

医用画像における細胞認識技術

Cell Recognition Technologies in Medical Images

三村 勇介*
Yusuke MIMURA

尾崎 雄一*
Yuichi OZAKI

一谷 修司*
Shuji ICHITANI

平澤 宏祐*
Kosuke HIRASAWA

要旨

コニカミノルタは、医療・ライフサイエンス分野で注目されている生体分子の標識材料として、従来の蛍光色素より輝度と褪色性に優れた“蛍光ナノ粒子”を開発した。この蛍光ナノ粒子を用いて、細胞あたりに発現する疾患に特定のタンパク質量を高感度に定量解析する技術を構築し、病理標本を解析するサービスを開始した。本サービスの細胞のタンパク質を定量化するという利点は、病理診断分野においては、従来の診断法(Diamino Benzidine [DAB]法など)では判別がつかず、同じ症例として扱っていた患者を層別化することで高い治療効果が見込める治療薬などを選択することが可能となり、特にがん治療において期待が高まっている。また、創薬分野においては、薬が与える影響を対象のタンパク質の発現量で測定することが可能となり、今までより治験の成功確率を上げ、対象検体数を減らすことが期待されている。開発したシステムでは、染色した病理組織標本を顕微鏡で撮影し、その明視野画像から細胞核を抽出し、蛍光画像から蛍光ナノ粒子の輝点を抽出し、それら2つの抽出結果を用いて細胞あたりのタンパク質量を求める。

本システムは、標本中の多数の細胞を正確に解析するため、画像認識により細胞核を抽出する技術と、蛍光ナノ粒子の輝点を抽出し定量する技術とで構成される。細胞核を抽出する技術では、細胞核領域をSupport Vector Machine (SVM)手法を用いた機械学習のスコアを算出し、その後、多段階二値化、細胞核の輪郭部のエッジ抽出、エッジとH-minima手法による重畳核の分割処理を適用することで、個々の細胞を正確に抽出する。また、蛍光ナノ粒子の輝点を抽出し定量する技術では、組織染色で発生する自家蛍光の除去を前処理として行うことで高精度な検出を実現した。

乳がんを対象とした検証では、病理医による指導のもと、作業者の目視による細胞核の抽出結果を正解として精度を算出した結果、細胞核抽出では、F値91%、輝点抽出では、F値90%となった。また、細胞核あたりのタンパク質量では目視判定による正解とシステムによる自動抽出結果の相関係数が0.97、細胞面積あたりのタンパク質量の相関係数は0.99となった。今後は他のがん種へも対象範囲を広げていく予定である。

Abstract

Konica Minolta provides a technology service using fluorescence nanoparticles to quantitatively analyze specific proteins per cell with high sensitivity. In support of that technology service, we have developed fluorescence nanoparticles for the marking of biomolecules whose brightness and resistance to discoloration exceed those of conventional fluorescent pigments.

Konica Minolta's technology service deals with both pathological diagnosis and drug development. In pathological diagnosis, highly effective medicines can be selected for individual patients by classifying patients according to their disease or disorder since the disorder (or disease) was not determined by conventional diagnosis method such as DAB (diamino benzidine) method. This is especially promising in the treatment of cancer.

In drug development, the effects of a drug can be measured by the number of proteins observed, and this increases the probability of successful clinical treatment trials aimed at reducing variety of specimens of interest.

In our cell nucleus extraction processing, we employed three processes FUNCTIONS: machine learning using a support vector machine (SVM), extraction of the contour edge of the nucleus, and segmentation processing using the edge extraction and H-minima methods. In the bright point extraction process, preprocessing removes the auto-fluorescence generated by tissue staining.

In verifying the system as applied to breast cancer, manual calculation of the system's accuracy was 100%, the F measure of cell nucleus extraction was 91%, and that of bright point extraction was 90%. Further, manual extraction and that of our system correlated highly, with a protein per cell nucleus correlation value of 0.97 and an amount of protein per cell area of 0.90.

We are now expanding our work to encompass other types of cancer.

