

研究種目：若手研究 (B)

研究期間：2007～2008

課題番号：19700177

研究課題名 (和文) 文字特徴の局所的配置に基づく実環境下で有効な高速文字認識技術

研究課題名 (英文) Fast and effective character recognition in real environment based on arrangement of character features

研究代表者

岩村 雅一 (MASAKAZU IWAMURA)

大阪府立大学・工学研究科・助教

研究者番号：80361129

研究成果の概要：近年の文字認識研究において、従来のようにスキャナを用いて文字画像を入力するのではなく、カメラを用いて、情景画像中に存在する文字の認識・理解を行う試みが盛んに行われている。これまでの研究によって、一定レベルの文字認識性能が得られるようになってきているが、まだ十分とはいえない。また、計算速度においては実用的なレベルから程遠い。そこで本研究では、単純だが効果的な処理を組み合わせることにより、情景画像中の実時間文字認識が可能な手法を提案した。提案手法は安価な web カメラと接続したノートパソコン上で実時間で動作する。サーバーを利用した場合には1秒間に200～250文字程度の認識が可能である。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,800,000	0	1,800,000
2008年度	1,500,000	450,000	1,950,000
総計	3,300,000	450,000	3,750,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：文字認識、文字切り出し、特徴点、ハッシュ

1. 研究開始当初の背景

近年の文字認識研究において、従来のようにスキャナを用いて文字画像を入力するのではなく、カメラを用いて、情景画像中に存在する文字の認識・理解を行う試みが盛んに行われている。これまでの研究によって、一定レベルの文字認識性能が得られるようになってきているが、まだ十分とはいえない。また、計算速度においては実用的なレベルから程遠い。そこで本研究では、情景画像中の実時間文字認識が可能な手法を提案する。

まず、カメラを用いて文字を認識できると何が得られるのかについて考えてみよう。将来的には知能ロボットに文字を読ませることも考えられるが、もう少し近い未来の要望

としてインタフェース技術が挙げられる。近年、我々が生活をしている物理空間と、インターネットに代表される情報空間を融合する必要性が叫ばれている。最も基本となる機能の一つは、ポスターを見て関連サイトにアクセスするように、物理空間のモノを通して情報空間にアクセスすることであろう。このとき、誰でもがこの機能を利用できるようにするためには、インタフェースが可能な限り簡単でなければならない。すなわち、モノを指差すような感覚でアクセスできることが望ましい。我々が情報を欲しいと思う対象の多くは、その表面に文字が書かれていることを考えれば、文字は情報空間へのポイントとして非常に適しているといえる。

次に、現在の情景画像中の文字認識技術によってどのようなことが可能であるか、具体例を挙げながら説明する。文献[a]では文字の部分画像を含む様々な変形のパターンを人工的に生成し、学習しておく手法が検討されている。この手法はある程度の変形があっても認識する能力があるものの、認識性能においても、計算時間においても実時間処理が可能なレベルには程遠い。文献[b]では、文字認識の前処理として行われる文字切り出し処理(認識候補の文字領域を特定する処理)において、弱学習器をカスケード結合することにより、画像1枚につき3秒程度で文字領域を切り出すことに成功している。しかし、画像が射影変換(カメラで撮影した際に起こる幾何歪み)のような変形を受けている場合、切り出した文字画像を高速かつ高精度で認識する手法は知られていないため、実際に認識を行うためには様々な変形に対応した柔軟な文字認識手法が必要となる。本研究で実現しようとする高速で高精度な情景画像中の文字認識は、現在の技術水準では実現困難であり、チャレンジングな課題であるといえる。

[a] Y. Kusachi et al., “Kanji Recognition in Scene Images without Detection of Text Fields—Robust Against Variation of Viewpoint, Contrast, and Background Texture—”, Proc. ICPR2004, 1, pp.457–460 (2004)

[b] Chen et al., “Automatic Detection and Recognition of Signs from Natural Scenes”, IEEE Trans. Image Processing, 13, 1, pp.87–99 (2004)

2. 研究の目的

前述のように、本研究では情景画像中の実時間文字認識が可能な手法を提案する。

3. 研究の方法

本研究では3つの方法を検討した。

(1)1つ目は特徴点ベースの方法である。まず、初年度に研究代表者が既に提案している文字切り出し手法 (Feature Point Voting Method; FPV)と高速な検索原理(Locally Likely Arrangement Hashing; LLAH)を融合することで、高速な文字認識手法の開発を試みた。LLAH法は原著論文において、単語の重心を特徴点としたときの特徴点の配置に基づき、デジタルカメラで撮影された文書画像がデータベース中のどの文書画像と一致するかを高速に検索する用途に使用されていた。そのために、多数の安定した特徴点を必要とする。したがって、その方式をそのまま文字認識に適用しようとすると、どのように特徴点を抽出すればよいのかという検

