エッジ成分の空間的配置に着目した情景画像からの文字列抽出

北田　英樹†　　若原　　徹††

† 法政大学大学院情報科学研究科　〒 184–8584 東京都小金井市梶野町 3–7–2
†† 法政大学情報科学部　〒 184–8584 東京都小金井市梶野町 3–7–2
E-mail: †hideki.kitada.4b@stu.hosei.ac.jp, ††wakahara@hosei.ac.jp

あらまし　本論文ではエッジ成分の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，カラー画像を濃淡画像に変換し Roberts オペレータによりエッジ強度の平均 μ と標準偏差 σ を求め，それを基に 2 値化を行う．次に，得られた 2 値エッジ画像から黒画素連結成分をラベリングし，画素数が極めて少ない連結成分や縦横比の大きく異なる連結成分をノイズとして削除する．次に各連結成分の外接最小方形の重心の x 座標や y 座標，面積比等を用いた空間的配置を利用した方法で文字列候補領域を抽出し，同じ文字列であると判断された文字候補領域が 3 つ以上の場合を最終的に文字列であると判断し，抽出結果とする．本手法を ICDAR2003 の robust Reading and Text Locating dataset の SceneTrialTest に含まれる 249 枚の画像に適用した結果，再現率 57.8%，適合率 64.5%，F 尺度 60.9%を達成した．このスコアは ICDAR2003 Competition で優勝を収めた手法を上回る数値である．

キーワード　パターン認識，情景画像，文字列抽出，エッジ成分

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，情景画像からの文字列抽出に用いられてきた手法としては，ある判断基準に基づき決定された複数のしきい値から 2 値画像を作成し，文字領域と背景領域を良好に分割する複数枚の分解画像を得る適応しきい値法 [1]，情景画像において文字線とその背景は局所的に 2 値画像を構成するという性質を用いた局所的 2 値化法に基づく領域分割 [2]，非線形 SVM を用いた文字抽出を高速に行うために，輝度ヒストグラム形状に基づく識別と非線形 SVM を組み合わせた階層型識別器を用いた方法 [3]，カラー画像にエッジに基づく領域分割を適用し，ファジークラスタリングを行い，クラスタごとに 2 値化を行い，SVM により文字パターンを識別するエッジと領域分割に基づく手法 [4]，文字列の直線性，近接性，文字サイズの類似性といった文字列の一般的特徴を利用した方法 [5]，などがある．

しかし, [1] では明度による 2 値化を行うため，照明や天候の状況等により有効な 2 値画像が得られにくい場合がある．また, [2] では文字の存在する物体面の姿勢に起因する文字パターンのひずみに対応すること, [3] では欠落部の復元を含む文字列抽出が, [4] では SVM の学習の際，1 文字のみからなる文字パターンがあまり存在しないため，単独文字の抽出が課題としてあげられている．また, [5] では,「1」や「l」等の文字は抽出が困難といった問題がある．

本研究ではこれらの問題を解決するために，様々な文字パターンに対応したエッジ成分の空間的配置に着目した文字列抽出の手法を提案する．以下に提案手法の流れを示す．

（1） 情景画像を濃淡画像に変換する．
（2） 濃淡画像から Roberts オペレータによりエッジを抽出する．
（3） エッジ画像から雑音成分を除去する．
（4） エッジ成分の空間的配置に着目した方法で文字列らしい部分を抽出する．

本手法では，様々な英数文字を含む文字列に対応するため文字列の縦横比の大きい「1」,「l」等の文字が含まれる文字列の特徴や「gh」,「hy」等の文字列の重心の座標が大きく異なるような文字列の特徴についても考慮している．本手法を用いて ICDAR2003 の robust Reading and Text Locating dataset の SceneTrialTest(注1) に含まれる 249 枚の情景画像に対して実験を行った結果，再現率 57.8%，適合率 64.5%，F 尺度 60.9%を達成した．

以下，2. で実験用画像データについて説明する．3. では前処理，4. ではエッジ成分の空間的配置に着目した文字列抽出，について述べ，5. で実験結果を示す．6. では本研究における考察を述べ，7. でむすびとする．

(注1)：http://algoval.essex.ac.jk/icdar/datasets/

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 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 1280×960 から 307×93 までの様々であり，英数文字を含む．図 1 に，画像例を示す．

図 1　実験用画像データ例.