*開発統括本部 システム技術開発センター データサイエンス技術室

1 はじめに

コニカミノルタは、医療・ライフサイエンス分野で注目されている生体分子の標識材料として、従来の蛍光色素より輝度と褪色性に優れた“蛍光ナノ粒子”を開発した¹⁾。この蛍光ナノ粒子を用いて、細胞あたりに発現する疾患に特定のタンパク質量を高感度に定量解析する技術を構築し、2015年7月から病理標本を解析するサービスを開始した。本サービスのタンパク質を定量化するという利点は、病理診断分野においては、従来の診断法(Diamino Benzidine [DAB] 法など)では判別がつかず、同じ症例として扱っていた患者を層別化することで高い治療効果が見込める治療薬などを選択することが可能となり、特にがん治療において期待が高まっている。また、創薬分野においては、薬が与える影響を、対象のタンパク質の発現量で測定することが可能となり、今までより治験の成功確率を上げ、対象検体数を減らすことが期待されている。開発したシステムでは、染色した病理組織標本を顕微鏡で撮影し、その明視野画像から細胞核を抽出し、蛍光画像から蛍光ナノ粒子の輝点を抽出し、それらの2つの抽出結果を用いて細胞あたりのタンパク質量を求める。さらに標本全体の平均値や分布などの統計解析結果を算出し、がんの性状を分子レベルで診断することを可能にしている。

本システムでは、細胞あたりのタンパク質量を定量化するために、明視野画像から細胞核の抽出と、蛍光画像からナノ粒子の輝点の抽出を行う。細胞核あたり、または、細胞核面積あたりに同輝点が何個存在しているのか計算することで、蛍光ナノ粒子スコアと定義した指標を正確に算出する。ここで、細胞核抽出及び輝点抽出を手作業で行うことを考えると、膨大な時間と労力がかかるうえ、作業員間のバラつきから、安定した結果を得ることが難しくなる。よって、染色画像から細胞核と蛍光画像上の輝点を、それぞれ正確に抽出する画像処理技術を搭載した解析システムが必須である。また、システムには、抽出した細胞核と輝点から細胞毎の蛍光ナノ粒子スコアを出力する統計的な解析処理の機能も必要である。

本報では、細胞あたりのタンパク質を定量化するシステムの概要について説明する。そして、システム内で細胞核抽出と輝点抽出に用いている画像認識アルゴリズムを説明し、最後に病理医による指導のもと、作業員の目視による抽出結果と本システムが抽出した結果とを比較検証した内容について報告する。

2 システム

システム全体のフローをFig. 1に示す。細胞あたりのタンパク質量を算出することを目的とした本システムは、細胞核抽出、輝点抽出、蛍光ナノ粒子スコア算出処理の3つの処理機能を備えている。これらの機能について以下で説明する。

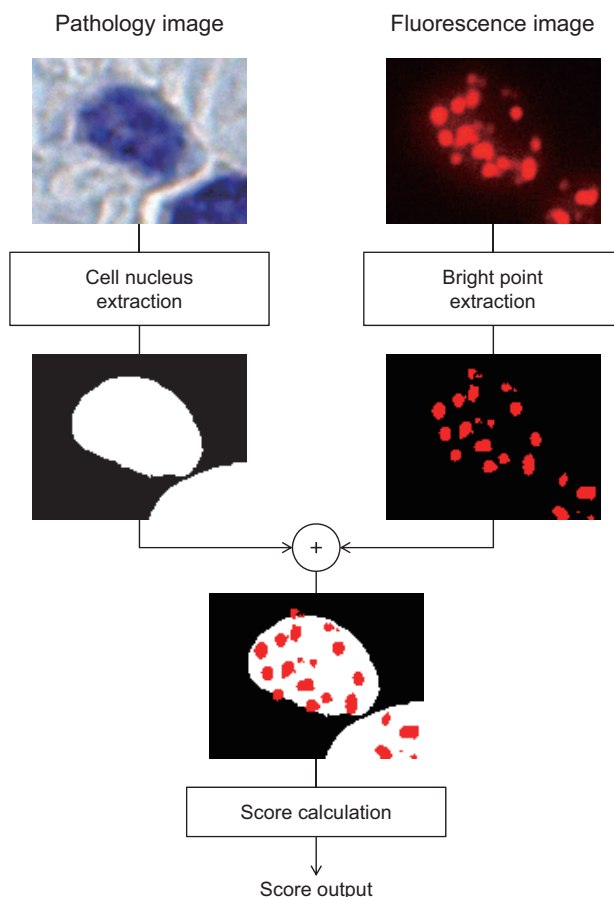


Fig. 1 Quantitative analysis of specific cell nucleus proteins.

