エッジ成分の方向分布と空間的配置に着目した情景画像からの文字列抽出

KITADA, Hideki / 北田, 英樹

(発行年 / Year)
2013-03-24
(学位授与年月日 / Date of Granted)
2013-03-24
(学位名 / Degree Name)
修士(理学)
(学位授与機関 / Degree Grantor)
法政大学 (Hosei University)

2012 年度修士論文

エッジ成分の方向分布と空間的配置に着目した情景画像からの文字列抽出

Character String Extraction from Scene Images Using Direction Distribution of Edge Components and their Spatial Arrangements

2013 年 2 月 18 日　提出

指導教官　　若原　徹　教授

法政大学大学院情報科学研究科情報科学専攻

学籍番号　11T0019　北田（キタダ）　英樹（ヒデキ）

Abstract

This paper proposes a method for character string extraction from scene images using direction distribution of edge components and their spatial arrangements. First, we detect edge components using Canny operator as applied to individual RGB channels and label the connected edge components. Second, we count the number of edge points of labeled edge component and calculate the maximum of x-coordinate, y-coordinate, and minimum of them. Third, we remove such edge components that have too few edge points, too large size of width and/or height, and show too large difference between width and height. Fourth, we generate character string candidates by adding up of edge components obtained through individual RGB channels. Fifth, we extract character strings based on spatial arrangements of edge components. Sixth, we extract isolated characters using direction distribution features of edge components. Finally, we extract character strings composed of concatenated characters and isolated characters based on tentative segmentation and character recognition by evaluating character likeness of individual components. Experimental results made on a total of 249 images extracted from ICDAR2003 robust reading and Text Locating dataset "SceneTrialTest" show that the proposed method achieves a recall rate of 71.3%, a precision rate of 64.5%, and an F measure of 67.8%.

内容概要

本研究では，カラー情景画像を対象に，RGB 成分ごとの微分画像に 2 値化処理を施し，雑音成分を除去して融合し，文字列候補領域となるエッジ成分を抽出し，エッジ成分の外接矩形の空間的配置の条件から文字候補領域を選択し，単一文字候補領域として残された領域や接触文字列候補領域に対して，方向分布特徴を用いた文字／非文字判定を用いて文字列を抽出する手法を提案した.特に，空間的配置条件を利用した手法では抽出が不可能であった接触文字列のエッジ成分においても抽出を可能とする手段として，仮分割を施して分割されたエッジ成分ごとに文字／非文字判定を行うことで，抽出精度の向上を図った.

提案手法を ICDAR2003 の公開データセットに適用し，再現率 71.3%，適合率 64.5%，F 値 67.8%を達成した.

目次

第 1 章　序論

1.1 研究の背景

近年，画像処理やパターン認識技術の研究によって視覚機能を備えた自律移動ロボットの実現の可能性や, 福祉情報工学の分野においては全国で 30 万人を超えるといわれている視覚障害者のための環境内文字読上げシステムの開発などに期待が寄せられている．また，伝票や書籍等の表面上に限られていた文書画像からの文字認識の技術を看板や標識等の 3 次元空間上に拡張することができれば自動車運転の支援や交通監視等の幅広い応用が考えられる．また，デジタルカメラが一般に急速に普及したことによって文字や文書の記録の仕方にも変化が起きている．しかし，画像のデータは容量が大きく数が多くなると整理等が困難であるという問題がある．画像データをテキストデータに変換することができれば容量が小さくなり整理等も容易である．それらを実現するためには，情景画像から早く正確に文字列を抽出することが不可欠である．

1.2 研究の目的

従来，情景画像からの文字列抽出の研究は行われている[1]．それらは大別して 3 つの手法に分かれる．第 1 に画像のエッジ成分を抽出し，文字の性質から文字列を抽出する手法．第 2 に画像のカラー情報を用いた手法，第 3 に上記 2 つの手法を組み合わせた手法とであるが，抽出の精度についてはまだ改善の余地がある．

本研究では, エッジ成分の方向分布と空間的配置に着目した情景画像からの文字列抽出の手法を提案する．まず，画像の RGB 成分に対してそれぞれエッジ成分を抽出して雑音成分を除去し，得られた 3 枚のエッジ画像の和を取ってエッジ画像とする．次に，エッジ画像からエッジの連結成分の外接矩形の中心間の距離や面積比等の空間的配置条件に基づいた文字列抽出を行い，抽出できない単一文字や接触文字については方向分布特徴を用いた文字／非文字判定を用いて文字列抽出を行う手法を提案する．

1.3 研究の構成

本章では，序論として本研究の背景や，どのような目的で行われたかを述べた．第 2 章で使用した画像データを紹介する．第 3 章では，Canny オペレータを用いたエッジ成分の抽出，雑音除去等の文字列候補画像作成，第 4 章でエッジ成分の空間的配置に着目した文字列抽出，第 5 章でエッジ成分の方向分布特徴を用いた文字／非文字判定，第 6 章で実験結果，第 7 章で考察，第 8 章で追加実験について述べ，第 9 章でむすびとする．

第 2 章　使用した画像データ

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．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[2]．図 1 はそれぞれの部門の画像例である．

図 1　(a) Text locating.　(b) Robust word recognition. (c) Robust character recognition.

本研究では，Text locating 部門で用いられた ICDAR2003 robust Reading and Text Locating dataset の SceneTrialTest に含まれる 249 枚の画像を抽出対象とする．これらの画像サイズは 1280×960 から 307×93 までの様々であり，英数文字を含む．また，実験に用いた各閾値を求めるための予備実験には同一の dataset の SceneTrialTrain に含まれる文字列を含む画像 247 枚の画像である．

図 2　実験用画像データの例.

公開されている画像は JPEG 形式であるが，本研究ではフリーソフトの IrfanView[3]を用いて PPM 形式に変換した．またプログラムで処理しやすいようにファイル名を通し番号付きの名前にリネームした．

第 3 章　文字列候補領域抽出

従来研究では,エッジ抽出はカラー画像から作られた1枚の濃淡画像から行われてきた．しかし，本研究ではエッジ抽出の安定化を図るために，RGB の各成分からエッジ強度を得て 2 値化を行い，それらの論理和により最終的なエッジ画像を生成する．得られたエッジ成分の連結領域が文字候補領域となり，それらが横に並んで文字列候補領域となる．

3.1 Canny オペレータによるエッジ抽出

画像の R 成分に Canny オペレータ[4]を施しエッジ成分を得る．

Canny オペレータの説明をする．

STEP1:ガウシアンフィルタによる平滑化を行う．

まず，画像全体をぼかすことにより雑音の影響を減らす．今回用いたガウシアンフィルタの係数は以下の通りである．

$$\frac{1}{159} \times \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix}$$

図 3　ガウシアンフィルタの係数

STEP2:エッジ強度と勾配方向を計算

Sobel フィルタで水平方向のエッジ強度f_x，垂直方向のエッジ強度f_yの計算を行う．

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

図 4　(a) x方向のオペレータ　　(b) y方向のオペレータ

座標(x, y)の画素値を$p(x, y)$，エッジ強度をfとする．

$$f_x = p(x+1, y-1) + 2 \times p(x+1, y) + p(x+1, y+1) - p(x-1, y-1) - 2 \times p(x-1, y) - p(x-1, y+1) \quad (1)$$

$$f_y = p(x-1, y+1) + 2 \times p(x, y+1) + p(x+1, y+1) - p(x-1, y-1) - 2 \times p(x, y-1) - p(x+1, y-1) \quad (2)$$

エッジ強度は以下の式により求める．

$$f = \sqrt{f_x^2 + f_y^2} \quad (3)$$

また，勾配方向の量子化を行う．

$$\theta = \mathrm{atan}\left(\frac{f_y}{f_x}\right) \tag{4}$$

0 度方向：$-0.4142 < \theta \leq 0.4142$

45 度方向：$0.4142 < \theta < 2.4142$

90 度方向：$|\theta| \geq 2.4142$

135 度方向：$-2.4142 < \theta \leq -0.4142$

STEP3:エッジの細線化処理

ぼかして太くなったエッジを細くするために細線化処理を施す．エッジと鉛直方向の隣接画素 2 つと最大でなければ 0 とする．

位置(x, y)のエッジ強度を$f(x, y)$とすると

①0 度方向

$f(x, y) > f(x-1, y)$ かつ$f(x, y) > f(x+1, y)$

上記の式を満たす場合

$f(x, y) = f(x, y)$

上記の式を満たさない場合

$f(x, y) = 0$

②45 度方向

$f(x, y) > f(x-1, y+1)$ かつ$f(x, y) > f(x+1, y-1)$

上記の式を満たす場合

$f(x, y) = f(x, y)$

上記の式を満たさない場合

$f(x, y) = 0$

③90 度方向

$f(x, y) > f(x, y-1)$ かつ$f(x, y) > f(x, y+1)$

上記の式を満たす場合

$f(x, y) = f(x, y)$

上記の式を満たさない場合

$f(x, y) = 0$

④135 度方向

$f(x, y) > f(x-1, y-1)$ かつ$f(x, y) > f(x+1, y+1)$

上記の式を満たす場合

$f(x, y) = f(x, y)$

上記の式を満たさない場合

$f(x, y) = 0$

STEP4:ヒステリシス閾処理

閾値 1 つでエッジかどうかを判定するのではなく，閾値 2 つを用いて適応的に判定する

手法．Th_u，Th_lを用いて以下の条件でエッジかどうかを判定する．

①Th_uより大きいエッジ強度の画素はエッジ

②Th_lより小さいエッジ強度の画素は非エッジ

③Th_l以上Th_u以下のエッジ強度の画素のうちエッジに結合している画素はエッジ

今回は画像全体におけるエッジ強度の平均 μ と標準偏差 σ を用いて，以下の式により 2 つの閾値Th_u，Th_lを求めて，エッジ強度を 2 値化する．

$$Th_u = \mu + \delta_1 \times \sigma \tag{5}$$

$$Th_l = \mu + \delta_2 \times \sigma \tag{6}$$

3.2 雑音成分の除去

得られたエッジ成分の黒画素連結領域にラベリング処理を施し，エッジ成分ごとに処理を行えるようにする．

ラベリング処理とは同一の画素値を持つ部分領域ごとに異なる番号を割り振る処理のことで，ラベル番号ごとに処理を行うことで特定の領域ごとに x 座標の最小値や最大値，y 座標の最小値や最大値，黒画素数や外接矩形等の情報を得ることができるようになる．

エッジ成分ごとに雑音成分か文字候補領域かを判定し雑音成分を除去する．ここで画像の縦の長さを H，横の長さを W，N を画像全体の画素数，n をエッジ領域の画素数，h，w をそれぞれエッジ成分の外接矩形の縦の長さ，横の長さ，cx，cy を外接矩形の中心の x 座標，y 座標，gx，gy を外接矩形の重心の x 座標，y 座標とする．各 x 座標で y 座標の最小値と最大値を求め，その差の平均を y-mean とする．雑音成分は以下の条件を満たすものとする．

$$\frac{n}{N} < \delta_3,\ \mathrm{n} < \delta_4 \times (h + w),\ \frac{h}{w} > \delta_5,$$

$$h > \delta_6 \times H,\ w > \delta_7 \times W$$

$$|cy - gy| > \delta_8 \times H,\ |cx - gx| > \delta_9 \times W,$$

$$y - \mathrm{mean} < \delta_{10} \times H \tag{7}$$

図 5 に処理例を示す．

図 5　　(a)　　(b)　　(c)　　(d)

図 5 で(a)は原画像．(b)は R 成分の濃淡画像．(c)は G 成分の濃淡画像．(d)は B 成分の濃淡画像である．この画像では文字列の背景部分が青に近い色であるため，B 成分の濃淡画像において文字列部分と背景部分をしっかりと分離できていることが分かる．

