

視覚障害者のためのカメラ映像内の文字情報読み上げシステム

清田 公保^{*1} 島川 学^{*1} 本田 幸代^{*1}

江崎 修央^{*2} 砂崎 由樹^{*2}

Text reading system based on extraction of a character in camera image for visually impaired person

Kimiyasu Kiyota^{*1}, Manabu Shimakawa^{*1}, Yukiyo Honda^{*1},
Nobuo Ezaki^{*2} and Yuki Susaki^{*2}

Abstract This paper proposes a text reading system by using active micro camera. This system is designed as information navigation for visually impaired person. The system is composed with a laptop PC, Network Camera and a command input device. At first, the system extracts character areas automatically, and characters are recognized by offline-character recognition application. Next, The text database of the layered structure is composed from captured menu image by using the system. A technical factor in this system is as follows. Because the user of the system is the blind, the captured menu image is not certain whether to contain an appropriate character area. Some characters are transformed such as inclination, too big or small. Therefore the proposed system extracts the character area including a small character first. The camera zooms to each candidate character area, and then character area of high resolution is extracted. The proposed system can be used to check items of menu and prices. We apply this system to restaurant menu reading system for blind person. The user opens the first page of the menu, and pushes [capture] button on the portable PC. In this time, the camera turned in the direction of the menu. Then system tries to make database of the menu automatically. Then, a menu item database is composed of the menu items that are extracted by offline character recognition. The database items are name of menu and its price. The system announces the menu items by voice synthesizer. The user is able to select the candidate item by using [yes] or [No] button on the control interface board in the system. This operation is very easy for the blind. From the viewpoint that the visually impaired person uses it, we discuss that the system design of menu reading system, and the presentation method of an efficient voice reading out.

Keywords : Text reading, visually impaired person, Ubiquitous computing, PDA and Mobile PC

1. はじめに

厚生労働省の平成 14 年度の発表によると、我が国における全国の 18 歳以上の在宅の身体障害者は 3,245,000 人（推計）で、前回（平成 8 年 11 月）の調査から 10.6%増加している。特に近年は、白内障や糖尿病による疾病や交通事故により中途失明する後天的視覚障害者の割合が年々増加し、高齢社会の進行と共にその深刻性が増してきている[1]。一方、情報メディアの発達と各種情報端末の低価格化に伴い、我が国では次世代の高度情報化社会に対応すべく e-Japan に代表される情報ハイウェイの整備・確立が急務となってきている。こうした中で、一般家庭における ADSL や光ブロードバンドネットワークに代表される情報網の整備が急速に進められており、視覚障害者の間でも、音声合成ソフトや点字キーボードなどを活用してコンピュータを利用したいという要望が挙がっている。しかし、視覚

障害者全体に対するパソコンの利用率は 5 %程度（1万5千人）を推移したままであり、視覚障害者に対する高度情報化社会への対応策として、情報機器の利用支援は深刻な課題である。一般に情報機器に不慣れな高齢者や障害者のパーソナルコンピュータ利用を支援する技術への取り組みは立ち後れており、デジタルデバイド（情報格差）は、今後益々拡大する傾向にある。このような問題に対して、我々は普段の筆記と同じ手法でコンピュータに日本語入力が可能なオンライン文字認識技術に注目し、視覚情報の欠如により筆跡のフィードバックなしで書かれた変形の大きな文字に対しても柔軟に対応可能な文字認識手法を提案してきた[2-4]。さらに本手法をペン入力型携帯移動端末に適用し、視覚障害者の屋外環境での利用を考慮したユビキタスコンピューティングにおける情報入力インタフェース部の基本技術を確立している（図 1）。本研究では、これまでに開発した視覚障害者が屋外で利用可能な情報機器端末として機能するユビキタスインタフェースを基礎として、これまでの個々の研究成果を融合した総合情報端末としてのユビキタ

