

高次局所自己相関特徴による 病理画像からの異常検出手法

栗原 司[†] 野里 博和^{††} 坂無 英徳^{††}
高橋 栄一^{††} 寺井 謙介[‡] 徳山 宣[‡]
蛭田 啓之[‡] 古谷 立美[†]

現在、日本ではがん患者が増加する一方で、がんの確定診断を行う病理専門医が不足しており、見落としや診断ミスの増加が懸念され、大きな社会問題となっている。そこで、病理専門医の負担軽減を目的として、高次局所自己相関(HLAC)特徴を用いた病理診断支援システムを提案してきた。本論文では、病理組織画像を背景・細胞質・核に3値化し、回転・反転を考慮したHLAC特徴を抽出することで正常組織画像を異常としてしまう誤検出の削減に成功した。

Anomaly Detection Technique from Pathology Image using Higher-order Local Autocorrelation Features

Tsukasa KURIHARA[†] Hirokazu NOSATO^{††}
Hidenori SAKANASHI^{††} Eiichi TAKAHASHI^{††}
Kensuke TERAJ[‡] Wataru TOKUYAMA[‡]
Nobuyuki HIRUTA[‡] Tatsumi HURUYA[†]

In this paper we propose a methodology for detecting symptom (cancer) patterns in stomach biopsy cell images. Modern methods suffer from the high false-positive rate. For improving the detection performance, we adopt the modified HLAC features. We revise the prototype of HLAC in order to enhance the robustness to the variations in rotation and symmetry of testing medical image. Subsequently, we apply modified HLAC feature extraction method on marking-three level stomach biopsy cell images. To validate the proposed approach, we conduct experiments and the results verified our method present promising detection performance with high true-positive as well as low false-positive rate.

1. はじめに

現在、日本では毎年64万人以上が新たながん患者と診断されており[1]、今後更に増え続けると言われている。それに伴い、がんの確定診断を行う病理専門医の診断が必要とされる件数も増加し、2005年から2015年にかけて1.7倍になるものと予測されている[2]。しかし、日本における病理専門医の人数は1374人で[3]、人口10万人当たり1人、アメリカでは7.9人しかおらず、深刻な問題となっている。そのため、病理専門医一人当たりの負担が大きくなってきており、見落としや診断ミスの増加が懸念されている。

そこで、病理専門医の負担軽減と診断の質や効率の向上を目指して、病理診断支援システムが提案されてきた。提唱されている方法では、病理組織画像から細胞質を切り出し、核や細胞質の構造的特徴を測定し、がんの自動診断を行っている[4][5]。しかし、これらの手法では、がんの特徴を定義して診断を行っているため、未知のがんは検出できない問題や、病理組織画像から細胞を切り出して診断を行っているため、切り出せない場合は診断することが出来ない問題、切り出された細胞をひとつひとつ診断しているため処理が遅い問題などがあった。

そこで本研究では、この問題を克服する別アプローチとして高次局所自己相関(Higher-order Local Autocorrelation :HLAC)特徴[6]を用いた病理診断支援システムを提案している[7][8]。この手法では細胞を切り出すことなく病理組織画像から特徴抽出をすることが可能であり、特徴量の計算に積和しか使わないため処理速度が速いなどのメリットがある。また、提案手法では、正常な病理組織画像を使い、“正常であること”を学習し、“正常でないこと=異常”として異常検出を行う。実験の結果、がんを含む病理組織画像を異常と判別し、自動診断への適応の可能性を示した。

しかし、幾つかのサンプルが正常な病理組織画像であるにも関わらず、異常と判別されてしまう誤検出が見られた。医師への負担軽減を実現させるためには出来るだけ誤検出は抑える必要がある。

そこで、本論文では病理組織画像を部位ごとに明確化と、病理専門医の視点を反映した特徴抽出が可能となる画像処理法を新たに導入し、誤検出の抑制を試みる。方法としては病理組織画像を背景・細胞質・核の3領域に分割し、各領域の組織の重要度に応じてレベル値で3値化を行う。

