

THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPANESE EDITION)

**IEICE** | **電子情報通信学会**  
**D** | **論文誌** 情報・システム

VOL. J104-D NO. 8

AUGUST 2021

本PDFの扱いは、電子情報通信学会著作権規定に従うこと。

なお、本PDFは研究教育目的（非営利）に限り、著者が第三者に直接配布することができる。著者以外からの配布は禁じられている。

**情報・システムソサイエティ**

一般社団法人 **電子情報通信学会**

THE INFORMATION AND SYSTEMS SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

全方位カメラでの撮影とその後処理による視覚障害者の写真撮影支援

岩村 雅一<sup>†a)</sup>      平林 直樹<sup>†b)</sup>      程 征<sup>†c)</sup>      南谷 和範<sup>††d)</sup>  
 黄瀬 浩一<sup>†e)</sup>

Photography for People with Visual Impairment by Photo-Taking with Omni-Directional Camera and Its Post-Production

Masakazu IWAMURA<sup>†a)</sup>, Naoki HIRABAYASHI<sup>†b)</sup>, Zheng CHENG<sup>†c)</sup>,  
 Kazunori MINATANI<sup>††d)</sup>, and Koichi KISE<sup>†e)</sup>

あらまし 本論文では、全盲の視覚障害者が独力で写真を撮れる全く新しいシステム VisPhoto を提案する。スマートフォンとソーシャル・ネットワーキング・サービス (SNS) の普及により、晴眼者 (目が見える人) は写真撮影と共有が手軽にできるようになった。実は視覚障害者も同じように写真を介したコミュニケーションを望んでいるが、独力での写真撮影は容易でない。特に、(全盲の) 視覚障害者にとって、カメラを被写体に向けるのは難しい。この問題の解決を目指して、写真撮影に用いるスマートフォンの向きを音声や振動で利用者に指示する写真撮影支援システムが提案されている。それに対して、本論文で提案するシステムは、全方位カメラがカメラの周囲 360 度の景色を撮影できることを利用して、利用者の周囲を全方位画像として撮影する。そして、その後処理として、利用者の希望を反映し、かつ見た目が良い領域を「写真」として切り取る。提案手法のシステムの有用性と操作性を 8 名の全盲の視覚障害者に、撮影した「写真」の質を 10 名の晴眼者にそれぞれ検証してもらい、いずれも肯定的な評価を得た。

キーワード 視覚障害者, 写真撮影, 全方位カメラ, 物体検出, ウェブユーザインタフェース

1. ま え が き

Jayant ら [1] によると、視覚障害者<sup>(注1)</sup>の約 70% が最近写真を撮ったと報告している。視覚障害者が写真を撮る目的と対象は、晴眼者と同じく、友達や家族、旅行、イベントなどでの記念写真が過半を占める。特に最近ではソーシャル・ネットワーキング・サービス (SNS) の普及により、自分の撮った写真を他の人に見てもらいたいといった欲求もあるようだ [1]~[3]。他にも、[4]~[8] のようなスマートフォンアプリで文字を読んだり、色を判別したり、目の前の物体を判別し

たり、遠隔地にいる晴眼者に写真を見てもらって写っているものの情報を得たりという、視覚障害者に特有の写真撮影もあるという。

視覚障害者の多くが写真を撮りたいと思っている反面、視覚障害者の写真撮影にはチャレンジが伴う。視覚障害者の写真撮影を支援する既存研究の中には、撮影時のフレーミングや照明、焦点 [2]、構図 [9] に焦点を当てたものがある。しかし、その実、(全盲の) 視覚障害者はいかにして撮影対象をカメラに収めるかという、より初歩的な問題で困っている。

恐らく視覚障害者の写真撮影を助けてくれる最も身近な存在は、iPhone に搭載された iOS の読み上げソフト VoiceOver であろう。これは、写真を撮る際に、フレーム内に写る人の数をしゃべって教えてくれる。人に限らず、様々な対象の写真撮影を実現する方法も研究されているが、それらは「もうちょっと左」のように、カメラを向ける方向を音声やバイブレーションで利用