*1: 熊本電波工業高等専門学校

*2: 鳥羽商船高等専門学校

*1: Kumamoto National College of Technology

*2: Toba National College of Maritime Technology

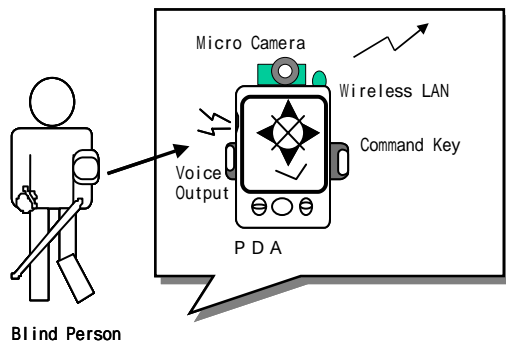


図1 視覚障害者のためのユビキタスコンピューティングシステムのイメージ

Fig.1 Ubiquitous computing system for visually impaired person.

ス情報ナビゲーションシステムの実用化を目指す。

今回は、視覚障害者が実際に情報端末を用いて利用したい機能のひとつとして挙げられた、交通標識・看板、レストランのメニューや値札などの文字情報をリアルタイムに読み上げるシステムの検討を行う。これまでに視覚障害者向けに周囲の環境から文字情報を抽出し、文字情報を提示する研究が報告されている[5]。本研究では、さらに具体的な視覚障害者への音声提示法までの検討を行うために、視覚障害者向けのレストランなどのメニュー読み上げシステムの基本設計を新たに考察する。特に情報機器に不慣れた視覚障害者に対しては、音声出力が有効な情報提示であるため、音声機能を用いた効

率的な情報提示の一手法について検討する。

2. システムの概要

2.1 メニュー読み上げシステムの基本設計

図2に、提案する視覚障害者のための文字情報読み上げシステムの基本構成を示す。本システムのハードウェアにおける構成要素は、ノートパソコンなどの携帯可能な情報端末とパン・チルト・ズーム機能付きのカメラである。システムの処理の流れは、画像のキャプチャ、文字候補領域の抽出、文字認識、文字情報の読み上げを順に行っていくが、対象となる文字の大きさやフォントは様々である。特に初期キャプチャ画像に含まれる文字のサイズは小さいことが予想されるため、まず小さいサイズの文字領域を専用の文字領域抽出手法で抽出し、候補文字領域にカメラのパン・チルト・ズーム機能を利用してズームインした後、改めて文字認識用の画像をキャプチャすることとしている。文字認識のためには別の手法で文字領域を抽出し、オフライン文字認識を利用して文字認識し音声合成ソフトにより音声出力される。

2.2 文字領域の抽出

視覚障害者が文字情報を知りたい対象物に携帯端末に備え付けられたマイクロカメラを向けたとき、初期段階でキャプチャされる画像は、晴眼者が撮影するような認識対象の文字のみが完全に含ま

