

領域の複雑さと連続性による 3V-3 景観画像中文字の抽出法

塩 昭 夫 (NTTヒューマンインタフェース研究所)

1. まえがき

景観画像中から文字を抽出する方法について述べる。人間は、身の周りの文字をいとも簡単に検出し認識することができる。しかし、これを工学的に実現するのは容易でない。

従来の検討では、文字の書かれたプレート(例えば車のナンバープレートなど)の形、色、大きさなどに関する知識を有効に使ってプレートを検出した後、プレート内の文字を切り出す方法が提案されている⁽¹⁾⁽²⁾。しかし、これらの方法は、プレートに関する知識を前提としている為、これらの知識が使えない対象には適用できない。

本報告では、文字プレート検出を行わずに直接文字を検出する方法として、ラベル付けされた領域の複雑さと、これらが文字列として出現することに着目した文字検出法について述べる。

2. アルゴリズム

本手法では、認識対象に依存しない方法を検討する立場から、

- (a)文字領域のコントラスト、
- (b)文字領域の複雑さ、
- (c)文字列の形成、

に関する知識を利用する。

本手法では、図1の手順に従い以下の処理を行う。

(1) 入力ボタンをコントラストに基づく動的2値化法⁽³⁾により2値化する(上記(a)の知識を利用)。

(2) 得られた2値ボタンから1×1画素の孤立点を除去したのち、黒点・白点のそれぞれに対してラベリング処理を行い連続した黒・白領域のそれぞれにラベルをつける。

(3) 得られた各ラベル領域(単に領域と呼ぶ)のうち大きさが極端に大きいもの(128×128画素以上)を除去する(雑音除去処理)。

(4) 白黒境界長に関する複雑度:

$$C_1 = [\text{外接矩形内の白黒境界の全長}] / C_0$$

を各領域*l*に関して求め、

$$C_1 > C_{th}$$

なるラベル(*l*)の領域を削除する(上記(b)の知識を利用)。

ただし、 C_0 は C_1 の最大値(市松模様の場合に相当)、 C_{th} は背景ノイズと文字領域を分離するためのしきい値を示す。この処理により動的に2値化されたボタンに現れる背景雑音が除去できる。

(5) 残った領域に関し、各領域が文字列を構成している可能性があるかどうかを判定する。まず、各領域間の上下左右の位置関係を求める。例えば、領域Aが領域Bの左側にあるための条件は、領域A、Bの外接矩形の左上と右下の頂点の座標をそれぞれ(X_{a1}, Y_{a1})(X_{a2}, Y_{a2})、(X_{b1}, Y_{b1})(X_{b2}, Y_{b2})とすると、上下の重なりに関して、

$$\text{Max}(Y_{a2} - Y_{b1}, Y_{b2} - Y_{a1}) /$$

$$\text{Min}(Y_{a2} - Y_{a1}, Y_{b2} - Y_{b1}) > V_{th}$$

文字列方向に関して、

$$X_{a1} < X_{b1} \text{ かつ}$$

$$X_{b1} - X_{a2} < H_{th} * \text{Min}(X_{a2} - X_{a1}, X_{b2} - X_{b1})$$

である。この様に得られた位置関係から縦または横方向にNヶ以上の領域が連なる時、これを文字列(

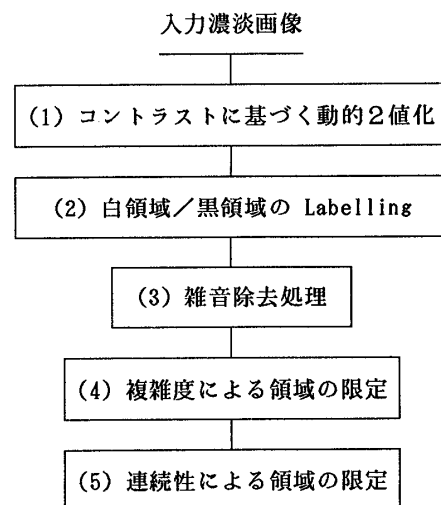


図1 文字検出処理手順

以下の実験では、横書きのみを対象としたと判定し、文字列に関係しない領域を削除する（上記知識(c)を利用）。

ただし、 V_{th} , H_{th} , N はともにパラメータであり、実験では $V_{th}=0.8$, $H_{th}=-4.0$, $N=5$ とした。

3. 実験

(i) 使用データ : テレビカメラ入力 of 15 枚の画像（いずれも車のナンバープレートを含む: 512×512 画素 $\times 8$ bits）を使用。