検証実験では、あらかじめ病理専門医に診断された胃生検組織画像を用いて、交

[†] 東邦大学
TOHO University

^{††} 独立行政法人産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

[‡] 東邦大学医療センター 佐倉病院
Toho University Sakura Medical Center

差検定法にて検証を行った。がんを含む病理組織画像と正常の病理組織画像を高精度に判別するとともに、誤検出の削減に成功した。

2. 病理診断支援システムと従来手法

2.1 病理診断

現在の医療の中で、がんの確定診断として行われている病理診断は、臨床医により適切な治療を行うために重要な役割を担っている。病理診断では、患者の生死にかかわる“がん”の診断が行われており、その診断の精度・迅速性が最も重要視されている。病理標本は一般的に、ヘマトキシリン・エオジン(Hematoxylin-Eosin:HE)染色により生体組織が染色されている。ヘマトキシリンは青紫色の色素で、病理組織内の核を青紫色に染める。一方、エオジンは淡紅色の色素で、核以外の細胞質、繊維、赤血球などをそれぞれの性質に応じた濃淡各種の赤色に染める。病理医は、病理画像から HE 染色の色や濃淡で各組織を見分け、核の大きさ、核と細胞質のバランス、組織の構築の乱れなどを手掛かりにして、総合的の診断を行っている。

2.2 従来技術の問題点

従来手法では一般的に、病理組織画像から細胞をひとつひとつ切り出し、あらかじめ定義されたがん細胞組織の特徴と照らし合わせ、異常検出を行っている。しかし、がん細胞には様々な種類が存在し、予め全てを定義することは不可能に近い、そのため定義されていない未知のがんは検出することが出来ない問題があった。また、画像中に含まれる膨大な量の細胞をひとつひとつ切り出して診断しているので処理に時間がかかる問題、病理組織標本は、臓器・組織の一部から針やメス、内視鏡によって採取されるため、組織構造の一部が破壊されたり、染色のムラなどにより濃淡が不明瞭だったりする場合は、核や細胞質を切り出すことが出来ない問題が存在した。

Cosato らの報告[9]でも、病理組織画像に含まれる核の 70%しか輪郭の切り出しは行えず、残りの 30%は特徴抽出の対象から漏れてしまっている。つまり、対象から漏れた核の中にがん細胞が含まれていた場合、見落とすことになってしまう。病理診断では組織構造が破壊されたり、輪郭が不明瞭だったりする細胞も分け隔てなく全てを対象として診断を行っている。核や細胞の切り出し処理を必要とする従来の技術では、病理医と同等のレベルで、すべての核や細胞を対象とした自動診断を実現することは難しい。

3. 提案手法

3.1 HLAC 特徴を用いた異常検出

提案手法では、胃生検組織画像を用いた実験より、病理細胞画像から細胞や核を

切り出すことなく特徴抽出を行い、がんの病理組織画像を異常として検出した。しかし、幾つかのサンプルが正常な病理組織画像であるにも関わらず、がんを含む病理組織画像より逸脱度が高くでてしまう誤検出が見られた。

そこで、本論文では、誤検出を抑えるべく、病理組織画像を部位ごとに明確化と、病理専門医の視点を反映した特徴抽出が可能となる画像処理法を提案する。ほうほうとしては病理組織画像を背景・細胞質・核の3領域に分割し、各領域の組織の重要度に応じたレベル値で3値化を行う。

3.2 提案病理診断支援システムの工程

図1に示すように、提案手法は大きく分けて学習フェーズと検査フェーズの2つに分かれている。学習フェーズでは、正常な病理組織画像の特徴を学習する。まず、正常な病理組織画像を本論文で提案する手法で、背景・細胞質・核の3領域にクラスタリングし、それぞれの領域に病理医の視点を反映したレベル値を与える。そして、その3値画像から回転・反転不変 HLAC 特徴抽出[10]を行い、得られた特徴を主成分分析して正常部分空間を得る。

検査フェーズでは、テスト病理組織画像の特徴と、学習フェーズで得られた正常部分空間から逸脱度を計算する。まず、学習フローと同様に、テスト病理組織画像を3領域に分離し、回転・反転不変 HLAC 特徴抽出で特徴を得る。そして得られた特徴と正常部分空間の垂線距離を逸脱度として異常検出を行う。



図1 提案手法フロー

Fig. 1 Flowchart of proposal technique.

