

看板認識システムの構築

Construction of a signboard recognition system

梶田 健史†
Takeshi Kabata

渡部広一†
Hirokazu Watabe

河岡 司†
Tsukasa Kawaoka

1. はじめに

本研究の目的は、人の手を借りずに自律的に行動できる「お使いロボット」の実現である。そして、一人でお使いするには、街中にある看板から情報を得ることが不可欠であると考えられる。このような理由から本稿では、看板認識システムについて報告している。そして、看板に書かれている文字列は、複数行で書かれている場合が多い。本研究では特に、この複数行文字列を一行ずつに分離するという新たな機能について報告する。評価は、作成した看板の中に書かれている文字列をOCRでどれだけ認識できたかで行う。

2. 看板認識システム

2.1 看板認識システムの流れ

以下に看板認識システムの流れを示す。

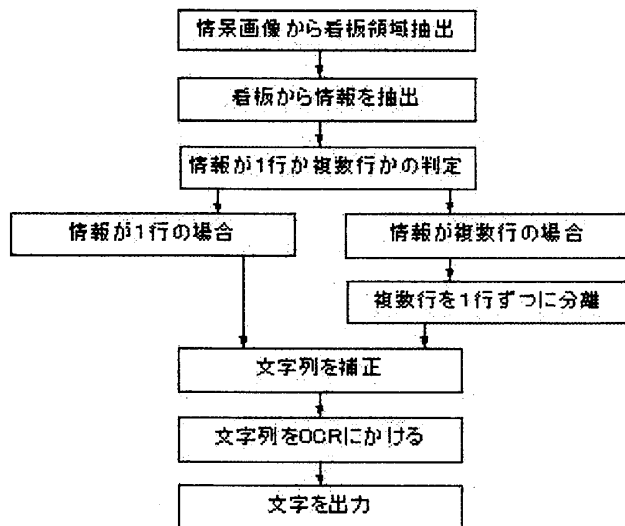


図1 看板認識システムの流れ

デジタルカメラで撮影された情景画像に対して、看板領域の抽出を行い、その看板から文字列を抽出する。文字列が複数行なら、1行ずつに分離する。そして、抽出された文字列を文字補正システムにかけて、OCRで認識できるようにする。最後に、認識された文字をテキスト形式で出力する。以上が、本研究で構築する看板認識システムである。

2.2 情景画像からの看板切り出し

情景画像からの看板抽出は、文献[1]の成果である「情景画像からの看板の切り出しと看板抽出プログラム」を利用する。

看板を含んだ情景画像(図2)において、看板は安定背景領域(色の変化が少ない領域)の一つと考えられる。その安定背景領域を情景画像から推定し、抽出する。ただし、安定している領域が必ずしも看板領域とは限らない(図4)。看板が含まれている画像において、以下の条件を満たしているものを看板領域とする。

1. 看板が比較的に画像の中心に位置すること。
2. 誘目性を保つためにある程度の大きさがあること。

この条件を考慮すると、図3が看板領域であることがわかる。



図2 情景画像

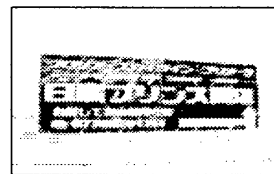


図3 安定背景領域1

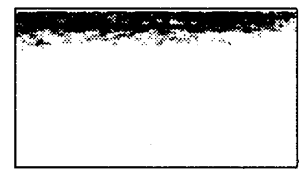


図4 安定背景領域2

2.3 看板領域からの情報抽出

看板領域からの情報抽出は、文献[1]の成果である「情景画像からの看板の切り出しと看板抽出プログラム」を利用する。

切り出された看板領域の中で、画素の色情報を調べ、看板領域の色情報とその看板に書かれている情報領域の色情報との対比関係を利用して、看板領域の中の情報(図5)を抽出する。

†同志社大学大学院 工学研究科
Graduate School of Engineering, Doshisha University



図5 抽出された情報領域の一部

看板領域において、看板領域の代表色範囲に入っていない画素からなる領域を、情報の存在する可能性が高いと考え、情報領域候補と呼ぶ。代表色範囲とは、色情報のヒストグラムにおいて、領域を構成している主な明度、彩度、色相の範囲と定義する。例えば、図6のようなヒストグラムがあったときに、度数の最大値の1/2までの度数をもつ色情報の範囲を代表色範囲とする。

