

オートマトン制御付き最適セグメンテーション法とその実環境文字認識への応用

内田 誠^{†a)} 酒井 恵[†] 岩村 雅一^{††} 大町真一郎^{†††}
黄瀬 浩一^{††}

FSA-Guided Optimal Segmentation and Its Application to Camera-Based Character Recognition

Seiichi UCHIDA^{†a)}, Megumi SAKAI[†], Masakazu IWAMURA^{††},
Shinichiro OMACHI^{†††}, and Koichi KISE^{††}

あらまし 本論文では、動的計画法 (DP) と有限状態オートマトン (FSA) の組合せに基づいた、一次元信号の最適セグメンテーション手法を提案する。具体的には、信号の性質 (例えば信号の値が高い区間と低い区間が交互に繰り返すと言った性質) を FSA 表現した上で制約条件としてセグメンテーション問題に組み込み、その制約下での大局的最適セグメンテーションを DP により効率的に求める。FSA の導入により、信号の性質と一致しないセグメンテーション結果は排除され、精度の向上が見込める。更に、FSA 状態と各区間の対応結果によって各区間の意味付けも可能となる。本論文では本手法の詳細を述べるとともに、更にある種の実環境文字画像認識タスクに適用することでその有効性を評価する。

キーワード セグメンテーション, 有限状態オートマトン, 動的計画法, 実環境文字認識, 情報埋込

1. ま え が き

一次元信号のセグメンテーション問題は様々な応用において検討されてきた。例えば、Lloyd-Max 最適量子化のための信号値の確率分布関数の分割 [1]、濃淡値・カラーヒストグラム分割に基づく画像領域分割 [2] や 2 値化・多値化 [3]、走査線上の濃淡値変化の階段関数近似に基づく画像圧縮 [4]、運動軌跡分割 [5]、周辺分布のセグメンテーションによる文書画像からの文字行抽出 [6]、テキストの話題分割 [7]、音声スペクトルからのフォルマントの分離検出 [8] が挙げられる。音声認識・文字列認識における recognition-based

segmentation [9]~[11] や曲線の多角形近似 [12] も、広義の一次元信号セグメンテーション問題としてとらえられる。

セグメンテーションは、信号をある一定の性質をもつ区間に分割する問題として定義され、信号の階段関数による近似問題に帰着されることが多い。上述の 2 値化・多値化 [3] はその典型例である。階段関数の区間境界と区間値がともに最適化すべき制御変数となり、原信号と階段関数間の最小二乗誤差が目的関数となる。

最適セグメンテーション問題の解法の一つに、動的計画法 (dynamic programming, 以下 DP) がある。実際、上で示した文献の多くでは、DP が利用されている。この理由は、(i) 厳密解が求まる、(ii) パターン長を I 、区間数を N とするとただか $O(I^2N)$ の計算量で済む、(iii) 様々な目的関数を扱える、といった性質を DP が具備するためと考えられる。

ところで、対象とする信号に規則性があれば、それを援用することが有効と予想される。例えば、上述の文書画像からの文字行抽出問題において、信号すなわち周辺分布 (黒画素数の水平射影ヒストグラム) は文

[†] 九州大学大学院システム情報科学研究科, 福岡市
Faculty of Information Science and Electrical Engineering,
Kyushu University, Fukuoka-shi, 819-0395 Japan

^{††} 大阪府立大学大学院工学研究科, 堺市
Graduate School of Engineering, Osaka Prefecture University,
Sakai-shi, 599-8531 Japan

^{†††} 東北大学大学院工学研究科, 仙台市
Graduate School of Engineering, Tohoku University, Sendai-shi,
980-8579 Japan

a) E-mail: uchida@is.kyushu-u.ac.jp

字行の部分で高く、行間において低くなり、それを繰り返す。したがって、階段近似した後の区間値も高低を繰り返すはずであり、それ以外の結果（例えば、区間値が徐々に低くなるような結果）は誤ったセグメンテーションとして排除できるであろう。

そこで本論文では、こうした信号特有の変化規則を有限状態オートマトン（finite state automaton, 以下 FSA）を用いて表現し、それを DP による最適セグメンテーションの枠組みに組み込んだ手法を提案する。FSA の組込みの利点は次の 2 点である。

- 信号固有の変化規則を FSA による制約条件として最適化過程に組み込むことができれば、その規則から逸脱するような分割結果を解空間から排除でき、正しい分割結果が得られやすくなる。

- FSA の状態と信号の各区間の対応結果から、各区間の意味付けが可能となる。例えば、ある状態に対応する区間を注目領域として分離抽出できる。

FSA に基づく DP セグメンテーション法は、従来にもいくつか検討されている。それらと本手法の差異については、2. において本手法の詳細を説明した後に、改めて述べることとする。

本手法の性能を定量的に評価するために、文献 [14] で提案されている文字パターンの認識実験を試みる。この文字パターンには、実環境での機械可読性を高めるために独特の工夫がなされている。具体的には、この文字パターンを横断する線分上の濃淡値変化境界から文字のカテゴリー情報が抽出できるようにデザインされている。本論文では、本手法によるセグメンテーションを用いて、この文字パターンを認識した結果について報告する。

2. FSA を組み込んだ DP セグメンテーション

2.1 基本的なセグメンテーションアルゴリズム

本節では、DP による基本的なセグメンテーションアルゴリズムについて述べる。このアルゴリズムは従来より用いられているものである。今回提案する FSA を組み込んだものについては、次節以降で説明する。

処理対象とする一次元信号を $f(x)$ 、 $x_s \leq x \leq x_e$ と表記する。図 1 (a) はその模式図である（図中の白抜き矢印については後述）。本論文でのセグメンテーションとは、この $f(x)$ を同図 (c) のように $(N-1)$ ステップからなる階段関数で最適近似し、その各段の両端点として境界を検出する問題である。ここで N は（両

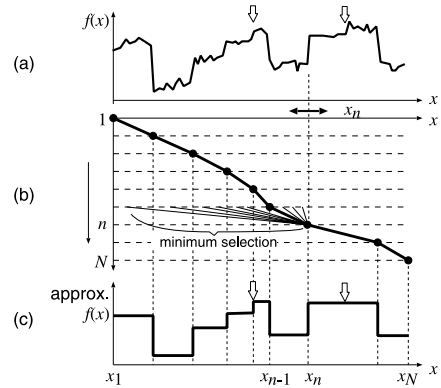


図 1 一次元信号の最適セグメンテーションの基本 (a) 一次元信号 $f(x)$, (b) DP による最適解探索過程, (c) 階段関数近似された信号及び最適境界 x_n

Fig. 1 Basic strategy of DP-based optimal segmentation of a one-dimensional signal. (a) One-dimensional signal $f(x)$. (b) DP-based search for optimal segmentation. (c) Segmentation result.

