

情景内カラー文字の最適2値化と変形耐性認識

横林, 実 / YOKOBAYASHI, Minoru

(発行年 / Year)

2008-03-24

(学位授与年月日 / Date of Granted)

2008-03-24

(学位名 / Degree Name)

修士(理学)

(学位授与機関 / Degree Grantor)

法政大学 (Hosei University)

修士論文

情景内カラー文字の最適 2 値化と変形耐性認識

提出日：2008 年 1 月 29 日

指導教官 若原徹

法政大学大学院情報科学研究科情報科学専攻
学生証番号 06T0024 横林実

概要

This paper addresses the most challenging problem of optimal binarization and distortion-tolerant recognition of color characters in scene images. Regarding binarization of color characters, we propose and compare two methods. One is based on selection of one optimal projection axis in the RGB color space using Otsu's criterion and decision between figure and background using the black-white ratio along the image periphery and common characteristics of characters. The other is based on clustering in the HSV color space and generation of binary images by dichotomizing those clusters. Next, regarding distortion-tolerant recognition of binarized characters we compare our global affine transformation (GAT) correlation method against the well-known tangent distance method. From experiments made on a total of 698 character images extracted from the ICDAR 2003 robust OCR dataset, a correct binarization rate of 75.3% is obtained by the method based on Otsu's criterion and decision of figure and background. Moreover, both recognition methods of the GAT correlation and the tangent distance have achieved a competing recognition rate of around 91%. However, against rotation ranging from -45 degrees to +45 degrees as artificially applied to template images the GAT correlation method is found to be much robuster than the tangent distance method.

この論文では、情景内のカラー文字の耐性認識と最適な 2 値化について取り組む。カラー文字の 2 値化について、私達は 2 つの手法の提案と比較を行う。1 つは RGB カラー空間で大津基準を利用した最適な投影軸の選択と、文字らしさと縁の白黒比を利用した「図と地」の判定に基づいている。一方は HSV 空間でクラスタリング手法の適応し、クラスタを二分することによって 2 値画像を生成する。次に 2 値化された文字の耐性認識について、私達は大局的アフィン変換 (GAT) 相関法とよく知られている接距離の比較を行った。国際会議 ICDAR2003 で使用された robust OCR データの中から合計 698 枚を選びだした文字画像の実験から、大津基準と図と地の判定に基づく手法により、正しい 2 値化率 75.3% を得られた。さらに、GAT 相関法、接距離の両認識手法が約 91% という競合している認識率を得た。しかしながら、テンプレート画像を強制的に -45 度から +45 度の範囲で回転させた画像をテスト画像とした実験では、GAT 相関法の方が接距離より耐性があることがわかった。

目次

概要	2
目次	3
第 1 章 序論	5
第 2 章 ICDAR 2003 の画像データベース	7
第 3 章 情景内カラー文字の最適 2 値化	10
3.1 カラー空間での大津基準による 2 値化	10
3.2 カラー空間でのクラスタリングによる 2 値化	12
3.3 その他の 2 値化	13
3.3.1 カラー空間・投影軸探索・閾値の組み合わせ	13
3.3.2 カラー空間での大津基準をローカルに適用した 2 値化	15
3.3.3 HSV カラー空間での大津基準による 2 値化	17
3.3.4 WMFAD(Weighted Median Filter based Anisotropic Diffusion)	18
第 4 章 情景内文字の変形耐性認識	21
4.1 前処理とテンプレート画像	21
4.2 接距離(Tangent Distance)の適用	22
4.3 GAT 相関法の適用	22
第 5 章 実験結果	25
5.1 大津基準による 2 値化	25
5.2 クラスタリングによる 2 値化	26
5.3 変形耐性認識	27
第 6 章 考察	29
6.1 大津基準による 2 値化の課題	29
6.2 クラスタリングによる 2 値化の課題	29
6.3 変形耐性認識の課題	31
6.4 今後の展開と持論	32
6.4.1 2 値化	32
6.4.2 変形耐性認識	35
第 7 章 むすび	37
実験画像一覧	38
誤認識画像一覧	42
謝辞	43

参考文献	44
------------	----

第 1 章 序論

文字認識には半世紀近い活発な研究の歴史があり、パターン認識の中で最も実用化が進んだ技術の一つである。近年では、従来のスキャナ入力による紙の上の手書き文字や活字の認識から、インターネット上の Web ドキュメントの認識・理解やカメラ・ビデオで入力した画像内文字の認識へと応用が広がり、新たな技術的課題も明らかになっている。

最近の文字認識・文書理解の研究動向については、既に優れたサーベイ報告がある[1]、[2]、[3]。

この新しい機運の中で、2003 年に開催された第 7 回文字認識・文書理解国際会議 ICDAR2003 では、カラー情景画像からの、(1) Text locating、(2) Robust word recognition、(3) Robust character recognition、についてそれぞれ別個のデータベースを用意したコンペティションの実施結果が報告された[4]。実際に参加者があったのは課題(1)のみであった。使用されたデータベースは後に公開されている[5]。その後、公開されているデータベースを利用し課題(3)に挑んだ研究が報告された[6]。

本研究では、上記の公開データベース[5]から ICDAR2003 robust OCR dataset を用いて、課題(3)の情景内カラー文字の個別文字認識を扱う。このデータベースは切出された 1 文字のみの画像から成るが、多様な複雑背景や配色、劣化要因および字体を有する文字画像であり、挑戦的な認識対象となっている。

本論文では、この情景内カラー文字を対象に最適 2 値化と変形耐性認識の 2 つの課題に分けて検討結果を述べ、現状レベルと残された課題について考察する。

カラー文字の 2 値化については、文字が様々に彩色されていたり、複雑な背景を持つ、さらにぼけ等の画像劣化を含む場合、極めて難しい課題となる[3]。最近の研究として、文章の 2 値化手法を提案した Gatos[7]らが、Niblack[8]、Sauvola[9]などの 2 値化手法と比較する研究が発表されているが、課題(3)のようなカラー画像かつ 1 文字に区切られた画像の 2 値化には適していない。

本論文で提案する 2 値化法は 2 種類である。1 つ目はカラー空間での大津基準[10]を適用した 2 クラス分類[6]を行い、2 クラスのいずれが文字部分であるかの「図と地」の判定処理を行う。「図と地」の判定処理には黑白比率と文字線幅や連結成分数に基づく文字らしさの評価を利用する。2 つ目は HSV カラー空間で k 平均法[11]によりクラスタリングを行い、得られたクラスタ群を二分して複数の 2 値画像を作成する。その際、連結成分数、文字部分の割合、文字幅、円形度などから文字らしさを評価し、作成される 2 値画像の枚数を抑制する。

次に、2 値化された文字の認識を行う。情景内文字は様々な字体、デザインによる変形・歪みがあり、またぼけ等の画像劣化を含む。このため、少なくとも、アフィン変換（回転・

伸縮・せん断、平行移動) やぼけに対して耐性を有する認識手法が必要となる。本論文では、アフィン変換に対して耐性のある手法として GAT 相関法 (Global Affine Transformation Correlation) [12]および接距離 (Tangent Distance) [13]を、正しく 2 値化された画像の認識に適用して性能の比較を行う。さらに、前処理として施す位置と大きさの正規化処理[14]で吸収できない回転成分に対する頑健性を調べるため、回転したテンプレート画像間のマッチング実験より両手法の回転耐性を評価する。

実験では、ICDAR2003 robust OCR dataset から抽出した総数 698 枚のカラー文字画像を評価用に用いた。認識の際は、標準テンプレート画像として単一字体の英大小文字および数字の計 62 枚のみを用いた。

上記 698 枚のカラー文字画像に対する 2 値化実験では、大津基準を適用した 2 クラス分類後に「図と地」の判定処理を適用する手法が正しい 2 値化率 75.3%を達成した。一方のクラスタリングに基づく 2 値化法は、クラスタ群を二分して文字らしい 2 値化パターンのみを生成する制御に未だ課題があることが明らかとなった。しかし、人間が目視によりクラスタ群を最適に二分すれば正しい 2 値化画像をほぼ得ることができるため、クラスタリングに基づく 2 値化法が有効である可能性を確認した。

変形耐性認識の実験では、正しく 2 値化された文字パターンに対する認識率で、接距離が 91.6%、GAT 相関法が 90.3%を達成した。 -45 度 $\sim+45$ 度の範囲での回転耐性の評価実験では GAT 相関法の優位性が明らかとなった。

以下、章 2 で ICDAR2003 robust OCR dataset について紹介し、章 3 では最適 2 値化、章 4 では変形耐性認識について述べる。章 5 は実験結果であり、章 6 で現状レベルと今後の課題について考察する。

なお、上記に加え、研究で使用してみた 2 値化を節 3.3 に、持論を含めた今後の展開を節 6.4 に記す。

第 2 章 ICDAR 2003 の画像データベース

2003 年英国エディンバラで開催された第 7 回文字認識・文書理解国際会議 ICDAR2003 で、カラー情景画像を対象に Robust reading competitions が企画され、

- Text locating
- Robust word recognition
- Robust character recognition

の 3 部門について別個のデータベースが用意された。実施結果が同会議中に報告された[4]が、応募があったのは Text locating 部門のみであった。使用されたデータベースは公開されてダウンロード可能である[5]。図 2.1 はそれぞれの部門の画像例である。



左 : Text locating 中央 : Robust word recognition 右 : Robust character recognition

図 2.1 各部門で公開されている画像例

本研究では、Robust character recognition 部門で用いられた ICDAR2003 robust OCR dataset を認識対象とする。このデータベースは 1 文字のみを含む情景内カラー文字画像であり、“Sample”、“TrailTrain”、“TrialTest”に分かれている。それぞれの画像枚数は Sample : 854 枚、TrrailTrain : 6185 枚、TrialTest : 5430 枚、合計 12469 枚から成る。しかし、この画像の中には図 2.2 のような、今回の実験対象としている英大小文字と数字以外の画像も含まれている。