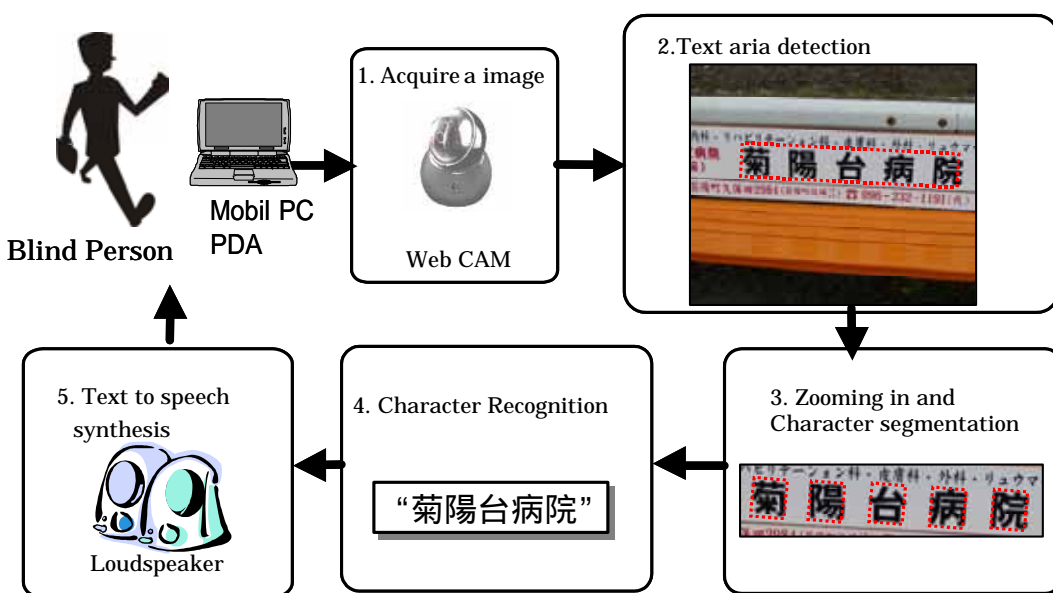


図2 視覚障害者のための音声情報読み上げシステム

Fig.2 Text reading system based on extraction of a character in camera image for visually impaired person.

れる可能性は低く、文字が含まれないか含まれていても認識できない程度に小さく撮影される可能性のほうが高いと考えられる。そこで、はじめにシステムは小さな文字領域抽出に有効な抽出方法を利用して候補文字領域を抽出する。

今回、領域抽出には、モルフォロジ演算の Tophat 処理を応用した手法および小領域 (32×32pixel) に対する 2 値化を利用した方法の 2 種類を併用している[6]。これらの処理によって候補文字領域が検出されたとき、キャプチャされた画像内には、文字情報があると判断し、それぞれの領域にズームインする。

ズームアップされた画像中には候補文字領域のみが含まれているため、これらの文字を 1 文字ごとに抽出するために Sobel エッジを利用した手法を 2 種類、RGB 各カラーチャンネルに対して 2 値化を行う手法の合計 3 種類の手法[7]を並列に適用し、候補領域を文字ごとに切り分ける。ここで、各領域に対して文字認識処理を施し、文字であることが確定されたならば文字情報として音声出力する。文字である可能性が低い場合には音声出力は行わず他の文字候補領域の検索を行う。

3. 文字領域抽出のプロトタイプによる実験

2. で述べたテキスト文字の抽出手法の実用能力を検証するために、実時間で文字領域を抽出するプロトタイプを試作した。ズーム、チルト機能およびオートフォーカス機能がある Logitech QVR-1 カメラをラップトップ PC に接続したもので携帯可能である。

図 3 に本プロトタイプシステムのカメラでキャプチャーされたカメラ映像の例を示す。利用者が、ある地点において現在の周りの情景の文字情報を得たがっているとき、ある方向に向かってカメラを向けてシステムのキャプチャボタンを押すことによりカメラ画像をが取り込まれる。このときカメラは文字領域抽出アルゴリズムにより、文字領域を自動的に探索する。我々が提案しているプロトタイプシステムは、誤り指数を最小にするのにいくつかのテキスト検出手法を用いている。システムは分類された文字候補領域の目標位置を一時記憶しており、画像をさらに拡大した後に、ソーベルメソッドを用いて文字認識のために 1 文字ずつ文字を抽出する。1 フレームあたりのソーベルメソッドによる画像処理時間は中の 300[ms]程度を要している。このプロトタイプシステムは実時間で動作可能であり、図 3 (a)の初期画像より、図 3 (b) ~ (d)の各



(a) First captured image



(b) Zoomed in image 1



(c) Zoomed in image 2



(d) Zoomed in image 3

図 3 提案システムのプロトタイプによる画像キャプチャーの一例 (駅舎構内の看板文字)

Fig. 3 Examples of a captured image by the prototype system.

小領域を 200[ms](インテル Pentium4 1.4GHz プロセッサ使用)のオーダーで、ズームングして切り出すことが出来た。しかし、切り出し結果からも分かるように抽出精度はまだ完全ではない。画面の状況や照明などの外的な環境状態により、最適な切り出しの閾値が刻々と変化することが大きな原因である。実時間処理に影響を及ぼさない範囲で複数の切り出し処理を併用し、候補文字領域の中から文字らしさの評価値を適用することにより文字枠の再構成と絞り込みを行うことで改善していく予定である。