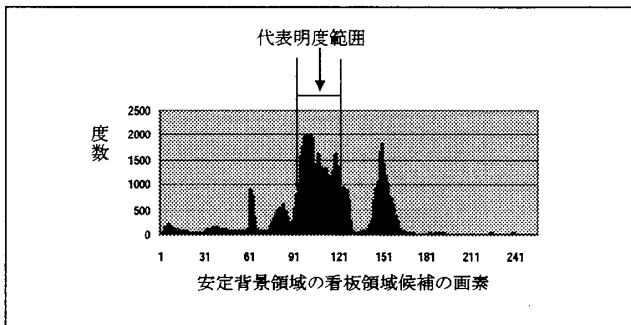


図6 看板領域の代表色範囲

情報領域候補の画素の明度のヒストグラムを調べる。ここで、各画素の明度と看板領域の代表明度の対比関係を反映するために、度数の替わりに、式(1)を要素として、対比ヒストグラムを作成する。

$$\text{度数} \times \text{ABS}(L-a) \quad \dots \text{式(1)}$$

ABS()は絶対値を求める関数を表す
L:明度
a:看板領域の代表色範囲の中間値

度数(図7)だけではなく、看板の色情報との対比関係も考慮する対比ヒストグラム(図8)を用いると、より正確な結果が得ると考えられる。

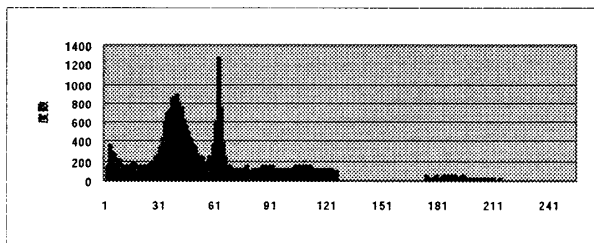


図7 明度ヒストグラム

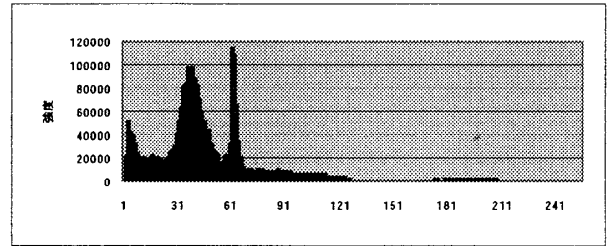


図8 明度対比ヒストグラム

色情報(明度、色相、彩度)のヒストグラムを用いて、強度が最大値の1/2以下になる所までの範囲に入っている画素の集合を連続区域とする。この連続区域が、情報領域となる。また、その連続区域は記録しておく。

そして、記録された連続区域に入っている画素の総数を数える。数えた画素の総数が情報領域候補の全画素数の1/2より少ない場合、その連続区域以外の範囲において、対比ヒストグラムを用いて、また同じ処理を行う。この処理は連続区域に入る総画素数が情報領域候補の総画素数の1/2より多くなるまで繰り返し、一つまたは複数の情報領域代表明度範囲を確定する。

確定された各代表明度範囲において、彩度、色相の対比ヒストグラムを作成し、明度と同じ処理を行う。結果は一つの情報領域の代表明度範囲について、一つまたは複数の代表彩度範囲、代表色相範囲を確定できる。

2.4 文字補正

文字列の補正は、文献[2]の成果である「歪み補正システム」を利用する。

歪み補正を行う過程の中では、その文字列を構成している文字は未知である。そこで、文字列全体が有している形状特徴と文字間余白の特徴を測定することで、その文字列が有している歪みを検出し、補正を行う(図9)。

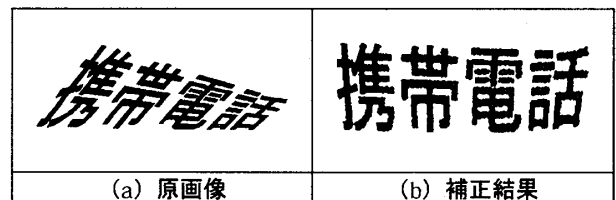


図9 歪みのある文字列の補正

2.5 複数行文字列の分離

文字補正システムは、一行の文字列、または、各行の文字列の長さが同じで両端がそろっている複数行の文字列しか適切に補正できない。それ以外の複数行文字列を補正しようとする和不適切に補正される。また、OCRには複数行認識する機能はあるが、歪みがある複数行の文字列は精度良く認識できない。よって、より正確にOCRで認識するためには、複数行文字列を1行ずつに分離して、それぞれを補正してOCRで認識する必要がある。