図 2.2 英大小文字と数字以外の画像例

英大小文字と数字以外を除くと画像枚数は表 2.1 のようになる。

表 2.1 各データベースの画像枚数

データベース名	公開されている 画像枚数	英大小文字と数字以外 を除いた画像枚数
Sample	854	851
TrailTrain	6185	6113
TrialTest	5430	5379
Total	12469	12307

今回、上記のデータベース **Sample** より、英大小文字と数字の画像 851 枚から、人間が認識できる画像 698 枚を抽出した。

図 2.3 に画像例を示す。

図 2.3 より、人間ならばカラー情報や類推により正しく認識できるものの、計算機による最適 2 値化や認識はかなり困難な画像群であることが分かる。



図 2.3 実験に用いた画像例

公開されている画像は JPEG 形式であるため、プログラムで処理しやすいように予め PPM 形式に変換をした。

今回の実験に使用した画像全 698 枚は、以前、背景や劣化の具合によって 7 つのグループに分け実験を行った[15]。グループに分け実験を行うことは、どのような画像がうまく 2 値化できないかを調べるのに有効である。以下、以前実験したときに使用した、分け方を記す。

画像全 698 枚を、背景や劣化の具合によって 7 つのグループに分け実験を行った。それぞれのグループと画像の枚数は表 2.2（次ページ）のとおりになる。

表 2.2 それぞれのグループの画像の枚数

Group	Number of images
Clear	199
Background design	130
Multi-color character	54
Nonuniform lighting	40
Little contrast	37
Blurring	210
Serious distortion	28
Total	698

図 2.4 にそれぞれのグループの画像の例を提示する。

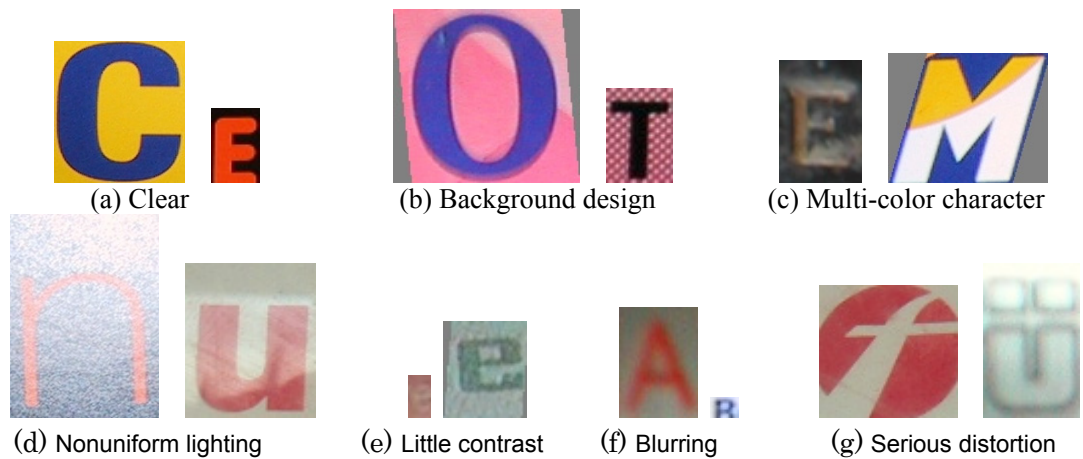


図 2.4 それぞれのグループの画像の例

- (a) の”Clear”は文字と背景の 2 色から成る画像。
- (b) の”Background design”は文字は 1 色から成っているが背景はいくつかの色でデザインされている画像。
- (c) の”Multi-color character”は文字と背景ともにいくつかの色で作られている。
- (d) の”Nonuniform lighting”は光が反射している画像。
- (e) の”Little contrast”は文字と背景のコントラストの差が小さい画像。
- (f) の”Blurring”はぼけている画像。
- (g) の”Serious distortion”は文字の形が特殊である画像。

以上のようにノイズの程度、劣化具合により 7 つのグループに分けたが、カテゴリ化の見直しは必要かと思われる。なお、本論文の最後の方の実験画像一覧に、実験データで使用了全 698 枚の画像の番号を記載しておく。

第3章 情景内カラー文字の最適2値化

この章では、本論文で提案するカラー空間での大津基準による2値化（節3.1）とクラスタリングによる2値化（節3.2）を記述する。節3.3には利用してみたさまざまな2値化方法を記述する。

3.1 カラー空間での大津基準による2値化

濃淡文字画像の最適2値化を行う手法として、大津の2値化基準[6]がよく知られている。

大津基準では、濃淡文字画像のヒストグラムに対して濃淡レベル k を閾値として2クラス分類した際の、クラス間分散 σ_B^2 と総分散 σ_T^2 の比 η を最大化する最適な閾値 k^* を用いることにより2値化を行う。

すなわち、次式により最適閾値 k^* を決定する。但し、 σ_W^2 はクラス内分散を表す。

$$\begin{aligned}\sigma_T^2 &= \sigma_B^2 + \sigma_W^2, \\ \eta &= \sigma_B^2 / \sigma_T^2 \rightarrow \max \text{ for } k.\end{aligned}\quad (1)$$

ここで η をクラス間分離度と呼ぶことにする。

本節では、このクラス間分離度を3次元カラー空間に拡張して適用し、情景内カラー文字の2値化を実現する。以下に詳細を述べる。

まず、カラー文字画像の各画素について RGB 値を3次元 RGB カラー空間の1点にプロットする。

こうして作成された3次元カラー空間内の点分布を平面で2分割するため、分割平面の直交軸へ点分布の投影を行う。

図3.1に、3次元 RGB カラー空間での投影軸設定の考え方を示す。

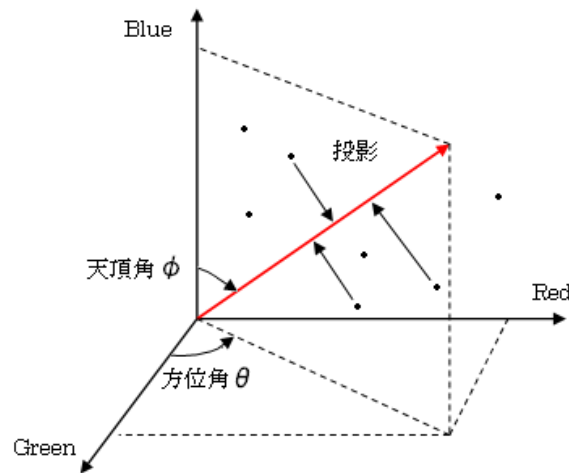


図 3.1 3次元 RGB カラー空間での投影軸の設定

図 3.1 より、投影軸は天頂角 ϕ と方位角 θ で指定され、各画素の RGB 値を軸上に投影した値 H は次式で算出される。

$$H = R \sin \phi \sin \theta + G \sin \phi \cos \theta + B \cos \phi, \quad (2)$$

$$0^\circ \leq \phi, \theta \leq 180^\circ.$$

天頂角 ϕ と方位角 θ で指定される投影軸毎に値 H のヒストグラムを作成して、大津基準によるクラス間分離度 $\eta(\phi, \theta)$ を算出する。

これより、最適な分割平面は次式：

$$\eta(\phi^*, \theta^*) = \max_{\phi, \theta} \eta(\phi, \theta). \quad (3)$$

を満たす投影軸 (ϕ^*, θ^*) およびその軸上の最適閾値 K^* により定まることになる。

実際には、 ϕ と θ をそれぞれ 1° 刻みで振ることにより、総数 180×180 個の投影軸から、式(3)を満たす最適な投影軸および閾値を決定した。

上記による方法で 2 値化された画像を文字部を黒くするために図と地の判定を行う。図と地の判定処理として、2 種類の方式を提案する。以下、文字と判定された領域を黒、背景と判定された領域を白と名づけて、それぞれの方式での文字部（黒）と背景部（白）が満たすべき条件を記す。

方式 1：画像全体での黒／白比の方が画像枠部分での黒／白比より大きい。

方式 2：①画像枠部分で黒／白比が閾値 α より小さい。この場合、黒領域の最大連結成分のみを文字部として残す。または、②黒領域の最大連結成分を収縮処理して消失する回数の方が、白領域の最大連結成分を収縮処理して消失する回数より少ない、この場合、黒領域の最大連結成分のみを文字部として残す。

但し、方式 2 において、収縮処理はモルフォロジ演算の **erosion**[16]を指す。また、黒領域の最大連結成分のみ残すのは、今回対象とした情景内文字が英数字で、小文字の「i,j」を除き単連結図形であることによる。

