法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-05-09

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

(出版者 / Publisher)
法政大学大学院情報科学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編
(巻 / Volume)
8
(開始ページ / Start Page)
161
(終了ページ / End Page)
166
(発行年 / Year)
2013-03
(URL)
https://doi.org/10.15002/00009859

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

北田　英樹
Hideki Kitada
法政大学大学院情報科学研究科情報科学専攻
E-mail: hideki.kitada.4b@stu.hosei.ac.jp

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%.

1. はじめに

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 万人を超えると言われている視覚障害者のための環境内文字読み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字認識の技術を看板や標識などの 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．それらを実現するためには，情景画像から速く正確に文字列を抽出することが不可欠である．従来，多くの手法が提案されてきた[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空間的配置条件に基づき，文字らしい連結成分が横に並んだ領域を文字列として抽出する．ただし，孤立した単一文字や文字間が接触している文字列は上記の判定基準から外れてしまうため，文字列として抽出できない．そこで接触文字列候補領域については横方向の仮分割を施してほぼ 1 文字単位に分割して単一文字候補領域群を生成する．そして，孤立した単一候補文字領域，および分割して得られた単一候補文字領域群に対して，方向分布特徴を用いた文字／非文字の判定を行う．この判定処理で文字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．ICDAR2003 で用いられた公開データセットに提案手法を適用して実験を行い，本研究の有用性を示す．

以下，**2.** で実験用画像データについて説明する．**3.** で文字列候補領域抽出，**4.** でエッジ成分の空間的配置に着目した文字列抽出，**5.** でエッジ成分の方向分布特徴を用いた文字／非文字判定，**6.** で実験結果，を述べ，**7.** で考察を加え，**8.** でむすびとする．

2. 実験用画像データ

実験に用いた画像データは ICDAR2003 の robust Reading and Text Locating dataset[2]の SceneTrialTest に含まれる 249 枚の画像である．これらの画像サイズは 1280×980 から 307×93 まで様々であり，英数文字を含む．図 1 に，画像例を示す．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 SceneTrialTrain に含まれる文字列を含む画像 247 枚の画像である．

図 1　実験用画像データの例．

Supervisor: Prof. Toru Wakahara

3.文字列候補領域抽出

従来技術では，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 の各成分毎にエッジ強度を得て 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[3]を施しエッジ強度を得る．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，以下の式による Canny オペレータの 2 つの閾値Th_u，Th_lを求めて，エッジ強度を 2 値化する．

$$Th_u = \mu + \delta_1 \times \sigma \quad (1)$$

$$Th_l = \mu + \delta_2 \times \sigma \quad (2)$$

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．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．

$$\frac{n}{N} < \delta_3,\ \mathrm{n} < \delta_4 \times (h + w),\ \frac{h}{w} > \delta_5,$$
$$h > \delta_6 \times H,\ w > \delta_7 \times W$$
$$|cy - gy| > \delta_8 \times H,\ |cx - gx| > \delta_9 \times W,$$
$$y - \mathrm{mean} < \delta_{10} \times H \quad (3)$$

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．

$$K_X > L_X,\ K_x < L_x,\ K_Y > L_Y,\ K_y < L_y,$$
$$(K_X - K_x) \le \delta_{11} \times (L_X - L_x),$$
$$(K_Y - K_y) \le \delta_{12} \times (L_Y - L_y) \quad (4)$$