3. 前　処　理

3.1　カラー画像から濃淡画像へ変換

実験に用いる画像データはカラーの JPEG 形式であるが，本研究ではフリーソフトの IrfanView(注2) を用いて ppm 形式に変換した．さらに NTSC 係数による加重平均法によって濃淡画像へ変換する．

赤の画素値を R，緑の画素値を G，青の画素値を B，新しい画素値を Y とすると次式で求められる．

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \qquad (1)$$

図 2 に，カラー画像から濃淡画像への変換例を示す．

(a)　　(b)

図 2　カラー画像から濃淡画像への変換例.
(a) 原画像．(b) 濃淡画像.

3.2　Roberts オペレータによるエッジ抽出

3.1 で得られた濃淡画像から Roberts オペレータ [8] によりエッジ強度を求め，2 値化画像を得る．

x 方向の差分を Δx，y 方向の差分を Δy とする．また座標 (x, y) における画素値を $f(x, y)$ とすると，Δx，Δy の値は以下の式によって求める．

$$\begin{aligned} \Delta x &= f(x, y) - f(x+1, y+1) \\ \Delta y &= f(x+1, y) - f(x, y+1) \end{aligned} \qquad (2)$$

(注2)：http://www.irfanview.com

座標 (x, y) におけるエッジ強度を $power(x, y)$ とすると，その値は以下の式によって求める．

$$power(x, y) = \sqrt{\Delta x^2 + \Delta y^2} \qquad (3)$$

次に $power$ の平均 μ，標準偏差 σ を求める．画像に含まれる全画素数を $total$ とすると，その値は以下の式によって求める．

$$\mu = \frac{1}{total} \sum_{x,y} power(x, y) \qquad (4)$$

$$\sigma = \sqrt{\frac{1}{total} \sum_{x,y} \left(power(x, y) - \mu \right)^2} \qquad (5)$$

ここで μ，σ，は画像の大半を占める背景部分によって決まっている．よって，この μ，σ で定まる背景部分を白くすればよい．

2 値化のしきい値 Th を次式で定める．

$$Th = \mu \times \alpha + \sigma \qquad (6)$$

本研究では，予備実験より $\alpha = 2.0$ の値を用いた．

図 3 に，図 2(b) の濃淡画像からのエッジ抽出結果を示す．

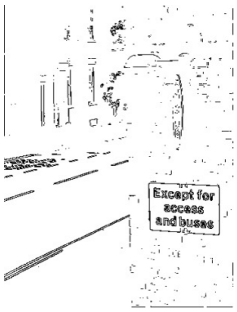


図 3　Roberts オペレータによるエッジ抽出.

3.3　黒画素連結成分のラベリングと雑音除去

3.2 で得られたエッジ画像に含まれる黒画素の連結成分をラベリングする．各ラベルの黒画素数を数え，全画素数の 1/10000 未満のものは雑音とし，白画素にする．

図 4 に，雑音除去した画像例を示す．

図 4　黒画素の少ない連結成分を除去した画像.

また，各連結成分の縦の長さと横の長さを比べ，縦の長さが横の長さの 8.5 倍以上，または横の長さが縦の長さの 3.0 倍以上違う場合に雑音と判断して白画素にする．特に縦に長い連結成分の除去は 4. でも述べるが「ｉ」や「l」などの文字と似ているため誤って抽出してしまうことを防ぐ目的がある．

図 5 に，縦あるいは横に長い連結成分の抽出例を示す．

図 5　縦に長い連結成分 (赤色) と横に長い連結成分 (黄色) の抽出例．

4. エッジ成分の空間的配置に着目した文字列抽出

4.1 文字列らしさ

本研究では下記のような「文字列らしさ」を属性として仮定した．抽出対象の文字列は英数文字であるため，

（1） 各文字は横方向に直線的に並んでいる．
（2） 各文字の外接最小方形の面積が近い．
（3） 各文字と文字の間隔がある程度近い．
（4） 各文字の大きさが大きく異ならない．

