

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 10 日現在

機関番号：11301

研究種目：基盤研究（B）

研究期間：2010～2012

課題番号：22300194

研究課題名（和文） ウェアラブルな文字認識視覚補助デバイスに関する研究

研究課題名（英文） Development of Wearable Vision Assistive Device
with Scene Text Recognition Feature

研究代表者

後藤 英昭 (GOTO HIDEAKI)

東北大学・サイバーサイエンスセンター・准教授

研究者番号：40271879

研究成果の概要（和文）：

環境中の紙媒体や看板などの文字を自動的に検出し、文字認識と音声合成によってその情報を視覚障害者に伝えられるような、身につけられる視覚補助デバイスを実現するために、必要とされる要素技術を開発した。従来手法よりも高速・高精度な文字領域自動検出・追跡手法や、リアルタイム文字認識のための高速パターン認識、および、音響信号による文字位置提示手法を開発した。これらを元に試作機を構築し、使用感や改良すべき部分などの知見を得た。

研究成果の概要（英文）：

We have developed some building-block technologies required for realizing wearable vision assistive devices that automatically detect text on paper documents, signboards, etc. in the environment, translate it into speech by Optical Character Recognition and Speech Synthesis, and present the contents to the visually-impaired users. The technologies include fast and precise automatic text region detection and tracking, fast pattern matching for real-time character recognition, and text location guidance using sound signals. We have also developed a prototype system, and obtained the knowledge about the usability and functions need to be improved.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	2,600,000	780,000	3,380,000
2011年度	2,400,000	720,000	3,120,000
2012年度	1,500,000	450,000	1,950,000
年度			
年度			
総計	6,500,000	1,950,000	8,450,000

研究分野：総合領域

科研費の分科・細目：人間医工学・リハビリテーション科学・福祉工学

キーワード：福祉・支援機器，文字認識，パターン認識

1. 研究開始当初の背景

視覚に障害を持つ人々にとって生活における切実な問題として、電子化されていない紙媒体や看板などの文字が読めないことが挙げられる。文字により記号化された情報は人間の生活において重要な役割を占めてお

り、文字情報を利用できないことによる QoL (Quality of Life, 生活の質) の低下は非常に大きい。WHO の調査(2008 年)によると、世界で 3 億人以上が視覚障害をもち、そのうち全盲が約 4,500 万人居るといふ。我が国でも今後、社会的に高齢化が急速に進むに伴い、

視覚の弱った高齢者の補助が不可避であると考えられる。すなわち、先天的あるいは事故・病気等で視覚を失った人々のみならず、社会全体として視覚補助の手段を考えていくことが急務である。

工学的ならびに医学的見地からの視覚補助の手段としては、1. 手術などにより視覚機能そのものの改善を図る方法、2. 埋め込みの撮像デバイスを用いて神経に直接電気信号を加える方法、および、3. 触覚や音声などの代替感覚により画像情報を伝える方法などがある。老化による視覚喪失や先天的な視覚異常には、1の方法では限界がある。2は脳に対する手術が必要であり、高いリスクを伴う。また、世界の視覚障害者の87%が発展途上国に居住しており、高額な費用がかかるこの手法は現実的ではない。3の手法は、伝達できる情報量がやや限られるという問題があるが、非侵襲であり安全性が高いという、他に代えがたいメリットを有しており、コストの面でも有利である。オランダの研究グループによる「The vOICe」では音響信号を、国内開発の「オーデコ」では電気刺激による触覚を用いている。しかしながら、画像から音・触覚への変換では分解能が低く、文字のように細かい対象物を識別することは困難である。文字情報を利用者に伝えるには、文字認識(OCR)により画像中の文字を認識した後、音声合成で読み上げるといった方法が必要となる。

