

混合正規分布モデルを用いた 経時観測蛍光画像からの細胞核の検出と追跡手法

瀬尾 茂人^{1,a)} 間下 以大² 前田 栄³ 竹中 要一¹ 石井 優^{3,4,5} 松田 秀雄^{1,4}

受付日 2013年1月30日, 再受付日 2013年3月21日,
採録日 2013年4月14日

概要: 様々な分子や細胞を可視化するバイオイメーjing技術の発展は目覚ましく, 医学・生物学研究を進めるうえでますます重要性を増している. またハイスループット化も進んでいるため, 得られた画像・動画を自動的・客観的に解析するための情報処理技術の開発が喫緊の課題となっている. 本研究では, 蛍光タンパク質を用いて細胞周期の可視化を行ったイメージングデータから細胞核の自動検出と追跡を行うために, 混合正規分布モデルを用いた方法を提案し, その有効性を評価するとともに得られた結果の考察を行う.

キーワード: 蛍光顕微鏡画像, 混合正規分布モデル, トラッキング, バイオイメーjingインフォマティクス, コンピュータビジョン, 細胞周期

Segmentation and Tracking Method of Cell Nuclei for Time-lapse Fluorescent Microscopy Images Based on Gaussian Mixture Model

SHIGETO SENO^{1,a)} TOMOHIRO MASHITA² SAKAE MAEDA³ YOICHI TAKENAKA¹
MASARU ISHII^{3,4,5} HIDEO MATSUDA^{1,4}

Received: January 30, 2013, Revised: March 21, 2013,
Accepted: April 14, 2013

Abstract: Biological imaging technologies have been rapidly advancing for several years and have become an important part in the field of biology. Time-lapse imaging techniques enable the monitoring of multiple cellular functions using live cell assays. In order to process massive time-lapse images and perform quantitative analysis, automated image analysis (including cell segmentation and tracking trajectories, defining and retrieving various statistics) is required. In this study, we proposed a segmentation and tracking method of cell nuclei for time-lapse fluorescent microscopy images based on Gaussian mixture model.

Keywords: fluorescence microscopy, Gaussian mixture model, tracking, bioimage informatics, computer vision, cell cycle

¹ 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Osaka University, Suita, Osaka 565-0871, Japan

² 大阪大学サイバーメディアセンター
Cybermedia Center, Osaka University, Toyonaka, Osaka
560-0043, Japan

³ 大阪大学大学院医学系研究科
Graduate School of Medicine, Osaka University, Suita, Osaka
565-0871, Japan

⁴ 大阪大学免疫学フロンティア研究センター
WPI-Immunology Frontier Research Center, Osaka University,
Suita, Osaka 565-0871, Japan

⁵ 大阪大学大学院生命機能研究科
Graduate School of Frontier Biosciences, Osaka University,
Suita, Osaka 565-0871, Japan

a) senoo@ist.osaka-u.ac.jp

1. はじめに

バイオイメーjingとは細胞・組織または個体レベルでタンパク質等の分布・局在をとらえ, その動態を画像として解析する技術のことであり, 高解像度の顕微鏡等の光学技術や蛍光タンパクやゲノム組み換え等の遺伝子工学的技術の進展にともない, 医学・生物学研究を進めるうえで主要な要素技術となりつつある. 細胞や粒子を観察するためには, 位相差顕微鏡や微分干渉顕微鏡, また蛍光標識を施したサンプルに対して共焦点レーザー顕微鏡等が利用される. 生体内の細胞やタンパク質を観測するライブイメージングには多光子励起顕微鏡も利用される. さらにこれらの

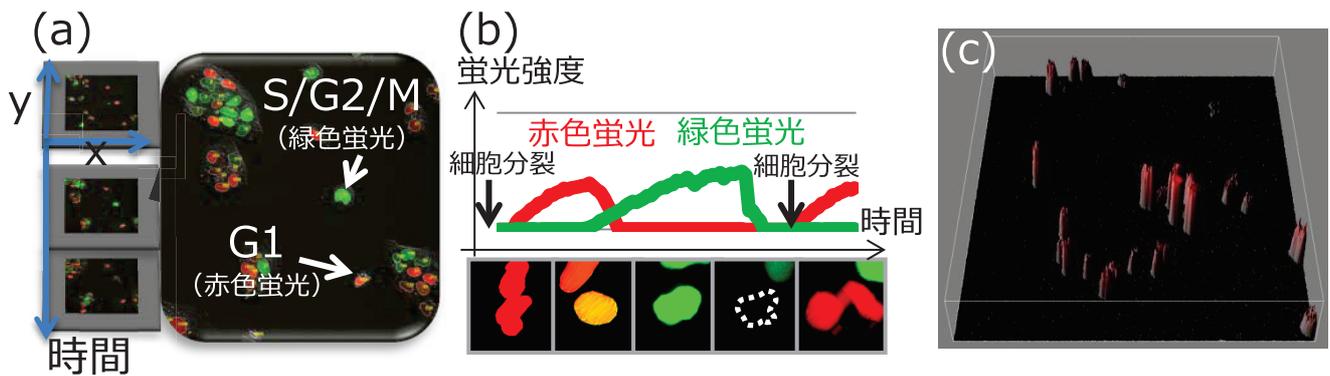


図 1 多色蛍光プローブとタイムラプスイメージングによる細胞周期の可視化

Fig. 1 Visualization of cell cycle using multi-color fluorescence and time-lapse imaging.

高機能な顕微鏡をコンピュータによって制御し、一定時間ごとに撮影することで長時間の変化を観察するタイムラプス撮影や、それを複数の視野で並行して行うことが可能となっている。このようなバイオイメージングデータの高分解像度化・ハイスループット化が進む現在、画像処理技術による細胞や粒子の自動検出・自動追跡の技術の開発は喫緊の課題となっている。

経時観測データから細胞や粒子の追跡を自動で行うためには、細胞領域と背景を区別するセグメンテーション（空間的解析）と、時間経過前後で同一のオブジェクトを対応付けるトラッキング（時間的解析）が必要になる。バイオイメージングのデータは、一般に低コントラストでノイズの大きい画像・動画になることが多い。また、あいまいな形状のオブジェクトが接触して多数存在することや、細胞の増殖や分裂、融合や死、視野内への出入りによる密度の変化等が自動での処理を困難にしている [16]。最も単純なセグメンテーションの方法は閾値処理であるが、ノイズに弱くエラーが多い [6]。テンプレートマッチング [15] は細胞の形があまり変化しない場合には有効である。分水嶺 (Watershed) 法は接触した対象を区別することができるが、過剰に分割しすぎる傾向があることが指摘されており、前処理後処理を工夫する必要がある [32]。さらに発展した方法として、たとえばレベルセット法 [8] は、動的に変化する輪郭を扱うことができるが、こちらは複数のオブジェクトを 1 つのものとして認識する傾向がある。トラッキングの方法も様々なものが提案されている。最も単純な方法は最近隣 (Nearest neighbor) 法であり、時間の前後の画像において最も距離が近いものを同一の対象であるとする方法であるが、密集状況下では精度が得られない。Mean shift 法のように対象の色ヒストグラムを特徴量とし、反復法で追跡を行う方法 [30] や、そのほかにも、より洗練された方法として、オプティカルフロー [12] やグラフベース [24] の方法、パーティクルフィルタ [29] 等が用いられる。このように様々な方法やツールが提案されているが、万能の方法はいまだなく、良い精度を得るためには、観測対象や利用

