

令和 5 年 6 月 12 日現在

機関番号：24405

研究種目：基盤研究(B) (一般)

研究期間：2020～2022

課題番号：20H04212

研究課題名(和文) 視覚障害者のための物体認識・検出システム

研究課題名(英文) Object Recognition/Detection System for People with Visual Impairment

研究代表者

岩村 雅一 (Iwamura, Masakazu)

大阪公立大学・大学院情報学研究科 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)：周囲の視覚情報を視覚障害者に提供するスマートフォンアプリが普及している。これらの多くは視覚障害者自らが写真を撮ることを前提としているが、視覚障害者にとって写真撮影は困難である。そのため、我々は視覚障害者が自ら写真を撮影しなくても良い「受動的な情報取得」を提案した。この枠組みを導入することで、視覚障害者は能動的に写真撮影する必要はなくなるものの、膨大な周囲の視覚情報が取得できなくなるため、本研究では、視覚情報の要約方法を検討した。まずは視覚情報を意味的や時間的に要約する手法ならびに質問応答により回答する手法を検討した。さらに、個人毎に違った情報を提示する要約手法と棒グラフの要約手法を検討した。

研究成果の学術的意義や社会的意義

視覚障害者向けのスマートフォンアプリは、これまで視覚障害者が独力で出来なかったことを出来るようにする大きな可能性を秘めている。特に写真を撮影したり、カメラを向けることにより、対象の情報を得るアプリは、認識技術の発展に伴い、大きな役割を果たしている。その反面、写真撮影に困難を感じている視覚障害者は少なくないが、この問題はこれまで見過ごされてきた。本研究では、このような対象にカメラを向ける必要があるスマートフォンアプリの難点に焦点を当てて、解決方法を模索して先鞭を付けた。

研究成果の概要(英文)：Smartphone apps that provide people with visual impairment with visual information about their surroundings have become widespread. Many of these apps are based on the premise that people with visual impairment take pictures by themselves, but for people with visual impairment, taking pictures is one of the difficult things. Therefore, we proposed a "passive information acquisition" that does not require people with visual impairment to take pictures by themselves. By introducing this framework, people with visual impairment do not have to actively take pictures, but they can acquire a huge amount of visual information about their surroundings. In this research, we first examined methods for summarizing visual information semantically and temporally as well as methods that accept user's questions. In addition, we examined a method that summarizes custom information for each individual and a method that summarizes bar charts.

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 視覚情報 受動的な情報取得 視覚情報 全方位カメラ 要約 質問応答 情報推薦 棒グラフ

1. 研究開始当初の背景

人は多くの情報を視覚から得ており、その割合は8割とも言われる。ただしこれは、「目が見える人(晴眼者)」に限った話である。全盲(視力が全くない)や弱視(視野に欠損があったり、視力が著しく悪い)というハンディキャップを負った、いわゆる視覚障害者は、視覚から十分な情報が得られないため、視覚以外の感覚器で外界の情報を受け取ったり、ボランティアなど周囲の人たちの援助を受けて、足りない情報を補っている。

この状況を改善するため、パターン認識技術を用いて視覚障害者の「目」を作る取り組みが盛んに行われている。特に最近では、TapTapSeeやMicrosoftのSeeing AIなどのスマートフォンのアプリを用いた物体認識や文字認識サービスにより、視覚障害者は独力で目の前の「もの」が何か分かるようになってきた。

これらのサービスが視覚障害者の生活を改善していることに疑いの余地は無いが、これによって視覚障害者がアクセスできるようになったのは、周囲の視覚情報の一部に過ぎない。実はスマートフォンアプリによって認識できるのは、認識したい対象がどこにあるかわかっている場合のみである。換言すれば、Whereは既