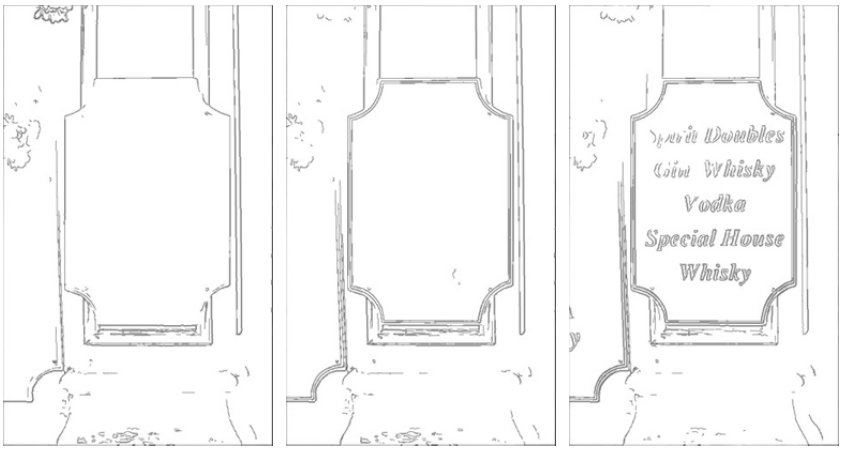


図 6　　(a)　　(b)　　(c)

図 6 で(a)は図 4 の(a)R 成分の濃淡画像からエッジ成分を抽出した画像．(b)は図 4 の(b)G 成分の濃淡画像からエッジ成分を抽出した画像．(c)は図 4 の(c)B 成分の濃淡画像からエッジ成分を抽出した画像である．文字列のエッジ成分はB成分に含まれていることが分かる．

本研究では英数文字を抽出対象の文字列とし，文字列は横方向に並んでいると仮定しているため，「i」，「l」，「1」等より縦方向に長い連結領域を持つエッジ成分は文字列ではないとし除去している．また，画像に占める文字の割合があまりに高すぎると考えられるものや，あまりに低すぎると考えられるものを除去している．

一方，情景画像中には筆記体のように接触して書かれている文字列や，文字同士が離れていても，照明や背景等の雑音成分の影響で文字と文字のエッジ成分が連結してしまうこともあるため，横方向に長い連結領域は除去せずに文字列候補として残しておく．

上記の処理を，G 成分，B 成分に対しても同様に施す．この処理により得られた 3 枚のエッジ画像を 1 枚の文字列候補画像として融合する．融合の仕方は単純に黒画素の論理和を取る．さらに後の処理でエッジ成分の方向分布特徴を用いた文字／非文字判定を行うため，「A」や「O」，前述した「B」や「8」等の単一文字で複数のエッジ成分を持つもののエッジ成分を同一のラベルにする必要がある．第 K 番目の連結領域に第 L 番目の連結領域が含まれる条件として以下の全てを満たす場合，第 L 番目の連結領域のラベル番号を第 K 番目の連結領域のラベル番号に置き換える．ただし，K_Xを第 K 番目の連結領域の x 座標の最大値，K_xを第 K 番目の連結領域の x 座標の最小値，L_Xを第 L 番目の連結領域の x 座標の最大値，L_xを第 L 番目の連結領域の x 座標の最小値，それぞれ y 座標をK_Y，K_y，L_Y，L_yとする．

$K_X > L_X$，$K_x < L_x$，$K_Y > L_Y$，$K_y < L_y$，

$(K_X - K_x) \leq \delta_{11} \times (L_X - L_x)$，

$(K_Y - K_y) \leq \delta_{12} \times (L_Y - L_y)$ (8)

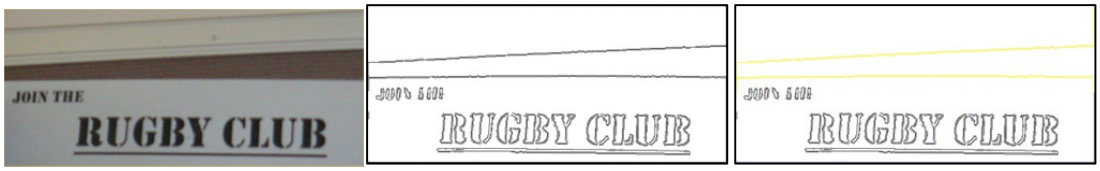
ここでさらに 1 つのエッジ成分の外接矩形の中に複数のエッジ成分を含むものを雑音成分であるとして除去する．

K_{in}を第 K 番目のエッジ成分に含まれるエッジ成分の数として以下の条件を満たすものとする．

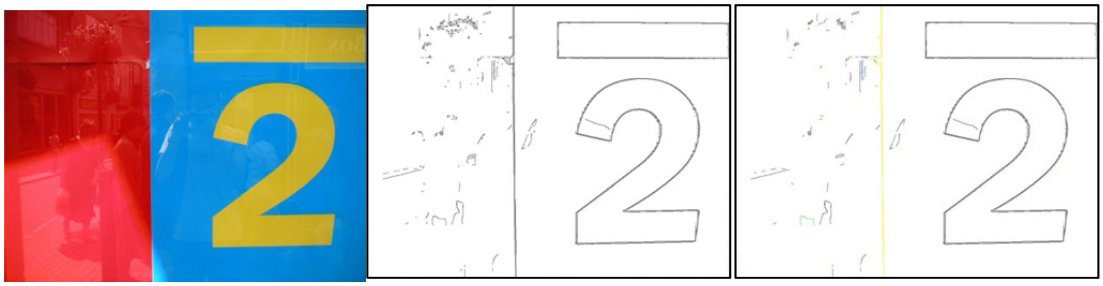
$K_{in} \geq \delta_{13}$ (9)

以下図 7 に雑音成分の除去例を示す．

(1-a) (1-b) (1-c)

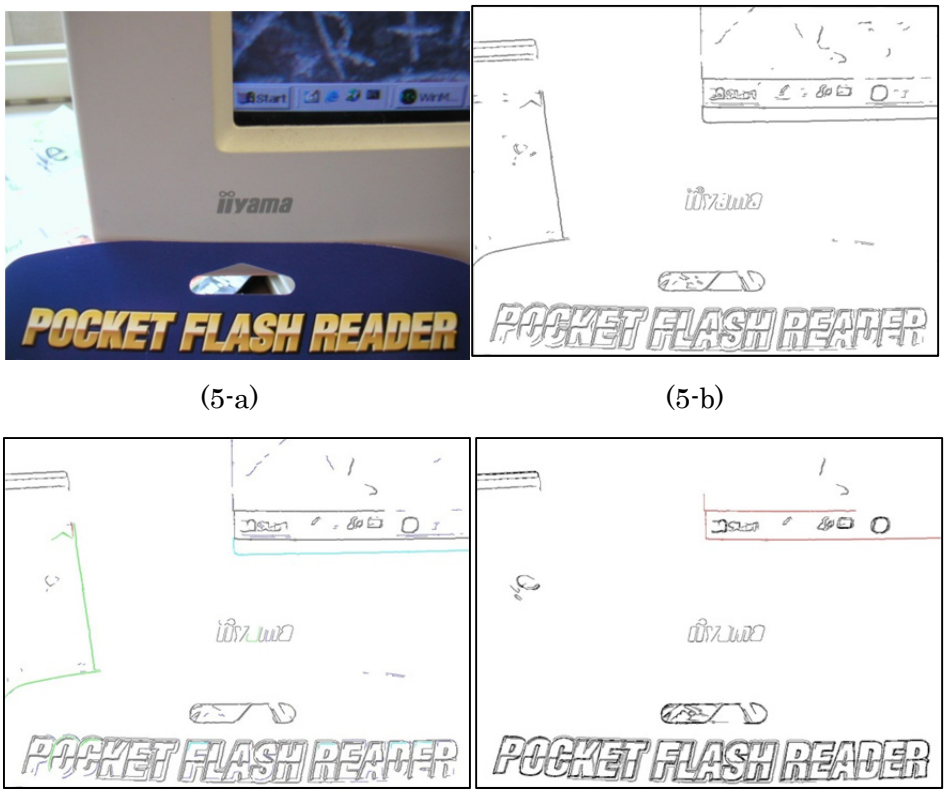


(2-a) (2-b) (2-c)



(3-a) (3-b) (3-c)

(4-a) (4-b) (4-c)



(5-a) (5-b)

(5-c) (5-d)

図 7　雑音成分の除去例.

図 7 で(a)はそれぞれ原画像. (b)がエッジ画像. (c), (d)が雑音成分を除去した画像である. (1)は細かい雑音成分を除去した例である．これは原画像に対して R 成分でエッジを抽出した画像(b)に含まれる細かい点のような雑音成分を除去した画像(c)である．(2)は画像の横の長さに非常に近い長さを持つエッジ成分を雑音成分として除去した例である．原画像に対して B 成分でエッジを抽出した画像(b)に含まれる画像の横の長さに非常に近い長さを持つエッジ成分を黄色に塗りつぶして雑音成分として表示した画像(c)である．(3)は画像の縦の長さに非常に近い長さを持つエッジ成分を雑音成分として除去した例である．原画像に対して B 成分でエッジを抽出した画像(b)に含まれる画像の縦の長さに非常に近い長さを持つエッジ成分を黄色に塗りつぶして雑音成分として除去した画像(c)である．(4)は縦に長いエッジ成分を除去した例である．原画像に対して G 成分でエッジを抽出した画像(b)に含まれる縦に長いエッジ成分を赤く塗りつぶして雑音成分として除去した画像(c)である．また, 連結成分の各 x 座標で y 座標の最小値と最大値を求め, その差の平均が閾値より少なかった場合のエッジ成分を緑色で塗りつぶして雑音成分として除去している．さらに，エッジ成分の黒画素数が外接矩形の縦と横の長さの和の閾値倍未満のエッジ成分を青色で塗りつぶして雑音成分として除去している．(5-c)はエッジ成分の黒画素の重心の座標と外接矩形の中心の座標が大きく異なる雑音成分を除去した例である．原画像に対して B 成分でエッジを抽出した画像(b)に含まれる，エッジ成分の黒画素の重心の座標と外接矩形の中心の座標

が大きく異なる雑音成分を水色で塗りつぶして雑音成分として除去している．(5-d)は原画像に対して雑音成分を除去した画像の論理和を取った画像に含まれる閾値以上のエッジ成分を含むエッジ成分を除去した例である．(5-d)で赤く塗りつぶされたエッジ成分の外接矩形中には多くのエッジ成分を含むため，このエッジ成分は雑音成分として除去されることになる．

第 4 章　エッジ成分の空間的配置に着目した文字列抽出

本研究では英数文字から成る文字列を抽出対象とするため，下記の 4 点を「文字列らしさ」の属性として仮定した．

(1)各文字は横方向に直線的に並んでいる．

(2)各文字の外接矩形の面積が近い．

(3)各文字間の距離が近い．

(4)各文字の大きさが近い．

連結成分が同じ文字列であるかを上記の英数文字列の特徴を基に判断する．各エッジ成分の外接矩形について以下の値を判定に用いる．左側のエッジ成分の外接矩形の幅w_L，高さh_L，それぞれ右側のw_R，h_R，外接矩形の中心の x 座標と y 座標を左右それぞれg_x^L，g_y^L，g_x^R，g_y^Rとする．また，外接矩形の y 座標の最小値をy_b^L，y_b^Rとする．左側の連結成分と右側の連結成分が同一の文字列であるかの判定に用いた式を以下に示す．

(a)　$\dfrac{max(w_L \times h_L,\ w_R \times h_R)}{min(w_L \times h_L,\ w_R \times h_R)} < \delta_{14}$

(b)　$0 < g_x^R - g_x^L < \delta_{15} \times max(w_L,\ w_R)$

(c)　$\left|g_y^L - g_y^R\right| < \delta_{16} \times max(h_L,\ h_R)$

(d)　$|h_L - h_R| < \delta_{17} \times max(h_L,\ h_R)$

(e)　$\left|g_y^L - g_b^R\right| < \delta_{18} \times max(h_L,\ h_R)$　or　$\left|g_y^R - g_b^L\right| < \delta_{18} \times max(h_L,\ h_R)$

