

視覚障害者のための環境内文字情報読み上げシステムの開発

Development of Text Information Reading System in Natural Scene Images for Blind Person

1061002 (登録番号)

研究代表者 鳥羽商船高等専門学校制御情報工学科 准教授 江崎 修央

[研究の目的]

近年、画像中からの文字領域抽出・認識を利用したシステム開発が注目を集めている。情景画像中から文字領域を抽出し、認識処理を利用することで、旅行先での自動翻訳システムや図書の自動データベース化などに利用されつつある。情景画像中の文字情報を読み上げることができれば、視覚障害者にとって必要な情報が得られる可能性が広まる。そこで視覚障害者向けのシステム用技術としても注目されており、バスの行き先読み上げシステムなどが報告されている。我々もこれまでに文字領域抽出手法に重点を絞って研究を行ってきた。この文字情報読み上げシステムと GPS によるヒューマンナビゲーションシステムを組み合わせることで、視覚障害者の単独外出を支援するシステムが構築できると考えている。

本研究では視覚障害者向けの情景画像中の文字情報読み上げシステムの実用化の第一歩としてレストランのメニュー読み上げを行うシステムの開発を目標とし、基本設計、文字領域の抽出手法の実験とメニューデータベースの作成およびメニュー選択に関する検討を行った。

[研究の内容, 成果]

1. メニュー読み上げシステムの概要

まず、レストランにおけるメニュー読み上げを行うにあたり、対象の店舗の形態を大きく「定食屋」系と「居酒屋」系に分けて考える。「定食屋」系とは、一人が一品頼む場合を主に想定している店舗形態、「居酒屋」系とは、多人数で複数の品を頼む形態であると定義する。「定食屋」系の店舗では、食べ物の注文は、一度で済むことが多いのに対して、「居酒屋」系の店舗では、食べ物や飲み物を複数回に渡って注文するという特徴がある。つまり、「定食屋」系の店舗では、メニューの内容を一度店員に聞けばよいので比較的質問しやすいが、「居酒屋」系の店では、メニューが多岐に渡る上に、注文するたびに質問することとなる。

本システムの構成要素は、図1に示すとおりノート型パソコンなどの小型携帯端末、ネットワークカメラ、コンピュータとへの情報入力を行うハンドコントローラである。動作の概要は以下に示すとおりである。①ユーザが読み上げさせたいメニューを前に広げてハンドコントローラの **Capture** ボタンを押して画像を撮影する。②文字領域抽出処理によりメニュー画像から文字領域を抽出する。③切り分けられた文字領域に対して認識処理を行う。④認識結果から階層構造のメニューデータベースを作成する。⑤システムはユーザに対して音声で順次選択を促す。⑥メニューの選択にはハンドコントローラのボタンを利用する。

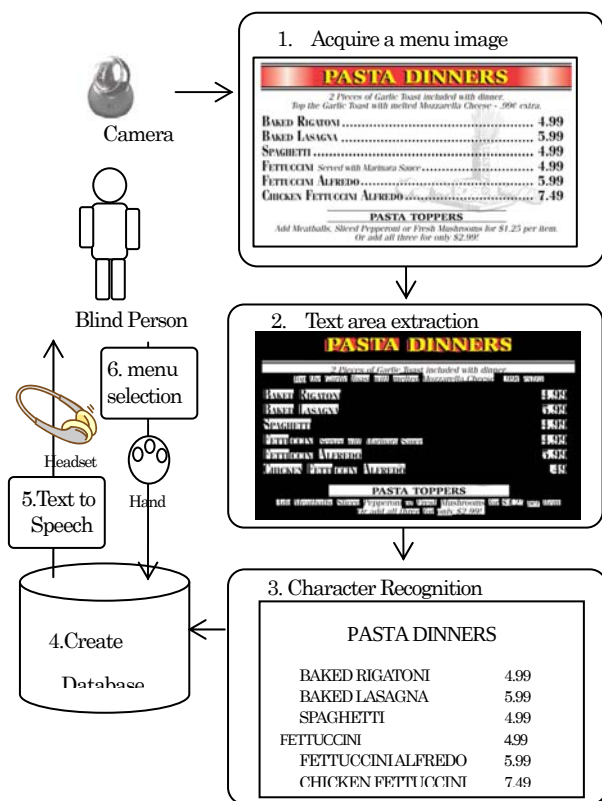


図1 システム概要図

2. メニュー画像からの文字領域抽出

2.1 対象画像

今回対象としたメニューはインターネット上で取得した居酒屋のメニュー画像である。店舗数は10、画像数は35枚である。図2にメニュー画像の一例を示す。

一概にメニューといっても、文字のみからなる画像、食べ物写真が掲載されている中に文字列が記入されている画像などさまざまなものがある。従って、既存のOCRソフトウェアでは文字領域と非文字領域の切り分けができないので、そのままでは認識処理には利用できない。