端点を含めた) 境界数に相当する。

関数 $f(x)$ の階段関数近似すなわち境界検出は、次の目的関数 $J = J(x_1, \dots, x_n, \dots, x_N)$ の最小化問題として定式化できる。

$$J = \sum_{n=2}^N \int_{x_{n-1}}^{x_n} |f(x) - \bar{f}_{x_{n-1}, x_n}| dx \quad (1)$$

ただし、 $x_{n-1} < x_n$ とし、更に境界条件として $x_1 = x_s$ 及び $x_N = x_e$ を課する。上式から分かるように、区間 $[x_{n-1}, x_n]$ の信号値 $f(x)$ 、 $x_{n-1} \leq x \leq x_n$ と平均値 \bar{f}_{x_{n-1}, x_n} が二乗誤差の意味で最も近くなるように区間境界 x_2, \dots, x_{N-2} を定める問題となる^(注1)。

目的関数 J の最小化問題は、 x_n を $n=2$ から $N-1$ まで順次決定していくという多段決定過程として考えることができる。一般に多段決定過程上での最適決定問題は DP により効率的に解けることが知られている。具体的には以下の DP 漸化式

$$g_n(x) = \min_{x' < x} \left[g_{n-1}(x') + \int_{x'}^x |f(\tau) - \bar{f}_{x', x}| d\tau \right] \quad (2)$$

を、 n を 2 から N まで増やしなが、すべての

(注1): ある決められた区間を一定値で近似する場合、その一定値として平均値 \bar{f}_{x_{n-1}, x_n} を用いるのが二乗誤差基準の意味で最良である。このため、式 (1) では平均値 \bar{f}_{x_{n-1}, x_n} からのずれを誤差基準としている。

$x \in [x_1, x_N] = [x_s, x_e]$ において計算することで、 J を最小化する最適境界 x_n を求めることができる。ここで $g_n(x)$ は第 n 境界 x_n を位置 x に設定する場合の最小累積誤差であり、したがって $g_N(x_N)$ が J の最小値となる。また DP 漸化式の初期条件は

$$g_1(x) = \begin{cases} 0 & \text{if } x = x_1 \\ \infty & \text{otherwise} \end{cases}$$

とする。図 1 (b) にこの漸化式計算の様子を示す。式 (2) 計算時に最小値を与えた x' を x_N から遡る処理 (バックトラック処理) により、最適境界が $x_N \rightarrow \dots \rightarrow x_n \rightarrow x_{n-1} \rightarrow \dots \rightarrow x_1$ の順に求まる。図 1 (c) が得られた最適境界を示す階段関数である。

正数 κ_1, κ_2 を用いて式 (2) 中の範囲制約 $x' < x$ を $x - \kappa_2 < x' < x - \kappa_1$ と変更すれば、階段近似の最適性を保ったまま、最小区間幅を κ_1 、最大区間幅を κ_2 に設定できる。図 2 はこれらの値を設定した時の DP 漸化式計算の様子である。現在の境界 $x = x_n$ について、 \times 印のついている位置には直前の境界 $x' = x_{n-1}$ を設定することは不可能になる。このように x' の選択範囲を制限するだけで区間幅制御を実現できる。

2.2 FSA による信号モデル

本節では FSA によって信号値の変化規則をモデル化する方法について述べる。理解を容易にするため、本節並びに次節 2.3 では、「高 → 低 → 中」の 3 レベルを繰り返す階段関数を例として説明する。ただし、実際に処理対象 $f(x)$ となるのは、この原信号に様々なノイズが重畳したもとする。図 3 (a) (図 1 (a) に同じ) はこの $f(x)$ の例である。このようにノイズのために区間境界は曖昧になっている。また直流分の局所的な変動により、本来同じレベルの 2 区間が大きく

異なった信号値をもっている。

本手法の目標は、信号 $f(x)$ を原信号の境界と同一箇所で区間分割し、各区間の原信号での意味 (例では、高、中、低) 付けを行うことにある。この意味付けのために、FSA の状態と区間の対応付けを行う。図 3 (a) の例ならば、高、中、低に対応して $\Omega = \{H, M, L\}$ の 3 状態を準備し、同図 (b) のような状態と区間の対応付けを目指す。次節 2.3 では、この目標達成のための、対応付けの一評価法について述べる。

FSA の状態遷移については、レベルの変化規則に応じて定める。したがって本例では図 3 (c) の FSA を準備する。初期状態を $s_0 \in \Omega$ 、最終状態を $s_F \in \Omega$ とする。信号が高で始まり、中で終わるとすれば、 $s_0 = H$ 、 $s_F = M$ となる。また後の議論のために状態 s に先行可能な状態の集合 $\Phi(s)$ を定義する。もし $s' \in \Phi(s)$ ならば、 s' から s への遷移が可能である。本例では、 $\Phi(H) = \{M\}$ 、 $\Phi(M) = \{L\}$ 、 $\Phi(L) = \{H\}$ となる。

2.3 FSA による評価

本節では、FSA の状態と区間の対応付けの評価法について述べる。前節で述べたように、各区間は図 3 (b) のように一つの状態に対応づけられる。ここで、各区間の原信号における (ノイズ重畳以前の) レベルに一致した状態に対応づけるための工夫が必要である。すなわち、直流分の局所的な変動により見かけの区間値が大きく変わっていても、正しく本来の状態に対応づけるように評価法を設計する必要がある。

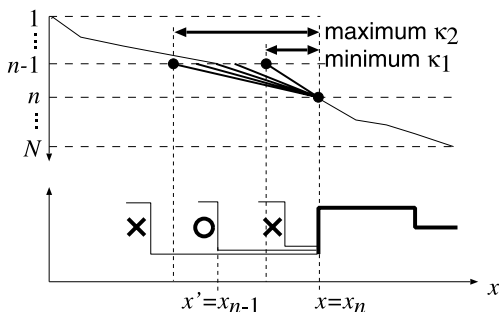


図 2 最大・最小区間幅の設定
Fig. 2 Specification of minimum and maximum segment lengths by κ_1 and κ_2 .

