

# 大域データ対応付けの反復実行による 細胞追跡精度の改善手法

藏重 昂平<sup>1</sup> 瀬尾 茂人<sup>1</sup> 間下 以大<sup>2</sup> 竹中 要一<sup>1</sup> 松田 秀雄<sup>1</sup>

**概要:** 近年、イメージング技術の発展により、細胞の挙動の経時観察が可能になった。大量の観察画像から細胞の挙動を自動的に分析することは重要な課題となっている。

細胞の挙動を解析する細胞追跡手法に、大域データ対応付けを用いた手法がある。大域データ対応付けでは、フレーム単位の対応付けにより短い細胞移動の軌跡であるトラックレットを生成し、細胞移動の仮説の尤度に基づきトラックレットの対応付けを行うことで、全体の軌跡を最適化する。しかし、細胞移動の仮説の適切な尤度計算は難しく、トラックレット対応付けの間違いが生じる問題がある。

本研究では、大域データ対応付けを一定フレーム領域でスライドしながら反復実行することによる、細胞追跡精度の改善手法を提案する。重力下におけるブラウン運動のシミュレーションデータおよび、濃度勾配に従い遊走する破骨細胞の観察画像に対して実験を行い、本手法により細胞追跡の精度が向上したことを確かめた。

**キーワード:** 細胞追跡, 大域データ対応付け, 反復実行

## 1. はじめに

近年、イメージング技術の発展により、生物学的現象を観察することが可能となった。多くの観察データが得られるようになり、扱うデータの爆発的な増加をもたらした。細胞増殖、分裂、および遊走といった挙動を解析することは、生物学や医学の研究に重要である。大量のデータを手で解析するためには膨大な時間がかかり困難であるため、コンピュータによる自動化が必要とされている。

細胞の挙動を解析するために、様々な解析手法が提案されてきた。画像処理での解析手法の1つに、トラッキングを用いた手法がある。トラッキングは一般的に、セグメンテーション、アソシエーションという2つの画像処理ステップからなる。セグメンテーションステップで画像から各細胞の検出をし、アソシエーションステップでフレーム単位で同一の細胞を対応付けることで、細胞移動の軌跡を得る。しかし、フレーム単位の対応付けでは、母娘関係の正確さの面で高い精度を達成することはノイズに影響されやすく困難である。例えば、誤検出による偽陽性 (False Positive, FP) のセグメンテーションが近くに現れると母娘関係のエラーを引き起こす可能性がある。その問題を解決するためにはグローバルな時間的情報が必要であり、本研究の対象

である大域データ対応付け (Global Data Association) が提案された [2]。大域データ対応付けを用いた細胞追跡では、まずフレーム単位の対応付けに基づいて検出反応をリンクすることにより短い細胞移動の軌跡であるトラックレットを生成する。次にトラックレットの対応付けを最大事後確率 (Maximum a posteriori, MAP) 問題として定式化する。MAP 問題は線形計画法により解く。この時、トラックレットの対応付けは、細胞移動の仮説およびその尤度に基づき線形計画法を解くことで行われるが、尤度の計算はトラックレットのデータに依存するため、正解の対応付けよりも尤度の合計が大きくなるような対応付け仮説が存在し、間違いが生じてしまうという問題がある。

本研究では、大域データ対応付けを反復実行することにより、トラックレットの対応付けの間違いを減らし、より高い精度で細胞追跡を行うことを目的とした手法を提案する。提案手法では、画像の全フレームに存在するトラックレットではなく、一定分のフレームに存在するトラックレットに対して大域データ対応付けを行う。これをフレーム全体にわたり反復して実行し、反復して得られた対応付け仮説の頻度に応じて対応付けを行う。また、本手法の有効性を示すために、シミュレーションデータによる実験および実データによる実験を行い、提案手法により細胞追跡の精度が改善されたことを示す。

<sup>1</sup> 大阪大学大学院情報科学研究科

<sup>2</sup> 大阪大学サイバーメディアセンター

## 2. 大域データ対応付けを用いた細胞追跡手法

本章では、本研究の従来手法とする、木構造の大域データ対応付け [2] を用いた細胞追跡手法について紹介する。

### 2.1 手法の概要

この手法では、まず画像のセグメンテーションにより細胞の検出をする。次に、アソシエーションにより細胞の対応付けを行い、短い細胞移動の軌跡であるトラックレットを生成する。大域データ対応付けにより生成したトラックレットの対応付けを行い、細胞移動の軌跡を得る。

### 2.2 トラックレット生成

フレーム単位の対応付けによって得られる長い細胞移動の軌跡は、短いものに比べるとノイズや他の細胞と対応付けられる可能性が高くなり、間違いを含む可能性が高くなる。したがって、セグメンテーションされた細胞から信頼性の高いトラックレットを生成する。追跡対象の周辺に他にターゲットになり得る細胞がなく、連続したフレーム間で、ターゲットとその候補の細胞のユークリッド距離が十分小さい細胞を対応付けすることで生成したトラックレットは信頼性が高いと言える。1 フレーム後の距離が大きく離れている、検出されなかった偽陰性 (False Negative, FN) の細胞により候補が存在しない、細胞同士が接触して複数の細胞が近くに存在する場合、その時点で対応付けを終了する。図 1 にトラックレットの生成例を示す。TP は正しく検出された真陽性 (True Positive) の細胞である。対応付けされなかった 1 つの細胞もトラックレットとし、生成されたトラックレットの集合を  $X = \{X_i\}$  と表す。