学会では近年、カメラで撮影された看板や文書の認識に関心が集まっているが、看板と印刷文書をシームレスに扱え、しかもリアルタイム処理を目指したものは見あたらない。文字認識が可能なカメラ付き携帯電話などは市販されているが、このようなデバイスでは晴眼者が認識対象にカメラを向け、注意深くカメラの撮影状況を調整しなければならず、全く視覚の無い者にとっては実用的ではない。様々な環境下で、利用者の自由な行動に伴って文字情報を利用できるようにするためには、リアルタイムで動作し、身に付けられる(ウェアラブルな)視覚補助装置に適した処理の開発が必要である。

2. 研究の目的

視覚障害者に自立的な文字識別・理解の能力を提供し、QoL(生活の質)の著しい向上を図るために、ウェアラブルな文字認識視覚補助デバイスの実現を目指して、看板や文書等のシーン文字のリアルタイム抽出やトラッキング(追跡)などの要素技術を開発し、理論的に考察を行う。実際にシステムを試作して、物の見え方やデバイスの使われ方、および開発した手法の性能を調査・評価することによって、学術的知識の蓄積、および早期製品化のための技術蓄積・提供を行う。

3. 研究の方法

2010年度は、ウェアラブルな画像記録システムによる画像データの収集と分析、シーン文字抽出・認識に関する種々の手法の調査、および、既開発のシーン文字領域検出・追跡(トラッキング)手法の改良などを行う。

まず、USBカメラなどの小型カメラとヘッドマウントディスプレイ(HMD)、超小型PCを用いて、ウェアラブルな画像記録システムを構築する。視覚障害者と晴眼者の両方による利用を想定して、実際の利用状況に近い環境で動画データを収集する。ヘッドマウントカメラによる物の見え方や、利用者の動きの傾向などを調査し、ヘッドマウントカメラ特有の問題点や、ウェアラブルデバイスに必要なとされる技術要件の洗い出し・整理を行う。

我々の研究グループで既開発のロボット用のシーン文字検出・追跡手法などをベースとして、より高精度で安定した手法の開発を行う。まずは、ロボットに搭載のカメラのように、動きがよくコントロールされた環境で実用的となる手法を開発することで、福祉用などの自律ロボット向けの文字認識機能の確立を目指す。

さらに、当研究室で開発した既存手法を基礎として、よりシーン文字の抽出に適した動的二値化手法または多値しきい値処理手法を開発し、性能評価を行う。日本語文字認識では3千字種以上の識別が必要で、欧文より処理時間が必要であるが、ウェアラブルカメラではリアルタイムな文字認識が必要と考えられることから、高速・高精度な日本語文字認識手法の開発も手掛ける。

2011年度には、ヘッドマウントカメラに適したシーン文字検出・追跡手法の開発を行い、視覚補助デバイスのプロトタイプを作成し、その性能評価を行う。高精細動画像も扱えるようなワイドレンジで、ロバスタな処理を開発する。最終年度には、視覚障害者向けの視覚補助デバイスを試作するとともに、フィールドテストを通しての使いやすしいユーザインタフェースの開発や、システム全体の性能評価を行う。

まず、前年度の開発成果を基に、ヘッドマウントカメラに適した、照明むらに強く、文字行の幾何変形にもロバスタな、シーン文字検出・追跡手法を開発する。既開発の高速文字行抽出手法をベースとして、より広い文字サイズに適應でき、精度も高い、リアルタイム文字行抽出処理を開発する。続いて、前年に開発した手法を基に、高速・高精度な文字領域トラッキング手法を開発する。

全盲の視覚障害者自身がシーン中の文字領域を見つけ、カメラの視野中央に入れて撮影できるようにするため、音響信号を用いた

ガイダンス機能を開発する。

視覚障害者により利用される文字・音声翻訳眼鏡などのアプリケーションを想定して、視覚による情報提示に頼らないユーザインタフェースを開発する。ヘッドマウントカメラとヘッドホンを装備したウェアラブルな視覚補助デバイスのプロトタイプを作成する。文字認識には、市販の文字認識ライブラリを用いる他、我々が既開発の文字認識サーバ(ウェブベース)も利用できるようにする。音声合成には、オープンソースのソフトウェアや、市販の音声合成ライブラリを利用する。