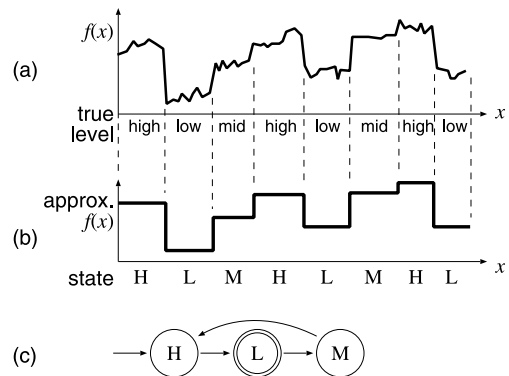


図 3 (a) 高、低、中の 3 レベルが繰り返す信号 $f(x)$ (図 1 (a) に同じ)、(b) 階段関数近似された信号及び区間と状態との対応、(c) レベル変化規則を表現する FSA
Fig. 3 (a) A signal with high, mid, and low levels. (b) Optimized correspondence between segment and state. (c) FSA representing the level transition of the signal.

最も単純には、今考えている区間 $[x', x]$ の信号値から何らかの特徴量を求め、それと状態 $s \in \Omega$ の整合性を評価すればよい。しかし、各区間の信号値はノイズ（特に直流分の局所的変動）により原信号値から大きく変化している可能性があり、1 区間単独での評価には限界がある。そこで本論文では、隣接区間の相対的な関係はノイズが重畳しても変化しにくいという点に着目した評価方法を採用する。具体的には、現在の区間 $[x', x]$ と直前の区間 $[x'', x']$ がそれぞれ状態 s と s' に対応するときの評価関数を用いる。

以下、この評価関数をコスト q と表し、その具体的な設計指針を図 3 の例を通して示す。今、前区間 $[x'', x']$ を $s' = H$ (高) に、現区間 $[x', x]$ を $s = L$ (低) に対応づける場合を考える。このとき、前区間の平均値 $\bar{f}_{x'',x'}$ と現区間の平均値 $\bar{f}_{x',x}$ を比べた結果、もし $\bar{f}_{x'',x'} \leq \bar{f}_{x',x}$ であれば、これは区間と状態の対応付けに矛盾があることになる。そこでこの場合は q を大きな値（本論文では ∞ ）として、この対応付けを許容しないようにする。一方、 $\bar{f}_{x'',x'} > \bar{f}_{x',x}$ であれば、 q を小さな値（本論文では 0）としてこの対応付けを許容すればよい。このように本手法では、区間平均 $\bar{f}_{x'',x'}$ 、 $\bar{f}_{x',x}$ の大小関係によって s' から s への状態遷移の可否を決定できるように $q \in \{0, \infty\}$ を設計する。

実際にはこの q を若干変更して用いる。上記のままであれば、コスト q は s と s' 並びに $\bar{f}_{x',x}$ と $\bar{f}_{x'',x'}$ から決定されるので、 x の決定は x'' すなわち二つ前の x に影響されることになる。このことは、 x_1, \dots, x_n, \dots の多段決定過程が単純マルコフ決定過程ではなくなることを意味し、計算量の増加を招く。そこで q を x'' -非依存とするべく、 $\bar{f}_{x'',x'}$ の代わりに $\bar{f}_{x'-\epsilon, x'}$ を近似値として用いる。ここで、 ϵ は正定数である^(注2)。実際には、図 4 に示すように、現区間 $[x', x]$ の平均値 $\bar{f}_{x',x}$ と前区間後半部 $[x'-\epsilon, x']$ の平均値 $\bar{f}_{x'-\epsilon, x'}$ の大小関係を見て、それが 2 状態 s', s がもつべき大小関係に合致するなら $q = 0$ 、合致しないなら $q = \infty$ とする。このようにコスト q の引数からは x'' は不要となるため、以下では $q(s, s', x, x')$ と表現する。

2.4 基本アルゴリズムへの FSA の組込み

FSA を組み込んだ DP アルゴリズムにおける DP 漸化式は以下ようになる。

$$g_n(x, s) = \min_{\substack{x' < x \\ s' \in \Phi(s)}} \left[g_{n-1}(x', s') \right]$$

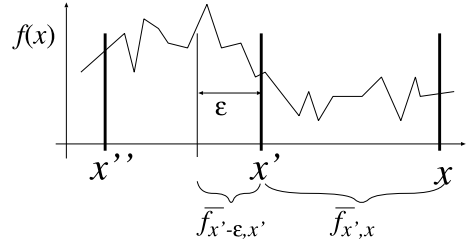


図 4 コスト $q(s, s', x, x')$ の評価対象領域

Fig. 4 Two segments, $[x' - \epsilon, x']$ and $[x', x]$, which are evaluated in $q(s, s', x, x')$.

$$+ \int_{x'}^x |f(\tau) - \bar{f}_{x',x}| d\tau + q(s, s', x, x') \quad (3)$$

このように、 $g_n(x)$ は状態 s を含めたものに拡張され、 $g_n(x, s)$ となる。FSA による事前知識は最小値選択における先行状態制約 $s' \in \Phi(s)$ と既述のコスト $q(s, s', x, x')$ として表現されている。この DP 漸化式 (3) を n を 2 から N まで増やししながら、すべての $x \in [x_1, x_N]$ と $s \in \Omega$ の組について計算する。ただし初期条件は

$$g_1(x, s) = \begin{cases} 0 & \text{if } x = x_1 \text{ and } s = s_0 \\ \infty & \text{otherwise} \end{cases}$$

となる。最小誤差 $\min J$ は $g_N(x_N, s_F)$ で得られ、最適境界及び各区間の状態はバックトラックにより得られる。

2.5 FSA 組込みの効果

1. でも述べたように、FSA の組込みには、(i) 高精度化と (ii) 各区間の意味付けの 2 点の効果が見込める。後者 (ii) については、本手法により各区間が一つの状態に対応づけられることから自明である。前者の高精度化について、以下、図 1 と図 3 の比較により説明する。

