

固有空間法とDTWによる古文書ワードスポッティング

寺沢 憲吾^{†a)} 長崎 健^{††} 川嶋 稔夫^{††}

Word Spotting for Historical Document Images with Eigenspace Methods and DTW

Kengo TERASAWA^{†a)}, Takeshi NAGASAKI^{††}, and Toshio KAWASHIMA^{††}

あらまし 歴史的文書のデジタルアーカイブの構築を考える場合、毛筆手書き文字に対する文書解析手法の開発は必要不可欠である。本論文では毛筆手書き文書画像に対するキーワード検索のための新しい手法として、文字認識手法によらず画像の部分マッチング問題として検索を行う方法を提案するとともに、提案手法の有効性を実験的に確認する。文字列画像をスリット状に切り出すことにより文字列画像はスリット画像のシーケンスとして表現され、更にこれに固有空間法を適用して低次元化することにより効率的なマッチングが可能となる。また、マッチングに際してDTW (dynamic time warping) を用いて文字の伸縮変形に対応させることにより、更に精度を高めることができる。江戸末期の毛筆文書画像を対象にキーワードの検索を行った実験では、平均適合率 73~93%を示した。

キーワード 毛筆手書き文書, ワードスポッティング, 文字切出し, 固有空間法, DTW

1. ま え が き

地域の図書館や資料館には多くの歴史的資料が貯蔵されており、これらをデジタルアーカイブとして公開し、広く世界に発信して一般の活用を促すことは、学術的文化的観点からのみならず経済的観点からも極めて有益である。実際にデジタルアーカイブを構築することを考える場合、資料をデジタル化して貯蔵する方法だけでなく、貯蔵された情報の中から必要な情報へ素早くアクセスする方法を提供することも主要な技術的課題となるが、歴史的文書のうち特に明治期以前のもは毛筆手書きで書かれたものが多いため従来の文字認識手法の適用が困難であり、自動的な文書解析を行うことができない。そのため現状では特に史料価値の極めて高い文献に対してのみ手作業でインデックス作成が行われているにとどまっている。こうした状況から、歴史的資料におけるデジタルアーカ

イブの対象を更に拡大することを考えた場合、歴史的文書画像に対する解析手法の研究開発の必要性は非常に高いものといえる。

一つの方法は、文字認識手法(OCR)により文書画像をテキスト形式に変換して取り扱うことである。しかし歴史的文書に対してOCRを適用することは極めて困難である。なぜなら毛筆文字は線幅が太く安定した細線化を行いにくいことに加え、崩し字体が多く用いられることや、更には保存状況による劣化などの問題もあるからである。その結果、OCRの第1段階である文書画像を文字単位に切り出すことからして難しく[1]、この精度を高める研究[2]が現在も行われており、いまだ発展途上である。

これに対し本論文では文字認識ではない別な方法、つまり、文書画像をテキスト形式に変換することなく画像形式のままに検索を行う手法について検討する。ここでは、文書画像中からある指定した文字列の部分と類似度の高い部分を検索し、抽出することを目的とした。これが可能となることにより、文書画像中から特定のキーワードを含む部分を抽出することができるほか、インデックス作成の作業支援や、あるいは翻刻者が解読できない文字列に遭遇した際にそれと同一の文字列が現れる別の文脈を提示することによる解読支

[†] 北海道大学知識メディアラボラトリー, 札幌市
Meme Media Laboratory, Hokkaido University, Sapporo-shi,
060-0808 Japan

^{††} 公立はこだて未来大学大学院システム情報科学研究科, 函館市
School of Systems Information Science, Future University-
Hakodate, Hakodate-shi, 041-8655 Japan

a) E-mail: terasawa@meme.hokudai.ac.jp

援なども行えると考えている。

1.1 関連研究

本論文で扱うような文字認識によらない文字列検索の研究としては、英語の手書き文書を対象に頻出する単語を抽出した Manmatha ら [3] によるものがあり、“ワードスポッティング” (word spotting) と名づけられている。Rath and Manmatha は、ワードスポッティングに適した四つの特徴量を提案し [4]、またこれに DTW (dynamic time warping) を適用することにより精度の向上を図っている [5]。また、Marinari ら [6] は主に活字で印刷された古文書を対象に、自己組織化マップを用いて連結成分を符号化し、符号列エディット距離を用いて単語間の対応を定める方法を提唱している。これらの研究はいずれも英語の文書を対象としているため、単語単位の抽出しが比較的容易に行われることを前提に、単語間の対応付けを行っている。

一方で、日本語や中国語のような言語では単語単位の抽出しが難しい。こうした言語を対象とした研究としては Yue Lu and Chew Lim Tan [7] が中国語の新聞記事を対象とした文字列検索の手法を提案しているものがあるが、これは活字で印刷されたものを対象としており、文字単位の抽出しを行うことが前提となっている。

崩し字やつづけ字などにより文字抽出しが困難な文書に対する文字抽出しを前提としない研究としては、探索範囲を単一文字に限定せずに抽出しすることとした近藤ら [8] の研究がある。ここでは文字幅に着目して探索範囲を切り出し、切り出した範囲と正規化した文字パターンとの間でテンプレートマッチングを行うという方法がとられている。また、筆者らは文献 [9], [10] において、毛筆手書き文書を対象としたワードスポッティングのために、文字列をスリット状に切り出して固有空間法を適用する方法及びこれに DTW を適用して文字列の伸縮に対するロバスト性を付加する方法を提唱した。

本論文では、[9], [10] で提唱した手法について述べるとともに、「蘭亭序」「亜国来使記」の 2 種類の毛筆文書画像を対象にこの手法の精度に対する定量的な評価を行い、その有効性を検証する。

なお本手法は画像の類似度に基づくスポッティングであるため、文字の書体が比較的安定している文書に対象が限定されてしまう。しかし、「亜国来使記」のような江戸期の公文書は「御家流」と呼ばれる書体を用

