

低品質文字列を認識するための文字間空隙特徴の利用

Recognition of Character Strings in Low-quality Images Using Inter-character Features

石田 皓之*
Hiroyuki Ishida

高橋 友和*
Tomokazu Takahashi

井手 一郎*
Ichiro Ide

村瀬 洋*
Hiroshi Murase

1 まえがき

カメラ付き携帯機器の普及に伴い、それらの機器を入力とする文書認識技術が注目を集めている [1]。しかし、ぼけやモーションブラー、解像度低下などの影響により隣接する文字同士が接触するという問題があり、文字列を文字単位に分割することは容易ではない。低品質画像中の文字の切り出し手法は多く提案されており [2]、主要なアプローチの一つに認識に基づく分類 (recognition-based segmentation) 手法 [3, 4] がある。ただし、認識と分割のいずれも独立の処理とすることは容易でなく、いずれかの処理を先に行おうとすると、双方の処理間の依存性に影響をきたす。例えば、個々の文字認識の結果をもとに分割とすると、文字の境界が曖昧となり、過分割をもたらす恐れがある。

本稿の手法では、分割を個々の文字認識の副産物として扱うのではなく、分割と認識の相互依存性を考慮する。そのため、個別文字の類似度と分割の適切度を統合し、それらの線形和を指標として評価する。また、分割の適切度の指標として、文字間空隙特徴を利用する方法を提案する。

2 文字間空隙の特徴

文字間空隙とは、図 1 に示されるように、隣接する 2 文字間の領域である。分割の適切度として、文字間空隙がどの 2 文字間に属しているかを測る。左側の文字を l 、右側の文字を r とすると、低解像度の場合は以下の特徴が観測されるはずである。

- (i) 空隙の左半分は文字 l の右端と類似している。
- (ii) 空隙の右半分は文字 r の左端と類似している。
- (iii) 空隙領域を左から右に進むに従い、文字 l の特徴から文字 r の特徴へとしだいに化する。

特徴 (iii) は、隣接する 2 文字が各々ぼやけることにより、双方のグレースケール値が加算されて生じる現象である。以上の特徴 (i)–(iii) を満たしたとき、その空隙は l – r 間にあるものと認識する。本稿の手法では、直交部分空間法 [5, 6] と類似したアプローチにより文字間空隙を認識する。

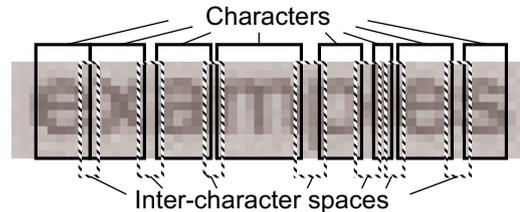


図 1 文字と文字間空隙。文字間空隙領域は、文字領域と 1 列分オーバーラップしている。

2.1 評価尺度

本手法では、空隙領域を列単位に分解し、それらの列ベクトルが文字間直交部分空間上で張る三角形の面積和を評価尺度として利用する。ここで文字間直交部分空間とは、隣接する 2 文字の組ごとに作成される 2 次元の部分空間である。その基底は左側文字 l の右端の特徴ベクトル $\phi^{(l)}$ と右側文字 r の左端の特徴ベクトル $\phi^{(r)}$ を直交化したものとなっている。本節冒頭で述べた特徴 (i), (ii) は、文字間直交部分空間上に投影したベクトルの長さに対応するので、これらをそれぞれ OA, OB と表す。さらに特徴 (iii) は、文字間直交部分空間上に投影したベクトル同士の開きを表しているため、これを $\sin \angle AOB$ と表す。これらの評価値の積は

$$OA \times OB \times \sin \angle AOB \quad (1)$$

となる。これは、ベクトルがなす面積の定義と一致する。よって、面積を評価尺度として扱う。

2.2 文字間直交部分空間の作成

左側の文字が l 、右側の文字が r である文字間空隙の直交部分空間を $\mathcal{I}^{(l)(r)}$ と表し、 $\mathcal{I}^{(l)(r)}$ への射影行列を $\mathbf{W}^{(l)(r)}$ とする。 $\mathbf{W}^{(l)(r)}$ は以下の処理によって構成される。

まず各カテゴリ c につき、多数の学習サンプルの平均をとることにより、左端 $\phi^{(c|)}$ と右端 $\phi^{(c)}$ のグレースケール特徴のベクトルを求める。これらのベクトルは、要素平均が 0、ノルムが 1 となるように正規化するものとする。 $\mathbf{W}^{(l)(r)}$ を構成するためには、 $\phi^{(l)}$ と $\phi^{(r)}$ を用い、以下の相関行列 P を計算する。

$$P = \frac{1}{2} \left(\phi^{(l)} \phi^{(l)\top} + \phi^{(r)} \phi^{(r)\top} \right) \quad (2)$$