2.5.1 文字列の概形

図10のような複数行文字列のおおまかな概形を求めることによって、走査する行間白画素の範囲を求める。

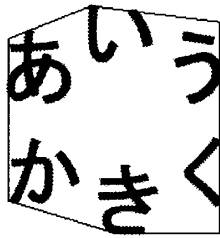


図 10 文字列の概形

2.5.2 複数行文字列の行間白画素走査と分離

概形の左端の直線の各画素から、傾きを変えて白画素を走査する。走査する範囲は、文字列の概形の中とする。左端から右端までの走査で、白画素以外の画素を通らなかった直線を構成している画素だけを記録しておく。ここで、傾きの範囲は、上限左側の直線の傾きの値を a 、上限右側の直線の傾きの値を b 、下限左側の直線の傾きの値を c 、下限右側の直線の傾きの値を d とする。傾きの走査範囲を $\min < \tan(\theta) < \max$ として \max , \min を以下のように定義する。

- ・ $a > d$ なら $\max = a$, $a < d$ なら $\max = d$
- ・ $b > c$ なら $\min = c$, $b < c$ なら $\min = b$

行間白画素を黒で塗りつぶした画像が図 11 の a である。そして、この行間白画素の主軸を検出したものが図 11 の b である。その主軸を基準に分離した画像が図 11 の c と d になる。

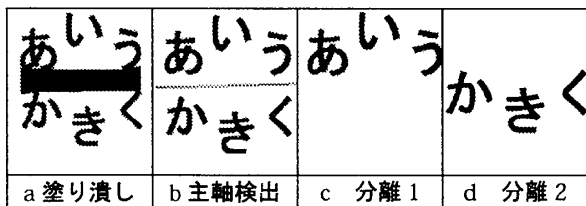


図 11 主軸の検出と分離画像

3. 実験と考察

A4 の白い紙に MS ゴシックで文字列を斜めに書いた看板を、周りの背景が複雑になるように窓ガラスに貼り付けて、それを 9 方向の視点からデジタルカメラで撮影したものを本研究で構築した看板認識システムでどのくらい認識できるか検証した。9 方向は、図 12 のように定義する。作成した看板は、11 枚で 99 個の画像データにより評価を行う。なお、看板領域が抽出できたもので評価を行う。

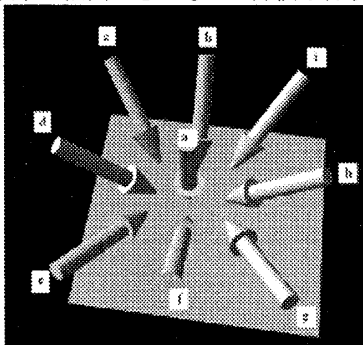


図 12 9 方向の視点

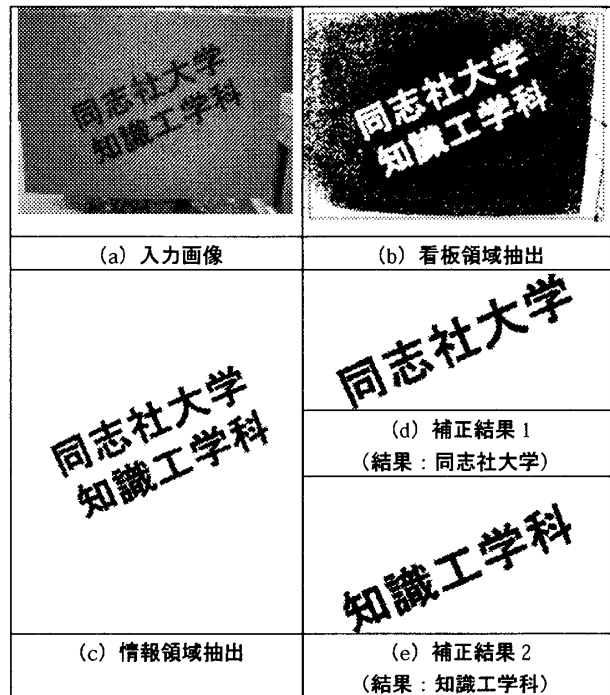


図 13 実験の例 (正面からの撮影)