いることと公式に定められており、このように比較的書体が安定した文書は膨大な量が保存されている。したがって、画像の類似度に基づくスポッティングの適用範囲は限界があるとはいえ十分に広く、本研究もこうした文書のみを対象として考えることとする。

2. スリット抽出しによるワードスポッティング

この章では、文書画像をスリット状に切り出すことによって、ある文字列画像からそれに類似した文字列画像を検索する方法の手順について述べる。

2.1 前処理

はじめに、入力画像に対して前処理を施す。前処理は、

- しきい値処理により、背景を消去
- 文字行の抽出し
- 文字行の中心位置の正規化

の 3 段階で行う。

しきい値処理は、画素値が一定のしきい値以下の (すなわち黒に近い) ピクセルのみを有効成分として抽出し、それ以外の部分は背景とみなすという形で行う。すなわち、 $I(x, y)$ を入力画像、 $\tilde{I}(x, y)$ を処理後の画像 (x, y は画像の座標を表す) として、

$$B(x, y) = \begin{cases} 1 & (I(x, y) \leq t) \\ 0 & (I(x, y) > t) \end{cases} \quad (1)$$

$$\tilde{I}(x, y) = (255 - I(x, y)) \times B(x, y) \quad (2)$$

とする。ここで t は入力画像の性質に応じて定めるしきい値であり、 $B(x, y)$ は 2 値化画像に相当する。こうして得られる $\tilde{I}(x, y)$ は 2 値化画像とは異なって文字部分の画素の濃淡情報はそのまま残されており、この後の解析もすべてグレースケールの領域で行う。このような方法を採用したのは、毛筆文書画像においてはペン字画像と異なり、画素値の濃淡に文字識別のために有益な情報が含まれている可能性があるためである。なお、式 (2) で画素値の反転を行っているのは、背景部分を 0 とした方が今後の議論が容易になるためである。以下ではこの $\tilde{I}(x, y)$ を黒画素値と呼ぶことにする。

次に、文字行の抽出しを行う。文字行とは文字が読み書きされる方向、すなわち縦書きの文書であれば垂直方向、横書きの文書であれば水平方向に連続する文字列の改行までの 1 単位である。本手法は縦書き横書きいずれの文書に対しても適用可能であるが、以下で

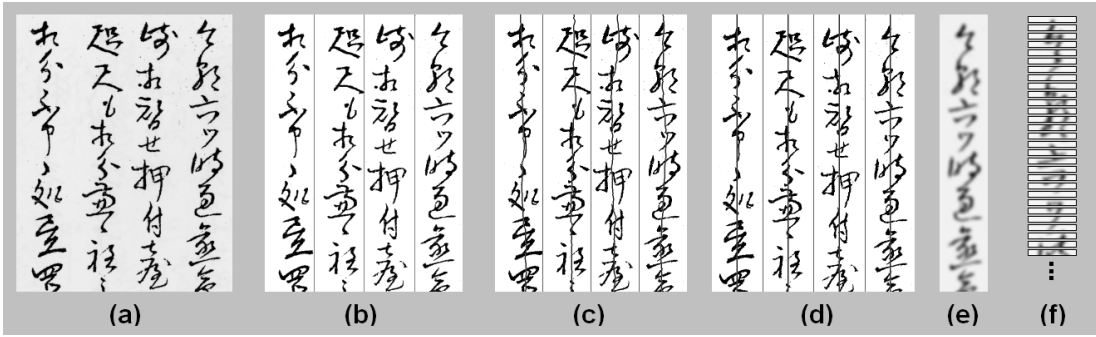


図 1 前処理及びスリット切出し。(a) 入力画像, (b) 背景除去及び文字行切出し, (c) 中心位置推定, (d) 中心位置正規化, (e) ガウス平滑化, (f) スリット切出し

Fig.1 Preprocessing and dividing into slits. (a) input image, (b) background removal and line separation, (c) center position estimation, (d) center position normalizing, (e) Gaussian smoothing, (f) dividing into slit images.

は縦書きの文書を前提に説明する。すなわち、垂直方向を文字行方向とする。なお横書きの文書を取り扱う場合ははじめに x, y 座標を入れ換えればよい。

日本語の文書は英語の文書と異なり単語の切出しが容易でないことは既に述べたが、文字行に切り出すことは比較的容易である。文字行方向の黒画素値の射影ヒストグラム

$$H(x) = \sum_y \tilde{I}(x, y) \quad (3)$$

を作成し、 $H(x)$ が極小となる x 座標を文字行と文字行との境界位置と定めることにより、文書画像を文字行単位に切り出すことができる。なお、このようにして切り出された文字行は幅が一定しないため、幅が狭い画像について左右に余白を追加することで、文字行の幅を一定にそろえておく。ここまでの処理結果の例を図 1 (b) に示す。

最後に、切り出された各文字行について中心位置の正規化を行う。けい線のない紙に書かれた文書は文字位置が左右に揺れる場合がしばしば見られる。この影響を除くため、移動平均法によって文字行の中心位置を推定し、これを正規化する。まず、各文字行画像の各垂直位置 y_0 について、前後 n ピクセルを含めた黒画素値の重心の x 座標、すなわち、

$$C_x(y_0) = \frac{\sum_x \sum_{y \in [y_0-n, y_0+n]} \tilde{I}(x, y) \times x}{\sum_x \sum_{y \in [y_0-n, y_0+n]} \tilde{I}(x, y)} \quad (4)$$

を求める (図 1 (c))。 n は大きすぎると補正による効

果が小さくなる一方で、小さすぎると文字形状自体を崩してしまう過補正が発生するため、文書画像の性質に応じて個別に定める必要がある。今回対象とする文書画像については、 n を 1 文字程度の長さを設定すると良好な結果が得られた。次に各文字行画像について、この重心位置が行の中心にそろうように水平ピクセルラインを再配置する。すなわち、

$$\hat{I}(x - x_c, y) = \tilde{I}(x - C_x(y), y) \quad (5)$$

とする。ここで x_c は正規化文字行画像の中心の x 座標を表す定数である。これにより、中心位置が正規化された文字行画像 $\hat{I}(x, y)$ が得られる (図 1 (d))。