2.1 細胞核抽出処理

細胞核抽出処理は、細胞の形状を認識するための明視野画像から色やエッジを用いて、個々の細胞核領域及び位置を抽出する処理である。病理標本の明視野画像には以下の3つの傾向が見られる。

1つ目は、細胞核の染色濃度のバラつきである。病理検体には、様々ながん種や採取法、保管法の異なる検体が存在する。さらに、細胞核も多様であるため、細胞核は全て一様な染色濃度に染色されない。よって、同じ検体内でも染色が濃い細胞核や、薄い細胞核があり、染色濃度の濃淡のバラつきが大きい場合がある。

2つ目は、細胞核の染色のムラ及び中抜けである。細胞核には上述した様に、多様な細胞核が存在している。その中には、細胞核内部が均等に染色されていない染色のムラを有する細胞核や、細胞核の輪郭の一部のみ染色され、他の部分が染色されていない中抜け状に染色された細胞核が存在する場合がある。

3つ目は、細胞核の重畳である。病理検体の細胞核には、細胞核同士が重畳している場合がある。そのため、細胞核と細胞核の境界部分が検出困難となり、個々の細胞核領域に分割し辛い場合がある。

この様に病理標本の明視野画像には、様々な特徴の細胞核が存在する。本システムの細胞核抽出処理では、上述した3つの傾向に対応したアルゴリズムを開発した。開発した細胞核抽出アルゴリズムのフローをFig. 2に示す。

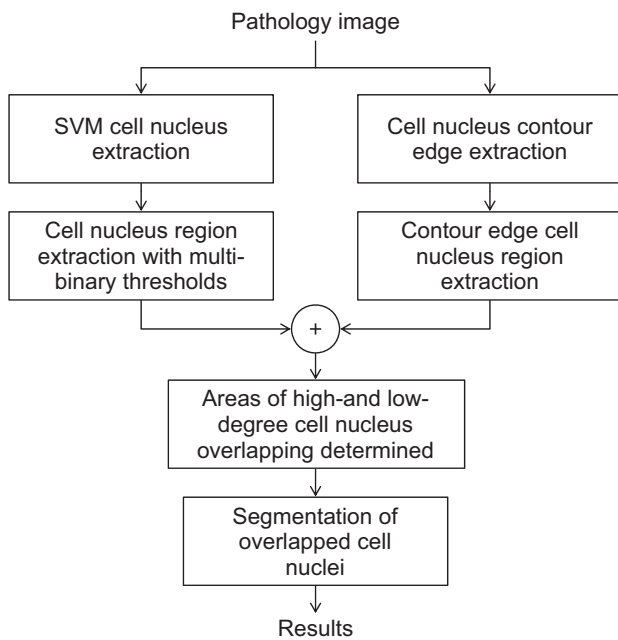


Fig. 2 Flow of the cell nucleus extraction algorithm (SVM: support vector machine).

