

ロバストカラーヒストグラムを用いた生体画像からの組織領域抽出

本多 英晴^{*†}, 横田 秀夫[†], 覚正 信徳[†], 吉田 学^{†#},
三島 健稔^{†#}, 姫野 龍太郎^{†#}, 大竹政光^{*†}

^{*} 明治大学大学院 理工学研究科
神奈川県川崎市多摩区東三田 1 - 1 - 1
e-mail: hideharu@riken.jp

[†] 理化学研究所 生体力学シミュレーション特別研究ユニット
埼玉県和光市広沢 2 - 1

[#] 埼玉大学 理工学部
埼玉県さいたま市桜区下大久保 2 5 5

1. はじめに

我々は3次元内部構造顕微鏡(3D-ISM)[1][2]から得られる生体フルカラー断面画像から、特徴量を色情報に限定した抽出法の研究を行っており、その有効性を示してきた[3]. 現在専門家の手作業に依存している領域抽出を自動化若しくは半自動化することにより、コンピュータ上に定量化された生体のモデルを構築が実現し、自由な視点での観察が可能となる. しかし、生体カラー画像においては抽出に有効な特徴量が不明であるため、抽出手法が確立していない. そこで、我々は色情報に特徴量を限定した組織領域抽出を目的として研究を行っている.

本稿では、我々が提案した、 $L^*a^*b^*$ 空間における教師画像から得られるカラーヒストグラムから、判別を行うために必要な出現頻度を推定したロバストカラーヒストグラムとそれを用いた抽出法及び、推定を制御するパラメータの設定方法について述べる.

2. カラーヒストグラム

本稿で用いるカラーヒストグラムとは、「画像の色を離散化し、それらの色が画像に出現する回数を数えたもの」[4]である. カラーヒストグラムは特定の色空間での表現を示すものではなく、多種の色空間での表現が可能である. カラーヒストグラムの例を図 1 (巻頭カラー図 1) に RGB 色空間での表現, 図 2 (巻頭カラー図 2) に $L^*a^*b^*$ 色空間での表現で示す. 図における色は出現頻度の高さを表している.

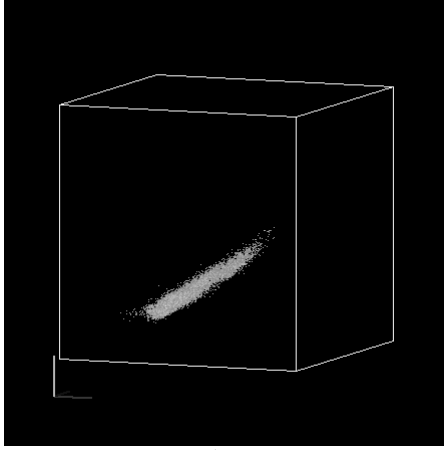


図 1 RGB 色空間での表現

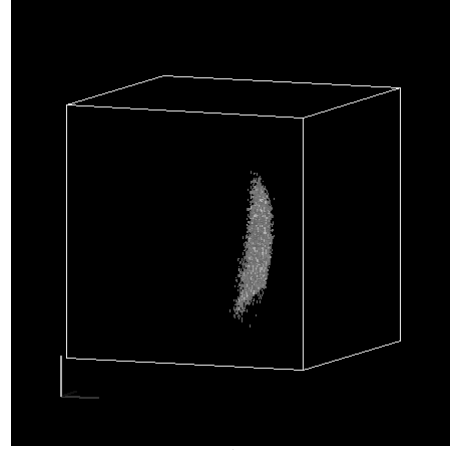


図 2 L*a*b*色空間での表現

3. ロバストカラーヒストグラムの頻度比較による抽出

3.1. カラーヒストグラムの頻度比較による抽出

本稿で提案した手法は，カラーヒストグラムにおける各画素の画素値の出現頻度を確率として扱い，各領域に出現する確率が高い画素値を選択することによる抽出である．これは，カラーヒストグラムの出現頻度を識別関数として用いて画素を抽出することに他ならない．ここで用いるカラーヒストグラムは後述のロバストカラーヒストグラムである．以下に出現確率の比較による組織領域抽出を詳述する．

画像に C 個の領域(クラス) w_i ($1 \leq i \leq C$) があるとき，画素値 x が w_i に出現する確率を $p_i(x)$ とする． $p_i(x)$ はカラーヒストグラムより計算する．画素値 x が w_i であると判別する識別関数を $g_i(x)$ とする． $g_i(x)$ は式(1)で表される．

$$g_i(x) = p_i - \sum_{j=1, j \neq i}^C p_j(x) \quad (1)$$

ただし， $\int p_i(x) = 1$

抽出のために 2 値化を行うには， $g_i(x) > 0$ である画素値 x を選択する． 3 クラスの例を図 3，図 4 に示す．図 3 は，カラーヒストグラムから得られる各クラスにおける画素値の出現確率を表したものである．図 4 は式(1)による識別関数であり，矢印で示した範囲が各領域として抽出される画素値である．このように選択された画素値を画像から選び出すことで，領域抽出を行う．