まず，ラベリングした黒画素連結成分を各ラベルごとに x 座標，y 座標の最大値，最小値を求めて得られる，エッジ成分の外接最小方形を考える．k 番目の連結成分と j 番目の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．ただし，連結成分の外接最小方形の縦の長さもしくは横の長さが 5 ピクセル未満のものとそれぞれ画像サイズの 0.8 倍，0.4 倍を超える大きさのものは対象としない．また，判定に用いる係数は予備実験より決定した．

以下，アルゴリズムの詳細を述べる．k 番目の連結領域の外接最小方形と j 番目の連結領域の外接最小方形の縦の長さを比べ，大きい方を $height(max)$ とする．それぞれの横の長さを比べ，大きいほうを $width(max)$ とする．それぞれ面積を比べ大きい方の面積を $s(max)$ とする．小さい方の面積を $s(min)$ とする．重心の x 座標をそれぞれ $g_x(k)$，$g_x(j)$ とする．重心の y 座標をそれぞれ $g_y(k)$，$g_y(j)$ とする．また，y 座標の最小値を $y\_min(k)$，$y\_min(j)$ とする．

図 6 に，文字列抽出に用いた連結成分のパラメータを示す．

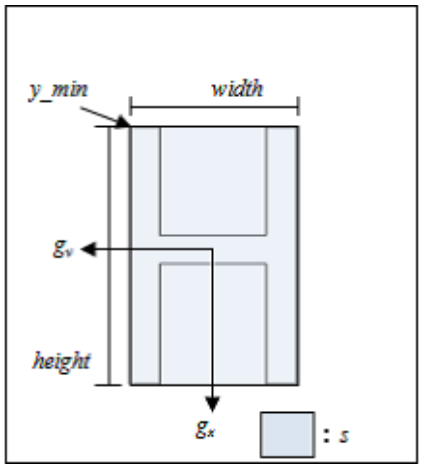


図 6　文字列抽出に用いた連結領域のパラメータ．

4.2 一般的な文字列の場合

① g_x の値を比べ，k 番目の連結成分より j 番目の連結成分が右側にあるものを対象とする．

$$g_x(k) - g_x(j) < 0 \tag{7}$$

② $s(max)$ が $s(min)$ の 3.0 倍未満である．

$$s(max)/s(min) < 3.0 \tag{8}$$

③ g_x の値を比べ，差が $width(max)$ の 1.5 倍以下である．

$$g_x(j) - g_x(k) \leqq 1.5 \times width(max) \tag{9}$$

④ g_y の値を比べ，差が $height(max)$ の 0.28 倍以下である．

$$|g_y(k) - g_y(j)| \leqq 0.28 \times height(max) \tag{10}$$

⑤ $height$ の値を比べ，差が $height(max)$ の 0.5 倍以下である．

$$|height(k) - height(j)| \leqq 0.5 \times height(max) \tag{11}$$

上記①-⑤を全て満たす場合に，k 番目の連結成分と j 番目の連結成分が同じ文字列に含まれる文字候補領域であるとする．

4.3 重心の y 座標間の差が大きくなってしまう文字列の場合

連結成分が図 7 のような場合には④の条件である，重心の y 座標の差が大きくなってしまい，文字列として正しく抽出されないことがある．そこで上記の①-③と以下の⑥-⑧を満たす場合も文字列であると判断する．

⑥ $g_y(k)$ と $y\_min(j)$ の差または $g_y(j)$ と $y\_min(k)$ の差が $height(max)$ の 0.2 倍以下である．

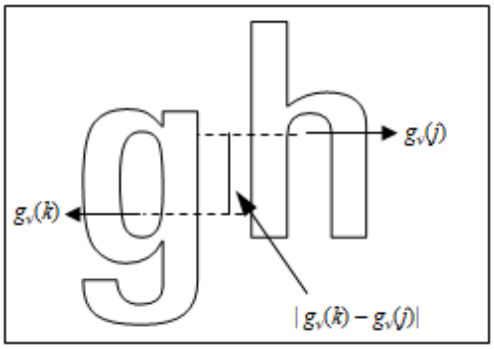


図 7　重心の y 座標間の差が大きくなってしまう例.

$$|g_y(k) - y\_min(j)| \leqq 0.2 \times height(max) \text{ または}$$
$$|g_y(j) - y\_min(k)| \leqq 0.2 \times height(max) \tag{12}$$