細胞核抽出処理では、以下の3つの処理を組み合わせで細胞核を抽出する。

1つ目の処理は、多段二値化処理である。本処理は、細胞核の染色の濃淡のバラつきに対応するため実装した処理である。この処理は、事前に明視野画像から抽出対象の細胞核と、それ以外の背景等の領域を、色等の特徴量を用いて学習する。学習には、SVM (Support Vector Machine) 手法による機械学習を用いる²⁾。そして、その学習結果を用いて、病理画像から画素毎のSVMスコアを示したグレースケール画像を作成する。その後、作成したグレースケール画像を複数の閾値を用いて二値化し、二値画像毎に、ラベリング処理を行う。ラベリング領域毎に面積や形状からその領域が細胞核か否かを判定し、条件を満たした領域を細胞核として抽出する。この処理では、複数の閾値を設定することにより、Fig. 3のように染色の濃淡のバラつきが大きい画像でも、細胞核を安定して抽出することができる。

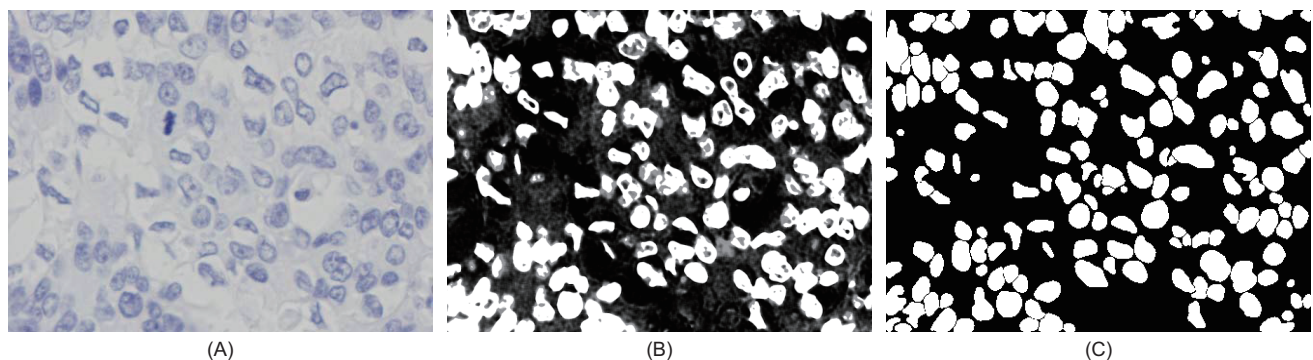


Fig. 3 Cell nucleus extraction using an SVM.

- (A) H stain pathology image of cell nuclei with uneven staining.
- (B) SVM score image in which cell nucleus pixels indicate brightness value.
- (C) Cell nucleus regions extracted.

2つ目の処理は、細胞核エッジの抽出処理である。本処理は、細胞核の染色ムラや中抜けに対応するため実装した処理である。本処理では、まず、細胞核の輪郭部のエッジの抽出を行う。エッジ抽出では、明視野画像で、注目画素から一定の範囲内でのエッジ強度の平均値を求める。その後、注目画素のエッジ強度が平均値より強い場合のみエッジを抽出する。その際、抽出したエッジはその方向も算出する。例として、Fig. 4 (A)で示す様な染色にムラがある細胞核に、本エッジ抽出法を適用した結果をFig. 4 (B)に示す。そして、抽出したエッジの方向及び面積から、抽出したエッジが細胞核の輪郭部かどうかの真偽の判別を行う。Fig. 4 (B)の画像から本手法により、細胞核の輪郭部のエッジを抽出した結果をFig. 4 (C)の画像で示す。その後、抽出した輪郭部のエッジに膨張処理を行った後、エッジで囲まれた閉領域を対象とし、形状から細胞核かどうかの真偽の判別を行い、条件を満たせば細胞核として抽出する。本処理では、エッジを用いることで、Fig. 4 (A)の様な染色にムラや中抜けが発生している細胞核も、Fig. 4 (D)で示すように細胞核を抽出することができる。

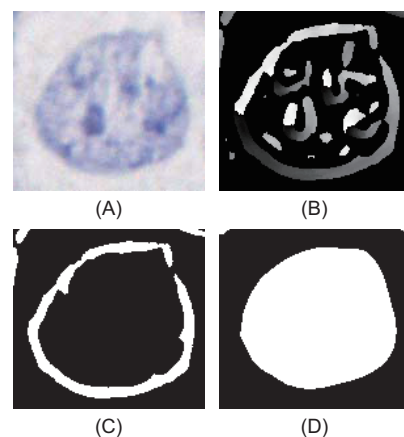


Fig. 4 Cell nucleus contour edge extraction.