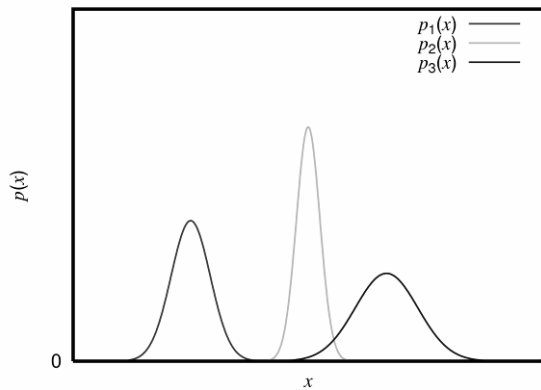


図 3 出現確率

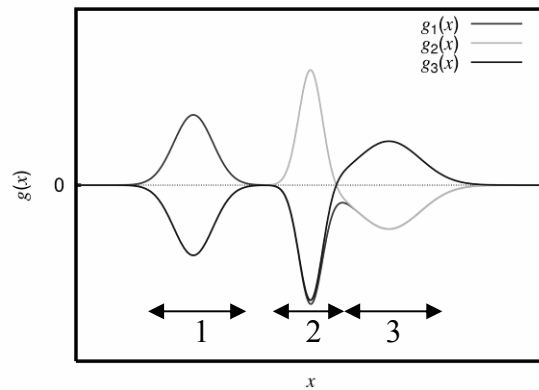


図 4 識別関数と抽出範囲

3.2. ロバストカラーヒストグラム

本稿で提案したロバストカラーヒストグラムとは、少数のサンプリングから判別、抽出に必要な色の出現頻度を推定したカラーヒストグラムである。教師画像から得られる教師カラーヒストグラムはサンプリングが少数であるため、判別を行うためにはサンプリングから抽出する全領域の色の出現頻度を推定する必要がある。そこで、カラーヒストグラムの出現頻度を用いて推定を行うことを提案し、推定したものをロバストカラーヒストグラムと呼ぶ。

3.2.1. ロバストカラーヒストグラムの推定

生体画像のカラーヒストグラムの分布は不明であるため、本論文ではロバストカラーヒストグラムの推定手法としてノンパラメトリック推定手法である Parzen Window 法[5][6]を用いる。Parzen Window 法とは、KB (Kernel-Based) 法とも呼ばれ、確率密度分布 $p(x)$ を式(2)のように推定する。

$$\tilde{p}(x) = \frac{1}{N} \sum_{i=1}^N g(x - x_i) \quad (2)$$

ここで、 $g(x)$ は核関数（カーネル）と呼ばれる密度関数であり、この関数の重ね合わせにより推定を行う。核関数には一様関数やガウス関数などが用いられる。生体画像の組織領域のカラーヒストグラムにおいて、出現頻度は出現頻度分布の中心に近いほど高く、中心から外れるほど出現頻度は低くなる。また、出現頻度が高い色に近い色はその出現頻度も高い。このことから、核関数 $g(x)$ には、 x に近いほど値が大きく、 x から遠いほど値が小さくなるガウス関数を用いることが適当だと考えられるため、本稿ではこれを用いる。

核関数にガウス関数に用いて推定を行う確率密度分布を式(3)に示す。図 5 にガウス関数のグラフを示す。

$$\tilde{p}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\mathbf{x} - \mathbf{x}_i)^2}{2\sigma^2}\right\} \quad (3)$$

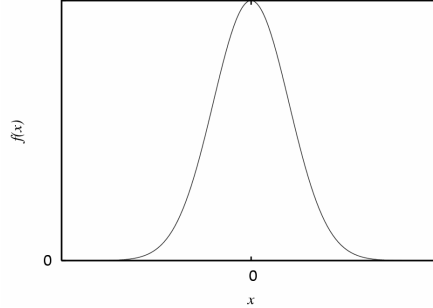


図 5 ガウス関数

式(3)における $\mathbf{x}$ は画素値， $p(\mathbf{x})$ は出現頻度（確率）として扱う．以上から，ロバストカラーヒストグラムの画素値 (l, m, n) の出現頻度 r_{lmn} と，教師カラーヒストグラムの画素値 (i, j, k) の出現頻度 t_{ijk} の関係は式(4)で表される．図 6 にロバストカラーヒストグラムの推定の様子を示す．黒棒がサンプリングにより得られた出現頻度，赤線が各出現頻度について核関数により得られた出現頻度，緑線が赤線の重ね合わせにより得られたロバストカラーヒストグラムの出現頻度を示している．

$$r_{lmn} = \frac{1}{N} \sum \frac{t_{ijk}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{d((l, m, n), (i, j, k))^2}{2\sigma^2}\right\} \quad (4)$$

ただし， N は TCH の色数， $d((l, m, n), (i, j, k))$ は色差を表す

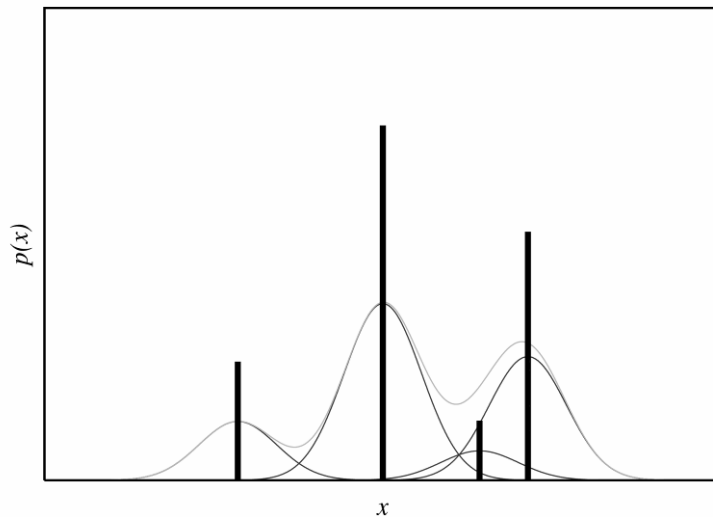


図 6 ロバストカラーヒストグラムの推定

3.2.2. パラメータ σ の影響及び設定法

ここで、式(3)の σ について述べる． σ は推定を制御するパラメータであり、 σ が過大の場合はバイアスが大きくなりよい推定がされず、一方で σ が十分に大きくない場合は個々の学習データに強く依存することになり、この場合もよい推定がされない．このパラメータを決定するためには、未知の真の確率密度分布 $p(\mathbf{x})$ と推定した確率密度分布 $\tilde{p}(\mathbf{x})$ との距離を小さくすることが必要になるが、 $p(\mathbf{x})$ は未知であるためにその距離を求めることができない．