2.2 平滑化及びスリット切出し

前処理済みの画像に対し、ノイズに対する頑健性を付与するためにガウスフィルタによる平滑化を行った後、これをスリット状に切り出す (図 1 (e) 及び (f))。このようにして画像を切り出すことにより、文字列画像をスリット画像のシーケンス (画像列) としてとらえることができるようになる。

ここでガウスフィルタによる平滑化を行う際のパラメータ σ の値及びスリット状に切り出す際の切出し幅については考慮して定める必要があるが、これらについては 3.3 及び 3.4 で検討する。

2.3 特徴量ベクトルの記述

切り出した画像列に対し、各スリット画像を低次元の特徴量ベクトルで記述することを考える。本手法では、特徴量ベクトルの記述には、主成分分析 (固有空間法) を用いる。

画像における固有空間法の最もよく知られている適

用例は顔認識におけるそれであり, Turk and Pentland [11], [12] が顔画像の集合に対して主成分分析を適用して得られた固有顔 (Eigenface) を用いることにより効率の良い顔認識が可能であることを示したのをはじめとして, 数多くの研究が行われている. ここではまず画像に対する固有空間法の適用法について, 簡単にその概要を示す.

M 枚の画像があり, 各画像は N 画素をもつものとする. 各画像に対し, その画素値を並べて N 次元列ベクトルとして表現したものを x_i とする. 各画像から平均画像 $c = (1/M) \sum x_i$ を除去し, それを並べた行列を作成し,

$$A = (x_1 - c \quad x_2 - c \quad \cdots \quad x_M - c) \quad (6)$$

とおく. これから共分散行列

$$C = AA^T \quad (7)$$

を作成し, C に対して固有値問題を解いて, 固有値と固有ベクトルを得る. これらを固有値の大きい順に並べ換え, 上位の固有ベクトルのみを基底として各画像の低次元表現を得る. すなわち, 固有ベクトルを固有値の大きい順に v_1, v_2, \dots として, 適当な次元 d までのものを順に並べた行列

$$F = (v_1 \quad v_2 \quad \cdots \quad v_d) \quad (8)$$

をつくり,

$$y_i = F^T(x_i - c) \quad (9)$$

とする.

これにより, 低次元 (ここでは d 次元) の画像の表現 y_i が得られたので対応付けの問題を解くことが容易となる.

実際には, 一般に $M \ll N$ であるため AA^T の N 次元固有値問題を直接解くことはせず, 代わりに $A^T A$ の M 次元固有値問題に帰着させてから解くという方法がとられる. すなわち, AA^T の固有ベクトル行列を V , $A^T A$ の固有ベクトル行列を U とすると $AU = VD$ が成り立つ (D は AA^T の固有値の平方根を対角成分に並べた $N \times M$ 行列) ため, U から V を導くことができる.

以上が通常の固有空間法のプロセスであるが, これをそのまま文字列画像から作成したスリット列に適用すると, 対象文書の長さに比例してスリット数 (= 画像の枚数. 前述のプロセスの M に相当) が増加し, 固

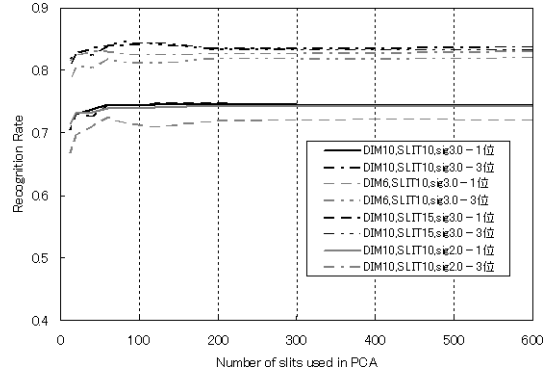


図2 基底作成に用いるスリット数と, 認識率の比較
Fig.2 Impact of various number of slits used in PCA.

有値問題の計算が極めて高コストになるという問題が発生する. しかし幸いなことに, 今回取り扱うような文字列画像のスリット列はある程度以上スリット数が増えてもほとんど固有画像が変化せず, その結果得られる特徴量ベクトルも変化しないという性質をもっている. それを実験的に確認したものが図2であり, 基底作成に用いるスリット数を変化させながら, 次章で述べるような様々な条件下での認識率の変化の様子を調べた結果が示されている. 図から, いずれの場合においても認識率は基底作成に用いるスリット数が50~100程度の早期に立ち上がり, それ以上スリット数を増やしてもほとんど認識率には影響しないことが分かる. このことから, 今後の実験においては基底作成に用いるスリット数は最大200スリットとし, これより大きい数のスリットを扱う場合は冒頭の200スリットのみから固有空間の基底を作成し, それ以降のスリットについては, こうして作成された基底を用いて特徴量ベクトルに変換することとした. 作成される固有画像の例 (実際は正負の値をもつ実数であるが, それをグレースケールに可視化したもの) を図3に示す.

2.4 特徴量ベクトルの系列による対応付け

前節により, 文字列画像を特徴量ベクトルの系列に変換する手法が得られた. この節では, これを用いて文書画像中からクエリー部分と類似度の高い部分を検出する方法について述べる.