- (A) Pathology image.
- (B) Edge angles indicated by pixel values.
- (C) Result of contour edge cell nucleus extraction.
- (D) Cell nucleus region detection.

3つ目の処理は、細胞核分割処理である。本処理は、細胞核の重畳に対応するため実装した処理である。まず、前述した多段二値化処理及び細胞核エッジの抽出処理で抽出した細胞核領域が、個々の細胞核領域なのか、それとも、細胞核同士が重畳している領域かを形状や面積で判断する。その後、重畳している細胞核と判断された領域に対しては、個々の細胞核領域に分割する処理を行う。分割アルゴリズムのフローチャートを Fig. 5 に示す。本分割処理では、多数の細胞核が重畳した領域と、2個又は3個程度の少数の細胞核が重畳した領域、それぞれに対応した分割処理を用いる。まず、多数の細胞核が重畳している領域を、少数の細胞核が重畳した領域まで分割する処理を行う。本分割処理では、重畳した細胞核領域から細胞核エッジの抽出処理で抽出した輪郭エッジを、細胞核と細胞核の境界線として用い、多数の細胞核が重畳した領域を少数の細胞核が重畳した領域へ分割する。例として、Fig. 6 (A) で示す様な細胞核が重畳した領域に、本分割処理を行った結果を示す。Fig. 6 の画像では、Fig. 6 (C) の多数の重畳した細胞核領域から、Fig. 6 (B) の細胞核の輪郭エッジ境界線を差し引いた後に、形成処理を行うことで、Fig. 6 (D) の様な、少数の細胞核が重畳している領域まで分割を行っている。次に、少数の細胞核が重畳している領域を、個々の細胞核領域に分割する処理を行う。本分割処理では、H-minima 手法^{3), 4)}により seed 点を検出して分割を行っている。Fig. 6 では、Fig. 6 (D) の様な少数の細胞核が重畳している領域を、H-minima を用いた分割で Fig. 6 (E) のように個々の細胞核領域まで自動で分割することに成功した。

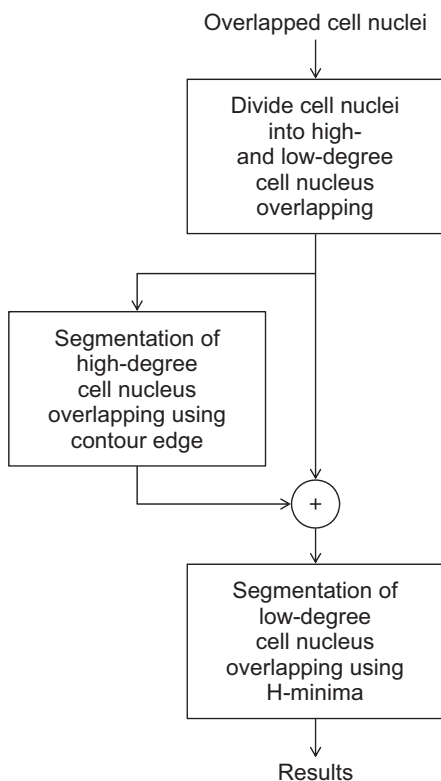


Fig. 5 Segmentation of overlapped cell nuclei.

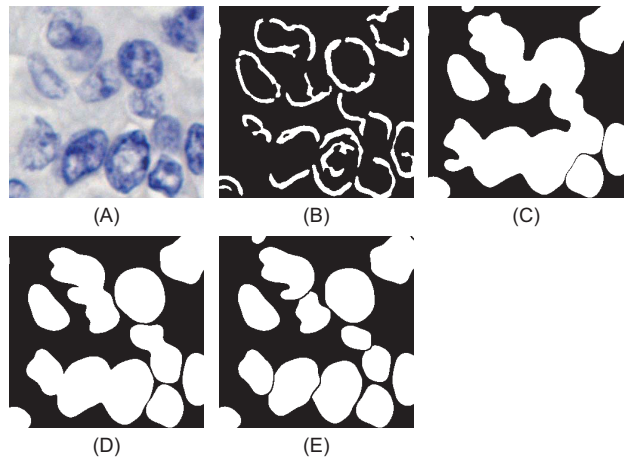


Fig. 6 Contour edge and H-minima methods combined in the segmentation of overlapped cell nuclei.