図 3.2（次ページ）に、方式 1 と方式 2 での 2 値化処理例を示す。但し、方式 2 での閾値 $\alpha = 0.3$ とした。

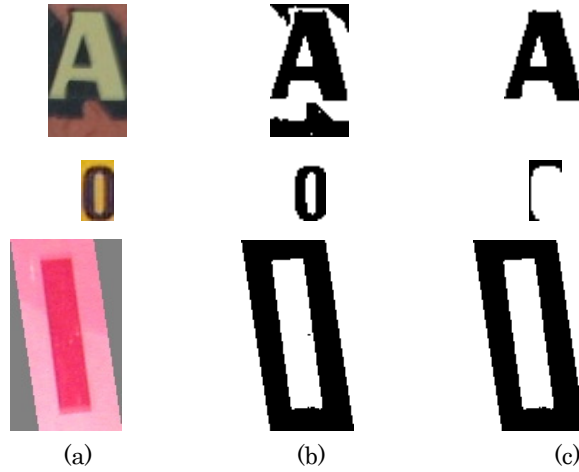


図 3.2 図と地の判定処理の例

(a) カラー画像 (b) 方式 1 (c) 方式 2

3.2 カラー空間でのクラスタリングによる 2 値化

特徴空間中で距離が近いデータをまとめ、 k 個のグループ（クラスタ）に分割するクラスタリング手法に k 平均法がある[11]。 k 個のクラスタは評価基準

$$J = \sum_{i=1}^k \sum_{x \in X_i} \|x - \bar{x}_i\|^2. \quad (4)$$

を最小化するように分割される。 $\bar{x}_i$ はクラスタ X_i に含まれる点の重心である。

この k 平均法を HSV カラー空間に適用して k 個のクラスタに分割し、得られたクラスタ群を二分することで 2 値画像を生成する。生成される 2 値画像の総数は

$$\sum_{i=1}^{k-1} k C_i. \quad (5)$$

となる。

クラスタ群の二分により生成される 2 値画像に対して、文字らしさを判定して 2 値画像の生成数を抑制する。以下の 4 つの尺度を用いて制御を行った。

- ① 雑音除去後の連結成分数
- ② 画像全体に対する文字部分の割合
- ③ 文字幅
- ④ 円形度

上記①は英大小文字と数字は最大で連結成分数 2 となるので、連結成分数 3 以上となった場合は画像を作成しないようにした。②、③、④は黒画素の数、周囲長などを利用し、あまりにも文字らしくない画像を作成しないように制御を行った。

3.3 その他の 2 値化

この章では、節 3.1、3.2 以外に行った 2 値化法、フィルターの手法を記しておく。

3.3.1 カラー空間・投影軸探索・閾値の組み合わせ

節 3.1 の 2 値化法はカラー空間から投影軸を 1 つ探し出し、その軸で閾値を決め 2 値化する方法である。この手法は以下の 3 つの選択を行っている。

- ① カラー空間の選択
- ② 投影軸探索の基準の選択
- ③ 閾値の選択

上記①ー③はさまざまな選択があり、組み合わせにより何通りもの 2 値化法を行うことができる。今までに利用してみたそれぞれの手法を表 3.1 に記す。

表 3.1 利用してみたカラー空間・基準・閾値

カラー空間	投影軸探索の基準	閾値
・ RGB 空間 ・ HSV 空間	・ 分離度 ・ 分散 ・ エントロピー	静的 ・ 大津の 2 値化 ・ Niblack[8] ・ Sauvola[9] 動的 ・ move_th ・ move_thx ・ move_thxe

表 3.1 の赤字で書いてある方法の組み合わせが節 3.1 の手法となる。統計は取っていないが、結果的にこの組み合わせが 1 番よかったと言ってよいだろう。なお、動的閾値の move_th、move_thx、move_thxe の方法は以下の通りである。

元画像の各注目画素を中心とする $W \times W$ サイズの平均濃淡値をローカル平均値とする。

move_th

ローカル平均値を注目画素の閾値とする方法。

move_thx

全体の平均値とローカル平均値の低い方を注目画素の閾値とする方法。

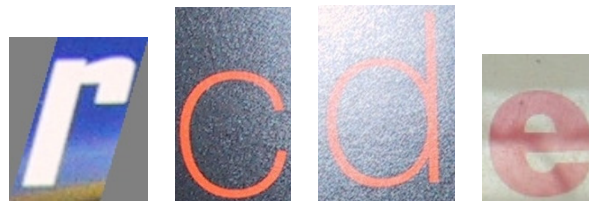
move_thxe

注目画素の閾値 Th は次式によって決まる。

$$Th = lmean_mean + ALPHA \times lmean_std. \quad (6)$$

$lmean_mean$ はローカル平均値の平均値、 $lmean_std$ はローカル平均値の分散、 $ALPHA$ は重みである。この Th と注目画素のローカル平均値を比べる方法。

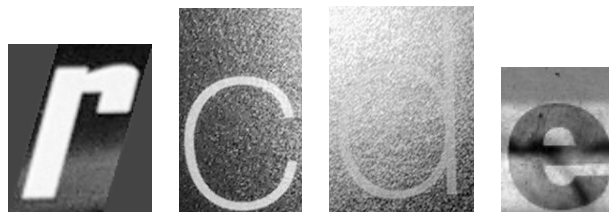
図 3.3 にカラー画像からそれぞれの動的閾値を適用した画像例を示す。



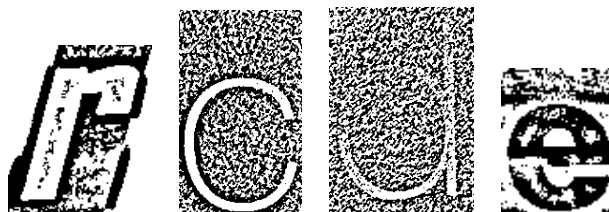
カラー画像



最大分離度軸の2値化 (節 3.1 の方法)



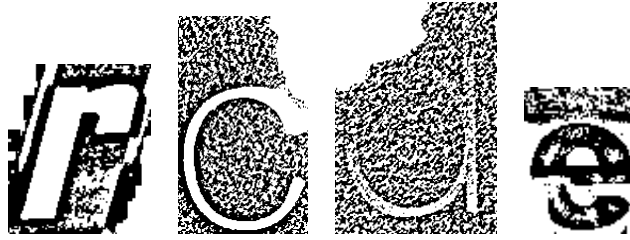
最大分離度軸のグレースケール画像



グレースケール画像に move_th を適用



グレースケール画像に move_thx の適用



グレースケール画像に move_thxe を適用

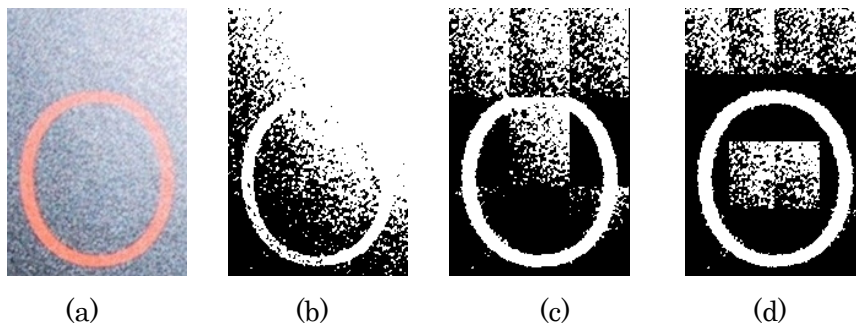
図 3.3 動的閾値を適用した画像例

図 3.3 は特にうまくいかなかった画像を表示しているので、もちろんうまく 2 値化できた画像もある。しかし、全体的に節 3.3 の手法よりうまくいかなかったと言ってよいだろう。

3.3.2 カラー空間での大津基準をローカルに適用した 2 値化

画像を $m \times m$ に区切り、一区切り毎に節 3.1 を適用した方法である。つまり $m \times m$ に区切った場合、一区切り毎の投影軸と閾値が決定され、1 枚の画像に $m \times m$ 個の投影軸と閾値が決定される。

図 3.4 にこの手法を適用した画像例を示す。



(a)カラー画像 (b)画像全体で大津基準を適用した画像(節 3.1 の方法)
(c)3×3 に区切り大津基準をローカルに適用した画像
(d)4×4 に区切り大津基準をローカルに適用した画像

図 3.4 大津基準をローカルに適用した画像例

図 3.4 の(d)を利用すれば、正しい 2 値画像を得ることができそうである。

図 3.5 に背景とみなすべき区切りを示す。

図 3.5 より、●の区切りを背景とみなせれば、正しい 2 値画像（文字 ‘O’）を得ることができる。

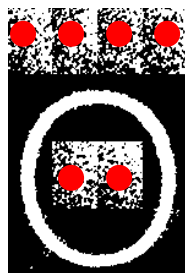


図 3.5 ●の部分为背景と判定すれば 2 値化がうまくいく図 3.4 (d)の画像

そこで、 4×4 で区切ったローカルに大津基準を適用した 2 値化を行い、連結製分数が多い区切りを背景（真っ白）にして、そこを基点に区切りの縁をしらべ上下左右の 4 周辺の区切りを反転するかしなかの判定を行っていく。背景にする所がなかった場合は一番左上の区切りを基点とする方法を考えた。

図 3.6 にこの連結製分数が多い区切りを背景とみなした 2 値化法を行った画像例を示す。



図 3.6 連結製分数が多い区切りを背景とみなした 2 値化法の画像例

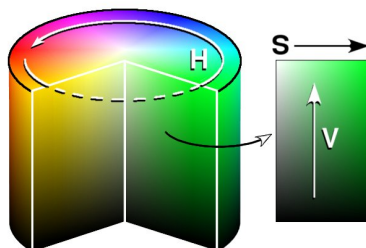
図 3.6 の左のハレーションが載っている画像は節 3.1 の方法ではほとんどうまく 2 値化することができなかった画像であるが、この方法によってうまく文字部 ‘n’ を抽出することができた。しかしながら、全体的にはうまくいかなかったと言ってよいだろう。

大きな問題点として 2 つある。1 つ目は区切るサイズ $m \times m$ の値をいくつにすればよいかわからない。実験に用いた画像には小さい画像も多く、小さい画像には区切ることで、かえってうまくいなくなってしまう。2 つ目は背景とみなすべき区切りは連結製分数が多くなるとは限らない問題である。そのために、図 3.6 の中央と左の画像はうまく 2 値化が出来なかった例である。

