

機関番号：24403

研究種目：若手研究（B）

研究期間：2009年度～2010年度

課題番号：21700202

研究課題名（和文） 情景中の重要な文字情報を実時間で抽出・提示するシステムの試作

研究課題名（英文） Test production of a system picking up and presenting important character information in scenes in real time

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（和文）：本研究では、情景中のあらゆる文字を実時間で認識し、利用者に必要な文字情報を自動的に抽出して利用者に提供するシステムの実現を目指して、特にカメラで撮影された文字を実時間で認識する手法の開発を行った。前述の目的を達成するために必要な、人に装着可能で広範囲の文字認識が実時間で行える手法の開発には至らなかったものの、利用者がカメラを認識対象に向けるという条件の下では高い認識性能と関連情報をユーザーに提示するシステムを作成できた。この成果は2010年6月に行われた国際ワークショップ(Ninth IAPR International Workshop on Document Analysis Systems (DAS 2010))で IAPR Nakano Award(最優秀論文賞)を受賞した。

研究成果の概要（英文）：In this research, in order to realize a system which recognizes all characters in scenes in real time and only information important is presented to the user, we developed real-time recognition methods of characters captured with a camera. While we did not succeed to realize a real-time camera-based character recognition method for a wide-angle camera, we succeeded to realize a system recognizing text that pointed out with a camera and presenting related information to the text working in real time. We were awarded the IAPR Nakano Award (best paper award) at the Ninth IAPR International Workshop on Document Analysis Systems (DAS 2010) for the result.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,900,000	570,000	2,470,000
2010年度	1,500,000	450,000	1,950,000
年度			
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

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：カメラ、射影歪み、アフィン不変なテンプレートマッチング、実時間、文字認識、視覚障害者

## 1. 研究開始当初の背景

身の回りに溢れる文字は有益な情報源であるが、視覚障害者や外国人はそれらを楽しむことができない。そのため、文献[a]では、

視覚障害者が携帯するカードに反応して、音声での案内を提供するシステムが提案されている。しかし、このシステムには、あらかじめ音声案内装置が設置されている場所で

しか使用できないという問題がある。

また、文献[b]では、カメラと文字認識装置と翻訳機の連携による文字翻訳システムが提案されている。ところが、ここで使用されている文字認識技術は真正面から撮影した文字画像しか認識できないため、翻訳したい文字を利用者が上手に真正面から撮影しなければならない。このように、場所や対象を選ぶシステムや、利用者に過度な負担を強いるシステムは利便性が良くない。申請者はこれらのシステムが実際に使用されている様子を見たことがないが、使い勝手の善し悪しが普及するかどうかを決める一つの重要な要素であるのは間違いないと思われる。

[a] 江口 弘, "音声誘導案内システム," 特願平 4-225837 (1992)、特開平 6-63070 (1994).  
[b] Yasuhiko Watanabe et al., "Translation camera," Proc. ICPR'98, pp.613-617 (1998).

## 2. 研究の目的

本研究の目的は、情景中のあらゆる文字を実時間で認識し、利用者に必要な文字情報を自動的に抽出して利用者に提供するシステムの実現である。

想定しているのは以下の状況である。利用者が街を歩くときに小型のカメラとパソコンを携帯し、認識装置が常に周囲の文字を認識し続ける。認識した文字情報が利用者にとって重要であるかどうかは、蓄えてある過去の事例やあらかじめ設定した条件に基づいて判断する。重要と判断された情報は、画像や音声等で利用者に伝える。それ以外の操作を利用者は行わない。

このシステムが実現すれば、視覚障害者にとっては盲導犬のように周囲の状況を確認する手助けになり、外国人にとっては読解不可能な外国語の翻訳機になる。さらに健常者にとっても、人間には発見困難な情報を瞬時に発見(繁華街で多数の看板の中から目的の店を探す場合など)したり、利用者が見落とししている重要な情報を提供することができ、幅広い応用が見込める。

## 3. 研究の方法

本研究で実現するシステムは、[i]カメラを用いた実時間文字認識処理と[ii]文字認識結果から利用者に必要な情報を取捨選択する処理で構成される。

まず、[i]については、課題が二つある。最初の課題は、頑健な文字認識をいかに実現するかである。カメラを用いて文字を撮影した場合、斜めから撮影すると、撮影する角度によって異なる文字画像が得られるため、変形した文字画像の認識が課題となる。もう一つの課題は、どのようにして実時間性を担保するかである。頑健な認識処理を行おうとす

れば、その分だけ時間が必要になる。

[ii]の文字認識結果の取捨選択処理には、利用者の嗜好や行動パターンを反映した知的な処理が求められる。しかし、この処理に要求される必要最低限の要件は、抽出して欲しい情報を利用者があらかじめ登録しておく、単純な照合を行うものであり、比較的単純な処理でも十分である。従って、本研究では[i]に注力した。