両図 (a) の信号は同じものであるが、セグメンテーション結果は FSA 制御の有無により両図において異なっている。図 1 (c) は、FSA 制御なし、すなわち信号が「高 → 低 → 中」を繰り返すという規則性を考えずにセグメンテーションを行った場合の模式図であるが、白抜き矢印で示した箇所のように、微小なノイ

(注2): ϵ の値は固定でもよいし、 $\epsilon(s')$ のように状態 s' に応じて切り換えてもよい。後述する実験ではこの切換を用いている。最小及び最大区間幅 κ_1, κ_2 についても同様の切換が可能である。

ズにより区間が過剰検出されたり、変化の小さい隣接区間ではそれらが 1 区間に統合されて検出されるなど、問題が生じやすい。一方、FSA を組み込んだ場合、図 3 (b) が示すように、区間平均が「高 → 低 → 中」を繰り返す結果だけを許容するという制約から、こうしたノイズなどの影響は極小化されやすい。したがって、高精度なセグメンテーション結果が得られるものと期待される。

2.6 境界数 N の決定

応用によっては検出線上の境界数 N が未知であり、その最適推定を行う必要がある。本論文では、 N を想定される最小値 N_{\min} から最大値 N_{\max} まで変えながら、それぞれの N で境界検出を行い、最終的なコスト $g_N(x_N, s_F)$ が最小となる N を境界数として決定するという方針を採る。

この N の決定法は、単純な試行錯誤過程ではあるものの、DP の性質のために非常に効率的に行うことが可能である。具体的には、境界数を N と仮定したときの結果が得られていたとすると、式 (3) を用いて $g_N(x, s)$ から $g_{N+1}(x, s)$ を求めさえすれば、境界数を $N + 1$ としたときの結果が求まったことになる。すなわち、 N を増やす度に DP 計算を最初からやり直す必要はなく、 $N = N_{\max}$ までの DP 計算を 1 回行えば、最適な N の決定に必要な $g_N(x_N, s_F)$ ($N \in [N_{\min}, N_{\max}]$) の値がその途中の過程ですべて求まる。

なお、 N の値によっては、FSA に受理されるようなセグメンテーションが存在せず、 $g_N(x_N, s_F)$ の値が無限大になる場合もある。これはむしろ好都合であり、多くの場合、現在の境界数 N が不適切であることを意味する。

2.7 他手法との関連

DP に基づく最適セグメンテーションに FSA を組み込むという方式は、前出の recognition-based segmentation [9] ~ [11] に見られるものである。しかし、以下の 2 点において、本手法の新規性が認められる。

- 隣接する区間について、その区間の性質（本手法においては区間平均値）と状態対応を評価するコスト q により、モデルとの整合性を評価する点が独特である。

- FSA と DP を組み合わせた最適セグメンテーションを、信号の最適階段関数近似に基づく境界検出に応用した例がほかにない。

以上 2 点のうち、第 1 点の新規性を明確にするべく、recognition-based segmentation で用いられるコ

スト q について述べる。このコストは、各区間と状態の整合性を評価するコスト並びに隣接する状態の組合せ (s, s') を評価するコストの和として定義されることが多い。すなわち q は $q(s, x, x') + q(s, s')$ のように表され、本手法の $q(s, s', x, x')$ とは異なる。なお、文字列認識 [9] では、 s 及び s' は隣接する区間が対応する文字（若しくは部首）であり、 $q(s, x, x')$ はそのパターンの区間の文字 s との非類似度、 $q(s, s')$ は文字 s と s' の隣接コストすなわちパイグラムに相当する。連続音声認識 [10], [11] も同様である。文献 [13] では、recognition-based segmentation を二次元的な境界検出に拡張しているが、FSA による評価形式はこれら一次元の場合と同様である。

この従来の考えに基づくコスト q を信号 $f(x)$ のセグメンテーション問題に適用することも可能である。この場合、 $q(s, x, x') + q(s, s')$ のうち主として第 1 項を利用することになる。具体的には、状態 s の標準的な信号値 v_s が存在するとして、 $q(s, x, x') = \|\bar{f}_{x',x} - v_s\|$ のように q を定義することになる。しかしながら、 v_s を用いて各区間単独で評価する方式では、2.3 でも述べたように、信号の直流成分の全体的・局所的な変動により、適切な結果が得られない場合がある。例えば図 3 (a) の信号では、 x に従って直流成分が増加しているため、1 回目の中レベル区間と 2 回目の低レベル区間が同じような値になっている。この場合、標準値 v_s を用いた手法では正しくセグメンテーションできないことになる。一方、2.3 で述べた q では、隣接する区間の平均 $\bar{f}_{x'',x'}$, $\bar{f}_{x',x}$ の大小関係（実際には $\bar{f}_{x'-\epsilon,x'}$, $\bar{f}_{x',x}$ の大小関係）だけを評価に用いているので、こうした状況を回避することが可能である。その結果、図 3 (b) のように、上述の 2 区間については、ほぼ同じ区間平均値であってもそれぞれ M と L に対応づけることができる。なお、従来方式である $q = q(s, x', x)$ と本論文で提案する $q = q(s, s', x, x')$ については、以下 3. において実験的に比較検討する。

最後に HMM との関連について述べる。一般に DP と HMM は非常に関連しており、実際 HMM で利用されるビタビアルゴリズムは DP そのものである。また DP のコストの与え方を確率的に拡張したものを HMM と見ることもできよう。しかし、音声認識などで用いられている通常の HMM の考え方をセグメンテーション問題に転用するのはさほど単純ではない。通常の HMM では x が $x + 1$, $x + 2$ と一つ進むごとに状態遷移を起こすため、結果的に 1 サンプル点単位

での評価に基づく最適化問題を扱っていることになる。このためセグメンテーションのように区間を単位とする最適化問題は扱いにくい。別の方針で（例えば x_n を直接観測変数とするように）定式化できたとしても、単調性制約 ($x_{n-1} < x_n$) の厳密な導入は困難であろう。また、本手法のコスト q のような if-then ルール的な評価は陽には扱えなくなるため、評価の自由度が低下する。更に大量の学習データが必要になるという問題も発生する。このように問題は多くあるが、HMM のような学習機能の導入は興味深く、応用によっては精度向上に役立つ可能性もある。今後の課題としたい。