2.2 文字領域の抽出手法

対象画像には比較的小さな文字が多く含まれているため、これまでに研究を行ってきた文字領域抽出手法の中でも小さな文字検出に強い4つの手法を利用した。以下にそれぞれの手法の特徴を述べる。

8color2手法は、対象画像を 32×32 ピクセルの小領域に対象画像を分割し、RGBそれぞれのカラーチャンネルに対して判別分析による2値化を行うことで、画像全体を8色に減色した後、文字領域を抽出する手法である。隣り合う文字領域は基本的に同じ色から構成されていることが多いという性質を利用した手法である。

B-Binary2はRGBカラーチャンネルを利用するのではなく、HSB表色系の明度に対して2値化を行う方法である。これも8Color2手法と同様に、小領域に分割したあと処理を行っている。メニューに限らず文字領域は人間が認識しやすいように背景領域と文字領域のコントラストをはっきりさせるので、明度で2値化を行うことは、背景領域と文字領域の切り分けに有効な手段である。

Closing-Opening処理は、Morphology演算の基本処理手法であるClosingを行った画像とOpeningを行った画像の差分を取る処理である。この処理により、文字領域は水平に接続した領域として出力される。



図2 メニュー画像の例 [1]

2.3 抽出結果

文字領域抽出の評価値として *Precision*, *Recall* を利用した。 *Precision* とは抽出した領域中に文字領域がどの程度含まれているかという指標である。 *Recall* は対象画像中の文字領域

をどの程度抽出できたかという手法である。*Precision* と *Recall* はそれぞれ相反する指標であるので、それぞれを考慮した評価値 f も利用する。今、抽出した文字領域を E 、正解文字領域を T 、 E と T の共通領域を c とすると、*Precision* と *Recall* および f はそれぞれ下記のように計算される。

$$precision = c / |E| \quad (1)$$

$$recall = c / |T| \quad (2)$$

$$f = (0.5 / precision + 0.5 / recall)^{-1} \quad (3)$$

表 1 に抽出結果を示す。表中で“+”で表してある結果は、2 つもしくは 3 つの手法で出力された領域の論理和を取った領域である。表 1 の結果から、最も高い *Precision* となったのは、B-Binary2 の手法で 78.5% であった。ただし、B-Binary2 だけでは *Recall* の値が低く、総合的な評価値 f としては 54.8% にとどまった。 f が最も高かったのは、B-Binary2 と Closing-Opening の論理和をとった結果で 63.7% であった。メニューを完璧にデータベースに登録することを考えると、高い *Recall* が必要である。多少 *Precision* が下がったとしても、抽出領域が本当に文字領域かどうかを文字認識処理結果により判断できると考えるからである。

表 1 各手法による文字領域抽出率

	<i>precision</i>	<i>recall</i>	f
8color2 (8)	69.4	44.1	53.9
B-Binary2 (B)	78.5	42.1	54.8
Closing-Opening(CO)	66.4	48.3	55.9
8 + CO	61.5	62.4	62.0
B + CO	66.1	61.4	63.7
8 + B + CO	61.1	63.1	62.1

3. 投票制を用いた文字認識

3.1 複数の認識処理系による投票制について

2 章で説明した文字領域の切り分け処理により、1 文字ずつの文字画像データが入力として

与えられる。その画像に対して認識処理を行う前に、複数の画像処理手法を組み合わせた前処理を行うことにより、いくつかの文字画像が生成される。これら別々の前処理を施された画像に対して文字認識処理を行うと、同じ文字画像にもかかわらず別の認識結果を得る可能性がある。ただし、前処理に関してどの手法の組み合わせが適切かは元画像によって違う為、今回は下記の組み合わせを利用した認識系を構築した。