- (A) Pathology image of overlapped cell nuclei captured.
- (B) Cell nuclei contour edges extracted.
- (C) High-degree overlapping cell nucleus regions extracted.
- (D) High-degree overlapping cell contour-edge segmented.
- (E) Segmentation of low-degree overlapping cell nucleus regions using the H-minima method in each cell nucleus region.

2.2 輝点抽出処理

輝点抽出処理は、蛍光ナノ粒子から発せられている輝点を蛍光画像から抽出する処理である。本処理では、細胞から発生する自家蛍光を、輝点として誤抽出することを防ぐため、自家蛍光を低減するノイズ低減処理を前処理として行う。蛍光画像上では、蛍光ナノ粒子による輝点は高周波成分を有し、自家蛍光は低周波成分を有している。この特徴を利用し、低周波除去を輝点抽出の前処理として行うことで、自家蛍光のノイズを低減している。その後、ノイズ低減を行った画像から以下の手順で輝点抽出を行う。

- (1) 画像から、極大点を検出する。
- (2) 検出した極大点から、最も近い極小点を検出する。
- (3) 極大点と極小点の差を算出し、差が一定以上の極大点を輝点として検出する。

Fig. 7 に輝点抽出処理の一連の画像を示す。

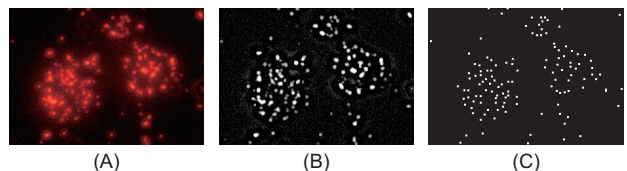


Fig. 7 Bright point extraction.
(A) Fluorescence image captured.
(B) Low-frequencies removed.
(C) Bright points extracted.

2.3 蛍光ナノ粒子スコア算出処理

蛍光ナノ粒子スコア算出処理では、細胞核抽出処理で抽出した細胞核と、輝点抽出処理で抽出した輝点から、蛍光ナノ粒子スコアである細胞核1個あたりの輝点数 Bn (Bright point numbers per cell nucleus) 及び細胞

の面積あたりの輝点数Ba (Bright point numbers per cell nucleus area) を算出する。算出した後、蛍光ナノ粒子スコアをもとに、対象のタンパク質量を換算し定量化する。蛍光ナノ粒子スコアを求める式を以下に示す。

$$Bn [a.u.] = \frac{1}{n} \sum_{i=1}^n CNb(i)$$

$$Ba [\mu m^2] = \frac{\sum_{i=1}^n CNb(i)}{\sum_{i=1}^n CNa(i)} \times 100$$

CNb (i): Bright points that exist in the i-th cell nucleus

CNa (i): Area of the i-th cell nucleus

3 評価条件及び評価結果

3.1 評価条件

開発したシステムの評価条件を Table 1 に示す。評価は、細胞核抽出処理と輝点抽出処理では、病理医の指導により作業者が目視で抽出した細胞核と輝点を正解画像とし、適合率 (P 値)、再現率 (R 値) 及び、適合率と再現率の調和平均 (F 値) を評価値とした。また、蛍光ナノ粒子スコア算出処理は、本システムで自動抽出した結果から算出した値と、正解画像から算出した値との相関係数と傾きを算出し比較した。

Table 1 Evaluation conditions.

Pathology image	Cancer type	Breast
	Stain	Hematoxylin
Fluorescence image	Marker	Ki67
Microscope	Microscope camera	Olympus DP73
	Image size	1600 x 1200
	Microscope power	x40

3.2 評価結果

細胞核と輝点の抽出結果を Table 2 に示す。

Table 2 Evaluation of cell nucleus and bright-point extraction. Note that the three values of the cell nuclei and those of the bright points compare closely to each other.