本稿では、実画像を用いてパラメータ σ の設定を行う．本手法の検討を実験的に行った結果では、ロバストカラーヒストグラムの推定に複数の中から特定の σ を用いた場合に一致度が最大となる結果が得られている[3]．このことから、妥当な推定が可能である σ が存在していることが考えられる．カラーヒストグラムの分布が真の確率密度分布である場合、断面にかかわらずこのカラーヒストグラムを用いた抽出は有効であることが予想できる．ここで、断面にかかわらず高精度な抽出が可能である特定のロバストカラーヒストグラムが存在すると仮定する．このロバストカラーヒストグラムの発見を目指し、抽出の試行により σ を設定する．複数の σ で推定した複数のロバストカラーヒストグラムについて教師画像とは別の正解画像を評価に用い、抽出結果を式(5)に示す正解画像との一致度を算出する．目指すロバストカラーヒストグラムはどの断面においても高い一致度で抽出できるため、評価画像においても高精度で抽出できるはずである．そこで、最大の一致度で抽出が可能であるロバストカラーヒストグラム及び σ を本稿では妥当に推定できたものとして採用する．なお、一致度は以降の実験において抽出結果の評価に用いる．

$$\text{一致度} = \frac{n(X \cap T)}{n(X \cup T)} \times 100$$

(5)

X : 抽出結果領域 T : 正解領域
n(Z) : 領域内の画素数

実験において上述のように仮定した、断面にかかわらず有効に抽出できる特定のロバストカラーヒストグラムが存在の検証を行う．

3.2.3. 色空間の選定

提案手法では、どの色空間での表現のカラーヒストグラムでも用いることが可能である．本論文では、 $L^*a^*b^*$ 色空間のカラーヒストグラムを用いて抽出を行う．RGB色空間では、デバイスにおける各信号の強さを表しているだけで、色差の定義がされていないために信頼できる色差を用いることができない．一方、 $L^*a^*b^*$ 色空間においては、人間の知覚に基づいて色の差異の大きさが色空間での距離と一致するように定義されており、色差は信頼できるものである．以上の理由により、 $L^*a^*b^*$ 色空間を採用する．

3D-ISM から得られる画像の画素値は、NTSC 信号の CCD カメラで撮影した後にコンピュータに取り込んだ RGB 値である．この RGB 値を Adobe 社の NTSC ICC プロファイルに従い、式(6)のように $L^*a^*b^*$ 値へ変換し、四捨五入により整数に離散化したものを用いる[7]．

$$L^* = 116 \left[f \left(\frac{Y}{Y_n} \right) \right] - 16 \quad a^* = 500 \left[f \left(\frac{X}{X_n} \right) - f \left(\frac{Y}{Y_n} \right) \right] \quad b^* = 200 \left[f \left(\frac{Y}{Y_n} \right) - f \left(\frac{Z}{Z_n} \right) \right] \quad (6)$$

ただし、

$$f(X) = \begin{cases} X^{\frac{1}{3}} & \text{if } X > 0.008856 \\ 7.7870X + \frac{16}{116} & \text{otherwise} \end{cases}$$

$$X_n = 0.964218 \quad Y_n = 1 \quad Z_n = 0.824921$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.634415 & 0.185226 & 0.144577 \\ 0.310974 & 0.591614 & 0.097412 \\ -0.001175 & 0.055557 & 0.770538 \end{bmatrix} \begin{bmatrix} \text{linear R} \\ \text{linear G} \\ \text{linear B} \end{bmatrix}$$

$$\text{linear R} = \left(\frac{R}{255} \right)^{2.199219} \quad \text{linear G} = \left(\frac{G}{255} \right)^{2.199219} \quad \text{linear B} = \left(\frac{B}{255} \right)^{2.199219}$$

式(4)における色差 $d((l, m, n), (i, j, k))$ は $L^*a^*b^*$ 空間においては、 $L^*a^*b^*$ 表色系での色差 ΔE_{ab}^* とする． ΔE_{ab}^* は式(7)で表される[8]． (L^*_1, a^*_1, b^*_1) と (L^*_2, a^*_2, b^*_2) との色差 $d((L^*_1, a^*_1, b^*_1), (L^*_2, a^*_2, b^*_2))$ は、式(8)で表される[8]．

$$\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad (7)$$

$$d((L^*_1, a^*_1, b^*_1), (L^*_2, a^*_2, b^*_2)) = \sqrt{(L^*_2 - L^*_1)^2 + (a^*_2 - a^*_1)^2 + (b^*_2 - b^*_1)^2} \quad (8)$$

4. 実験

4.1. 基礎実験

本節では、ロバストカラーヒストグラムを用いた抽出手法について、人工画像を用いてその有効性を検証する．4.1.1では、閾値抽出ができない画像についてカラーヒストグラムの比較による抽出の有効性、4.1.2では、ロバストカラーヒストグラムの推定に関する有効性を検証する．

4.1.1. 基礎実験 1

図 7 (巻頭カラー図 3) に対象画像 (256×256 pix.) を示す．関心領域を左側 128×

256 pix. とする．この画像の画素値分布は，関心，非関心領域ともに $0 \leq L \leq 100, -128 \leq a^* \leq 127, -128 \leq b^* \leq 127$ となっている．このような画像では，適切に閾値を設定することができないため，閾値を設定する手法では関心領域のみを抽出することができない．

