

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⁴ 松田 秀雄^{1,4}

概要：さまざまな分子や細胞を可視化するバイオイメージング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。

1. はじめに

バイオイメージングとは細胞・組織または個体レベルでタンパク質などの分布・局在を捉え、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伴い、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などが利用される。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、増殖や融合、視野内への出入りによる密度の変化などが自動での処理を困難にしている。様々な方法やツールが提案されているが、万能の方法は未だなく、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[1]。

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。本研究の対象である Fucci (Fluorescent Ubiquitination-based Cell Cycle Indicator) [2] 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)、S/G2/M期では緑色の蛍光を発することで、個々の細胞周期の進行を可視化することが出来る。細胞周期は細胞の増殖や分化、がん細胞の挙動などに密接に関わることが知られており[3]、様々な実験条件下で経時観察を行うと、薬剤の影響や細胞の特性の変化を調べる事ができる。Fucci細胞の追跡を手動で行うためのツールとしてはImageJ[4]、また、半手動で行うツールとしてImageJのプラグインLineageTracker[5]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、細胞核の検出(セグメンテー

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

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

³ 大阪大学大学院医学系研究科
Graduate school of Medicine, Osaka University

⁴ 大阪大学免疫フロンティアセンター
World Premier International Research Center Immunology
Frontier Research Center, Osaka University

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

ション)と、前の画像から次の画像間で同一細胞を対応付ける(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、円形・楕円形の細胞核1つ1つが、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、細胞核の数と大きさを求める問題を、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みなすことが出来る。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みなせる。本研究の対象のデータでは、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従来の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。それに比較して提案手法では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。

2. 混合正規分布モデルと EM アルゴリズム

2.1 混合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[6]。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。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(例えば色空間)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。セグメンテーション以外においても、顕微鏡の点像分布関数(Point spread function)[7]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[8]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。

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。

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

式1は有限混合分布モデルであり、 g は仮定する確率分布の数、 $f_k(x|\theta_k)$ 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 は全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、 $g, \pi_k, \theta_k = \{\mu_k, \sigma_k^2\}$ となるが、成分数 k 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[9]。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。欠損値とは各データが各成分に所属する確率(事後確率)であり、これをEステップで計算し、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 $S = \{s_1, s_2, \dots, s_n\}$ とする時、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1回目のEステップでは欠測値の事後確率(x_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挙げられるが、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出来るだけ低いような 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みなすことができ、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みなせる。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り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）とみなし、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振る。割り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り当てを行う。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り当て

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とみ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 $t-1$ の際の細胞核の数とする。すなわち時刻 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

の時4つの分布から一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 $t-1$ でど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てのではなく、ハンガリー法 [11] を用いて全体として最適になるように割り当てる。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り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・割り当てを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り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り当てが、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時には、 $t-1$ 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り当て

ブートストラップサンプリングと最適割り当てについて、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_g である。ここで I_g は行列であり、行列の各要素 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、その割り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_g から、まずは g_{t-1} 回サンプリングして得られたデータを $\{s_{11}, s_{12}, \dots, s_{1g_{t-1}}\}$ 、 $k (k = 1, \dots, g_{t-1})$ を $t-1$ 時点の各成分分布の番号とすると、それぞれのデータの $t-1$ 時点の各成分分布への所属確率は、

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

となる。この確率は k 行 i 列 ($k = i$) の正方向列となる。ここで、サンプリングしたデータ s_{1i} と成分分布 f_k が 1 対 1 に対応するように割り当てを行う。この時ハンガリー法 [11] を用いて、割り当てられた組み合わせの確率の積が最大となるように選択し、各 s_{1i} に対応する割り当て a_{1i} を