する装置に応じてそれぞれ専用に調整された個別のアルゴリズムが必要であることが示唆されている [20]。

バイオイメージングにおいては、さらに物体の形状以外のもの、細胞の状態や機能を観察するための技術も発展している。蛍光波長の異なるタンパク質を細胞の状態特異的に発現させることでその細胞の状態を可視化したり、蛍光共鳴エネルギー移動 (Fluorescence resonance energy transfer: FRET, またはフェルスタ共鳴エネルギー移動) と呼ばれる現象を利用して、特定の物質が結合したことを蛍光波長の変化として可視化 [13] したりすることが可能になっている。このような観測データにおいては、前述のバイオイメージングデータ一般の問題に加えて、より輪郭が弱く曖昧で、さらにコントラストや色相が変化するという問題に対処する必要がある。たとえば本研究の対象である Fucci (Fluorescent Ubiquitination-based Cell Cycle Indicator) [28] (図 1(a)) は、細胞周期依存的に交互に生産・分解される 2 つのタンパク質の発現を利用した蛍光プローブであり、細胞核が G1 期で赤 (オレンジ) 色、S/G2/M 期では緑色の蛍光を発するようにすることで、個々の細胞周期の進行を可視化することができる。図 1(b) は色を強調したものであるが、細胞核の色が赤から黄色、黄色から緑へと変化していることが分かる。細胞周期は細胞の増殖や分化、がん細胞の挙動等に密接に関わることが知られており [27]、様々な実験条件下で経時観察を行うことで、薬剤の影響や細胞の特性の変化を調べることができる。Fucci 細胞の追跡を手動で行うためのツールとしては ImageJ [3]、また、半手動で行うツールとして ImageJ のプラグイン LineageTracker [7] が実装されている。しかしながら様々な条件下で撮影された大量の動画から客観的・定量的に解析を行うためには、精度良く細胞の自動追跡を行う方法が求められる。

本研究で提案する方法は、Fucci を導入した細胞群のタイムラプス撮影データを対象とし、細胞核の検出 (セグメンテーション) と前の画像から次の画像間で同一の細胞を対応付ける (トラッキング) という問題を、混合正規分布

モデルのパラメータ推定の問題として帰着させる方法である。本研究の対象となるデータでは、細胞核1つ1つを位置平面 xy と蛍光強度 z の空間上の2次元正規分布に近似できるため(図1(c)), 画像から細胞核の数と大きさを求める問題を、混合正規分布の数とパラメータを推定する問題と同じと見なすことができる。さらにトラッキングについても、前のフレームと次のフレームで同一の細胞核を対応付ける問題は、EM アルゴリズムによってパラメータを推定する問題と見なせる。本研究の対象のようなデータでは、細胞核の輪郭があいまいで、さらに時間経過と細胞の状態に依存して蛍光強度の強さが大きく変化するため、蛍光強度の閾値を用いるような方法ではセグメンテーションが難しく、その結果トラッキングの精度も悪化する。提案手法では、細胞核を分布としてとらえるため蛍光強度の変化に頑健であり、またセグメンテーションとトラッキングを相互依存的に行うため高い精度を実現することができる。

以下2章では混合正規分布とEM アルゴリズムの数理モデルについて述べ、3章に混合正規分布モデルを用いた経時観測蛍光画像からの細胞核の検出と追跡の手法を示す。4章では実際に Fucci プローブを導入した細胞の経時観察データに対して提案手法を適用し精度の評価を行うとともに、実験条件の差が細胞の増殖や細胞周期の状態に与える影響を定量化するために本研究が有用であることを示す。

2. 混合正規分布モデルとEM アルゴリズム

2.1 混合正規分布モデルと画像処理

混合分布モデルは、複数の確率分布の加重和の形で表されるモデルであり、混合正規分布モデルは成分分布に正規分布を仮定する。データ構造が未知な場合における密度推定や統計的分類法(クラスタリング)のための手法として広く利用されている[1]。クラスタリングは似ているものを集め、似ていないものを分けることであるが、画像処理の分野でもセグメンテーション等に用いられてきた[11], [21]。クラスタリングによるセグメンテーションは、画像平面における各画素の特徴量を特徴空間(たとえば色空間)へ写像し、特徴空間上で分類を行ったのち、画像平面へ逆写像を行って領域を分割する。その結果として、前景と背景を分離したり、複数の対象を区別したりすることができる。 k -means 法を用いたクラスタリングによるセグメンテーションは、混合正規分布モデルを用いる方法とよく似た方法である[23], [26]。また顕微鏡の Point Spread Function (PSF; 点像分布関数)[31]としての利用や、圧縮やレンダリングの方法として画像を正規分布の重ね合わせで表現する[25]等、画像処理の分野でも混合分布モデルの応用は枚挙に暇がない。

バイオイメージングの画像処理においても、正規分布へのフィッティングを用いて物体の検出を行う方法は事例が多い。特に微小な粒子を対象としたデータに対してよく用

いられており、コロイドを2次元正規分布で近似して検出を行う方法[4]や、正規分布をPSFとして適用し蛍光タンパク質の分子を検出する方法[18]等が提案されている。本研究の検出・追跡対象である細胞核の画像は、これらの粒子よりも大きく形状もいくらか不定ではあるものの、基本的には円形・楕円形の形状であり、2次元正規分布での近似が可能である。

2.2 混合正規分布モデルとEM アルゴリズム

本節では、提案手法においても利用する一般的な混合正規分布モデルとそのパラメータを求めるためのEM アルゴリズムについて述べる。

混合分布モデルは、観測されたデータが複数の母集団分布からの実現値であるという仮定の下で、異なる複数の確率分布(混合正規分布の場合は正規分布)の混合として記述するモデルである[22]。

$$f(s|\theta) = \sum_{k=1}^g \pi_k f_k(s|\theta_k) \quad (1)$$

式(1)は観測データである変数 s に対する有限混合分布モデルであり、 g は仮定する確率分布の数、 $f_k(s|\theta_k)$ は第 k 成分の確率密度関数、 θ_k はその密度関数のパラメータである。 π_k は第 k 成分分布の混合比率であり、各分布の重みとなる($\sum_{k=1}^g \pi_k = 1$)。 f_k はすべて正規分布の密度関数を用いる。

事前情報がない場合、推定の必要なパラメータは、 g , π_k , $\theta_k = \{\mu_k, \sigma_k^2\}$ となるが、成分数 g を推定することは θ_k を推定することとは性質が異なるので、まずは、 g を固定して θ_k を推定する方法について述べる。

混合正規分布のパラメータの推定にはEM アルゴリズムを用いるのが一般的である[5]。EM 法による混合分布モデルのパラメータ推定は、E ステップとして欠測値の期待値推定、M ステップとして対数尤度の最大化という手続きを、収束条件を満たすまで反復することで達成される。このモデルの欠測値とは、各データが本来どの成分に属するかということであり、これをEステップで各成分に所属する確率(事後確率)として計算する。Mステップでは事後確率が与えられた下でのパラメータの推定値の更新を行う。

観測データを $S = \{s_1, s_2, \dots, s_n\}$ とするとき、対数尤度関数は

$$L(\Theta|S) = \sum_{i=1}^n \log \sum_{k=1}^g \pi_k f_k(s_i|\theta_k) \quad (2)$$

で与えられる($\Theta = \{\pi_1, \dots, \pi_g, \theta_1, \dots, \theta_g\}$ とし、 $\theta_k = \{\mu_k, \sigma_k^2\}$ は混合比率以外のパラメータとする)。

第 l 回目のEステップでは欠測値の事後確率(s_i の第 k 成分への所属確率)を

$$P_{ki}^{(l)} = \frac{\pi_k^{(l-1)} f_k(s_i|\theta_k^{(l-1)})}{f(s_i|\Theta^{(l-1)})} \quad (3)$$

で推定する．そして M ステップで第 k 成分分布の混合比率，平均，分散を更新する．

パラメータを推定した後のデータの分類は，事後確率

$$P_{ki} = \frac{\hat{\pi}_k f_k(s_i | \hat{\theta}_k)}{f(s_i | \hat{\Theta})} \quad (4)$$