3. 実験

3.1 実験タスクの概要

本手法は汎用的であり、1. で述べたような文書画像からの文字行抽出などへの応用が見込めるが、本論文では文献 [14] で提案されている文字パターンに適用した結果を述べる。この文字パターンは、もともと黒色（濃淡値 0）の文字線と白色（255）の背景からなる白黒文字画像に、実環境下でも安定して認識されるような工夫を施したものである。具体的には、文字の 카테고리を表す平行じまからなる模様（以下、複比パターン）が白黒文字画像上に重畳印刷されている。図 5(a) は複比パターンを文字画像 “A” に重畳して印刷した例である。一番上と一番下のしま（濃淡値 175）はガイドと呼ばれ、複比パターンの存在領域を表す。また、ガイドに挟まれた領域は幅 l_1, l_2, l_3 をもつ三つのしまからなる。幅 l_2 のしまのみが有色であり、ここでは文字線と背景の濃淡値がそれぞれ 80 と 215 に変化する。

文字カテゴリを複比パターンのしまの幅 l_1, l_2, l_3 により表現する。具体的には次の複比

$$r = \frac{(l_1 + l_2)(l_2 + l_3)}{l_2(l_1 + l_2 + l_3)} \quad (4)$$

により、文字カテゴリを表現する。図 6 のように、カテゴリごとに異なる複比パターンを重畳印刷しておけば、複比 r を抽出することで、各文字のカテゴリ $c \in C$ (C はカテゴリ集合) を確定できることになる。

複比は以下のように本手法を用いることで抽出できる。まず上下二つのガイド及び文字を縦断するような直線を引く（図 5(b) 中の p ）。以下、この直線を検出線と呼ぶ。次に検出線上の濃淡値変化を $f(x)$ として本手法によりしまの境界を求め（詳細は 3.2 にて後述）、それから検出線上でのしまの幅 l'_1, l'_2, l'_3 を求める。その際、検出線が文字のどの部分を通過するかは未知であるため、境界数 N を 2.6 の手法により推定する必要がある。こうして求めたしまの幅を l_1, l_2, l_3 の代わりに用いて、式 (4) により複比 r を求める。複比は射影不変であり、またしまはもともと平行なので、検出線が上下ガイドを通過する限り、どのような角度から撮影しても、原理的には常に埋め込んだ複比が r として求まる。倒立や鏡像にも不変である。

実際にはデジタル画像特有の離散化誤差のため、求まる複比 r は埋め込んだ複比（真の複比）の値とは厳密には一致しない。加えて、境界数 N の誤推定も起こる。そこで認識の際には、求めた r を埋込時に想定した $|C|$ 種の複比のうちで最も近いものに量子化することで、各検出線での複比を定める。更に、文字画像上に P 本の検出線を引き、それぞれで得られた P 個の（量子化後の）複比の値について多数決を行うことで、最終的なカテゴリを決定する。より洗練された推定・統合処理も考えられるが、本論文では実現の容易さから多数決を採用した。

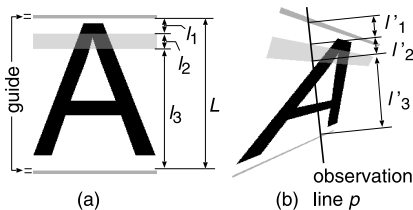


図 5 (a) 複比パターンが重畳印刷された文字画像及び (b) その射影変換例

Fig.5 (a) A character image “A” printed with a cross ratio pattern. (b) Projective distortion.



図 6 複比パターンが重畳印刷された英大文字

Fig.6 Character images printed with different cross ratio patterns.

3.2 境界検出の詳細

検出線上の濃淡値変化 $f(x)$ の境界には、複比パターンのしまによる境界と、文字線と背景の境界が含まれている。前者は複比の計算に必要であるが、後者は不要である。したがって境界の区別が必要になる。この際に、各領域の意味付けが可能という、FSA 導入の効果を生かすことができる。

検出線上の濃淡値変化を FSA 表現するために、もともとの文字画像を図 7(a) のように 10 領域に分けて考える。領域 B0, B1, B2, B3 は背景領域である。いずれももともとは同じ濃淡値をもつ領域であるが、複比パターンのしまを境界として 4 領域に区別している。また B+ は背景領域に複比パターンの中央のしまが重畳された領域である。同様に F1, F2 は文字線領域であり、F+ はそこに複比パターンの中央のしまが重畳された領域である。残る G1, G2 は上下のガイドの領域である。ここで、例えば同図 (a) の検出線 p を考えると、(b) のように領域を遷移することになる。他の検出線では当然異なる領域を遷移する。しかし、どのように検出線を引いても、それが上下ガイドを通過する限り、同図 (c) のような遷移規則に従うことが分かる。以下では、領域を状態とみなし、(c) を検出線上の濃淡値変化を表現する FSA と考える。状態集合は $\Omega = \{B0, B1, B2, B3, B+, F1, F2, F+, G1, G2\}$ 、初期状態 s_0 は B0、最終状態 s_F は B3 となる。

この FSA に基づき、コスト q を 2.3 に述べた方法により定める。すなわち、各状態の標準的な値は一切用いずに、状態の組 (s, s') がとるべき濃淡値の大小関

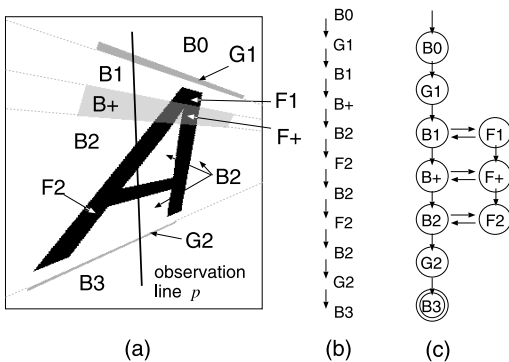


図 7 検出線上の濃淡値変化の FSA 表現 (a) 各領域への状態の割当, (b) 検出線 p 上の状態変化, (c) すべての状態変化を受理する FSA

Fig. 7 FSA representation of the transition of the intensity level along an observation line p .