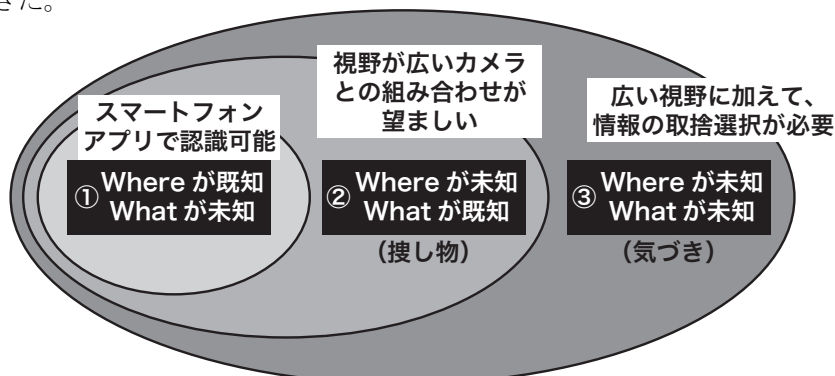


図1: 視覚障害者がパターン認識技術を利用して獲得できる周囲の視覚情報

知だが、Whatは未知な場合のみである。どこにあるかわからないものは、そもそも認識対象にならず、認識技術の恩恵を受けることはできない。しかし、図1に示すように、我々の日常には前述のどこにあるかわかるもの(図中の①)だけでなく、どこにあるかわからないもの(図中の②、③)が多数存在する。図中の②はWhereが未知でWhatが既知なパターンであり、図中の③はWhatもWhereも未知なパターンである。

②番のパターンは、特定のものを探し出すタスクに該当する。形があり、触れられるものであれば手探りでも探せるが、文字を読まないと判断がつかない場合や探索範囲が広い場合などは、認識技術に頼らざるを得ない。以前学会で、建物の廊下に居る視覚障害者がスマートフォンの文字読み上げソフトを用いて、尋ねたい部屋を探すというデモ動画を見たことがある。まさしく探し物(探し人)である。部屋には表札があり、視覚障害者が表札に書いてある人名を一部屋ずつスマートフォンでチェックする様子が動画に収められていた。各部屋の表札の正確な場所はわからないので、視覚障害者が手探りで探し当てていた。労を惜しまなければ、時間を気にしなければ、確かにこの方法で目的の部屋を探し出すことは可能であるが、非現実的である。

③番のパターンは、予期せず意味のある情報に遭遇する場合であり、ユーザにとっての「気づき」が生じる場合と考えられる。道を歩いていて、新しい店が出来ているのに気付くとか、突然自転車に迫ってきて、危険を感じるというのもこれに該当する。

②番と③番のパターンでも視覚障害者が認識技術を用いて周囲の視覚情報にアクセスできるようにするにはどうすれば良いだろうか? 2つのことがその答えになると考えている。1つはカメラデバイスである。一度に広い範囲の情報を認識技術で調べるために、スマートフォンのカメラのように視野が狭いカメラではなく、魚眼レンズ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利用することが望ましい。もう1つは、情報処理機構である。一度に得られる情報が多いということは、不要な情報も多いということである。したがって、得られた情報の中から不要なものを捨て、本当に必要なものだけを残す取捨選択が必要になる。その実現の鍵は晴眼者の視覚情報処理機構にあると考えている。晴眼者は多量の視覚情報が目から入っても苦にならない。それは、目から入った視覚情報は視野で自動的に取捨選択されるためである。同様の機構が実現できれば良い。

2. 研究の目的

本研究の目的は、認識結果をいかに視覚障害者に伝えるかという問題に焦点を当て、解決を試みることである。現在利用可能なスマートフォンアプリを用いた認識サービスは、利用者が指定した対象のみを認識する。しかし、視覚障害者に、より多くの情報に接する機会を提供するに