(f)　$|h_L - h_R| < \delta_{19} \times max(h_L,\ h_R)$

(g)　$\left|g_y^L - g_y^R\right| < \delta_{20} \times max(h_L,\ h_R)$

(h)　$\dfrac{h_L}{w_L} < \delta_{21} and/or \dfrac{h_R}{w_R} < \delta_{21}$

(i)　$\dfrac{max(w_L \times h_L,\ w_R \times h_R)}{min(w_L \times h_L,\ w_R \times h_R)} < \delta_{22}$

(j)　$0 < g_x^R - g_x^L < \delta_{23} \times max(w_L,\ w_R)$

上記の条件を以下に示す 3 つの場合に分け適用する．

(1) アセンダー(h，b，k)等とディセンダー(p，q，y)が並ぶ文字列の場合

上記の条件の(*a*)，(*b*)，(*e*)，(*f*)，(*g*)を満たす場合に文字列と判定する．

(2)「i」や「l」等の縦に長い文字列が含まれる場合

上記の条件の(*c*)，(*d*)，(*h*)，(*i*)，(*j*)を満たす場合に文字列と判定する．

（3）一般的な文字列の場合

（1），（2）の条件を満たさない一般的な文字列の場合は(*a*)，(*b*)，(*c*)，(*d*)，(*e*)を満たした場合に文字列と判定する．最終的につながったエッジ成分の数が 2 以上である場合，文字列であると判断する．

第 5 章　エッジ成分の方向分布特徴を用いた文字／非文字判定

前章までの処理では，1 文字が単一の連結エッジ成分に対応していると仮定している．しかし，情景画像中には接触文字も多く存在する．また，上記の処理では，2 文字以上横に並んだ文字列を抽出対象としているため，孤立した単一文字の抽出はできない．そこで，孤立した単一文字の候補領域，および接触文字列候補領域に横方向の仮分割を施してほぼ 1 文字単位に分割した単一文字候補領域群に対して，方向分布特徴を用いた文字／非文字の判定を行う．この判定により，文字らしいとして受理された場合は，それらも文字列として抽出する．

5.1 特徴抽出

文字／非文字の判定に用いた特徴は加重方向指数ヒストグラム特徴[5]を簡略化した輪郭方向分布特徴である．

まず，サイズ 120×80 の画像を 96 個の正方ブロック(サイズ 10×10)に分割する．ブロックごとにエッジ成分の輪郭方向の 4 方向別ヒストグラムを算出する．このヒストグラムに対し，2 次元ガウスフィルタをかけ，96 次元ベクトルへと次元圧縮したものを輪郭方向分布特徴とする．

図 8 に 2 次元ガウスフィルタの重み係数，図 9 に輪郭方向分布特徴の概念図を示す．

0.000	0.009	0.017	0.009	0.000
0.009	0.057	0.105	0.057	0.009
0.017	0.105	0.194	0.105	0.017
0.009	0.057	0.105	0.057	0.009
0.000	0.009	0.017	0.009	0.000

図 8　2 次元ガウスフィルタの重み係数.

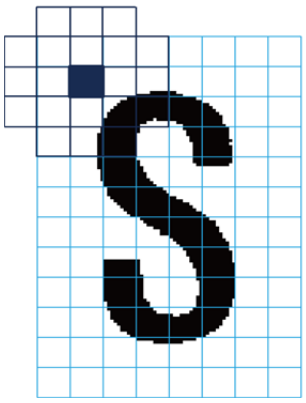


図 9　輪郭方向分布特徴の概念図.

5.2 改良投影距離法

本節では，改良投影距離法とその基礎となるベイズ識別関数について説明する．

5.2.1. 固有値・固有ベクトル

本研究で学習に用いたデータセットは the Chars74K のデータセット[6]である．図 10 に例を示す．

図 10　学習に用いたデータセットの例.

このデータセットは 0 から 9 の 10 種類, アルファベットに対しては大文字の A から Z の 26 種類，そして小文字の 26 種類，計 62 種類の文字がそれぞれ 1016 種類のフォントで描かれたデータセットである．前節で述べた特徴抽出を行うため，各文字サンプルに Canny オペレータを施してエッジ抽出を行い，位置と大きさの正規化を施し，さらに Hilditch の細線化処理[7]を行った．この細線化画像に対して輪郭方向分布特徴の抽出を行った.

まず，位置と大きさの正規化であるが，提案手法ではモーメント法に基づく手法を採用した．任意の画像に対し，縦 120，横 80 の画像に正規化を施す．まず位置の正規化についてであるが画像内の黒画素領域の *x* 座標，*y* 座標を足しこんでいき黒画素数で割ることにより，黒画素の重心の座標(gx，gy)を求める．そして重心が画像の中心(40，60)にくるように平行移動する．ここで重心を求める操作が画像の 1 次モーメント算出に相当する．次に大きさの正規化であるが，画像の黒画素領域の重心(*gx*，*gy*)を求めて，各黒画素から重心までの平均距離 rad を算出する. 予め定めた正規化半径 RADIUS と rad の値が等しくなるように，黒画素部分を重心の周りに一様に伸縮する．この操作が大きさの正規化となる．各黒画素から重心までの平均距離 rad を求める操作が画像の 2 次モーメント算出に相当する．なお，提案手法では正規化半径 RADIUS は予備実験により 25.0 の値を採用した.

位置と大きさの正規化を施した画像において Hilditch の細線化処理を施す.

続いて，固有値と固有ベクトルを求める．まず，同一カテゴリに属する特徴ベクトル $f_i(i = 1,2,\dots,96)$を用いて平均g_0を求める．次に各特徴ベクトル$f_i(i = 1,2,\dots,96)$から平均g_0を引いて，平均特徴ベクトルからの差異を表すベクトル$\tilde{f}_i$を次式により求める.

$$\tilde{f}_i = f_i - g_0 \tag{9}$$

差異を表すベクトルに対する共分散行列を次式により求める.

$$\tilde{M} = \frac{1}{N}\sum_{i=1}^{N} \tilde{f}_i \tilde{f}_i^t \tag{10}$$

ここで，行列$\tilde{M}$の固有値と固有ベクトルの計算には特異値分解[8]を用いて，計算の効率化を図った．

5.2.2. 擬似ベイズ識別関数

擬似ベイズ識別関数は，ベイズアプローチから得られる最適識別関数を計算量・記憶容量を削減するために変形して得られる識別関数で，次式で定義される．

$$g(X) = (N + N_0 + 1)\ln\{1 + \frac{1}{N_0\sigma^2}[\|X - M\|^2\} - \sum_{i=1}^{k}\frac{(1-\alpha)\lambda_i}{(1-\alpha)\lambda_i + \alpha\sigma^2}\left[\Phi_i^T(X - M)\right]^2]\} + \sum_{i=1}^{k}\ln\left((1-\alpha)\lambda_i + \alpha\sigma^2\right) - 2\ln P(\omega) \tag{11}$$

ただし，

$$\alpha = \frac{N_0}{N+N_0} \tag{12}$$

ここで，N は各クラスの学習標本の大きさ，N_0は特徴ベクトル X の事前分布の共分散行列の信頼度を表す定数，$P(\omega)$はクラスの事前確率であり，λ_i，φ_iはそれぞれ X の共分散行列の第 i 固有値と固有ベクトル，k は識別に用いる固有ベクトルの数である．

5.2.3. 改良投影距離法

前節で定義した擬似ベイズ識別関数をさらに近似することを考える．すなわち，すべてのクラスで共分散行列の行列式，$P(\omega)$，N，N_0が等しいと仮定すると，式の第 2 項，第 3 項は定数項となり，第 1 項の係数$(N + N_0 + 1)$はクラス間で共通となる．この結果，次式で定義される改良投影距離によって，式の擬似ベイズ識別関数を近似することができる．

$$g(X) = \|X - M\|^2 - \sum_{i=1}^{k}\frac{(1-\alpha)\lambda_i}{(1-\alpha)\lambda_i + \alpha\sigma^2}\left\{\Phi_i^T(X - M)\right\}^2 \tag{13}$$

ここで，αは[0, 1)の実数パラメータであり，識別率を最大化するように設定する．
また，識別に用いる固有ベクトルの数は，固有値の累積寄与率を用いて定めることとした．すなわち，固有値を降順にソートして，次式で定義される固有値の累積寄与率ηが一定値を超える最小の次元数を k とした．

$$\eta = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{j=1}^{N}\lambda_j} \geq \beta \tag{14}$$

ここで学習データの認識実験により，パラメータの値を決定していく．

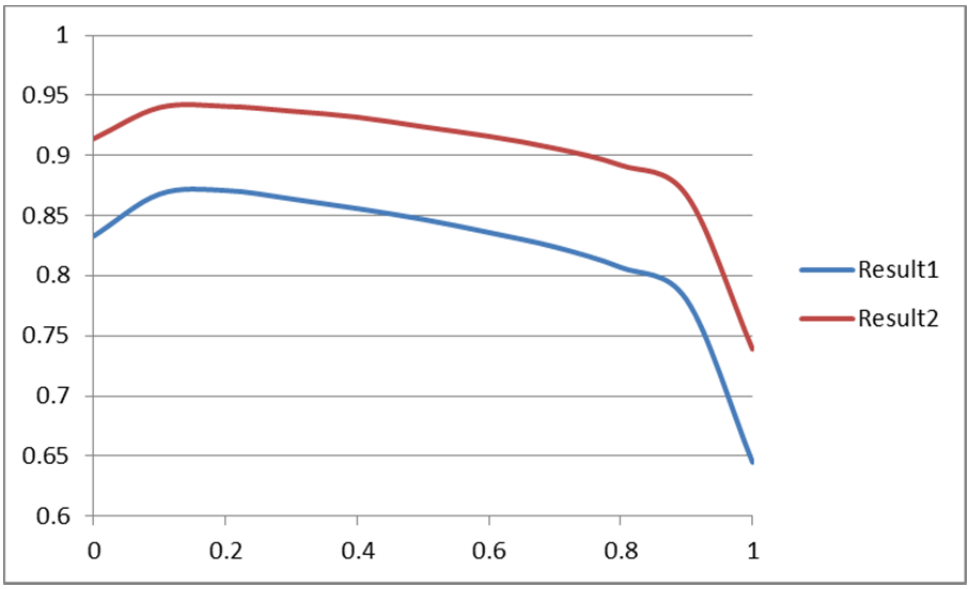


図 11　学習データの認識率のグラフ.

図 11 はパラメータ α の値を 0.0 から 0.1 刻みで変化させたときの認識率のグラフである. Result1 はアルファベットの大文字と小文字を区別する場合，Result2 が大文字と小文字を区別しない場合の学習データに対する認識率の値である．今回の場合，α の値が 0.2 のときに大文字と小文字を区別しない場合の認識率が 94.1%と最大になった．このため，提案手法では α の値は 0.2 を採用した.

5.3. 孤立した単一文字領域の判定

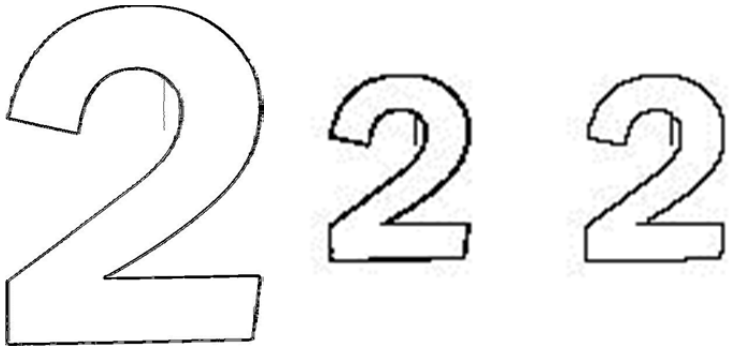
前章で文字列として抽出されなかったエッジ成分の中から，まず孤立した単一文字候補領域と考えられるものを対象に文字／非文字判定を行った.