4. メニュー音声読み上げシステムへの応用

4.1 メニュー画像からのメニュー品目の抽出

視覚障害者でも、白杖を利用し単独で自由に屋外を移動することが可能である。しかし、レストランや喫茶店などで注文をする際、初めての店ではメニュー内容を店の人に読み上げて貰う必要がある。人に頼らずに自由に情報が読み取れる携帯の読書機の開発が待たれるが、文字情報の読み取り技術にはまだ多くの課題がある。プロトタイプによる文字領域の切り出し処理の実用性が確認できたので、今回は視覚障害者がユビキタスアプリケーションのな

かで利用したい機能のひとつであるレストランや喫茶店にあるメニューの自動読み上げシステムを検討する。特殊フォントや手書きメニューに対しては、高度なオフライン文字認識技術が必要であるが、一般的なフォントで印刷されたメニューであれば、市販の汎用 OCR 認識エンジンでもある程度の認識精度が期待できる。そこで、今回はレストランや食堂にあるメニューを対象に、プロトタイプにより得られた文字領域画像から文字認識に有効な特徴画像の生成を行い、既存の OCR 認識エンジンに入力として与えることとした。

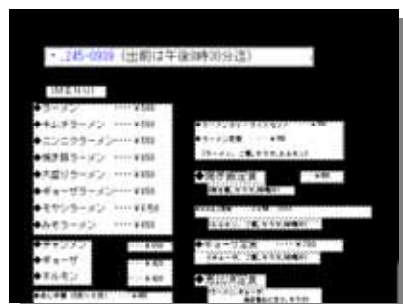
図 4 にメニューの画像キャプチャと認識結果の自動読み上げの概念図を示す。机に置かれたメニューに対して、カメラを 30[cm]程度離れたところで、システムの[Capture]ボタンを押して画像データを取り込む(図 4 (a))。画像から文字領域候補を抽出し、領域毎に文字認識処理を行う(図 4 (b))。得られた文字認識結果より、文字枠(文字サイズ)の大きさと位置情報を基にメニュー項目をブロック毎に分け、項目毎に階層構造のデータベースを構築する(図 4 (c))。

4.2 メニュー情報の音声出力提示法

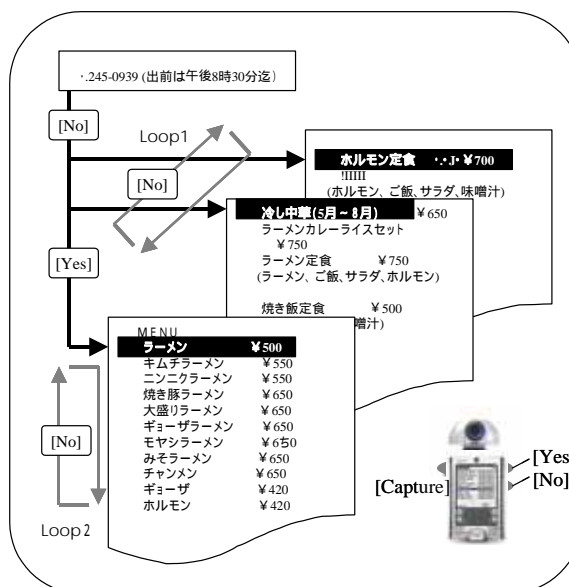
典型的なメニューとして、メニューの分類が上部



(a) Text aria extraction



(b) A result of character recognition



(c) Database Restructuring for Menu item reading system

図 4 メニューの画像キャプチャの一例(メニュー項目の抽出と読み上げ)

Fig. 4 Examples of a menu item extraction from a captured menu list image .

にあり、メニューの名前、説明及び価格が順に並んでいるものを基本形とする。このような構造配置を利用して、文字枠情報と配置座標、インデックス記号(先頭に規則的に出現する記号)から自動的に階層的なメニューを作成する。ここで作成されたメニューデータベースを基にトップダウンにカテゴリ項目毎に音声出力ソフトによって項目を読み上げ、キー入力インタフェースの[Yes]/[No]キーを利用して、ユーザ自らが選択し、メニュー項目を絞り込むことによって内容や価格を知ることができる。また、階層毎に音声読み上げの声を「男性音」と「女性音」を交互に切り替えるようにした。