	Precision value	Recall value	F measure
Cell nuclei	93%	90%	91%
Bright points	89%	90%	90%

Table 2 の結果から、細胞核抽出、輝点抽出共に F 値 90% を超える高い抽出精度であることが分かった。Fig. 8 にシステムで算出した蛍光ナノ粒子スコアと、作業者の目視での抽出結果を用いた蛍光ナノ粒子スコアを比較したものを示す。グラフから、細胞あたりのタンパク質量 Bn と細胞面積あたりのタンパク質量 Ba 共に相関係数 0.97 以上の高い相関結果を得られることが検証できた。特に Ba の結果では、相関係数 0.99、傾き 0.99 となり、システムによる結果でも、目視による結果とほぼ同等の結果を得ることができている。

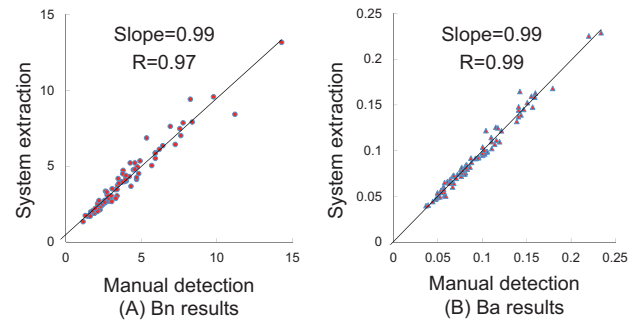


Fig. 8 High correlation of system extraction and manual detection in fluorescence nanoparticle scores Bn and Ba.

4 考察

本システムによる抽出された細胞核と輝点から算出された、細胞あたりの輝点数 Bn 及び細胞面積あたりの輝点数 Ba は、病理医の目視による結果との相関係数が 0.97 以上及び傾きもほぼ 1 であるということがわかった。このことから、本システムによって算出される蛍光ナノ粒子スコアは、目視による抽出とほぼ同じ結果を得られることがわかり、本システムでの自動解析の有効性が確認できた。

5 まとめ

本報では、蛍光ナノ粒子を用いて染色した病理検体の画像解析システムを紹介し、特に、システム内で使われている細胞核抽出処理と輝点抽出処理のアルゴリズムについて詳しく紹介した。本システムの細胞核抽出処理では、明視野画像の染色のムラ、染色の中抜け、細胞核の重畳に対応したアルゴリズムを搭載している。また、輝点抽出処理では、細胞からの自家蛍光の影響を受けないアルゴリズムを搭載している。開発したシステムの性能を評価した結果、細胞核抽出処理及び輝点抽出処理共に、抽出精度で F 値 90% 以上を達成した。また、蛍光ナノ粒子スコアは、病理医の目視による抽出結果と自動抽出結果の相関係数が 0.97 以上とほぼ同等の結果を得られた。このことから本システムによる自動抽出機能を利用して、蛍光ナノ粒子を用いて染色された病理検体の解析が可能であることがわかった。

本報告で検証対象とした乳がん以外への展開を考えると、重畳している細胞核が多い等の、新たな特徴が出てくる可能性がある。今後、それらの特徴に対応することで、新たながん種に対応したシステムを開発していく予定である。

●参考文献

- 1) H.Goda, H.Okada, H.Nakano: Development of a Quantitative Diagnostic Method of Estrogen Receptor Expression Levels by Immunohistochemistry Using Fluorescent Nanoparticles, KONICAMINOLTA Tech.Rep., Vol.11, 68-72 (2014) [in Japanese]
- 2) V. Vapnik and A. Lerner: Pattern recognition using generalized portrait method, Automation and Remote Control, Vol24, 774-780 (1963)
- 3) Jung C, Kim C: Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization, IEEE Trans Biomed Eng., Vol.57 (10), 2600-2604 (2010)
- 4) Luc Vincent: Morphological Segmentation for Textures and Particles, Digital Image Processing Methods, Chapter2, 43-102 (1994)