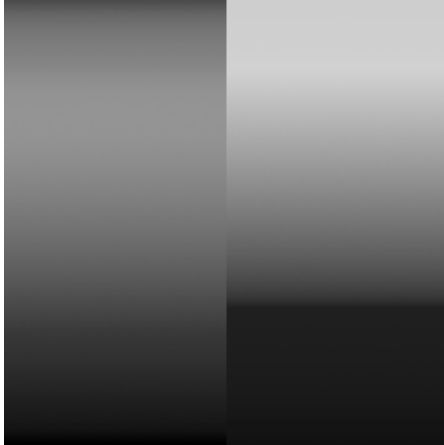


図 7 基礎実験 1 対象画像

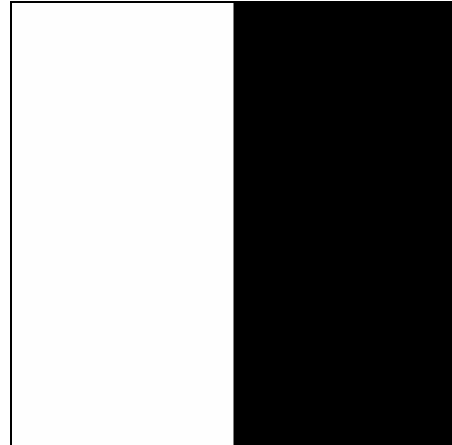


図 8 提案手法抽出結果

正解画像の関心領域，非関心領域からカラーヒストグラムを作成し，提案手法であるカラーヒストグラムの比較を用いて，関心領域を抽出した．結果を図 8 に示す．白が関心領域，黒が非関心領域と判別されたことを表し，一致度は 100%であった．このことから，領域内の出現頻度を正しく表現しているカラーヒストグラムを与えることが可能であれば，精度の高い抽出ができることが示された．

4.1.2. 基礎実験 2

図 9 (巻頭カラー図 4) に対象画像 (256×256 pix.) を示す．中央部の領域 128×128 pix. を関心領域とする．関心領域の $L^*a^*b^*$ 値はそれぞれ $N(60, 1), N(32, 5), N(32, 5)$ ，非関心領域の $L^*a^*b^*$ 値はそれぞれ $N(40, 1), N(16, 5), N(16, 5)$ に従っている． $N(\mu, \sigma^2)$ は，式(9)に示される平均 μ ，分散 σ^2 である正規分布を表す．また，関心領域を抽出するための教師画像を図 10 (巻頭カラー図 5) に示す．教師画像は対象画像と分布がほぼ同じであるが，各画素値は奇数のみに制限している．そのため，分布は似ているが僅かに異なること，また，対象画像に含まれる画素値の全てが教師画像に含まれてはないという特徴がある．このように対象画像，教師画像ともに生体画像を模擬し，生体画像における抽出の問題として設定した．

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (9)$$

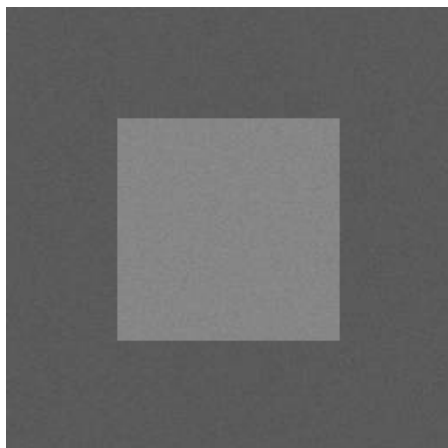


図 9 基礎実験 2 対象画像

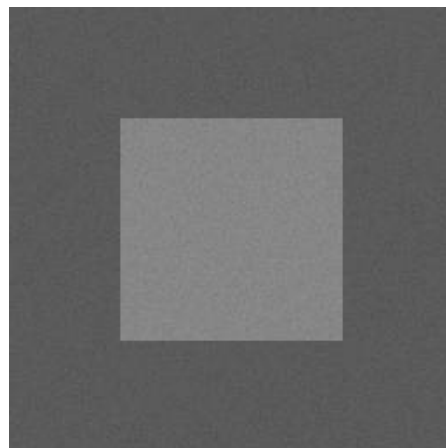


図 10 基礎実験 2 教師画像

教師画像から作成した教師カラーヒストグラムを用いて抽出した結果を図 11（巻頭カラー図 6）に示す．一致度は約 12.1%であった．これから，教師カラーヒストグラムには判別に必要な関心領域内全ての色の出現頻度が記述されていないために，判別ができていないことがわかる．

次に， $\sigma=1.0$ と設定して教師カラーヒストグラムから出現頻度を推定したロバストカラーヒストグラムを用いて抽出を行った．抽出結果を図 12に示す．一致度は 100%であった．これは，判別に必要であるが欠損していた出現頻度が $L^*a^*b^*$ 色空間での周囲の色の出現頻度から推定が行われ，正しく判別できたことを示している．以上から，少数サンプリングのカラーヒストグラムから出現頻度の推定したロバストカラーヒストグラムを用いた提案手法は有効であると言える．

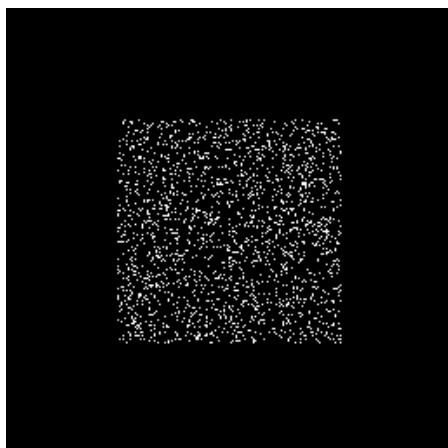


図 11 TCH による抽出結果

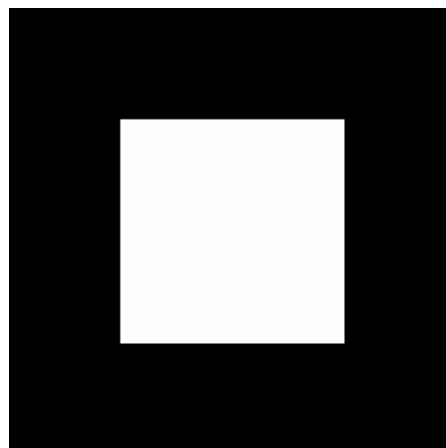


図 12 RCH($\sigma=1.0$)による抽出結果

4.2. 生体画像への適用

本節では，生体画像における提案手法の有効性を検証する．扱うデータは，3D-ISM から得られた昆虫の画像データである．これらの画像から筋肉組織領域を抽出することを目的とする．昆虫に関する研究では，昆虫内部の組織構造の観察が必要になる．

