

G-010

バーチャルスライドの特徴量分布可視化 - 改善 - Visualization of feature distributions in whole slide images - Improvement -

長瀬 瑛人[†]

Akito Nagase

芝浦工業大学[†]

Shibaura Institute of Technology

高橋 正信[†]

Masanobu Takahashi

湘南藤沢徳州会病院[‡]

Shonan Hujisawa Tokushukai Hospital

中野 雅行[‡]

Masayuki Nakano

1. はじめに

病理組織診断は、病理医が病理組織標本を観察して行われるが、その客観化や診断精度向上のため、画像解析による診断支援の研究が進められている。近年では、標本をデジタルデータ化したバーチャルスライド (Whole Slide Image) が普及してきており、遠隔病理診断などに活用されると共に、標本全体を画像解析し診断支援情報を提供する試みが進められている。

我々はこれまで、非癌部との判別が難しい早期肝癌の診断支援を目的として肝組織のバーチャルスライドの核密度や円形度などの特徴量を自動抽出する手法を実現した[1]。核密度の算出にあたっては核の小さな切れ端をどの程度まで核と見做すかで計測される核密度が変動する問題があった。例えば、面積に固定閾値を用いた場合、標準的な核面積が 10%異なると計測される核密度が 11%変化する[2]。核面積は同じ標本中であっても標準偏差で 20%から 30%近く変動するため、面積に固定閾値を用いると核面積の変動がそのまま核密度に反映されてしまう。この問題は小領域ごとに標準的な核面積を求め、その一定割合を各小領域内での面積の閾値とすることで解決した。

本研究ではさらに、従来は固定されていたパラメータの一部を標本あるいは処理単位 (ブロック画像) ごとに自動的に補正することで精度を改善した。また、N/C 比に類似した特徴量として擬似 N/C 比を自動算出する機能を実現すると共に、実用的な処理時間 (3.9 秒/mm², 大きな手術標本でも 1 時間以下) を実現したので報告する。

2. 処理の流れ

処理の大まかな流れは、①処理単位 (ブロック画像) へ分割、②空領域・リンパ球領域を判定し処理対象から除外 (ブロック単位)、③核候補位置抽出、④核輪郭抽出、⑤標準的核面積算出、⑥核の選別、⑦核輪郭の再抽出、⑧空領域・リンパ球領域判定 (ブロックを分割した小領域単位) ⑨特徴量算出、⑩可視化である。

3. エネルギー関数の改善

3.1 エッジ強度による補正

核の抽出と選別にはエネルギー関数を用いている。(1)式は核の自動選別に用いるエネルギー関数である。核輪郭は L 個の輪郭点で表現されており、各項はそれぞれエッジ強度、輪郭点間距離、輪郭線の角度 (曲率) を反映している。

$$E_{radial} = \sum_{l=1}^L (\alpha E_{img} + \beta E_{dist} + \gamma E_{ang}) \quad (1)$$

このうち、最初の項 (E_{img}) は核輪郭部でのエッジ強度に比例するが、図 1 に示すようにエッジ強度は画像により大きく異なるため、輪郭形状が同じであっても核エネルギーが異なるという問題がある。そこで、ブロック画像ごとに

輪郭部の平均エッジ強度を求め、その大きさに反比例するようにブロックごとに α を補正することでこの問題を解決した。



図 1 核の画像例

3.2 核サイズによる補正

距離エネルギー (E_{dist}) は輪郭点間の距離 (d_l) を用いて (2)式で算出され、輪郭点間の距離が (d_{min}, d_{max}) の範囲を越えにくくするように働く。パラメータ d_{min}, d_{max} は従来は固定されていたが、ブロック画像ごとに求めた標準的核サイズに比例するように改善した。

$$E_{dist} = \max(d_l - d_{max}, d_{min} - d_l, 0) \quad (2)$$

4. 空領域・リンパ球領域の判定

核密度の算出にあたっては、肝細胞がほとんど無い領域を除くのが望ましい。従来の手法では組織が無い領域 (空領域) とリンパ球が密集している領域 (リンパ球領域と呼ぶ) を抽出して処理対象外とする手法を実現したが、本研究では標本ごとにパラメータを自動補正などの改善を施した。なお、判定はブロック単位とブロックを $N \times N$ に分割した小領域単位で行う。前者の結果はブロック全体を処理対象から除外するために用いられ、後者の結果は特徴量算出から除外するために用いられる。