<sup>†</sup> 大阪府立大学大学院工学研究科, 堺市  
 Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Sakai-shi, 599-8531 Japan

<sup>††</sup> 大学入試センター, 東京都  
 National Entrance Examination Center, 2-19-23 Komaba, Meguro-ku, Tokyo, 153-8501 Japan

a) E-mail: masa@cs.osakafu-u.ac.jp  
 b) E-mail: hirabayashi@m.cs.osakafu-u.ac.jp  
 c) E-mail: zheng386@hotmail.com  
 d) E-mail: minatani@rd.dnc.ac.jp  
 e) E-mail: kise@cs.osakafu-u.ac.jp

DOI: 10.14923/transinfj.2020JDP7069

(注1): [2] に倣い、本論文では「視覚障害者」という言葉を目視による状況の確認や、写真の内容を識別することを困難とする人々のことを指す。



図1 正距円筒図法の画像の例



図2 通常のカメラで撮影したような画像の例

者に指示するものである [1], [9]~[12]. この方法の難点は, 利用者が指示どおりに動くように頑張らないといけないことと, 時間がかかることである. 例えば, 猫を撮ろうとカメラの向きを修正している間に猫がどこかに行ってしまうかもしれないと限らない. このように, 「写真を撮るときは撮影対象にカメラを向ける」という, 晴眼者には普通のことが, 視覚障害者には難しい.

そこで本論文では, 利用者が撮影対象にカメラを向けなくてもよい撮影方法として, 全方位カメラを使う方法を提案する<sup>(注2)</sup>. 全方位カメラ (全天球カメラともいう) とは, 文字どおり, カメラから見て全ての方向の写真が一度に撮れるカメラである. 全方位カメラを使えばカメラの周囲の写真を一度に撮影できるため, 利用者が撮影対象にカメラを向ける必要がない<sup>(注3)</sup>. しかし, 全方位カメラで撮影した画像は図1のような正距円筒図法で表現されているため, ひずんでいる. これは, カメラの周囲 360 度の景色を無理矢理 1 枚の画像に収めたためであり, ひずみは画像の場所によって異なる. 全方位カメラで撮影した写真を見るための専用ビューアもあるが, 一覧性に難があるため, 図2のような通常のカメラで撮影したような画像の方が好まれる. そこで, 図1のような画像から図2のような画像を切り出すことで, 「写真撮影」を実現する. 全方位カメラで撮影した画像から, 「写真」として切り出す際に我々が注目するのは, 画像中の「もの (物体)」の情報である. すなわち, 写真に入れたいものと入れたくないものを利用者に尋ね, 望んだ写真を切り出す. その際, 写真の美しさも考慮に入れる.

提案手法の長所は以下の 3 点である.

- (1) 利用者がカメラを被写体に向ける必要がない.
- (2) カメラの向きの補正に時間がかかる従来手

(注2): 本論文は [13]~[15] に基づく.

(注3): 本研究で使用する全方位カメラは二つの広角カメラで撮影した画像を自動的に上手くつなぎ合わせてくれるため, 利用者がカメラの境目を気にする必要はない.

法 [1], [9]~[12] とは異なり, 原理的には動いている被写体も撮影できる.

(3) 利用者への実時間フィードバックを返す必要がない. 既存手法は利用者へに実時間でフィードバックを返すために, できる処理は計算負荷の軽いものに限られる. しかし, 提案手法の後処理は計算サーバを用いたオフライン処理であるため, 計算負荷の比較的重い処理が実施可能である. 本論文では, 物体検出と写真の美しさを評価する二つのニューラルネットワークを用いた. これらをスマートフォンなどのポータブルデバイスで実行すると計算時間がかかるため, 既存手法で用いることはできない.

なお, 本研究は大阪府立大学 大学院工学研究科 倫理委員会の承認を受けて実施した.