図4(c)を用いて「みそラーメン」を探す選択例を示す。

- (1) ユーザが、読み取りたいメニューに対してカメラを向けてシステムの[Capture]ボタンを押す。
- (2) 「現在、メニューリストを作成中」の音声メッセージが出力される。
- (3) 「メニュー項目を読み上げます」の音声メッセージが出力される。
- (4) 「Top メニュー、245-0939 (出前は午後 8 時 30 分迄) (男性音)」
- (5) [No]ボタンを押す。
- (6) 「 MENU ラーメン ¥500」(女性音)
- (7) [No]ボタンを押す。
- (8) 「冷し中華(5月~8月) ¥650」(女性音)
- (9) [No]ボタンを押す。
- (10) 「ホルモン定食 ¥700」(女性音)
- (11) [No]ボタンを押す。
- (12) 「 MENU ラーメン ¥500」(女性音)
- (13) [Yes]ボタンを押す。
- (14) 「 MENU ラーメン ¥500」(男性音)
- (15) [No]ボタンを押す。
- (16) 「キムチラーメン ¥550」(男性音)
- ・
- [No]ボタンを押す。(同様にして5回繰返す)
- ・
- (17) 「みそラーメン ¥650」(男性音)
(希望のメニューの価格確認)

5. まとめ

本報告では、視覚障害者のコピキタス情報処理システムの機能の一つとして、レストランなどにおけるメニューリストをカメラで採取し、画像の中から文字情報を抽出し、音声読み上げのための階層構造のデータベースを利用したメニュー自動読み上げシステムの基本構成について述べた。キャプチャ画像からの文字領域の切り出しアルゴリズムの有効性を確認するために、ノートパソコンおよびパンチルトズーム機能付きの Web カメラ(Logitech: QVR-1)を利用したカメラ画像からの文字情報読み上げシステムのプロトタイプを作成し、携帯用ラッ

ブトップ PC を用いたプロトタイプによる文字切り出し機能の実時間による動作を確認した。さらに、オフライン文字認識エンジンを付加して、視覚障害者が単独で文字情報を得たい任意の空間上にカメラを向けて[Capture]ボタンを押すことにより、マイクロカメラのアクティブ機能によりキャプチャされた画面上にある文字列を自動的に走査し、メニュー情報を自動読み上げるシステムの基本構成と音声読み上げ提示法について示した。

今後は、メニューの標準的な基本パターンをさらに拡張して、知識処理を用いた効率の良い音声によるメニュー項目検索法を追加し、被験者実験による評価試験を行う予定である。

6. 参考文献

- [1] 厚生労働省社会・援護局障害保健福祉部 ; <http://www.mhlw.go.jp/houdou/2002/08/h0808-2.html>
- [2] 清田公保, 江崎修央, 柳井貴志, 山本眞司: “視覚障害者のためのオンライン日本語入力インタフェースの基本設計”, 信学論(A), Vol.J79-A, No.2, pp.310-317 (1996).
- [3] 清田公保, 櫻井敏彦, 山本眞司: “ストローク代表点の相対位置関係に基づく視覚障害者用オンライン文字認識”, 信学論(D-II), Vol.J80-D-II, No.3, pp.715-723 (1997).
- [4] 清田公保, 森田智之, 堂下万由美, 山本眞司: “視覚障害者のためのオンライン日本語入力システムの開発”: JJME(医用電子と生体工学), Vol.35, No.4, pp.42-49 (1997)
- [5] 鈴木悠司, 平岩裕康, 竹内義則, 松本哲也, 工藤博章, 山村 毅, 大西 昇, “視覚障害者のための環境内の文字情報抽出システム”, 電気学会論文誌(C), vol.124-C, no.6, pp.1280-1287, 2004-6.
- [6] N. Ezaki, M. Bulacu and L. Schomaker, “Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons”, International conference on pattern recognition 2004
- [7] 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”, International conference on Document Anarysis and Recognition 2005.

謝 辞

本研究を進めるにあたり、適切なお助言と実験にご協力いただいた Groningen 大学の Dr.M. Bulacu 並びに Prof. L. Schomaker. に感謝します。また、本研究の一部は、文部科学省・科研費基礎研究(C)(課題番号: 17500387)の助成を受けた。