細胞内タンパク質局在の多様性に適合する顕微鏡画像の分類

蕪山典子† 立野玲子‡ 後藤敏行† 影井清一郎† 富樫卓志§ 菅野純夫§ 恒川隆洋¶
横浜国立大学† 東京都臨床医学総合研究所‡ 東京大学医科学研究所§ 富士通株式会社¶

１．はじめに

遺伝子の機能を探るポストゲノム研究において，タンパク質が細胞内小器官のどこに局在するかを知ることは，タンパク質を作る遺伝子の機能推定の手がかりとなる．約 3 万 4 千個と特定されたヒト遺伝子の機能推定にはコンピュータを用いたタンパク質局在部位自動認識システムによる網羅的な処理が望まれている．

我々はこれまで 1 種類の遺伝子を導入した細胞のタンパク質は必ず同じ部位に局在し，その画像パターンは同一であることを前提に研究を進めてきた[1][2]．しかし生物系での研究が進むにつれて，1 種類の遺伝子を導入してもタンパク質の生成や輸送速度には細胞の個体差があり，様々な像が観察されることが明らかになった．そこで本研究では，多様な像が混在する画像に対して高精度な認識を可能とする手法について検討する．

２．タンパク質局在

ヒトのガン細胞由来である HeLa 細胞にタンパク質局在判定対象の遺伝子を導入する際，その遺伝子にあらかじめタンパク質を蛍光発色する機能を組み込んでおくと蛍光顕微鏡下において局在部位の像を観察することができる．図 1 は小胞体(ER)，ゴルジ体(Gol)，細胞膜(Mem)，ミトコンドリア(Mito)，核(Nuc)，ペルオキシソーム(Per)，細胞質(Cyto)の代表的なタンパク質局在像である．しかし図 2 のように 1 種類の遺伝子を導入した場合でも細胞毎に明らかに異なる像が観察された．遺伝子の局在部位は，細胞個々の局在部位だけでなくこれらの像の出現分布によって特徴付けられる．

３．認識アルゴリズム

本研究では固有空間法をベースとした認識手法で画像から検出された細胞個々のクラスを決定した後，全細胞のクラス出現分布を作成し，学習画像から作成したクラス出現分布と比較することで，その画像のタンパク質局在部位を決定する．図 3 に今回提案する学習・認識処理の流れを示す．

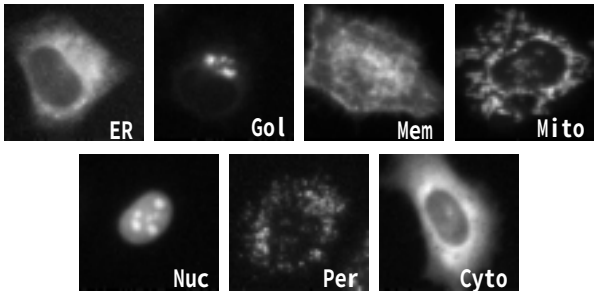

図 1．代表的なタンパク質局在像

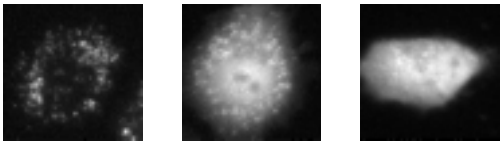

図 2.ペルオキシソーム像の多様性

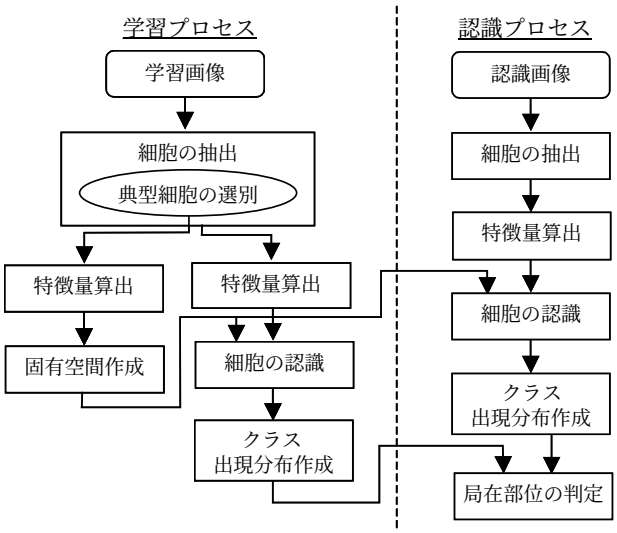

図 3．学習・認識処理の流れ

学習画像では画像内の細胞位置が自動検出され，一定サイズの細胞領域が切り出される．自動検出された細胞には典型的な像やタンパク質輸送途中の像，さらには顕微鏡焦点外のボケた像なども含まれている．典型的な像だけから固有空間を作成するため，選別された細胞領域から算出した特徴量によって共分散行列を構築し，固有空間を作成する．また学習画像から検出した，ボケた像を含む全細胞領域からも特徴量を算出し，固有空間に射影して認識を行いその結果からそれぞれのクラス出現分布を作成する．

認識画像でも，学習処理と同様，画像内の細胞位置を自動検出し，切り出された細胞領域から特徴量の算出を行う．その後固有空間に射影し，まず細胞個々の認識クラスを得，さらに画像内の全

---

*Images classification accommodating multiformity in subcellular localization of proteins