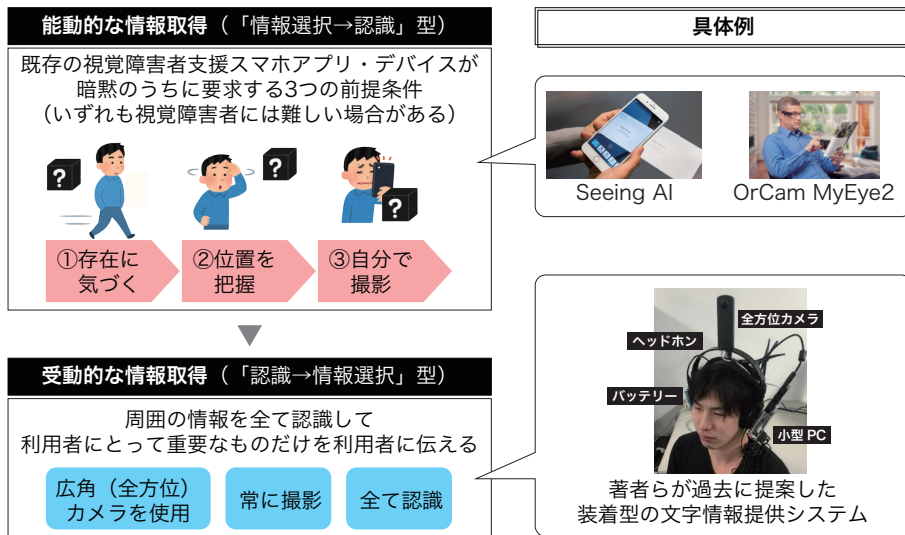


図 2: 能動的と受動的な視覚情報の取得

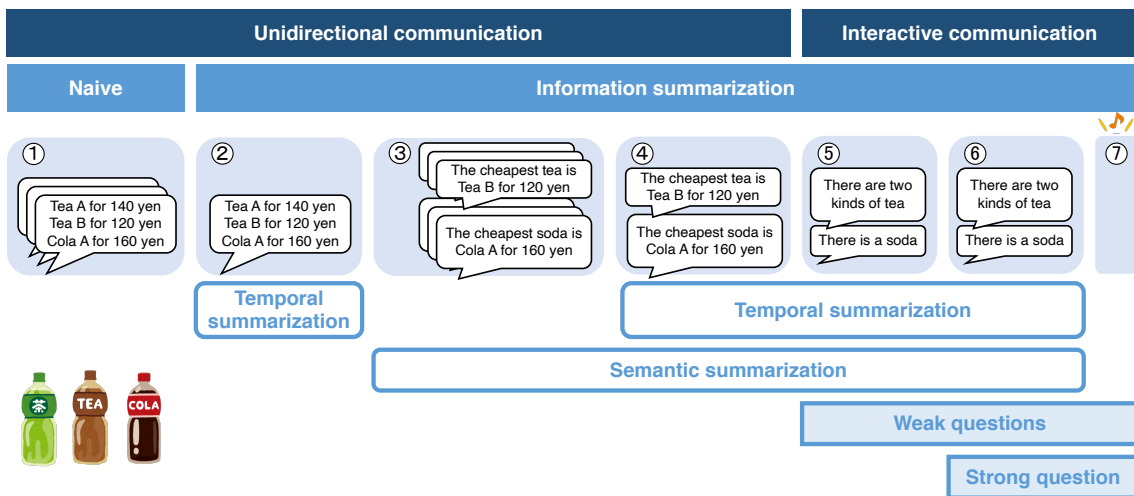


図 3: 検討した 7 種類の情報伝達手法の概要。①が要約しないで認識結果を全て伝える手法、②は時間的な要約を施す手法、③は意味的な要約を施す手法、④が時間的と意味的な要約の併用である。⑤～⑦は質問応答システムを用いた対話的な手法であり、⑤は④に加えて弱い質問のみを受け付ける場合、⑥は④に加えて弱い質問と強い質問を受け付ける場合、⑦は⑥において認識結果を伝えなくて、ベルだけを鳴らす場合である。

は、より広い範囲を自動的に認識する必要がある。これにより、多量に発生する認識・検出結果は、情報量としては多過ぎるので、その中から重要なものを取捨選択して提供する必要がある。その際、取捨選択は、単にどのような認識結果が得られらるかという「外的要因」だけで決まるのではなく、利用者がどのような状態かという「内的要因」を反映したものであるのが望ましい。

このように、認識・検出技術を用いた視覚障害者において、最初から絞り込まれた情報のみを認識するという方式から、多量の情報を認識した上で必要な情報のみを選択して提供するという方式へのパラダイムシフトを目指している。

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 つ目では、図 2 に示すように、視覚障害者が自ら写真を撮影しなくても良い「受動的な情報取得」を提案し、その際に問題となる周囲の視覚情報の要約方法を検討した。2 つ目では、周囲の視覚情報の要約方法において、利用者（視覚障害者）の好みを反映するために、情報推薦技術を利用した方法を提案した。3 つ目では、グラフの情報を要約する際に、グラフの具体的な数値を言わずに相対的な関係のみを伝える手法を提案した。

4. 研究成果

4. 1 受動的な情報取得における情報の要約

詳細は [1] に譲り、以下に概要のみを示す。

4. 1. 1 実験概要

視覚障害者 9 名（全盲 6 名、弱視 3 名）に音声案内システムを使用して疑似的な買い物をして貰った。買い物は視覚障害者にとって困難でありそして、音声案内システムが読み上げる商品名と価格から最も安い商品を探すというタスクを実験参加者にして貰った。このタスクを課した理由は、情報を正確に受け取る必要があるタスクでの所要時間とスコア（商品選択の正確さを表す）を測り、定量的に評価を行うためである。実験では、図 3 に記載の商品の名前と価格を伝える情報の要約方法 7 種類を①から⑦の番号順に 2 回ずつ試してもらった。

