

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

辻 智彦[†] 堀松 晃[†] 岩村 雅一[†] 黄瀬 浩一[†]

[†] 大阪府立大学大学院工学研究科

〒 599-8531 大阪府堺市中区学園町 1-1

E-mail: †tsuji@m.cs.osakafu-u.ac.jp, †{masa,kise}@cs.osakafu-u.ac.jp

あらまし ユーザーにとって快適なアプリケーションを考えたとき、インタフェースには即応性が要求される。我々は高速なカメラベース文字認識技術を提案しているが、これを利用すれば新たな対話型インタフェースになり得ると考えている。本デモでは Web カメラで撮影した環境中の文字を認識し、関連した情報を提供するシステムを提案する。キーワード カメラベース文字認識、実時間処理、情報提供システム、リンクアンカー化

A Webcam-Based Interface Which Makes Arrangements of Characters to Anchor Texts

Tomohiko TSUJI[†], Akira HORIMATSU[†], Masakazu IWAMURA[†], and Koichi KISE[†]

[†] Graduate School of Engineering, Osaka Prefecture University

1-1 Gakuen-cho, Naka, Sakai, Osaka, 599-8531 Japan

E-mail: †tsuji@m.cs.osakafu-u.ac.jp, †{masa,kise}@cs.osakafu-u.ac.jp

Abstract It is necessary for an interface to have a rapid response in order to realize a comfortable application for the user. We have proposed a real-time camera-based character recognition method, and we think that there is a possibility to realize a new interactive interface with it. In this paper, we demonstrate a system which recognizes characters captured with a webcam and offers related information.

Key words Camera-Based Characters Recognition, Real-Time Processing, Information Service System, Making to Link Anchor

1. はじめに

ユーザーの要件に応じて必要な情報を素早く提示することはユーザーインタフェースにとって利便性にかかわる重要な要素である。我々は実時間カメラベース文字認識 [1] を提案しているが、これを用いて環境中の文字を実時間で認識できるとインタラクティブな情報取得インタフェースが実現するのではないかと考えている。

本稿では実時間カメラベース文字認識を用いた情報提供システムを提案する。概要は図 1 に示す通りであり、撮影した環境中の文字を認識し、文字の並びに応じた情報を提供する。任意の文字列を指定して翻訳や Web 検索などを行うことができ、これを我々は「環境中テキストのリンクアンカー化」と呼ぶ。情報はテキストによるものに限らず、画像や音声といったマルチメディアのサービスも可能であり、近年多機能化している携帯電話のように新たなサービス配信ツールとしての発展が期待できる。

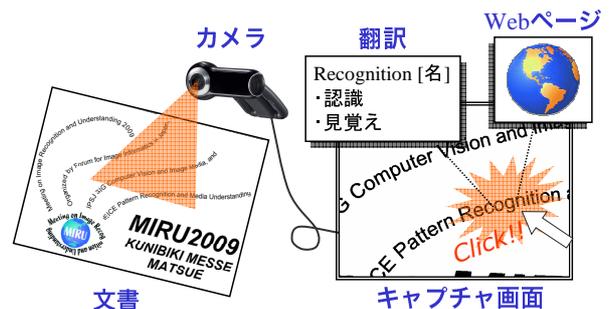


図 1 アプリケーションの概要

2. 高速文字認識の特色

今回、高速なカメラベース文字認識には我々が提案した手法 [1] を用いる。この手法では射影歪みに頑健な認識を実時間で行うことができ、サーバーを用いた場合 1 秒間に約 200 文字が認識できることを確認している。市販の Web カメラとノートパソコンでもインタラクティブなインタフェースが実装可能であり、手軽に持ち運ん