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．

$$K_{in} \ge \delta_{13} \quad (5)$$

図 2 より，この画像に対しては B 成分において文字列部分のエッジがよりはっきりと抽出されていることが分かる．

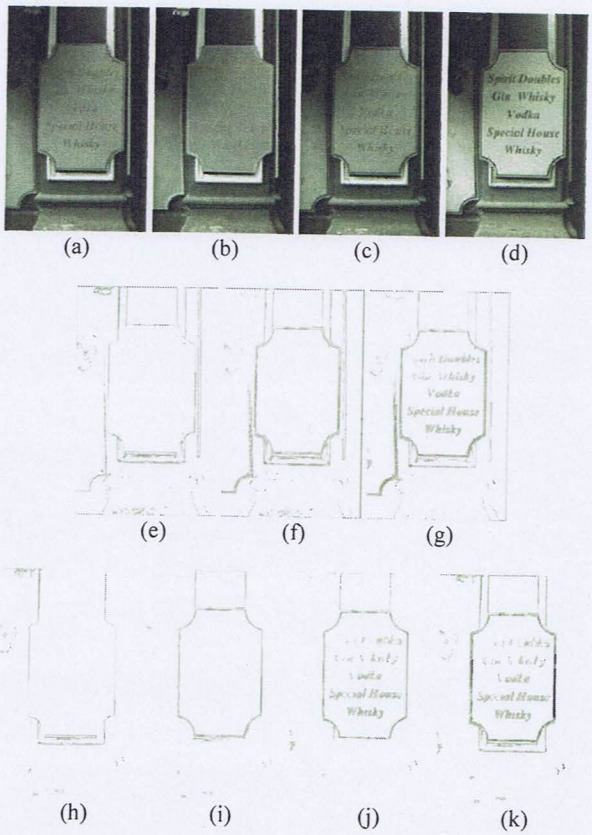


図 2　エッジ抽出処理の画像例．(a)原画像．(b)R 成分．(c)G 成分．(d)B 成分．(e)R 成分のエッジ画像．(f)G 成分のエッジ画像．(g)B 成分のエッジ画像．(h)R 成分で雑音成分を除去した画像．(i)G 成分で雑音成分を除去した画像．(j)B 成分で雑音成分を除去した画像．(k)R，G，B 成分を融合した画像．

4. 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．

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

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

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

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

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

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

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

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

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

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

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)を満たせば文字列と判定する．

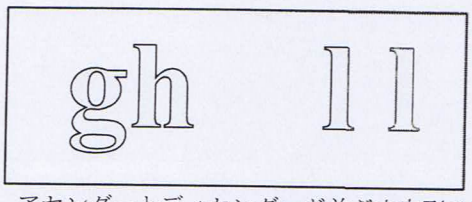


図 3　アセンダーとディセンダーが並ぶ文字列の場合，縦に長い文字列が含まれる場合．

5. エッジ成分の方向分布特徴を用いた文字／非文字判定

上記の処理では，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[4]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 4 に 2 次元ガウスフィルタの重み係数，図 5 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 4　2 次元ガウスフィルタの重み係数．

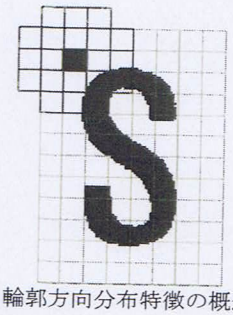


図 5　輪郭方向分布特徴の概念図．

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[7]である．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[5]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．

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$f_i(i=1,2,\ldots,96)$を用いて平均g_0を求める．次に，各特徴ベクトル$f_i(i=1,2,\ldots,96)$から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．

$$\tilde{f_i} = f_i - g_0 \tag{6}$$

$\tilde{f_i}$に対する共分散行列$\tilde{M}$を次式により求める．

$$\tilde{M} = \frac{1}{N}\sum_{i=1}^{N}\tilde{f_i}\tilde{f_i}^t \tag{7}$$

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[6]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．

$$\begin{aligned} g(X) = (N+N_0+1)\ln\{1+\frac{1}{N_0\sigma^2}[\|X-M\|^2\} \\ -\sum_{i=1}^{k}\frac{(1-\alpha)\lambda_i}{(1-\alpha)\lambda_i+\alpha\sigma^2}\left[\Phi_i^T(X-M)\right]^2]\} \\ +\sum_{i=1}^{k}\ln\left((1-\alpha)\lambda_i+\alpha\sigma^2\right)-2\ln P(\omega) \end{aligned} \tag{8}$$

ただし，

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

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ω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 i 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(8)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(8)の擬似ベイズ識別関数を近似することができる．

$$g(X) = \|X-M\|^2 - \sum_{i=1}^{k}\frac{(1-\alpha)\lambda_i}{(1-\alpha)\lambda_i+\alpha\sigma^2}\left\{\Phi_i^T(X-M)\right\}^2 \tag{9}$$

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 k 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(10)で定義される固有値の累積寄与率ηが一定値βを超える最小の次元数を k とした．

$$\eta = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{j=1}^{N}\lambda_j} \geq \beta \tag{10}$$