本研究では大きく分けて4つの研究に取り組んだ。

### (1) 射影変換に頑健な高速文字認識手法

本研究課題を開始する前に、射影変換を受けた文字の高速な認識方法を既に開発していた。そこで、まずこの既存の方法を改良した。具体的には、従来手法は認識可能なフォント数を増加させると認識率が低下するという問題を解決した。次に、この文字認識手法の出力を組み合わせることで単語認識を実現する手法を開発した。そして、これらの技術を利用して、図1に示すノートパソコンで動作する文字の関連情報を即座に提示するシステムを開発した。これは利用者がカメラを向けた方向にある単語を認識し、その単語の翻訳と、単語にあらかじめ関連づけられた画像や音声を表示するものである。例えば、Hawkという単語を認識したときには、鷹の画像を表示したり、鷹の鳴き声を再生したりするといった具合である。このシステムを用いれば、利用者はwebページのリンクをクリックするのと同様の手軽さで、知りたい情報入手することができる。この認識システムの処理の概要を図2に示す。

この成果を今年度(2010年6月)国際ワークショップ(Ninth IAPR International Workshop on Document Analysis Systems (DAS 2010))で成果発表したところ、IAPR Nakano Award(最優秀論文賞)を受賞した。また、本研究の紹介記事がSPIE Newsroomに掲載された(<http://spie.org/x43601.xml>, DOI: 10.1117/2.1201012.003308)。

### (2) 人に装着可能なカメラによる情景中の文字認識

情景中のあらゆる文字を抽出するためには、利用者にカメラを持ってもらい、実時間で認識する必要がある。そのために利用するカメラとしては、当初Point Grey ResearchのLadybugなどの全方位カメラを想定していたが、人に持ってもらうには重いことと高価であることから、たまたま利用可能であったアイトラッカー付属カメラで情景中の文字を認識することを試みた。その結果、実時間で動作するシステムを開発することができた。しかし、ピントが合っていない認識対象に対しての文字認識は容易ではなく、自動ピント合わせ装置付きのカメラを使用するか、

ぼけに頑健な認識手法が必要ということがわかった。

### (3) 局所特微量ベースの文字認識手法

(1)で開発したシステムは文字切り出しが成功したときには高速かつ高精度に文字認識が可能であるが、そうでなければ効果を発揮しないものであった。そこで、多少時間がかかっても切り出しできなかった文字も認識できるように、局所特微量を利用した文字認識手法を試みた。その結果、(1)のシステムが英数字だけを対象としていたのに対して、この手法は図7のような漢字混じりの日本語を図8のように認識することができ、(1)のシステムを補完するものであることがわかった。

### (4) (1)のシステムとスペルチェッカーの統合

文字を相当斜めから撮影すると、文字が潰れてしまい、文字認識がかなり困難になる。そこで、文字認識で失われた情報を単語辞書を利用して補完するために(1)で開発したシステムにスペルチェッカーを組み込んだ。これにより、単語によっては大幅に認識率が向上した。

このように本研究の[i]に関しては、利用者がカメラを認識対象に向けるという条件の下では高い認識性能と関連情報をユーザーに提示するシステムを作成できた。しかし、情景中から文字情報を発見してユーザーに提示するという目的のためには人に装着可能で広範囲の文字認識が実時間で行える必要がある。この点について、前述の手法(2)で、広範囲を撮影可能なカメラを使用したものではないが人に装着可能な装置を用いて認識実験をしたところピントの問題などが明らかになった。従って、一般にピント合わせが難しいと考えられる広範囲を撮影可能なカメラを利用する文字認識においてはぼけた文字に対する認識が必須であると考えられる。そのため、広範囲の文字認識が可能なシステムを前提とする[ii]に関してはほぼ手つかずのままである。本研究で実現できなかったシステムを今後いかに実現するかを引き続き検討したいと考えている。

## 4. 研究成果

ここでは前節の手法(1)と(3)の実験結果を示す。

手法(1)のシステムのクラス(アフィン変換を受けると類似する字種が同じクラスになるように統合したもの)単位の認識率とそれに要する時間をそれぞれ図3、4に示す。図中の「改良なし」は本研究開始前の性能を表しており、「改良あり」は本研究の成果と

して得られた性能を表している。本研究によって認識率が大幅に向上し、処理時間は減少したことがわかる。また、このクラス認識結果を利用した単語認識を行った。単語単位の認識率とそれに要する時間をそれぞれ図5、6に示す。これらの図から、単語単位の認識が高速かつ高い精度で行える事がわかる。