3.3 病理組織画像特有の性質を特徴抽出に活かす 3 値化

本論分では、病理組織画像を部位ごとに明確化して、特徴を捉えやすくするために、背景・細胞質・核の 3 領域にクラスタリングし、病理専門医が診断時に着目する組織の重要度に応じて各領域にレベル値を設定し 3 値化する手法を提案する。

3.3.1 3 値化手法

背景・細胞質・核の 3 領域にクラスタリングするために図 2 で示すような、核と背景領域を別々に抽出し、それらの抽出された領域を統合する手法を提案する。

(1) 核領域の抽出

病理組織画像において核領域は濃い青紫色であり、その他の領域はピンクや白といった色になるため、RGB 色空間の R 軸にて R 成分の含有量を比較すると、核領域と他の領域に大きな差があることが予備実験にて判明した。そこで、病理組織画像から核領域を抽出するために、RGB 色空間の R 軸成分を大津の自動閾値選定法 [11]により 2 値化処理を行う。この処理により病理組織画像が核とその他の領域に分割される。

(2) 背景領域の抽出

病理組織画像において背景領域は輝度値が高く、その他の領域はピンクや青紫といった赤成分が強く作用した色情報を持っている。そのため、筆者らは、背景領域の抽出手法として、輝度と赤成分の差分である YUV 色空間の V 成分を用いた抽出手法を報告している [12]。しかし、この手法で分割した画像を病理専門医に確認してもらくと、その他の領域が大きく取られ過ぎてしまっていることが分かった。しかし、染色ムラなどの影響などもあり、固定された色空間の軸を利用して、背景とその他の領域を分離するのは難しい。そこで、画像ごとに空間軸を変化させることができ、統計的なクラスタリングが可能となる主成分分析を用いたクラスタリングを適用する。

方法としては、病理組織画像の RGB 色空間で主成分分析を行い、算出された第 1~3 主成分軸の各主成分得点を 0~255 にスケールする。次に、各軸においてスケールして得られた値を背景領域の値が 0、その他の領域の値が 1 になるように大津の自動閾値選定法にて 2 値化することで 3 つの 2 値画像を生成する。そして、これらを論理和演算することで合成し、背景領域とその他の領域に分離させる。

(3) 画像の統合

核とその他、背景とその他に分離された 2 つの 2 値画像を統合し、互いのその他に分割領域に含まれている細胞質を抽出する。抽出の結果は図 2 に示すような核・細胞質・背景の 3 つの領域に分離された病理画像が生成される。

3.3.2 レベル値の設定

がん細胞の変化は核の大きさ・形・並びに現れやすい。そのため、病理専門医は病理画像診断において核と細胞質の情報を相対的に診断しているが、主に核を重点

的に診ている。そこで病理専門医の視点を反映させるために HLAC 特徴の抽出時に核の情報を重視できるように、背景・細胞質・核に重要度に応じたレベル値を設定する。本論文では、背景領域を 0、細胞質領域を 2 と固定して核を様々なレベル値に設定して予備実験を行った。結果を図 3 のグラフに示す。がんと非がんを最も区別できた {背景, 細胞質, 核} = {0, 2, 10} の組み合わせを採用する。

3.4 HLAC 特徴抽出と回転・反転不変特徴への再構築

HLAC 特徴抽出では 3 値化された対象病理組織画像を $f(x)$ とすると、変位方向 (a_1, \dots, a_N) に対して、 N 次自己相関関数

$$x_N(a_1, \dots, a_N) = \sum_r f(r) f(r + a_1) \dots f(r + a_N) \quad (1)$$

で定義される。

本論文では、1 つの対象画像に対し、高次局所自己相関関数係数の次数を $N = 2$ とし、次変位方向 a を参照点 r の周囲 3×3 画素の領域に限定することで、図に示すような 35 個の局所パターンマスクから算出した 35 次元のベクトル $x_i (i = 1, \dots, 35)$ を HLAC 特徴とする。

