

コスト最小化による2次元画像からの文字領域抽出

4L-2

行天 啓二 馬場口 登 北橋 忠宏

大阪大学産業科学研究所

1 まえがき

近年の文書画像処理技術の発展に伴い、その一要素技術である文字領域抽出においても、現在まで様々な手法が提案されている [1]。しかし、従来の文字領域抽出手法の大半は対象画像の書式に関する知識を利用しているため、ポスターやチラシ等、書式のない文書画像に適用することはできない。

書式に依存しない文字領域抽出を実現するためには、あらゆる文字において満たされる、文字の一般的特徴を積極的に利用しなければならない。この観点に基づき、筆者らは一般的文字特徴の一つである文字配置特徴を利用した、コスト最小化による文字領域抽出手法を提案している [2]。

本稿では、一般的文字特徴として新たに文字形状特徴を導入し、複数の連結領域により構成される文字に対しても適用可能である改良手法を提案する。

2 一般的文字特徴

本手法では、印刷文字の一般的文字特徴として、表1に示す形状特徴と配置特徴を考慮した文字領域抽出を試みる。

表 1: 一般的文字特徴

文字形状特徴	
正方形性	文字の外接矩形の形状は正方形に近い
線分の密集	文字は複数の短線分の密集により構成される
文字配置特徴	
近接性	同一文字列内の隣接文字は比較的近接している
サイズ的一致	同一文字列内の隣接文字はほぼ同一サイズである
直線的配置	同一文字列内の3文字はほぼ直線的に並んでいる
等間隔配置	同一文字列内の3文字はほぼ等間隔な2つの空白部を形成する

3 コスト最小化による文字領域抽出

3.1 解空間

コスト最小化手法を用いる場合、まずコスト関数が定義される解空間を定義する必要がある。本手法では、図1のように、画像中の連結領域と文字領域の候補となる領域(以後文字候補領域と呼ぶ)に関するモデル中のパラメータを解空間として利用する。

本モデルは、画像中の連結領域を表すノード $\{C_j\}$ と文字候補領域を表すノード $\{V_i\}$ 、及びノード間のリンク g_{ij}, r_{ik} により構成される。但し、 g_{ij}, r_{ik} は $V_i - C_j$ 間、 $V_i - V_k$ 間の強度である。 g_{ij} は、 C_j に相当する連結領域が V_i に相当する文字候補領域の一部である確度を表し、 r_{ik} は、 V_i と V_k に相当する文字候補領域が互いに文字列をなす確度を表す。また、 p_i は V_i に相当する文字候補領域が文字領域である確度を表す。パラメータ g_{ij}, r_{ij}, p_i により構成される空間を G, R, P とする。コスト関数は、これらの空間上で定義される。

各ノードは、領域情報として外接矩形の座標位置、線分の密集度等に関する情報を有する。但し、 C_j の領域情報は、画像中の連結領域より得られる。 V_i の領域情報は、 V_i とリンクを有する C_j の領域情報の統合により得られ、 $\{g_{ij}\}$ の値に応じて変動する。

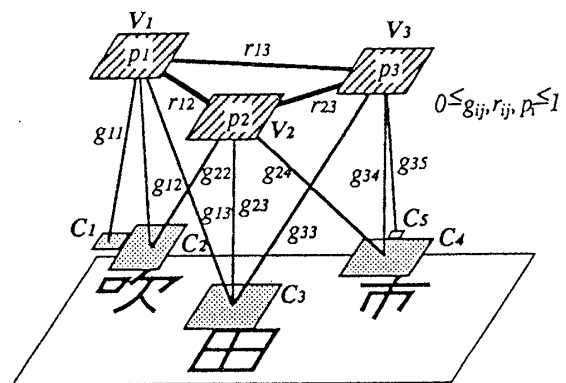


図 1: 連結領域-文字候補領域モデル

3.2 一般的文字特徴の特徴量

2で挙げた文字特徴をコスト関数に反映させるため、各特徴の特徴量を導入する。紙面の都合上、詳細な定義は省略するが、これらの特徴量は $\{V_i\}$ の領域情報を基に算出される。但し、 $\{V_i\}$ の領域情報自体が $\{g_{ij}\}$ によ

A Cost Minimization Approach to Extracting Character Regions from Two Dimensional Image,
Keiji GYOHTEN, Noboru BABAGUCHI, Tadahiro KITA-HASHI,
The Institute of Scientific and Industrial Research, Osaka University

り変動するため、これらの特徴量は、 G 上の関数と捉えることができる。各特徴量を以後、正方形度 $q_i(G)$ 、線分密集度 $w_i(G)$ 、近接度 $c_{ij}(G)$ 、サイズ一致度 $m_{ij}(G)$ 、直線度 $l_{ijk}(G)$ 、等間隔度 $e_{ijk}(G)$ と表記する。

