

令和 3 年 6 月 22 日現在

機関番号：24403

研究種目：基盤研究(B) (一般)

研究期間：2017～2019

課題番号：17H01803

研究課題名(和文) 全方位カメラを用いた周辺情報認識による視覚障害者支援

研究課題名(英文) Assistance for People with Visual Impairment by Recognizing Surrounding Information Using an Omnidirectional Camera

研究代表者

岩村 雅一 (Iwamura, Masakazu)

大阪府立大学・大学院工学研究科・准教授

研究者番号：80361129

交付決定額(研究期間全体)：(直接経費) 13,600,000円

研究成果の概要(和文)：本研究では、コンピュータに文字や物体を認識させる従来の視覚障害者補助システムの拡張を試みた。従来のシステムには、認識したいものが視覚障害者の目の前にあるという暗黙の前提があった。しかし、視覚障害者が知りたいものは目の前にあるものだけに限らない。そこで、視覚障害者にとって位置の分からないものも認識出来る枠組みの一例として、全方位カメラを用いる探し物探索システムを提案した。視覚障害者に参加してもらった実験により、全方位カメラを用いる方が従来のスマートフォン型のカメラよりも使いやすいことを確認した。また、物体認識の認識精度を向上する手段として、深層学習に用いるShakeDropという正則化を提案した。

研究成果の学術的意義や社会的意義

コンピュータに文字や物体を認識させて視覚障害者の目の代替を目指す試みが盛んになっている。既存の方法は、利用者である視覚障害者が目の前にあるものをスマートフォン付属のカメラで撮影して、それが何であるかを伝えるものが主流である。しかし、視覚障害者が知りたいものは目の前にあるものだけに限らないのではない。この考えから、周囲が一度に撮影できる全方位カメラを用いたシステムを考案した。しかし、撮影範囲が増えると、認識で得られる情報も増える。時として膨大な認識結果を全て視覚障害者に伝えれば、逆に混乱を招く恐れがある。そこで本研究では、全方位カメラを用いる場合のより良い認識結果の提示方法などを検討した。

研究成果の概要(英文)：In this research, we attempted to extend the conventional assistive system for the visually impaired, which allows a computer to recognize text and objects. The conventional systems have the implicit assumption that the target object or text to be recognized is right in front of the user. However, what a visually impaired person wants to know is not limited to be in front of the person. Therefore, we proposed an assistive system for "looking for something" using an omnidirectional camera as an example of a framework that can recognize objects whose location is unknown to the user. In this context, we confirmed that the omnidirectional camera is easier to use than the conventional smartphone-type camera through experiments with visually impaired people. We also proposed a regularization method called ShakeDrop for deep learning to improve the recognition accuracy of object recognition.

研究分野：パターン認識、視覚障害者支援

キーワード：視覚障害者 全方位カメラ 視覚情報 物体検出 探し物 深層学習

1. 研究開始当初の背景

「文字認識技術を使用した視覚障害者の補助」というアイディアはかなり昔から多くの人が考えてきたため、ある意味古典的な研究テーマである。しかし、それでもなお、このテーマの研究が必要とされるのは、認識技術が視覚障害者の暮らしにほとんど役立っていないからである。文字認識技術を使用して視覚障害者を補助する最近の研究を図1にまとめる。[A]～[D]のうち、[A]と[B]は視覚障害者自身が文字の位置を特定できると仮定している。[A]は、文字の場所はわかるが読めない、失語症の人を対象としているので、この仮定は妥当だが、全盲の人には使えない。一方、文字行をうまく指でなぞれるように工夫した[B]は、そもそもどこに文書があるのかがわからないと使えないという点でお粗末である。[C]と[D]は文字が視覚障害者の正面にあると仮定する。この仮定が成り立つ場面もあるが、歩行時には使えないなど、使用場面が限定される。