特に飛行に関する研究では，筋肉組織がどこに，どのように分布しているかを知ることが重要になる．しかし，昆虫は外骨格であり，通常の解剖では内部組織を潰してしまうため，内部組織を観察することが困難である．そこで，3D-ISM により内部組織を保存したまま得られた画像から，昆虫から筋肉組織を抽出することにより，筋肉組織の観察を可能にする．昆虫の筋肉組織を抽出することは昆虫研究にとって非常に有用であるため，この意義は大きい．

4.2.1. 実験対象画像

本節の実験では 3D-ISM により得られた，タガメ（カメムシ目・コオイムシ科）の連続断面を扱う．XY 方向の分解能は $80\mu\text{m}$ ，Z 方向の分解能は $20\mu\text{m}$ であり，総撮影枚数は 4061 枚であった．本実験では，Z 方向に 4 枚に 1 枚に間引いた画像(Z 方向分解能 $80\mu\text{m}$, 総数 1016 枚)を利用し，そのうち，筋肉組織が存在している連続断面画像 177 枚を扱う．

4.2.2. 正解画像

教師画像及び抽出結果の評価のために，手作業により抽出した画像を正解画像として用いた．3D-ISM では，切切断面を撮影するという構造により Z 方向で色の変化が発生する場合がある（具体的には下層ほど色が薄くなる）．このことから，教師とする画素値は，XY 方向だけではなく，Z 方向にもサンプリングしたものが有効であると考えられる．そのため，ボリュームデータにおける教師カラーヒストグラムを生成する画像は，XYZ 軸に直交する 3 断面の抽出結果画像を教師画像とする．ここでは， $(X,Y,Z)=(307,271,89)$ における直交 3 断面を用いた．Z 軸については中央値である 89，XY については $Z=89$ (89 枚目)における抽出結果の重心である値としてそれぞれ 307,271 を採用した．上記の値における，XY 断面，XZ 断面，YZ 断面の画像とその抽出結果である教師画像を図 13（巻頭カラー図 7）に示す．これらの画像から関心領域と非関心領域の 2 つの教師カラーヒストグラムを得た．

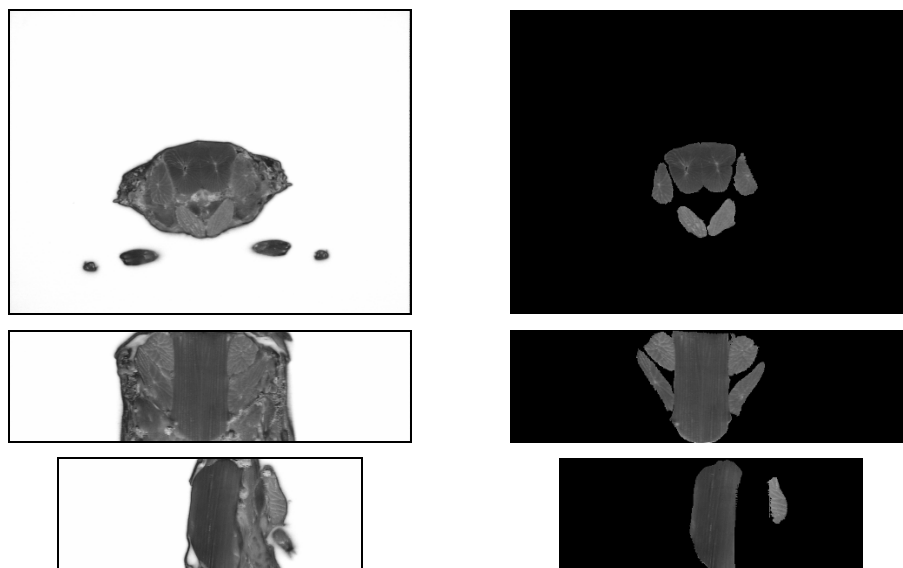


図 13 直交 3 断面 左：元画像 右：教師画像 (上：XY, 中：XZ, 下：YZ)

4.2.3. 実験

次に抽出を行うために必要であるロバストカラーヒストグラムを教師カラーヒストグラムから推定する．本実験では，断面にかかわらず高精度な抽出が可能である特定のロバストカラーヒストグラムが存在すると仮定し，その発見を目指す． σ を $0.1 \leq \sigma \leq 1.0$ (0.1刻み)， $1.0 \leq \sigma \leq 10.0$ (1.0刻み)， $10.0 \leq \sigma \leq 50.0$ (10.0刻み) に設定して推定したロバストカラーヒストグラムを用いて，抽出の試行により評価を行う．

177 枚を約 3 等分する断面である 60 枚目，118 枚目を評価画像として用いてロバストカラーヒストグラムの評価を行った．図 14 に評価画像，図 15 に上述の設定値の σ で推定を行ったロバストカラーヒストグラムを用いて抽出を行った結果の一致度を示す．ただし， $\sigma=0$ は教師カラーヒストグラムを表す．このグラフから，全体的な高低の差はあるものの，どちらの断面とも一致度の傾向は同じであり， $\sigma=1.4$ 前後で最も高い一致度を得ることがわかる．図 16（巻頭カラー図 9）に $\sigma=1.4$ のロバストカラーヒストグラムによる評価画像の抽出結果を示す．なお，図示はしていないが，全断面においても同様な結果が得られた．