スリット画像列の特徴量ベクトルの系列を $\{y(t)\}$ (t はスリット番号) とし, クエリー画像は $t_0 \leq t \leq t_0 + \tau$ の範囲に含まれているものとする. このとき, クエリー画像列 $A = \{y(t) \mid t_0 \leq t \leq t_0 + \tau\}$ と, t'_0 を起点とする同じ長さの画像列 $B = \{y(t) \mid t'_0 \leq t \leq t'_0 + \tau\}$

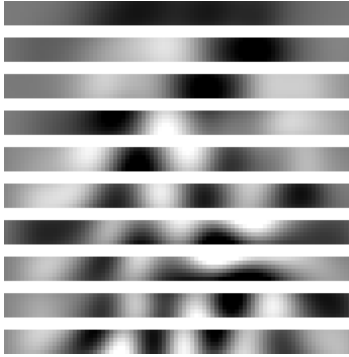


図 3 主成分分析により作成される固有画像の例 . 上から順に第 1 固有画像, ..., 第 10 固有画像
Fig. 3 Examples of eigenslits.

との間の距離を

$$D(A, B) = \sum_{0 \leq t \leq r} |y(t_0 + t) - y(t'_0 + t)| \quad (10)$$

で定め, 小さい $D(A, B)$ を与える B をクエリー画像と類似度の高い画像と定義する. ここで $|y(t_0 + t) - y(t'_0 + t)|$ は各スリットにおける特徴量ベクトル間の距離を表し, この定義の方法もいくつかの候補が考えられるが, 本研究では最も単純な L1-ノルム (マンハッタン距離)

$$\begin{aligned} & |y(t_0 + t) - y(t'_0 + t)| \\ &= \sum_i |y_i(t_0 + t) - y_i(t'_0 + t)| \end{aligned} \quad (11)$$

(y_i はベクトル y の第 i 成分を表すものとする) を採用することとした.

B の始点 t'_0 を変化させながら $D(A, B)$ を計算し, 最も類似度の高い画像を第 1 位検出画像, 以下第 2 位検出画像, 第 3 位検出画像, ... として出力することとする. これにより, 文書画像から, クエリーとする部分と類似度の高い部分を検出する方法が得られた.

2.5 Dynamic Time Warping

前節までで文字画像列から類似画像を検出する方法が得られたが, ここでは更にそれを拡張し, 文字列の縦方向の伸縮変形に対応するために DTW (dynamic time warping) を導入することを考える. DTW は主に音声認識の分野で発達した手法で, 二つの時系列信号が入力されたときに, それぞれの時間軸を非線形に変形させながら最も良い対応がとれる時間対応を探し, その時間対応のもとでの類似度を出力するものである. 本研究で取り扱う文書画像検索においても, 文字行方

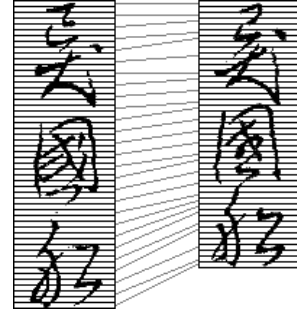


図 4 文字列画像に対する DTW の適用イメージ
Fig. 4 DTW application to text images.

向の軸 (縦軸) を時間軸とみなすことによりスリットに分割された文字画像を時系列信号とみなすことができ, DTW を適用することが可能となる (図 4). 以下では DTW の概要と, 文字画像に対する適用法について述べる.

時系列信号 $A = \{y(t) \mid \alpha_1 \leq t \leq \alpha_n\}$ と $B = \{y(t) \mid \beta_1 \leq t \leq \beta_m\}$ に対し, DTW により時間伸縮を調整した距離 $D(A, B)$ を次のように定義する.

$$D(A, B) = \min \left[\frac{\sum_{\theta=1}^k |y(i_\theta) - y(j_\theta)|}{k} \right] \quad (12)$$

ここで $(i_1, j_1), \dots, (i_k, j_k)$ は対応付けの経路を表し,

$$\begin{aligned} (i_1, j_1) &= (\alpha_1, \beta_1) \\ (i_k, j_k) &= (\alpha_n, \beta_m) \\ (i_\theta, j_\theta) &= \begin{cases} (i_{\theta-1} + 1, j_{\theta-1}) \\ (i_{\theta-1}, j_{\theta-1} + 1) \\ \text{or } (i_{\theta-1} + 1, j_{\theta-1} + 1) \end{cases} \end{aligned} \quad (13)$$

を満たすものとする. k は経路長を表す. 式 (12) における \min の算出は, あらゆる可能な経路の中で最小のものを求める. 可能な経路としては上式を満たす限り無限の伸縮を許容するというものでは必ずしもなく, ある一定の範囲に収まるもののみを考える場合が普通である. 本研究では, 経路は常に次式

$$1/\alpha (i_\theta - i_1) \leq j_\theta - j_1 \leq \alpha (i_\theta - i_1) \quad (14)$$

を満たすものという制約を課した. ここで α は伸縮比を表す. 以下の検証では $\alpha = 1.2$ と設定して実験を行う.

3. 最適パラメータの決定

前章で導入した手順を実際の実装する場合、いくつかのパラメータを決定することが必要となる。この章ではそれらについて検討を行う。

3.1 評価手法

最適なパラメータを適切に検討するためには、システムの性能に関する定量的な指標が必要である。ここではその指標として、王羲之「蘭亭序」の2通りの写本(図5)に対し、写本Aのある部分をクエリーとして写本Bから同一の部分が検出できるか否かを調べ、正しく検出できた割合を認識率として性能評価の指標に用いる。「蘭亭序」の2通りの写本はいずれも28行からなり、321文字が含まれる。これを1文字当りの解像度をおよそ 200×200 ピクセル程度でスキャンした画像を基礎に、それを人工的に様々なレベルに低解像度化したものを対象に、検証を行う。

なお、このサンプルデータに対しては前処理の背景除去の段階で文字列に重なっている印影を画像から除去することができなかったが、これはノイズとしてそのまま残して実験を行った。したがってこのタスクはやや難しい部類のタスクであるといえる。

固有空間及び基底の作成にあたっては、写本Aの冒頭200スリットのみを用い、ここで得た基底を全画像に対して適用して特徴量ベクトルを作成した。その後、写本Aの任意の連続する440ピクセル分の領域(約2文字程度の長さ)に相当)をクエリーとして写本