試作したデバイスの性能とユーザインタフェースの使い勝手をフィールドテストによって評価する。テスト結果より、改良が必要な点や、新たに得られた知見をまとめる。その結果を元にして、デバイスとインタフェースの改良を行う。また、当研究グループで開発した視覚補助デバイスを一般利用者のモニターに供し、使用感や要望を聞き取り調査し、改良が必要な点などを明らかにする。

4. 研究成果

視覚障害者に自立的な文字識別・理解の能力を提供し、QoL(生活の質)の向上を図るために、ウェアラブルな文字認識視覚補助デバイスの実現を目指して、看板や文書等のシーン文字のリアルタイム抽出やトラッキング(追跡)、高速文字認識などの要素技術を開発し、理論的・実験的な考察を行った。主たる成果を以下に示す。

(1) 高速・高精度文字領域抽出・トラッキング(追跡)手法の開発

我々の研究グループで既開発の、ウェアラブルカメラ用の文字抽出・トラッキング手法(2008, 2009年発表)をベースとして、より高精度で安定した手法の開発を行った。初年度は、パーティクルフィルタにおける類似度評価に、色分布とエッジ方向ヒストグラムを組み合わせた Bhattacharyya 距離を導入することによって、文字領域追跡の精度を大幅に改善し、同一の文字領域から生じる重複した画像の数を減らすことができた。屋内環境では、文字認識処理にかかる文字候補画像の数を、従来比で約 41%まで削減できた。

続いて、一般物体の追跡に使用されているピラミッド型 Lucas-Kanade 法を用いて、まず画像中の特徴点を追跡した上で、その結果に文字領域検出結果を加味し、文字領域を選択的に追跡する手法を開発した。文字領域検出を先に行う従来手法と比較して、同一の文字領域から生じる重複した画像の数を約半数(平均約 1.5 枚)に抑えることができた。

さらに、この手法に改良を加え、同一の文字領域から生じる重複した画像の数を約 1/3 に抑えつつ、同約 3 倍の速度を実現した(図

1)。
開発した手法を文字認識視覚補助デバイスに応用することで、重複した音声出力が減り、デバイスの利便性が大幅に改善される。



図 1. 文字領域検出・追跡の様子

(2) 音響信号を用いたシーン文字位置提示手法の開発

視覚障害者が環境中の看板などを見つけ、認識に適した文字画像を撮影できるようにするため、検出された文字候補領域を仮想的音源として、三次元音響によって利用者に提示する手法を開発した(図 2)。上下方向の位置に応じて周波数を変化させる手法を頭部伝達関数(HRTF)と組み合わせることで、HRTF 単体と比較して、文字位置の正答率を 2.6 倍(約 75%)まで大幅に改善できた。

次に、上下方向の位置を提示するのに、HRTF を用いた従来の音場再生型に代わり、基準音程を付加した記号的な音響信号を用いる手法を開発した。また、文字領域を大きな順に二領域まで逐次提示する仕組みも取り入れた。これらの改良により、文字位置の正答率を 46.3%(従来法)から 98.3%まで大幅に向上させ、実用的な精度を実現できた。

視覚障害者にとっての聴覚は、周囲の障害物などを探知するための重要な感覚となっていることから、ヘッドホンで耳を塞ぐことは好ましくない。提案手法では骨伝導型ヘッドホンを用いてこの問題に対処した。また、音場よりも記号的表現の方がより正確に文字位置を把握できることが明らかになった。

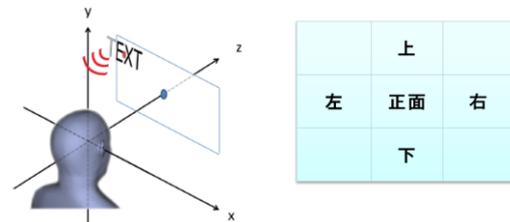


図 2. 音響信号による文字位置提示

(3) 「文字読み上げカメラ」の試作と利便性・課題の分析

(2)のシーン文字位置提示手法に文字認識エンジンと音声合成処理を組み合わせ、「文字読み上げカメラ」のプロトタイプを開発した。このカメラを利用することで、視覚障害者が看板等の文字を撮影し、音声により内容を知ることができる。文字領域の検出に