図1: 文字認識技術を使用して視覚障害者を補助する最近の研究

そこで我々は、既存研究にありがちな、「視覚障害者自身に文字を指し示してもらおう」という考えを排除して、図2に示す視覚障害者補助システムを作成した[10]。これは小型の全方位カメラで利用者周辺の画像を撮影し、その中からコンピュータが文字を発見して、読み上げて利用者に伝えるウェアラブルデバイスである。立体音響を使用して、読み上げた音声は文字のある方向から聞こえるという工夫を施した。学会で最初にこれを発表した際に、耳を覆うヘッドホンではなく、骨伝導イヤホンを使うことを勧められ、骨伝導イヤホンを使うバージョンも作成して発表した[9]。

後者の発表をきっかけに、自らが視覚障害者であり、かつ工学の博士号を持つ企業研究者2名と知り合うことができた。彼らと話をする中で、晴眼者には思いもよらない事情が見えてきた。一番驚いたのは、生まれつき全盲、もしくは早いうちに視力を失なった人は、ずっと文字に頼らずに生活してきたので、「そもそも文字情報がどう役に立つのか見当がつかない」という意見をもらったことであった。これは、文字認識研究者が抱きがちな、「文字さえ(コンピュータで)読めれば視覚障害者は幸せになれる」という考えが幻想であることを痛切に実感する機会となった。さらに、「認識結果を全て伝えられても情報量が多過ぎて困る」といった問題点が浮かび上がってきた。

2. 研究の目的

本研究では、周囲の文字や物体の情報を認識して、利用者である視覚障害者に伝えるシステムの拡張を試みる。従来のシステムには、認識したいものが視覚障害者の目の前にあるという暗黙の前提があった。しかし、視覚障害者が知りたいものは目の前にあるものだけに限らない。そこで、視覚障害者にとって位置の分からないものも認識出来る枠組みの一例として、全方位カメラを用いる捜し物探索システムを提案する。



図2: 申請者らが提案した装着型視覚障害者補助システム[10]

3. 研究の方法

本研究は大きく 3 つの部分で構成される。1 つ目は、視覚障害者がコンピュータに周囲の文字や物体の情報を認識させて周囲の視覚情報を得る場面の 3 分類である。2 つ目は、前述の 3 分類の 1 つの代表例として、探し物を取り上げ、全方位カメラを用いる探し物探索システムを提案する。3 つ目は、物体認識の認識精度を向上する手段として、深層学習に用いる正則化を 2 つ提案する。

4. 研究成果

4. 1 視覚情報取得方法の 3 分類

コンピュータビジョン技術を用いる既存のスマートフォンアプリは、晴眼者が取得している視覚情報の一部を利用者である視覚障害者に提供しているに過ぎない。今後これを拡張するための道標として、視覚障害者がコンピュータビジョン技術を用いて視覚情報を取得するタスクを 3 つに分類 (カテゴリ化) する。

スマートフォンアプリの利用目的はその物体が何かを知ることであり、物体の写真を撮影することで実現できる。ただし、そのために利用者は物体の位置を把握しなければならない。従って、「物体が何か」(What と表記する) が未知で、「物体がどこにあるか」(Where と表記する) が既知な場合のみ、スマートフォンアプリが使用できる。この考え方を拡張することで、コンピュータビジョン技術を用いて取得できる視覚情報は以下の 3 つに分類できる。