⑦ $height$ の値を比べ，差が $height(max)$ の 0.2 倍以下である.

$$|height(k) - height(j)| \leqq 0.2 \times height(max) \tag{13}$$

⑧ g_y の値を比べ，差が $height(max)$ の 0.5 倍以下である.

$$|g_y(k) - g_y(j)| \leqq 0.5 \times height(max) \tag{14}$$

4.4　重心の x 座標間の差が大きくなってしまう場合

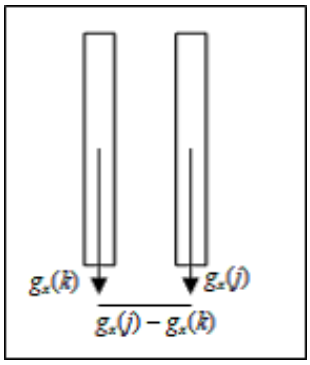


図 8　重心の x 座標間の差が大きくなってしまう例.

図 8 のように数字の「1」や，アルファベットの「i」,「I」,「l」等に相当する連結成分の場合は，$width$ の大きさが小さいため，文字列として抽出されないことがある．そこで，上記の①，④，⑤と以下の⑨ - ⑪を満たす場合も文字列であると判断する.

⑨ k 番目，j 番目の連結成分で少なくともどちらか一方の縦の長さが横の長さの 2.5 倍以上である場合を対象とする.

$$height(k)/width(k) \leqq 2.5 \text{ または}$$
$$height(j)/width(j) \leqq 2.5 \tag{15}$$

⑩ $s(max)$ が $s(min)$ の 8.0 倍未満である.

$$s(max)/s(min) < 8.0 \tag{16}$$

⑪ g_x の値を比べ，差が $width(max)$ の 3.0 倍以下である.

$$g_x(j) - g_x(k) \leqq 3.0 \times width(max) \tag{17}$$

4.5　文字候補連結成分の統合

①-⑤または①-③，⑥-⑧または①，④，⑤，⑨-⑪をそれぞれすべて満たす場合に k 番目の連結成分と j 番目の連結成分は同じ文字列に含まれる文字であると判断する．今，それぞれの文字列候補領域は隣合う 2 文字が同じ文字列であるという判定しか行っていないため，1 つの文字列として抽出するためには各連結成分を統合していく必要がある．ここで，k 番目の連結成分と j 番目の連結成分が同じ文字列であることを $instring[k][j] = 1$ と表現すると，$instring[n][k] = 1$ または $instring[j][n] = 1$ である連結成分 n が存在した場合 n 番目の連結成分も k 番目の連結成分と j 番目の連結成分と同じ文字列だと判断する.

最終的につながった連結成分の数が 3 以上である場合，文字列であると判断する.

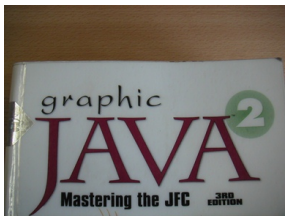
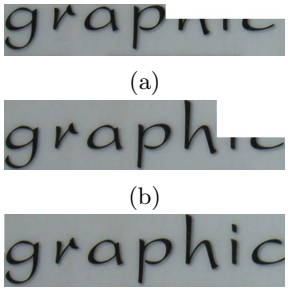


図 9　実験に用いた画像の例.



(a)

(b)

(c)

図 10　図 9 の「graphic」部分からの文字列抽出結果．(a)4.2 のみ使用．(b)4.2，4.3 を使用．(c) 本手法のすべてを適用.

図 9 は，今回の実験に用いた画像の一枚である．この画像に本手法の 4.2 のみを適用して文字列抽出をした結果が図 10(a) である．ただし，ここでは結果を見やすくするため文字列「graphic」の部分のみ表示している．「ph」が同じ文字列と判断されていないことが分かる.

表 1　同一データに対する抽出結果.

手法	適合率	再現率	F 尺度
芦田 [4]	0.55	0.46	0.50
J.Kim [6]	0.56	0.64	0.59
W.Pan [7]	0.73	0.79	0.76
本手法	0.65	0.58	0.61

本手法の 4.2，4.3 を適用して文字列抽出をした結果が図 10(b) である.「ph」の部分が同じ文字列として判断されていることが分かる，しかし「hi」の部分が同じ文字列として判断されていないことが分かる．本手法の全てを適用して文字列抽出をした結果が図 10(c) である.「hi」,「ic」が同じ文字列として判定され,「graphic」という一つの文字列として正しく抽出されていることが分かる.