が最大の成分分布 k に割り当てる．

成分数の推定 (g をいくつに設定するかという問題) の代表的な方法としては，統計的仮説検定と情報量基準があげられるが，本研究では情報量基準 BIC (Bayesian Information Criterion) を用いる．BIC はモデルが多くの項を持つことに対してペナルティを課するもので，成分分布の個数が必要以上になることを避ける意図がある [10]．最適な g の推定を行う場合は，任意の幅の g を与え，それぞれの g の下で EM アルゴリズムを実行しパラメータ $\hat{\Theta}_g$ と尤度を計算する．そして BIC を用いて，尤度が最も高く自由度はできるだけ低いような g の下でのモデルを最適なモデルとする．

3. 提案手法

本章では，混合正規分布モデルを用いた経時観測蛍光画像からの細胞核の検出と追跡手法を述べる．

3.1 概要

本研究では，経時観測蛍光画像からの細胞核の検出と追跡という問題を，混合正規分布モデルのパラメータ推定として定式化を行う．図 1(c) のように，本研究の対象となるデータでは，円形・楕円形の細胞核の蛍光強度を観測した画像が，位置平面 xy と蛍光強度 z の空間上において，2次元混合正規分布の密度関数に近似できる．よって細胞核の数と大きさを求める問題が，画像から混合正規分布の数とパラメータを推定する問題と同じと見なすことができ，さらにトラッキングについても，前のフレームと次のフレームで同一の細胞核を対応付ける問題は，EM アルゴリズムによってパラメータをフィッティングする問題と見なせる．ただしトラッキングの際には前後のフレームにおける細胞核の位置の移動，蛍光の強度の変化に対応する必要がある．このためにブートストラップ法と最適割当てのためのハンガリー法を導入する．画像の座標と蛍光強度を標本集団と密度 (蛍光が強い位置ほど標本が密集している) と見なし，細胞核候補の座標をランダムサンプリングしつつ各成分分布のパラメータ推定の初期値として割り当てる．割当てを行う際にはスコアの合計が最大となるようハンガリー法を用いて割当てを行う．この手順をブートストラップ法で繰り返し行うことにより，セグメンテーションとトラッキングを行う際の妥当な初期値を得ることができる．

提案手法の主となるアイデアは，以下の，

STEP.1. ブートストラップサンプリングと最適割当て

STEP.2. 混合正規分布のパラメータ推定

を繰り返すことにあるので，本節ではその概要を示す．実

際のデータとしては赤・緑の 2 色の蛍光画像を対象とするが，ここでは簡単のため蛍光は 1 種類であるとして説明を行う．

図 2 を用いて提案手法の概要を示す．図 2(a) を時刻 $t-1$ の蛍光画像とする．画像には A, B, C, D の 4 つの細胞核が観測されている．時刻 t になると細胞核 A, B, C はそれぞれ右上へ移動している．このような例において細胞核の検出と追跡の問題を考える．また図 2(c) は時刻 t の画像を位置空間 xy と蛍光強度の 3 次元に表示したものである．

STEP.1 ではまず，図 2(c) が標本集団の確率密度を示しているを見なしてランダムサンプリングを行う．蛍光の強さを頻度に近似するため，蛍光強度の大きな位置の座標がサンプリングされる確率が高い．ここで，時刻 t における細胞核の数は推定対象であり判明していないので，仮に時刻 $t-1$ の際の細胞核の数 g_{t-1} とする．すなわち図 2(d) のように，時刻 t においても 4 つの細胞核があると仮定し，4 回復元抽出を行う．このとき，4 つの分布から 1 点ずつサンプリングされることを期待する．これを，時刻 $t-1$ の混合正規分布モデルに対して，式 (4) のように事後確率を求める．次にサンプリングされたデータが時刻 $t-1$ での成分分布に由来していたかを割り当てるが，ここでは単に最大の成分分布に割り当てるのではなく，ハンガリー法 [19] を用いて全体として最適になるように割当てを行う．当然同一の分布から複数個のデータがサンプリングされることも起こるが，4 個を 1 セットとしたサンプリングと割当てを何度も行うことで，時刻 t のそれぞれの細胞核を示す分布からサンプリングされたデータは，時刻 $t-1$ においてそれぞれが所属していた成分分布へと割り当てられる割合が高くなる．図 2(e) は，4 個をサンプリング・割当てを 50 回繰り返した結果である．多少のばらつきはあるが，細胞核の移動前後で対応するように割り当てられていることが分かる．STEP.2 では，STEP.1 でサンプリングされたデータとその割当てを初期値として，EM アルゴリズムによる混合正規分布のパラメータ推定を行う．図 2(e) の時点では多少ばらついていた割当てが，EM アルゴリズムを反復するうちに逐次割り当て直され，EM アルゴリズムが収束したときには， $t-1$ 時点の成分分布から t 時点の成分分布へとパラメータの更新も完了している (図 2(f))．これにより，各成分分布の標準偏差内を細胞核領域と近似すると，細胞核の検出と追跡が同時に行われたことになる．

3.2 アルゴリズムの詳細

3.2.1 ブートストラップサンプリングと最適割当て

ブートストラップサンプリングと最適割当てについて，アルゴリズムを示す．ここでの入力は，時刻 $t-1$ における混合正規分布モデルのパラメータ Θ_{t-1} と成分分布数 g_{t-1} ，時刻 t における画像 I_t である．ここで I_t は行列であり，行列