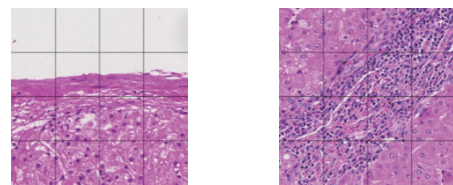


図 2 空領域とリンパ球領域

4.1 空領域の判定

同じバーチャルスライド内では、空領域の画素値はほぼ一定である。そこで、バーチャルスライドを縮小した画像を用い、画素値のヒストグラムを利用して閾値を決め、空領域の判定を行った。具体的には、縮小バーチャルスライドの RGB 各色について最頻値を求め、その一定の割合 (R_{em}) を各色の閾値とする。3 色全てで画素値が閾値以上となる画素を空と判定した。そして、領域内に空と判定された画素が一定割合以上ある領域を空領域とした。今回は、この割合を 30%とした。

4.2 リンパ球領域の判定

リンパ球領域の平均画素値を他と比較すると R の値が低いことがわかった。そこで、基準となる R の画素値を決め、

その一定割合 (R_{ly}) を閾値として空を除く領域の R の平均が閾値以下の場合をリンパ球領域と判定した。基準となる R の画素値は標本全体の色合いによって適応的に設定する必要がある。そこで、縮小バーチャルスライドのうち 4.1 で判定された空領域を除いた領域の R のヒストグラムを求め、その最頻値を基準とした。

5. 擬似 N/C 比

病理組織診断において重要な特徴量として N/C 比がある。これは標本中の細胞核と細胞質の面積比で、一般的に癌部ではこれが高くなる。N/C 比の算出には細胞面積を求める必要があるが、細胞膜の抽出が容易ではないためその自動算出は難しい。そこで、核面積と核密度を組み合わせ、N/C 比に近い指標として用いることのできる特徴量 (擬似 N/C 比) を新たに求めた。擬似 N/C 比は(3)式で算出する。

$$\text{擬似 N/C 比} = \text{標準的核面積} \times \text{核密度} \quad (3)$$

これは領域全体に対する核の面積比に相当するが、細胞質の割合が領域ごとにほぼ一定と仮定すると、N/C 比にほぼ比例する。従って、同じ標本中での N/C 比の相対的な大小関係を知るために利用できると考えている。

6. 実験

6.1 バーチャルスライド

HE 染色された早期肝細胞癌の手術標本 20 枚を NanoZoomer (浜松ホトニクス) を用い 20 倍の対物レンズでデジタル化したバーチャルスライドを用いた。ブロックサイズは $0.29\text{mm} \times 0.29\text{mm}$ とした。

6.2 空領域・リンパ球領域判定の精度評価

空領域の精度評価には、各バーチャルスライドから 1 ブロックずつ選択した 20 枚の画像を 4×4 分割した計 320 領域を用いた。予め作成した正解との誤差が最小となるようにパラメータ R_{em} を最適化した結果、 $R_{em} = 0.92$ のとき空領域の正解率 100%、非空領域の正解率 99.4% が得られた。

リンパ球領域の精度評価には、同様に分割した領域のうちリンパ球 (非リンパ球) 領域のみからなるものを用いた。リンパ球領域の面積の割合を 7% と仮定し、標本全体で正しく識別されるブロックの割合 (面積比正解率) が最大となるようにパラメータ R_{ly} を最適化した結果、 $R_{ly} = 0.81$ のとき面積比正解率 95.3% (リンパ球領域正解率 50.5%、非リンパ領域正解率 98.7%) が得られた。面積比正解率が最大となるように閾値を設定したため、面積比の小さなリンパ球領域の正解率が低めとなったが、特徴量算出から除外するため小領域ごとに判定する場合 (ステップ⑧) はリンパ球領域の正解率が高い方が望ましい。そこで、小領域ごとの判定に用いる閾値は $R_{ly} = 0.92$ (面積比正解率 81%、リンパ球, 非リンパ球領域とも正解率 81%) とした。