- **カテゴリ (i) : What が未知、Where が既知**
このカテゴリの視覚情報は、関心のある物体がどこにあるか分かっているため、利用者が自力で撮影できる。そのため、[1, 5, 4] のようなコンピュータビジョン技術を用いる既存のスマートフォンアプリで取得できる。
- **カテゴリ (ii) : What が既知、Where が未知**
このカテゴリの視覚情報を使った代表的なタスクは探し物である。探し物の際には、利用者は何を探しているかを知っているが、その場所が分からない。利用者は位置が分からない探し物 (関心のある物体) の写真を撮れないため、カテゴリ (i) と同様にスマートフォンアプリを利用できない。もし無理矢理にでも写真を撮ろうとするならば、スマートフォンを様々な方向に動かしながら写真や動画を撮影し、偶然探し物が写るのを期待するのみであろう。そのため、利便性の向上のために、魚眼カメラや全方位カメラなどの広い視野を持つカメラデバイスを用いる方が良いと考えられる。カテゴリ (i) とは異なり、利用者が何を探しているか分かっているため、スマートフォンアプリには探し物の位置を知らせることが期待される。
- **カテゴリ (iii) : What、Where とともに未知**
このカテゴリでは、利用者は関心のある物体が何か、それがどこにあるのかのどちらも分かっていない。つまり、利用者がその視覚情報の存在に気づいていない場合である。そのため、利用者はスマートフォンアプリが利用者に視覚情報を提供することを期待していない。しかし、もし提供されれば、その視覚情報は利用者にとって有益であることが期待される。概念的には、利用者が潜在的に興味を持っていると思われる反面、利用者にとって予想外の商品を提示することが期待されるという点で、Amazon.com などの e-commerce サイトで用いられる情報推薦システムに似ている。そのため、このカテゴリの代表的なタスクは利用者にとって価値があるものの予想外の視覚障害を見つけることである。このカテゴリの視覚障害を提供するシステムが実際に使われるとすれば、スマートフォンアプリが利用者の周囲にある視覚情報をできるだけ多く得ることが求められる。そのため、カテゴリ (ii) と同様、広い視野を持つカメラデバイスを用いる方が良いと考えられる。他のカテゴリと大きく異なる点は、スマートフォンアプリが提供する視覚情報の量が過多になる可能性があることである。すなわち、スマートフォンアプリが利用者にとって価値のある視覚情報を大量に見つけてしまう可能性がある。多量の情報はむしろ混乱を生むだけであるため、利用者に提供する視覚情報を適切に取捨選択する必要がある。

4. 2 全方位カメラを用いる探し物探索システム

ここでは、4.1 で述べた 3 つのカテゴリの中からカテゴリ (ii) に注目し、カテゴリを代表するタスクである探し物について、適切なカメラシステムと回転ナビゲーション手法を議論する。

4. 2. 1 プロトタイプシステム

探し物タスクにおいて、コンピュータビジョン技術を基盤とするプロトタイプシステムを実装する。これは、利用者が対象物体に以下の 3 つのステップで到達できるように誘導する。

- Step 1: 物体検出
- Step 2: 回転ナビゲーション
- Step 3: 前進ナビゲーション

表 1: 回転ナビゲーション手法の 5 段階評価

回転ナビゲーション手法		A	B	C	D	E	F	G	平均
音声	左右のみ (LR)	4	2	3	2	1	4	2	2.6
	目標の方向への角度 (AG)	4	4	3	5	3	4	5	4.0
	クロックポジション (CP)	5	2	5	3	4	5	3	3.9
音	断続的なビープ音 (IB)	3	5	4	4	4	3	5	4.0
	音程の高低 (PT)	2	4	4	4	5	2	4	3.6

実装したプロトタイプシステムは 2 種類あり、2 種類のカメラシステムのいずれかとノートパソコン (MacBook Pro) から構成される。1 つ目のカメラシステムは全方位カメラと電子コンパス、深度カメラから成る。2 つ目のカメラシステムは、疑似スマートフォンである。スマートフォンの内蔵カメラではなく、ウェブカメラを使用した。2 つのカメラシステムの違いはカメラだけであり、公平な比較のために、いずれのカメラシステムでも同一の電子コンパスと深度カメラを使用した。

4. 2. 2 提案する回転ナビゲーション

以下の通り、本論文では、以下の 5 手法を考察する。最初の 3 つは声によるナビゲーション手法であり、後の 2 つは音によるナビゲーション手法である。

左右のみ (Left or Right; LR) は、目標の方向が右なのか、左なのかを約 1.5 回/秒の間隔で繰り返し利用者に伝える手法である。ただし、目標の方向が利用者から 15 度以内に近づいたときは「正面」と伝える。

目標の方向への角度 (Angle; AR) は、LR のように目標の方向が左か右かを伝えた後に、目標の方向までの相対的な回転角度を繰り返し伝える。ただし、利用者の正面は常に 0 度とする。LR と同様に、目標の方向から 15 度以内のとき、「正面」と伝える。

クロックポジション (Clock Position; CP) は、AG と似ているが、角度では無く、クロックポジションを用いるところが異なる。目標の方向から 15 度以内のとき、「正面」と伝える。