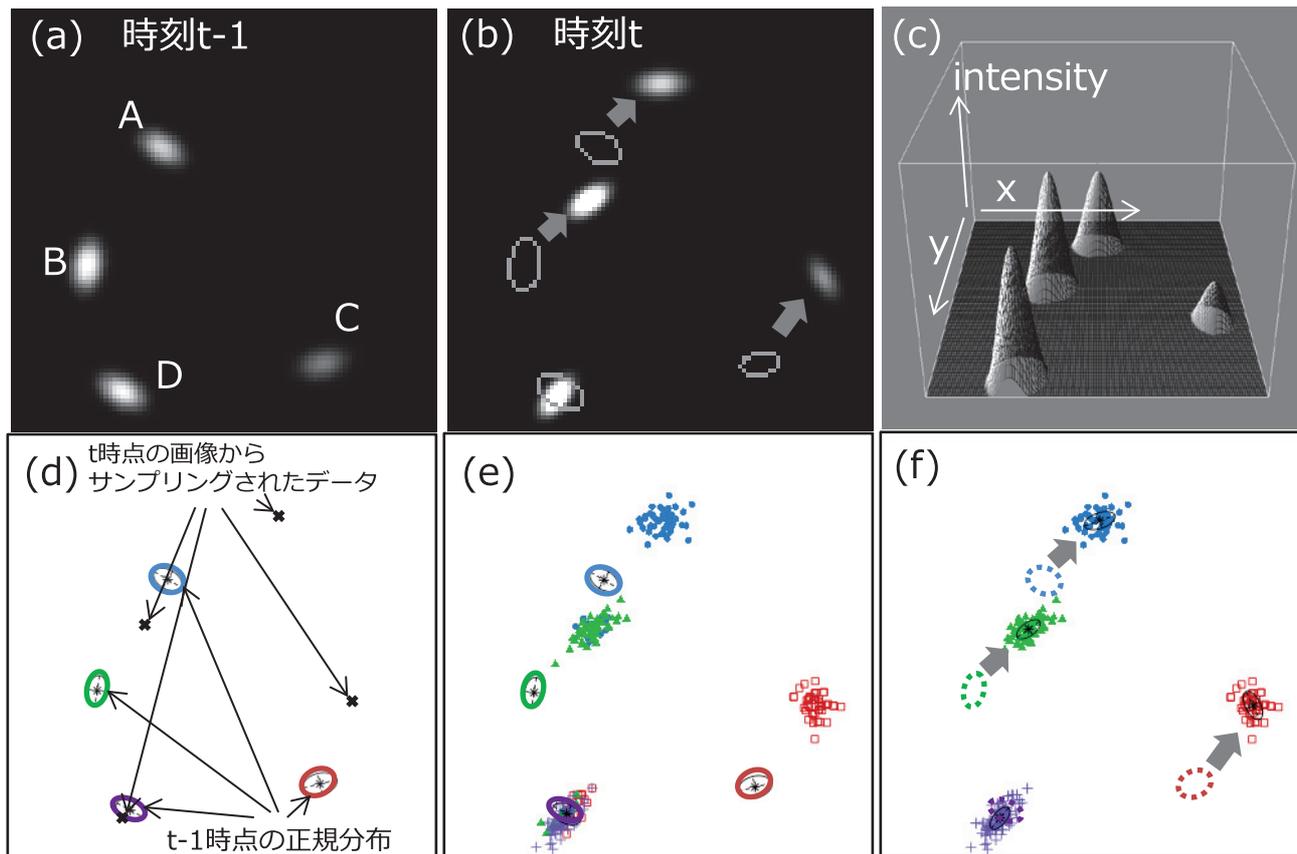


図 2 提案手法の概要

Fig. 2 Outline of proposed method.

の各要素 i_{xy} は該当するピクセルにおける蛍光強度を示す。ただし確率密度と近似するために $\sum_x \sum_y i_{xy} = 1$ を満たすように正規化を行う。出力は、サンプリングされたデータ $S = \{\{s_{11}, s_{12}, \dots, s_{1g_{t-1}}\}, \dots, \{s_{b1}, s_{b2}, \dots, s_{bg_{t-1}}\}\}$ と、その割当てを示すベクトル A である。ただし b はブートストラップの回数を示す。

I_t から、まずは j 回 ($j = g_{t-1}$) サンプリングして得られたデータを $\{s_{11}, s_{12}, \dots, s_{1j}\}$, k ($k = 1, \dots, g_{t-1}$) を時刻 $t-1$ の各成分分布の番号とすると、それぞれのデータの時刻 $t-1$ の各成分分布への所属確率は、

$$P_{kj} = \frac{\pi_k f_k(s_{1j} | \theta_k)}{f(s_{1j} | \Theta_{t-1})} \quad (5)$$

となる。この確率は k 行 j 列 ($k = j = g_{t-1}$) の正方行列となる。ここで、サンプリングしたデータ s_{1j} と成分分布 f_k が 1 対 1 に対応するように割当てを行う。このときハンガリー法 [19] を用いて、割り当てられた組合せの確率の積が最大となるように選択し、各 s_{1j} に対応する割当て a_{1j} を得る。単に最大確率へと割当てを行うと、図 2 (b) のように細胞核がかつて別の細胞核があった位置へ移動した場合うまく割り当てられないためである。一般のトラッキングの際にも全体が最適になるように割り振る方法は nearest neighbor を用いるよりロバストであるとされる [2]。

この g_{t-1} 個のサンプリングと割当てをセットとして、 b 回ブートストラップを行うことで、ノイズに対してもロバストで、また t 時点での成分分布数が不明なまま割当てを行うことができる。 A は S と同じ長さのベクトルであり、 S のデータそれぞれが時刻 $t-1$ のどの成分分布由来であるかの仮の割当てを示す。時刻 $t-1$ と時刻 t で細胞核の数が異なる場合は追加の処理が必要となるが、その際の手順は後述する。

3.2.2 混合正規分布のパラメータ推定

混合正規分布のパラメータ推定についてアルゴリズムを示す。ここでの入力、STEP.1 の出力であるサンプリングされたデータ S と、その割当て A であり、出力は時刻 t における混合正規分布のパラメータ $\hat{\Theta}_t$ である。 $\Theta = \{\pi_1, \dots, \pi_g, \theta_1, \dots, \theta_g\}$ であり、混合比 π_k は細胞核の蛍光強度、 $\theta_k = \{\mu_k, \sigma_k^2\}$ の μ_k は細胞核の中心座標、 σ_k は細胞核の面積と同義となる。初期値からパラメータ推定を行う手順はごく一般的であり、2 章で述べたとおりである。

3.3 その他の処理

前節に提案手法の主となる部分について説明を行った。ここでは対象とするデータに固有な処理等、実際のデータに適用するための比較的細かい処理についての記述を行う。

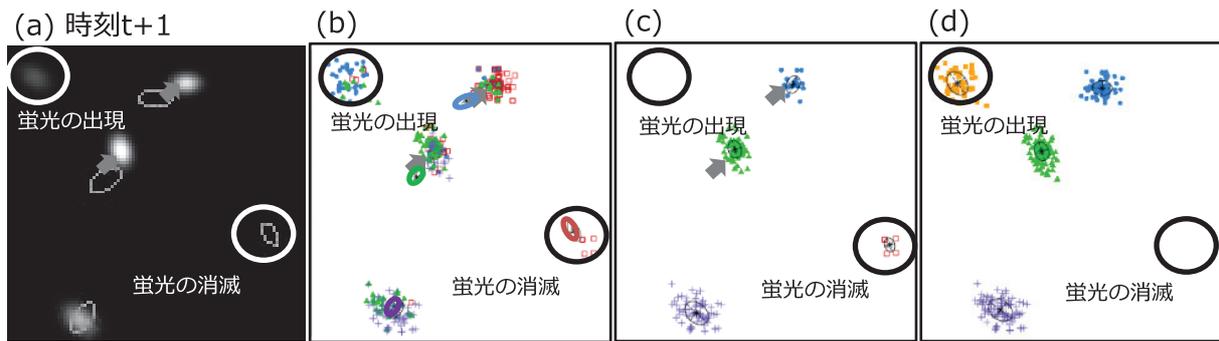


図 3 細胞核の出現と消滅

Fig. 3 Appearance and disappearance of cell nuclei.

3.3.1 初期値について

提案手法は、時刻 $t-1$ の混合正規分布のパラメータを用いて時刻 t のパラメータ推定を行うものであり、時刻 1 の画像については例外的な処理を行う必要がある。本研究では k -means 法と混合正規分布モデルを組み合わせた方法により初期値を得る。すなわち k -means 法の k を、混合正規分布モデルの成分分布数 g と同値と見なし、 k -means 法の結果のクラスターを EM アルゴリズムの初期割当てとして解く方法である。手順としては、細胞数の上限 g_{limit} を与え、 $k (= g)$ を 1 から g_{limit} まで逐次増加させながら、 k -means 法、そして EM アルゴリズムによる混合正規分布のパラメータと尤度の推定を行い、最も良い BIC になる g のときのパラメータを時刻 1 の解とする。

3.3.2 細胞数の変化について

細胞を追跡する問題において細胞数の変化は一般的な問題である。細胞分裂による増殖や細胞死による減少のほか、視野内への出入りも頻繁に起こりうる。本研究の対象である Fucci の観測データにおいて特徴的な現象は、図 1(b) に示すように細胞分裂時には、赤・緑のどちらの蛍光も消失するという点である。蛍光が消失している間は背景と区別がつかず、細胞分裂に成功すると一見何も無いところから再び蛍光を発生し始める（細胞分裂に失敗して細胞死を起こすと蛍光は消失したままである）。このような細胞の出現と消滅に対応するために、STEP.1.a と STEP.2.b の手続きを導入する。図 3 を用いて処理の概要を述べる。図 3(a) は図 2 の次の時刻 $t+1$ の画像である。図 3(a) 左上には新たに蛍光を発生させた細胞核が、右下には蛍光が消失しつつある細胞核がある。このような場合に STEP.1 を行うと、図 3(b) のようになる。