手法(3)の再現率と適合率を図9に示す。その結果、ひらがなやカタカナのように図形的に単純な認識対象に対する認識性能は若干劣るものの、図形的に複雑な字種の多い漢字は紙面の正面から撮影した場合は再現率97%、適合率98.4%という一定の認識性能を得た。

## 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)

[雑誌論文] (計1件)

岩村 雅一, 堀松 晃, 丹羽 亮, 黄瀬 浩一, 内田 誠一, 大町 真一郎, 段階的な枝刈りによるアフィン不変な文字認識、電気学会論文誌(D), **131**, 7 (2011-7)、査読有

[学会発表] (計12件)

- ① 小林 拓也, 岩村 雅一, 黄瀬 浩一, 局所特徴の位置関係を用いた情景画像中の文字認識、電子情報通信学会技術研究報告, PRMU2010-275, pp.223-228、2011年3月10日、つくば市
- ② 浅田 伸彦, 岩村 雅一, 黄瀬 浩一, 文字誤認識の傾向を考慮したスペルチェッカーによる単語認識の精度向上、電子情報通信学会技術研究報告, PRMU2010-268, pp.183-188、2011年3月10日、つくば市
- ③ Takuya Kobayashi, Masakazu Iwamura, Koichi Kise, Recognition of Affine Distorted Characters by Using Affine-Invariant Local Descriptors, Proceedings of the 2nd China-Japan-Korea Joint Workshop on Pattern Recognition (CJKPR2010), pp.74-77, 2010 Nov 4, Fukuoka Japan
- ④ 岩村 雅一, 辻 智彦, 黄瀬 浩一, カメラで撮影した単語画像の実時間認識、画像の認識・理解シンポジウム(MIRU2010)論文集, IS1-31, pp.247-254、2010年7月27日、釧路市
- ⑤ Masakazu Iwamura, Tomohiko Tsuji, Koichi Kise, Memory-Based Recognition of Camera-Captured Characters, Proceedings of the 9th IAPR International Workshop on Document Analysis Systems (DAS2010), pp.89-96, 2010 Jun 10, Boston U.S.A.

- ⑥ 辻 智彦, 岩村 雅一, 黄瀬 浩一、リアルタイム単語認識技術を利用したカメラベース情報取得システム、電子情報通信学会技術研究報告, PRMU2002-216, pp.51-56 2010年2月18日、東京
- ⑦ 岩村 雅一, 辻 智彦, 黄瀬 浩一、カメラ撮影文字の事例に基づく実時間認識、電子情報通信学会技術研究報告, PRMU2009-222, pp.87-92、2010年2月18日、東京
- ⑧ Masakazu Iwamura, Tomohiko Tsuji, Akira Horimatsu, Koichi Kise, Real-Time Camera-Based Recognition of Characters and Pictograms, Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR2009), pp. 76-80, 2009 Jul 27, Barcelona, Spain
- ⑨ Masakazu Iwamura, Tomohiko Tsuji, Akira Horimatsu, Koichi Kise, Real-Time Recognition of Camera-Captured Characters in Complex Layouts, Proceedings of the Third International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2009), pp.53-60, 2009 Jul 25, Barcelona, Spain
- ⑩ 岩村 雅一, 辻 智彦, 堀松 晃, 黄瀬 浩一、レイアウト非依存な実時間カメラベース文字認識、画像の認識・理解シンポジウム (MIRU2009) 論文集, OS6-2, pp.174-181、2009年7月21日、松江市
- ⑪ 辻 智彦, 堀松 晃, 岩村 雅一, 黄瀬 浩一、文字の並びをリンクアンカー化する Web カメラベースインタフェース、画像の認識・理解シンポジウム (MIRU2009) 論文集, DS-3, pp.1863-1864、2009年7月21日、松江市
- ⑫ 岩村 雅一, 辻 智彦, 堀松 晃, 黄瀬 浩一、カメラで撮像した文字画像の実時間認識システム、第15回画像センシングシンポジウム (SSII09) 講演論文集, IS3-28、2009年6月10日、横浜市

[産業財産権]

○出願状況 (計2件)

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：国内優先権主張  
 番号：特願 2009-163924  
 出願年月日：2009年7月10日  
 国内外の別：国内

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：PCT 出願  
 番号：PCT/JP2010/51889  
 出願年月日：2010年2月9日  
 国内外の別：国外

