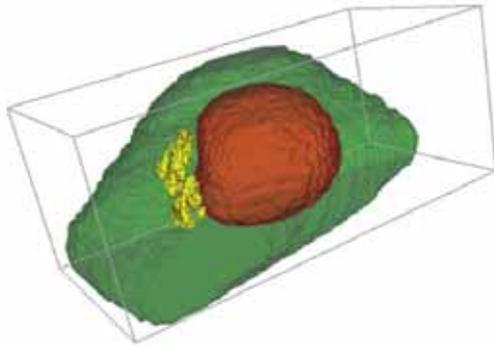


図1 HeLa細胞内のゴルジ体
(緑：細胞質、赤：核、黄色：ゴルジ体)

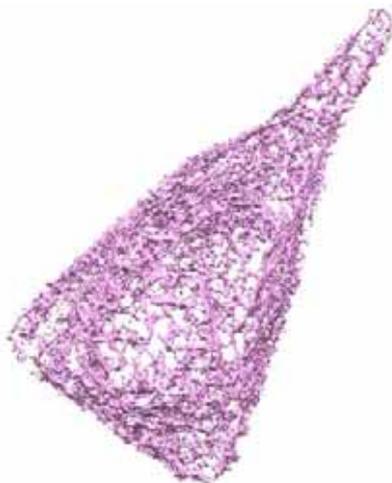


図2 HeLa細胞内のTubulin

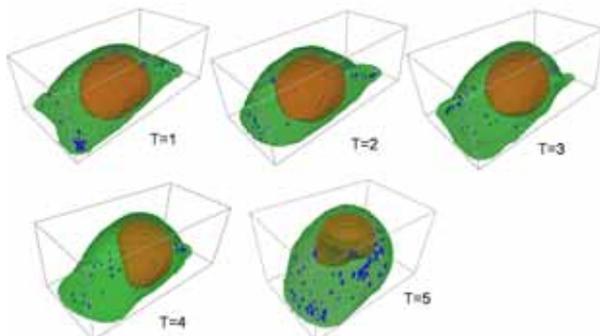


図3 細胞内で変化するLipid Droplet
(青：Lipid Droplet)

(2) 自動認識システムプロトタイプ開発
本研究では、共焦点レーザー顕微鏡で撮影した細胞の2次元～4次元のライブセル画像から、観察の対象とする細胞内オルガネラ及び細胞骨格の領域を自動認識するシステムの基本設計及びプロトタイプ開発を行った。

システムは、エンドユーザーが提示した時空間画像内の認識ターゲット(教師データ)の画像特徴を解析し、システムにあらかじめ用意された複数個の認識アルゴリズムの精度スコアを算出する。その後、トップスコアを算出したアルゴリズムを用いて大量の時空間画像からターゲット全体の形態認識を実現するものである。これによって、画像やターゲットに応じて、適宜アルゴリズムを選択できるため、精度の高い認識が可能となる。

本研究では特に、特徴間距離と輪郭形状間距離の2つの尺度の線形結合によって構成される評価関数が、エンドユーザーが示した教師データにより近い認識を実現できる可能性があることを示した。この結果は現在投稿中である。このシステムの実現により、時間や場所ごとに変化する細胞内のオルガネラや細胞骨格を、大量の画像から自動抽出することが可能となった。抽出結果を用いることにより、抽出対象を自由視点から観察することのみならず、体積や表面積の計測、他のオルガネラとの位置相関関係などの定量解析が可能となった。

なお、システムは極めて計算量が膨大となることから計算並列化の基本設計・実装も行った。現在までにLinuxクラスタマシン及び理研スーパー・コンパイド・クラスタ・システム(RSCC)によるテストランを完了している。

システムの成果物の一例として、複数枚の画像からオルガネラを自動認識した結果を図4に示す。この認識結果は、VCA Dシステムを用いて認識結果をボリュームレンダリングによる3次元表示したものである。前述のように、本システムは最も精度良い認識を実現するアルゴリズムを自動選択できるものである。この結果は、10種類の認識アルゴリズムをシステムが評価し、最適アルゴリズムと判断されたものを用いて作成したものである。各アルゴリズムの詳細は割愛するが、図5～図6及び表1に、評価例を示す。図5～図6は、各アルゴリズムが自動認識した結果とエンドユーザーが示した教師データとの類似性を、それぞれ特徴間距離、輪郭形状間距離によって評価したものである。表1はそれらの線形結合に基づく評価関数から得られた最終評価結果を示しており、システムはアルゴリズムA4が最適であると判断したことになる。