3.3 コスト関数

上述の解空間、特徴量を基にコスト関数を定義する。

最終的に得られる解においては、 $q_i(G)$ と $w_i(G)$ 、 $c_{ij}(G)$ と $m_{ij}(G)$ 、 $l_{ijk}(G)$ と $e_{ijk}(G)$ の値が十分大きい場合、それぞれ $p_i = 1$ 、 $r_{ij} = 1$ 、 $r_{ij}r_{jkl} = 1$ が成り立つ。そこで、これらの条件が満たされている場合に最小値をとる関数を、目的関数として用いる。 p_i 、 r_{ij} 、 $r_{ij}r_{jkl}$ に着目した目的関数 F_1 、 F_2 、 F_3 を以下に示す。

$$F_1(P, G) = \frac{1}{N} \sum_i^N D(p_i, T_q(q_i(G))T_w(w_i(G)))$$

$$F_2(R, G) = \frac{1}{N(N-1)} \sum_i^N \sum_j^N D(r_{ij}, T_c(c_{ij}(G))T_m(m_{ij}(G)))$$

$$F_3(R, G) = \frac{1}{N(N-1)(N-2)} \sum_i^N \sum_j^N \sum_{\substack{k \\ k \neq i \\ k \neq j}}^N D(r_{ij}r_{jkl}, T_l(l_{ijk}(G))T_e(e_{ijk}(G)))$$

但し、 N は V_i の数、 $T_x(x \in \{q, w, c, m, l, e\})$ ($0 \leq T_x \leq 1$) は各特徴量のしきい関数であり、各特徴が満たされている場合に 1 となる。 D は $D(a, b) = \{2(1-ab) + a - b\} / 3$ ($0 \leq a, b \leq 1$) であり、 $a = b = 1$ の時、最小値 0 をとる。また任意の a, b において $\partial D / \partial b < 0$ となる。すなわち、目的関数の最小化の過程において、 F_1, F_2, F_3 における D 中の T_x の積の増大、つまり各文字特徴の充足により、関数の値が減少する。従って、 $\{V_i\}$ は常に文字特徴が満たされるような形で形成されてゆく。

本問題において、最低限満たされなければならない条件として、以下のものを挙げることができる。

- $0 \leq g_{ij}, r_{ij}, p_i \leq 1$
- 文字列中では文字は 3 つ以上の文字と隣接しないので、 V_i において $0 \leq \sum_j^N r_{ij} \leq 2$
- 連結領域は 2 つ以上の文字領域に属さないので、 C_j において $0 \leq \sum_i^N g_{ij} \leq 1$
- 文字領域は L 個以内の連結領域により構成されると考えた場合、 $0 \leq \sum_j^M g_{ij} \leq L$
- $r_{ij} = 1$ の場合、 V_i, V_j が文字列をなしていると考えられるので、 $p_i = p_j = 1$

但し、 M は C_j の数である。以上の条件が満たされている場合に最小値をとる関数を、拘束条件として用いる。

コスト関数は、目的関数、拘束条件の線形和とする。本コスト関数の最小解が、本問題における解となる。本手法では、シミュレーテッド・アニーリング法によりコスト関数の最小解を導く。但し、図 1 のモデルの初期状態

において、 $\{V_i\}$ は、 $\{C_j\}$ と全く同一の領域情報を有するものを用意した。各パラメータの初期値は、 g_{ij} については $i = j$ ならば $g_{ij} = 1$ 、 $i \neq j$ ならば $g_{ij} = 0$ とした。また、 r_{ij} は $T_c(c_{ij}(G))T_m(m_{ij}(G))$ 、 p_i は $T_q(q_i(G))T_w(w_i(G))$ とした。また、 V_i, V_j ($i \neq j$) が互いに類似した領域情報を有している場合、一方を除去するようにした。

3.4 実験的検討

特定の書式を有さない、図 2(a) に示す文書画像に、本手法を適用した。その結果を図 2(b) に示す。ここで、各矩形は $p_i \geq 0.5$ の V_i を表す。つまり、この矩形が文字領域に相当する。矩形内の連結領域は、各矩形に相当する V_i と強度 $g_{ij} \geq 0.5$ でつながる構成要素であることを表す。また、矩形間を結ぶ線は、各矩形に相当する V_i, V_j が強度 $r_{ij} \geq 0.5$ でつながる文字列であることを示す。

図 2(b) より、一つまたは複数の連結領域により構成される文字候補領域が、一般的文字特徴を満たすように適切に形成されていることが分かる。また、一般的文字特徴のみを用いているため、異なるサイズの混在、文字列の傾きに全く影響されないことも併せて分かる。

大阪大学

大阪大学

産業科学研究所

産業科学研究所

北橋研究室

北橋研究室

(a)対象画像

(b)文字領域抽出結果

図 2: 文字領域抽出例

4 おわりに

文書画像の書式に依存しない文字領域抽出手法を提案した。本手法は、コスト最小化の概念に基づいており、最も適切に一般的文字特徴を満たす形で文字領域を抽出できることを特徴とする。

今後は、各種文書画像を対象とした文字領域抽出実験を行なった上で、目的関数中のしきい関数と文字領域抽出結果との相関性の解析、処理の並列化等によるコスト最小化の高速化について検討する予定である。

参考文献

- [1] 美濃：「文書画像処理の現状と動向」，信学会誌，76，5，pp.502-509(1993)。
- [2] 行天，馬場口：「制約充足型文字領域抽出の基礎検討」，信学技報，PRU92-119(1993)。