5. 実 験 結 果

今回の実験の評価は，次の 3 つの尺度を用いて行う.

(i) 再現率

全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 K であったとき，K/P で表される.

(ii) 適合率

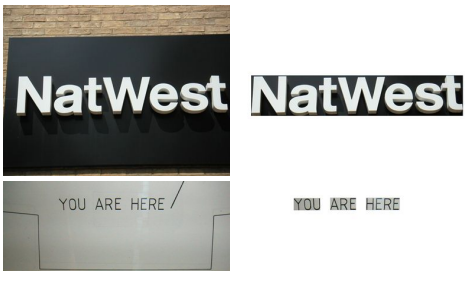
全テスト画像中から文字列領域として抽出された画素数が K'，その中で正しく抽出できた文字列領域の画素数が Q であったとき，Q/K' で表される.

(iii)F 尺度

再現率と適合率の調和平均で表される.

表 1 に，同一のデータを用いた実験結果の比較を掲げる．本手法は高い数値を達成していることが分かる.

以下に，文字列抽出例を示す.



(a)　(b)

図 11　文字列抽出例 1．(a) 原画像．(b) 文字列抽出結果.

図 11 の例は，本手法により入力した情景画像から，それに含まれる全ての文字列部分の外接最小方形のみをそれぞれ抽出することに成功した画像の例である．下段の「YOU」,「ARE」,「HERE」という 3 つの文字列もそれぞれ単語ごとの抽出に成功している．このような例では再現率と，適合率共に 100%に近い値を達成することができた.

図 12 の例は，入力した画像から文字列部分の外接方形を抽出することに成功した画像の例である．このような例の画像では再現率は 100%に近い値を得ることができるが，文字列と文字列の間の空白部分も抽出してしまっているため適合率は若干落ちる．図 12 上段の例では「keep」,「locked」の文字列が 1 つの文字列として抽出されている，下段の例では「MILL」,「ANTIQUES」の文字列が 1 つの文字列として抽出されてしまっている.

(a)　(b)

図 12　文字列抽出例 2．(a) 原画像．(b) 文字列抽出結果.

(a)　(b)

図 13　文字列抽出例 3．(a) 原画像．(b) 文字列抽出結果.

図 13 の例は，入力した情景画像から文字列部分の外接方形を抽出することに一部失敗した画像の例である．このような例の画像では文字を全て抽出できていないため，再現率は落ちてしまう．適合率は余計な背景部分を抽出していない場合は 100%に近い値を得られる．図 13 上段の例では「LET」の文字列が「TO」という文字列と共に 1 つの文字列として抽出されている，さらに下の方の丸い物体が並んでいる部分も文字列と判定され抽出してしまっている．このため適合率も落ちてしまう．下段の例では「FIRE」の部分のみしか抽出できていない.

図 14 の例は，入力した情景画像から文字列部分の外接方形を全く抽出することができなかった画像の例である．このような例の画像では再現率，適合率共に 0%である．図 14 上段の例は単一文字の例である．本手法では単一文字の抽出はできない．下段の例は文字列ではなく窓領域が文字列として誤って抽出されてしまっている.

(a) (b)

図 14　文字列抽出例４．(a) 原画像．(b) 文字列抽出結果．

6. 考　　察

本手法では従来では抽出が困難だった「1」や「l」等の文字を含む文字列も正しく抽出することに成功している．特に正しく文字列を抽出できたものは，図 11 のように背景領域と文字列領域の色がはっきりと異なり，エッジ成分が途切れたり分割されたりしていない場合であった．精度に関しては，未だ改良の余地があるが，エッジ成分の空間的配置を利用した本手法は情景画像から文字列を抽出するための有効な方法の一つであると考えられる．