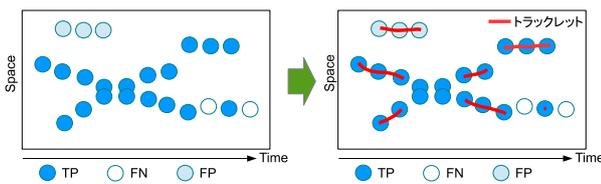


図 1 トラックレットの生成例

### 2.3 大域データ対応付け

この節では、木構造の大域データ対応付けについて説明する。 $T = \{T_k\}$  を、画像全フレームにわたる細胞軌道の木の仮説集合とする。それぞれの木  $T_k$  は、対応付けられたトラックレットで形成される。 $T_k$  に関する木構造の仮説は次の表記を用いて定義する。

(1)  $E_k = \{E_{k_i}\}$ . 木  $T_k$  の辺集合. 各エッジはトラックレットの順序リストとして定義される. つまり,  $E_{k_i} = \{X_{k_i}^j\}$ .  $X_{k_i}^j$  はエッジ  $E_{k_i}$  の  $j$  番目のトラックレットを表す. また,  $E_{k_0}$  は木のルートエッジを表す.

(2)  $B_k = \{B_{k_i}\}$ . 木  $T_k$  の分岐ノードの集合. 各分岐ノード  $B_{k_i}$  は親子関係を定義する.  $B_{k_i} = \{E_{k_{p_i}}, E_{k_{c_{i1}}}, E_{k_{c_{i2}}}\}$ .  $E_{k_{p_i}}$  は親を,  $E_{k_{c_{i1}}}, E_{k_{c_{i2}}}$  は子を表す.

(3)  $L_k = \{E_{k_{l_i}}\}$ . 木  $T_k$  のリーフエッジの集合.

$X$  を観測されたトラックレットの集合とすると、最善の仮説  $T^*$  を解くために事後確率を最大限にする。

$$\begin{aligned} T^* &= \arg \max_T P(T|X) \\ &= \arg \max_T P(X|T)P(T) \\ &= \arg \max_T \prod_{X_i \in X} P(X_i|T) \prod_{T_k \in T} P_{Tree}(T_k) \quad (1) \end{aligned}$$

式 (1) では、与えられた  $T$  に入力トラックレットの尤度は条件付独立であり、 $T_k \in T$  は互いに重なることはないと仮定する。つまり、 $T_k \cap T_l = \phi, \forall k \neq l$  である。観測されたトラックレット  $X_i$  の尤度は、次で表される。

$$P(X_i|T) = \begin{cases} P_{TP}(X_i), & \text{if } \exists T_k \in T, X_i \in T_k \\ P_{FP}(X_i), & \text{otherwise} \end{cases} \quad (2)$$

$P_{TP}(X_i)$  は  $X_i$  が真陽性である確率、 $P_{FP}(X_i)$  は偽陽性である確率を表す。 $P_{tree}(T_k)$  はマルコフ連鎖としてモデル化される。

$$\begin{aligned} P_{tree}(T_k) &= P_{ini}(E_{k_0}) \times \prod_{E_{k_i} \in T_k} P_{edge}(E_{k_i}) \\ &\quad \times \prod_{\{E_{k_{p_i}}, E_{k_{c_{i1}}}, E_{k_{c_{i2}}}\} \in B_k, B_k \in T_k} P_{div}(E_{k_{c_{i1}}}, E_{k_{c_{i2}}}|E_{k_{p_i}}) \\ &\quad \times \prod_{E_{k_{l_i}} \in L_k, L_k \in T_k} P_{term}(E_{k_{l_i}}) \quad (3) \end{aligned}$$

$P_{ini}$  は木のルートの始端確率で、 $P_{term}$  はリーフの終端確率である。 $P_{div}(E_{k_{c_{i1}}}, E_{k_{c_{i2}}}|E_{k_{p_i}})$  はエッジ  $E_{k_{p_i}}$  が 2 つのエッジ  $E_{k_{c_{i1}}}, E_{k_{c_{i2}}}$  に分裂するエッジ分裂確率である。マルコフ仮定のもとで、エッジ確率は次のように定式化される。

$$P_{edge}(E_{k_i}) = \prod_{j=1:N_{k_i}-1} P_{link}(X_{k_i}^j | X_{k_i}^{j-1}) \quad (4)$$

$P_{link}(X_{k_i}^j | X_{k_i}^{j-1})$  はトラックレット  $X_{k_i}^j$  と  $X_{k_i}^{j-1}$  が一緒に繋がる確率で、 $N_{k_i}$  はエッジ  $E_{k_i}$  のトラックレットの数である。 $X_{k_i}^0$  を  $E_{k_i}$  の最初のトラックレット、 $X_{k_i}^{end}$  を最後のトラックレットとする。マルコフ仮定のもとで、始端、終端、分裂確率はそれぞれ、次のように定式化される。

$$P_{ini}(E_{k_0}) = P_{ini}(X_{k_i}^0) \quad (5)$$

$$P_{term}(E_{k_{l_i}}) = P_{term}(X_{k_i}^{end}) \quad (6)$$

$$P_{div}(E_{k_{c_{i1}}}, E_{k_{c_{i2}}}|E_{k_{p_i}}) = P_{div}(X_{k_{c_{i1}}}^0, X_{k_{c_{i2}}}^0 | X_{k_{p_i}}^{end}) \quad (7)$$