係のみに基づいて 0 若しくは ∞ の値をとるように定めた。例えば区間値が小(暗)から大(明)に変化するならば、状態 (s, s') が $(B1, G1)$, $(B1, F1)$, $(F+, F1)$, $(B+, F+)$, $(B2, B+)$, $(B2, F2)$, $(B3, G2)$ の場合に q は 0 となり、それ以外の組では q は ∞ となる。

各検出線について、この FSA を組み込んだ 2.4 の DP アルゴリズムを適用し、更に、バックトラック処理により境界系列と状態遷移系列を求める。その状態遷移系列のうち、 $B1 \rightleftharpoons F1$, $B2 \rightleftharpoons F2$, $B+ \rightleftharpoons F+$ の計 6 種類の遷移にかかわる境界については文字線と背景間の境界として無視する。残る遷移にかかわる境界を用いて l'_1, l'_2, l'_3 を求めることで、その検出線が示す複比 r を抽出できる。

なお、複比パターンのしまの境界を単純なしきい値処理で検出するのは困難である。これは、実際に観測される濃淡値(すなわち撮影された文字画像の濃淡値)が照明等の影響により変動するためである。図 8 はカメラで撮影された文字画像について、一つの検出線上の濃淡値変化をプロットしたものである。白抜き矢印で示された部分は、背景でありもともとは同じ濃淡値をもつ箇所であるが、大きく異なった値となっている。またもともと階段状に変化しているはずの濃淡値が、細かく変動していることも分かる。

3.3 実験 試料

認識対象として、複比パターン付きの英語大文字 26 種類(図 6)の画像をプリンタを用いて白紙に印刷し、カメラで撮影した。図 9 に示すように、撮影方向については、カメラと紙面を正対させたものを 1 方向、カメラと紙面を正対させていないもの 8 方向(いずれも紙面法線に対しておよそ 45 度の撮影角度)の計 9 方向を想定した。したがって本実験では $26(\text{文字種}) \times 9(\text{角度}) = 234$ 画像を使用した。撮影環境を室内とし、カーテンを締め切って蛍光灯の光を

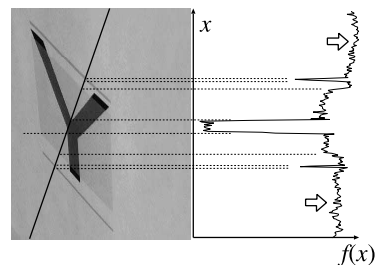


図 8 実環境中の濃淡値変動

Fig. 8 Contamination of intensity level by noise.

照明とした．画像サイズは図 9 のように撮影方向で異なり，最小で 300×225 ，最大で 493×370 ，平均 381×291 であった．なおカメラ撮影時はカラー画像であるが，これを濃淡画像（グレースケール）に変換したものを処理対象とした．

こうして撮影された画像を上下 2 方向から走査することで，まずガイドを検出した．この検出の際に限り，背景がもともと白地であることを利用したしきい値処理を用いた．次に，上下ガイドをそれぞれ 10 等分割し，九つの分割点を求めた．この上下それぞれの 9 分割点を互いにすべて結びつことで，各文字画像に $P = 81$ 本の検出線を定めた．

パラメータ ϵ については今回は実験的に $\epsilon = 5$ とした．ただしガイド状態 ($G1, G2$) については幅が細いことに配慮して $\epsilon = 3$ とした．同様に最小区間幅 κ_1 についても，ガイド以外で 5，ガイドで 3 とした．画像中のしめの幅を超えるような大きな値としない限り実験結果に大きな影響はなかったものの，タスクに応じた ϵ, κ_1 の自動設定法については今後の課題といえる．なお最大区間幅は設定しなかった ($\kappa_2 = \infty$)．

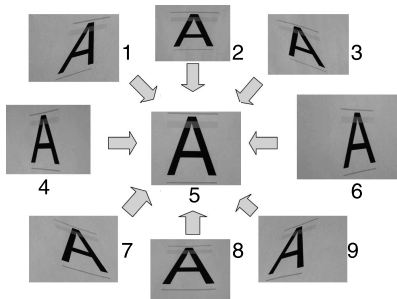


図 9 画像の撮影方向
Fig.9 Camera angles.

3.4 境界検出結果

境界検出の結果例を図 10 に示す．同図 (a) は FSA を組み込まず，式 (1) の最小化問題を単純に DP で解いた場合の結果であり，(b) は FSA を組み込んだ場合の結果である．FSA を組み込まない場合，不要な境界が二つ発生し，逆に下部ガイドに関する境界が二つとも検出されていない．一方，FSA を組み込んだ場合，すべての境界が正しく検出されている．この改善は，同図 (a) の誤った階段近似（5 回連続で平均濃淡値が減少）が FSA で受理されず，結果から排除されたためと考えられる．このように FSA により濃淡値変化モデルを導入することで，境界検出精度を向上させている．

次節以降の定量的評価からも明らかになるように，図 9 程度の射影変換ひずみに対して，本手法は多くの場合において正しい境界検出結果を与えた．射影変換ひずみを更に強くするべく撮影角度を 45 度以上に増加していくと，ある程度までは対応できていたが，徐々に性能は劣化した．これは，強い射影変換ひずみにより，画像中のガイド幅が κ_1 や ϵ 以下になり，ガイドを正しく検出できなくなったためである．

3.5 複比抽出成功率

次に，境界検出精度を定量的に評価した．本手法では，各検出線において複比 r を求め，それを $|C| = 26$ 種のうちで最も近いものに量子化している．そこで，この量子化後の複比 r が真の複比に等しければ，その検出線では正しい複比が抽出できたと考える．そのような検出線の割合を，複比抽出成功率として評価に用いた．図 11 及び図 12 は，それぞれ文字種別及び撮影角度別の複比抽出成功率である．詳細には，図 11 では $81(\text{本}) \times 9(\text{角度})$ のうち正しい複比を与えた検出線の割合，図 12 では $81(\text{本}) \times 26(\text{文字種})$ のうち正しい複比を与えた検出線の割合を成功率として示している．

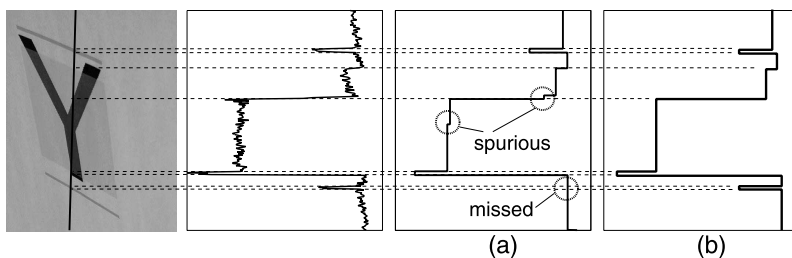


図 10 境界検出結果の例 (a) FSA なし (b) FSA あり
Fig.10 Segmentation results without FSA (a) and with FSA (b).