3.5 回転・反転不変特徴の再構成

病理組織画像は主に円や曲線状に構成されているおり、上下・左右・反転といった方向的な特徴が必要ないことから、抽出した HLAC 特徴を回転・反転不変 HLAC 特徴として再構築を行う [10]。方法としては、図 4 に示すような、抽出された 35 次元の HLAC 特徴 x_i から、回転や反転を考慮すると同じ相関関係とみなすことができる特徴の線形和を、8 次元の回転・反転不変特徴 y_i に再構築する。

梅山らの報告 [12] では、 $\pi/2$ 回転の回転不変を考慮した従来手法として、2 値画像での 25 次元の HLAC 特徴から 9 次元の不変特徴量の再構成について報告されている。本提案手法では、病理組織画像の構造に最適な $\pi/4$ 回転での回転・反転不変特徴として、35 次元の HLAC 特徴 $x_i (i = 1, \dots, 35)$ から 8 次元の回転・反転不変特徴 $y_i (i = 1, \dots, 8)$ を再構成する。

3.6 主成分分析を用いた正常部分空間の形成

本論文では、正常な病理組織画像を使い、正常であることを学習し、“正常でないこと = 異常” として異常検出を行つために、正常な病理組織画像の特徴から主成分分析を用いて正常部分空間 [13] を形成する。正常部分空間とは、学習用の正常の病理組織画像から抽出した特徴の固有ベクトルで張られる空間である。

まず、回転・反転不変特徴ベクトル y_i に主成分分析を適用し、正常部分空間を形成する主成分ベクトルを求める。この主成分ベクトルは、特徴ベクトル y_i の自己相関行列 R_y の固有値問題を解くことにより求めることができる。

$$R_y U = U \Lambda \quad (2)$$

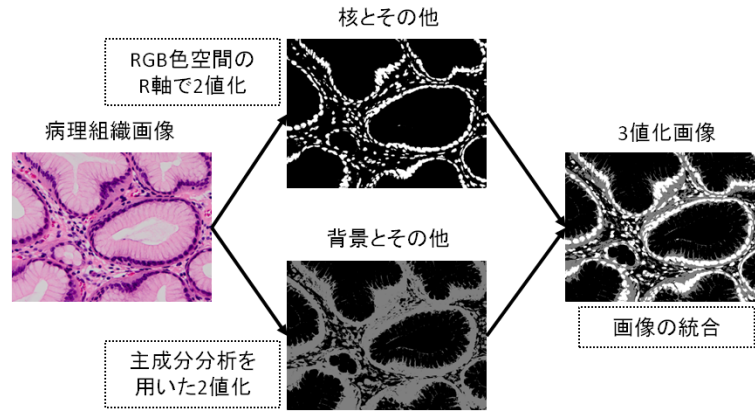


図2 3値化手法

Fig. 2 Method of making to three values.

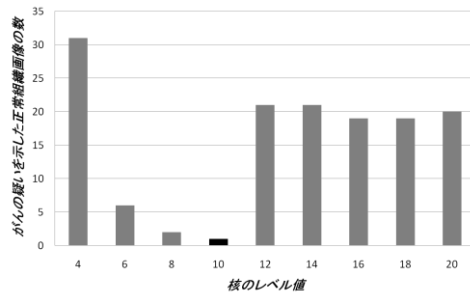


図3 予備実験：核のレベル値選定実験

Fig. 3 Preliminary experimental result

ここで、 U は固有値ベクトルを列とする行列であり、 Λ は固有値を対角要素とする対角行列である。

次に、固有ベクトルの中から正常部分空間を形成する主成分ベクトルとなる次元数 K を決定する。次元数 K は、主成分がどれだけ分析したデータの情報を表現することに寄与しているかを数値化した累積寄与率

$$\eta_K = \frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^M \lambda_i} \quad (3)$$

によって決定される。累積寄与率条件 C としたとき、正常部分空間は、累積寄与率 $\eta_K \geq C$ を満たす次元 K までの固有ベクトル $U_K = \{u_1, \dots, u_K\}$ を基底ベクトルとして定義される空間とする。

3.7 逸脱度による異常検知

本論文では、学習過程によって得られた正常部分空間を用いて、テスト病理組織画像から抽出した特徴と正常部分空間との垂線距離を逸脱度として異常検出の指標とする。この逸脱度は、正常部分空間の直交補正空間への射影成分の距離として次のように算出することができる。