3.3.3 HSVカラー空間での大津基準による2値化

この章では節 3.3.1 の表 3.1 の HSV カラー空間で大津基準を適用することで発生する問題点をふまえた手法を提案する。

図 3.7 に円柱の HSV カラー空間と RGB から HSV への変換式を示す(2008 年 1 月 Wikipedia より抜粋[17])。



$$H = \begin{cases} 60 \times \frac{G-B}{MAX-MIN} + 0, & \text{if } MAX = R \\ 60 \times \frac{B-R}{MAX-MIN} + 120, & \text{if } MAX = G \\ 60 \times \frac{R-G}{MAX-MIN} + 240, & \text{if } MAX = B \end{cases}$$

$$S = \frac{MAX - MIN}{MAX}$$

$$V = MAX$$

図 3.7 円柱の HSV カラー空間と RGB から HSV への変換式

H は色相、S は彩度、V は明度である。MAX は RGB 値の最大値、MIN は RGB 値の最小値である。実験に使用したカラー画像は RGB 値をそれぞれ 0～255 の値を持っているので、図 3.7 の HSV への変換式より変換を行うと、H:0～359(範囲外は 360 で剰余をとる)、S:0～1.0、V:0～255 の値をとる。この HSV カラー空間で大津基準を適用するには以下の問題点がある。

- ① H は角度を表している。
- ② H・S・V のそれぞれの距離比が違う。
- ③ MIN=MAX と MAX=0 の場合の H の定義。

H が例えば 0 と 359 の場合、距離は数値上では 359 離れていることになるが、H は角度なので正しい距離は 1 でなければならない。ユークリッド距離で距離を計算すると、本当はほぼ同じ色相であるにもかかわらず、1 番遠い色相であると計算されてしまう(上記①の

問題)。なお、節 3.3.1 の HSV 空間で大津基準を適用した 2 値化ではユークリッド距離を採用している。

②については、 $H \cdot S \cdot V$ にそれぞれに重みをかけて、距離を計算しなければならない。節 3.3.1 では HSV 変換後の値をそれぞれ 0～255 の範囲に直し、2 値化を行った。

③については、 $\text{MIN}=\text{MAX}$ と $\text{MAX}=0$ の場合は H の定義には意味がない。 $\text{MIN}=\text{MAX}$ 場合、図 3.7 の円柱の円の中心を通る軸上になる。しかしながら、HSV 空間で大津基準を適用するには H に値を入れなければならない。そこで節 3.2 の手法では $\text{MIN}=\text{MAX}$ と $\text{MAX}=0$ の場合は H を 0 とした。しかし、 H が 0 というのは、色相で赤を表しており、そこに誤差が発生してしまう。

上記の問題①をふまえて、HSV 変換後の値を、 $S \sin H \cdot S \cos H \cdot S \cdot V$ の 4 次元に拡張して軸に投影を行うという方法を提案した。4 次元空間から軸に投影する計算式は、節 3.1 の式(2)を拡張し

$$H = S \sin H \sin \phi \sin \theta \sin \rho + S \cos H \sin \phi \cos \theta \sin \rho + S \cos \phi \sin \rho + V \cos \rho, \quad (7)$$

$$0^\circ \leq \phi, \theta, \rho \leq 360^\circ.$$

となる。左辺の H は計算後の値であって、HSV の H でないことに注意していただきたい。右辺の H は HSV の H である。投影した軸の分離度を計算し、分離度が最大となる軸で大津の 2 値化により 2 値化を行った。

統計は取っていないが、節 3.1 の手法よりよくなかったと言ってよいだろう。

3.3.4 WMFAD(Weighted Median Filter based Anisotropic Diffusion)

WMFAD というのは、RGB 値を保持し、かつノイズを除去した画像を作成するための平滑化を行うことができる手法である[18]。残念ながら文献[18]は見つけることができなかったが、情景内カラー画像から文字列を探し出す研究で使われており、手法の説明と有効性が示されている[19]、[20]。

WMFAD による平滑化後の画素値は次式により求められる。

$$I(t+1) = I(t) + \lambda \sum_D C_D \nabla_D I(t) \quad (8)$$

$I(t)$ は注目画素の RGB 値、 $I(t+1)$ は平滑化後の RGB 値、 λ は定数、 D は上下左右の 4 方向、 $\nabla_D I(t)$ は 4 近傍の画素と注目画素の差、 C_D は 4 近傍に対する重みを表し、

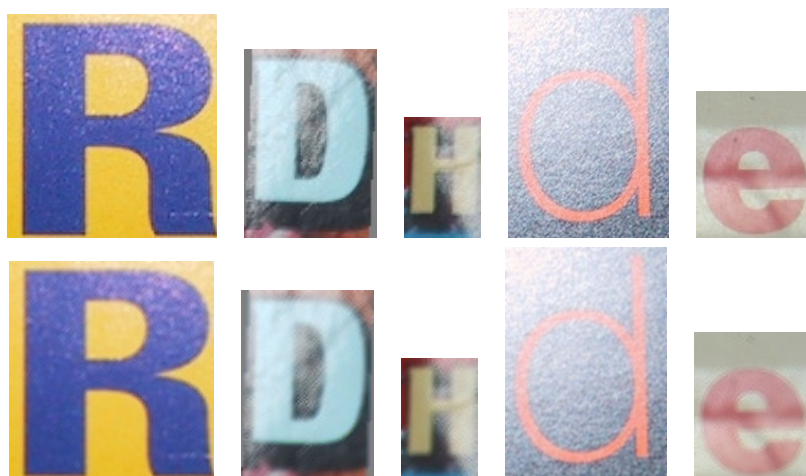
$$C_D = \exp(-(\nabla_D \text{median}(Y)/K)^2) \quad (9)$$

で表される。 $\text{median}(Y)$ は注目画素と 8 近傍の明度値にメディアンフィルタをかけた値、 K は定数を表す。

ここで、文献には書かれていなかったが、RGB 値がそれぞれ 0～255 の範囲で表されている場合、 λ 、 K の値によっては平滑化後の値が範囲外の値になってしまう。そのため今回の実験では $\lambda=0.4$ を $K=40$ とし、平滑化後の値が 255 を越えてしまう場合があるため、

その場合は λ を 0.1 減らして、再び計算するようにした。

図 3.8 に入力画像と平滑化後の画像例を示す。



上：入力画像 下：平滑化後の画像

図 3.8 入力画像と平滑化後の画像

図 3.8 より、平滑化後の画像は入力画像に比べて明度が滑らかになり、突飛なノイズがなくなっているのが分かる。

続いて、平滑化によって節 3.1 のカラー画像での大津基準による 2 値化に違いがあるかを調べた。

図 3.9 に平滑化していない画像と、平滑化した画像に大津基準による 2 値化を実行した画像例を示す。なお、図と地の判定は行っていない。



上：平滑化していない画像を 2 値化した画像 下：平滑化した画像を 2 値化した画像

図 3.9 平滑化した画像としていない画像の 2 値化による違い

図 3.9 より、1 番右の画像以外は詳しく見ないとわからないくらいの違いしかない。1 番

左の画像はかなりの違いが見られる。これは平滑化することによって、分離度が最大となる軸がまったく変わったためである。

まれに図 3.9 の 1 番左のようにまったく 2 値化後の画像が大きく変わるが、全体的には正しい 2 値化を得るために有効とは言えないと考えている。これは文献[19]、[20]は風景画像から文字列を探し出す研究であるのに対し、私達が対象としている画像は 1 文字に区切られている画像であることが、あまり有効であることが示されなかった大きな原因であると考えている。しかしながら、使い方によっては平滑化は有効な手段であると考えている。

第4章 情景内文字の変形耐性認識

4.1 前処理とテンプレート画像

今回、マッチングに用いるテンプレート画像としては、単一字体（HGP Gothic E）の英大小文字と数字の計 62 枚のみを用いた。

図 4.1 にテンプレート画像の例を示す。



図 4.1 テンプレート画像の例

2 値化文字に対して、前処理として、位置と大きさの正規化[14]を施す。具体的には、正規化後の画像サイズを 28×28 画素とし、まず黒領域の重心を画像中心に移動し、次いで重心からの黒画素への平均距離が一定値 $= 7.0$ になるように大きさを正規化する。さらに、 3×3 近傍での平均化操作を 1 回施して濃淡画像に変換する。

テンプレート画像についても、入力文字と同様に、位置と大きさの正規化および濃淡化操作を施した。

図 4.2 に、前処理を施したテンプレート画像の例を示す。



図 4.2 前処理を施したテンプレート画像の例

4.2 接距離(Tangent Distance)の適用

接距離(Tangent Distance)[13]は、マッチングすべき 2 枚の画像の各々について、変形成分毎に接ベクトル(Tangent Vector; TV)を作成しておき、それらの線形結合を用いて、画像間の最小距離を計算する方法である。

入力画像 $\mathbf{f}$ とテンプレート画像 $\mathbf{g}$ の各変形成分に対応する TV を列ベクトルとして並べた行列を $\mathbf{T}_f$ 、 $\mathbf{T}_g$ 、各 TV の重みを成分とするベクトルを α_f 、 α_g とすると、2 つの画像間の接距離 $D_T(\mathbf{f}, \mathbf{g})$ は

$$\begin{aligned} D_T(\mathbf{f}, \mathbf{g}) &= \min_{\alpha_f, \alpha_g} \|\tilde{\mathbf{f}} - \tilde{\mathbf{g}}\|, \\ \tilde{\mathbf{f}} &= \mathbf{f} + \mathbf{T}_f \alpha_f, \\ \tilde{\mathbf{g}} &= \mathbf{g} + \mathbf{T}_g \alpha_g. \end{aligned} \quad (10)$$

として定義される。各変形成分の TV は、標準偏差が σ であるガウシアンフィルタ

$$G_\sigma(x, y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (11)$$

に対応する変形作用素を施したものと画像との畳み込み演算を行うことにより作成できる。計算方法の詳細は文献[21]を参照して頂きたい。

今回は、水平移動、垂直移動、拡大・縮小、回転、水平方向への伸縮変形、ねじれ、細め・太めに対する計 7 種類の TV を作成した。また、標準偏差は $\sigma = 0.7$ 、畳み込み演算の近傍範囲は 19×19 とした。