式 (1) に式 (2.2)-(2.7) を代入した後、目的関数に対数をとったものが式 (8) になる。

$$\begin{aligned}
T^* = \arg \max_T \{ & \sum_{X_i \notin T_k, \forall T_k \in T} \log P_{FP}(X_i) \\
& + \sum_{X_i \in T_k, \forall T_k \in T} \log P_{TP}(X_i) \\
& + \sum_{X_{k_0}^0 \in E_{k_0}, E_{k_0} \in T_k, \forall T_k \in T} \log P_{ini}(X_{k_0}^0) \\
+ & \sum_{X_{k_i}^j, X_{k_i}^{j-1} \in E_{k_i}, \forall E_{k_i} \in T_k, \forall T_k \in T} \log P_{link}(X_{k_i}^j | X_{k_i}^{j-1}) \\
& + \sum_{\{X_{k_{p_i}}^{end}, X_{k_{c_{i1}}}^0, X_{k_{c_{i2}}}^0\} \in B_k, B_k \in T_k, \forall T_k \in T} \log P_{div}(X_{k_{c_{i1}}}^0, X_{k_{c_{i2}}}^0 | X_{k_{p_i}}^{end}) \\
& + \sum_{X_{k_{l_i}}^{end} \in E_{k_{l_i}}, \forall E_{k_{l_i}} \in L_k, L_k \in T_k, \forall T_k \in T} \log P_{term}(X_{k_{l_i}}^{end}) \} \quad (8)
\end{aligned}$$

式 (8) の MAP 問題を整数計画問題として定式化して解く。  $N_X$  を全シーケンスにおけるトラックレットの数、ベクトル  $\rho$  は起こりうる全ての仮説の尤度を、行列  $C$  は競合する仮説を避けるための制約を格納する。  $C$  の各行は  $2N_X$  の列があり、各列は 2 つのトラックレットの対応付けに関するトラックレットのインデックスを示す。ベクトル  $\rho$  と行列  $C$  のエントリは以下の 5 つの仮説に基づき計算する。  $h$  は新しい仮説のインデックスである。

#### (1) 始端仮説 (Initialization hypothesis)

トラックレット  $X_k$  の始端がシーケンスの始めに現れる、画像の境界付近に現れる場合、そのトラックレットは始端トラックレットの候補である。行列  $C$  と尤度  $\rho$  のエントリは次のように定義する。

$$C(h, i) = \begin{cases} 1, & \text{if } i = N_X + k \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

$$\rho(h) = \log P_{ini}(X_k) + 0.5 \log P_{TP}(X_k) \quad (10)$$

#### (2) 終端仮説 (Termination hypothesis)

トラックレット  $X_k$  終端がシーケンスの終わりに現れる、画像の境界付近に現れる場合、そのトラックレットは終端トラックレットの候補である。行列  $C$  と尤度  $\rho$  のエントリは次のように定義する。

$$C(h, i) = \begin{cases} 1, & \text{if } i = k \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

$$\rho(h) = \log P_{term}(X_k) + 0.5 \log P_{TP}(X_k) \quad (12)$$

#### (3) 移動仮説 (Translation hypothesis)

トラックレット  $X_{k_1}$  の終端とトラックレット  $X_{k_2}$  の始端の時空間の距離が一定以内の場合、  $X_{k_1} \rightarrow X_{k_2}$  はトラックレット移動の候補である。行列  $C$  と尤度  $\rho$  のエントリは次のように定義する。

$$C(h, i) = \begin{cases} 1, & \text{if } i = k_i \text{ or } i = N_X + k_2 \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

$$\rho(h) = \log P_{link}(X_{k_2} | X_{k_1}) + 0.5 \log P_{TP}(X_{k_1}) + 0.5 \log P_{TP}(X_{k_2}) \quad (14)$$

#### (4) 分裂仮説 (Dividing hypothesis)

トラックレット  $X_p$  の終端が細胞分裂発生イベントの位置に近い場合、そのトラックレットは親トラックレットの候補であり、トラックレット  $X_{c_1}$ ,  $X_{c_2}$  の始端が  $X_p$  に近い場合、それらのトラックレットは子トラックレットの候補である。行列  $C$  と尤度  $\rho$  のエントリは次のように定義する。

$$C(h, i) = \begin{cases} 1, & \text{if } i = p \text{ or } i = N_X + c_1 \text{ or} \\ & i = N_X + c_2 \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

$$\rho(h) = \log P_{div}(X_{c_1}, X_{c_2} | X_p) + 0.5 \log P_{TP}(X_p) + 0.5 \log P_{TP}(X_{c_1}) + 0.5 \log P_{TP}(X_{c_2}) \quad (16)$$

#### (5) 偽陽性仮説 (False positive hypothesis)

全てのトラックレットは偽陽性である可能性がある。トラックレット  $X_k$  が偽陽性の候補である場合、行列  $C$  と尤度  $\rho$  のエントリは次のように定義する。

$$C(h, i) = \begin{cases} 1, & \text{if } i = k \text{ or } i = N_X + k \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

$$\rho(h) = \log P_{FP}(X_k) \quad (18)$$

$N_X$  個のトラックレットで  $M$  個の仮説を作った後、式 (8) の MAP 問題は、任意の木が互いに重ならないという制約下で、  $\rho$  に対応する要素の和が最大になるような  $C$  の行の部分集合を選択することとみなすことができる。これは次のような整数計画問題として定式化される。

$$x^* = \arg \max_x \rho^T x, \text{ s.t. } C^T x = 1 \quad (19)$$

$x$  は  $M \times 1$  のバイナリベクトルで、  $x_k = 1$  は  $k$  番目の仮説がグローバル解で選択されたことを表す。制約  $C^T x = 1$  は、各トラックレットが対応付けされるか偽陽性トラックレットであることを保証する。図 2 に整数計画問題の例を示す。図 2 左のようなトラックレットを考え、可能性のある仮説に関して、定義に基づきベクトル  $\rho$  と行列  $C$  のエントリを計算する。その後、式 (19) の整数計画問題を解き、解を求める。オレンジ色の部分が解として選択されたものを表す。