* 名古屋大学大学院 情報科学研究科
Graduate School of Information Science, Nagoya University

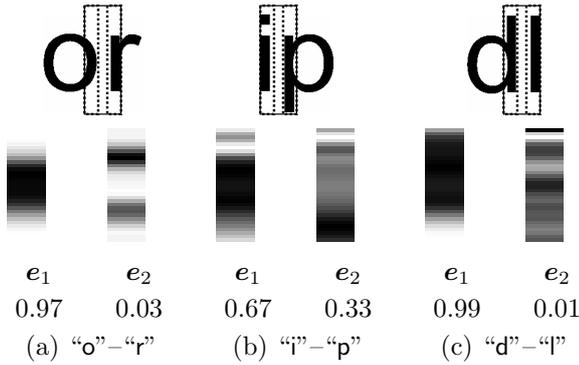


図2 文字間部分空間の固有ベクトル．下の数字は固有値を示す．

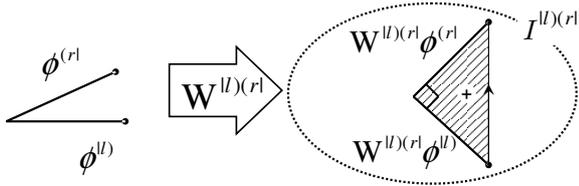


図3 部分空間 $\mathcal{I}^{(l)(r)}$ 上で正規化された学習ベクトル

この P の固有値と固有ベクトルを第2位まで求め、それぞれ $(\lambda_1, \lambda_2), (e_1, e_2)$ とする．固有ベクトルの例を図2に図示する．射影行列 $W^{(l)(r)}$ は以下によって求められる．

$$W^{(l)(r)} = \frac{1}{\sqrt{2}} \Lambda^{-1/2} B^T \quad (3)$$

$$\Lambda^{-1/2} = \text{diag}(\lambda_1^{-1/2}, \lambda_2^{-1/2}) \quad (4)$$

$$B = [e_1 \ e_2] \quad (5)$$

この操作により、以下の性質が満たされる．

$$\left| W^{(l)(r)} \phi^{(l)} \right| = \left| W^{(l)(r)} \phi^{(r)} \right| = 1 \quad (6)$$

$$W^{(l)(r)} \phi^{(l)} \perp W^{(l)(r)} \phi^{(r)} \quad (7)$$

これは、行列 $W^{(l)(r)}$ によって特徴ベクトル $\phi^{(l)}, \phi^{(r)}$ が直交化されたことを意味する．さらに、行列式が以下のように正になることを保障するものとする．

$$\det \begin{vmatrix} W^{(l)(r)} \phi^{(l)} & W^{(l)(r)} \phi^{(r)} \end{vmatrix} > 0 \quad (8)$$

これは、ベクトル $W^{(l)(r)} \phi^{(l)}, W^{(l)(r)} \phi^{(r)}$ がなす面積が正となることを意味する．もしこれが成り立たない場合には、ベクトル e_2 の符号を反転させ、上式を満たすように手で修正する．以上の処理を図示したものを図3に示す．

2.3 文字間空隙の認識

文字列画像中で n 列目から m 列目までを占める文字間空隙領域に対し、文字間直交部分空間 $\mathcal{I}^{(l)(r)}$ との類似度を計算する方法を述べる．