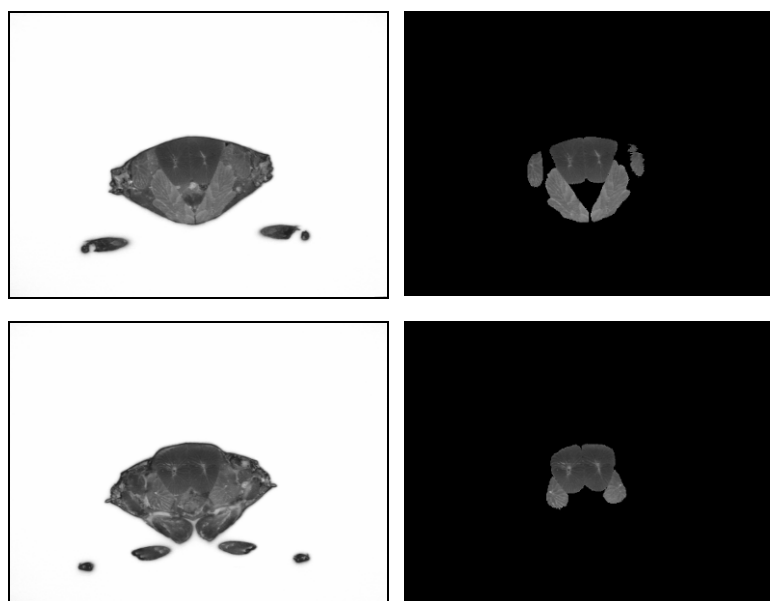


図 14 評価画像 左：元画像 右：正解画像

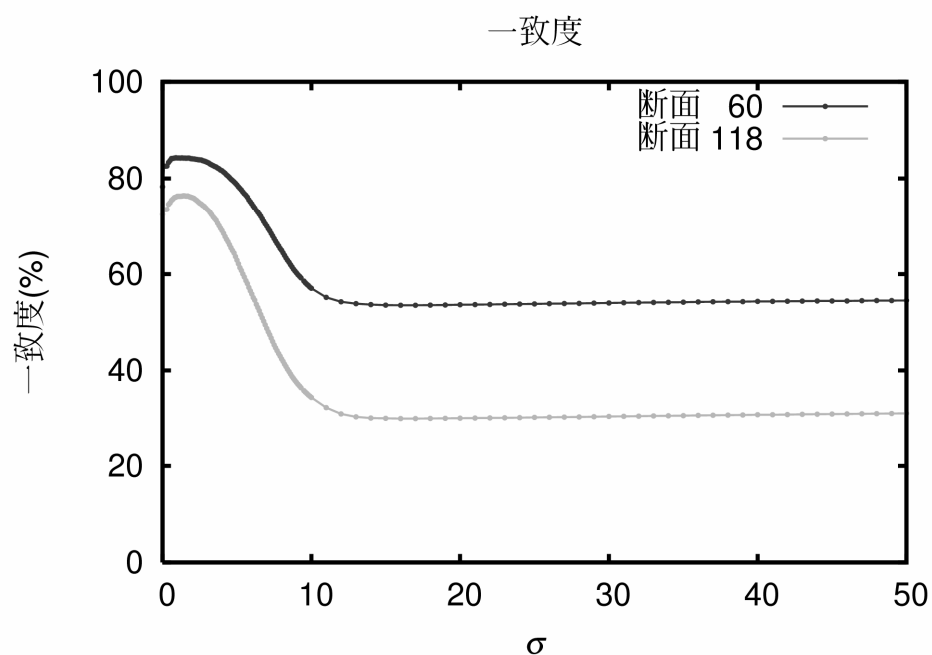


図 15 抽出結果一致度

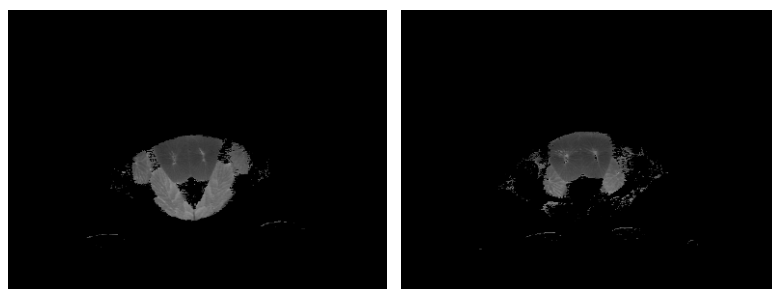


図 16 RCH($\sigma = 1.4$)による評価画像の抽出結果

以上より，妥当に推定されたロバストカラーヒストグラムは存在することが判明し，この連続断面画像からの抽出においては $\sigma = 1.4$ であった．これは，本手法における推定は有効であることを示している．図 17（巻頭カラー図 10）に断面 40, 80, 120, 160 の抽出結果画像，立体像を図 18（巻頭カラー図 11）を示す．