4. 1. 2 音声案内システム

情報の要約方法を評価するために、広角レンズを搭載したウェアラブルカメラを用いて音声案内システムを作成した。このシステムはウェアラブルカメラと骨伝導ヘッドセット、及び物理ボタンを組み合わせたものになっている。通常は、カメラで撮影した商品を認識する機能を物体認識手法を用いて実現する。しかし、物体認識には誤認識がつきものであり、誤認識が情報の要約に与える影響が無視できないと予想できる。そのため、ウェアラブルカメラを用いて、商品情報を含む QR コードを撮影し、その商品情報を骨伝導ヘッドセットを通して音声でアナウンスすることにした。骨伝導ヘッドセットは実験参加者の耳を塞がず、実験者からの指示を聞き取れるため、本実験で採用した。

ウェアラブルカメラは、実験参加者に帽子を被ってもらい、帽子のつば部分にマウント用クリップで固定する。また、大量の商品情報の読み上げによって実験参加者が不快に感じたり混乱した場合に、物理ボタンを押して実験参加者の任意で読み上げを一時停止することができる。QR コードは余白を含め一辺約 19cm で、A0 サイズのポスターに商品として 24 個印刷した。この 24 個の QR コードに対して①から⑦の各要約方法ごとに、異なる商品と価格の情報をランダムに割り当てた。その際、実際の買い物売場に似せるように同じカテゴリーの商品はまとめて配置した。QR コードを印刷したポスターから実験参加者までの距離は約 2m である。

4. 1. 3 主観評価

利用者にタスクを行う上で使い易かった要約方法の順番を答えてもらい、表 1 にまとめた。

まず、識別器から利用者に情報を要約して伝える 4 手法（すなわち、手法①から④まで；以後、一方向通信手法と呼ぶ）に焦点を当てる。ほぼ全ての参加者が手法③と④の使用が容易と述べている。両手法ともに意味的要約を用いており、これは意味的要約の有効性を示す。また、手法①と②、手法③と④をそれぞれ比較すると、大半の参加者が手法②と④を好むことが分かった。手法②と④は時間的要約を用いており、これは時間的要約の有効性を示す。

次に、全ての手法について考える。一方向通信手法（手法①から④まで）と比較すると、対話型通信手法（すなわち、手法⑤から⑦まで）の方が利用者に好まれていることが分かる。全ての参加者が手法⑦が最善であり、手法⑥が次善であると述べている。これらの手法は強い質問を使えるため、これは強い質問の有効性を示す。一方、6 人の参加者が手法⑤を 3 位にランク付けし、3 人がそれ以下にランク付けしている（すなわち、1 人が 4 位、2 人が最下位にランク付け）。これは、手法⑤が弱い質問のみを許可しているためである。つまりこれは、弱い質問を組み合わせることは利用者にとって常に快適であるとは限らないことを示唆している。

4. 1. 4 客観評価

客観的評価では、タスクを 2 回実行した際の平均所要時間と平均スコアにより手法の有効性を検証した。

タスクを完了するまでの平均時間については、情報伝達手法の違いよりも参加者の違いによる影響が大きいことがわかった。この方法を使用すると、最も安価な製品を得るために、参加者は「認識されているカテゴリーは何か？」と「(カテゴリー名)の最も安価なものは何か？」の 2 つの質問を繰り返し組み合わせる必要がある。このプロセスは、利用者がカテゴリーリストと、これまでの最も安価な製品の名前と価格を覚えておく必要があるため、比較的高い認知負荷を利用者に課す。一方、強い質問の使用を許可する手法⑥と⑦は、かなり少ない時間を必要とした。強い質問の存在のおかげで、参加者は強い質問を使用して最も安価な製品を容易に特定できた。これは対話型通信手法の力を示す。

平均スコアについて、まず一方向通信手法では、時間的要約を使用するかどうかで結果が異なった。時間的要約により一度だけ説明する手法②と④のスコアは低かったのに対し、時間的要約を使用せずに結果を繰り返し説明する手法①と③のスコアは高かった。このスコアの違いは、参加者が製品名を聞き逃したり忘れてきたときに再確認できないからと考えられる。これは時間的要約の典型的な欠点であるように思われる。次は対話型通信手法である。強い質問を使用できる手法⑥と⑦は、すべての参加者で 100% を達成した。これは強い質問の使いやすさを示す。対照的に、弱い質問のみが使用できる手法⑤のスコアは、手法⑥と⑦のスコアよりも低

表 1: 相対的な順位による主観評価。“>”は左側のものが右側より使いやすいことを示し、“=”は比較可能であることを示す。つまり、左側にあるほど使用が容易といえる。