## 2. 関連研究

### 2.1 写真撮影支援

既存研究における現在主流のアプローチは, 利用者が被写体にカメラを向けられるよう, 音声と振動のフィードバックを提供するものである [1], [9]~[12]. EasySnap [1], [10] は, 被写体にカメラを向ける際に役立つ iPhone アプリである. フレーム内のぼかしや明るさ, カメラの傾き, 顔とオブジェクトのサイズと位置などの情報を, 音声フィードバックで利用者に提示する. PortraitFramer [1] は EasySnap に似ているが, 複数の人物の顔が含まれる場合に顔の数やサイズと位置を利用者に提示する. Vazquez ら [12] は, カメラのロールを自動的に修正する機能をもつ手法を提案した. Balata ら [9] は, 被写体にカメラを向ける際に中央及び黄金比の構図に最適化するスマートフォン アプリを提案した.

### 2.2 写真の管理, 閲覧, 共有の補助

Frohlich ら [16], [17] は写真に音声録音を組み合わせたことの価値を探り, 周囲の環境音や音声による注釈

は多くの場合有益であることを見出した。これは視覚障害者向けの研究ではないが、示唆に富んでいる。上記の EasySnap [1] は、視覚障害者が利用可能な写真アルバムである。写真が撮影された場所と認識エンジンによって付与された写真の中身の情報が各写真に関連づけられているため、視覚障害者が写真の中身を把握して共有できる。Harada ら [2] は写真を閲覧できるように、環境音や音声メモの録音を追加し、撮影と録音が可能なモバイルアプリケーションを提案した。Adams ら [18], [19] は、写真撮影時の周囲の環境音を録音し、撮影日時と位置情報とともに写真に関連づけることで、目が見えなくても写真ライブラリを整理して「閲覧」できる VisSnap という iPhone アプリを提案した。

### 2.3 写真に関連したサービス

VizWiz [4], Be My Eyes [5], Envision AI [6], TapTapSee [7] などのスマートフォンアプリは、写真を通して視覚障害者が情報を取得する際に役立つ。ただし、そのためには視覚障害者が知りたい物体の写真を撮る必要がある。Gurari らの調査 [20] によると、VizWiz に寄せられた視覚障害者の写真とそれに関する質問のうち、~28% は回答不能という。その一因は写真の品質である。これは視覚障害者にとって写真撮影が難しいことを示している。Kacorri らは視覚障害者が撮影する写真に個人差があることから、利用者が撮影する少量の写真に物体認識の識別器を適応させることで、認識精度を向上させた [8]。しかし、これらはいずれも認識対象が目前にあることを前提にしているため、何らかの方法であらかじめ対象物体の位置を把握しておく必要がある [21]。

他にも幾つかのサービスが挙げられる。OrCam MyEye 2 [22] は、眼前のテキストを読み上げ、顔や商品を認識できるウェアラブルデバイスである。vOICe [23] は、利用者の目の前にあるシーンを認識できるウェアラブルデバイス（ソフトウェア）である。

### 2.4 全方位ビデオの通常視野への変換

提案手法に関連する研究分野には、全方位ビデオの通常視野に変換するものがある [24]~[29]。特に [26], [28] は物体検出を使っているという点で提案手法と関係がある。しかし、これらの研究の目的は、情報量の大きい通常視野のビデオの作成である。いずれの手法も利用者が望む領域を切り出すことを目的としていないという点で提案手法とは異なる。

### 2.5 ライトフィールドカメラ

提案システムは、撮影後に利用者が望む画像を生成するという点において、ライトフィールドカメラ [30] に似ている。代表的なメーカーには、2018 年にサービスを停止した Lytro [31] がある。Lytro のライトフィールドカメラはマイクロレンズアレイで構成されており、“ライトフィールド”を保存できるため、後で写真の焦点を自由に変更できる（リフォーカス）。例えば、人と犬が含まれている写真があったとき、写真撮影後に人物に焦点が合った画像、犬に焦点が合った画像、及び両方に焦点が合っていない画像を生成できる。提案システムと比較すれば、ライトフィールドカメラは全方位カメラに、リフォーカスされた画像は切り出した画像に対応する。

## 3. 提案手法 VisPhoto

本論文では、全方位カメラでの撮影とその後処理による視覚障害者の写真撮影を実現するシステムを提案する。提案手法の概要を図 3 に示す。提案手法は、(1) 全方位カメラでの写真撮影（図中の Step 1）と (2) 写真の切り出し操作（図中の Step 2~4）の二つの処理に大別できる。