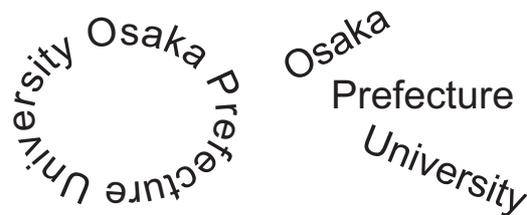


図 2 レイアウトの例

で環境中の文字から情報を得ることができる。また、この手法では 1 文字単位で認識を行っており、文字の並びや方向が揃っていないかったり文書内に複数の文字サイズや字体が存在していたりしても認識に影響しないという利点を持つ。これにより、図 2 のような並びの文字でもリアルタイムに、また斜めから撮影しても認識が行える。

提案システムは環境中のあらゆる文字列を利用できるが、同じ文字列でも書かれている場所やレイアウトによって文字の持つ意味が異なり、それに応じて提示すべき情報が変化することが考えられる。文字列の識別レベルは複数考えられるが、そのうちの 3 つについて述べる。1 つ目は同じ文字列は全て同じ意味を持つという考え方である。例えば図 3(a) のように異なるレイアウトでも文字列が同じ “Osaka Prefecture University” であれば同じ情報が提供される。2 つ目は文字列とそのレイアウトの組み合わせが意味を持つという考え方である。この場合、図 3(b) のように同じ文字列でもレイアウトが違えば別の情報が提示される。3 つ目は文字列が書かれている場所が意味を持つという考えである。この場合、図 3(c) のように雑誌に書かれている場合は Web ページ、看板上に書かれている場合は案内図、のように変化する。ここで、SIFT [2] のような特徴量を用いた画像認識システムを考えると、同じ画像かどうかを識別するため、同じ並びの同一文字列のみが同じ意味を持つとしか解釈されない。提案システムでは様々な識別レベルを持つことができ、柔軟な情報提供が行える。本稿では先述の 3 種類の識別レベルのうち、最初の 2 つのレベルで識別できるようにした。他にも字体ごとに提供される情報を変えろという方法が考えられる。

3. デモの概要

ここではデモとして環境中の文字を認識して関連情報を提供するシステムを紹介する。装置は図 1 のようにインタフェースを実装したノートパソコンに市販の Web カメラを接続したものである。

デモではユーザに提示する情報として翻訳、画像表示、Web リンク、音声案内の機能を実装した。翻訳は図 3(a) の認識レベル、それ以外は図 3 の (a) と (b) の 2 種類の認識レベルで情報を提供する。翻訳の機能は外国への旅行中に看板上に書かれた文章の意味が分からない場合な



(a)



(b)



(c)

図 3 識別レベルの例

どに役立つ。従来は辞書で 1 単語ずつ調べるのに対し、提案システムでは撮影だけで翻訳語が提示されるため非常に便利である。Web リンクの機能は環境中の文字列を撮影すれば URL を入力することなく関連サイトにアクセスできる。QR コードも URL を入力せずにアクセスできるが、あらかじめ貼り付けた限定的な媒体にしか利用できない。しかし、提案システムではそのような制約が無く、利便性が向上する。また、音声や画像を貼り付けることもでき、メッセージや音楽を再生することで宣伝媒体などとしての利用が期待できる。提案システムでは動物に関する記事に対して鳴き声や写真をリンクさせた。

キャプチャ画面に含まれる文字列を認識して全ての情報を提供することもできるが、その場合ユーザにとっては雑多な情報を受け取ることになり、たとえ即応性があったとしても不便である。そこで、提案手法ではユーザにとって有益な情報のみを取り出せるよう、任意の文字列を選択して関連情報が得られるようにした。選択はマウスでクリックする方法の他、カメラのアクションを利用する方法がある。これはトラッキングを用いてカメラの動きを追跡し、円を描いたり振ったりすることでユーザの意思が伝えられるという直感的なインタフェースである。

謝辞 本研究の一部は、科研費補助金若手研究 (B) (19700177, 21700202) の補助による。

文 献

- [1] 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, 2009.
- [2] David G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, Vol. 60, No. 2, pp. 91–110, 2004.