## 2.4 仮説の確率推定の実装

この節では、本研究における各仮説の確率推定の実装方

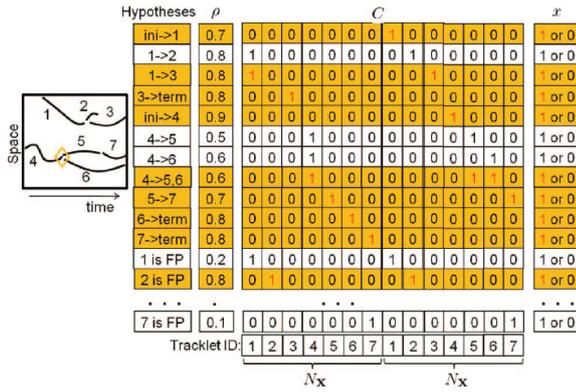


図 2 整数計画問題の例 ([2] より)

法について説明する。

(1) 偽陽性確率  $P_{FP}$

トラックレット  $X_i$  が誤検出である確率.  $\alpha$  を細胞検出器の誤検出率,  $|X_i|$  をトラックレット  $X_i$  を構成する細胞の数とする.

$$P_{FP}(X_i) = \alpha^{|X_i|} \quad (20)$$

(2) 真陽性確率  $P_{TP}$

トラックレット  $X_i$  が正検出である確率.

$$P_{TP}(X_i) = 1 - P_{FP}(X_i) \quad (21)$$

(3) 始端確率  $P_{ini}$

トラックレット  $X_i$  が始端トラックレットである確率.  $R_{i_0}$  はトラックレット  $X_i$  の最初の検出反応,  $dt_0(R_{i_0})$  は初期フレームから検出反応  $R_{i_0}$  が現れるまでの時間距離,  $ds(R_{i_0})$  は検出反応  $R_{i_0}$  と画像の境界との空間距離である.  $\theta_t$  は時間しきい値,  $\theta_s$  は距離しきい値で, このしきい値以内のトラックレットに関して確率を計算する. また,  $\lambda_1$  と  $\lambda_2$  は調整用のパラメータである.

$$P_{ini}(X_i) = \begin{cases} e^{-\frac{dt_0(R_{i_0})}{\lambda_1}}, & \text{if } dt_0(R_{i_0}) \leq \theta_t \\ e^{-\frac{ds(R_{i_0})}{\lambda_2}}, & \text{if } ds(R_{i_0}) \leq \theta_s \\ \eta, & \text{otherwise. } (\eta \approx 0) \end{cases} \quad (22)$$

(4) 終端確率  $P_{term}$

トラックレット  $X_i$  が終端トラックレットである確率.  $R_{i_{end}}$  はトラックレット  $X_i$  の最後の検出反応,  $dt_{end}(R_{i_{end}})$  は最終フレームから検出反応  $R_{i_{end}}$  が現れるまでの時間距離である.

$$P_{term}(X_i) = \begin{cases} e^{-\frac{dt_{end}(R_{i_{end}})}{\lambda_1}}, & \\ \text{if } dt_{end}(R_{i_{end}}) \leq \theta_t \\ e^{-\frac{ds(R_{i_{end}})}{\lambda_2}}, & \text{if } ds(R_{i_{end}}) \leq \theta_s \\ \eta, & \text{otherwise. } (\eta \approx 0) \end{cases} \quad (23)$$

(5) 連結確率  $P_{link}$

トラックレット  $X_j$  と  $X_i$  が繋がる確率.  $ds(R_{j_0} - R_{i_{end}})$  はトラックレット  $X_j$  の最初の検出反応とトラックレット  $X_i$  の最後の検出反応の空間距離,  $dt(R_{i_{end}} - R_{j_0})$  はトラックレット  $X_j$  の最初の検出反応とトラックレット  $X_i$  の最後の検出反応の時間距離である. また,  $\lambda_3$  は調整用のパラメータである.

$$P_{link}(X_j|X_i) = e^{-\frac{ds(R_{j_0} - R_{i_{end}})}{\lambda_3}}, \quad (24)$$

if  $dt(R_{i_{end}} - R_{j_0}) \leq \theta_t$  and  
 $ds(R_{j_0} - R_{i_{end}}) \leq \theta_s$

(6) 分裂確率  $P_{div}$

トラックレット  $X_i$  が細胞分裂する確率. 本研究において, 扱うデータの細胞は細胞分裂をしないため, 細胞分裂の検出はない. したがって, 分裂確率に関しては考慮しない.