正常部分空間への射影子は、 $P_{\perp} = U_K U_K^T$ として表され、それに対する直交補正空間への射影子は、単位行列を I とすると、 $P_{\perp} = I - P$ と表せる。このとき、直交補正空間への射影成分 d_{\perp} は、

$$d_{\perp}^2 = \|P_{\perp} y\|^2 = y^T (I - U_K U_K^T) y \quad (4)$$

と表すことができる。

この距離 d_{\perp}^2 を逸脱度として、予め設定する閾値 H より逸脱度が高ければ異常、低ければ正常とし、異常検出する。

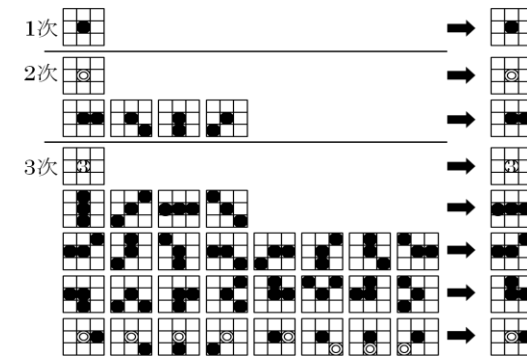


図4 回転・反転不変特徴への再構築

Fig. 4 Reconfiguration of features from HLAC features.

4. 検証実験

本論文では、提案する3値化の有効性を検証するために胃生検組織画像を用いて交差検定法にて検証実験を行った。

4.1 実験画像

実験画像として、予め病理医に診断された明らかな正常組織画像169枚と明らかながん組織画像6枚を用いた。検証実験で実験画像は顕微鏡倍率20倍で撮影され、 2560×1920 のjpeg形式で保存されたものである。

実験では画像を4分割して利用した。よって、1枚の画像から 1280×1920 の画像が4サンプル生成するので正常組織画像は676サンプル(169セット \times 4枚)、がん組織画像は24サンプル(6セット \times 4枚)となる。

4.2 実験と評価の方法

本実験は、画像データによる推定精度の偏りをなくするために、テスト用の正常組織画像を1セットずつ入れ替えながら169回繰り返す169分割交差検定法で行った。

評価方法として、169分割交差検定法において毎回、学習画像の逸脱度の平均値(μ_j)と標準偏差(σ_j)を算出し、変数 z を設け、 $H_j = \mu_j + z\sigma_j$ を計算しておき、 H_j の平均値を閾値 H として用いた。変数 z の値を変えることで閾値 H を可変することができる。テスト画像に対して、この閾値を越えた逸脱度の場合を、がんの疑いを示す異常として検出する。

本実験の評価では、テスト用がん画像がすべて異常と判別されている閾値において、テスト用正常組織画像の逸脱度が閾値を越えた場合を誤検出としてカウントし、誤検出数の比較により各手法の評価を行った。

4.3 検証実験 3値化手法の有効性の検証

提案する3値化手法の有効性を検証するために、グレースケール画像、YUVのV成分を用いた従来の3値化画像、主成分分析を用いた3値化画像の比較実験を行った。比較は、テスト用正常組織画像の逸脱度と、テスト用がん組織画像からは169分割交差検定法を行ったときの各サンプルの最小値を比較して行う。累積寄与率条件の C は、0.999, 0.9999, 0.99999の3つの条件のうち各手法において最もよい条件の結果を比較した。

図5に原画像と、目視による比較を行うため分割領域を判別しやすく変換した各手法の領域分割画像を示す。なお、提案手法の画像では、背景、細胞質、核の画素値をそれぞれ0, 127, 255として表示している。グレースケールは、原画像に最も近く、見た目でも組織の構造が明瞭である。YUVのVを用いた3値化手法では、細胞質領域が大きく取られすぎてしまっている事がわかる。提案手法では、核周りの細胞質領域を中心とした抽出に成功している。病理専門医による検証でも、提案手法の方が細胞質領域をより診断しやすく分割していると評価を受けた。