本手法で正しく文字列を抽出できなかったものの要因について考える．図 14 上段の例では，文字列のエッジ成分ははっきりと抽出できていたが単一文字であるため，抽出することに失敗している．下段の例では，文字列のエッジ強度よりも建物と空の部分のエッジ強度や窓の部分のエッジ強度が強く，正しく文字列を抽出することができず，3 つ以上窓のエッジが並んでいる箇所を文字列として抽出してしまっている．このように，照明や周囲の雑音等の影響により文字列のエッジ成分が正しく抽出できなかった場合に正しい文字列抽出に失敗している．また，エッジ成分が正しく抽出できていても 2 文字や単一文字の抽出には対応していないため，抽出することができなかった．さらに，ガラスに書かれている文字列，手書きの文字列の抽出に失敗してしまった例もある．上記の多くは同じ文字列であるのにエッジ成分が分割されている場合や，一部しか抽出できていない場合があったために正しい文字列の抽出に失敗している．また，文字列が繋がっている場合にはエッジ成分も繋がってしまうため，外接最小方形の横の長さが大きくなってしまい正しい文字列抽出に失敗している．

これらを改善するためには文字列領域のエッジ成分をより上手く抽出することが必要である．そのためにはカラー情報等を利用して，分割されたエッジ成分を補完することや，繋がったエッジ成分を分割することや 2 値化の際の最適なしきい値を動的に求めること等があげられる．また，2 値化で得られた文字列候補領域からより多くの雑音や背景領域を取り除くことで正しく文字列を抽出できる場合が増えるため，前処理の段階でより多くの雑音や背景領域を取り除くことが必要であると考えられる．そのためには背景領域のエッジ成分の特徴やカラー情報等を利用すること，サポートベクターマシン等の識別器による文字列領域と背景領域の識別が有効であると考えられる．

7. む　す　び

本研究では従来から行われている情景画像からの文字列抽出の精度の向上を目的として，エッジ成分の空間的配置に着目した手法を提案した．まず，カラー画像を濃淡画像に変換し，Roberts オペレータを用いてエッジ強度の大きい画素を黒，小さい画素を白としたエッジ画像を得る．次いで，得られたエッジ画像において黒画素の連結成分を考え，黒画素の少ないものや外接最小方形の縦横の比が大きく違うものに対しては文字ではない雑音と考え，抽出対象から除外する．最後に，残った黒画素連結成分で隣り合うものの外接最小方形の重心間の距離や面積比，縦横の長さの比等を用いた条件によって文字列の抽出を行う．

ICDAR2003 の robust Reading and Text Locating dataset の SceneTrialTest に含まれる 249 枚の画像に適用した結果，再現率 57.8%，適合率 64.5%，F 尺度 60.9%を達成した．

今後の課題として，2 値化の際の動的なしきい値の決定やカラー情報を利用した文字列のエッジ抽出やカラー情報や背景領域の特徴を利用した背景領域の削除，サポートベクターマシン等の識別器による背景領域と文字列領域の識別を用いて精度を上げることが挙げられる．

文　　献

[1] 松尾賢一, 上田勝彦, 梅田三千雄, “適応しきい値法を用いた情景画像からの看板文字列領域の抽出,” 信学論 (D-II), vol. J80-D-II, no.6, pp. 1617-1626, June 1997.

[2] 大谷淳, 塩昭夫, “情景画像からの文字パターンの抽出と認識,” 信学論 (D), vol. J71-D, no. 6, pp. 1037-1047, 1988.

[3] 山口拓真, 丸山稔, “階層型識別器を用いた情景画像からの文字抽出手法,” 信学論 (D-II), vol. J88-D-II, no.6, pp. 1047-1055, 2005.

[4] 芦田和毅, 永井弘樹, 岡本正行, 宮尾秀俊, 山本博章, “情景画像からの文字抽出,” 信学論 (D-II), vol. J88-D-II, no. 9, pp. 1817-1824, 2005.

[5] 劉詠梅, 山村毅, 大西昇, 杉江昇, “シーン内の文字列領域の抽出について,” 信学論 (D-II), vol. J81-D-II, no. 4, pp. 641-650, 1998

[6] J. Kim, S. Park, and S. Kim. “Text locating from natural scene images using image intensities.” Proc. of eightth International Conference on Document Analysis and Recognition, vol. 2, pp. 655-659, 2005.

[7] 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, 2009.

[8] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Third Edition, Pearson Prentice Hall, 2008.