は離散コサイン変換(DCT)を利用した特徴量(2008年発表済み)を利用しており、ビル内のシンプルな壁などの比較的単純な背景ならば、看板の文字を精度よく検出し、位置を音響信号で知ることが可能である。

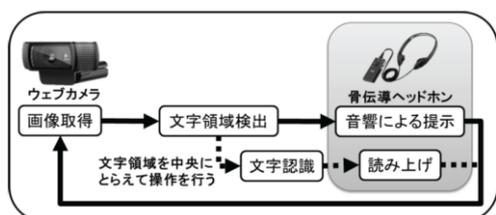


図3. 文字読み上げカメラのプロトタイプ

音響信号の工夫により、文字位置の高い正答率(98.8%)を得た。本プロトタイプを視覚障害者向けイベント「サイトワールド2012」に出展し、大勢の視覚障害者に試用してもらうことで、提案システムの有効性を確認するとともに、今後の改良や発展につながる有益なコメントや知見を多数得ることができた。なお、木々やビルなどの複雑な背景がある環境では文字領域検出の精度が不十分であるが、複雑背景でも高精度かつリアルタイムに動作する手法の開発は今後の課題である。利用者からは、用途が限定的であってもこのようなデバイスを早く利用したいという声が多数聞かれた。

(4) 高速日本語文字認識手法の開発

ビデオレートのリアルタイム文字認識を実現するために、正準判別分析(CDA)と二分探索木を用いて日本語文字認識の高速化手法を開発した。2011年度はさらに効率的な次元圧縮手法を開発、付加することで、前年度開発の手法と比べて約3.6倍の速度を得た。

多クラス判別分析(LDA)と二分探索木を用いて、日本語文字認識の高速化手法の改良を推し進めた。2011年度開発の手法では手書き文字データベース(ETL9B)で速度・認識精度とも大きく低下する問題があったが、改良型の手法ではこの性能低下を抑え、全数整合法と比べて33.9倍の速度を0.24%の精度低下で実現した(図4)。この処理速度は動画像(30fps)の1フレームあたり約198文字に相当し、リアルタイム文字認識の応用につながる。

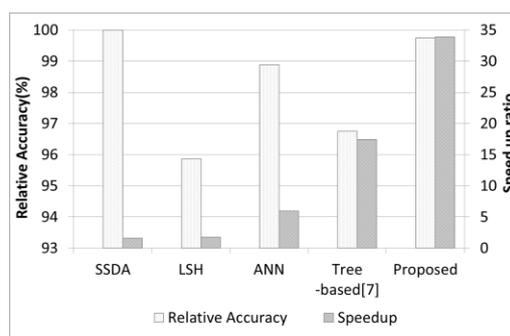


図4. 高速化手法の性能比較

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)

[学会発表] (計16件)

[1] 斎藤 亮, 後藤英昭, “視覚障害者のための音響による情景文字情報提示・読み上げシステム,” 電子情報通信学会2013年総合大会講演論文集 A-19-4, p.265 (2013.3.20, 岐阜大学).

[2] 佐々木崇裕, 後藤英昭, “高速日本語文字認識のための線形判別分析を用いた高精度クラスタリング手法,” 信学技報 パターン認識・メディア理解 PRMU2012-73, pp.19-24 (2012.12.13, 山形大学).