図 4.3 に数字 ‘9’ と文字 ‘R’ のそれぞれの変形の TV を示す。左から水平移動、垂直移動、拡大・縮小、回転、水平方向への伸縮変形、ねじれ、細め・太めの TV である。



図 4.3 それぞれの変形の TV

4.3 GAT相関法の適用

GAT 相関法[12]は、入力画像に対して、最適な大局的アフィン変換 (Global Affine Transformation; GAT) を施して、テンプレート画像との正規化相互相関値を最大化する画像マッチングの手法である。

まず、入力画像を $\mathbf{f} = \{f(i, j)\}$ 、テンプレート画像を $\mathbf{g} = \{g(i, j)\}$ と記し、次式を満たすよう

に線形変換する。

$$\begin{aligned}\sum_j \sum_i f(i, j) &= 0, \quad \| \mathbf{f} \| = 1, \\ \sum_j \sum_i g(i, j) &= 0, \quad \| \mathbf{g} \| = 1.\end{aligned}\quad (12)$$

この操作により、2 枚の画像 $\mathbf{f}$ と $\mathbf{g}$ の正規化相互相関値は内積 $(\mathbf{f}, \mathbf{g})$ に等しくなる。

次に、回転・伸縮・せん断を表す 2×2 行列 $\mathbf{A}$ と平行移動を表す 2 次元ベクトル $\mathbf{b}$:

$$\mathbf{A} = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}. \quad (13)$$

によるアフィン変換操作を $\text{Affine}[\bullet]$ と記すと、正規化相互相関値を最大化するための目的関数 Φ は次式で与えられる。

$$\begin{aligned}\Phi &\equiv (\text{Affine}[\mathbf{f}], \mathbf{g}) = \sum_j \sum_i \text{Affine}[\mathbf{f}](i, j) \times g(i, j) \\ &= \sum_j \sum_i f(i, j) g(\tilde{i}, \tilde{j}) \rightarrow \max \text{ for } \mathbf{A}, \mathbf{b}. \\ \tilde{i} &= a_{00} \times i + a_{01} \times j + b_0, \quad \tilde{j} = a_{10} \times i + a_{11} \times j + b_1.\end{aligned}\quad (14)$$

さらに、 Φ を $\mathbf{A}$ 、 $\mathbf{b}$ に関して微分可能とするために、Gauss 型核関数を用いた、次の目的関数 Ψ で置き換える。

$$\begin{aligned}\Psi &\equiv \sum_j \sum_i \sum_{j'} \sum_{i'} \gamma(i, j, i', j') f(i, j) g(i', j') G(\mathbf{A}, \mathbf{b}, i, j, i', j') \\ &\rightarrow \max \text{ for } \mathbf{A}, \mathbf{b}. \\ G(\mathbf{A}, \mathbf{b}, i, j, i', j') &= \exp\left(-\frac{(i' - \tilde{i})^2 + (j' - \tilde{j})^2}{2D^2}\right), \\ \tilde{i} &= a_{00} \times i + a_{01} \times j + b_0, \quad \tilde{j} = a_{10} \times i + a_{11} \times j + b_1.\end{aligned}\quad (15)$$

ここで、 D は Gauss 型核関数の拡がり制御パラメータであり、 γ は重なりを拘束する関数である。より詳しい内容は文献[12]を参照して頂きたい。

GAT 処理の流れを以下に示す。

1. まず GAT 処理を行う前に、入力画像と目的画像の相関値を取る
2. 最適な $\mathbf{A}$ 、 $\mathbf{b}$ を算出し、GAT 処理を行う
3. 再び相関値を求め、相関値が収束するまで、GAT で得られた画像を再び入力画像として、アフィン変換し続ける
4. 相関値が収束した時点で、GAT 処理を終了する

図 4.4 (次ページ) は GAT 処理の流れを示している。左上の画像が入力画像で、左下の画像が参照画像である。真ん中に、2 画像のずれがわかりやすいように 2 画像を重ね合わせた図を用意した。そしてその 2 画像に対して GAT 処理を行い、収束した画像が右の画像である。2 画像が GAT によって変形し、ずれがなくなっているのが分かる。

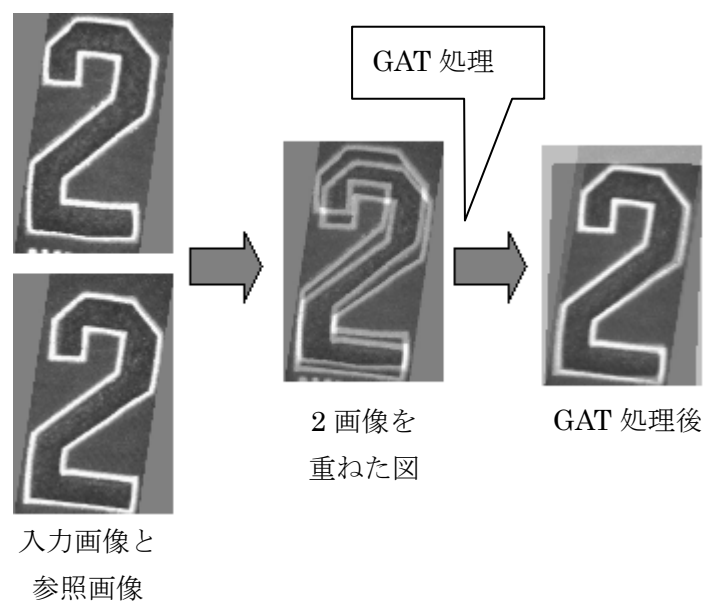


図 4.4 GAT 処理の流れ

第5章 実験結果

5.1 大津基準による2値化

計 698 枚のカラー文字画像の2値化处理結果を、目視により、次の3種類に分類した。

成功：文字部分のみが正しく黒となったもの。但し、微少雑音は許容する。

不十分：文字と背景が反転しているもの、あるいは文字以外に余分な黒部分の塊があるもの。

失敗：カラー空間での2クラス分類で失敗して、文字と背景が分離できなかったもの。

表 5.1 に、2 値化处理の分類結果を示す。

表 5.1 より、方式2により、方式1で「不十分」と判定された2値化画像が21枚救済されたことが分かる。一方、「失敗」と判定された画像は、2クラス分類での失敗であり、図と地の判定では救済できない。

表 5.1 2 値化处理の分類結果

図と地 の判定手法	分類毎の枚数 (%)		
	成功	不十分	失敗
方式1	505 (72.3)	71 (10.2)	122 (17.5)
方式2	526 (75.3)	50 (7.2)	122 (17.5)

図 5.1 に、図と地の判定に方式 2 を用いた 2 値化処理における分類結果の画像例を示す。

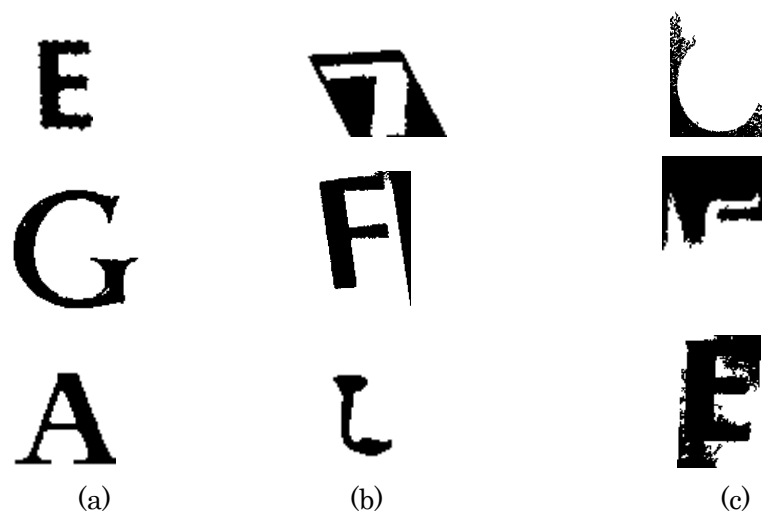


図 5.1 2 値化処理の分類結果の画像例（方式 2）

(a) 成功 (b) 不十分 (c) 失敗

5.2 クラスタリングによる 2 値化

カラー空間を RGB 空間から HSV 空間へ変換を行い、クラスタ数 $k=5$ として 2 値化を行い、認識処理に GAT 相関法を利用して、認識率の比較を行った。式(5)より計算すると、生成される 2 値画像は 30 枚となる。

クラスタリングによる 2 値化の認識処理には GAT 相関法を用いる。さらに GAT 相関法の性能を高めるために、節 4.1 による前処理を画像サイズ 120×80 画素、重心からの黒画素への平均距離が一定値 $= 25.0$ として大きさの正規化を行った。

以下の 4 つの方法の認識率を比較した。

- ① カラー空間での大津基準による 2 値化と図と地の判定方式 1 を利用した方法
 - ② k 平均法による 2 値化（節 3.2）の方法
 - ③ k 平均法によって作成された全 30 枚を利用して、1 番高かった相関値を利用する方法
 - ④ k 平均法によって作成された全 30 枚の 2 値画像の中から目視で 1 枚を選択する方法
- 表 5.2（次ページ）にそれぞれの認識率を示す。

表 5.2 より、 k 平均法による方法②、③は大津基準による方法①より大幅に認識率が下がっていることが分る。しかし、全 30 枚の中から目視で 1 枚を選択した方法④では認識率が大幅に向上している。これより、クラスタリングによる 2 値化法は、制御により正しい 2 値画像のみを選び出すことができれば、大津基準に基づく 2 値化法の性能を上回