討課題が浮かび上がった。さらに、文書画像では平均数百個の特徴点が存在したのに対して、文字からそれと同程度の特徴点を抽出することは困難であるという問題点もあった。そこで、検討の第一段階として、文字の輪郭から特徴点を抽出ことにした。しかし、輪郭から得られる特徴点は多くないため不足するため、従来のLLAH法程の情報量は得られない。そこで、従来法のように情報を単純に多数の特徴点から得ることを諦め、2種類の角や直線といった特徴点の種類を記述することで情報を得ることにした。実験の結果、相似変換に対してはロバストな認識を行えることが確認できた。この成果は電気学会 電子・情報・システム部門大会で発表した。

(2)2つ目は連結成分を正規化する方法である。アフィン変換を受けた図形を正規化する方法は既にLeuによって提案されていたが、この方法は回転方向の正規化はできないため、Leuの手法を適用した後に何らかの方法で回転方向の照合を行う必要があった。そこで、回転方向と字種の高速な照合を行う progressive pruning という方法を提案した(図1参照)。この成果はMIRU2008とDAS2008で発表した。

(3)3つ目は(1)の方法の発展版であり、(1)では相似変換にしか対応できなかったが、この方法によってその問題を解決できた(図2参照)。具体的には、Geometric Hashing(GH)と呼ばれる特徴点ベースの手法を連結成分の認識に使用し、さらに幾何学的不変量の計算原理を通常とは異なる方法で用いることである(図3参照)。この方法により、特徴点数を P としたときの計算量をGHでは $O(P^4)$ 必要だったのが提案手法では $P(O^2)$ に削減することができた。実験により、射影変換を受けた文字画像を約80%の認識精度で、1秒間に200~250文字の認識が可能であることを確認している。本研究の成果は、平成21年3月の電子情報通信学会PRMU研究会で発表した。また、年度が変わり、研究期間が終了してしまうが、SSII09、MIRU2009、ICDAR2009で発表予定である。

4. 研究成果

ここでは前節の(3)の方法の実験結果を示す。

図4に従来手法では認識できない文字行が直線でないようにデザインされた文書の例を示す。この文書画像を印刷して認識した。デジタルカメラを紙面から0度、30度、45度の3種類に傾けて撮影して、背景が写らな

いように紙面の部分だけ切り取った。切り取った後の 0 度、30 度、45 度の画像の大きさはそれぞれ 2470 x 1746、2278 x 1746、2038 x 1844 である。図 4 には 236 文字含まれている(カンマは除く)。内訳は、Arial が 168 文字、Arial Black Italic が 27 文字、MIRU のロゴの周囲の文字(フォント不明)が 41 文字である。このうち、Arial と MIRU のロゴの周囲の文字を登録して認識実験を行った。なお、表 1 の類似文字のリストで同じ箱に入っている文字はアフィン歪みを受けた場合に認識が難しいので、同一のクラスとした。

認識に要する処理時間と認識性能に影響するパラメータ S を $S=200$ と $S=20$ としたときについて、処理時間と認識結果を表 1 に示す。主な誤認識は「i」の認識失敗(「i」と「I」や「l」の混同)、「U」と「u」や「n」の混同、「E」と「m」の混同であった。 $S=200$ と $S=20$ を比べると、 $S=200$ は $S=20$ よりも認識率が高かったが、 $S=20$ の処理時間は $S=200$ の約 1/6 であった。以上より、提案手法は S が小さい場合は高速な認識が可能であること、若干処理時間を要するが、必要であれば S を大きくすることで、より頑健な認識が実現できることが確認できた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 4 件)

- ① 堀松 晃, 岩村 雅一, 黄瀬 浩一、点配置の一意性を利用した高速なアフィン不変形状認識、電子情報通信学会技術研究報告、PRMU2008-229, pp.127-132、平成 21 年 (2009)、査読無
- ② Akira Horimatsu, Ryo Niwa, Masakazu Iwamura, Koichi Kise, Seiichi Uchida, Shinichiro Omachi、Affine Invariant Recognition of Characters by Progressive Pruning、Proceedings of the 8th IAPR International Workshop on Document Analysis Systems (DAS2008)、pp. 237-244、平成 20 年 (2008)、査読有
- ③ 堀松 晃, 丹羽 亮, 岩村 雅一, 黄瀬 浩一, 内田 誠一, 大町 真一郎、アフィン不変な文字認識手法とその高速化、像の認識・理解シンポジウム(MIRU2008)論文集、IS5-10, pp.1450-1455、平成 20 年 (2008)、査読無
- ④ 堀松 晃, 岩村 雅一, 黄瀬 浩一、局所特徴の組み合わせに基づく相似不変な文字認識、電気学会 電子・情報・システム部門大会後援