(ii) 実験と考察 : 図2は、上記15枚に対する処理の進行と平均の雑音領域数の関係を示しており、処理の進行につれ雑音領域数が、単調に減少している。同図では、複雑度に関する閾値 $C_{th} = 0.2, 0.3, 0.4$ の場合を示すが、このうち $C_{th}=0.2$ および 0.3 の場合には、文字領域の若干の消失が起きた。 $C_{th}=0.4$ の場合の処理(3)を行った時点での平均雑音領域数(白領域・黒領域の合計)は 222、処理(4), (5)の後では、それぞれ 57, 17であった。なお、ラベリング処理直後の雑音領域数は、528 である。

図3は、処理結果の一例であり、(a)は原画、(b)はその2値化結果、(c)は複雑度による領域の限定処理結果、(d)は更に領域の連続性（文字列の形成）によ

って限定した結果である。(d)では、若干の雑音を残しているがほぼ文字領域のみに限定されている。

なお、図3では黒領域を示しているが、白領域に対しても同様の処理を行っている。

4. むすび

以上の結果は、文字領域のコントラスト、複雑さおよび文字列を形成すると言う基本的な性質のみで、文字領域の抽出がほぼ可能であることを示しており、本手法と文字認識処理との併用⁽⁴⁾により景観中文字の抽出・認識が可能である。今後は、より多様な画像に対する評価を行う予定である。

〔謝辞〕本研究の機会を与えて戴いた視覚情報研究部の小森部長、酒井リーダに感謝致します。また、日頃ご討論戴く同研究部の皆様に感謝致します。

〔参考文献〕

- (1)安居院ほか, 信学論J70-D, no.3, pp.560-566 (March 1987)
- (2)M.Takato et al., Proc.of int.workshop on industrial appl.of Machine vision and Machine intellegence, pp.76-79 (Feb. 1987)
- (3)塩, 信学論J71-D, No.5, pp.863-873(May 1988)
- (4)大谷, 宮原, 赤松, 信学総全大S.62-1502 (1987)

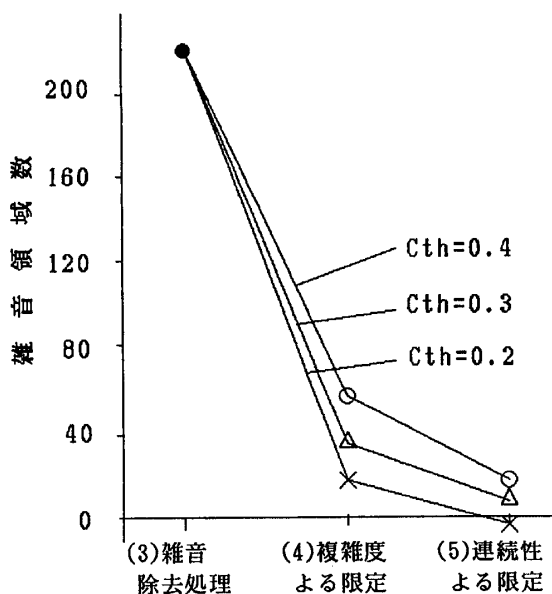


図2 文字検出評価結果

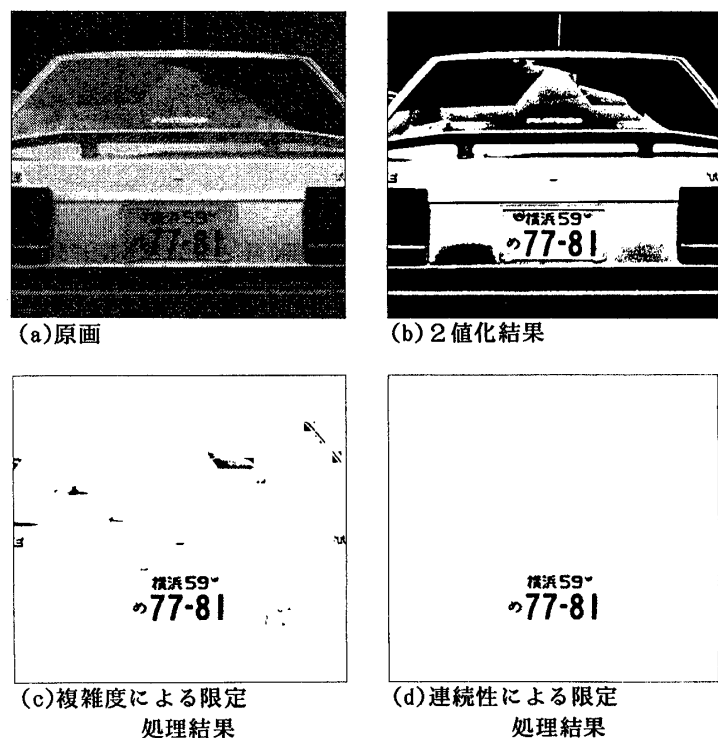


図3 文字抽出結果(例)