

情景画像からの文字情報読上げを利用した視覚障害者向けシステムに関する研究

Study on Text Reading System from Natural Scene for Blind Person

研究学生 インドラ ヘルマワン 指導教官 江崎 修央

1. はじめに

近年、情景画像からの文字抽出を利用した研究が盛んに行われている。この技術を利用したアプリケーションシステムとして、図書館における書籍画像データベースの自動検索、高速道路における案内板や標識の自動認識などが提案されている。このような研究が盛んに行われる背景には、高解像度なデジタルカメラやネットワークカメラの普及があると考えられる。

そこで、我々は3年前から視覚障害者のための環境内文字情報読上げシステムを提案し、主に文字領域抽出方法について研究を進めてきた[1-2]。本稿では文字領域抽出手法の精度向上とシステム実用化に向けた第一歩としてレストランメニューを読み上げシステムの実現を目指す。

2. 新しい文字領域抽出の提案

これまでに、開発してきた文字領域抽出はエッジなど幾何学的情報を利用した手法が有効であった。今回は画像上の色要素に着目した手法を提案する。まず、一般的にコンピュータ上で扱われているRGB表色系から、HSB表色系に各画素値を変換する。ここで、HSBとはH:色相, S:彩度, B:明度を示している。図1(a)に示した元画像に対して、HSBの各チャンネルのヒストグラムに対して判別分析手法[3]で2値化をおこなう(図1(b))。この2値化された画像およびその反転画像、それぞれの画像に含まれる結合要素に対してラベリングを行い(図1(c))、文字列抽出のルール[2]を適用し文字候補領域の抽出を行う(図1(d))。

本手法をICDAR2003RobustReadingCompetition[4]データセットに対して適用し、文字候補領域の抽出実験を行った結果は54%であった。これまでに提案しているSobelエッジに基づく方法の認識結果は64%であるので、良くないように見えるが、これらの手法から出力される領域のORをとると61%の認識率となった。



図1 色相の2値化による文字領域抽出の例

3. メニュー読み上げシステムの基本設計

システム実用化に向けて、我々はレストラン内のメニュー自動読み上げシステムの開発に着手した。システムの構成要素としては、ノートパソコンおよびネットワークカメラである。利用するネットワークカメラはパン・チルト・ズーム機能とオートフォーカス機能を備えている。また、ノートパソコンには音声出力ソフトウェアがインストールされており、利用者へのメニュー情報の提示は音声により行われるものとする。

大まかなシステムの流れとして利用者が席に着いたあと、メニューをカメラの前に設置し「キャプチャ」ボタンを押すことにより、メニューのデータベースが自動作成される。利用者は音声出力を利用して「Yes/No」方式の応答を行うことにより、好みのメニューの値段や説明を知ることができる。

4. メニューデータベースの自動作成

世の中にあるメニューにはいろいろなフォーマット、フォントがあり、すべてのメニューに対してロバストにデータベースを作成する手法を提案することは困難である。そこで、本稿ではインターネット上から取得したピザのメニューを取り上げてデータベースの自動構築実験を行った。

メニューに含まれる項目は図2に示すように「category」、「item」、「price」、「category option」、「item option」である。取得されたメニューの画像に対して各種画像処理を行い、カテゴリ抽出、項目候補領域の抽出、連続した点の取り除き、アイテムと値段の決定、オプションの決定の順で実行しメニューのデータベースを作成する(図3)。

今回は切り出したメニュー項目に対して認識処理は行わなかったため、データベースとして各項目の座標をXMLのファイル形式に保存した。

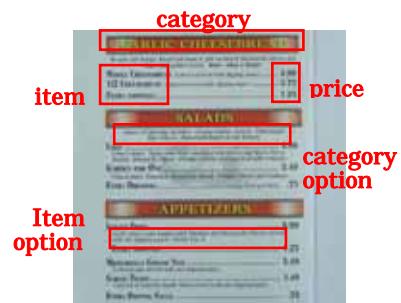


図2 メニューを構成する項目



(a) カテゴリの抽出 (b) item, price, optionの抽出

図3 メニューのデータベース化

5. まとめ

本研究ではHSBによる新しい文字領域抽出手法を提案した。また、視覚障害者向けシステム実用化の第一歩としてレストランメニューの自動データベース化を実装した。作成データベースにより、システムとユーザの簡単なやり取りにより項目選択が可能となる見通しが得られた。

参考文献

- [1] N.Ezaki, K.Kiyota, B.T.Minh, M.Bulacu and L.Schomaker, "Improved Text-Detection Methods for a Camera-based Text Reading System for Blind Persons", Proc. of ICDAR, 2005.
- [2] Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons, Nobuo Ezaki 他, International conference on pattern recognition 2004
- [3] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram", IEEE Trans. Systems, Man and Cybernetics, Vol. 9, 1979, pp. 62-69.
- [4] S.M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, "ICDAR 2003 Robust Reading Competitions", Proc. of the ICDAR, 2003, pp. 682-687.