(1) では、利用者が全方位カメラのシャッターボタンを押すことで、全方位画像を撮影する。提案手法が通常の写真撮影と異なるのは、シャッターボタンを押しただけで写真撮影が完結しない点である。すなわち、後処理である Step 2~4 において写真を切り出すまでが通常の「写真撮影」に対応する。

後処理である (2) では、利用者の希望を反映して写真を切り出す。全盲の視覚障害者が普段使用しているパソコンやスマートフォンでウェブインタフェースを操作することを想定している。そのため、スクリーンリーダーで読み上げられるように、ウェブページは主にテキスト情報で構成されている<sup>(注4)</sup>。サーバには複数の写真（全方位画像）がアップロードされることを想定しているため、Step 2 では処理したい写真を選択する。Step 3 では、選択した写真から物体検出技術で検出された物体の一覧を利用者に提示して、写真に含めたい（あるいは含めたくない）物体を選んでもらう。Step 4 では、利用者が希望した写真をダウンロードできるようにする。

(注4)：見出し語に HTML の h4 タグや h5 タグを付与することで、項目間を素早く移動して読み上げられるように工夫している。

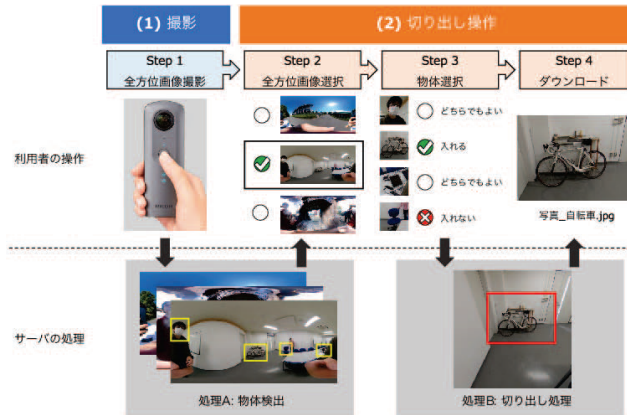


図3 提案システムの概要

以下、利用者の操作とサーバの処理に分けて、順を追って処理の詳細を説明する。

### 3.1 利用者の操作

#### 3.1.1 Step 1: 全方位カメラでの写真撮影

提案手法はシャッターボタンを押ただけで写真撮影が完結しないため、後処理の段階で、シャッターボタンを押したときはどういう状況で、何を撮りたかったのかを利用者が覚えておく必要がある。そこで本研究では、全方位カメラ上で動く専用アプリを開発して、アプリ上で写真撮影と写真撮影時の状況の記録とサーバへのアップロードができるようにした。写真撮影時の状況の記録については、Haradaら[2]のように、写真撮影時の音声を録音しておくことでこの課題の解決を図る。これは写真撮影と連動しており、シャッターボタンを押すと同時に録音が始まり、シャッターボタンを離すと録音が終わるようになっている。ただし、シャッターボタンを押している時間が10秒未満の場合は、最低でも10秒録音する。この理由は、利用者に何らかの音声情報を残すことを促すためである。また、環境音を長目に録音することで、利用者が写真撮影時の状況を思い出せる確率を上げるためでもある。

#### 3.1.2 Step 2: 全方位画像の選択

Step 2では処理したい写真を選択する。図4(a)に示すStep 2のウェブインタフェースでは、写真が新しいものから古いものという順番に並ぶ。その中から、写真の選択ボタンを押すことで写真を選ぶ。写真が見えない利用者でも写真が選択できるように、写真以外の情報として、Step 1で録音した音声、画像に写る物体のリスト(次項で述べる)、撮影日時を写真とともに表示する。既存研究で行われているように、写真を撮影