2.5 大域データ対応付けの問題点

大域データ対応付けにより, 偽陽性のセグメンテーションによる対応付けの間違いを減らすこと, 偽陰性のセグメンテーションにより途切れたトラックレットの対応付けが可能になった. しかし, トラックレット対応付けの仮説の尤度計算は, トラックレットのデータに依存し, 短いトラックレットが多い場合や, 偽陰性のセグメンテーションが多い場合に適切な値を計算することは難しい. 本来は正解となるべき対応付けでもトラックレット同士が離れており尤度が小さい値になる場合がある. 最適化した結果, 正解より尤度の和が大きくなるような対応付け仮説が存在し, 間違っただ対応付けが生じる問題がある.

3. 大域データ対応付けの反復実行による細胞追跡精度の改善手法

2章で, 大域データ対応付けを用いた細胞追跡手法とその問題点について述べた. 本章では, トラックレットの対応付けの間違いが生じる問題に対して, 大域データ対応付けの反復実行による細胞追跡精度の改善手法を提案する. 提案手法では, 同様の尤度計算方法でも対応付けの間違いを減らし, より高い精度の細胞追跡結果を得ることを目的とする.

3.1 手法の概要

トラックレットの生成までは従来手法と同様である. 提案手法では, 次に一定フレーム間で大域データ対応付けをずらしながら反復実行し, そのデータをもとにトラックレットの対応付けを行う. 最後に追跡結果の出力をし, 細胞移動の軌跡を得る.

尤度の定義により, 全フレームでの大域データ対応付けと一定フレーム間での大域データ対応付けでは, 解として

選ばれる対応付けの仮説に違いが生じる。一定フレーム間をずらしながら大域データ対応付け反復実行し、解として選ばれた回数が多い対応付け仮説は正しいと考え、その頻度に基づき対応付けを行う。

### 3.2 大域データ対応付けの反復実行

ここでは大域データ対応付けの反復実行について説明する。従来手法では、大域データ対応付けを画像全フレームのトラックレットに対して行うが、提案手法では一定分のフレームに存在するトラックレットを取り出し、その取り出されたトラックレット間で行う。アソシエーションの結果は出力せず、取り出されたトラックレット間における解の対応付け仮説の情報を行列  $N_{ij}$  に解として選ばれた対応付け仮説の回数を保存する。  $N_{ij}$  は次のように定義する。

$$N_{ij} = 0(\text{初期状態}) \quad (25)$$

$$N_{h_1 h_2} = N_{h_1 h_2} + 1 \quad (26)$$

$i$  はトラックレット, 始端仮説の集合 ( $i = \{X, ini\}$ ),  $j$  はトラックレット, 終端仮説, 偽陽性仮説の集合 ( $j = \{X, term, FP\}$ ) である。  $h_1, h_2$  は解の対応付け仮説を表す。例えば, トラックレット  $X_1$  の始端仮説は  $(h_1, h_2) = (ini, X_1)$ , トラックレット  $X_1, X_2$  の移動仮説は  $(h_1, h_2) = (X_1, X_2)$ , トラックレット  $X_2$  の終端仮説は  $(h_1, h_2) = (X_2, term)$ , トラックレット  $X_3$  の偽陽性仮説は  $(h_1, h_2) = (X_3, FP)$  と表す。

これを初期フレームから最終フレームに至るまで1フレームずつずらしながら反復実行し、保存した仮説の頻度をもとに整数計画法を解きアソシエーションを行い、細胞移動の軌跡を得る。

#### 3.2.1 トラックレットの取り出し

ここでは、トラックレットの取り出しについて説明する。  $t$  フレーム目の場合を考える。図3のように、トラックレットの出現フレームまたは消失フレームが  $t$  フレーム目から  $t + \beta$  フレーム目までに存在するトラックレットを取り出す。取り出されたトラックレット  $X'_i$  は式(27)で表される。  $dt_t(R_{i_0})$  は  $t$  フレーム目からトラックレット  $X_i$  の出現フレームまでの時間距離,  $dt_t(R_{i_{end}})$  は  $t$  フレーム目からトラックレット  $X_i$  の消失フレームまでの時間距離である。

$$X'_i = \{X_i | (0 \leq dt_t(R_{i_0}) < \beta) \vee (0 \leq dt_t(R_{i_{end}}) < \beta)\} \quad (27)$$

このトラックレット  $X'_i$  に関して、従来手法と同様にトラックレット対応付け仮説の尤度を計算し、整数計画問題を解き対応付けの解を求める。その解の対応付け仮説の情報は保存しておく。始端確率および終端確率の計算方法については、初期フレームと最終フレームの定義が変わるた

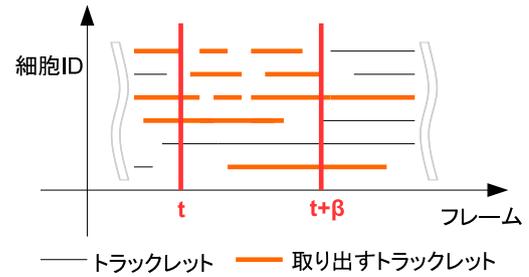


図3 トラックレットの取り出し例

め、それぞれ以下のように計算方法を変更した。

#### (1) 始端確率 $P_{ini}$

$dt_t(R'_{i_0})$  は  $t$  フレーム目からトラックレット  $X'_i$  の最初の検出反応が現れるフレームまでの時間距離である。  $t$  フレーム以前から存在するトラックレットの始端確率に関しては1と定義した。

$$P_{ini}(X'_i) = \begin{cases} 1, & \text{if } dt_t(R'_{i_0}) \leq 0 \\ e^{-\frac{dt_0(R'_{i_0})}{\lambda_1}}, & \text{if } 0 < dt_t(R'_{i_0}) \leq \theta_t \\ e^{-\frac{ds(R'_{i_0})}{\lambda_2}}, & \text{if } ds(R'_{i_0}) \leq \theta_s \\ \eta, & \text{otherwise. } (\eta \approx 0) \end{cases} \quad (28)$$