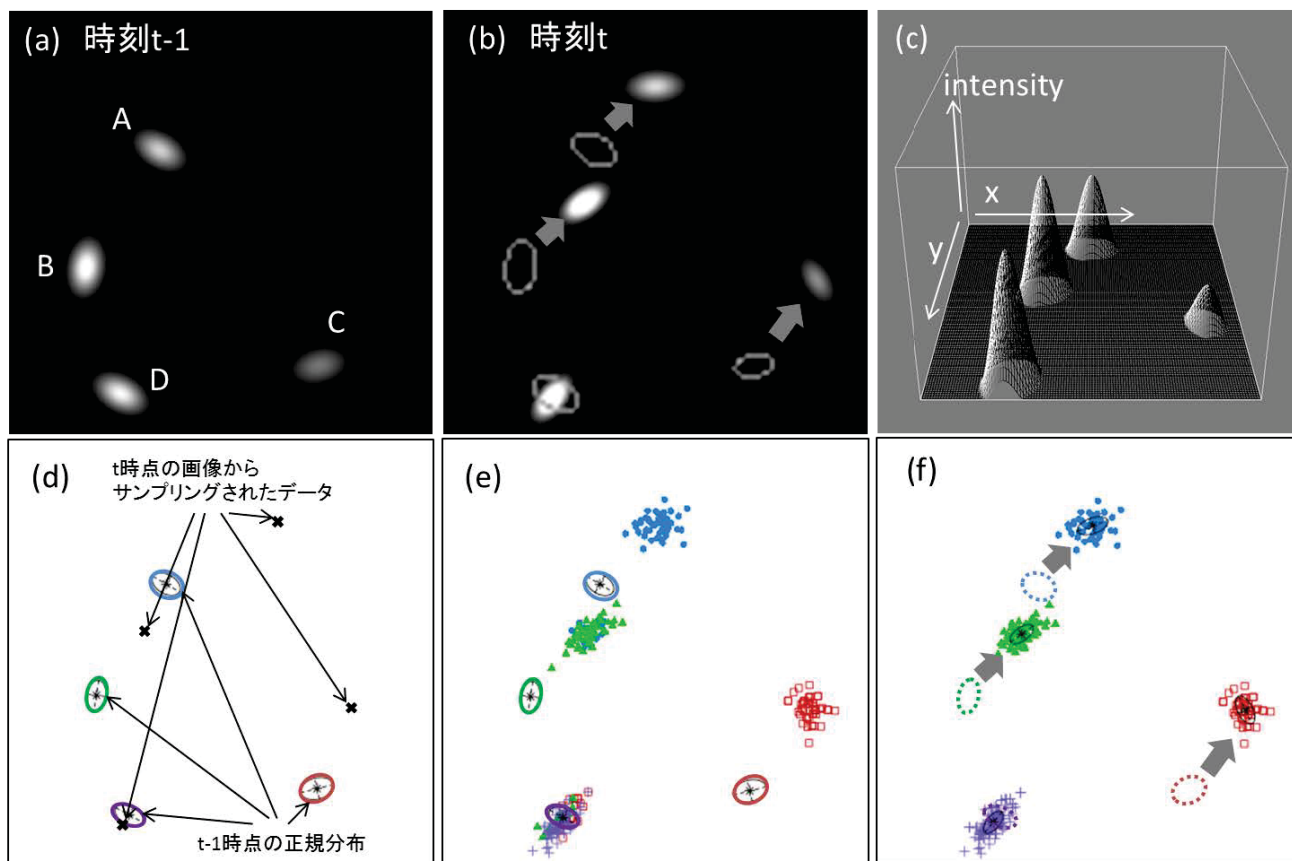


図 1 提案手法の概要

得る．単に最大確率へと割り当てを行うと，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に上手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 [12] ．

この g_{t-1} 個のサンプリングと割り当てをセットとして， b 回ブートストラッピングを行うことで，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り当てを行うことが出来る． 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 $t-1$ 時点のどの成分分布由来であるかの仮の割り当てを示す． $t-1$ 時点と t 時点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．ここでの入力は，STEP.1. の出力であるサンプリングされたデータ S と，その割り当て 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 章で述べた通りである．

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．

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 と同値とみなし，k-means 法の結果のクラスターを EM アルゴリズムの初期割り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 の時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他，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

するということである。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（細胞分裂に失敗すると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.1a と STEP.2b の手続きを導入する。

STEP.1.a 割り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 の確率が閾値以上のデータ S_h と割り当て A_h 、閾値以下のデータ S_l と割り当て 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、割り当てが間違っているとみなす。信頼出来るデータ S_h と割り当て 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。

STEP.2.a 成分分布の追加と削除を行う

S_h と割り当て 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。またデータ S_l に対して、初期値を求める際と同様に、k-means 法・EM アルゴリズムによる混合正規分布の最尤推定を行い、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。