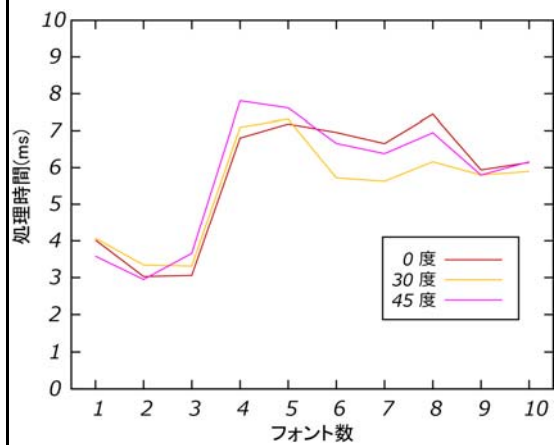
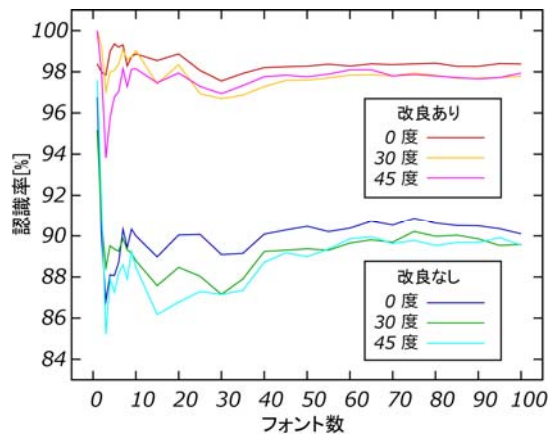
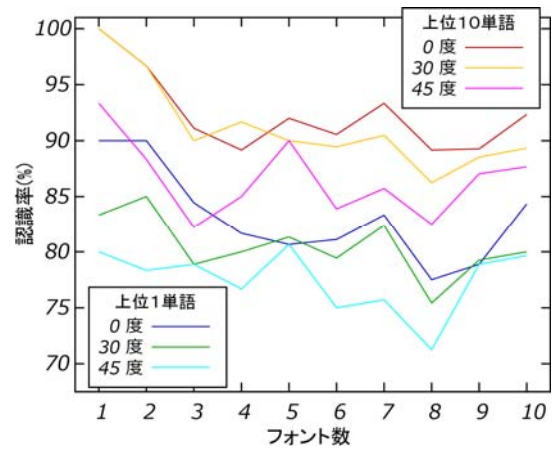
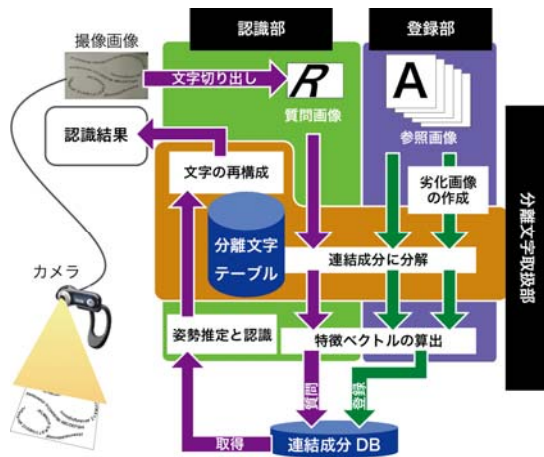
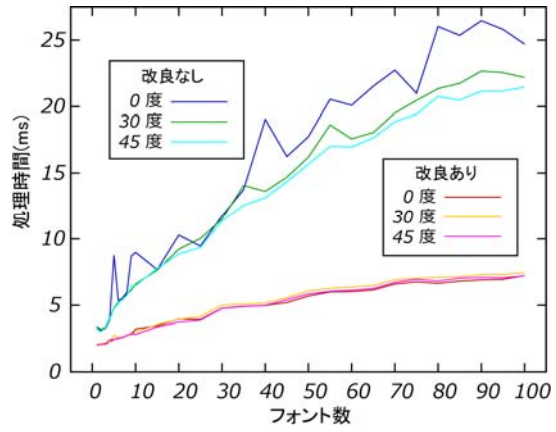
○取得状況 (計0件)

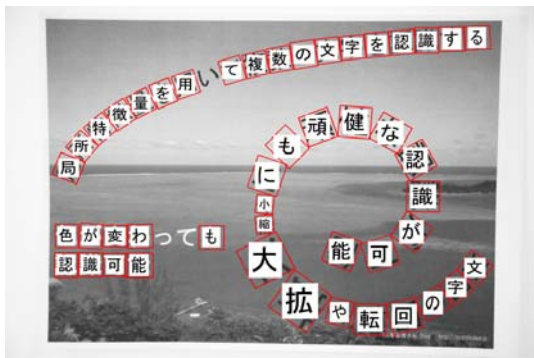
[その他]  
 ホームページ等  
<http://www.m.cs.osakafu-u.ac.jp/camocr/>

## 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





	Hiragana	Katakana	Kanji	Mixture
Recall	74.6	60.6	97.0	75.4
	63.4	50.7	92.2	71.3
Precision	91.4	100	98.4	98.2
	95.7	100	98.5	96.8