した位置の情報も表示するのが望ましいが、今回用いた全方位カメラはGPS情報を取得しないため、実現できていない。

#### 3.1.3 Step 3: 物体選択

Step 3では、選択した写真から検出された各物体を写真に含めたいか、含めたくないか、あるいはどちらも良いかを利用者に入力してもらう。図4(b)に示すように、利用者に物体の一覧を提示する際には、カメラから見た物体の方向をクロックポジションで示す。クロックポジションとは、方向をアナログ時計の短針にたとえて表す方法で、正面が12時、右が3時、背面が6時、左が9時のように表す。全方位画像との対応は、画像の横軸を考えたとき、中央が12時、両端が6時に相当する。図4(b)の例では、12時の方向にbicycle、3時にchair、5時にchairとdiningtable、6時にpersonとcell phoneといった具合で検出されている。画面に表示されている全方位画像は、検出された物体の中心位置に点が打ってある。各物体には「含める」、「どちらでもいい」、「含めない」の三つのラジオボタンが用意しており、初期状態では「どちらでもいい」が選択されている。

#### 3.1.4 Step 4: 写真のダウンロード

Step 4では、Step 3で利用者が入力した希望を反映した写真を切り出し、図4(c)のように画面に出力する。切り出した画像はダウンロードできる。その際、画像に含まれる物体の名前がファイル名に含まれる。これは、視覚に障害がある利用者が後から何を撮影したのかを理解するための工夫である。

### 3.2 全方位カメラとAndroidアプリ

本研究では、Step 1で使う全方位カメラとして、二



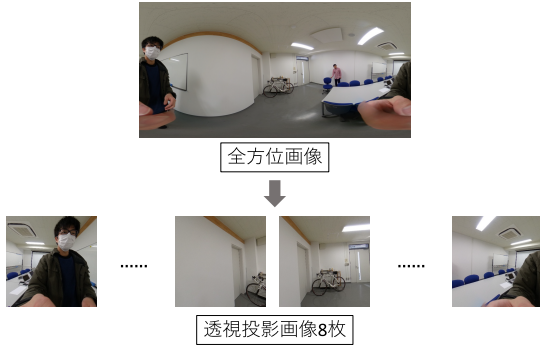


図5 全方位画像を8枚の透視投影画像に変換してから物体を検出する

像に投影して統合する。ただし、現在の実装は物体の中心位置のみを考慮して統合しているため、全方位画像に大きく写る物体は別の物体として扱われ、二つ以上検出されてしまう可能性がある。この解決は今後の課題である。

もう一つの実装は、物体検出アルゴリズムとして、Google Cloud API の物体検出機能<sup>(注5)</sup>である。また、全方位画像の赤道だけでなく、上下に30度ずつらした位置でも8枚の透視投影画像に変換している。すなわち、合計24枚の透視投影画像を作成する。それ以外は前述のYOLOを使う実装と同じである。

### 3.3.2 処理B：写真の切り出し

利用者の希望を反映した上で見た目が良い写真を2段階の処理により切り出す。すなわち、1段階目では全方位画像から大き目の透視投影画像を生成し、2段階目でその画像から所望の画像を切り出す。

1段階目では、「含める」と選択した(複数の)物体の重心を全方位画像上で求め、その重心を中心とする透視投影画像を生成する。すなわち、新しく生成される透視投影画像の中心の座標 $(x^C, y^C)$ は、

$$(x^C, y^C) = \frac{1}{|I_{in}|} \sum_{i \in I_{in}} (x_i^C, y_i^C), \quad (1)$$

与えられる。ここで $(x_i^C, y_i^C)$ は*i*番目の物体の中心の座標であり、 $I_{in}$ は「含める」と選択された物体の集合である<sup>(注6)</sup>。こうして生成する画像の画角は上下左右がそれぞれ90度であり、解像度は800×800で

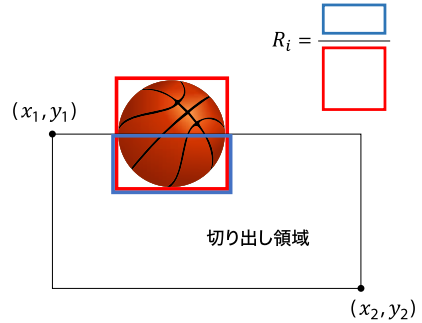


図6  $R_i$  の定義. 赤い矩形は物体(ボール)の外接方形、青い矩形は物体の領域と切り出し領域の共通部分を表す. このとき、 $R_i$  は赤い矩形の青い矩形に対する面積比として定義される