STEP.1.a 割当てされた際の信頼度が高いものと低いものを区別する。

STEP.1 の出力である S と A について、式 (5) の確率が閾値 T_p 以上のデータ S_h と割当て A_h 、閾値以下のデータ S_l と割当て A_l に分離する。 S_l は前時点の細胞核から極端に遠い場所にあるデータであり、割当てが間違っていると見なす。信頼できるデータ S_h と割当て A_h を用いて

STEP.2 を実行し、仮の混合正規分布のパラメータを得る (図 3(c))。

STEP.2.b 成分分布の追加と削除を行う。

S_h と割当て A_h を用いて得られた成分分布において、所属するデータが極端に少ない分布を削除する (図 3(d) 右下)。その判定には閾値 T_d と成分分布の混合比率とを用い、 $T_d > (\pi_k/g)$ となる成分分布が消滅したとする。またデータ S_l に対して、時刻 1 の画像を処理する際と同様に、 k -means 法・EM アルゴリズムによる混合正規分布の最尤推定で最も良い BIC となるモデルを計算し、その結果を成分分布として追加する (図 3(d) 左上)。

以上の手続きを追加することで、細胞核の出現・消滅に対応することができる。

3.3.3 2色の蛍光画像による検出と追跡について

本研究の対象であるデータは赤色蛍光と緑色蛍光の2色になっている。これらを考慮する方法としては、合成した画像をグレースケールとして扱う方法もあるが、今回は各色で個別に検出・追跡を行ったのちに結果を統合する方法を用いた。緑色で検出された細胞核の中心が赤色で検出された細胞核の標準偏差領域内にあること、またその逆の両方を同時に満たすときのみそれらを同一の細胞核とした。

4. 実験

肝臓癌の細胞に Fucci ベクターを導入し、30分間隔で約3日 (計150フレーム) タイムラプス撮影を行ったデータに対して本手法を適用した。画像は赤色蛍光と緑色蛍光を撮影したそれぞれ $1,024 \times 1,024$ ピクセルの12bit (4,096階調) として得られる。このうちコントロール (薬剤の投与なし) である視野1から10と、ある種の抗癌剤を投与した視野11から20の20視野分を用いる。

提案手法の実装には R 言語を用いた。並列化処理等による高速化は行っていない。Core i7-870 (2.93 GHz) の CPU を搭載した計算機 (メモリ: 16 GB, OS: Ubuntu) によって計算を行ったところ、20視野分の平均が233.62分 (CPU 時間)、標準偏差が27.12分であった。ただし、提案手法中 STEP.1 で行うブートストラップサンプリングの

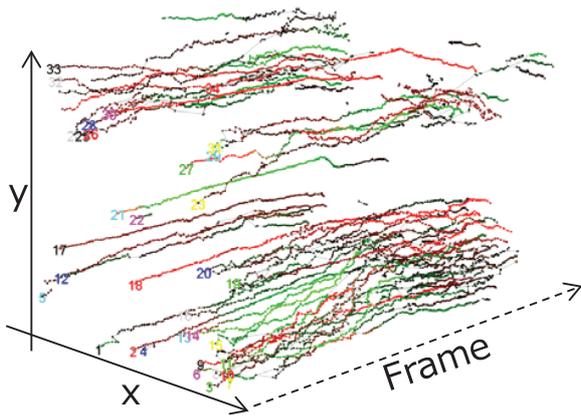


図 4 追跡の結果

Fig. 4 Tracking result.

回数 b は 500, STEP.1.a で用いる信頼度の閾値 $\log T_P$ は -20 , STEP.2.b で用いる消滅の判定の閾値 T_d は 0.05 とした. また時刻 1 の画像を処理する際の成分分布数の上限 g_{limit} は, 今回のデータにおいては十分な値である 100 に設定した.

図 4 は視野 1 に提案手法を適用した結果であり, 各細胞核の各フレームにおける細胞周期の位相を色で表示するとともに, 移動の軌跡を xy と t の時空間の 3 次元に表示したものである. このように細胞を個別に追跡することで, 各細胞が細胞周期のどの位相にどれくらいの時間滞在するかを調べることができる.

4.1 細胞核の検出精度

まず細胞核の検出精度を評価するために, 視野 1 と視野 11 に対して提案手法と既存のツールである LineageTracker [7] を適用した結果を比較した. LineageTracker は, Fucci のようなデータを追跡することを目的とした方法であり, 画像処理支援環境 ImageJ [3] のプラグインとして実装されている. このツールはセグメンテーションに Gaussian Maxima と Seeded growth 法 [9] を使用しており, ここでは Noise Tolerance のパラメータを, 高い値と低い値に設定した 2 種類を比較対象とする. 正解の作成は ImageJ 環境を用いて人手で行い, 細胞核の輪郭をなぞった結果を検出の正解とした. コントロールである視野 1 では 36 個の細胞が 3 日後には 78 個に, 抗癌剤を投与した視野 11 では 53 個の細胞が 72 個になる. 150 フレーム中の細胞核の延べの数は, 視野 1 では 9,660 個, 視野 11 では 10,210 個であった.

検出精度の評価を行うために, 正解に対して図 5 のような基準で自動検出の結果を分類した. 真の細胞核領域と検出した細胞核領域が相互に十分重なっているとき, これを TP (真陽性: True Positive) とする. 真の領域に検出した領域が十分に重ならないとき, 真の領域を FN (偽陰性: False Negative), 誤って検出した領域を FP (偽陽性: False Positive) とする. TP と FN の数の和は, 真の領域

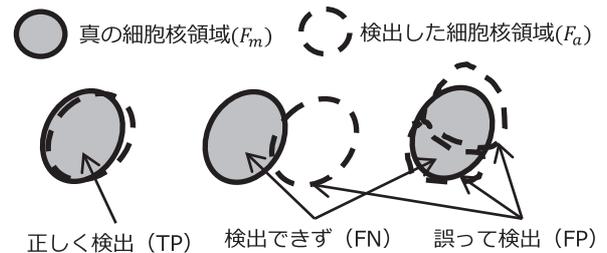


図 5 細胞核検出の基準

Fig. 5 Assessment criterion.

表 1 検出精度

Table 1 Detection accuracy.

視野 1	TP	FP	FN	Prec.	Rec.	F 値
EM	8,271	1,103	1,389	0.88	0.86	0.87
LT(H)	6,781	1,740	2,879	0.80	0.70	0.76
LT(L)	6,728	23,161	2,932	0.23	0.69	0.34
視野 11	TP	FP	FN	Prec.	Rec.	F 値
EM	8,712	1,503	1,498	0.85	0.85	0.85
LT(H)	7,708	2,104	2,502	0.79	0.75	0.77
LT(L)	7,697	19,323	2,513	0.28	0.75	0.41

の数とつねに等しい.

また真の領域と自動検出の領域の重なり判定には Jaccard 係数を用いる. 各細胞核の真の領域を F_m , 自動検出の結果の領域を F_a としたとき, その重なりを示す指標 J を,

$$J = \frac{F_m \cap F_a}{F_m \cup F_a} \quad (6)$$

とする. すなわち F_m と F_a のピクセルの積集合と和集合の割合であり, お互いが同じ大きさで重なっている部分が大きければ 1 に近づく. 重なりが小さかったり, 領域の大きさが極端に異なっていたりする場合は 0 に近づく. ここでは J が 0.8 以上になる領域を十分重なっているとした.

視野 1 と視野 11 のそれぞれ 150 フレームずつの赤色蛍光と緑色蛍光の画像に対して提案手法 (EM) と従来手法, LineageTracker (LT) を適用し, 上述の基準により分類した結果を表 1 に示す.