#### (2) 終端確率 $P_{term}$

$dt_{t+\beta}(R'_{i_{end}})$  は  $t + \beta$  フレーム目からトラックレット  $X'_i$  の最後の検出反応が現れるフレームまでの時間距離である。  $t + \beta$  フレーム以降まで存在するトラックレットの終端確率に関しては1と定義した。

$$P_{term}(X'_i) = \begin{cases} 1, & \text{if } dt_{t+\beta}(R'_{i_{end}}) \leq 0 \\ e^{-\frac{dt_{end}(R'_{i_{end}})}{\lambda_1}}, & \\ \text{if } 0 < dt_{t+\beta}(R'_{i_{end}}) \leq \theta_t & (29) \\ e^{-\frac{ds(R'_{i_{end}})}{\lambda_2}}, & \text{if } ds(R'_{i_{end}}) \leq \theta_s \\ \eta, & \text{otherwise. } (\eta \approx 0) \end{cases}$$

#### 3.2.2 反復実行とアソシエーション

前小々節では  $t$  フレーム目について考えたが、初期フレームから最終フレームに至るまで、1フレームずつずらしながら反復実行する。最終フレームまで終わると、取り出した部分ごとの解の対応付け仮説の数の和が  $N_{ij}$  に格納される。

最後に、反復実行して得られた全ての解の対応付け仮説を用い、トラックレットの対応付けを行う。大域データ対応付けと同様に整数計画問題を解くことで行うが、提案手法では、解として選ばれた対応付け仮説の頻度を最大化する。

また、フレームを部分ごとに取り出して反復実行するため、従来手法ではないはずの始端仮説、終端仮説も解として選ばれる。したがって、次のように始端仮説と終端仮説の数に重み付け処理をした。

$$N_{inij} = N_{inij} \times W \quad (30)$$

$$N_{iterm} = N_{iterm} \times W \quad (31)$$

上記の処理を加えた  $N_{ij}$  を次のように  $\rho'$  に代入する。  $h'$  は  $N_{ij}$  のエントリを表す。

$$\rho'(h') = N_{ij} (N_{ij} > 0) \quad (32)$$

$C'$  には、 $\rho'$  に対応する仮説の制約行列を大域データ対応付けと同様の定義で代入し、以下の整数計画問題を線形緩和して解く。

$$x^* = \arg \max_x \rho'^T x, \text{ s.t. } C'^T x = 1 \quad (33)$$

式 (33) で得られた解が提案手法の結果となる。

## 4. 実験

提案手法の有効性を示すために、シミュレーションデータに基づく実験および実データに基づく実験を行った。トラックレット生成までをそれぞれデータ加工、ツールを用いて行い、位置座標およびトラックレットの情報のあるデータを用いて手法のテストをした。本章では実験結果を示すとともに、実験結果に対する考察をする。

### 4.1 準備実験

#### 4.1.1 シミュレーションデータに基づく実験

本研究では、実データだけでなくシミュレーションデータに基づく実験を行った。なぜなら、細胞追跡を行うツールは数多くあるが、完璧に追跡できるものはないため、細胞追跡の結果が正しいかどうかの評価は、人の手で細胞を追跡して行うのが一般的である。しかし、大量のデータの作成は困難であり、正解データを作成するのは容易ではない。したがって、本研究では実データである破骨細胞の動きに近い運動のシミュレーションデータを作成したものを正解データとし、そのデータに実際に起こるようなセグメンテーションエラーを加えたものを用意し、手法のテストを行った。

##### 4.1.1.1 実験条件

従来手法で用いる誤検出率、調整用パラメータ、しきい値はそれぞれ次のように設定した。  $\alpha = 0.3$ ,  $\lambda_1 = 5$ ,  $\lambda_2 = 15$ ,  $\lambda_3 = 40$ ,  $\theta_t = 20$ ,  $\theta_s = 50$ 。また、提案手法で新たに用いるトラックレットを取り出すフレーム幅、重み付け値は次のように設定した。  $\beta = 50$ ,  $W = 0.3$ 。

##### 4.1.1.2 テストデータの作成

使用する破骨細胞の画像は、濃度勾配に従い画像上から下に動く。シミュレーションを行うにあたり、似たような運動である細胞の重力下におけるブラウン運動 [5] をモデルにした。モデルの運動方程式は次のように表される。任意の細胞  $i$  の位置ベクトルを  $r_i$ 、粒子に作用する重力を  $f_i^{(g)}$ 、細胞  $i, j$  間に作用する力を  $f_{ij}$ 、ランダム力を  $f_i^B$  と

する。

$$m \frac{d^2 r_i}{dt^2} = f_i^{(g)} + \sum_{j(\neq i)} f_{ij} + f_i^B \quad (34)$$

式 (34) を元に画面サイズを横 300[pixel]、縦 300[pixel] としてシミュレーションを行い、300 フレーム分の正解となるシミュレーションデータを作成し