断続的なビープ音 (Intermittent Beep; IB) は、角度の違いに反比例した間隔の断続音を用いる方法である。対象物体が 15 度以内、つまり利用者の正面に近づいたとき、両耳のイヤホンに約 8Hz のビープ音を再生する。

音程の高低 (Pitch; PT) は、音の高さの変化により目標の方向までの角度を伝える手法である。目標の方向の 15 度以内は IB と同じである。

4. 2. 3 実験

視覚障害者 7 名に参加してもらい、実験を行った。参加者の内訳は男性が 4 名、女性が 3 名であり年齢は 23 歳から 48 歳であった。7 名のうち、6 名は全盲で、残り 1 名は弱視であった。実験では以下の 4 つの項目を順に実施した。

実験 1: 5 つの回転ナビゲーション手法の比較

この実験では、4.2.1 の Step 1 (全方位カメラを用いた物体検出) と Step 2 (回転ナビゲーション) に従って、参加者に 5 つのナビゲーション手法をひとつずつ使ってもらった。認識対象としては、ノートパソコンやペットボトルのような検出しやすい物体カテゴリを指定して、室内に 1 つだけ置いた。手法を公平に比較するため、イヤホンの使用を前提にしている IB と PT だけでなく、全てのナビゲーション手法で参加者にイヤホンを使用してもらった。表 1 に参加者による各ナビゲーション手法の主観評価を示す。評価は 5 段階評価で行われ、大きな数字は高い評価を表す。表 1 から、ナビゲーション手法に関する参加者の評価が多様であることがわかる。すなわち、LR を除く 4 つのナビゲーション手法をそれぞれ最も高く評価した参加者が 1 名以上いた。

実験 2: カメラデバイスの選択

実験参加者に 2 つのカメラシステムをそれぞれ使ってもらい、4.2.1 で述べた 3 段階の手順に従って捜し物をしてもらった。その際、実験 1 で試したナビゲーション手法の中で、各実験参加者が最も良いと判断したナビゲーション手法を使用してもらった。イヤホンの使用は実験参加者の判断に任せた。表 2 に 2 つのカメラシステムに対する参加者の主観評価を示す。表 2 から、1 名 (参加者 C) は疑似スマートフォンの方が良いと、残りの 6 名は全方位カメラの方が良いと評価したことがわかる。これらの結果より、探し物において全方位カメラを使用する利点を確認できた。

表 2: カメラデバイスの 5 段階評価

カメラ	A	B	C	D	E	F	G	Average
全方位カメラ	4	4	4	5	5	3	5	4.3
疑似スマートフォン	3	2	5	3	3	1	2	2.7

4. 3 深層学習に用いる新しい正則化

PyramidSepDrop と ShakeDrop という 2 種類の正則化を提案した。

前者は PyramidNet [2] というニューラルネットワークに特化した手法である。その詳細は [11] に譲る。

後者は PyramidNet だけでなく、ResNet [3]、ResNeXt [6]、Wide ResNet [8] にも適用可能な手法である。その詳細は科研費 挑戦的研究 (萌芽) 2018 年度採択課題 「学習の阻害による深層学習の学習促進」*1 の報告書あるいは [7] に譲る。

参考文献

- [1] URL: <https://www.letsenvision.com/>.
- [2] Dongyoon Han, Jiwhan Kim, and Junmo Kim. “Deep Pyramidal Residual Networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- [4] URL: <https://www.microsoft.com/en-us/ai/seeing-ai>.
- [5] URL: <https://taptapseeapp.com/>.
- [6] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. “Aggregated Residual Transformations for Deep Neural Networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [7] Yoshihiro Yamada, Masakazu Iwamura, Takuya Akiba, and Koichi Kise. “ShakeDrop Regularization for Deep Residual Learning”. In: *IEEE Access* 7 (2019), pp. 186126–186136. DOI: 10.1109/access.2019.2960566.
- [8] Sergey Zagoruyko and Nikos Komodakis. “Wide Residual Networks”. In: *Proceedings of the British Machine Vision Conference (BMVC)*. 2016.
- [9] 岩村 雅一, 宮田 武嗣, 程 征, 田井中 溪志, 黄瀬 浩一. “視覚障害者のための環境文字情報提示システムの検討”. In: *情報処理学会 研究報告アクセシビリティ (AAC)*. Vol. 2016-AAC-1. 7. July 2016. URL: <http://id.nii.ac.jp/1001/00169969/>.
- [10] 宮田 武嗣, 岩村 雅一, 黄瀬 浩一. “立体音響で教える全方位単語感知システム”. In: *電子情報通信学会技術研究報告, PRMU2015-194*. Vol. 115. 517. Mar. 2016, pp. 179–184.
- [11] 山田 良博, 岩村 雅一, 黄瀬 浩一. “PyramidNet における確率的な正則化の効果の検証”. In: *電子情報通信学会技術研究報告*. Vol. 116. 461. Feb. 2017, pp. 35–40.