Precision (適合率) と Recall (再現率) は,

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \quad (7)$$

で与えられる評価基準であり, それぞれ検出したもののうちの正解の割合と, 真の領域のうち検出できたものの割合を示す. F 値はこれらの調和平均であり, 総合的な精度の指標とする. 提案手法は, 従来手法のどちらの閾値設定と比較しても良い精度を得られている. 一般に適合率と再現率はトレードオフの関係になるが, 真の領域が複数に分割されて検出された場合にはどちらも Jaccard 係数が閾値以下になることが多く, FP と FN として計上されるため, 今回は単純なトレードオフとはならない. LineageTracker の Noise Tolerance のパラメータを低く設定した場合, 細胞

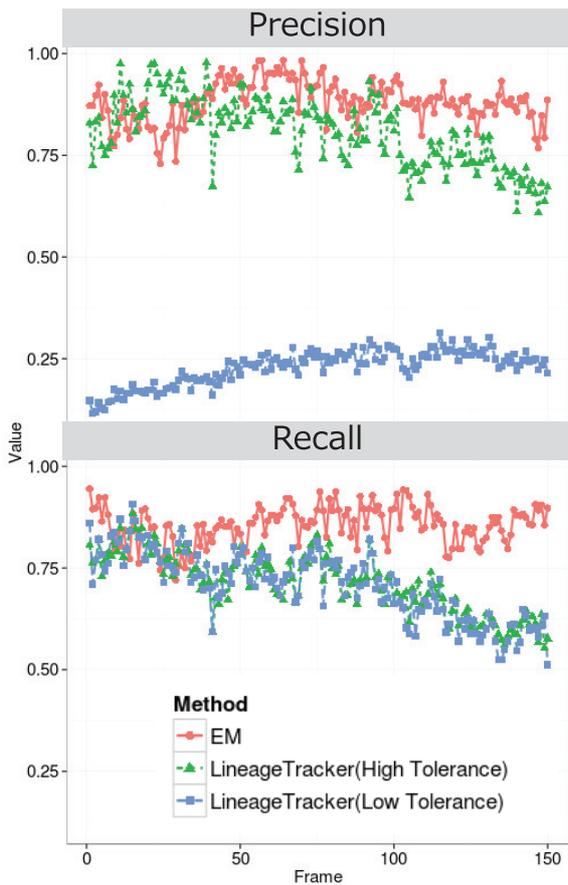


図 6 視野 1 での精度の推移
 Fig. 6 Accuracy of field of view #1.

核領域が大量に検出されるが、そのほとんどは真の領域が複数に分割されて検出されており、これが FP を増やしても TP が増えない理由である。

次に視野 1 の各フレームごとに Precision と Recall を計算した結果を図 6 に示す。上段が Precision の推移、下段が Recall の推移である。視野 1 は 1 フレーム目では 36 個の細胞が 150 フレーム目には 78 個に増殖するデータである。従来手法では時間の経過とともに Precision と Recall がともに下がっている。これは従来法の、輝度の極大値を探索して細胞核の中心を決定し、その後細胞核領域を少しずつ広げながら決定するという方法に起因する。分裂を行う細胞核に対して、分裂しないままの細胞では核が肥大化しており、細胞核の面積が大きくなる傾向がある。たとえば視野 1 の 1 フレーム目では、細胞核の面積の中央値は 1,214 ピクセル、最大値は 2,160 ピクセルであるが、150 フレーム目では最大値は 4,355 ピクセルとなる。このため 1 つの細胞核内に複数の極大値を特定してしまい、過剰分割が起こりやすくなってしまふ。また逆に、蛍光強度の高い細胞核と低い細胞核が接触しているとき、蛍光強度の高い方のみの極大値が採択されて 1 つの細胞核と判定してしまうケースも多く見られた。細胞が増殖すると接触する頻度も高くなり、これも精度の低下を招く要因である。それに

表 2 追跡精度

Table 2 Tracking accuracy.

視野 1	正解	不正解	正解率
EM	7,862	1,720	0.82
LT(H)	6,128	3,454	0.64
LT(L)	5,722	3,860	0.60
視野 11	正解	不正解	正解率
EM	8,127	1,896	0.81
LT(H)	7,128	2,894	0.71
LT(L)	6,587	3,436	0.66

対して提案手法では、前フレームの情報を用いて次フレームのパラメータ推定を行うため細胞核の大きさや輝度の変化に頑健であり、精度の低下を防ぐことができている。

4.2 追跡の精度

さらに追跡精度の評価を行うために、人手で細胞核の追跡を行った結果と自動追跡の結果を比較した。人手による追跡と自動追跡の結果が、前のフレームと後のフレームの両方において同一の領域を指しているときのみを正解とし、異なる領域を指す場合や対応する領域が検出されていない場合等はすべて不正解とする。自動追跡において、前後のフレーム間がともに誤検出した領域どうしである場合はどちらにも計上しない。表 2 にその結果を示す。

追跡の結果についても提案手法の方が良い精度が得られた。従来手法では、肥大化した細胞核を過剰分割してしまうことや、細胞核どうしが接触した際に蛍光強度の弱い方が検出できないことがあり、これによるセグメンテーション精度の低下が追跡の失敗の要因となっていた。提案手法では、これらの場合も比較的正確に追跡することができていたが、接触した状態のまま細胞核の出現と消滅が起こるときにうまく追跡ができなくなる頻度が高かった。

4.3 自動検出と追跡の有用性

次に、自動検出・追跡の有用性を示すために、コントロールである視野 1 から 10 と、ある種の抗癌剤を投与した視野 11 から 20 の検出・追跡結果の集計を行った。

図 7 はそれぞれの実験条件下での細胞の増殖率を示したもので、左側がコントロール（投薬なし）、右側が抗癌剤を投与した際のものである。増殖率（Growth rate）は、各視野の各フレームでの細胞核の数を、1 フレーム目の細胞核の数で割ることで計算した。折れ線グラフにより各視野での個別の増殖率の推移を示し、各フレームごとの分布を箱ひげ図で示している。この図より、各視野の増殖率にはばらつきがあるものの、全体的な傾向としては抗癌剤を投与すると増殖率が抑制されていることが分かる。

また、図 4 で示したように、各視野の細胞を個別に追跡することで、各細胞が細胞周期のどの位相にどれくらいの時間滞在するかを調べることができる。抗癌剤の投与の有

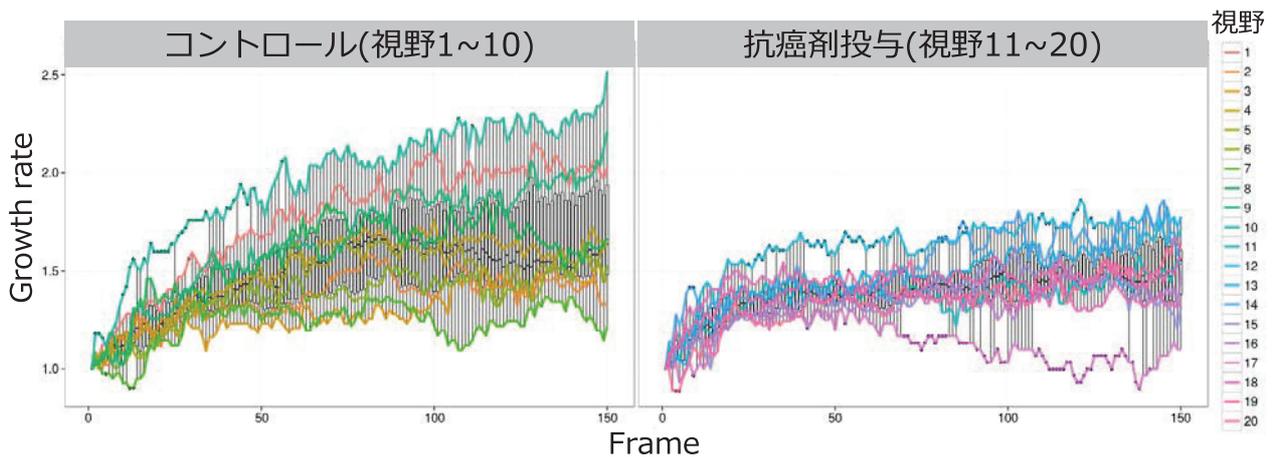


図 7 細胞の増殖率

Fig. 7 Difference of growth rate between two conditions.