単一文字候補領域から特徴ベクトルを抽出し，まず 62 種類の文字カテゴリそれぞれとの改良投影距離を算出する．次に，それらの中で最小の改良投影距離値を持つ文字カテゴリを求めて，その最小改良投影距離値が当該文字カテゴリの改良投影距離値の分布に含まれるかどうかを判定する．これには，当該カテゴリに属する学習サンプルの改良投影距離値の分布から予め算出した平均と標準偏差を用いて，次式により，単一文字候補領域が文字／非文字であるかを判定する.

$$g(X) < \mu + \delta_{24} \times \sigma \qquad (15)$$

すなわち，上式を満たす場合に文字であると判定した.

図 12 に，単一文字の文字／非文字判定の処理に用いた画像例を示す.



(a)　　(b)　　(c)

図 12　単一文字候補領域の文字／非文字判定の例.

図 12 で，(a)が元の単一文字候補領域のエッジ成分，(b)が(a)に対して位置と大きさの正規化を施した画像，(c)が(b)に対して Hilditch の細線化処理を施した画像である．例に用いた画像は数字の「2」である．(c)の細線化画像に対して 62 種類の文字カテゴリとの改良投影距離を算出した結果，最小値を持ったカテゴリは「2」であった．これに対して式の条件を満たしたため，この孤立したエッジ成分は単一文字領域であると判定される．

5.4. 接触文字列領域の判定

前章で文字列として抽出されなかったエッジ成分の中から，次に，横長の連結エッジ成分に対して，接触文字列領域であるかどうかの判定を行った．

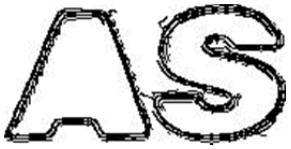
具体的には，横長の連結エッジ成分の w と h において$w > \delta_{25} \times h$を満たす場合に，そのエッジ成分は接触文字列候補として仮分割を行う．

ここで，仮分割とはエッジ成分を文字の平均縦横比を用いて以下の 2 通りの分割数で分割を行うことである．

$$K_1 = \left\lfloor \frac{w}{h \times \gamma} \right\rfloor \qquad K_2 = K_1 + 1 \tag{16}$$

分割して得られたエッジ成分ごとに前節と同様に文字／非文字判定を行う．分割した過半数のエッジ成分で文字であると判定されたものについて文字列領域であると判定する．K_1 個分割したもの，K_2個分割したものの内少なくともどちらか 1 つで文字列領域あると判定された場合に元の 1 つのエッジ成分は文字列領域であると判定する．

図 13 に，接触文字列の仮分割，文字／非文字判定の例を示す．



(a)

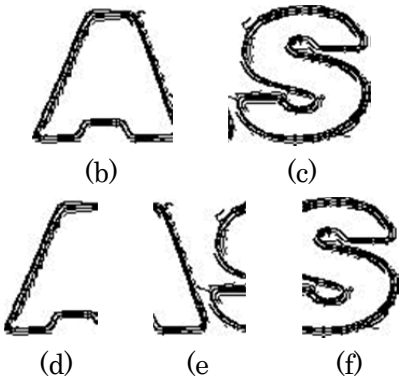


図 13　接触文字列の仮分割の例.

図 13 で, (a)が接触文字列のエッジ成分, (b), (c)が式(16)によって 2 分割された画像, (d), (e), (f)が式(16)によって 3 分割された画像である．これは「AS」という文字列がそれぞれ接触して抽出された例であるが，2 分割したことによって「A」,「S」の領域に分割できていることが分かる．分割した各画像に対して前述した文字／非文字判定を行う．2 分割では(b)は「A」, (c)は「s」との改良投影距離が最も近かった．この内，(b)が式(15)の条件を満たした．半数以上の分割領域で条件を満たしたため 2 分割で文字列と判定された．一方 3 分割では(d)は「L」, (e)は「B」, (f)は「3」との改良投影距離が最も近かった．が，式(15)の条件をいずれも満たさなかったため，3 分割では文字列として判定されなかった．どちらかの分割で文字列と判定された場合に文字列として抽出するため，このエッジ成分は文字列として抽出される．

第 6 章　実験結果

前章までの処理で文字列と判断されたエッジ成分に対して x 座標の最小値，最大値，y 座標の最小値，最大値をそれぞれ求め，原画像の画素値で出力することで文字列領域を抽出する．抽出された領域をデータセットに含まれる xml ファイルの Ground truth と照合し，精度を算出する．

今回の実験の評価は次の 3 つの尺度を用いて行う．

（1）再現率

テスト画像中に含まれる文字列領域の画素数を P, 正しく抽出できた文字列領域の画素数が A であったとき，A／P で表される．

（2）適合率

テスト画像中から文字列領域として抽出された画素数を A’，その中で正しく抽出できた文字列領域の画素数が Q であったとき，Q／A’で表される．

（3）F 値

再現率と適合率の調和平均で表される．

まず，表に予備実験として用いた SceneTrialTrain に対する実験結果を掲げる．

表 1 に，同一のデータセットに対する抽出精度の比較を掲げる．

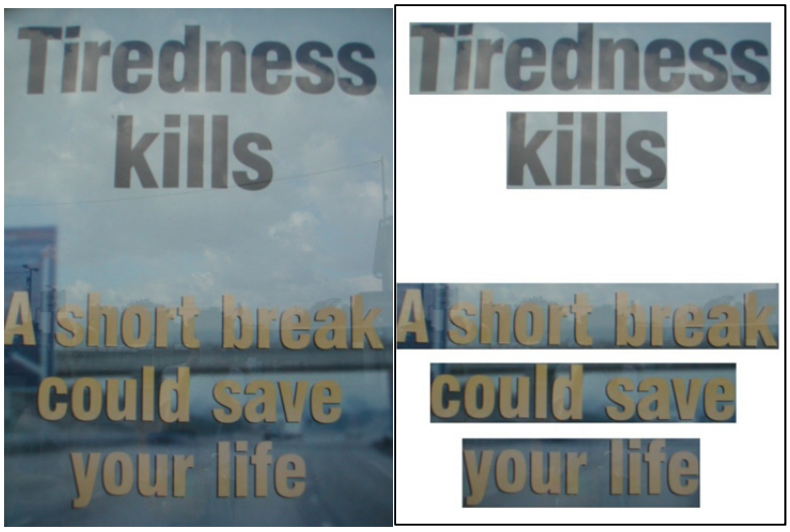
表 1　同一のデータセットに対する抽出精度の比較

手法	適合率	再現率	F 値
芦田ら[9]	55.0%	46.0%	50.1%
J.Kim ら[10]	56.3%	64.3%	60.0%
W.Pan ら[11]	78.7%	73.4%	76.0%
提案手法	64.5%	71.3%	67.8%

なお，文字列抽出に用いたパラメータ δ_1 から δ_{25} および改良投影距離で用いた β は SceneTrialTrain に対して行った予備実験により決定した．図 14 に SceneTrialTest に対して行った実験結果の例を示す．

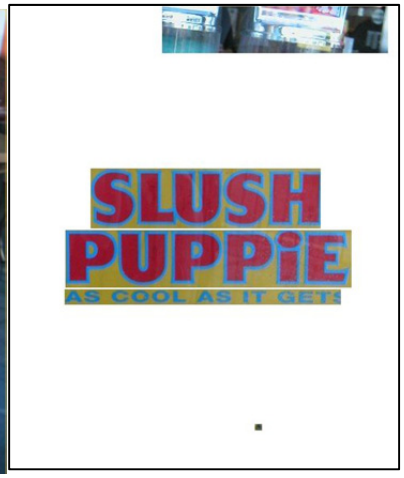
(1-a) (1-b)

(2-a) (2-b)



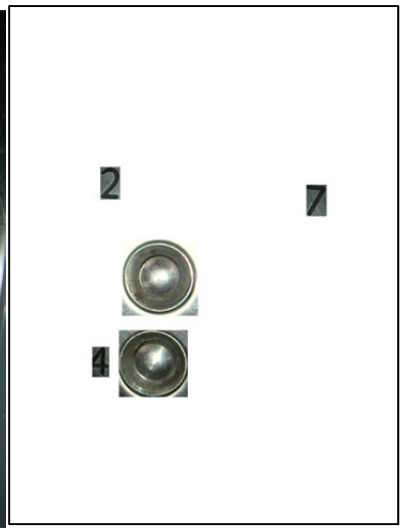
(3-a) (3-b)

(4-a)



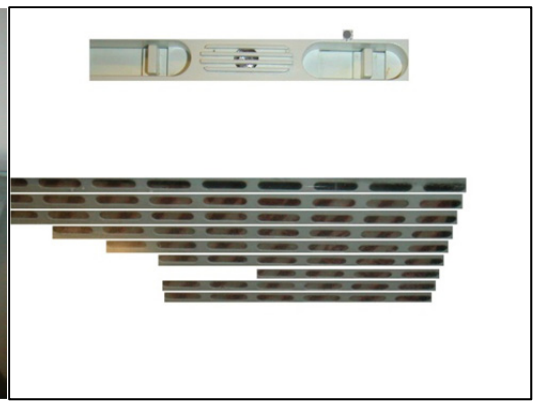
(4-b)

(5-a)

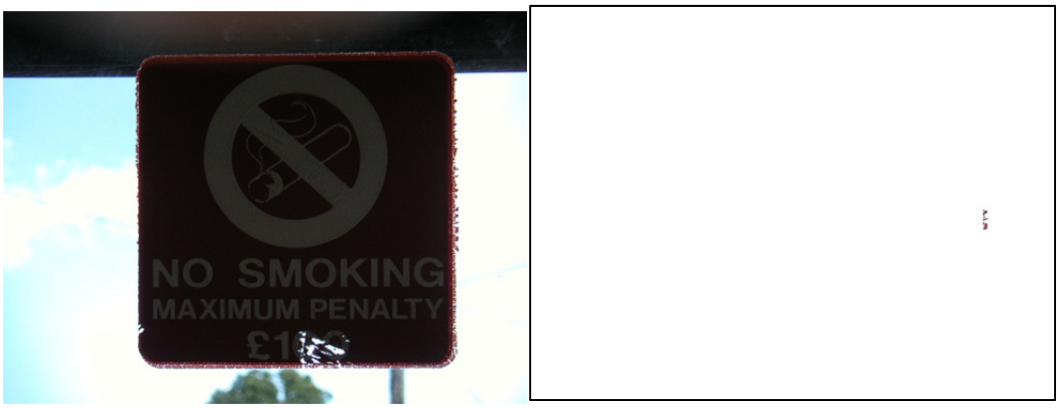


(5-b)

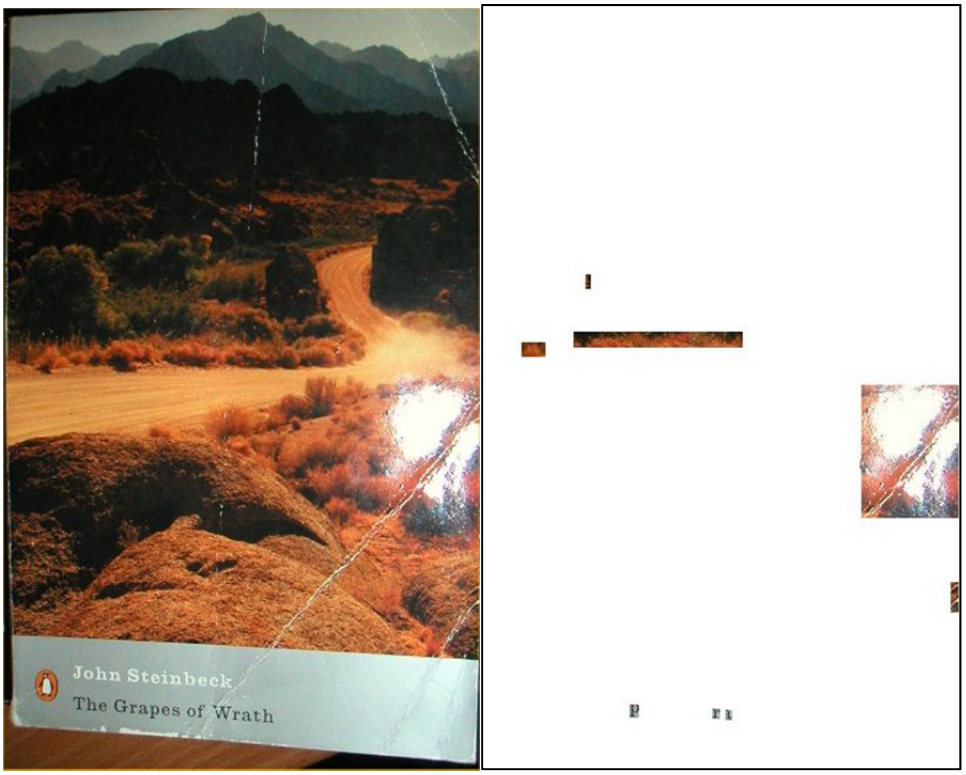
(6-a)



(6-b)



(7-a) (7-b)



(8-a) (8-b)

図 14　実験結果の例.