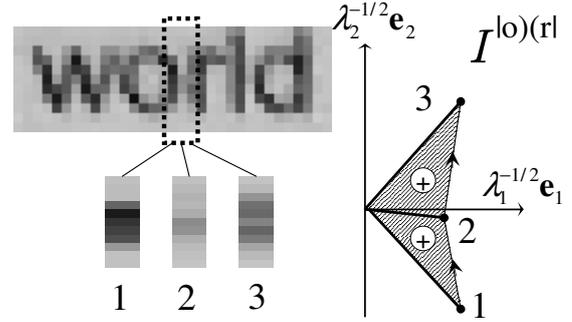


図4 類似度計算の図．文字間直交部分空間への類似度は投影ベクトルがなす面積和として評価される．

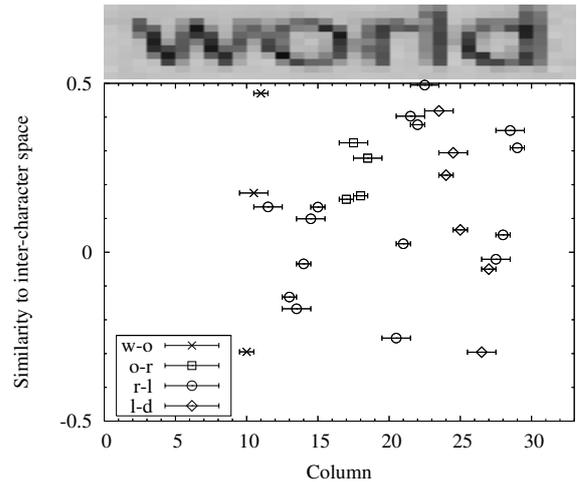


図5 文字間空隙の文字間直交部分空間に対する類似度．文字間直交部分空間 “w”-“o”, “o”-“r”, “r”-“l”, “l”-“d” に対する類似度を横棒でプロットした．

i 列目 ($n \leq i \leq m$) のグレースケール値を並べたベクトルを y_i とし、要素の平均値が0、ノルムが1となるように正規化する．これらのベクトルを $\mathcal{I}^{(l)(r)}$ に投影し、ベクトル群 $W^{(l)(r)} y_i$ を得る．文字間空隙 ($n \leq i \leq m$) と $\mathcal{I}^{(l)(r)}$ との類似度は、これらのベクトルがなす三角形の面積の和となる．したがって、以下の式で表される．

$$s_{(n,m)}^{(l)(r)} = \frac{1}{2} \sum_{i=n}^{m-1} \det \begin{vmatrix} W^{(l)(r)} y_i & W^{(l)(r)} y_{i+1} \end{vmatrix} \quad (9)$$

図4にて類似度計算の例を示す．図中左の3列からなる文字間空隙と、“o”-“r”間の文字間直交部分空間との類似度を面積計算によって求めている．このような類似度を文字列画像中の多数の文字間空隙に対して求めた結果を図5に示す．式(9)の評価値は、セグメンテーションすべき場所で高い値を示している．

2.4 固有ベクトルが2つ存在しない場合

これまで、第2固有ベクトルが存在する場合の処理を述べた．しかし、 $\phi^{(l)}$ と $\phi^{(r)}$ が非常に類似している場合には、固有ベクトルが1つしか得られないため、以上で

述べた処理を適用できない．図2の(c)がその例を示している．そのような組に対しては，過分割を避けるために0の類似度を与えるものとする．

$$s_{(n,m)}^{l(r)} = 0 \quad (\lambda_2 < \epsilon) \quad (10)$$

ここで， ϵ は十分小さな値とする．

3 文字列画像の認識

文字列の認識のため，本稿では候補文字ラティス法[4, 7]を用いて検証する．孤立文字の類似度評価は，低品質文字に対しても有効性が確認されている[8]部分空間法[5]による．従来法[4, 7]では，孤立文字評価によって得られた候補文字群から図6に示すような仮説グラフ(hypothesis graph)[9]を構築し，動的計画法によって文字列を認識している．本手法では，文字間空隙の類似度を併用することにより，認識率向上を図る．

3.1 部分空間法による個々の文字の認識

部分空間法による文字の類似度評価方法の概略を述べる．文字列画像中の m 列目から n 列目までを占める文字領域をベクトル化し， $z_{(m,n)}$ と表す．このベクトルは要素平均が0，ノルムが1となるように正規化するものとする．各カテゴリ c の部分空間 $S^{(c)}$ を学習データから主成分分析(PCA)により作成しておき，その固有ベクトルを $e_r^{(c)} \in S^{(c)}$ と表す．入力文字領域 $z_{(m,n)}$ とカテゴリ c との類似度は以下の式で与えられる．

$$s_{(m,n)}^{(c)} = \sum_{r=1}^R \left(e_r^{(c)\top} z_{(m,n)} \right)^2 \quad (11)$$

R は使用する固有ベクトルの次元数である．各領域 (m,n) につき，類似度が上位を占める文字が候補文字となる．文字列の横幅を W とすると，任意の m,n ($1 \leq m < n \leq W$)に対し候補文字を求める．