次にテストデータを作成するために、シミュレーションデータに対して、実際の画像で起こりうるセグメンテーションエラーを加えた。1つは、細胞同士の接触によるセグメンテーションエラーである。同一フレームにおける座標データを比較し、複数の細胞が 15[pixel] 以内に存在していた場合、その細胞を取り除くことにより再現した。2つ目は不十分なセグメンテーションによる偽陰性の細胞である。正検出率 (True Positive Rate, TPR) を設定し、正解データの細胞の個数からその値の個数分をランダムに取り出して使用することにより再現した。ここでは、TPR=0.7 と設定した。2つのセグメンテーションエラーを加えたものから、同一の細胞で連続したフレームに存在しているものを1つのトラックレットとみなし、トラックレット生成を行った。

また、使用したテストデータの検出精度を表1に示す。細胞総数は作成した正解データのうち画面内に存在する全細胞の数である。TP, FP, FN はそれぞれ、正検出の細胞の数、誤検出の細胞の数、セグメンテーションエラーを加えた細胞の数である。検出精度の基準には適合率 (Precision)、再現率 (Recall) [1] を用いた。適合率は検出結果のうち正しく検出できた数の割合、再現率は全ての正解の数に対して正しく検出できた数の割合であり、それぞれ次の式で計算する。

$$Precision = \frac{TP}{TP + FP} \quad (35)$$

$$Recall = \frac{TP}{TP + FN} \quad (36)$$

表 1 シミュレーションデータ検出精度

細胞総数	TP	FP	FN	Precision	Recall
6348	4443	0	1905	1.000	0.700

#### 4.1.2 実データに基づく実験

##### 4.1.2.1 実験条件

従来手法で用いる誤検出率、調整用パラメータ、しきい値はそれぞれ次のように設定した。  $\alpha = 0.3$ ,  $\lambda_1 = 5$ ,  $\lambda_2 = 15$ ,  $\lambda_3 = 40$ ,  $\theta_t = 20$ ,  $\theta_s = 50$ 。また、提案手法で新たに用いるトラックレットを取り出すフレーム幅、重み付け値は次のように設定した。  $\beta = 50$ ,  $W = 0.3$ 。

##### 4.1.2.2 使用画像

本研究では、濃度勾配により垂直下方向へ遊走する破骨細胞の画像を使用した。図4は使用した画像のある時点

でのスナップショットである。また、画像のフレーム数は272、サイズは横 416[pixel]、縦 311[pixel] である。

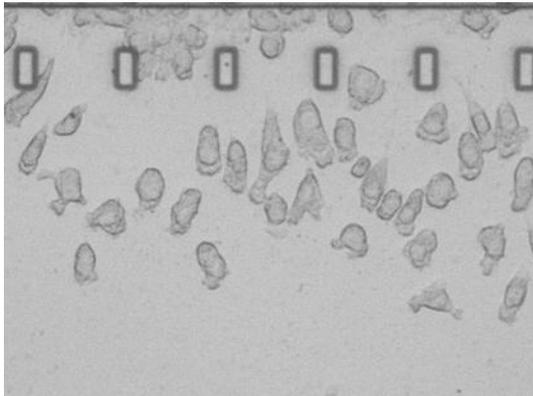


図 4 遊走する破骨細胞の観察画像

#### 4.1.2.3 テストデータの作成

細胞画像のセグメンテーションおよびトラックレット生成には、オープンソースの画像処理ソフトウェアである ImageJ のプラグイン、LineageTracker[4] を用いた。LineageTracker は、セグメンテーションの方法をいくつかから選択し、最近傍法によるアソシエーションのパラメータを調節することができる自動細胞追跡のツールである。セグメンテーションにはしきい値法を用い、1 フレーム後に 7.0[pixel] 以内にある細胞に関して対応付けを行った。このデータに対し、正解データから細胞の ID を付与し、トラックレットがどの細胞の一部であるかを判断できるようにした。また、同じ細胞 ID のもので出現フレーム昇順にトラックレットを並べ、トラックレット番号を順に付与した。同一細胞 ID で、トラックレット番号が連続したものが正しい対応付けであると判断する。

実験結果を評価するための正解データの作成には、ImageJ のプラグイン、MTrackJ を用いた。MTrackJ は細胞 1 つ 1 つの xy 座標、細胞 ID、フレーム番号を出力する手動トラックのツールである。このツールを使用して画像中の細胞を手動で追跡し、出力データを正解データとして使用した。

また、使用したテストデータの検出精度を表 2 に示す。細胞総数は MtrackJ で作成した正解データの全フレームにわたる細胞の数である。TP, FP, FN はそれぞれ、LineageTracker で作成したテストデータのうち正検出の細胞の数、誤検出の細胞の数、未検出の細胞の数である。

表 2 実データ検出精度

細胞総数	TP	FP	FN	Precision	Recall
11667	5495	484	6172	0.919	0.471

## 4.2 実験結果

### 4.2.1 評価基準

評価基準として、細胞追跡の精度評価に追跡精度 (Tracking Accuracy, TA), アソシエーションの精度評価に対応付け再現率 (Association Recall, AR), 対応付け適合率 (Association Precision, AP) [3] を用いた。細胞  $j$  の  $TA(j)$ ,  $AR$ ,  $AP$  はそれぞれ次の式で計算する。

$$TA(j) = \frac{L(i, j)}{L(j)} \quad (37)$$

$$AR = \frac{TP}{TP + FN} \quad (38)$$

$$AP = \frac{TP}{TP + FP} \quad (39)$$

$L(i, j)$  は細胞  $j$  のトラックレット  $i$  が細胞  $j$  に正しく対応付けられたフレームの長さ、 $L(j)$  は正解データの細胞  $j$  のフレーム総数を表す。全ての細胞についての平均を  $TA$  の値とする。