論文集、MC3-4、平成 19 年 (2007)、査読無

[産業財産権]

○出願状況 (計 1 件)

- ①名称：パターン認識方法
発明者：岩村雅一、黄瀬浩一
権利者：大阪府立大学
種類：特許
番号：特願 2009-29031
出願年月日：2009 年 2 月 10 日
国内外の別：国内

[その他]

ホームページ等

<http://www.m.cs.osakafu-u.ac.jp/~masa/index-j.shtml>

6. 研究組織

(1) 研究代表者

岩村 雅一 (MASAKAZU IWAMURA)
大阪府立大学・工学研究科・助教
研究者番号：80361129

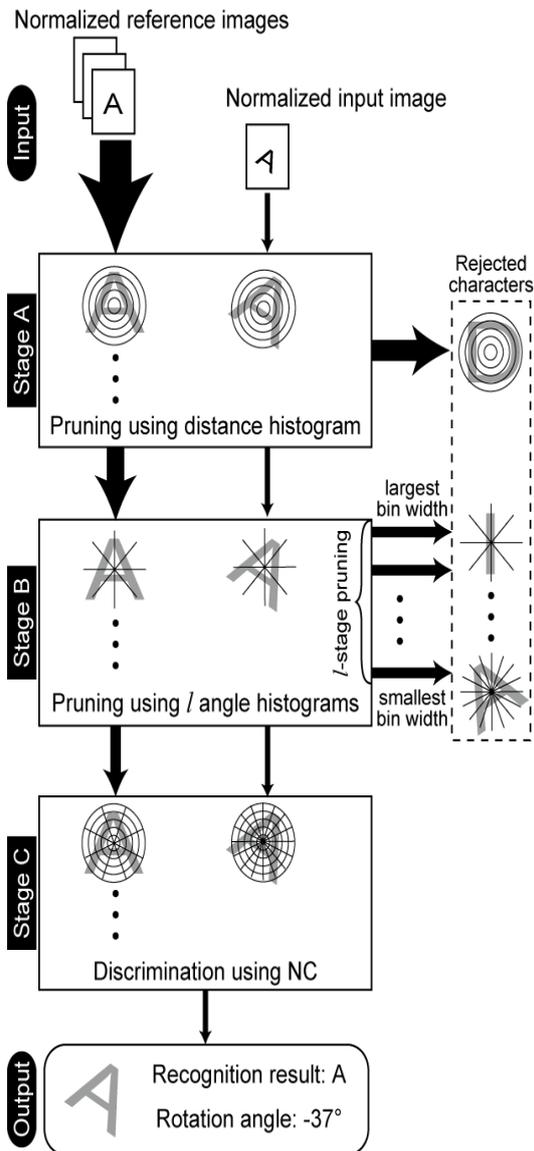


図1 (2)の手法の概要

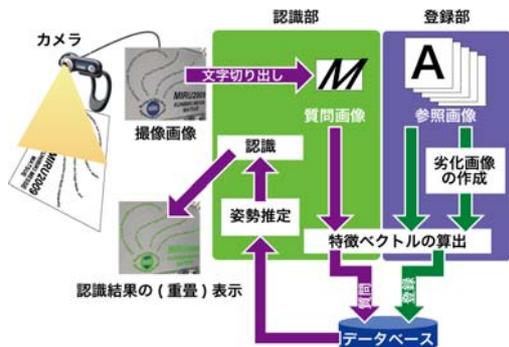


図2 (3)の手法の概要

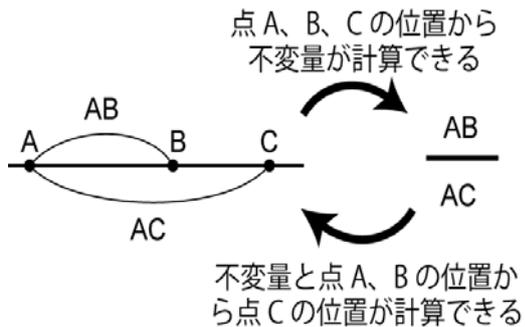


図3 (3)の手法で利用する原理



図4 従来手法では認識できない文字行が直線でないようにデザインされた文書

表1 アフィン歪みを受けた場合の類似文字のリスト.

O O o	6 9	C c	l l	S s	u n
W w	X x	N Z z	p d	q b	7 L V v

表2 図4の文字の認識結果

試行回数 S	200			20		
角度 (度)	0	30	45	0	30	45
処理時間 (ms)	7990	7990	7020	1300	1260	1140
認識率 (%)	94.9	90.7	86.4	86.9	81.8	76.3
リジェクト率 (%)	0.4	3.0	6.4	6.4	9.3	16.5
誤認識率 (%)	4.7	6.4	7.2	6.8	8.9	7.2