ある。

2段階目の処理では「写真」として切り出す領域を決定する。まず、写真として切り出す長方形領域を $\theta = (x_1, y_1, x_2, y_2)$ と定義する。ここで $(x_1, y_1)$ と $(x_2, y_2)$ はそれぞれ領域の左上と右下の座標である。更に $I_{in}$ と同様に、 $I_{out}$ を「含めない物体」の集合として、以下のエネルギー関数 $E(\theta)$ を定義する。

$$E(\theta) = \frac{E_{const}(\theta) + 1}{E_{aes}(\theta)} \quad (2)$$

$$E_{const}(\theta) = \sum_{i \in I_{out}} R_i + \sum_{i \in I_{in}} (1 - R_i), \quad (3)$$

ここで $E_{aes}(\theta)$ はview finding network (VFN) [36], [37]の出力スコアである。VFNは写真の美しさを評価する深層ニューラルネットワークであり、美しい写真に対して大きなスコアを出力する。また、 $R_i$ は、図6に示すように、*i*番目の物体と切り出し領域の共通部分の面積比である。式(2)はNelder-Mead法により最小化され、目的関数を最小化する長方形領域 $\theta$ を求める。長方形領域 $\theta$ は、後の実験でiPhoneの写真と比較するため、縦横比をiPhoneと同じ4:3になるように制限したが、画像の大きさに制限は設けなかった。

## 4. 評価

提案手法の評価のために、3種類の調査を行った。一つ目は、システムの有用性や操作性等を検証するためのものであり、全盲の視覚障害者8名に参加してもらった。二つ目は、視覚障害者が独力で写真を撮影できることを確認するためのものであり、視覚障害者10名に参加してもらった。三つ目は、提案手法で出力した写真の質を検証するためのものであり、晴眼者10

(注5) : <https://cloud.google.com/vision/docs/object-localizer>

(注6) : 選択された「含めたい物体」が広い範囲に分布している等、一枚の透視投影画像に入り切らない場合には、この方法では含めたい物体を全て含む透視投影画像にならない。

表1 実験1の参加者の基本情報

	年齢	性別	視覚障害の状況
VI-1	26	女性	生まれつき弱視。7歳から病患で失明。
VI-2	27	男性	8歳まで健常。8歳から視野が狭くなり、10歳で失明。
VI-3	33	女性	6歳まで健常。6歳から弱視。13歳で失明。
VI-4	28	男性	生まれつき左目は盲、右目は弱視。7歳で失明。
VI-5	29	男性	2歳まで健常。2歳で失明。
VI-6	23	女性	2歳まで健常。2歳で失明。
VI-7	47	男性	生まれつき弱視。6歳頃失明。
VI-8	28	女性	生まれつき右目は盲、左目は弱視。25歳で失明。

名に参加してもらった。

#### 4.1 実験1：システムの有用性、操作性の検証

全盲の視覚障害者8名に参加してもらい、システムの有用性、操作性の検証を検証した。本実験で用いた提案システムはYOLOを使う実装であり、全方位カメラはRicoh Theta Vを用いた。実験参加者の基本情報を表1に記す。参加者は全員全盲で点字使用者であった。

実験は、実験参加者に提案手法の説明を行った後、提案システムを数日間貸し出し、提案システムを自由に使用してもらう形式で行った。提案手法を独力で操作できるように、実験参加者には3.2で示したVisPhoto専用アプリを起動する手順を伝えたが、撮影方法の指示は与えなかった。実験参加者には提案手法の説明の前に写真撮影に関する質問にインタビュー形式で答えてもらった。そして、カメラの返却時に提案手法に関するアンケートに回答してもらった。詳細は付録の表A-1に記すことにして、以下に概略を記す。

##### 4.1.1 写真撮影に関する質問への回答

• Q1の「どれくらい写真を撮りたいですか?」という質問(5段階評価:5が撮りたい、1が撮りたくない)に対して、6名の参加者が5または4と答えた。

• Q3の実際に写真撮影する頻度は、ほとんどの参加者が月に1回または年に数回程度と答えた。

• Q4の「写真を撮る目的は何ですか?」という質問では、視覚障害者自身が体験を覚えておくためと体験を他者に共有するための2通りの回答があった(表A-2参照)。

• Q5の「うまく写真を撮れますか?」という質問には5名が否定的に答え、Q7の「現在使っている写真撮影機器に満足していますか?」という質問には6名が否定的に答えた。したがって、実験参加者はおお