6.3 特徴量の精度評価

特徴量の精度評価には癌部、非癌部の肝細胞領域からランダムに 1 枚ずつサンプリングした計 40 枚のブロック画像を用いた。病理診断支援システム[3]を用いて作成した正解との平均絶対比率誤差を表 1 に示す。核密度と円形度の誤差はそれぞれ 6.7%、0.97% であった。エッジ強度補正や核サイズによる補正を行わない場合は誤差は増大しており、ブロック画像ごとのパラメータの自動補正は有効であ

った。また、特徴量算出から除外する核面積の閾値を固定とした場合も核密度の誤差は増大しており、標準的核面積を用いた閾値の設定が有効であることがわかる。

表 1 特徴量の平均絶対比率誤差

	核密度	円形度
エッジ強度補正なし	13.4%	1.09%
核サイズによる補正なし	6.7%	0.99%
核面積の閾値固定	10.4%	1.02%
全て適用	6.7%	0.97%

6.4 特徴量分布可視化

ブロック画像ごとに算出した特徴量の可視化にはイコライズを利用する。予め階級数を決めると、各階級の頻度が均一になるように各ブロックに自動で階級 (色) が割り当てられる。図 3 に原画像と特徴量 (核密度、円形度、核面積、擬似 N/C 比) の分布画像例 (階級数 7) を示す。ブロックの色が暗いほど核密度が高い (円形度が低い、面積が大きい、擬似 N/C 比が高い) ことを示している。シアンは空領域、黄色はリンパ球領域を示している。こうした分布画像により、特徴量が他と異なる領域を容易に知ることができる。図 3 の暗い領域 (癌部) を見ると、同じ結節内であっても特徴量のばらつきが大きい。小領域の特徴量のみを計測する場合は ROI の選択が問題となるが、全体の分布を可視化することで、こうしたサンプリングの問題も解決できる。

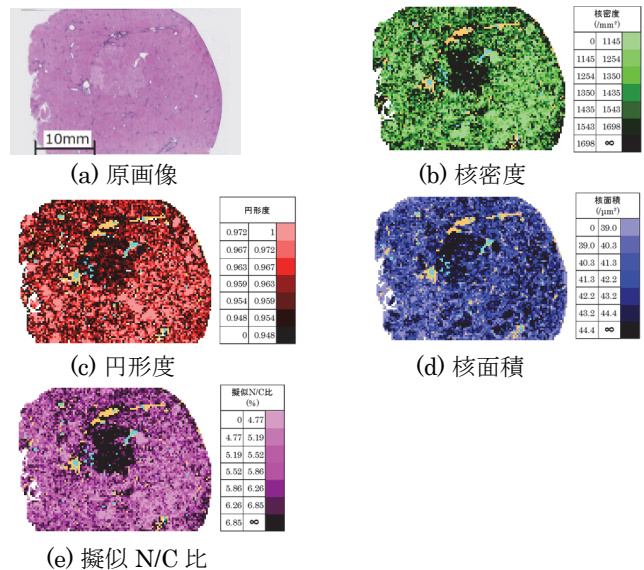


図 3 バーチャルスライドと特徴量分布

7. まとめ

バーチャルスライドの特徴量分布を自動的に可視化する手法の精度を改善するとともに、N/C 比に類似した新たな特徴量 (擬似 N/C 比) の算出を実現した。計算時間 (Core i7-3930K, 6 コア) は、図 3 のバーチャルスライド (空領域・リンパ球領域を除く処理ブロック数: 8210) の場合で約 44 分 (3.9 秒/mm²) であった。

【参考文献】

- [1] 藤井, 他: 2014 年電子情報通信学会総合大会, D-16-1, 2014.
- [2] A. Nagase, et. al.; Bio-Medical Materials and Engineering, vol.26, pp.s1335-s1344, 2015.
- [3] 谷本, 他: 信学技報, MI2009-63, 2009.