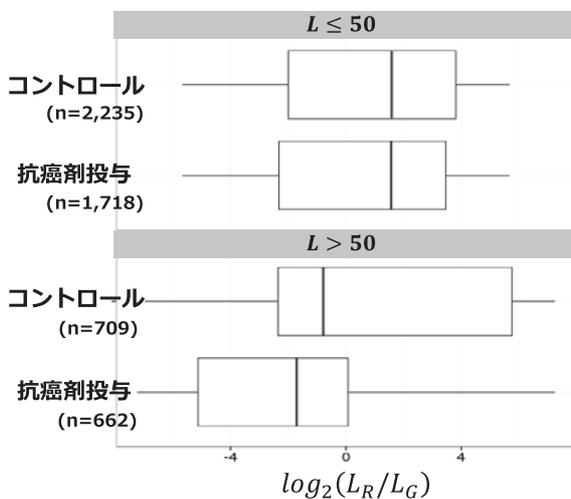


図 8 細胞周期の滞在時間比率

Fig. 8 Difference of cell cycle states between two conditions.

無による細胞周期の滞在時間比率を、各 10 視野分ずつ集計した結果が図 8 である。各細胞について、蛍光が観測できる期間である分裂後から分裂前まで（追跡を開始した時点から追跡を終了した時点まで）のフレーム数を L とし、赤色蛍光の方が強かった数を L_R 、緑色蛍光の方が強かった数を L_G とする。 $L = L_R + L_G$ である。この比の対数を細胞周期の滞在時間比率として、 $\log_2(L_R + \delta / L_G + \delta)$ とする。ただし、 δ はゼロ除算等を避けるための小さな定数 (= 1) とする。この値が正の方向に大きいほどその細胞が赤 (G1 期) にとどまっていた時間が長く、負の方向に大きいほど緑 (S/G2/M 期) にとどまっていた時間が長かったことを示す。図 8 は各細胞の追跡できた時間 L に応じて集計したものであり、図上は短期間しか追跡できなかったもの ($L \leq 50$)、図下は長期間にわたって追跡できたもの ($L > 50$) である。短期間しか追跡できなかった原因としては、追跡の失敗のほか、細胞の分裂から分裂までが観測時間内 (150 フレーム) に収まらなかった場合や細

胞核の視野内外への出入りがある。ちなみに今回のデータにおいて、細胞が正常に分裂する際に要する時間は約 2 日 (= 96 フレーム) である。図 8 から、追跡が短期間であった場合の滞在時間比率には違いがないが、長期間追跡ができた場合には、コントロールに比べ抗癌剤を投与した際に緑 (S/G2/M 期) の滞在時間比率が高い細胞が多いことが分かる。投与した抗癌剤は S 期における DNA 合成を阻害して癌細胞の増殖を抑制する作用を持つことが知られているが、このように細胞の追跡を行い結果を集計することで、実際に S 期から先の状態に進めない細胞が増加していることが分かる。

本節では細胞の増殖率と細胞周期の時間比率の変化を示したが、細胞の検出と追跡を行えば、その結果から様々な統計量を得ることができ、薬剤の効果等をより定量的に評価することが可能になる。ただし、図 7 で示したように、実験条件が同じ場合においても視野ごとによりばらつきが存在するため、実験条件に真に有意な差があるかどうかを判断するためには、大量のデータを処理することが非常に重要である。しかしながら、すべてのデータに対して人手で検出や追跡を行うのはほぼ不可能である。実際に、今回のデータにおいても 1 視野分が 150 枚の経時観測画像であり、150 枚には延べ 10,000 前後の細胞が写っているため、人手で細胞の数を数え追跡を行うには 1 視野分でさえ多大な時間と労力を要している。計測装置もコンピュータ制御可能になった現在、観察する視野数を増やしデータを大量に得ること自体は比較的容易になりつつあるため、その大量のデータを客観的に解析できる自動検出・追跡の方法は有用性が高いといえる。

5. おわりに

本論文では、多色蛍光イメージングによる経時観測データから細胞核の自動検出と追跡を行った事例を紹介した。混合正規分布モデルと EM アルゴリズムを用いる方法によ

り検出と追跡を行い、またその際にブートストラップ法によるサンプリングを繰り返し行うこと、ハンガリー法による最適割当てを行うことで、ノイズの影響を軽減し、細胞核が大きく移動したときや密集した状況下での検出・追跡における精度の低下を防ぐことができた。対象の形状だけではなく、状態や機能までも可視化するイメージング技術はこれからさらに発展し、かつハイスループット化されることが予想される。現在細胞核という、正規分布で近似できる対象に限定されるとはいえ、自動検出と追跡を精度良く行う本研究の手法は有用であると考えられる。

本論文では蛍光画像からの細胞核の検出と追跡という問題を、混合正規分布のパラメータ推定の問題として定式化を行い、その妥当性と潜在的な有効性を示すことを主眼とした。より実用的に高い精度を目指すためには、たとえば、各色蛍光での検出・追跡結果を統合する際に、全体が最適となるように対応付ける [2] ことで、さらなる精度の向上が期待できる。また前後のフレームにおける細胞核の数の変化に関して、本論文では k-means 法と BIC を利用したが、BIC よりも GIC (一般化情報量基準 [17]) や EIC (Extend Information Criteria [14]) を応用することで、現実の問題の解と数理モデルの最適値とを、より一致させることができる可能性があると考えられる。