第 7 章　考察

7.1 考察

本研究で正しく文字列が抽出できなかった例としては,
①エッジ成分が正しく抽出できていないもの,
②文字のエッジ成分が背景領域のエッジ成分と連結しているもの,
③一文字あるいは接触文字の場合に，非文字と判定されてしまうもの,
がある．例えば実験結果の図の(5)は 1 から 9 までの数字が単一文字として，またボタンのようなものがあるが,「2」,「4」,「7」のみ単一文字として抽出することに成功したが，他の数字については単一文字として抽出することに失敗してしまった．また，ボタンのようなものを単一文字として誤って抽出してしまっている．これは「O」,「o」,「0」等の文字と形状が似ているためであると考えられる．(6)は上の方にある「3」,「2」,「１」,「0」がそれぞれ単一文字として抽出されなかった．また，下方にある「BLANCO」という文字列については接触文字列として抽出されず，失敗してしまった．さらに，穴の部分が横方向に似たエッジ成分が並んでしまっているため文字列として誤って抽出されてしまう結果となってしまった．(7)は文字列部分のエッジ成分が全く抽出できなかったために，文字列部分が全く抽出できなかった例である．(8)は文字列部分のエッジが一部しか抽出できず，文字列として存在しているが，単一文字として一部の文字のみ抽出することに成功した例である．

①のエッジ成分が正しく抽出できていないものについては局所的な 2 値化を行うこと，②の文字のエッジ成分が背景領域のエッジ成分と連結してしまう場合についてはカラー情報を利用すること，そして③の文字／非文字の判定についてはより良い特徴量や識別法を適用することで精度を向上させていくことが課題である．

再現率については向上させることができたが，適合率については下がってしまった．これは雑音成分の除去の段階で雑音成分として除去されず，文字列候補領域として円形の雑音成分や縦に長い雑音成分が「0」,「O」，や「i」,「j」,「1」,「l」等と誤って判定されてしまった場合が多かったためである．そのため，文字として頻繁に判定されやすいものについてはより厳しい閾値を設定することや，他の判定方法と組み合わせた手法を新しく提案することが課題である．

また，横方向に長い連結成分を全て文字列候補領域として残しているため，横方向に長いエッジ成分を持つ雑音領域と文字のエッジ成分が接触してしまい，文字列として抽出されず，仮分割処理を加えたためにかえって再現率が下がってしまう例も存在した．明らかに文字ではないと判断できる横方向に長い連結成分は除去する等の改善が必要である．
さらに，本研究では仮分割において全ての文字列候補領域に対して等間隔で分割を施したため，必ずしも個々の文字領域ごとに分割されない．そのため，多くの接触文字列が文字列として抽出されない結果となってしまったと考えられる．よって本研究の制度を向上さ

せるためには，文字全体の縦横の比で等分割する仮分割ではなく，個々の文字ごとに分割できるような処理を組み合わせることが課題となる．

一方，適合率を向上させるためには，より正確に雑音成分のみを除去していくことが必要である．そのためには情景画像中に含まれる背景領域の特徴や雑音成分の特徴をつかみ，雑音成分のみを除去できるような処理が必要である．

最後に，本研究では多くの閾値が存在するが，その決定の仕方は単一文字の縦横比以外の閾値については予備実験を繰り返し行う中で，値を変えながら最も抽出精度が高くなる値に設定した.ヒューリスティックに定めた閾値が少ない処理の方がより一般的であるため，出来る限り閾値の数を減らした処理を行うことも今後の課題である．

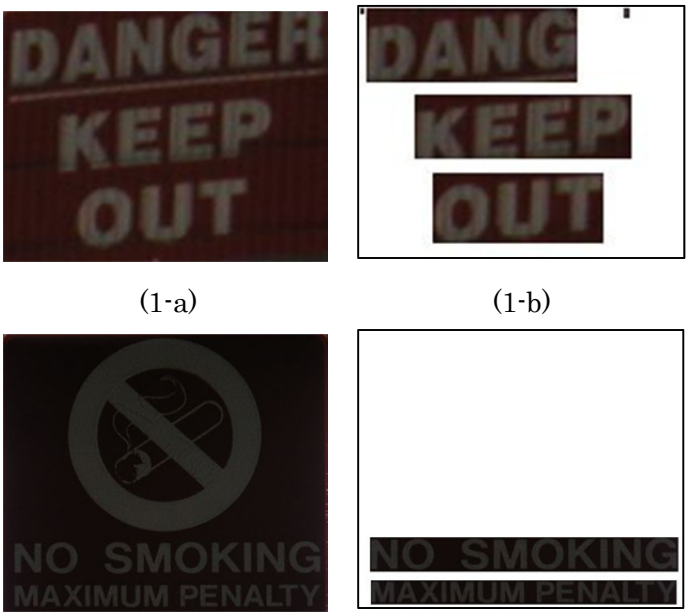
7.2 検討事項

検討事項①　局所的な 2 値化を行った場合

Canny オペレータによる 2 値化を行う際，提案手法では画像全体のエッジ強度の平均と標準偏差の値を用いて Canny オペレータの閾値を求めていた．しかし，図 15 に示すような画像の場合，背景部分に非常に強いエッジ強度を持つ場合があり，文字列部分のエッジ成分が抽出できない場合がある．

図 15　背景のエッジ強度が強い場合の例．

図 15 のような画像の場合，強いエッジ強度を持つ領域が背景領域同士のエッジになってしまい，文字列のエッジ成分が全く抽出できずに文字列抽出に失敗してしまう．左の画像では建物の屋根の部分と背景部分のエッジ強度が非常に強く，文字列のエッジ成分は全く抽出できない結果となった．右の画像では逆光で撮影された画像であると考えられるが，看板部分と背景の空の部分のエッジ強度が非常に強く，文字列のエッジ成分が全く抽出できない結果となった．しかし，どちらの画像でも目視で確認できるように文字列領域とその近傍でのエッジ強度には一定の差があると考えられる．このような画像の場合に実験的に手動で画像を切り出して抽出実験を行った結果を図 16 に示す．



(1-a) (1-b)

(2-a) (2-b)

図 16 手動で切り出した画像に対する抽出結果.

図 16 より文字列を含む一部分の領域でエッジ抽出を行うことによって，文字列のエッジが正しく抽出され文字列が抽出できることが分かる.

そこで，画像全体のエッジ強度の平均，標準偏差を用いるではなく，特定の窓を設けてその窓の中に存在する画素のエッジ強度の平均，標準偏差を用いて 2 つの閾値を設定することとした.

今回用いた窓の大きさは画像の縦の長さ H と横の長さ W を基準にして設定した．縦(H×0.1)横(W×0.1)の大きさの窓とした．窓の中に存在する画素のエッジ強度の平均 μ と標準偏差 σ を用いて

$$Th_u = \mu + \delta_{25} \times \sigma \quad (15)$$

$$Th_l = \mu + \delta_{26} \times \sigma \quad (16)$$

上記の式によりエッジ判定をした.

提案手法と同様に R，G，B それぞれの値のグレースケール画像に局所的 2 値化手法を用いた Canny オペレータによりエッジ成分を得る．得られた画像を提案手法と同様に処理を施し，文字列抽出を行う．抽出結果は以下の表の通りである.

表 2 局所的 2 値化手法を用いた文字列抽出結果.

適合率	再現率	F 値
61.5%	56.1%	58.7%

局所的に 2 値化を行うことにより，多くのエッジ成分を抽出してしまい背景領域のエッジ成分も抽出してしまう場合が多く，文字列のエッジ成分と接触してしまうことで文字列

が抽出できずに再現率が下がってしまう場合や，背景領域のエッジ成分を文字列として抽出してしまう割合も増加し，適合率も下がってしまう結果となってしまった．全ての画像において局所的な 2 値化を行うべきではないことや，適切な窓の大きさの求め方等困難な問題があるが，画像によって局所的な 2 値化を行うべきか，全体的な情報を用いて 2 値化を行うべきか判断することが可能になれば抽出精度は確実に向上すると考えられる．

検討事項②

K-means クラスタリングによる文字列領域の 2 値化

入力画像の全ての画素を HSI カラー空間に投影し，K-means クラスタリングにより K 個のクラスタに分け，それら K クラスタを 2 分割する組合せにより網羅的な 2 値化候補画像を生成する手法を検討する．

HSI カラー空間への投影

RGB カラー空間から HSI カラー空間に変換を行う．ただし，R, G, B ∈ [0,255]と同様に，H, S, I ∈ [0,255]の値域となるようにスケール変換を施した．HSI カラー空間とは，色相(Hue)，彩度(Saturation)，輝度(Intensity)の 3 つの成分からなる色空間である．

$$
\begin{aligned}
&I = \max(R,G,B), \quad m = \min(R,G,B),\\
&if \quad I = 0 \quad or \quad I = m \quad then \quad S = 0, \quad H = \text{不定}\\
&else \quad S = \frac{I-m}{I} \times 255,\\
&r = \frac{I-R}{I-m}, \quad \mathrm{g} = \frac{I-G}{I-m}, \quad b = \frac{I-B}{I-m},\\
&if \quad R = I \quad then \quad h = \frac{\pi}{3}(b-g),\\
&if \quad G = I \quad then \quad h = \frac{\pi}{3}(2+r-b),\\
&if \quad B = I \quad then \quad h = \frac{\pi}{3}(4+g-r),\\
&if \quad h < 0 \quad then \quad h = h + 2\pi,\\
&H = h \times 255. \qquad (17)
\end{aligned}
$$

画像サイズ $W \times H$ のカラー画像の全画素を(17)式に従って HSI カラー空間に投影すると，総数 $W \times H$ 個の画素が HSI カラー空間に散布することになる．

K-means クラスタリング

K-means クラスタリングとは非階層的クラスタリングの手法である．K-means クラスタリングの手法を以下に示す．

Step 1: $W \times H$ 個の点からランダムに K 個の点を選択し，クラスタ中心の初期値$\{\mu_k^{(\tau=0)}\}$

$\{\mu_k^{(\tau)}\}_{k=1}^{K}$ とする．次いで，全てのデータ点をそれぞれ最短距離のクラスタ中心に割り振ってグループ分けをする．各グループが 1 つのクラスタとなる．

Step 2: 各グループに含まれるデータ点の平均を求め，当該クラスタのクラタ中心の更新値とする．$\tau = \tau + 1$ となり，更新されたクラスタ中心を改めて $\{\mu_k^{(\tau)}\}_{k=1}^{K}$ と記す．

Step 3: 全てのデータ点をそれぞれ最短距離のクラスタ中心に割り振ってグループ分けをする．グループ分けに変動がなかった場合は，それら *K* 個のクラスタを出力して終了する．変動があった場合は，*Step 2* に戻る．

K-means クラスタリングにより得られた HSI カラー空間内の各クラスについて，当該クラスタに属する画素群を元のカラー画像に逆投影すると，各々1 枚の分離画像が生成される．これら *K* 枚の分離画像の和集合が元の画像となる．K-means クラスタリングの結果は，*K* 個のクラスタ中心の選び方に依存するため，本研究では K-means クラスタリングにマルチスタート探索を適用した．マルチスタート探索とは，最適化問題の解が一般に初期値に強く依存してしまうため，初期値を複数設定してそれぞれの解を求めておき，それらから最適解を選択する手法である．今回の K-means 法へのマルチスタート探索の適用では，複数解の中からクラスタ内分散の和が最小となるクラスタリング結果を採用した．

K クラスタの 2 分割による 2 値化候補画像の生成

K 枚の分離画像群を網羅的に 2 つのグループに 2 分割して，一方を文字列候補部分(黒)，他方を背景候補部分(白)として，複数の 2 値化候補画像を生成する．

上記 2 分割の全ての組合せにより生成される 2 値化候補画像の総数 N_{binary} は次式の通りである．

$$N_{binary} = \sum_{i=0}^{K-1} {}_K C_i = 2^K - 2 \tag{18}$$

図 17 に，1 枚の入力画像に対する 2 値化候補画像の生成例を示す．ただし，クラスタ数は *K*=5 とした．

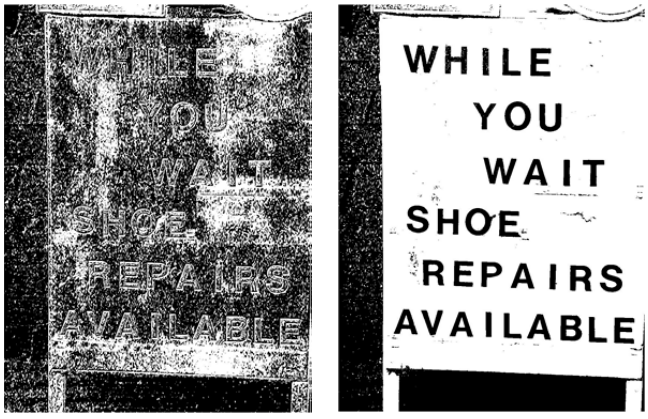
(1-a) (1-b)

(2-a) (2-b)

図 17　クラスタリング結果例.