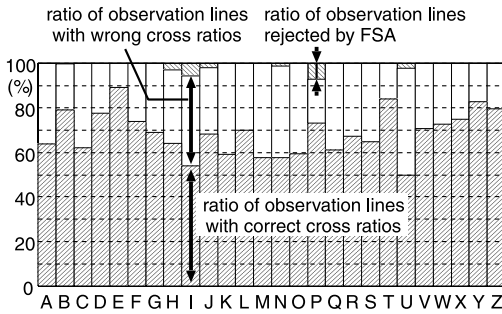


図 11 文字種別複比抽出成功率
Fig. 11 Extraction accuracy of cross ratios for each category.

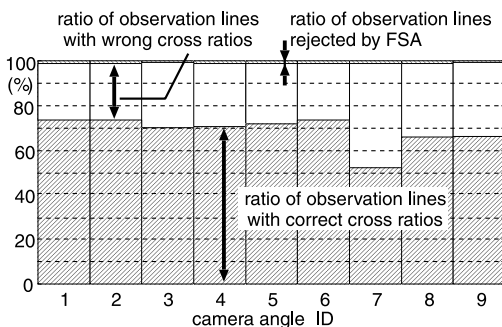


図 12 撮影方向別複比抽出成功率．横軸の撮影角度 ID は図 9 中の数字に対応．
Fig. 12 Extraction accuracy of cross ratios for each camera angle.

いずれの図からも半数以上の検出線において正しい複比が抽出されていることが分かる．平均では 69.2% の成功率であった．複比が正しいということは、その検出線上のすべての境界検出に成功していることを意味している．したがって、検出線ではなく、各境界を単位とした境界検出率（それぞれの境界が検出できているか否か）を考えると、69.2% よりもはるかに高い値になっていると考えられ、本境界検出手法は十分な精度をもつものと考えられる．

複比抽出の失敗要因には次の三つがあった．

- 第 1 は、境界数 N の誤推定である．本実験で用いた全検出線 81×234 本のうち、正しい N が推定されたのは、80.9% であった．すなわち、2 割の検出線では境界が過剰若しくは過少検出されており、その影響があったといえる．

- 第 2 は、境界位置を間違っても FSA に受理される場合が存在するためである．すなわち、FSA によるチェックを通過してしまう場合である．例えば、境

界位置の検出誤差が比較的小さな場合が相当する．

- 第 3 は、濃淡値変化 $f(x)$ が FSA に従わない場合である．例えば文字線・背景境界と複比パターン境界が一致してしまった場所は、濃淡値変化が図 7(c) の FSA で想定されているものにならず、境界が求まらない．ただし、複比パターン重畳の時点で、こうした一致はなるべく起こらないように配慮したため、結果的にあまり大きな問題にはなっていない．実際、図 11、図 12 には、FSA で受理されないために複比抽出に失敗した検出線の割合も表示されているが、非常に小さいことが分かる．

3.6 文字認識率

3.1 で述べたように、本手法では全 81 本の検出線で得られる複比の多数決で認識カテゴリーを決定する．こうして得られた文字の認識率は $99.6\% = 233/234$ であった．すなわち誤った認識結果はわずかに 1 例であった．前節で述べたように、半分以上の検出線で正しい複比が求まったことは、ほとんどの文字画像で多数決により正しい複比が選出されることを意味しており、本結果を裏づけている．

失敗した 1 例は、“O” の “N” への誤認識であった．81 本の検出線のうち、27 本で “N” 用の複比が、26 本で “O” 用の複比が求まっており、多数決の結果 “N” が選ばれていた．図 6 より分かるように、両文字の複比パターンは類似しており、したがってこれらの複比の差異は小さいため、微小な検出ずれの影響により誤認識となった．

3.7 比較実験

本手法で用いたコスト q 以外にも、従来法の考えに基づいたコストを考えることもできる．そこで、2.7 で述べた recognition-based segmentation のコスト $q(s, s') + q(s, x, x')$ について、本セグメンテーション問題において利用可能な第 2 項すなわち $q(s, x, x') = \|\bar{f}_{x',x} - v_s\|$ を実際に用いて境界検出実験を行った．各状態の標準的な信号値 v_s としては、最も単純には印刷時の濃淡値をそのまま使うことが考えられる．しかし、今回は照明による影響に配慮し、撮影した画像中の最大、最小濃淡値がそれぞれ背景 ($s \in \{B0, B1, B2, B3\}$)、文字線 ($s \in \{F1, F2\}$) となるように正規化した上で、各 v_s を適応的に定めた．

このように極力有効に $q(s, x, x')$ が機能するように工夫したものの、複比抽出率は本手法 (69.2%) の半分にも及ばない 28.3% であった．文字認識率も $36.8\% = 86/234$ とかなり低いものであった．この理

由としては、図 8 に関して 3.1 で述べたように、もともと濃淡値をもつ部分であっても、非均質な照明等による全体的・局所的な変動により、異なる濃淡値として観測されることが多いことが考えられる。逆にいえば、そういう濃淡値変動があっても、本手法では多くの場合に正しい複比抽出結果が得られており、コスト $q(s, s', x, x')$ の有効性を裏づけている。

4. む す び

本論文では、一次元信号の最適セグメンテーションのために、動的計画法 (DP) に基づくアルゴリズムを提案した。本アルゴリズム最大の特徴は、オートマトン (FSA) を導入することで、モデルベースのセグメンテーションを可能にしている点である。これまでも FSA を導入したセグメンテーション法は提案されているが、本手法では隣接区間の信号値の相対関係を表現するために FSA を導入している点に新規性がある。

本手法の性能を定量的に評価するために、実環境での機械可読性を高めるためにしま模様が重畳されている文字パターンの認識実験を試みた。この文字パターンは、それを横断する線分上の濃淡値変化境界からそのカテゴリー情報が抽出できるようにデザインされている。本手法によりこの境界検出を行ったところ、実環境下での照明変動があっても、多くの場合正しい境界が得られることが分かった。更に検出された境界からカテゴリー情報を抽出し、文字認識を行った。実環境文字認識はスキャナベースの文字認識に比べ非常に困難なタスクとされているが [15]、9 種類の角度から撮影された英語大文字 234 文字画像のうち 233 文字を正しく認識できることを確認した。