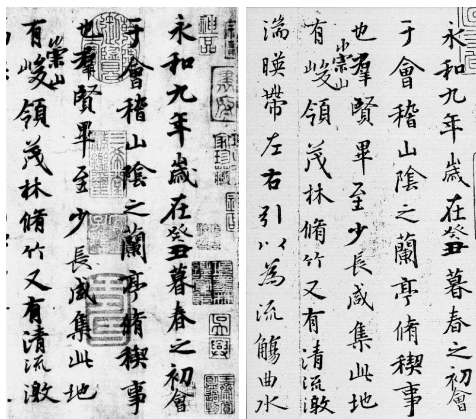


図5 評価実験に用いる画像(部分):王羲之「蘭亭序」より(A:神龍半印本,B:張金界奴本)

Fig. 5 Tested images: "Ranteijo" written by Ogishi.

Bに対して検索を行い、対応部分が1位に検出される割合(1位認識率)及び3位以内に検出される割合(3位認識率)を調べた。ここで検出画像がクエリー画像に対する対応画像であると判定する条件は、検出画像とクエリー画像が50%以上オーバーラップしていることとした。

なお、この章の評価は各種パラメータの最適値を推定する目的であるため、ここではDTWを適用せず、通常のコサイン類似度による認識率を基準に評価を行った。

3.2 固有空間の次元の決定

まず、固有空間法で低次元特徴量ベクトルを作成する際の固有空間の次元数を決定するため、特徴量の次元を変えながら認識率を確認する実験を行った。ここでは文字解像度を60ピクセル、スリット数を10(すなわちスリット幅6ピクセル)とした。その結果を図6に示す。図の横軸は、特徴量記述に用いた固有空間の次元数を表す。また、図7には固有空間の次元ごとの寄与率を示している。

図6から、提案手法による検索は部分空間の次元数が10次元程度までの間は次元数につれて増加し、それ以上次元を増やしても認識率の向上には限界があることが分かる。したがってここでは次元 $d=10$ を採用することとし、これ以降の評価はこれに基づいて行う。

3.3 解像度とスリット幅の決定

次に、文字列検索における画像の解像度及びスリット切出しの際のスリット幅の影響を確認するため、解像度とスリット幅を変えながら認識率の変化を調べる実験を行った。

解像度については1文字当り200ピクセルの原画像をソフトウェア的に10~50%に縮小し、1文字当り20~100ピクセルの画像を合成した。スリット幅は、1

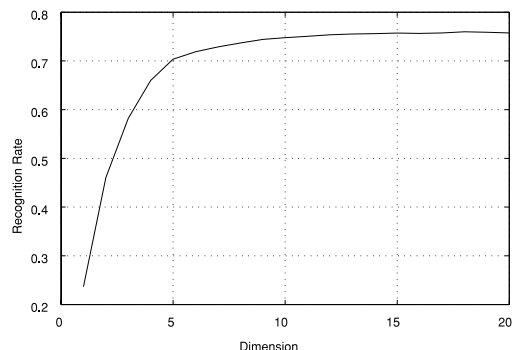


図6 固有空間の次元と認識率

Fig. 6 Recognition rate vs. dimensionality.

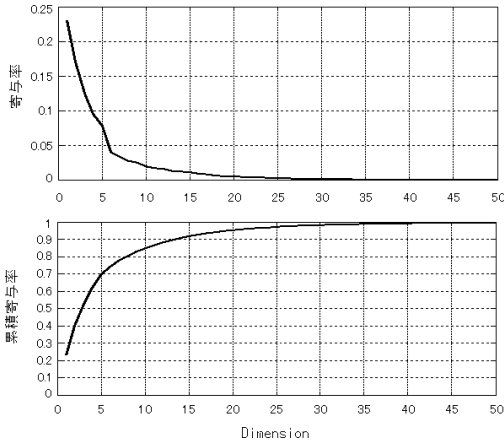


図 7 固有空間の次元ごとの寄与率及び累積寄与率
Fig. 7 Eigenvalue proportions and cumulative proportions.

文字当りのスリット数が 4~20 で、かつスリット幅 2 ピクセル以上のケースを試した。

結果を図 8 に示す。等高線図において色の濃い部分が認識率の高い部分である。図の横軸は 1 文字当りの解像度、縦軸は 1 文字当りのスリット数を表す。文字解像度が 60 ピクセルでスリット数が 10 の場合、スリット幅が 6 ピクセルであることを意味している。また、図 8 には等高線図を解像度 60 ピクセルで切断した断面と、スリット数 10 で切断した断面も併せて示した。

これらの図から、解像度は 1 文字当り 60 ピクセル程度、スリット数は 1 文字当り 10 スリット程度とるとほぼ認識率は最大に到達し、それ以上解像度を上げたりスリット幅を細かくしたりしても、効果は限定的であることが分かる。

3.4 ガウス関数の分散の決定

ここではガウスフィルタを適用する際の σ の値に関する実験を行う。前節の結果を踏まえスリット数は 10 で固定し、文字解像度を前節同様に変えながら、 σ に対する認識率の変化を見た。その結果を示したものが図 9 である。