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出来る。

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時のみそれらを同一の細胞核とした。

4. 実験

4.1 データについて

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。このうちコントロールである視野1と、ある種の薬剤を投与した視野2の2視野分を個別のデータセットとして用いた。人手で細胞核の検出と追跡を行った結果を正解とした時、視野1では36個の細胞が3日後には78個に、視野2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2では10,210個であった。

以降では細胞核の検出と追跡の精度とを評価するために、人手で検出と追跡を行った結果と比較した。また比較対象として、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、ImageJのプラグインとして実装されているLineageTracker[5]のソフトウェアを用いた。このソフトウェアはセグメンテーションにGaussian Maxima

と Seeded growth 法 [13] 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。

4.2 細胞核の検出精度

検出精度の評価を行うために、人手で細胞核の検出を行った結果を正解とし、以下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時、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時、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。また1つの真の領域に対して複数の検出した領域が対応した時、これも FN と FP とみなす。TP と FN の数の和は、真の領域の数と常に等しい。

視野1と視野2のそれぞれ150フレームずつ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、

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \quad (6)$$

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、真の正解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、今回は真の領域が複数に分割して検出された場合には FP と FN として計上するため、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。

4.3 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

表1 検出精度

視野1	TP	FP	FN	Precision	Recall	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2	TP	FP	FN	Precision	Recall	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

表 2 追跡精度

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 2	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

両方において同一の領域を指している時のみを正解とし、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全て不正解とする。自動追跡において、前後のフレーム間が共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、特に密集した状態でセグメンテーションの精度が下がるため、そのような場合に追跡も失敗することが多かった。提案手法では密集状態でも細胞核の移動については比較的正確に追跡することが出来ていたが、密集状況下で細胞核の出現と消滅が起こるときにうまく追跡が出来なくなる頻度が高かった。

5. おわりに

本研究では、多色蛍光イメージングによる経時観測データのために、混合正規分布モデルと EM アルゴリズムを用いて細胞核の検出と追跡を同時に行う方法を提案した。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、またハンガリー法による最適割り当てを行うことで、ノイズの影響を軽減し、また、細胞核が大きく移動した時や密集した状況下での検出・追跡においての精度の低下を防ぐことが出来た。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、正規分布で近似出来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、例えば、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 [12] 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 [14]) や EIC (Extend Information Criteria [15]) 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出来る可能性があると考えられる。

参考文献

- [1] 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).
- [2] Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H. and et al.: Visualizing spatiotemporal dynamics of multicellular cell-cycle progression., *Cell*, Vol. 132, No. 3, pp. 487-498 (2008).
- [3] 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).
- [4] Collins, T. J.: ImageJ for microscopy., *Biotechniques*, Vol. 43, No. 1 Suppl, pp. 25-30 (2007).
- [5] Downey, M. J., Jeziorska, D. M., Ott, S., Tamai, T. K., Koentges, G., Vance, K. W. and Bretschneider, T.: 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).
- [6] Baudry, J.-P., Raftery, A. E., Celeux, G., Lo, K., Gottardo and Raphael: Combining Mixture Components for Clustering, *Journal of computational and graphical statistics a joint publication of American Statistical Association Institute of Mathematical Statistics Interface Foundation of North America*, Vol. 9, No. 2, pp. 332-353 (2010).
- [7] 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).
- [8] Papas, M., Jarosz, W., Jakob, W., Rusinkiewicz, S., Matusik, W. and Weyrich, T.: Goal-Based Caustics, Vol. 30, pp. 503-511 (2011).
- [9] 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).
- [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] Kuhn, H. W.: The Hungarian method for the assignment problem, *Naval Research Logistics Quarterly*, Vol. 2, No. 1-2, pp. 83-97 (1955).
- [12] 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 (2011).
- [13] Fenistein, D., Lenseigne, B., Christophe, T., Brodin, P. and Genovesio, A.: A fast, fully automated cell segmentation algorithm for high-throughput and high-content screening., *Cytometry Part A the journal of the International Society for Analytical Cytology*, Vol. 73, No. 10, pp. 958-964 (2008).
- [14] Konishi, S. and Kitagawa, G.: Generalised information criteria in model selection, *Biometrika*, Vol. 83, No. 4, pp. 875-890 (1996).
- [15] 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).