初期画像として用意する画像は次の 5 つとした。それらは、「元画像 (カラー画像)」、「グレースケール画像」、「RGB の青要素のみの画像」、「RGB の R のみ」、「RGB の G のみ」である。上記で述べた 5 つの画像を初期画像として、それぞれに以下に示す 5 種類の前処理をすべての組み合わせで施すこととする。5 種類の前処理は「2 値化」、「メディアンフィルタによるノイズ除去」、「階調反転」、「ガンマ補正」、「モルフオロジ演算によるトップハット処理」である。これら 5 つの前処理の組み合わせ数は $2^5=32$ 通りとなる。つまり、5 つの初期画像それぞれに対して、32 通りのすべての組み合わせの前処理を施すことにする。これにより、5 (初期画像数) $\times$ 32 (前処理の組み合わせの数) = 160 通りの画像が生成されることとなる。今回利用した OCR 処理系は AI ソフトの「活字文書 OCR ライブラリ」である。このソフトウェアは、1 つの画像に対して 1 つの候補文字列を出力するので、この投票制を利用した認識処理系の投票総数は 160 票となる。最終的には、投票結果の多い順から上位 10 候補を出力する。

3.2 文字認識実験結果

投票制による文字認識処理の結果が図 3 である。グラフの横軸は順位、縦軸は認識率を表している。1 位の結果は 28% だが、2 位、3 位と累積するごとに認識率が上がっていった。1 位ではまだ認識精度が低いものの、2 位まで考慮すれば認識精度は 38%、10 位まで累積結果で認識率は 50% 程度に上がることを確認した。認

識率はフォントによるばらつきが大きいので、
今後はフォントに依存しない認識処理を検討する
必要がある。

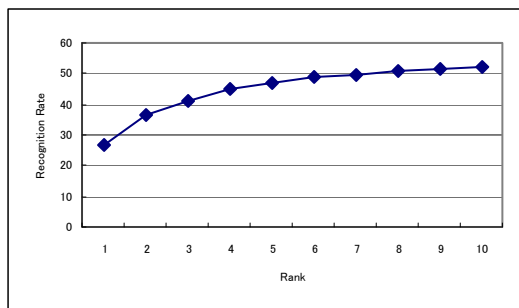


図3 投票制による文字認識処理の結果

4. メニューの読み上げ方法に関する検討

店舗における多くのメニューは、「うどん」「丼」「定食」などに分類されている上で「きつねうどん 400 円」「肉うどん 600 円」というメニュー項目などが記載されている。しかし、分類項目は店によって違う。特に最近新規開店する店舗では「一品料理」の分類項目に対して「すぐに出る料理」といった表記を行ったり、メニュー自体も「昔ながらの屋台ラーメン」といった具合に消費者の心をつかむようなキャッチフレーズ入りのメニューも多く見られるようになってきた。メニューの名前に関しては視覚障害者にとってもただ「ラーメン」と読み上げられるよりも「昔ながらの屋台ラーメン」といった名前のほうが楽しめると思うが、それらの分類項目があまりに奇抜なものであると、はじめの段階でどこに希望のメニューがあるのか想像しにくくなると考えられる。そこで、我々は一般的な店舗の典型的な分類項目を整理し、そのテンプレートのどこに該当のメニューを振り分けるかを自動的に決定するメニューデータベース作成機能の実装を目指す。

今回の実験におけるメニューの選択方式として、「Yes/No」選択方式について検討を行った。

「Yes/No」選択方式とは、「Yes」と「No」に対応した2つのボタンのみを押すことによってメニューを選択していく方式である。「Yes」は

「決定」を意味し、次の階層に進む場合や最終的なメニュー決定に利用される。「No」は「次に進む」ことを意味し、同階層の別の項目を読み上げる動作を行う。

〔今後の研究の方向、課題〕

今後は、文字領域抽出、認識精度の向上を目指していく必要がある。また、メニューの読み上げに関しても、さらに効果的な階層構造を検討を行う。

〔成果の発表、論文等〕

- 1) 江崎修央, 砂崎由樹, 中西 航, 清田公保, 視覚障害者向けレストランメニュー画像からのメニュー読み上げシステム, 電子情報通信学会ヒューマンインフォメーション研究会技術報告 (ITE-HI), ITS2006-82, IE2006-267, pp.41-46 (2007)
- 2) K.Kiyota, N.Ezaki, K.Itoh, L.Hirasaki, S.Yamamoto, Pen-based Electronic Note-taking System to support study of visually impaired person, the 13th Conference of the International Graphonomics Society (IGS2007), pp.205-207 (2007)

〔参考文献〕

- [1] 村さ来ホームページ,
<http://www.murasaki-net.co.jp/>