謝辞 本研究は JSPS 科研費 24651229 の助成を受けたものです。

参考文献

- [1] Baudry, J.-P., Raftery, A.E., Celeux, G., et al.: Combining Mixture Components for Clustering, *Journal of Computational and Graphical Statistics*, Vol.9, No.2, pp.332–353 (2010).
- [2] Bise, R., Yin, Z. and Kanade, T.: Reliable Cell Tracking by Global Data Association, *2011 IEEE International Symposium on Biomedical Imaging From Nano to Macro*, pp.1004–1010, IEEE (2011).
- [3] Collins, T.J.: ImageJ for Microscopy, *Biotechniques*, Vol.43, No.1 Suppl, pp.25–30 (2007).
- [4] Crocker, J.: Methods of Digital Video Microscopy for Colloidal Studies, *Journal of Colloid and Interface Science*, Vol.179, No.1, pp.298–310 (1996).
- [5] Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society Series B Methodological*, Vol.39, No.1, pp.1–38 (1977).
- [6] Dima, A.A., Elliott, J.T., Filliben, J.J., et al.: Comparison of Segmentation Algorithms for Fluorescence Microscopy Images of Cells, *Cytometry Part A the Journal of the International Society for Analytical Cytology*, Vol.79, No.7, pp.545–559 (2011).
- [7] Downey, M.J., Jeziorska, D.M., Ott, S., et al.: Extracting Fluorescent Reporter Time Courses of Cell Lineages from High-throughput Microscopy at Low Temporal Resolution, *PLoS ONE*, Vol.6, No.12, p.e27886 (2011).
- [8] Dzyubachyk, O., Niessen, W. and Meijering, E.: Advanced Level-set Based Multiple-cell Segmentation and Tracking in Time-lapse Fluorescence Microscopy Images, *Biomedical Imaging: From Nano to Macro*, pp.185–188, IEEE (2008).
- [9] Fenistein, D., Lenseigne, B., Christophe, T., et al.: A fast, Fully Automated Cell Segmentation Algorithm for High-throughput and High-content Screening, *Cytometry A*, Vol.73, No.10, pp.958–964 (2008).
- [10] Fraley, C. and Raftery, A.E.: Model-Based Clustering, Discriminant Analysis, and Density Estimation, *Journal of the American Statistical Association*, Vol.97, No.458, pp.611–631 (2002).
- [11] Greenspan, H., Ruf, A. and Goldberger, J.: Constrained Gaussian Mixture Model Framework for Automatic Segmentation of MR Brain Images, *IEEE Trans. Medical Imaging*, Vol.25, No.9, pp.1233–1245 (2006).
- [12] Hand, A.J., Sun, T., Barber, D.C., et al.: Automated Tracking of Migrating Cells in Phase-contrast Video Microscopy Sequences Using Image Registration, *Journal of Microscopy*, Vol.234, No.1, pp.62–79 (2009).
- [13] Imamura, H., Huynh Nhat, K.P., Togawa, H., et al.: Visualization of ATP Levels Inside Single Living Cells with Fluorescence Resonance Energy Transfer-based Genetically Encoded Indicators, *Proc. National Academy of Sciences of the United States of America*, Vol.106, No.37, pp.15651–15656 (2009).
- [14] Ishiguro, M., Sakamoto, Y. and Kitagawa, G.: Bootstrapping Log Likelihood and EIC, an Extension of AIC, *Annals of the Institute of Statistical Mathematics*, Vol.49, No.3, pp.411–434 (1997).
- [15] Kachouie, N.N., Fieguth, P., Ramunas, J. and Jervis, E.: Probabilistic Model-Based Cell Tracking, *International Journal of Biomedical Imaging*, Vol.2006, pp.1–10 (2006).
- [16] Kanade, T., Yin, Z., Bise, R., et al.: Cell Image Analysis: Algorithms, System and Applications, *2011 IEEE WACV*, pp.374–381, IEEE (2011).
- [17] Konishi, S. and Kitagawa, G.: Generalised Information Criteria in Model Selection, *Biometrika*, Vol.83, No.4, pp.875–890 (1996).
- [18] Kubitscheck, U., Kuckmann, O., Kues, T. and Peters, R.: Imaging and Tracking of Single GFP Molecules in Solution, *Biophysical Journal*, Vol.78, No.4, pp.2170–2179 (2000).
- [19] Kuhn, H.W.: The Hungarian Method for the Assignment Problem, *Naval Research Logistics Quarterly*, Vol.2, No.1-2, pp.83–97 (1955).
- [20] Meijering, E., Dzyubachyk, O., Smal, I. and Van Cappellen, W.A.: Tracking in Cell and Developmental Biology, *Seminars in Cell Developmental Biology*, Vol.20, No.8, pp.894–902 (2009).
- [21] Nacereddine, N., Tabbone, S., Ziou, D. and Hamami, L.: Asymmetric Generalized Gaussian Mixture Models and EM Algorithm for Image Segmentation, *20th International Conference on Pattern Recognition ICPR 2010*, Vol.6, No.2, pp.4557–4560 (2010).
- [22] Nakamura, N.: *Methods for Multidimensional Data Analysis*, Kyoritsu Shuppan Co., Ltd. (2009).
- [23] Ng, H., Ong, S., Foong, K., et al.: Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm, *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, Vol.71, No.2, pp.61–65 (2006).
- [24] Padfield, D., Rittscher, J. and Roysam, B.: Coupled Minimum-cost Flow Cell Tracking for High-throughput Quantitative Analysis, *Medical Image Analysis*, Vol.15, No.4, pp.650–668 (2011).

- [25] Papas, M., Jarosz, W., Jakob, W., et al.: Goal-based Caustics, *Computer Graphics Forum*, pp.503-511, Wiley Online Library (2011).
- [26] Ray, S. and Turi, R.H.: Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation, *Proc. 4th International Conference on Advances in Pattern Recognition and Digital Techniques*, pp.137-143, Narosa Publishing House (1999).
- [27] Sakaue-Sawano, A., Kobayashi, T., Ohtawa, K. and Miyawaki, A.: Drug-induced Cell Cycle Modulation Leading to Cell-cycle Arrest, Nuclear Mis-segregation, or Endoreplication., *BMC Cell Biology*, Vol.12, No.1, p.2 (2011).
- [28] Sakaue-Sawano, A., Kurokawa, H., Morimura, T., et al.: Visualizing Spatiotemporal Dynamics of Multicellular Cell-cycle Progression, *Cell*, Vol.132, No.3, pp.487-498 (2008).
- [29] Shen, H., Nelson, G., Kennedy, S., et al.: Automatic Tracking of Biological Cells and Compartments Using Particle Filters and Active Contours, *Chemometrics and Intelligent Laboratory Systems*, Vol.82, No.1-2, pp.276-282 (2006).
- [30] Yang, X., Li, H. and Zhou, X.: Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy (2006).
- [31] Zhang, B., Zerubia, J. and Olivo-Marin, J.-C.: Gaussian Approximations of Fluorescence Microscope Point-spread Function Models, *Applied Optics*, Vol.46, No.10, pp.1819-1829 (2007).
- [32] Zhou, X., Li, F., Jun, Y. and Wong, S.T.C.: A Novel Cell Segmentation Method and Cell Phase Identification Using Markov Model, *IEEE Trans. Information Technology in Biomedicine a Publication of the IEEE Engineering in Medicine and Biology Society*, Vol.13, No.2, pp.152-157 (2009).



瀬尾 茂人 (正会員)

大阪大学大学院情報科学研究科助教。2006年大阪大学大学院情報科学研究科修了。博士(情報科学)。同年同研究科助手。2007年より現職。JSBi, MBSJ, IEEE 各会員。



間下 以大 (正会員)

大阪大学サイバーメディアセンター助教。2006年3月大阪大学大学院基礎工学研究科システム科学分野博士後期課程修了。博士(工学)。2008年4月より現職。IEEE, 電子情報通信学会, ヒューマンインタフェース学会, パー

チャルリアリティ学会各会員。



前田 栄

大阪大学大学院医学系研究科外科学講座消化器外科学大学院生。2004年群馬大学医学部医学科卒業。初期研修, 外科専攻医を経て, 2011年大阪大学大学院入学。現在, 大阪大学大学院医学系研究科免疫細胞生物学に在籍し, イメージングによるがん細胞の可視化に取り組んでいる。



竹中 要一 (正会員)

大阪大学大学院情報科学研究科准教授。2000年大阪大学大学院基礎工学研究科修了。博士(工学)。同年大阪大学大学院基礎工学研究科助手。2002年大阪大学大学院情報科学研究科助教授。2007年より現職。電子情報通信学会, 言語処理学会, JSBi, ISCB, IEEE 各会員。



石井 優

大阪大学大学院医学系研究科/生命機能研究科教授。1998年大阪大学医学部医学科卒業。2000年大阪大学大学院医学系研究科助手(2005年, 医学博士)。2006年米国国立衛生学研究所客員研究員。2009年大阪大学免疫学フロンティア研究センター特任准教授。2011年同准教授。2013年より現職。



松田 秀雄 (正会員)

大阪大学大学院情報科学研究科教授。1987年神戸大学大学院自然科学研究科修了(学術博士)。同年同大学工学部助手。1994年大阪大学基礎工学部助教授。2002年より現職。JSBi, ISCB, IEEE CS, ACM 各会員。