図 17 でクラスタリングした結果を視覚的に分かりやすくするために，同一のクラスタには同一の画素値を設定し濃淡画像を作成した．図 18 に 2 値化候補画像の例を示す.

(1-a) (1-b)



(2-a) (2-b)

図 18　2 値化候補画像の例.

図 18 で(a)は生成された 2 値化候補画像の内, 文字列部分と背景領域がきちんと分離され

ていない例である．(b)は文字列部分と背景領域がきちんと分離されている例である．K=5 の場合，2 値化候補画像は 30 枚となる．それぞれにおいて提案手法と同様に黒画素連結領域にラベリング処理を施し，雑音成分を除去し，文字列領域を抽出し，単一文字領域や接触文字列領域に対して，文字／非文字判定処理を施す予定であったが，30 枚の画像には正しい画像候補は数枚しか存在せず，多くが無駄な処理になってしまうことが問題になってしまうことや，画像によって最適な K の値が異なるため，動的に K の値を設定することができれば良いが，K の値によっては 2 値化候補画像の数も非常に多くなってしまうことや最適な K の値を選び出すことも困難であるため，断念した．

検討事項③

画像の濃度等高線情報を用いた文字列領域抽出

同一のデータセットを用いた従来研究で非常に高い精度である W.Pan らの研究では画像の濃度等高線情報を用いている．画像における濃度等高線は閉曲線を成し，画素値が 256 階調の画像においては 256 通りの濃度等高線が存在する．等高線情報の例を図 19 に示す．

0	8	7	3	5	6
6	2	5	8	9	5
7	3	6	2	8	5
7	4	3	2	1	2

(a)原画像.

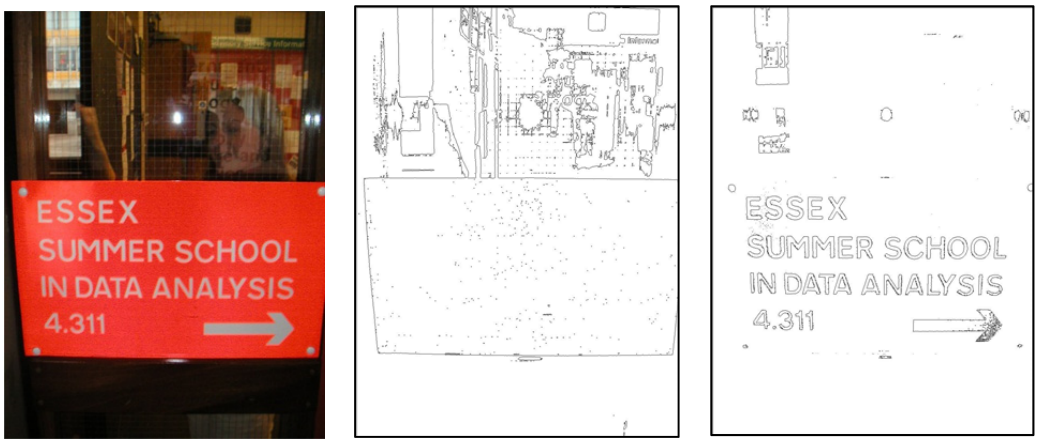
0	8	7	3	5	6
6	2	5	8	9	5
7	3	6	2	8	5
7	4	3	2	1	2

(b)画素値 5 以上の等高線.

図 19

図 19 で(a)の原画像に対して(b)の等高線情報は必ず閉領域を成すことが分かる．

入力画像に対する等高線情報の例が以下の図 20 である．



(a)　　(b)　　(c)

図 20　等高線情報の例.

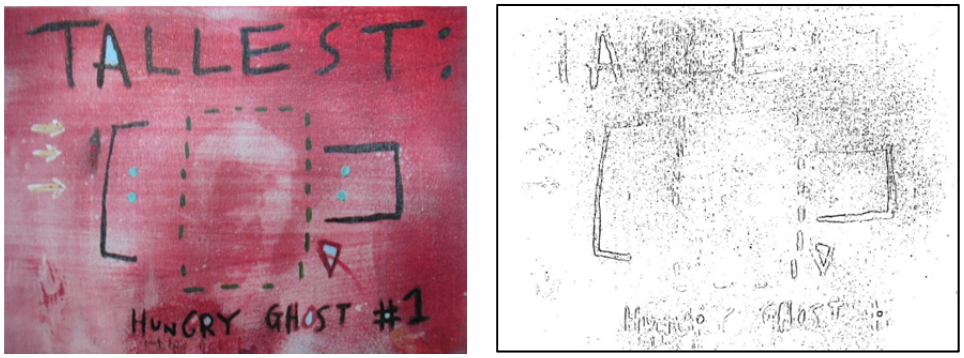
図 20 で(a)は原画像．(b)は画素値 62 の等高線．(c)は画素値 151 の等高線である．等高線情報ではエッジ強度と異なり，画素値の差が小さい場合でも輪郭線を抽出することができ

る．この例では画素値 151 付近で文字列の等高線情報が抽出できることが分かる．

これら 256 通りの濃度等高線画像において黒画素部分を等高線領域としたときに，256 枚の画像を重ね合わせ黒画素の重なり度合いを調べる．文字列部分は背景領域に比べ，多くの等高線が存在していると考え，重なっている黒画素数の平均，標準偏差を求めて以下の式により文字列候補領域 Tを選び出す．

$$T = \mu + \delta_{27} \times \sigma \tag{19}$$

図 21 に等高線情報を用いた文字列の輪郭線候補領域を抽出した結果を示す．



(a)　　(b)

図 21 等高線情報を用いた輪郭線候領域の抽出結果例．

図 21 で(a)は原画像．(b)は(19)式の条件を用いて求めた輪郭線候補領域の画像である．この画像では背景部分の輪郭線も多く抽出されてしまっていることが分かる．

この方法では文字列と背景部分で等高線の高さの差が小さい場合の文字列領域の輪郭線が抽出できなかった．結果的にはエッジ強度に近い用い方をしているからだと考えられる．しかし，文字の特徴として一定の太さの閉領域を成すという点では等高線情報は文字らしい特徴を捉えた情報であると考えられるため，有効な等高線情報を得ることができれば提案手法で，文字と背景部分のエッジ強度が近いために抽出が困難であった文字列の輪郭線も抽出することが可能となり，抽出精度の向上が期待できるため，今後，濃度等高線情報の活用も課題となる．

検討事項④

単一の濃淡画像に対する処理を行った場合

提案手法においてエッジ成分を抽出する際に，RGB 成分それぞれの濃淡画像からエッジ成分を抽出するのではなく，従来のように 1 枚の濃淡画像を作成し，それに対して同様の処理を施した場合について考える．

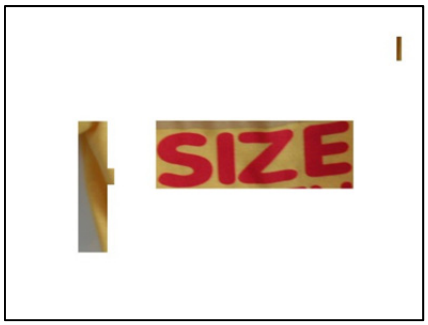
まず，R，G，B それぞれの値の平均を新しい画素値として濃淡画像を作成した場合の結果である．

次に NTSC 係数による加重平均法により濃淡画像を作成した場合の結果である．NTSC 係数法とは Yを新しい画素値とすると以下の式により変換される．

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (20)$$

次に中間値法により濃淡画像を作成した場合の結果である．中間値法とはR，G，Bの最大値と最小値の中間の値を用いて濃淡画像を作成する方法である．図22にそれぞれの方法での文字列抽出結果を示す．

(1-a)



(1-b)

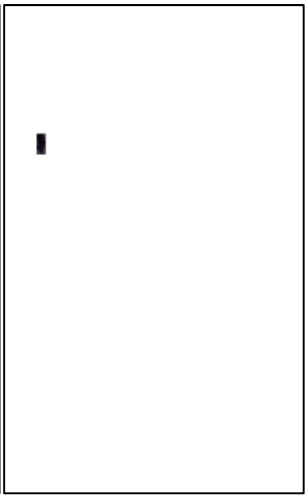
(1-c)

(1-d)

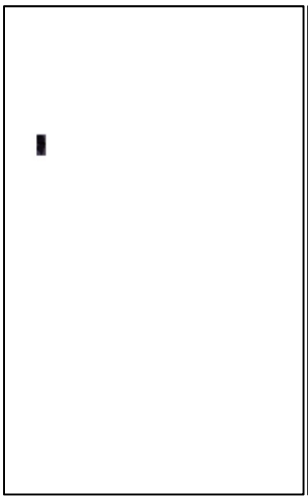
(1-e)

(2-a)

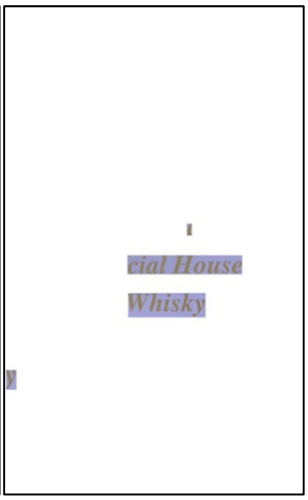
(2-b)



(2-c)

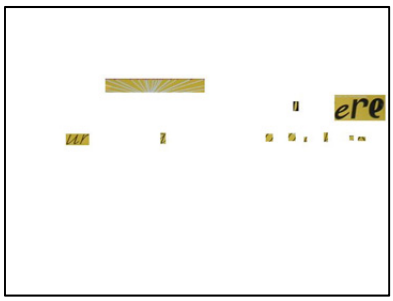


(2-d)



(2-e)

(3-a)



(3-b)

(3-c)

(3-d)

(3-e)

図 22　提案手法の結果と 1 枚の濃淡画像に対して抽出処理を行った結果例.