5.3. 孤立した単一文字領域の判定

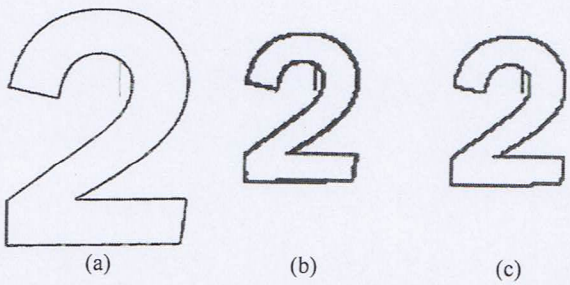
前章で文字列として抽出されなかったエッジ成分の中から，まず孤立した単一文字候補領域と考えられるものを対象に文字／非文字判定を行った．

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．

$$g(X) < \mu + \delta_{24} \times \sigma \tag{11}$$

すなわち，上式を満たす場合に文字であると判定した．

図 6 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 6　単一文字候補領域の文字／非文字判定の例．

図中，(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(11)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$w > \delta_{21} \times h$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γ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．

$$K_1 = \left\lfloor\frac{w}{h\times\gamma}\right\rfloor \qquad K_2 = K_1 + 1 \tag{12}$$

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．分割した半数以上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どちらかが文字列領域で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 7 に，接触文字列の仮分割，文字／非文字判定の例を示す．

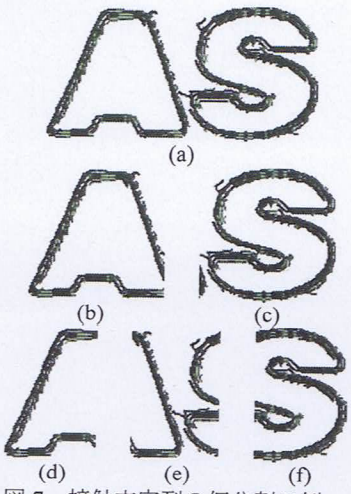


図 7　接触文字列の仮分割の例．

図中，(a)が接触文字列のエッジ成分，(b)，(c)が式(12)によって 2 分割された画像，(d)，(e)，(f)が式(12)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おおよそ「A」，「S」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 分割した例では 1 つ以上，3 分割した例では 2 つ以上の分割領域で文字と判定された場合に元の接触文字列を文字列領域として抽出する．この例では，2 分割では(b)は「A」，(c)は「s」との距離が最も近かった．この内，(b)が式(11)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」との距離が最も近かった．が，式(11)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．

6. 実験結果

今回の実験の評価は，以下に定義する 3 つの尺度：再現率，適合率，F 値を用いて行う．

(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．

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[8]	55.0%	46.0%	50.1%
J.Kim[9]	56.3%	64.3%	60.0%
W.Pan[10]	78.7%	73.4%	76.0%
本手法	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_{24}および改良投影距離で用いたα，βは SceneTrialTrain に対して行った予備実験により決定した．

図 8 に，文字列抽出結果の例を示す．

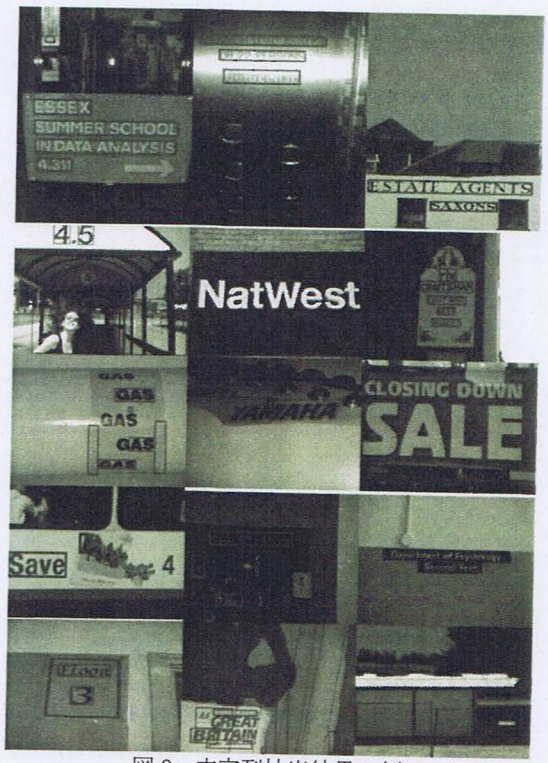


図 8　文字列抽出結果の例．

7. 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．

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領域が「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必要である．

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再現率が下がってしまう例も存在した．明らかに文字列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．よって本研究の精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で，より正確に文字判定を検討す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．恣意的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．

8. むすび

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．

今後の課題として，エッジ成分の 2 値化の際に局所的に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．

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

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.