情報領域の抽出は、全画像 99 枚中、75 枚が情報領域の抽出に成功し、約 76% の精度で抽出された (図 14)。

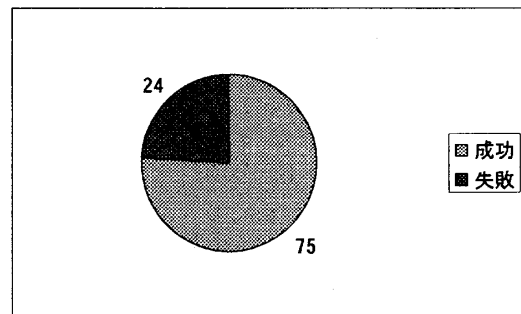


図 14 情報領域の抽出制度

複数行分離の評価は、情報領域が抽出された 75 個の画像データを用いて行う。この 75 個のうち、正確に 1 行ずつ文字列を分離できたものを正解とする。81% の精度で複数行分離成功していることがわかる (図 15)。

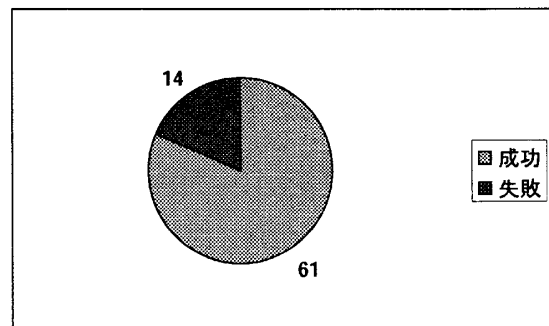


図 15 複数行文字列の分離の精度

複数行の分離が成功した 61 枚の画像データで、文字補正の評価を行う。この評価は、OCR で正しく文字補正されたかどうかで行う (図 16)。認識率は約 82%である。

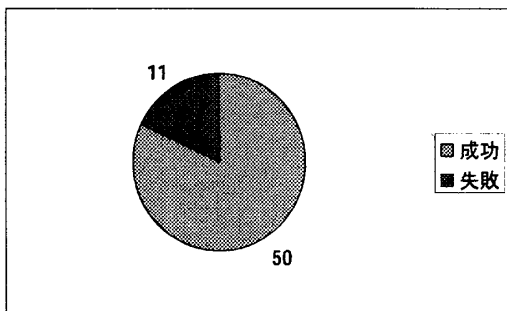


図 16 文字補正の精度

最後に、各視点別の評価を表 1 に、システム全体の評価を図 17 に示す。

表 1 視点別の OCR における認識精度

視点	a	b	c	d	e	f	g	h	i
認識率	72%	64%	37%	55%	27%	82%	36%	45%	36%

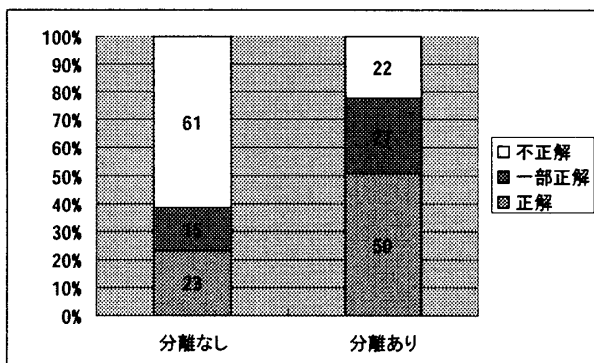


図 17 OCR による認識結果

一部正解のものも含めると複数行文字列の分離を行わないときの精度が約 38%、行ったときの精度が 78%となり、約 40%の精度が上がっていることになる。よって、複数行分離は看板に書かれている文字列を認識するのに効果的であるといえる。しかし、表 1 をみるとわかるように、斜めから撮影した画像は、認識結果の精度があまり良くなかった。本研究で構築した看板認識システムにおいて利用した補正システムは、補正を 2 次元的に行っている。斜めから撮影したものは、奥行きがでて 3 次元的になるので、補正が失敗したと考えられる。

4. おわりに

本研究では、デジタルカメラで撮影した情景画像から、看板領域を抽出し、その看板領域から看板の中に書かれている文字列を抽出して、その文字列を OCR で文字認識し、

文字列をテキストで出力する看板認識システムの構築を実現した。

本研究では、複数行文字列の分離という新たな機能について報告した。この機能をシステムに組み込みことによって、認識精度を 23%から 50%まで上げることに成功した。

謝辞

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクトにおける研究の一環として行ったものである。

参考文献

- [1] 周景龍, 渡部 広一, 河岡 司, 「情景画像からの看板領域と看板中の情報抽出—安定背景領域の推定—」, 信学技報, PRMU2002-217 pp.37-42, 2003.
- [2] 萩原 敏浩, 渡部 広一, 河岡 司, 「ひずみを有する活字文字列における形状補正手法」, 電子情報通信学会論文誌 D-II, J86-D-II, No.2, pp.262-271, 2003.