今後の課題としては、本手法の他の応用が挙げられる。1. で述べたように、一次元信号のセグメンテーションは様々な応用があり、それらの中には本手法の FSA 制御が生かされる場合も多いと考えられる。1. で述べた周辺分布のセグメンテーションによる文字行抽出はその一例である。運動軌跡分割についてもより知的な分割法の実現につながる可能性がある。更に本論文で比較対象としてきた recognition-based segmentation についても、本手法で採用した評価方式を導入することで、隣接要素間の物理的関係 (音声では調音結合、文字列ではリガチャ) を考慮した新たな方式が実現できる可能性がある。また本手法は本来が階段関数になっていない信号についても応用可能であるが、その

際のコスト q の設計についても一考の余地があると考えている。更に 2.7 で述べたように、HMM に類した学習機能の導入も興味深い課題である。

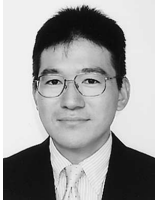
謝辞 本研究の一部は科研費 (若手 (B), No.17700198)、並びに科学技術振興機構「実用化のための可能性試験」(FS) によった。

文 献

- [1] A.K. Jain, "Image data compression: A review," Proc. IEEE, vol.69, no.3, pp.349-389, March 1981.
- [2] R.M. Haralick and L.G. Shapiro, "Survey: Image segmentation techniques," Comput. Vis. Graph. Image Process., vol.29, pp.100-132, 1985.
- [3] 大津展之, "判別および最小二乗規準に基づく自動しきい値選定法," 信学論 (D), vol.J63-D, no.4, pp.349-356, April 1980.
- [4] 坂東幸浩, 西 修巧, 鎌田清一郎, "ヒルベルト走査を用いた時空間領域分割による高速動画画像圧縮法," 映像学誌, vol.53, no.4, pp.559-564, April 1999.
- [5] R. Mann, A.D. Jepson, and T. El-Maraghi, "Trajectory segmentation using dynamic programming," Proc. ICPR, vol.1 of 4, pp.331-334, 2002.
- [6] S. Messelodi and C.M. Modena, "Automatic identification and skew estimation of text lines in real scene images," Pattern Recognit., vol.32, no.5, pp.791-810, May 1999.
- [7] 越仲孝文, 奥村明俊, 磯谷亮輔, "HMM の変分ベイズ学習によるテキストセグメンテーション及びその映像インデキシングへの応用," 信学論 (D), vol.J89-D, no.9, pp.2113-2122, Sept. 2006.
- [8] L. Welling and H. Ney, "Formant estimation for speech recognition," IEEE Trans. Speech Audio Process., vol.6, no.1, pp.36-48, Jan. 1998.
- [9] R.G. Casey and E. Lecolinet, "A survey of methods and strategies in character segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol.18, no.7, pp.690-706, 1996.
- [10] H. Sakoe, "A generalized two-level DP-matching algorithm for continuous speech recognition," IECE Trans., vol.E65, no.11, pp.649-656, Nov. 1982.
- [11] H. Ney and S. Ortman, "Progress in dynamic programming search for LVCSR," Proc. IEEE, vol.88, no.8, pp.1224-1240, Aug. 2000.
- [12] J.-C. Perez and E. Vidal, "Optimum polygonal approximation of digitized curves," Pattern Recognit. Lett., vol.15, pp.743-750, 1994.
- [13] 平 英二, 高山誠悟, 内田誠一, 迫江博昭, "モデル当てはめによる書棚画像解析," 信学論 (D-II), vol.J87-D-II, no.2, pp.565-573, Feb. 2004.
- [14] 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, "カメラによる文字認識のためのカテゴリー情報の埋込に関する検討," 信学論 (D), vol.J89-D, no.2, pp.344-352, Feb. 2006.
- [15] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, "デジタルカメラによる文字・文書の認識・理解," 信学誌, vol.89,

no.9, pp.836-841, Sept. 2006.

(平成 18 年 10 月 10 日受付, 19 年 1 月 22 日再受付)



内田 誠一 (正員)

平 2 九大・工・電子卒。平 4 同大大学院修士課程(情報)了。セコム(株)勤務を経て、現在、同大システム情報科学研究院知能システム学部門准教授。博士(工学)。画像パターン・時系列パターンの解析・認識に関する研究に従事。平 15 本会 PRMU 研究奨励賞, 平 18 MIRU 長尾賞各受賞。IEEE, 情報処理学会, 映像情報メディア学会各会員。



酒井 恵 (学生員)

平 18 九大・工・電気情報工卒。現在、同大システム情報科学府修士課程。画像パターン認識に関する研究に従事。



岩村 雅一 (正員)

平 10 東北大・工・通信卒。平 15 同大大学院博士課程了。同年同大大学院工学研究科助手。平 16 阪府大大学院工学研究科助手。現在助教。博士(工学)。パターン認識, コンピュータビジョンに関する研究に従事。IEEE, 情報処理学会, Pattern Recognition Society 各会員。



大町真一郎 (正員)

昭 63 東北大・工・情報卒。平 5 同大大学院博士課程了。同年同大情報処理教育センター助手。平 8 同大工学部助手。平 11 同大大学院工学研究科助教授。現在准教授。博士(工学)。その間, 平 12~13 米国ブラウン大学客員助教授。パターン認識, コンピュータビジョン, 並列処理, 文字認識システムの開発などの研究に従事。IEEE, 情報処理学会, 人工知能学会, Pattern Recognition Society 各会員。



黄瀬 浩一 (正員)

昭 61 阪大・工・通信卒。昭 63 同大大学院博士前期課程了。同年同大大学院博士後期課程入学。平 2 阪府大・工・電気助手。平 5 同情報助手, 平 8 講師, 平 11 助教授を経て, 平 17 同大大学院工学研究科教授。現在に至る。博士(工学)。その間, 平 12 年~13 ドイツ人工知能研究センター客員教授。文書画像解析, 情報検索などの研究に従事。情報処理学会, 人工知能学会, 電気学会, IEEE, ACM 等各会員。