図から、解像度が高くなるほど、最適な σ の値が大きくなることが分かる。これはスケールスペースの理論から解釈しても妥当な結果である。

具体的には、

$$\sigma = \frac{1 \text{ 文字当りの解像度}}{20} \quad (15)$$

と設定すると、おおむね最適な結果を得ることができ

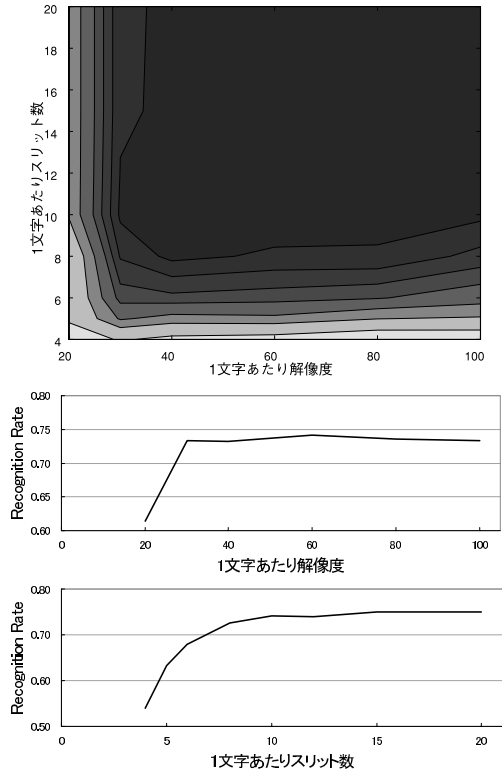


図 8 解像度とスリット数に対する 1 位認識率の等高線図、及びその [スリット数=10] における断面図と [解像度=60] における断面図

Fig. 8 Contour plot of recognition rate by resolutions and number of slits. The cross section is plotted below the contour plot.

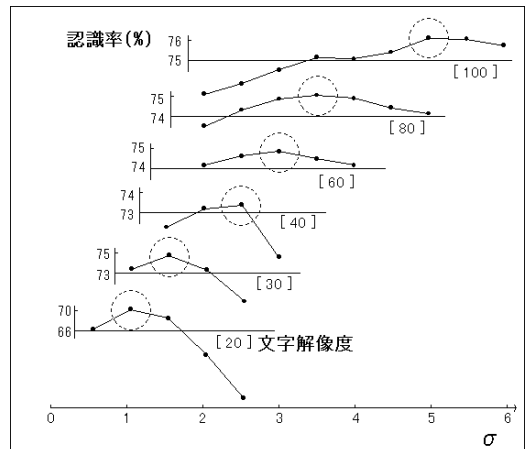


図 9 解像度別のガウス関数の σ の値と 1 位認識率の関係を文字解像度ごとに示したものの。図中破線で囲まれた部分は、その解像度における最大認識率

Fig. 9 Recognition rate vs. Gaussian parameter for various resolutions.

るといえる。

3.5 考 察

以上から、1文字当りの解像度を60~80ピクセル程度、スリット数は1文字当りで10スリット(すなわち幅は6~8ピクセル)とし、ガウスパラメータは式(15)で定めると良好な結果が得られることが分かる。

1文字当り60から80ピクセル程度という解像度を説明するために具体的に例示すると、これは郵便番号枠(幅5.7mm)いっぱいの文字を300dpiでスキャンした程度に相当する。実際にデジタルアーカイブの構築を考える際にも、この程度の解像度を得ることは特に問題ない水準であるといえるだろう。

更に、仮にこれ以下の解像度の画像データしか得られない場合も、単純に画像を拡大して1文字当りのピクセル数を60~80程度まで増やすことにより認識率をある程度向上させることができる。表1はそれを示したもので、解像度200ピクセルの原画像を縮小して解像度を20ピクセルとした画像と、それに拡大処理を施して解像度を60または80ピクセルとしたものの認識率を比較したものである。ここにおける拡大処理とは一切の補間をせず、300%拡大であれば同じ画素値のピクセルを3×3個並べるという極めて単純なものである。表1から、このような単純な処理であるにもかかわらず、解像度を60または80ピクセルまで拡大することにより、解像度20ピクセルのままでは検索を行うのに比べて認識率が若干向上していることが分かる。

低解像度画像を単純に拡大するだけで、情報量としては変化していないにもかかわらず、認識率の向上が得られる理由は次のように考えられる。顔認識の研究などでも広く知られているとおり、主成分分析を用いて画像の照合を行う方法においては、位置合せの正確さが、照合の精度に大きく影響する。本手法においては前処理の中心位置正規化のステップで位置合せを行うとともに、主成分分析を適用する前にガウスフィルタで平滑化を行うことにより、位置ずれの影響を低減させている。低解像度の画像に対して拡大処理を行った後にこれらの手続きを適用するということは、原

像に対して計算領域をサブピクセル領域まで拡張した上でこれらの手続きを適用することと等価である。この計算領域拡張の効果で、より正確な位置合せ、より適切な平滑化が行われ、その結果として認識率の改善効果が得られるものと考えられる。

4. 実験結果

4.1 実験 1

前章で得られたパラメータを実際に用いて、改めて王羲之「蘭亭序」の認識率を調べる実験を行った。固有空間は10次元、解像度は1文字当り80ピクセル(解像度200の原画像を40%に縮小)、スリット数は1文字当り10(すなわちスリット幅8ピクセル)、ガウス関数の $\sigma = 4.0$ と設定し、任意の連続する約2文字程度の長さの領域(原寸で440ピクセル、縮小時で176ピクセル)をクエリーとして、DTWを適用した場合と適用しない場合の両方について、1位認識率及び3位認識率を調べた。その結果が表2である。表から、DTWの導入により認識率が向上していることが分かる。この効果は、次節の実験で再度検証する。

4.2 実験 2

これまでの実験は王羲之「蘭亭序」の2種類の写本に対して行ってきたが、本手法がこの文献に限って適用可能なものではなく一般性をもつ手法であることを示し、またワードスポッティングに対して有効であることを示すために、「亜国来使記」(図10)に対し、キーワード画像を指定して全文に対して検索を行う実験を行った。

「亜国来使記」は安政元年(1854年)に書かれた松前藩家老・松前勘解由の日記であり、その全文は182ページ、1553行、25148文字からなる。これから、文中に25回以上登場する人名(表3)を対象に、再現率と適合率を評価した。なお、人名抽出にあたっては田畑[13]による翻刻結果を使用し、各キーワードの画像の該当領域への対応付けは、手作業で個別に割り当てるという方法により行った。これらを用いて各キーワードの各画像をクエリーとして全文に対する検索を行い、検出される画像領域が正解領域と誤差が約2文

表1 低解像度画像の拡大による認識率の改善効果
Table 1 Improvement by image enlargement.