3.2 類似度の統合

仮説グラフ中の候補文字が決まれば，それらを結ぶ経路の中で類似度和が最大となるものを探索することによって，文字列の認識結果が得られる．

経路中で j 番目 ($1 \leq j \leq J$)の文字カテゴリを c_j とし，その左端の列番号を m_j ，右端の列番号を n_j とする． J はその経路における文字数である．ここで，

$$m_1 < n_1 < m_2 < n_2 < \dots < m_J < n_J \quad (12)$$

である．文字の類似度和 S_1 は以下の式で定義する．

$$S_1 = \sum_{j=1}^J (n_j - m_j + 1) s_{(m_j, n_j)}^{(c_j)} \quad (13)$$

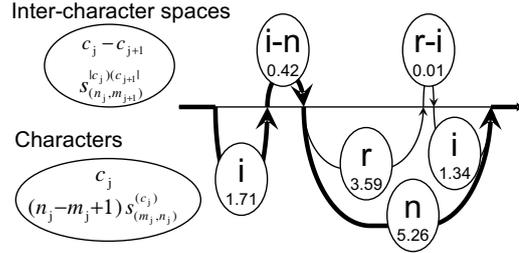


図6 入力文字“in”に対して構成した仮説グラフ．グラフ下側の円は候補文字とその類似度を示す．グラフ上側の円は両端の候補文字の組に対する文字間空隙の類似度を示す．

これは，各文字の類似度 ($0 \leq s_{(m_j, n_j)}^{(c_j)} \leq 1$) を，文字領域の横幅で重み付けた和となっている．一方，文字間空隙の類似度和 S_2 は以下の式で定義する．

$$S_2 = (n_J - m_1 + 1) \sum_{j=1}^{J-1} \left[s_{(n_j, m_{j+1})}^{(c_j)(c_{j+1})} - 1 \right] \quad (14)$$

この S_2 は負となるため，文字間空隙特徴は不適切な分割を抑制するためのペナルティとして作用する．これらを統合した類似度 S は重み k を用いて次式で与えられる．

$$S = S_1 + kS_2. \quad (15)$$

認識結果の文字列 (c_1, c_2, \dots, c_J) は，この S を最大とする組が採用される．重み k は，実験的に定める必要がある．

4 実験

本節にて，提案手法の有効性を評価する．デジタルカメラ(DMC-FX9)を用い，233単語が印刷された用紙を20回にわたって撮影し，合計4,660単語のテストデータを収集した．カテゴリ数は英数字62文字(A-Z, a-z, 0-9: Ariel フォント)とした．テストデータの文字列画像の縦幅は12 pixelとなった．

学習時においては，CG画像である文字テンプレートから学習データを生成する[10]．切り出し時の位置ずれに対応するため，各カテゴリにつき，文字を水平方向および垂直方向に移動させた625個の学習データ(32×32 pixel)を生成した，これらの学習データから，2.1で述べた文字の左端 $\phi^{(c)}$ と右端 $\phi^{(c)}$ のグレースケール特徴，3.1節で述べた固有ベクトル $e_r^{(c)}$ ($1 \leq r \leq R = 5$)を求め，文字間空隙の直交部分空間 $\mathcal{I}^{l(r)}$ と文字の部分空間 $S^{(c)}$ を構成した．なお，式(10)における ϵ は0.02とした．

様々な k に対し認識精度を求めた結果を図7に，認識結果の文字列の例を図8に示す．認識精度の指標として，

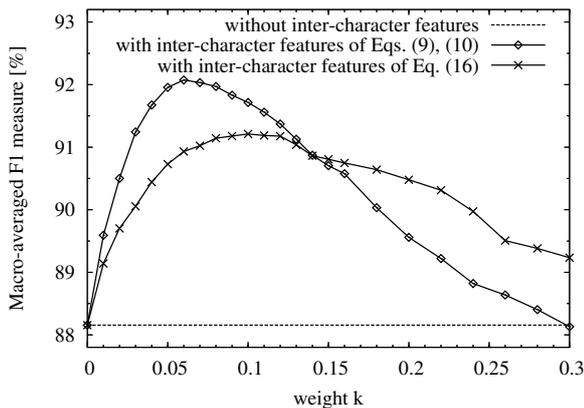


図7 認識精度のグラフ。

文献[11]にて紹介されているマクロ平均 F_1 値を用いる。これは、各テスト文字列に対し、含まれる文字の適合率 p と再現率 r から F_1 値 ($F_1 = 2pr/(p+r)$) を計算し、全テスト文字列にわたって平均したものである。結果によると、 $k < 0.06$ の範囲では、 k を増加させるごとに認識精度の改善がみられた。これは、文字間空隙特徴を利用することで、誤ったセグメンテーションが抑制されたためと考えられる。しかし、 $k > 0.06$ の範囲では認識精度が低下し、 $k \geq 0.3$ では文字間空隙特徴を使わない場合 ($k = 0$) の精度を下回った。これは、文字間空隙特徴の信頼度が文字特徴ほど高くないことを示唆している。以上より、文字間空隙特徴の利用は有効であるが、文字の類似度と組み合わせるときの比率を適切に設定する必要があることが明らかになった。