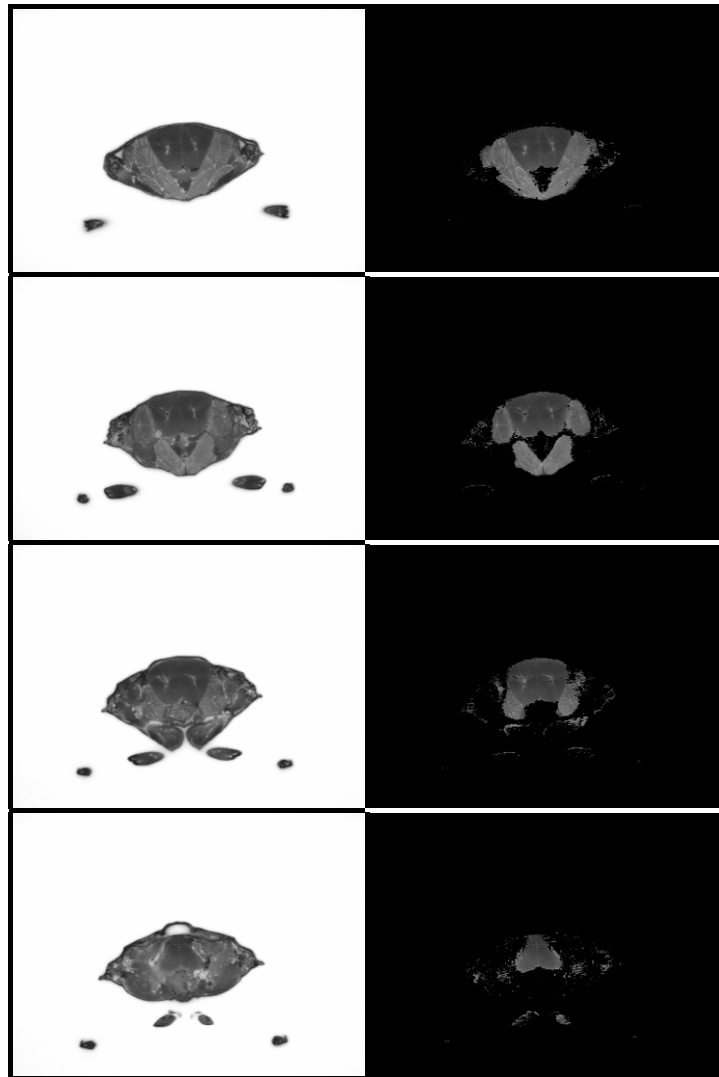


図 17 RCH($\sigma=1.4$)による抽出結果 (右：元画像)

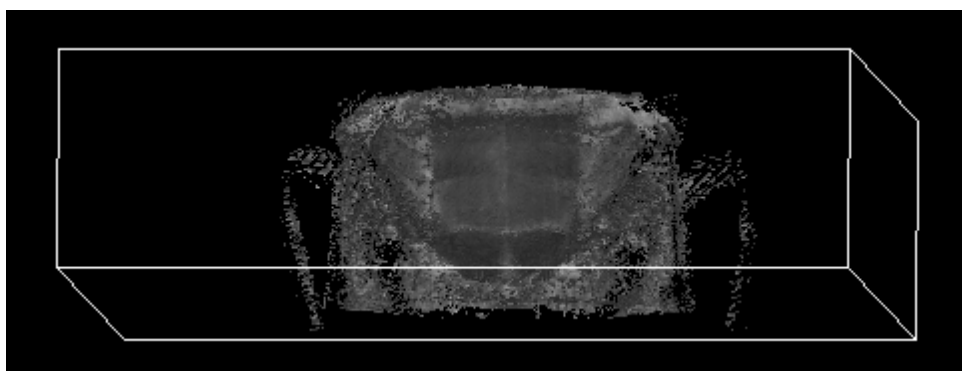


図 18 タガメ筋肉組織立体像

4.2.4. 考察

断面に関わらず、高い一致度での抽出が特定のロバストカラーヒストグラムを用いて可能であることから、推定が正しくできていることがわかった。本節では、タガメを用いた実験において提案手法を用いることで高精度な抽出を実現できたことについて考察を行う。

提案手法の処理は全て色空間上で行っているため、高精度な抽出ができたことは色空間上で色の判別を正しくできたことに他ならない。そこで、色空間上での判別結果について考察する。4.2.3の実験において色空間上での判別結果 (a^*b^* 断面, $0 \leq L \leq 100$ 10 刻み) を図 19 (巻頭カラー図 12) に示す。赤色が関心領域と判別した領域、青色が非関心領域と判別した領域である。また、明るさは頻度の大きさ、つまり可能性の高さを表している。左に全断面の正解との重ね合わせを示す。白点が関心領域、濃灰点が非関心領域、淡灰点が関心領域及び非関心領域に重複する領域を表す。どの a^*b^* 断面においても、関心、非関心の正解領域はそれぞれ判別結果の明るい部分、つまり可能性が高いと判別した部分にほぼ一致し、正しく判別できていることがわかった。しかしながら、関心領域及び非関心領域で重複している部分も少なからず存在しているためにエラーが発生していることがわかった。