解像度	1位認識率(%)	3位認識率(%)
20(原寸)	70.02	79.93
60(300%拡大)	71.67	82.45
80(400%拡大)	71.37	82.30

表2 「蘭亭序」に対する認識率
Table 2 Recognition rate for "Ranteijo."

計算条件	1位認識率(%)	3位認識率(%)
DTWなし	75.59	84.18
DTWあり	78.10	84.43

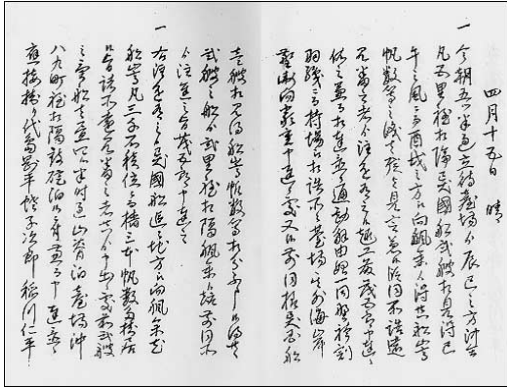


図 10 「亜国来使記」: 安政元年 (1854 年) に書かれた幕府役人の日記
 Fig. 10 “Akoku Raishiki” written in the mid-19th century.

表 3 実験対象としたキーワード
 Table 3 Keywords used in the experiment.

キーワード	出現回数
又左衛門	165
ウリヤムス	73
井上富左右	25
石塚官蔵	25

字分 (20 スリット) 以内の位置に検出されればその検出を適合と判断するという基準で評価した。精度評価尺度には、Rath [5] と同じく平均適合率 (average precision) を用いた。平均適合率は情報検索の評価において頻りに用いられる指標であり、あるクエリーに対しデータベース内から検索を行ったときに作成される順位リストを上から順に調べ、適合するデータが出現した時点までの適合率を順次計算し、得られたすべての適合率の平均値として定義され、その値は再現率-適合率曲線の下の部分の領域の面積に相当する。

実験対象としたキーワードの各クエリー画像に対し平均適合率を算出し、それを更にキーワードごとに平均値を算出した結果をまとめたものが表 4 である。また、参考のため、このうちキーワード「又左衛門」をクエリーとした場合の一部 (DTW ありの場合で平均適合率が 165 個所中 20 位, 40 位, …, 160 位となった計 8 個所) についての再現率-適合率曲線を図 11 に示した。同様の評価基準で実験を行った Rath [5] において George Washington collection に対する結果が良質な画像に対してで 40.98%, 劣化の著しい画像に対してで 16.50%と報告されていることを考えると、実験対象が違うので単純な比較はできないが、本手法

表 4 「亜国来使記」に対する平均適合率 (%)
 Table 4 Average-precision for “Akoku Raishiki.”

キーワード	DTW なし	DTW あり
又左衛門	66.09	79.39
ウリヤムス	79.77	93.68
井上富左右	60.32	83.50
石塚官蔵	56.88	73.75

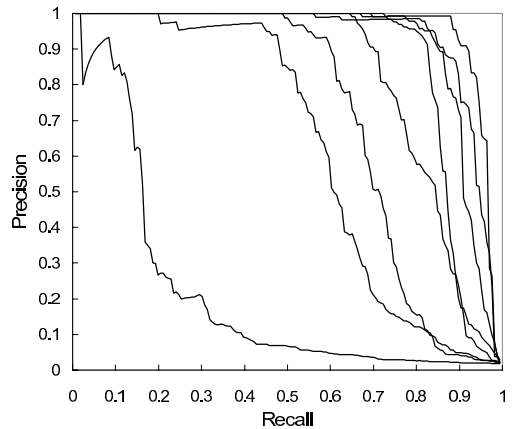


図 11 「又左衛門」に関する再現率-適合率曲線 (一部)
 Fig. 11 Recall-precision curve for keyword “Matazaemon.”

による検索の精度は一定の有効性を示す水準に達していると評価することができる。またここでも、DTW を適用することが検索の精度を高めているということが確認できる。

また、表 4 の内訳として、キーワード「又左衛門」のすべての出現個所 (165 個所) について、その出現画像をクエリーとした際の平均適合率を図示したものが図 12 である。縦軸は平均適合率であり、横軸はクエリーの ID であるが、ここではグラフを見やすくするため、DTW ありの場合の平均適合率の順にソートして ID を付け直している。図より、ほぼすべてのクエリーについて DTW を導入することで平均適合率が向上していることが分かる。実際、DTW により平均適合率が低下したのは 165 個所中 4 個所のみであり、そのいずれもが平均適合率が低いところ、すなわちクエリー画像の性質自体が良くないと思われるところに集中している。

また、図 12 からは、165 個所のクエリーに対する精度は平均値の周りに均等に分布しているのではなく、平均以上の精度をもつクエリー画像が過半数を大きく上回る一方で、極端に精度が悪いデータが数個存在して、平均を押し下げていることが分かる。特に DTW

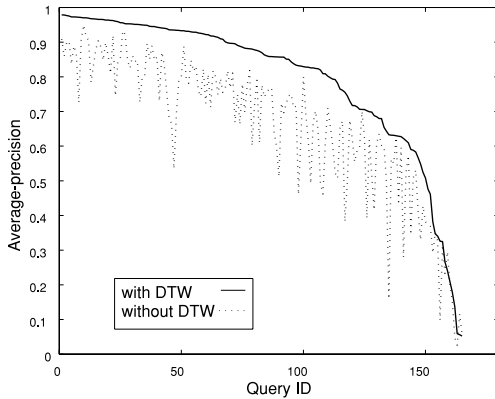


図 12 又左衛門 165 種の平均適合率
Fig. 12 Average-precision for keyword
“Matazaemon.”