[3] Faustin Pegeot and Hideaki Goto, “Scene Text Detection and Tracking for a Camera-equipped Wearable Reading Assistant for the Blind,” The 11th Asian Conference on Computer Vision (ACCV2012), Workshop on Detection and Tracking in Challenging Environments (DTCE) (2012.11.6, Korea). 査読有

[4] Takahiro Sasaki, Akira Saito, and Hideaki Goto, “Wearable Reading Assistant Device with Scene Text Locator for the Blind,” 10th IAPR Workshop on Document Analysis Systems (DAS2012) Short Paper Booklet, pp.1-2 (2012.3.27, Australia). 査読有

[5] 佐々木崇裕, 後藤英昭, “文字列矩形の配列を利用した情景画像中の文字領域抽出,” 電子情報通信学会2012年総合大会講演論文集 D-12-24, p.118 (2012.3.21, 岡山大学).

[6] 斎藤 亮, 佐々木崇裕, 後藤英昭, “視覚障害者のための音響による情景中文字位置提示システム,” 電子情報通信学会2012年総合大会講演論文集 A-19-14, p.315 (2012.3.20, 岡山大学).

[7] 蘇武陽平, 後藤英昭, “高速日本語文字認識のための正準判別分析を用いた高精度クラスタリング手法,” 信学技報 パターン認識・メディア理解 PRMU2011-94, pp. 37-42 (2011. 10. 7, 幕張メッセ).

[8] 後藤英昭, “Six-year Operation and Study of the OCR Web Service Using WeOCR - What have changed and what did not -,” 信学技報 パターン認識・メディア理解 PRMU2011-95, pp. 43-48 (2011. 10. 7, 幕張メッセ).

[9] 佐々木崇裕, 後藤英昭, “文字矩形の配列を利用した情景画像中の文字領域検出,” 平成 23 年度電気関係学会東北支部連合大会 1D05, p. 121 (2011. 8. 25, 東北学院大学).

[10] Yohei Sobu and Hideaki Goto, “Binary Tree-Based Accuracy-Keeping Clustering Using CDA for Very Fast Japanese Character Recognition,” Proceedings of the 12th IAPR Conference on Machine Vision Applications (MVA2011), pp. 299-302 (2011. 6. 14, Japan). 査読有

[11] 佐々木崇裕, 後藤英昭, “視覚障害者のための音響による情景中文字探知システム,” 電子情報通信学会 2011 年総合大会講演論文集 A-19-11, p. 291 (2011. 3. 14, 東京都市大学).

[12] Binghua Wang and Hideaki Goto, “Scene Text Detection and Tracking for Wearable Camera System,” 信学技報 PRMU2010-156, pp. 47-52 (2011. 1. 20, 立命館大学).

[13] Yohei Sobu, Hideaki Goto, and Hirotomo Aso, “Binary Tree-Based Precision-Keeping Clustering for Very Fast Japanese Character Recognition,” Proceedings Image and Vision Computing New Zealand 2010 (IVCNZ 2010) (Nov. 8, 2010, New Zealand). 査読有

[14] Hideaki Goto, “A Feasibility Study on Distributed Cooperative OCR Systems Using WeOCR,” Proceedings of the IADIS International Conference WWW/INTERNET 2010, pp. 42-48 (Oct. 16, 2010, Romania). 査読有

[15] Binghua Wang and Hideaki Goto, “Scene Text Detection and Tracking for Wearable Camera System,” 平成 22 年度電気関係学会東北支部連合大会 2A20, p. 39 (2010. 8. 27,

八戸工業大学).

[16] 蘇武陽平, 後藤英昭, 阿曾弘具, “高速文字認識のためのパターン辞書構成法に関する研究,” 平成 22 年度電気関係学会東北支部連合大会 1D12, p. 115 (2010. 8. 26, 八戸工業大学).

[その他] (計 2 件)

[1] 「文字読み上げカメラ」プロトタイプ出展, 視覚障害者向けイベント「サイトワールド 2012」(2012 年 11 月, 東京).

[2] 文字認識ウェブサービス WeOCR ホームページ, <http://weocr.ocrgid.org/>

6. 研究組織

(1) 研究代表者

後藤 英昭 (GOTO HIDEAKI)

東北大学・サイバーサイエンスセンター・准教授

研究者番号 : 40271879