図 22 で(a)は原画像．(b)は提案手法で抽出処理を行った結果．(c)は RGB 値の平均で濃淡画像を作成した画像に対して抽出処理を行った結果．(d)は NTSC 係数を用いた加重平均法により作成した濃淡画像に対して抽出処理を行った結果. (e)は RGB 値の最大値と最小値の中間の値を用いて作成した濃淡画像に対して抽出処理を行った結果である．(1)は提案手法の再現率が低くなってしまった例である．提案手法においても文字列部分のエッジ成分は抽出できていたが，R，G，B 成分の論理和を取る段階で接触文字列になってしまい，後の処理でも接触文字列として抽出することに失敗してしまったために下の部分の「QUEEN」という文字列が抽出できない結果になった．一方，1 枚の濃淡画像に対して抽出処理を行った場合についてはどの方法を用いても正しく文字列抽出をすることができた．特に(b)，(c)においては再現率，適合率共に 100%に近い値を出すことができた．(2)は提案手法の再現率が最も高い値となった例である．この例では(b)(c)においては文字列部分のエッジ成文は全く抽出できなかった．これは濃淡画像を作成する際にこの画像のように B 成分に大きな値

が出ていてもR，G，Bそれぞれの値の平均または加重平均を用いているため値がならされてエッジが抽出できなくなってしまったためであると考えられる．提案手法では第 3 章でも述べたように B 成分から文字列部分のエッジ成分を抽出できたため，論理和を取ることにより文字列抽出を最も高い割合で成功させることができた．また，RGB 値の最大値と最小値の中間値を用いる方法でも文字列のエッジ成分を一部抽出することに成功した．これは濃淡画像の画素値を，平均ではなく最大値と最小値で求めているため平均するよりも大きな値で取ることを可能とし，背景部分をより白に近い色にすることができたためであると考えられる．(3)は提案手法において再現率が下がってしまった例である．この画像では黒で書かれた文字列に白い線が放射状に繋がっているため，提案手法では白の部分と黒の部分のエッジ成分が連結して抽出されてしまい，うまく文字列を抽出することができなかった．しかし，1 枚の濃淡画像に対して抽出処理を行った場合はどの方法においてもほとんど変わらない再現率，適合率を出すことができた．これは，平均や加重平均，中間値を取る際に放射状の線と文字列部分のエッジ強度が大きくならない値にうまくぼけた濃淡画像を作ることができたためであると考えられる．

また，処理時間については提案手法が約 84 分，単純平均法が約 42 分，NTSC 係数を用いた加重平均法が約 39 分，中間値法が約 43 分となった．1 枚の濃淡画像に対して処理を行うことで処理時間をおよそ半分の時間に短縮することができる．

検討事項⑤

エッジ抽出手法に Roberts オペレータ[12]を用いた場合

Roberts オペレータは図 23 のようなオペレータである．

0	0	0
0	1	0
0	0	-1

(a)x方向のオペレータ

0	0	0
0	0	1
0	-1	0

(b)y方向のオペレータ

図 23

f_xをx方向のエッジ強度，f_yをy方向のエッジ強度としたとき，エッジ強度fは以下の式により求める．

$$f = \sqrt{f_x^2 + f_y^2} \tag{21}$$

Canny オペレータと同様に画像全体におけるエッジ強度の平均μと標準偏差σを用いて，以下の式により 2 値化の閾値を定め処理を行い，エッジ成分を得る．

$$Th_r = \mu + \delta_{28} \times \sigma \tag{22}$$

後の処理は提案手法と全く同様である．表 3 に Roberts オペレータを用いた文字列抽出結果を掲げる．

表 3 Roberts オペレータを用いた文字列抽出結果.

適合率	再現率	F 値
68.7%	58.0%	62.9%

Roberts オペレータで文字列抽出を行った結果, 再現率が大幅に下がる結果となってしまった．これは得られるエッジ成分が太く，接触文字列として多くの文字列が抽出されない結果となってしまったと考えられる.

検討事項⑥

HSI 色空間で Canny オペレータを適用した場合

検討事項②で行った RGB 色空間から HSI 色空間への変換を行い，Canny オペレータを H，S，I それぞれに適用し，提案手法と同様に文字列抽出処理を行った結果を示す．なお HSI 色空間への変換は検討事項②で用いたように $H,\ S,\ I \in [0, 255]$の値域となるようにスケール変換を施した．表 4 に HSI 色空間でエッジ成分を抽出し文字列抽出を行った結果を掲げる.

表 4 HSI 色空間でエッジ成分を抽出し文字列抽出を行った結果.

適合率	再現率	F 値
49.6%	60.2%	54.4%

HSI 色空間で文字列抽出を行った場合, 抽出精度は低下したが RGB 空間では得られなかった文字列のエッジ成分が得られた場合もあるため, うまく RGB 空間と併用することやエッジ抽出の仕方を工夫することで抽出精度を向上させることができる可能性があると考えられる.

第８章　追加実験

提案手法を筆者が撮影した情景画像 66 枚に対して適用し，日本語への対応や様々な状況での抽出状況の様子を調べる．

8.1. 実験に使用した画像データ

追加実験に使用した画像は筆者が東京都小金井市で撮影した情景画像 66 枚であり，一部日本語文字を含む様々な画像が存在する．
以下に例を示す．

図 24　追加実験用画像データ

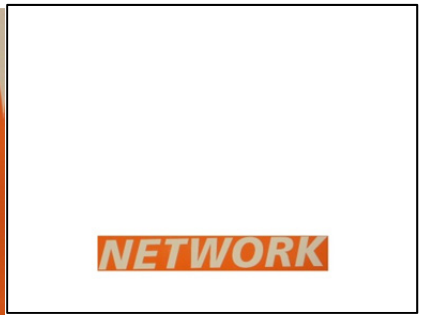
8.2. 実験結果

提案手法を追加実験用画像データに適用した結果を示す．

(1-a)　　(1-b)

(2-a)

(2-b)



(3-a)

(3-b)

(4-a)

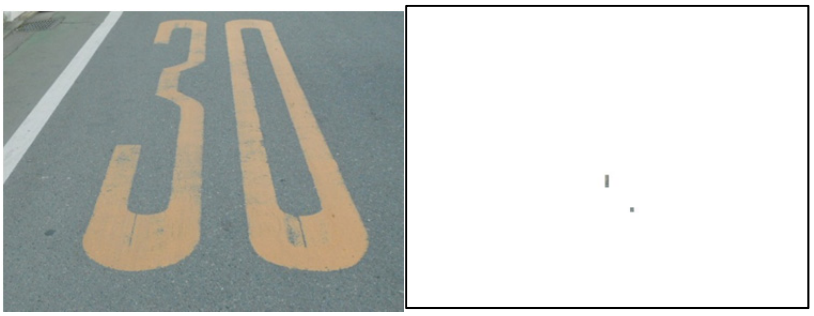
(4-b)

(5-a)

(5-b)

(6-a)

(6-b)



(7-a)　　(7-b)

図 25　文字列抽出の追加実験に対する結果の例(a)原画像．(b)抽出結果．

(1)は道路標識を撮影した 1280×960 の画像である．これは日本語文字列に対してうまく抽出できた例である．この画像に対して提案手法が有効であった理由は，文字と背景部分のエッジ成分が正しく抽出できたためである．また，「止まれ」という文字列がほとんどのアルファベットや数字と同様に 1 つの文字に対して 1 つのエッジ成分を持つためである．(2)コンビニエンスストアを撮影した 1280×960 の画像である．これは文字列が一部うまく抽出できなかった例である．背景部分のエッジ強度が強く，文字列部分のエッジ成分が抽出できなかった部分が存在したためである．(3)は郵便局を撮影した 1024×768 の画像である．これは「JP」という文字列が「J」という文字の背景部分が「P」という文字を構成しており，全体の背景色で「J」という文字が構成されているために，1 つのエッジ成分として抽出されてしまった．また，エッジ成分が文字の形と異なるために仮分割処理を施しても文字列として抽出することができなかった．(4)はある会社の看板を撮影した 640×480 の画像である．この画像に含まれている文字列は，文字列部分が背景色と同じ色であり，ある方向の光源から光が当たったときの影で表現されている．この例では文字列部分のエッジは背景色と全く同じであるため，抽出できなかったが影の部分のエッジ成分が抽出できたため結果的に文字列部分の抽出に成功した例である．(5)はラーメン屋さんを撮影した 640×480 の画像である．これは日本語文字列に対してうまく抽出できなかった例である．この例では文字列のエッジ成分は正しく抽出できたが，「ラーメン」という文字列は横に長い「ー」と 1 つの文字で複数のエッジ成分を持つ「ラ」と「ン」という文字を含むため，正しく抽出できなかった．(6)はある会社の看板を撮影した 640×480 の画像である．これは日本語文字列と英数文字列が，背景色が白の部分は文字の色が黒，背景色が緑の部分は文字の色が白となっている複雑に混在している例である．「住」という文字は「へん」の部分と「つくり」の部分が分離しているが，縦に長いエッジ成分は英数文字の「 l 」，「i」，「1」等の場合と類似しているためであると考えられるが正しく抽出することができた．また，背景部分の緑色の三角形の領域と接触してしまっている文字は抽出することができなかった．(7)は道路に書かれた制限速度表示を撮影した 640×480 の画像である．この例では文字列部分のエッジ成分がしっかりと抽出できたが，途中でエッジ成分が途切れてしまう箇所が見られた．このため「30」という文字列は抽出できなかった．これは道路の表面の凹凸の影響

であると考えられる．

次に，画像を拡大又は縮小した際の抽出率の違いについて考える．以下に画像を拡大又は縮小した際の抽出結果例を示す．

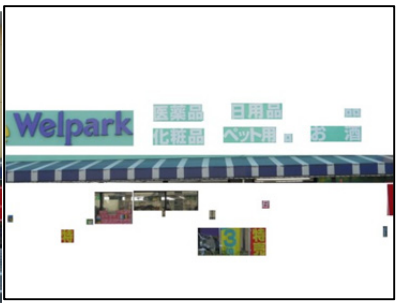
(1-a)

(1-b)

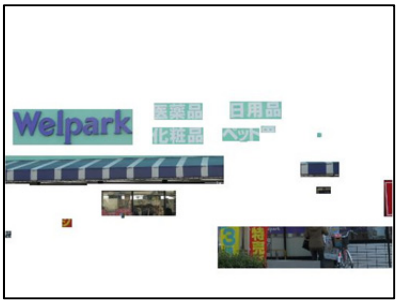
(1-c)

(1-d)

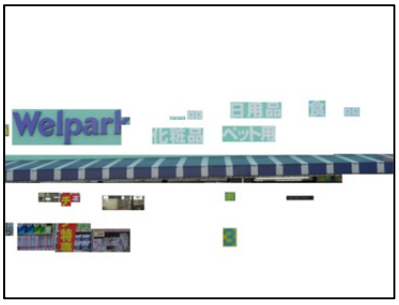
(2-a)



(2-b)

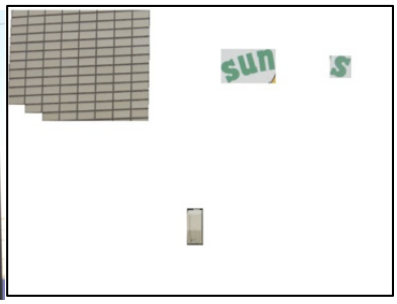


(2-c)

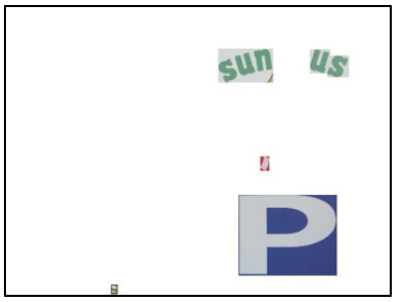


(2-d)

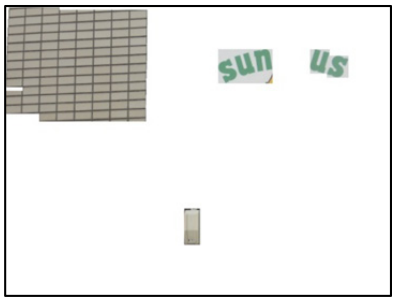
(3-a)



(3-b)



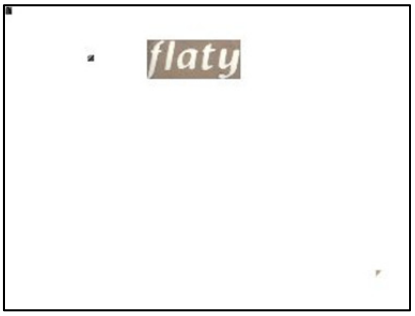
(3-c)