また、ここでの TP は正しくトラックレットの対応付けができた数、FP は間違っただ対応付けされた数、FN は正しく対応付けされなかった数を表す。例えば、図 5 のようなトラックレットを考える。このトラックレットの正しい対応付けは図 6 に示すように、1 の始端仮説、1 → 2, 2 → 3 の移動仮説、3 の終端仮説、4 の偽陽性仮説であり、この時 TP の数は 6, FP と FN の数は 0 となる。図 7 のように間違っただ対応付けされた場合、TP の数は 1, FP の数は 6, FN の数は 5 となる。

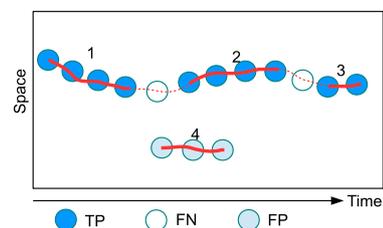


図 5 トラックレット例

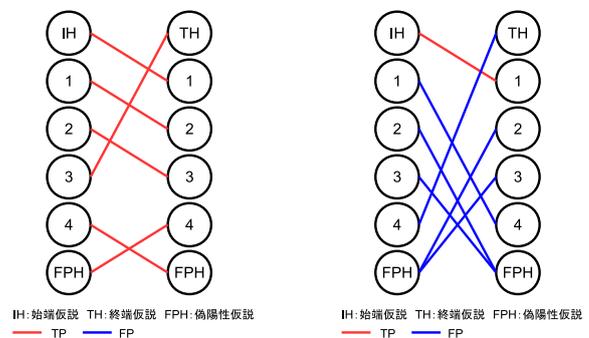


図 6 図 5 の正しい対応付け 図 7 図 5 の間違っただ対応付け

### 4.2.2 結果

表 3, 表 4 にそれぞれシミュレーションデータ、実データに対し従来手法および提案手法を実行した結果を示す。

表 3 シミュレーションデータ結果

手法	TP	FP	FN	AR	AP	TA
従来手法	1241	46	37	0.964	0.971	0.977
提案手法	1271	7	7	0.990	0.990	0.985

表 4 実データ結果

手法	TP	FP	FN	AR	AP	TA
従来手法	450	345	424	0.566	0.515	0.570
提案手法	483	300	391	0.617	0.553	0.585

### 4.3 考察

表 3, 表 4 から, シミュレーションデータ, 実データともに評価基準 AR, AP, TA 全てにおいて, 提案手法が従来手法の数値を上回っていることがわかる. トラックレットの対応付けの間違いが減り, 正しい対応付けが増えたため, その分追跡精度が改善したと考えられる.

シミュレーションデータは正解データの細胞から 7 割をランダムに選び作成したが, 同一細胞のトラックレットが時空間距離が少なく存在しているため, 実データに比べて数値が高くなったと考えられる. 実データはセグメンテーションの精度が高いものではなく, 全細胞のうち正しく検出できていた割合が表 2 より 0.471 であった. 同一細胞のトラックレットでもフレーム差が離れて存在しているものが多く, トラックレット対応付け仮説の尤度計算がうまくできなかったため, シミュレーションデータに比べて数値が低くなったと考えられる. セグメンテーション精度が高いものを用意できれば, それだけ良い結果が得られると考えられる.

### 5. おわりに

本論文では, 大域データ対応付けの反復実行による細胞追跡精度の改善手法を提案した. シミュレーションデータに基づく実験, 実データに基づく実験の結果から, トラックレット対応付けの間違いが減り, 細胞追跡の精度が向上したことが示された.

従来手法と比べて数値上は精度が向上したが, 従来手法で追跡できていた箇所全てが提案手法でも追跡できていた訳ではなかった. 提案手法では, 大域データ対応付けの反復実行により得られたトラックレットの対応付け仮説の頻度をもとに整数計画問題を解き, その解で対応付けを行ったが, 頻度が多いものが必ず正しいというわけではないということであり, この点に改善の余地がある.

また, 本論文で扱った破骨細胞のデータは遊走するのみであったが, 破骨細胞は融合することが知られている. 細胞の融合, 分裂, 細胞死といった他の様々な現象も扱えるように拡張することも, 今後の課題として挙げられる.

### 参考文献

- [1] Ryoma Bise, Kang Li, Sungeun Eom, and Takeo Kanade. Reliably tracking partially overlapping neural stem cells in dic microscopy image sequences. In *MICCAI Workshop on Optical Tissue Image Analysis in Microscopy, Histopathology and Endoscopy (OPTMHIS)*, No. CMU-RI-TR-, pp. 67–77, September 2009.
- [2] Ryoma Bise, Zhaozheng Yin, and Takeo Kanade. Reliable cell tracking by global data association. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1004–1010. IEEE, 2011.
- [3] Robert Canavan, Dr. Claire McCullough, and William J. Farrell. Track-centric metrics for track fusion systems. In *FUSION*, pp. 1147–1154, 2009.
- [4] Mike J. Downey, Danuta M. Jeziorska, Sascha Ott, T. Katherine Tamai, Georgy Koentges, Keith W. Vance, and Till Bretschneider. Extracting fluorescent reporter time courses of cell. *PLoS ONE*, June 2011.
- [5] 佐藤明. HOW TO 分子シミュレーション-分子動力学法, モンテカルロ法, ブラウン動力学法, 散逸粒子動力学法. 共立出版, 2004.