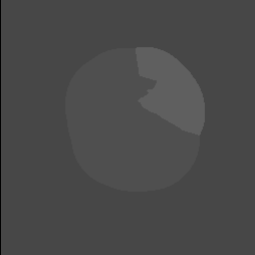
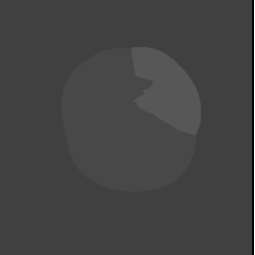
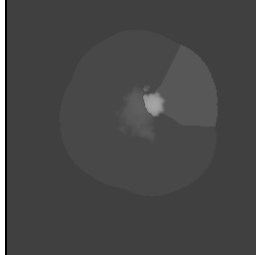
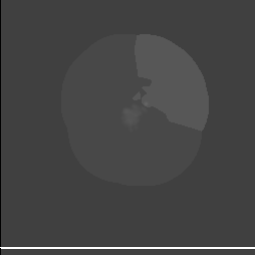
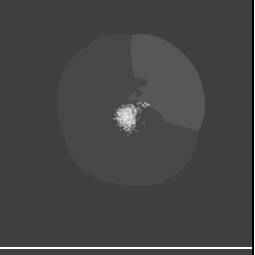
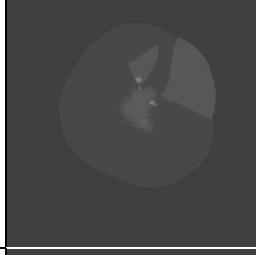
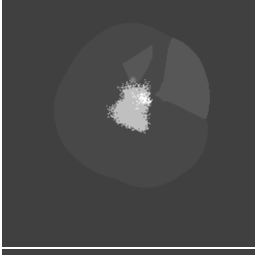
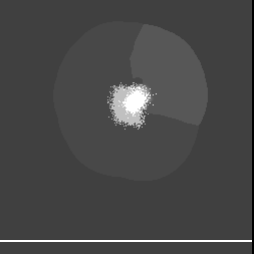
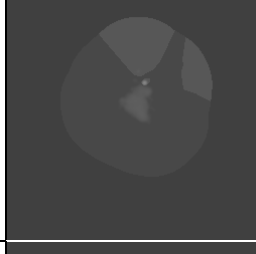
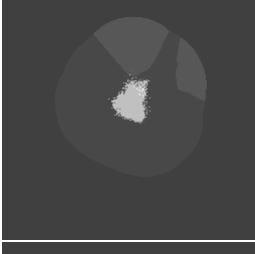
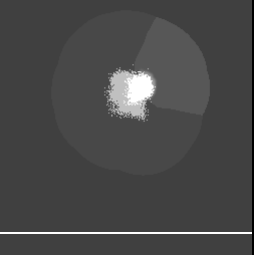
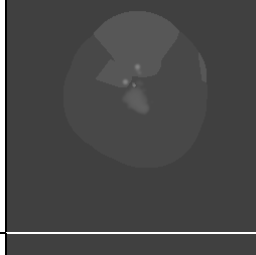
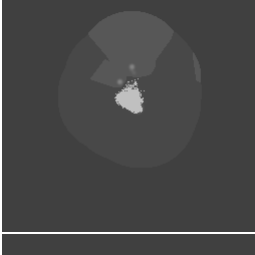
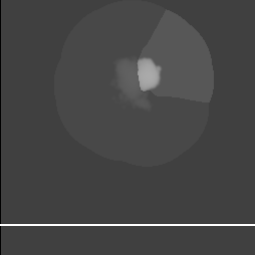
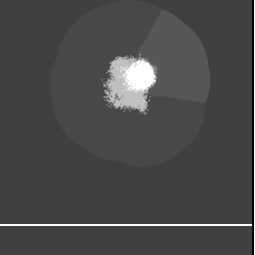
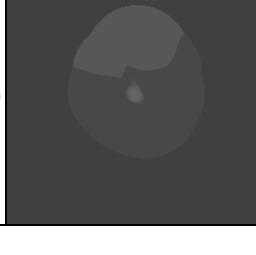
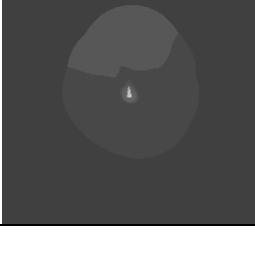
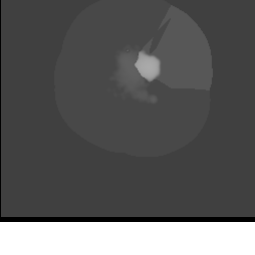
L*	判別結果	正解との重ねあわせ	L*	判別結果	正解との重ねあわせ
0			60		
10			70		
20			80		
30			90		
40			100		
50					

図 19 L*a*b*色空間における判別結果

5. まとめ

本論文では、色情報に特徴量を限定した抽出を目的とし、カラーヒストグラムの出現頻度比較による抽出を提案した。また、教師画像から得られるカラーヒストグラムから、Parzen Window 法を用いて判別に必要な出現頻度の推定したロバストカラーヒストグラムを提案した。

断面にかかわらず高い一致度での抽出が可能である特定のロバストカラーヒストグラムが存在することを明らかにし、評価画像を用いて妥当と考えられるパラメータを設定することが可能であることを示した。以上から、ロバストカラーヒストグラムの正しい推定及びそれを用いた高精度な抽出が可能であることを示した。本研究により、抽出手法の確立及び生体モデル構築への貢献が期待される。

6. 謝辞

本稿の実験における計算は RIKEN Super Combined Cluster(RSCC)の利用により遂行された。ここに記して謝意を表する。

7. 参考文献

- [1] 横田秀夫, 工藤謙一, 樋口俊郎 他, “3 次元内部構造顕微鏡による凍結生体試料の観察と計測,” 低温生物工学会誌 Cryobiology and Cryotechnology, Vol. 44, No. 1, pp.1-9, 1998.
- [2] 横田秀夫, 中村佐紀子, 川口龍平 他, “3 次元内部構造顕微鏡による生体の侵襲的イメージング,” Medical Imaging Technology, Vol. 20, No. 6, pp. 660-665, 2002.
- [3] 本多英晴, 竹本智子, 横田秀夫 他, “カラーヒストグラムを利用したフルカラー連続生体画像からの抽出法の検討,” 理研シンポジウム:生体形状情報の数値化及びデータベース構築研究, pp. 56-70, 2005
- [4] Michel J. Swain, Dana H. Ballard, “Color Indexing, International Journal of Computer Vision”, Vol. 7, No. 1, pp. 11~32, 1991.
- [5] Parzen E, “Mathematical Considerations in the Estimation,” Technometrics, Vol. 3, pp. 167-190, 1961.
- [6] Parzen E, “On estimation of a probability density function and mode,” Ann. Math. Stat., Vol. 33, pp. 1065-1076, 1962.
- [7] International Color Consortium, “Specification ICC.1:2004-10 Image technology colour management – Architecture, profile format, and data structure,” ICC, 2004
<http://www.color.org/ICC1V42.pdf>.
- [8] 川上元朗, 小松原仁, 「新版 色の常識 第2版」 日本規格協会, 1999.