今回開発したシステムは、我々のみならず、細胞生物・分子生物学の研究者にも使いやすい

いよう設計しているため、当該研究分野のより一層の発展の一躍を担えると確信している。

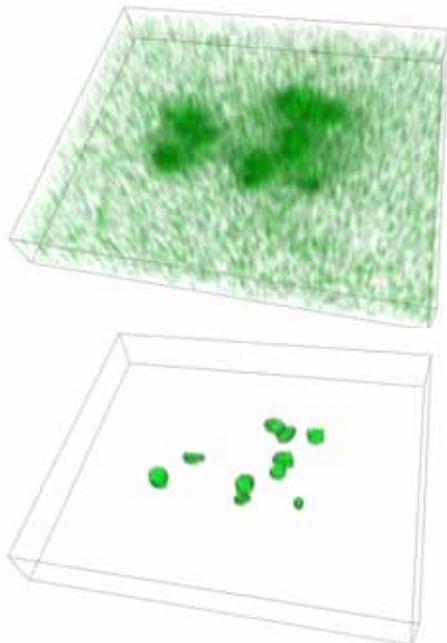


図4 Living cell images of botanical cells with marked regions of Golgi apparatus (up). The automatic segmented region as a Golgi apparatus (bottom).

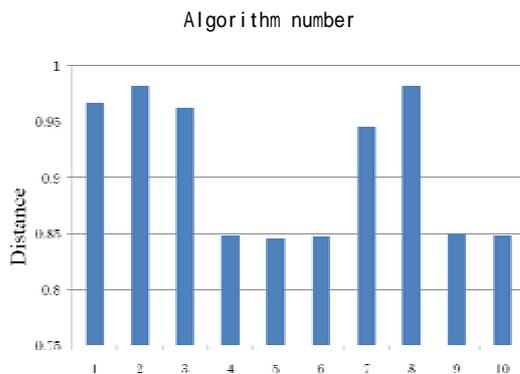


図5 特徴間距離

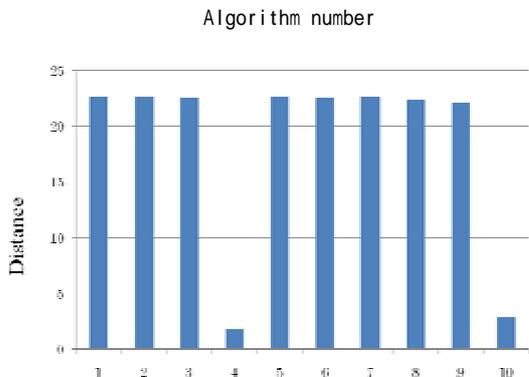


図6 輪郭形状間距離

表1 Performance ranking

Algorithm Number	Normalized Similarity	Performance Ranking
A1	1.48	8
A2	1.73	10
A3	1.41	7
A4	-3.05	1
A5	-0.52	3
A6	-0.49	5
A7	1.13	6
A8	1.70	9
A9	-0.50	4
A10	-2.91	2

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

S. Takemoto, H. Yokota, R. Himeno, T. Mishima: "Segmentation of Anatomical Structure by Using a Local Classifier Derived from Neighborhood Information", Advances in Intelligent and Soft Computing, Springer, 2009 (in press) 査読あり

〔学会発表〕(計 6 件)

竹本智子, 生物の内部構造理解・解析のための画像領域抽出, 理研シンポジウム VCADシステム研究 2007, 2007年10月, 埼玉

竹本智子, ライブセルイメージにおける領域抽出, 理研シンポジウム 生物研究基盤ツールとしてのライブセルモデリング, 2008年3月, 埼玉

Satoko Takemoto, Segmentation of Anatomical Structure by using a Local Classifier Driven from Neighborhood Information, IEEE International Conference on Human System Interactions, 2008年6月, Krakow, Poland

Satoko Takemoto, Image Segmentation for Live Cell Modeling, 1st Joint Workshop on Computational Science, 2008年7月, 埼玉.

Shin Yoshizawa(発表者), Live Cell Modeling Project, The 9th international Conference on Systems Biology, 2008年8月, Gothenburg, Sweden.

竹本智子, 領域抽出: 画像特徴と形状の評価によるオブジェクト抽出, 理研シンポジウム VCAD システム研究 2008, 2008

年 11 月，埼玉

〔産業財産権〕

出願状況（計 1 件）

名称：画像処理装置、画像処理方法およびプログラム

発明者：竹本智子、横田秀夫

権利者：独立行政法人 理化学研究所

種類：特許

番号：特願 2009-110683

出願年月日：2009 年 4 月 30 日

国内外の別：国内

6 . 研究組織

(1)研究代表者

竹本 智子 (TAKEMOTO SATOKO)

独立行政法人理化学研究所 生物基盤構築
チーム 研究員

研究者番号：00450403