る可能性があることが分かる。

表 5.2 上記①－④による 2 値化の認識率

2 値化方法	認識率
①	78.9%
②	62.5%
③	67.8%
④	88.3%

5.3 変形耐性認識

前節 5.1 で方式 2 により「成功」と分類された計 526 枚の 2 値文字画像に対して、接距離と GAT 相関法を適用して認識実験を行った。

表 5.3 に、接距離と GAT 相関法の認識率を示す。表中、単純相関法は、位置と大きさの正規化処理を施してから、単純に相関マッチングを行った場合を指す。

表 5.3 より、接距離と GAT 相関法の認識精度は前者が上回るもののほぼ同等で、単純相関法における誤認識の約 50%を救済していることが分かる。

表 5.3 2 値化成功文字に対する認識率

認識手法	認識率
単純相関法	80.4 %
接距離	91.6 %
GAT 相関法	90.3 %

今回、前処理として位置と大きさの正規化処理を施したため、平行移動と伸縮変形への耐性要求は軽減されている。これに対し、回転成分については適当な前処理がないため、認識手法で吸収する必要がある。

そこで、この回転耐性を評価するため、62 枚のテンプレート画像を－45 度～＋45 度まで 5 度刻みで回転したものと元の正立テンプレート画像の間でマッチングを行い、最適マッチングで得られる正規化相互相関値の平均値を回転角毎に調べた。

図 5.2（次ページ）に、単純相関法、接距離、GAT 相関法での回転耐性の評価結果を示す。

図 5.2 より、接距離、GAT 相関法とも単純相関法と比較して明らかに回転耐性で優れるが、特に回転角が 20 度を越えると GAT 相関法の優位性が顕著である。

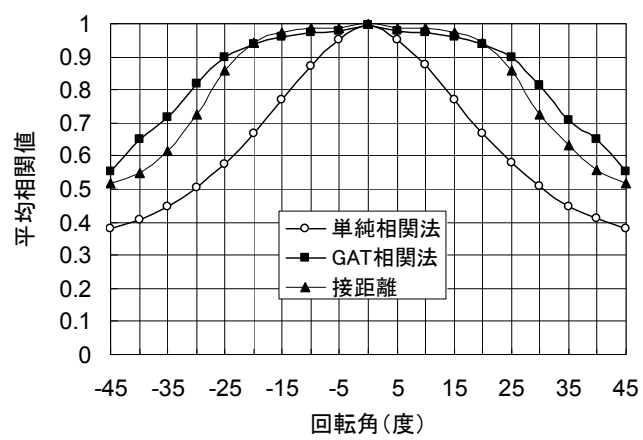


図 5.2 相関値に基づく回転耐性の評価

第6章 考察

6.1 大津基準による2値化の課題

節5.1で2値化結果を3種類に分類したが、「失敗」と「不十分」と分類されたものに対する今後の課題を考察する。

「失敗」については、節3.1で述べたRGBカラー空間での2クラス分類としての最適分離平面の選択という手法の限界が考えられる。失敗した画像例の分析から次のような問題点が明らかになっている。

- ① 大津基準で決定された閾値と異なる閾値で正しい2値化が得られることがある。
- ② 画像全体に一律で1つの投影軸と閾値を適用することに限界がある。局所的な最適化が必要な場合がある。
- ③ 多色文字については、カラー空間での2クラス分類では対応できない場合がある。
- ④ カラー空間の選択に任意性があり、どんな空間を採用すべきか明確な指針がない。

上記の①、②については今後、基礎的な検討が必要である。③、④についてはカラー空間での多クラスへのクラスタリングや複数閾値の選択で対応できると期待される[3]が、最終的な2値化のためには、後段の図と地の判定と連携した最適組合せの問題に帰着する。

「不十分」については、節3.2で述べた図と地の判定処理の高度化が必要である。基本的には「文字らしさとは何か」について定量的に評価できる有効な基準を案出することになる。ただ、次のような性質も留保条件が付くところに「図と地の判定」の難しさがある。

- ① 文字を含んで矩形で切出した画像において、画像枠の部分は背景に属する割合が大きい。
- ② 文字に属する領域はほぼ一定の線幅である。
- ③ 文字の連結成分数は、面積の大きいものに限定すると、極めて少数である。

上記を考慮した上で、複数の特徴量を用いた、より統計的な判定方式の導入が有力と期待される。

6.2 クラスタリングによる2値化の課題

クラスタリングによる2値化は以下の問題点がある。

- ① カラー空間の選択
- ② 作成された複数枚の2値画像からの正しい2値画像の選択
- ③ クラスタリングによる問題

上記①は、大津基準による 2 クラス分類の課題と同じである。文献[22]はクラスタリングにおいて、どのカラー空間を利用すれば、正しく 2 値化できるかを検証している。

②は大津基準による 2 クラス分類後の図と地の判定と同じように、文字らしさを定量的に評価できる有効な基準を考案する必要がある。

③はクラスタリング手法の問題であり、2 つある。1 つ目はクラスタ数 k を予め決めておかなければならないことである。今回の実験では $k=5$ として 2 値化を行ったが、クラスタ数を少なくして、作成される 2 値画像をなるべく抑えた方が、正しい 2 値画像を選ぶことが簡単である。2 つ目はクラスタの初期中心点の選択法によってクラスタリング結果が変動するために、正しい 2 値画像が作成されるかが左右されてしまう問題である。

図 6.1 に入力画像とクラスタリング結果の画像例を示す。

クラスタ数 $k=5$ でクラスタリングした結果であり、クラスタリング後の画像は 5 種類の色で描かれており、同じ色のところが同じクラスタであるということを示している。なお、5 種類の色はランダムで決定されているので色の意味はない。図 6.1 の上はクラスタに違いがあるが、二分化によって正しい 2 値画像が作成することができる。下は二分化によって正しい 2 値画像ができたりできなかったりということが発生してしまう。



上：クラスタに違いがあるが、二分化によって正しい 2 値化画像ができる画像例

下：クラスタに違いがあり、二分化によって正しい 2 値化画像ができない画像例

図 6.1 入力画像とクラスタリング結果の画像例

この初期中心点の問題はクラスタ数 $k=3$ でクラスタリングを行っても発生し、クラスタ数を少なくした方が、クラスタの結果がばらつくようである。

図 6.2 (次ページ) にクラスタ数 $k=5$ と 3 の違いの画像例を示す。

図 6.2 の 1 番左の画像がクラスタ $k=5$ でクラスタリングした結果であり、毎回同じようなクラスタの結果が現れた。しかし、クラスタ数 $k=3$ でクラスタリングした結果(1 番左以外の画像 4 枚)は毎回クラスタの結果が変わってしまった。



左：クラスタ数 $k=5$ その他：クラスタ数 $k=3$

図 6.2：クラスタ数が違うクラスタリングの画像例

クラスタリングの問題であるクラスタ数を考えると、できるだけクラスタ数を少なくして、生成される 2 値画像を減らした方が有効であると考えられる。しかしながら、クラスタ数を減らすと、クラスタの結果がばらつくという問題が発生してしまう。この問題を解決するには文字らしさの考慮や、カラー空間のプロットを考慮した初期中心点の最適な決定法が求められる。

6.3 変形耐性認識の課題

節 5.2 で述べたように、正しく 2 値化された文字画像に対して、接距離の認識率が 91.6%、GAT 相関法の認識率が 90.3%に止まっている。

図 6.3 に、接距離と GAT 相関法でともに誤認識となった 2 値化文字の例を示す。



図 6.3 両手法ともに誤認識となった 2 値化文字の例

図 6.3 より、誤認識の要因として、(1)類似文字の存在、(2)テンプレートと異なる字体、(3)大きな変形、があることが分かる。ただ、これらは文字認識一般に指摘される要因でもある。

情景内文字認識は、未知の字体や変形が出現する open world での認識課題であるため、多数の学習サンプルを用いた統計的手法は適用しにくい。この場合、接距離において細め・太めの TV を加えたように、予想しうる変形成分についてのモデル化により変形耐性を高め

ることは有効である。また、限られた学習サンプルを用いて、予想される変形に対する多様体を空間内に内外挿するアプローチにも大きな期待がある[21]。

6.4 今後の展開と持論

この章では節 6.1、6.2、6.3 の考察をふまえ、さらに今まで研究してきた経験から今後の展開を述べる。

6.4.1 2 値化

図 6.4 に文字が多色になっている画像とその画像を節 3.1 の方法で 2 値化を行った画像を示す。



左：テスト画像 右：節 3.1 の方法で 2 値化を行った画像

図 6.4 多色文字の画像例

図 6.4 の 2 値化の結果は 1 つの閾値によって一方が白と黄色、もう一方が灰色と青というふうに分かれ、正しい 2 値化（大文字 ‘M’）を得ることができた。しかし、正しい 2 値化ができたのはまったくの偶然であると言える。

また、図 6.5 の画像は図 6.4 の左のテスト画像を強制的に色を変えた画像である。



図 6.5 図 6.4 のテスト画像の色を強制的に変えた画像

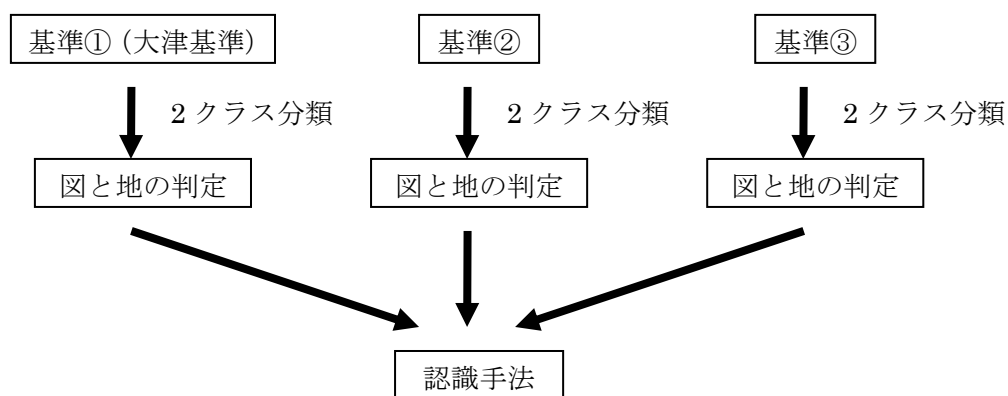
アルファベットを知っている人間なら図 6.5 を見て文字部を当てることができるだろう。しかし、アルファベットを知らない人間に文字部が言い当てることができるだろうか？ある人は赤い部分、ある人は白い部分と言うかもしれない。アルファベットを知っている人間が文字部を当てることができるのは大文字 M という文字の存在を知っているからだ。つまり、図 6.5 のような画像の場合、文字部を当てる（正しい 2 値化）を行うにはアルファベ