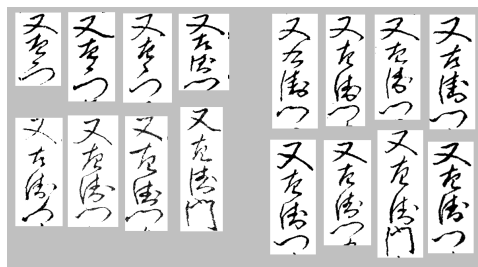


図 13 極端に精度の悪いクエリー画像 (左) 及び平均的に良好なクエリー画像 (右)
Fig. 13 Query images of low precision (left) and good precision (right).

ありの場合については 165 個所のうち 111 個所において平均適合率が 80%以上である一方、平均適合率が 50%を下回る個所が 15 個所存在した。これらのクエリー画像による極端な精度の違いの原因を確認するため、精度が極端に悪いデータと比較的良好なデータのサンプルを分けて図示したものが図 13 である。これらを観察すると、書体の変化が極めて大きい場合(「又左衛門」の「衛」の字)や、文字列の伸縮が極めて大きい場合、あるいは前処理におけるしきい値処理の問題で文字が全体にかすれた場合などにおいて精度が大きく下がっていることが分かる。その一方で、その他の変形に対してはおおむね良好に対応できていることが分かる。

5. む す び

本論文では、毛筆手書きで書かれた文書画像に対し、ある文字列の画像領域をクエリーとして与えて、それと類似度の高い画像領域を検索することにより対応

する文字列を検出する方法について述べた。画像をスリット状に切り出した上で固有空間法を適用することにより、文字列画像検索はシーケンスのマッチング問題として解くことが可能となり、また DTW を導入することにより、検索の精度が高まることが確認された。この手法は文字切出しが極めて困難な崩し字書体に対しても適用可能であり、毛筆手書き文字からのキーワード検索に対して高い精度を得ることができることを示した。

今後は、本手法に適した前処理についての研究を進めて手法の更なる高精度化を図るとともに、本手法を更に拡張し、同一人物が同一筆跡で文書を書いた場合にとどまらず、異なる筆跡の間で検索を行うことへの適用可能性などについて研究を進めていく予定である。

文 献

- [1] 山田奨治, 柴山 守, “古文書を対象にした文字認識の研究,” 情報処理, vol.43, no.9, pp.950-955, Sept. 2002.
- [2] 坪井昭憲, 八村広三郎, 吉村ミツ, “江戸期版本画像からの文字切り出しの試み,” 情処学研報, no.2005-CH-66, 2005.
- [3] R. Manmatha, C. Han, and E.M. Riseman, “Word spotting: A new approach to indexing handwriting,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp.631-637, 1996.
- [4] T.M. Rath and R. Manmatha, “Features for word spotting in historical manuscripts,” Proc. Int. Conf. on Document Analysis and Recognition, pp.218-222, 2003.
- [5] T.M. Rath and R. Manmatha, “Word image matching using dynamic time warping,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp.521-527, 2003.
- [6] S. Marinai, E. Marino, and G. Soda, “Indexing and retrieval of words in old documents,” Proc. Int. Conf. on Document Analysis and Recognition, pp.223-227, 2003.
- [7] Y. Lu and C.L. Tan, “Word spotting in Chinese document images without layout analysis,” Proc. IEEE Int. Conf. on Pattern Recognition, pp.30057-30060, 2002.
- [8] 近藤博人, 松本隆一, 柴山 守, 山田奨治, 荒木義彦, “文字切出しを前提としない古文書標題認識,” 情処学研報, no.2003-CH-57, 2003.
- [9] K. Terasawa, T. Nagasaki, and T. Kawashima, “Eigenspace method for text retrieval in historical document images,” Proc. Int. Conf. on Document Analysis and Recognition, pp.437-441, 2005.
- [10] 寺沢憲吾, 長崎 健, 川嶋稔夫, “古文書を対象としたワードスポッティング,” 画像の認識・理解シンポジウム MIRU2005, pp.522-529, 2005.
- [11] M.A. Turk and A.P. Pentland, “Eigenfaces for recog-

niton,” Journal of Cognitive Neuroscience, vol.3, no.1, pp.71–86, 1991.

- [12] M.A. Turk and A.P. Pentland, “Face recognition using eigenfaces,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp.586–591, 1991.
- [13] 馬場 修, 田畑幸三郎, “安政元年箱館湊日米応接日記 ; 市立函館図書館所蔵 (非売品), 函館, 1972.
(平成 17 年 7 月 14 日受付, 18 年 1 月 17 日再受付)



寺沢 憲吾 (正員)

平 10 東大・工・土木卒 . 平 12 同大大学院修士課程了 . 平 18 公立はこだて未来大学大学院博士後期課程了 . 同年北大博士研究員 . 画像情報検索に関する研究に従事 .



長崎 健 (正員)

平 4 北大・工・情報工卒 . 平 6 同大大学院修士課程了 . 平 9 同大学院博士課程単位取得退学 . 平 9 (株) ビュージー入社 . 平 6~9 まで日本学術振興会特別研究員 . 平 12 よりはこだて未来大講師 . 平 18 同助教授 . コンピュータビジョンの研究に従事 . 日本ロボット学会, 情報処理学会各会員 .



川嶋 稔夫 (正員)

昭 55 北大・工・電子卒 . 昭 57 同大大学院修士課程了 . 同年苫小牧高専助手 . 昭 59 北大工学部助手 . 平 2 同講師 . 平 4 同助教授 . 平 12 はこだて未来大教授となり現在に至る . コンピュータビジョン, センサフュージョンに関する研究に従事 . 工博 . 情報処理学会会員 .