むね良い写真を撮れているとは思っておらず、満足もしていない。

• Q6から、全参加者が写真を撮るためにiPhoneを使用している。1名はRicoh Thetaも使用している。

• Q8の写真撮影に失敗する原因は、おおむね既発表文献で挙げられたとおりの内容であった。以下に列挙する。距離感をつかめない。ピントが合わない。動くものに対して適切なタイミングでシャッターボタンを押せない。被写体が写真に含まれない、あるいは被写体の全体が写真に含まれない。撮影した写真が暗い。撮影した写真に指が含まれる(表A-2参照)。

• Q11で撮影した写真をどう整理するかを尋ねたところ、ファイル名に日付や場所の情報を加える等、ファイル名を変更するという回答が多かった(表A-6参照)。

• Q12の「写真に写るものの名前がファイル名に入っていたら便利ですか?」という質問に対しては全員が「はい」と回答した。

##### 4.1.2 提案手法VisPhotoに関する質問への回答

• Q13では、全参加者が簡単に提案システムの使い方を理解できたと回答した。

• Q14の「実際にVisPhotoを使いたいと思いますか?」という質問(5段階評価:5が思う、1が思わない)の回答は、5が5名、3が2名、1が1名であった。

• Q15の「VisPhotoがあれば、もっと写真を撮りたいと思いますか?」という質問(5段階評価:5が思う、1が思わない)に対して、6名が肯定的に答えた。

• Q16の「VisPhotoにどんな機能があれば便利ですか?」という質問に対しては、様々な意見が寄せられたが、実際に切り出した「写真」に含まれる物体が何かを教えて欲しいという意見が多かった(表A-7参照)。

• Q17~19について、写真を撮影した瞬間のことを覚えておくことに関して、ほとんどの参加者は我々が予想していたよりも比較的長く覚えていられると回答した。具体的には、以下のとおりである。

– Q17の「写真撮影時に撮りたかったものを覚えていらっしゃいますか?」という質問に対しては1名を除いて肯定的な回答であった。

– Q18の「写真撮影時に撮りたかったものをどれくらいの期間覚えていられると思いますか?」という質問に対しては、2~3日が3名、1週間が3名、1か月が2名であった。

– Q19の「どうすれば写真の内容を長く覚えてお



けると思いますか?」という質問に対しては、写真のファイル名の工夫や音声メモが有用という回答が多かった(表 A-8 参照)。

結果として、提案システムの有用性は肯定的に評価された。

#### 4.2 実験 2: 視覚障害者が撮影した写真の検証

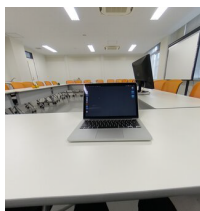
視覚障害者 10 名(全盲 7 名, 弱視 3 名)に参加してもらい、提案手法を用いて写真撮影できるかを検証した。着席した実験参加者の前のテーブルに認識対象が置かれており、実験者は実験参加者に認識対象が何であるかを伝える。実験参加者は手探りで認識対象の位置を確認するなどして、提案手法で撮影してもらった。撮影対象はノート PC, モニタ, キーボード, マウス, ドーナツとジュース, めいぐるみ, 自転車, ブロccoli, パナナとオレンジ, りんごの 10 種類で、実験参加者にはこの順に 1 枚ずつ、合計 1 名当り 10 枚の写真撮影してもらった。なお、撮影対象が自転車の場合のみ、実験参加者には立って撮影してもらった。撮影された写真は著者の 1 人が目視で確認して、対象物体の全体が写真に収まっていれば成功と判定し、それ以外は失敗と判定した。実験参加者は写真撮影(図 3 の Step 1~4)のみを行い、全方位カメラの起動などの操作は実験者が行った。実験参加者には、全方位カメラは両手ではなく片手でもつことと、全方位カメラの上をもつと指が含まれる可能性があるの、なるべく下の方をもつことを指示した。本実験で用いた提案シ