ID	Preference
P1	⑦ = ⑥ > ③ > ④ > ① > ② > ⑤
P2	⑦ = ⑥ > ⑤ > ④ > ③ > ② > ①
P3	⑦ > ⑥ > ④ > ⑤ > ③ > ② > ①
P4	⑦ > ⑥ > ⑤ = ④ > ③ > ② = ①
P5	⑦ > ⑥ > ⑤ > ④ > ② > ③ > ①
P6	⑦ = ⑥ > ⑤ > ③ > ④ > ② = ①
P7	⑦ = ⑥ > ⑤ > ③ > ④ = ② > ①
P8	⑦ > ⑥ > ⑤ > ④ > ③ > ② > ①
P9	⑦ > ⑥ > ④ > ③ > ① > ② > ⑤

かった。

4. 2 情報取得による周囲の視覚情報の要約

4.1 で述べた手法は、利用者が誰であっても同じ要約結果を伝える。しかし、利用者によって知りたい情報は異なるはずである。そこで、情報推薦技術を用いて利用者の欲しい情報を反映した要約を提供するシステムを提案し、提案システムを評価する実験を行った。実験では、視覚障害者が一人で街を歩いているときに飲食店に遭遇したとき、その飲食店の情報を伝えるかどうかを情報推薦技術によって決めるというシナリオで実施した。あらかじめ晴眼者 1000 人に飲食店と食べ物の好みを回答してもらい、収集したデータに基づいて、視覚障害者の利用者に伝えるべきかを決める。実験には 21 名の視覚障害者に参加してもらい、要約しないで全ての情報を伝える手法などと提案手法を比較した。その結果、要約しないで全ての情報を伝える手法が良いと答えた参加者が多く、提案手法の有効性を確認できなかった。この原因は、今回の実験では利用者に提示される情報があまり多くなかったため、そもそも要約の必要が無かったためと考えられる。そのため、問題設定を見直して、改めて検討したい。

4. 3 グラフの要約

4.1 と 4.2 では、周囲の視覚情報の要約を目指したが、ここでは論文やウェブページに含まれるグラフをどのように要約するかを検討した。[2] など、グラフを認識する手法がいくつか提案されているため、我々はグラフの数値はわかっているものとして、それを読み上げ音声で利用者である視覚障害者に伝えるときに、認知負荷をかけずに要約する方法を検討した。提案手法は、グラフの具体的な数値を言わずに相対的な関係のみを伝えることに主眼を置いている。この研究は現在も継続中である。

参考文献

- [1] Masakazu Iwamura, Takaaki Kawai, Keigo Takashima, Kazunori Minatani, and Koichi Kise. “Acquiring Surrounding Visual Information Without Actively Taking Photos for People with Visual Impairment”. In: *Proc. Joint International Conference on Digital Inclusion, Assistive Technology & Accessibility (ICHP-AAATE 2022)*. Vol. 13341. Lecture Notes in Computer Science. https://www.doi.org/10.1007/978-3-031-08648-9_27. Springer International Publishing, July 14, 2022, pp. 229–240. DOI: 10.1007/978-3-031-08648-9_27.
- [2] Prerna Mishra, Santosh Kumar, Mithilesh Kumar Chaube, and Urmila Shrawankar. “ChartVi: Charts summarizer for visually impaired”. In: *Journal of Computer Languages* 69 (2022), p. 101107. ISSN: 2590-1184. DOI: <https://doi.org/10.1016/j.cola.2022.101107>.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Masakazu Iwamura, Takaaki Kawai, Keigo Takashima, Kazunori Minatani, and Koichi Kise
2. 発表標題 Acquiring Surrounding Visual Information without Actively Taking Photos for People with Visual Impairment
3. 学会等名 Joint International Conference on Digital Inclusion, Assistive Technology & Accessibility (ICCHP-AAATE 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 河合 隆哲, 高嶋 慶伍, 岩村 雅一, 南谷 和範, 黄瀬 浩一
2. 発表標題 視覚障害者の受動的な情報取得に向けての情報要約に関する調査
3. 学会等名 第181回ヒューマンインタフェース学会研究会
4. 発表年 2021年

1. 発表者名 河合 隆哲, 高嶋 慶伍, 岩村 雅一, 南谷 和範, 黄瀬 浩一
2. 発表標題 視覚障害者の受動的な情報取得に向けての情報要約に関する調査
3. 学会等名 第181回ヒューマンインタフェース学会研究会
4. 発表年 2021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
研究分担者	南谷 和範 (Kazunori Minatani) (90551474)	独立行政法人大学入試センター・研究開発部・教授 (82616)	

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
---------	---------