(3-d)

(4-a)

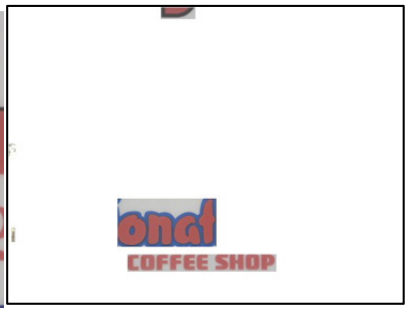
(4-b)



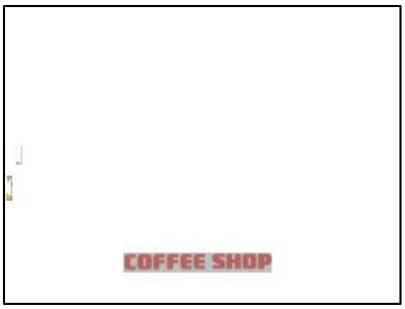
(4-c)

(4-d)

(5-a)



(5-b)



(5-c)

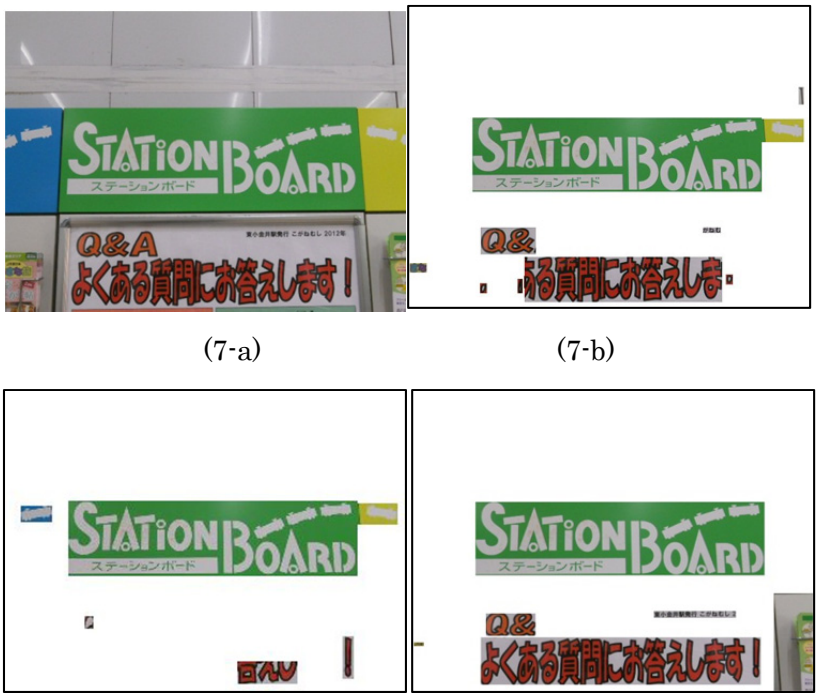
(5-d)

(6-a)

(6-b)

(6-c)

(6-d)



(7-a)　　(7-b)

(7-c)　　(7-d)

図 26　画像を拡大縮小した際の抽出結果．(a)は原画像．(b)は等倍の抽出結果．(c)は縦横それぞれ 1/2 倍した画像の抽出結果．(d)は縦横それぞれ 2 倍した画像の抽出結果．

(1)は自動販売機を撮影した 1024×768 の画像である．この画像を縮小したものはオリジナルの画像から抽出したものに比べ適合率は向上していると考えられる．逆にこの画像を拡大したものはオリジナルの画像から抽出したものに比べ再現率が向上していると考えられる．特にこの画像においては，拡大した場合には缶に書かれている「creamy cafe」というとても小さな 文字列についても抽出することに成功している．(2)はドラッグストアを撮影した 1024×768 の画像である．この画像ではオリジナルの画像が再現率，適合率共に最も高い結果となった．拡大しても再現率が向上しなかった理由として考えられることは漢字の「医薬品」という文字列を拡大したことによって文字を構成しているエッジ成分が分離してしまったためであると考えられる．ただし，拡大した画像においては旗に書かれている「3」という単一文字を正しく文字として判定できていることが分かる．(3)はコンビニエンスストアの看板を撮影した 1024×768 の画像である．この画像では「sunkus」という文字列に含まれる「k」という文字が極めて大きさや色や形の面でかなり特殊な表現で書かれているため，どの場合でも抽出できなかった．この場合では縮小した際に「P」という単一文字を正しく抽出できたために，縮小した画像において適合率，再現率が最も高くなる結果になった．また，この例で見られるようにレンガ模様が提案手法で仮定した文字列らしさを満たすために，誤って文字列として抽出されてしまうことが多い．(4)は建物の看板を撮影した 640×480 の画像である．この画像ではオリジナルの画像が最も適合率，再現率が高くなる結果になった．この画像では縮小した際に平仮名文字列の「みや」のエッジ成

分が抽出できずに最終的な文字列抽出に失敗している．また拡大した際には雑音成分のエッジ成分を多く抽出することになってしまい，適合率を下げる結果となってしまった．(5)はファミリーレストランの看板を撮影した 640×480 の画像である．この画像では拡大した画像において再現率が最も高い結果となった．この画像では「jonathan's」という文字列が青い枠の中に赤い文字で書かれている．オリジナルの画像では一部，縮小した画像ではほとんどが連結した 1 つのエッジ成分としてこの文字列を抽出してしまっているが，拡大した画像では青い枠の部分のエッジと赤い文字列の部分のエッジが分離されて抽出されていた．そのため，赤い文字列部分のエッジ成分を利用して抽出することができたと考えられる．(6)は店を撮影した 640×480 の画像である．この画像ではオリジナルの画像に拡大縮小を施しても再現率や適合率にほとんど影響がなかった例である．この画像はレンガ模様やガラス枠，柱など横方向に似たエッジ成分が多く抽出されてしまう画像であり，文字列として誤って抽出されてしまう背景領域が多い画像例のひとつである．(7)は駅に掲げてあるボードを撮影した 640×480 の画像である．この画像では拡大した際に再現率が最も高くなる結果になった．この画像は日本語文字列の「よくある質問にお答えします！」の部分が黒い枠の中に書かれた赤い文字で書かれている．縮小した際，枠のエッジ成分と文字のエッジ成分が連結してしまったためにうまく抽出することができなかったと考えられる．また，拡大した画像においては「東小金井駅発行　こがねむし　2012 年」という非常に小さい文字列も「東小金井駅発行　こがねむし　2」という部分ではあるが抽出することに成功している．

また処理時間については，オリジナルの画像 66 枚に対して約 31 分，縮小した画像に対して約 9 分，拡大した画像に対しては約 77 分という結果となった．やはり画像の大きさが小さい程，処理をする画素の数が少なくなるため処理速度は速くなることが分かる．縮小した画像に対しても 1 枚の画像を処理するために 8 秒以上かかることになるため，リアルタイム処理は難しいことが分かる．

追加実験によって提案手法が漢字や平仮名，カタカナ等の日本語文字に対しても一部有効であることが示された．また，画像を拡大縮小した際には多くの場合，縮小処理を施すと余分なエッジ成分を抽出しなくなるため適合率が向上し，再現率が低下する．また拡大処理を施すと文字列のエッジ成分を正しく抽出できる割合が多くなり再現率が向上し，余分なエッジ成分も抽出してしまうようになるため適合率が低下する結果となることが予測される．今回 1 つの文字に対して複数のエッジ成分が存在する文字「ラ」,「ン」等や縦書きの文字列に対しては抽出することができなかった．縦書きの文字列に対しては，空間的配置条件を変更することで抽出が可能であると考えられる．また，日本語の文字は chars74K の学習データに存在しないため文字／非文字判定はできない．提案手法を日本語文字にも対応させるためには，文字列が横に並んでいるという条件だけでなく縦に並んでいる可能性も考慮する必要がある．また日本語文字は漢字を含めるとカテゴリ数が非常に多いため文字／非文字の判定には学習データを必要としない特徴を用いることや漢字の一般的な特徴を用いることが必要であると考えられる．

第９章　むすび

本研究では，自律移動ロボットの目や環境内文字読み上げシステム等に応用するためのカラー情景画像からの文字列抽出を目的として，画像の RGB 成分ごとの微分画像に 2 値化処理を施し，雑音成分を除去して融合し，文字列候補領域となるエッジ成分を抽出し，エッジ成分の外接矩形の空間的配置の条件から文字候補領域を選択し，単一文字候補領域として残された領域や接触文字列候補領域に対して，方向分布特徴を用いた文字／非文字判定を用いて文字列を抽出する手法を提案した．特に，空間的配置条件を利用した手法では抽出が困難であった接触文字列のエッジ成分においても抽出を可能とする手段として，仮分割を施して分割されたエッジ成分ごとに文字／非文字判定を行うことで，抽出精度の向上を図った．

提案手法を ICDAR2003 の公開データセットに適用し，再現率 71.3%，適合率 64.5%，F 値 67.8%を達成した．単一文字や接触文字の抽出にも成功した．また，追加実験により一部の日本語文字を含む情景画像に対しても提案手法が有効であることを示すことができた．

今後の課題として，エッジ成分の 2 値化の際に局所的な閾値を決めることで，エッジ成分の抽出精度の向上を図ることや，カラー情報を利用した文字列のエッジ成分と背景や雑音領域のエッジ成分の分離，さらにはより良い，文字としての特徴量や文字／非文字の識別法を求めること，雑音成分や背景領域の特徴を用いた非文字成分の正確な除去が挙げられる．

謝辞

学部生時代から継続して研究してきた内容の論文が完成に至るまで，ご指導いただいた若原教授には大変お世話になりました．また同若原研究室の先輩方や同期や後輩の皆さんにも大変お世話になりました．ありがとうございました．

2013 年 2 月 18 日

参考文献

[1] J. Liang, D. Doermann, and H. Li, “Camera-based analysis of text and documents: a survey". *International Journal on Document Analysis and Recognition*, vol. 7, pp. 84-104, 2005.

[2] The ICDAR 2003 Robust Reading Datasets.
http://algoval.essex.ac.uk/icdar/Datasets.html.

[3] http://www8.plala.or.jp/kusutaku/iview/

[4] J. Canny, “A computational approach to edge detection”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 679-698, 1986.

[5] 大岡信治，栗田昌徳，原田智夫，木村文隆，三宅康二，加重方向指数ヒストグラム法による手書き漢字・ひらがな認識,”信学論(D), vol. J70-D, no. 7, pp. 1390-1397, July 1987.

[6] T. E. de Campos, B. R. Babu, and M. Varma, “Character recognition in natural images". *Proc. of the International Conference on Computer Vision Theory and Applications*, vol. 2, pp. 273-280, Lisbon, Portugal, Feb. 2009.

[7] Hilditch, C. J. : Linear Skeletons from Square Cupboards, Machine Intelligence 4, edited by B.Meltzer et al., University Press, Edinburgh, pp.403−420, 1969.

[8] 金谷健一著，“これなら分かる応用数学教室－最小二乗法からウェーブレットまで - ”, 6.3 節，共立出版，2003.

[9] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, “ICDAR 2003 robust reading competitions". *Proc. of Seventh Int. Conf. on Document Analysis and Recognition*, vol. 2, pp. 682--687, Edinburgh, Aug. 2003.

[10] J. Kim, S. Park, and S. Kim, “Text locating from natural scene images using image intensities". *Proc. of Eighth International Conference on Document Analysis and Recognition*, vol. 2, pp. 655-659, Seoul, Aug. 2005.

[11] W. Pan, T.D. Bui, and C.Y. Suen, “Text detection from natural scene images using topographic maps and sparse representations". *Proc. of International Conference on Image Processing*, pp. 2021-2024, Cairo, Nov. 2009.

[12] R. C. Gonzalez and R. E. Woods, Digital Image Process in, Third Edition, Pearson Prentice Hall, 2008.