比較手法として、文字間特徴の類似度を文字間直交部分空間上の面積ではなく、単純に $\phi^{(l)}$ および $\phi^{(r)}$ との正規化相関によって測った場合の精度も評価した。この場合、2.における特徴 (i), (ii) のみを評価していることになる。類似度の評価式は次式で表される。

$$s_{(n,m)}^{(l)(r)} = \frac{1}{2} \left(\phi^{(l)\top} \mathbf{y}_n \right) \left(\phi^{(r)\top} \mathbf{y}_m \right) \quad (16)$$

結果を図7中に示す。式(16)を用いて文字間空隙の類似度を計算した場合も改善がみられたが、提案手法と比較すると認識精度のピークがやや低くなっている。式(9)と式(16)の違いは、 \mathbf{y}_n と \mathbf{y}_m の差異を考慮するか否かである。式(16)は、 \mathbf{y}_n と \mathbf{y}_m がほぼ同一の場合でも、 $s_{(n,m)}^{(l)(r)}$ が高い値を示し得るため、文字内部の不適切な場所でのセグメンテーションが抑制されにくい傾向があると考えられる。

5 おわりに

本稿では、カメラ入力の低品質文字列画像を認識する手法を提案した。認識およびセグメンテーションの精度を高めるため、文字特徴と文字間空隙の特徴を併用した。

正答語	world	rooms	on	but
入力画像				
$k = 0$	worldl	iroorns	on	lbut
$k = 0.06$	world	rooms	on	but
$k = 0.30$	world	rooms	oln	but
$k = 0.10^{(*)}$	world	rooms	on	lbut

図8 テストデータの例と認識結果。ただし(*)で示した項は、式(16)で類似度を評価した結果を示す。

文字間特徴を評価するにあたり、隣接する文字の端のグレースケール特徴から直交部分空間を作成し、その空間への射影ベクトルの面積を評価値とした。実験により、文字特徴と文字間空隙特徴を組み合わせることで認識精度の向上が可能であることを確認した。また、特徴を組み合わせる際の比率によっては高い改善率が得られた。今後は、特徴の組み合わせ方の改良を検討している。複数フレームを統合し、認識精度を向上させることも検討している。

謝辞 本研究の一部は日本学術振興会科学研究費補助金の支援により行われた。

参考文献

- [1] J. Liang et al., "Camera-based analysis of text and documents: a survey," IJDAR, vol.7, no.2-3, pp.84-104, July 2005.
- [2] R. Casey and E. Lecolinet, "A survey of methods and strategies in character segmentation," IEEE Trans. on PAMI, vol.18, no.7, pp.690-706, July 1996.
- [3] J. Sun et al., "Grayscale feature combination in recognition based segmentation for degraded text string recognition," Proc. 1st CBDAR, pp.39-44, Seoul, Korea, August 2005.
- [4] S. Wachenfeld et al., "Recognition of screen-rendered text," Proc. 18th ICPR, pp.1086-1089, Hong Kong, China, August 2006.
- [5] E. Oja, "Subspace methods of pattern recognition," Research Studies, Hertfordshire, UK, 1983.
- [6] 河原智一 他, "直交相互部分空間法を用いた顔認識," 情報研報, CVIM-151, pp.17-24, November 2005.
- [7] 村瀬洋 他, "候補文字ラティス法による枠無し筆記文字列のオンライン認識," 信学論 (D), vol.J-68-D, no.4, pp.765-772, April 1985.
- [8] 柳詰進介 他, "携帯デジタルカメラにより撮影された動画画像からの低解像度文字認識," 信学論 (D-II), vol.J89-D, no.2, pp.323-331, February 2006.
- [9] T. Breuel, "Segmentation of handprinted letter strings using a dynamic programming algorithm," Proc. 6th ICDAR, pp.821-826, Seattle, USA, September 2001.
- [10] H. Ishida et al., "Recognition of low-resolution characters by a generative learning method," Proc. 1st CBDAR, pp.45-51, Seoul, Korea, August 2005.
- [11] Y. Yang, "An evaluation of statistical approaches to text categorization," Journal of Information Retrieval, vol.1, no.1-2, pp.69-90, April 1999.