表1と図6に各手法を用いた検証実験の結果を示す。閾値は、 $\mu_j + 0\sigma_j$, $\mu_j + 1\sigma_j$, $\mu_j + 2\sigma_j$, $\mu_j + 3\sigma_j$ の4つの条件のうちテスト用がん画像がすべて異常と判別できる最大の物と比較した。表1が示す結果から、主成分分析を用いた3値化から特徴を抽出することで誤検出数が大幅に減少することが確認できる。図6は横軸がサンプル番号で縦軸が逸脱度のグラフである。グラフを確認しても、V成分を用いた3値化手法に比べ、主成分分析を用いた3値化の結果の方が、正常組織画像とがん組織画像の逸脱度に大きな差異があることが確認できる。よって、提案手法の方が病理診断支援向きの画像処理法であると言える。

5. おわりに

本論文では病理組織画像からの異常検出における誤検出を抑えることを目的として、病理組織画像の細胞質領域をより原型に近い形で抽出する3値化処理法を提案した。提案手法では病理組織画像を背景・細胞質・核の3つの領域にクラスタリングして、病理専門医の視点を反映したレベル値としてそれぞれの領域に設定した。

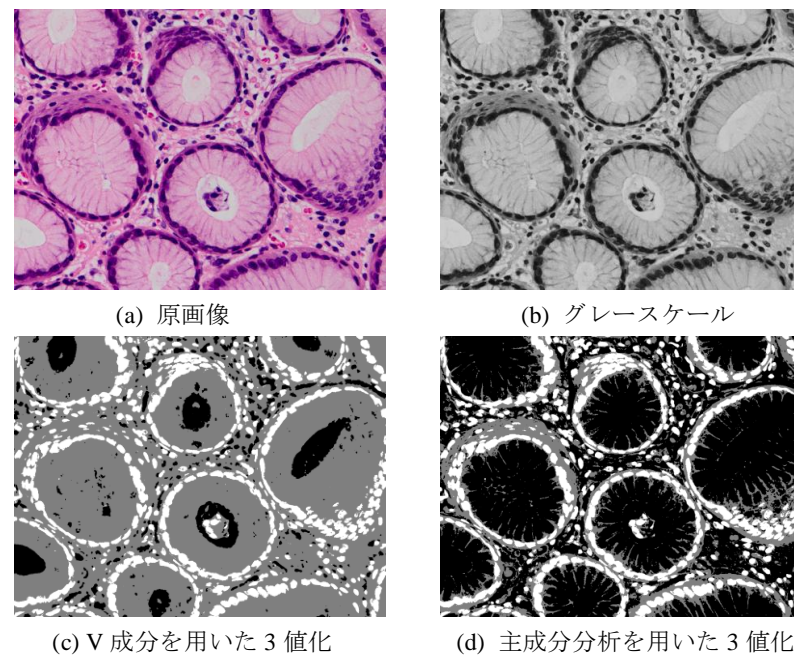
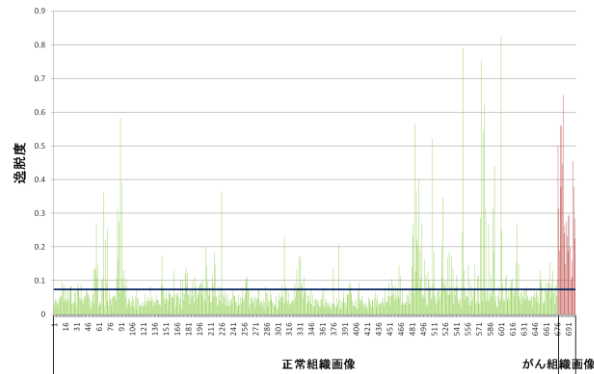


図5 画像処理法の比較

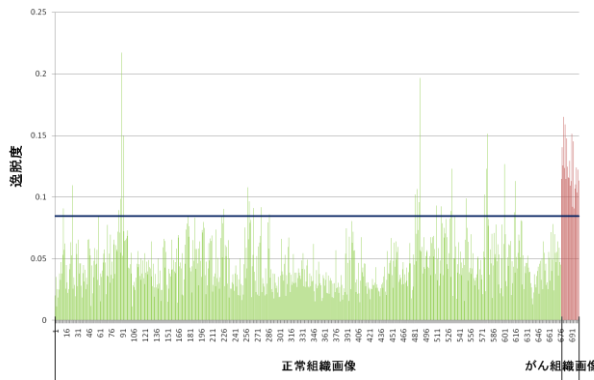
Fig. 5 Comparison of image processing methods.