ットの知識がなければならない可能性がある。よって、1 番の最適な 2 値化を行うには文字の知識を使うことだと私は考えている。

しかしながら、今現在、2 値化に文字の知識を入れるといった具体的なアイデアはない。そこで私が考えているアイデアは主に 2 つである。

- ① 大津基準以外の基準を採用し何枚かの 2 値画像を作成する方法
- ② 色情報と画像の位置情報を組み合わせたクラスタリング

節 3.1 の大津基準を用いた 2 値化と GAT 相関法で誤認識した画像の中から 80 枚ほどを分析した結果、約 4 割は閾値を変えることによって、より正しい 2 値化ができるようになった。つまり、上記①のような手法でより正しい 2 値化ができる可能性があるということだ。①の手法で大津基準以外の基準を 2 つ利用した場合、フローチャートにすると図 6.6 のようになる。



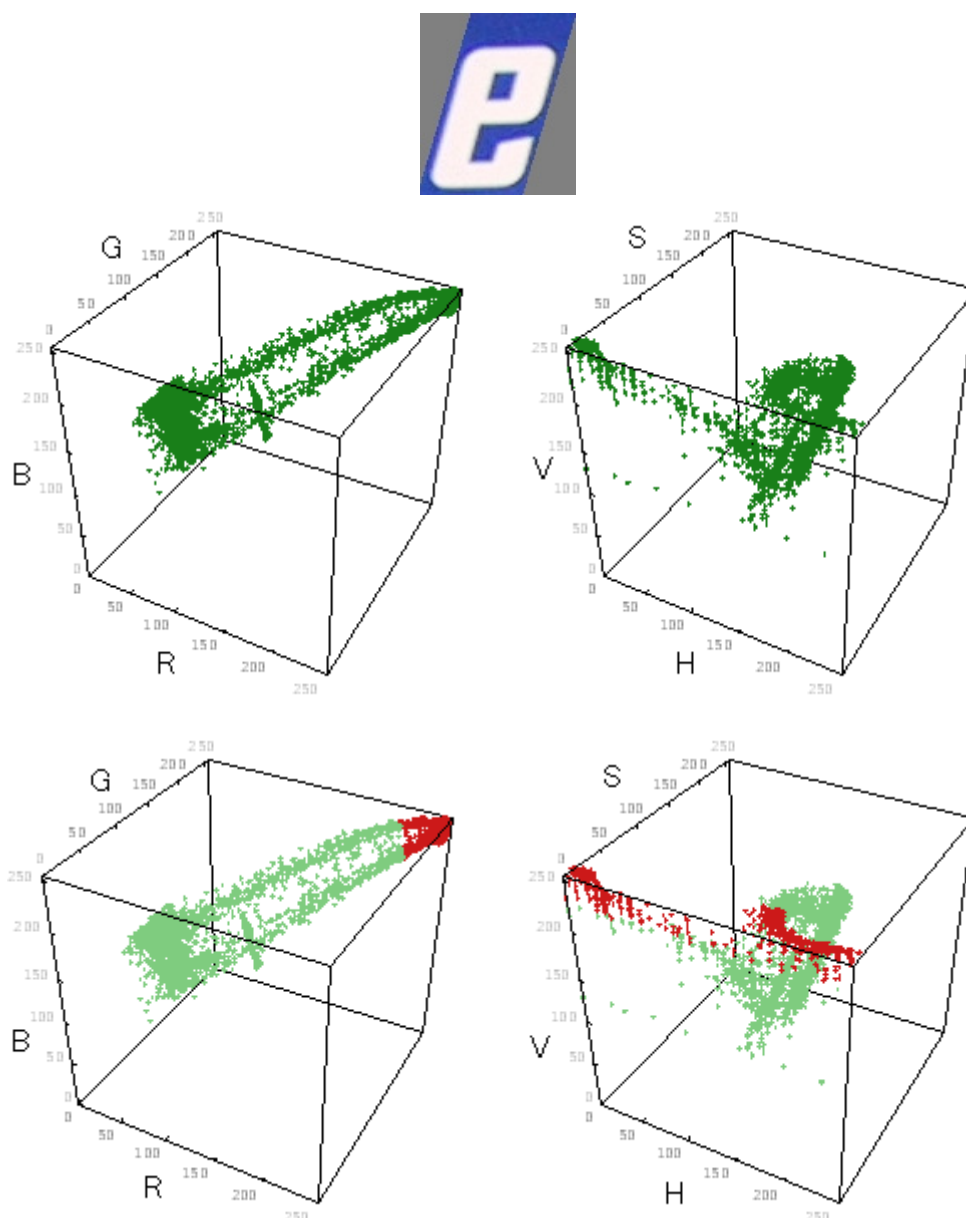
3 枚の 2 値画像の 1 番高い相関値に認識

図 6.6 大津基準以外の基準を利用したフローチャート

上記②は、例えば、カラー空間に RGB 空間を利用した場合、RGB+x, y を加え、5 次元でクラスタリングを行うということになる。その場合、距離を計算する場合に色と位置の比をどうするかという問題や初期中心点の決定はどうするかという問題がある。さらに、1 番難しいのは作成された画像から正しい 2 値画像の選択法をどうするかであろう。アルファベットや数字はある基準によって作成されたものではないので、文字らしさを掴むことは簡単ではない。

いずれにしても、画像をカラー空間で視覚的に見ることは有効である。その際には Mathematica の利用を勧める。

図 6.7 (次ページ) に Mathematica で 3D 表示した例を示す。



上：テスト画像 中：カラー空間(RGB と HSV)の 3D 表示
 下：文字部（赤）と背景（エメラルド）で表示したカラー空間(RGB と HSV)の 3D 表示
 図 6.7 Mathematica による 3D 表示

図 6.7 の下の文字部と背景を違う色で表示した図を見ると、RGB 空間の方が文字部が集まっていることがわかり、このテスト画像の 2 値化を行うには RGB 空間の方がよいであろうことがわかる。

Mathematica は表示だけでなく、あらゆる計算や画像処理などもでき、画像をよく解析することができるので、新しい 2 値化のアイデアが生まれるきっかけになるかもしれない。

6.4.2 変形耐性認識

節 6.3 の図 6.3 をより詳しく解析すると、まずテンプレート画像と異なる字体がある。
図 6.8 にテンプレート画像と異なる字体の画像例を示す。



図 6.8 テンプレート画像と異なる字体の画像例

このようにテンプレート画像とテスト画像の字体が異なる場合は、単純に新しい字体を加えるという案がある。しかし、字体を加えることによって認識率が低下してしまうという結果がある[15]。しかし、文献[15]はすべての 2 値画像をテスト画像としているので、テスト画像を正しい 2 値画像のみにした場合は認識率が向上する可能性はある。

次に類似文字の存在である。図 6.9 に類似文字の存在のために誤認識した画像例を示す。



図 6.9 類似文字の存在のために誤認識した画像例

図 6.9 の左図の正解文字は ‘J’ である。しかし、文字 ‘I’ との正規化相互相関値が高くなってしまい誤認識してしまった。図 6.9 の右図の正解文字は ‘T’ である。しかし、文字 ‘I’ との正規化相互相関値が高くなってしまった誤認識してしまった。一見、なぜ文字 ‘T’ より文字 ‘I’ の方が正規化相互相関値高くなるのかと不思議に思う。これは、図 6.9 の右図の ‘T’ の横棒が細く、縦棒が太いことが原因である。部分的な変形に対応するために、接距離による接ベクトルを部分的に作成することにより正しく認識できるかもしれない。

最後に図 6.10 にテスト画像がテンプレート画像に比べて変形が大きいために誤認識してしまった画像例を示す。



図 6.10 変形が大きいために誤認識してしまった画像例

大きな変形に対応するためには GAT 相関法の向上、特徴量の工夫などが期待される。

続いて GAT 相関法と接距離の認識の違いについて調べる。節 5.3 の表 5.3 より GAT 相関法と接距離の認識率の差はほとんどない。接距離の方がわずかに 7 枚多くを正認識しただけである。

図 6.11 に GAT 相関法のみで誤認識となった画像例、図 6.12 に接距離のみで誤認識となった画像例を示す。



図 6.11 GAT 相関法のみで誤認識となった画像例



図 6.12 接距離のみで誤認識となった画像例

図 6.11 と図 6.12 を見比べてみても、GAT 相関法と接距離の認識の差ははっきりわからない。しかし、誤認識した画像を詳しく調べると、若干ではあるが、接距離が文字の太め細めの変形に対応していることにより、文字の太め細めの変形に対して強いと思われる。

本論文の最後の方、誤認識画像一覧に誤認識した画像名を載せておく。

第7章 むすび

カメラ・ビデオ入力による情景画像から文字領域を抽出して認識する技術は、ロボットの目や自動ナビゲーションなどで中核的役割を担うものと期待される。

本論文では公開データベース ICDAR2003 robust OCR dataset から1文字ずつ切出されたカラー文字画像 698 枚を使用して、最適 2 値化と変形耐性認識の検討を行った。