*1 <https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-18K19785/>

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Yamada Yoshihiro, Iwamura Masakazu, Akiba Takuya, Kise Koichi	4. 巻 7
2. 論文標題 Shakedown Regularization for Deep Residual Learning	5. 発行年 2019年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 186126 ~ 186136
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/access.2019.2960566	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計11件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 谷口 善信, 岩村 雅一, 黄瀬 浩一
2. 発表標題 タッチレスなタッチパネル操作デバイスの提案
3. 学会等名 インタラクション2020
4. 発表年 2020年

1. 発表者名 河合隆哲, 岩村雅一, 黄瀬浩一
2. 発表標題 視覚障害者のための周辺情報選択システム
3. 学会等名 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM)
4. 発表年 2019年

1. 発表者名 Masakazu Iwamura, Yoshihiko Inoue, Kazunori Minatani, Koichi Kise
2. 発表標題 Suitable Camera and Rotation Navigation for People with Visual Impairment on Looking for Something Using Object Detection Technique
3. 学会等名 17th International Conference on Computers Helping People with Special Needs (ICCHP2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Yoshihiro Yamada, Masakazu Iwamura, Koichi Kise
2. 発表標題 ShakeDrop Regularization
3. 学会等名 Proc. 6th International Conference on Learning Representation (ICLR) Workshop (国際学会)
4. 発表年 2018年

1. 発表者名 岩村 雅一, 南谷 和範, 井上 慶彦, 程 征, 内海 ゆづ子, 黄瀬 浩一
2. 発表標題 全方位カメラと認識技術を用いた視覚障害者支援システム
3. 学会等名 情報アクセシビリティをめぐる諸問題に関する研究集会
4. 発表年 2019年

1. 発表者名 山田良博, 岩村雅一, 黄瀬浩一
2. 発表標題 ResNetsに対する新たな正則化手法ShakeDropの提案
3. 学会等名 第21回画像の認識・理解シンポジウム(MIRU2018) 予稿集
4. 発表年 2018年

1. 発表者名 井上 慶彦, 岩村 雅一, 黄瀬 浩一
2. 発表標題 全方位カメラを用いた物体検出とトラッキング - 視覚障害者支援システムの実現に向けて -
3. 学会等名 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM)
4. 発表年 2018年

1. 発表者名 Masakazu Iwamura, Naoyuki Morimoto, Keishi Tainaka, Dena Bazazian, Lluís Gomez, Dimosthenis Karatzas
2. 発表標題 ICDAR2017 Robust Reading Challenge on Omnidirectional Video
3. 学会等名 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (国際学会)
4. 発表年 2017年

1. 発表者名 山田 良博, 岩村 雅一, 黄瀬 浩一
2. 発表標題 PyramidNetに対する新たな確率的正則化手法ShakeDropの提案
3. 学会等名 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU) 研究会
4. 発表年 2017年

1. 発表者名 Yoshihiro Yamada, Masakazu Iwamura, Koichi Kise
2. 発表標題 Deep Pyramidal Residual Networks with Stochastic Depth
3. 学会等名 画像の認識・理解シンポジウム217
4. 発表年 2017年

1. 発表者名 山田良博, 岩村雅一, 黄瀬浩一
2. 発表標題 画像分類精度を向上させる追加タスクの検討
3. 学会等名 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU) 研究会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------