画像処理手法	利用した閾値	誤検出数
グレースケール	$\mu + 0\sigma$	263
V成分を用いた3値	$\mu + 0\sigma$	186
主成分分析を用いた3値	$\mu + 2\sigma$	27

表1 各手法における誤検出数
Table. 1 False positive results.



(a) V成分を用いた3値化手法の実験結果



(b) 主成分分析を用いた3値化の実験結果

図6 検証実験の結果

Fig. 6 Result of experiment.

参考文献

- 1) Matsuda T, Marugame T, Kamo K, Katanoda K, Ajiki W, Sobue T, and The Japan Cancer Surveillance Reserch Group. Cancer incidence and incidence rates in Japan 2003: based on data from 13 population-based cancer registries in the Monitoring of Cancer Incidence in Japan (MCIJ) project. Japanese Journal of Clinical Oncology, 39: 850-8, 2009
- 2) 厚生労働省がん研究助成金「がん生存者の社旗的適応に関する研究 平成14年度研究報告」
- 3) 厚生労働省大臣官房統計情報部「平成20年(2008)医師・歯科医師・薬剤師調査の概況」
- 4) C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," Technical Report, TR-05-09, Rensselaer Polytechnic Institute, Mar.2005.
- 5) E. Cosatto, M. Miller, H. P. Graf and J. S. Meyer, "Grading Nuclear Plemorphism on Histological Conference in pattern Recognition (ICPR2008), pp.1-4, Tampa, USA, Dec.2008"
- 6) N. Otsu and T. Kurita, "A new scheme for practical flexible and intelligent vision systems," Proc. IAPR Workshop on Computer Vision, pp.431-435, Tokyo, Japan, Oct.1988.
- 7) 野里博和, 坂無英徳, 村川正宏, 樋口哲也, 大津展之, 寺井謙介, 蛭田啓之, 亀田典章, "高次局所自己相関特徴を用いた病理組織診断支援技術の研究," 情報処理学会研究報告 MPS-73, vol.2009, no.19, pp.161-164, Mar.2009.
- 8) 栗原司, 野里博和, 坂無英徳, 村川正宏, 古谷立美, 樋口哲也, 大津展之, 寺井謙介, 蛭田啓之, 亀田典章, "高次局所自己相関特徴を用いた病理画像からの異常検出," 2009年映像情報メディア学会冬季大会講演予稿集, p5-2, Dec.2009.
- 9) E. Cosatto, M. Miller, H. P. Graf and J. S. Meyer, "Grading Nuclear Pleomorphism on Histological Micrographs," Proc. of the 19th International Conference on Pattern Recognition (ICPR2008), pp.1-4, Tampa, USA, Dec.2008.
- 10) 野里博和, 栗原司, 坂無英徳, 村川正宏, 小林匠, 古谷立美, 樋口哲也, 大津展之, 寺井謙介, 蛭田啓之, 亀田典章, "病理画像からの異常検出における高次局所自己相関特徴の抽出手法に関する研究," 第13回画像の認識・理解シンポジウム論文集, pp.1739-1746, July.2010.
- 11) 大津展之, "判別および最小2乗基準に基づく自動しきい値選定法," 電子通信学会論文誌, vol.J63-D, no.4, pp.349-356, Apr.1980.
- 12) 梅山神二, 栗田多喜夫, "高次局所自己相関特徴を用いた図形の回転不変特徴の構成," 情報処理学会第45回(平成4年後期)全国大会講演論文集, 2, pp.323-324, Mar.1993
- 13) Takuya Nanri and Nobuyuki Outu, "Unsupervised Abnormality Detection in Video Surveillance," Proc. IAPR Conf. on Machine Vision Application (MVA2005), pp.574-577, Tsukuba, Japan, May.2005.