† Noriko Kabuyama, Toshiyuki Gotoh, Seiichiro Kagei (Yokohama National University)
‡ Reiko Tachino (The Tokyo Metropolitan Institute of Medical Science)
§ Takushi Togashi, Sumio Sugano (The Institute of Medical Science, The University of Tokyo)
¶ Takahiro Tsunekawa (Fujitsu Limited)

細胞の認識クラスから画像のクラス出現分布を作成する．学習画像の出現分布と認識画像の出現分布の相関によりタンパク質局在部位の判定を行う．

４．実験

HeLa 細胞に局在部位が既知の 7 種類の遺伝子を注入し約 24 時間培養した後，蛍光顕微鏡に接続した CCD カメラで細胞を撮影した．各遺伝子それぞれ 160 枚の画像(696×520 画素×16bit)を撮影し，学習，認識には共に各 80 枚の画像を使用した．

認識のクラスには 7 種類の典型的な像の局在部位に加え，特にパターンが多様なペルオキシソームはタンパク質が飽和した像のクラス（Per2）を追加した．また細胞培養上避けられない細胞死（アポトーシス）による誤認識を防ぐため，アポトーシスクラスを作成し計 9 クラスとした．

学習画像では画像内に多数存在する細胞の位置を自動検出するため，予め用意した局在検出用モデルとのテンプレートマッチングを行った．相関値の局所最大値を細胞位置とし，細胞位置を中心に平均的な細胞の大きさである 90×90 画素領域を細胞領域として切り出した．

切り出された細胞領域から特徴量(表 1)を算出し正規化した後，各クラスの典型的な細胞像を目視で選別し，それぞれ 63 細胞の学習用細胞の特徴量から共分散行列を求め固有空間を作成した．

認識の固有空間次元数の決定には，学習用細胞を用いて固有空間の次元数を 1～39 次元と変化させた認識実験を行い，最良の認識率が得られた次元数(11 次元)を採用した．

学習画像から検出された全細胞領域について，まずそれぞれの特徴量を固有空間に射影し，マハラノビス距離により認識クラスを決定した．このときマハラノビス距離と固有空間の次元数から得られるカイ二乗確率が全クラスにおいて 0.1 以下の細胞は，誤認識の可能性があるとして棄却した．さらに各クラスの細胞の認識結果からそのクラスの出現分布を作成した(表 2)．

認識画像でも，学習画像と同様に細胞位置を検出した後，細胞領域を切り出し特徴量を算出した．細胞毎に固有空間に射影しマハラノビス距離から認識クラスを決定した．学習画像での出現分布の作成と同様，カイ二乗確率が 0.1 以下の細胞は棄却した．

1 種類の遺伝子を細胞に導入し，培養したウェル容器からは撮影位置を変更すれば複数枚の画像が取得できる．複数枚では撮影・処理時間が増大するが細胞数の増加により認識率の向上が望めるため，今回の実験では同一ウェルから 2 枚の画像を撮影したと仮定した．2 枚の画像内全細胞の認識結果からウェルのクラス出現分布を作成し，学習画像の出現分布との相関を求め最大相関値のクラスをタンパク質局在部位の判定結果とした．表 3 はこのときの実験結果である．

表 1．実験に用いた 39 種の特徴

カテゴリー	数
濃度統計量	7
エッジ要素特徴	12
形状特徴	8
テクスチャ特徴	9
ランレングス統計量	3

表 2．学習画像から作成した出現分布

	クラス出現分布							
入力	ER	Gol	Mem	Mito	Nuc	Per1	Per2	Cyto
ER	0.96	0	0.02	0	0	0.02	0	0
Gol	0.02	0.96	0	0	0.01	0.01	0	0
Mem	0.21	0.01	0.7	0	0.01	0.06	0	0.01
Mito	0.1	0	0	0.65	0.01	0.24	0	0
Nuc	0	0.01	0	0	0.99	0	0	0
Per	0.07	0	0.05	0	0.01	0.27	0.27	0.33
Cyto	0.11	0	0.03	0	0.03	0	0.09	0.74

表 3．実験結果

	局在部位の判定結果							
入力	ER	Gol	Mem	Mito	Nuc	Per	Cyto	再現率
ER	40	0	0	0	0	0	0	1
Gol	0	40	0	0	0	0	0	1
Mem	2	0	38	0	0	0	0	0.95
Mito	1	0	0	38	0	1	0	0.95
Nuc	0	0	0	0	40	0	0	1
Per	0	0	0	0	0	37	3	0.925
Cyto	0	0	0	0	0	0	40	1
適合率	0.93	1	1	1	1	0.97	0.93	

平均再現率：0.975，平均適合率:0.976

５．まとめ

多様な像を見せる細胞内小器官におけるタンパク質局在顕微鏡画像の認識において，細胞個々の認識結果だけではなく画像(ウェル)内の認識クラスの出現分布を用いることにより平均再現率，平均適合率ともに 97%以上を得ることができた．

今後認識対象を，学習画像では認められない部位に局在する可能性の高い局在部位未知遺伝子に広げる予定である．

謝 辞

本研究は，NEDO 委託研究「ゲノム機能解明のための細胞画像自動解析システムの研究開発」によるものであり，関係各位に深く感謝する．

参考文献

[1]蕪山他：ゲノム機能解析のためのタンパク質細胞内局在の自動認識，第 63 回情報処理学会全国大会講演論文集，2001.

[2]蕪山他：細胞内小器官へのタンパク質局在パターンの自動認識，第 8 回画像センシングシンポジウム講演論文集，2002.