2 値化処理では、大津基準を適用した手法とカラー空間にクラスタリングを適用した手法の比較を行った。大津基準を適用した 2 値化法では「図と地」の判定処理に黒白比率、黒連結成分の幅や成分数などの文字らしさを用いることにより、正しい 2 値化率 75.3%を達成した。クラスタリングを適用した 2 値化法はクラスタ群の二分法を改良できれば大きな可能性があることを示した。さらに、2 値化実験結果から「失敗」や「不十分」の要因を分析して今後の課題を述べた。

変形耐性認識では、正しく 2 値化できた文字パターンに対して GAT 相関法および接距離を適用した結果、約 91%の認識率を達成した。また、位置と大きさの正規化処理で吸収できない回転成分については GAT 相関法が耐性で優れることを示した。

今後、章 6 で指摘した課題について具体的な解決策が提案されることを期待する。

実験画像一覧

実験画像の番号をグループごとに載せておく。元画像は JPEG 形式で保存されているので、画像名は数字.jpeg である。以前実験したときのグループの分け方で記述してあるが、今回使用した画像は全 698 枚である。

合計 (698 枚)

clear (199 枚)

37	39	69	70	71	72	73	74
75	76	77	78	79	80	81	82
83	84	85	86	87	88	133	134
135	136	137	138	139	140	141	142
143	144	182	183	186	187	188	189
190	191	192	193	194	195	196	197
198	199	200	201	202	203	204	205
206	207	208	209	210	211	212	213
214	215	216	217	218	219	220	221
222	223	224	225	226	227	228	229
230	231	232	233	234	235	236	237
238	239	240	241	242	243	244	245
246	247	248	249	250	251	252	253
254	255	256	257	258	259	260	261
262	263	264	265	266	267	268	634
673	675	676	677	678	679	680	681
682	683	684	685	686	687	688	689
690	691	692	693	694	695	696	697
698	699	700	701	702	703	704	705
706	730	731	732	733	734	735	736
737	738	739	740	741	742	743	744
745	747	748	749	750	751	752	753
754	755	756	757	758	759	760	761
818	819	820	821	822	824	825	826
827	831	832	833	834	835	836	

Background design (130 枚)

1	13	14	15	16	17	18	19
20	21	22	23	24	25	38	40
41	47	48	49	50	51	52	53
54	55	56	57	58	59	112	113
114	152	154	156	158	160	280	281
282	329	330	331	332	333	334	335
336	337	338	339	340	341	342	343
344	345	346	347	348	349	350	351
352	353	354	355	356	357	358	359
360	367	368	369	370	638	639	640
641	648	649	650	651	652	653	654
655	791	792	793	794	795	796	797
798	799	800	801	802	803	804	805
806	807	808	809	810	811	812	823
837	838	839	840	841	842	843	844
845	846	847	848	849	850	851	852
853	854						

Multi-color character (54 枚)

64	65	66	67	68	111	269	270
272	285	286	707	708	709	710	711
712	713	714	715	716	717	718	719
720	721	722	723	724	725	726	727
728	729	762	763	764	765	766	768
769	770	771	772	773	774	775	776
777	778	779	780	781	782		

Nonuniform lighting (40 枚)

26	27	28	29	30	31	32	33
34	35	36	115	116	117	185	379
635	636	637	642	643	644	645	646

647	667	668	669	670	671	672	674
783	784	785	786	787	788	789	790

Little contrast (37 枚)

273	274	275	276	277	278	279	284
427	428	430	433	437	440	441	442
468	470	472	493	495	496	514	517
530	533	534	535	536	537	545	546
547	548	815	816	817			

Blurring (210 枚)

42	43	44	45	46	89	90	91
92	93	94	95	96	97	98	99
100	101	102	103	104	106	107	109
110	119	123	125	126	127	128	129
130	131	145	146	147	148	150	151
161	162	163	164	165	166	167	172
177	178	179	287	288	289	290	291
293	294	295	296	297	298	300	302
304	305	306	307	310	311	313	314
316	317	318	319	320	321	322	323
324	325	326	327	328	361	362	363
364	365	366	372	373	374	375	376
377	380	381	382	383	384	385	386
387	388	389	390	391	392	393	394
395	396	397	400	401	402	403	404
405	406	407	408	411	412	413	414
415	417	419	420	425	443	444	445
448	449	450	475	497	519	520	524
528	529	549	550	551	552	553	554
555	556	557	558	559	560	561	562
563	564	565	566	567	568	569	570
571	572	573	574	575	576	577	578

579	580	581	582	583	584	585	607
608	609	610	611	612	613	614	615
616	617	619	620	621	622	623	625
626	627	628	630	631	632	633	828
829	830						

Serious distortion (28 枚)

3	4	5	6	8	9	12	118
168	371	378	586	587	588	591	592
656	657	658	659	660	661	662	663
664	665	666	767				

誤認識画像一覧

正しく 2 値化できた 526 枚を使用して誤認識となった画像番号を載せておく。

GAT 相関法のみで誤認識となった画像

16	23	25	125	160	172	233	276
313	353	437	534	641	673	685	740
742	754	760	782	790			

接距離のみで誤認識となった画像

86	88	111	114	273	383	394	533
677	689	708	723	735			

GAT 相関法と接距離で誤認識となった画像

3	9	55	65	68	130	167	298
304	336	372	377	378	448	570	619
622	623	640	643	648	653	664	688
706	718	730	731	732	736		

謝辞

日ごろ御指導頂いた若原教授に深い感謝の意を表します。忙しい時間の中、研究内容や論文の添削や発表など多くの時間を割いていただき、ありがとうございました。おかげで、この論文を仕上げることができ、さらに多くの事を学ぶことができました。

また、この論文を仕上げるにあたって協力をしてくださった若原研究室のメンバーに感謝いたします。ありがとうございました。

参考文献

- [1] 森, 澤木, “低品質文字の認識手法とその応用に関するサーベイ,” 信学技報, PRMU2001-275, March 2002.
- [2] D. Doermann, J. Liang, and H. Li, “Progress in camera-based document image analysis,” *Proc. 7th Int. Conf. on Document Analysis and Recognition*, vol. I, pp. 606-687, Edinburgh, Scotland, Aug. 2003.
- [3] 黄瀬, 大町, 内田, 岩村, “カメラを用いた文字認識・文書画像解析の現状と課題,” 信学技報, PRMU2004-246, March 2005.
- [4] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, “ICDAR 2003 robust reading competitions,” *Proc. 7th Int. Conf. on Document Analysis and Recognition*, vol. II, pp. 682-687, Edinburgh, Scotland, Aug. 2003.
- [5] <http://algoval.essex.ac.uk/icdar/Datasets.html>.
- [6] M. Yokobayashi and T. Wakahara “Binarization and Recognition of Degraded Characters Using a Maximum Separability Axis in Color Space and GAT Correlation,” *Proc. 18th Int. Conf. on Pattern Recognition*, Vol. II, pp. 885-888, 2006.
- [7] B. Gatos, I. Pratikakis and S. J. Perantonis “Adaptive degraded document image binarization,” *IEEE Trans. Pattern Recognition*, Vol. 39, pp. 317-327, 2006.
- [8] W. Niblack, *An Introduction to Digital Image Processin*, Prentice Hall, Englewood Cliffs, NJ, pp. 115-116, 1986.
- [9] J. Sauvola, and M. Pietikainen, “Adaptive document image binarization,” *IEEE Trans. Pattern Recognition*, Vol. 33, pp. 225-236, 2000.
- [10] N. Otsu, “A threshold selection method from gray-level histogram,” *IEEE Trans. Systems, Man and Cybernetics*, vol. SMC-9, pp. 62-69, 1979.
- [11] J. McQueen “Some methods for classification and analysis of multivariate observations,” *Proc. 5th Berkely Symposium on Mathematica Statistics and Probability*, pp. 281-297, 1967.
- [12] T. Wakahara, Y. Kimura, and A. Tomono, “Affine-invariant recognition of gray-scale characters using global affine transformation correlation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-23, pp. 384-395, 2001.
- [13] P. Simard, Y. LeCun, and J. Denker, “Efficient pattern recognition using a new transformation distance,” *Advances in Neural Information Processing Systems*, vol. 5, pp. 50-58, Morgan

Kaufmann, 1993.

- [14]R.G. Casey, "Moment normalization of handprinted characters," *IBM J. Res. Develop.*, vol. 14, pp. 548-557, 1970.
- [15]横林, "Binarization and Recognition of Degraded Characters Using a Maximum Separability Axis in Color Space and GAT Correlation," 卒業論文, pp. 14-15, The Faculty of Computer and Information Sciences, Hosei University, March 2006.
- [16]R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Second Edition, Chap. 9, Prentice Hall, 2002.
- [17]<http://ja.wikipedia.org/wiki/HSV%E8%89%B2%E7%A9%BA%E9%96%93>
- [18]M. Gabbouj, P. Haavisto and Y. Neuvo, "Recent advances in median filtering," E. Arikan(Ed.), *Communication, Control, and Signal Processing*, Vol. II Elsevier Science Publishers B, Ankara, Turkey, pp.1080-1094, 1990.
- [19]平山, 大町, 阿曾, "カラー情報を利用した情景画像中の文字列の高精度抽出," 信学技報, PRMU2004-247, March 2005.
- [20]K. Wang and J. A. Kangas, "Character location in scene images from digital camera," *IEEE Trans. Pattern Recognition*, Vol. 36, pp. 2287-2299, 2003.
- [21]堀田, 喜安, 宮原, "局所部分空間法と変形不変性を用いた画像パターン分類," 部分空間法研究会 2006, online publishing, pp. 111-118, Jul. 2006.
- [22]C. Thillou and B. Gosselin "Color text extraction from camera-based images: the impact of the choice of the clustering distance," *Proc. 8th Int. Conf. on Document Analysis and Recognition*, vol. I, pp. 312-316, 2005.