テムは Google Cloud API を使う実装であり、全方位カメラは Ricoh Theta Z1 を用いた。実験参加者の基本情報を表 2 に記す。

実験参加者が撮影した 100 枚の写真のうち、撮影に成功したのは 98 枚(成功率 98%)であった。成功例を図 7 に示す。失敗した 2 枚のうち、1 枚はマウスが検出されなかったことが原因である。もう 1 枚はドーナツとジュースを撮影するタスクにおいて、実験参加者が(提案手法の Step 3 の物体選択で)ドーナツの選

表 2 実験 2 の参加者の基本情報

	年齢	性別	視覚障害の状況
VI-9	23	女性	全盲 / 生まれつき全盲。
VI-10	30	男性	全盲 / 生まれつき全盲。
VI-11	27	女性	全盲 / 3 歳まで健常。3 歳で失明。
VI-12	28	女性	全盲 / 生まれつき全盲(光覚弁)。
VI-13	35	男性	弱視 / 埋まれつき弱視(右視力 0.15, 左 0.04)。20 歳頃から右目が 0.01, 左目が光覚弁。
VI-14	31	女性	弱視 / 18 歳まで健常。18 歳で右が 0.01 程度で中央の視野欠損, 左を失明。
VI-15	27	男性	全盲 / 先天性の緑内障により、生まれつき弱視で 5 歳から全盲。
VI-16	47	男性	全盲 / 生まれつき弱視(右が光覚弁, 左が 0.05)。10 代後半で両眼光覚弁になり、20 歳で失明。
VI-17	36	女性	全盲 / 生まれつき弱視(視力 0.02 程度で視野狭窄)。15 歳で失明。
VI-18	57	男性	弱視 / 未就学時は視力 0.9 程度。25 歳と 55 歳でそれぞれ発症した緑内障と網膜剥離により、現在両目共視力 0.02 程度で、視野がドーナツ状に約 10 度。



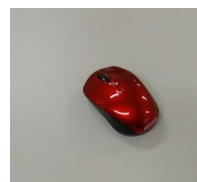
(a) ノート PC



(b) モニタ



(c) キーボード



(d) マウス



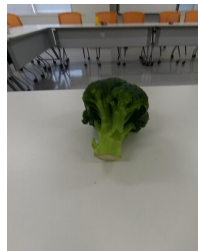
(e) ドーナツとジュース



(f) めいぐるみ



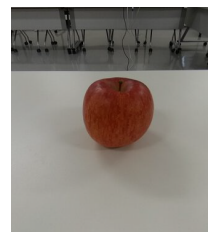
(g) 自転車



(h) ブロccoli



(i) パナナとオレンジ



(j) りんご

図 7 《実験 2》提案手法の写真撮影の成功例 (いずれも VI-15 が撮影)

択を忘れたため、図8に示すようにドーナツの一部が画面外にはみ出た。成功と判定した写真のうち、5枚には図9に示すように指が入っていた。これは、カメラをもつ位置がカメラレンズに近かったことが原因である。指が入った写真は、最初に撮影したノートPCの写真に多く、再度実験参加者に注意を促すことにより、1名の実験参加者(VI-10)を除いて再発しなかった。また、成功と判定した写真には、図10(a)に示すように対象物体が小さ目に写った写真もあった。美しい写真を切り出すために導入したVFNを用いない場合は図10(b)に示すように対象物体が大きくなるように切り取られるため、この現象はVFNを用いることの副作用といえる。

### 4.3 実験3：晴眼者が撮影した写真の質の検証

晴眼者10名に参加してもらい、提案手法で出力した写真の質を検証した。晴眼者に写真の質を評価してもらうのは、視覚障害者との対話の中で、撮影した写真は晴眼者とのコミュニケーションに使いたいという声があったからである。そのような目的であれば、写真を受け取った晴眼者が写真に違和感を抱かなければ成功と評価できる。

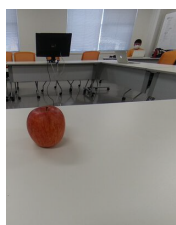


図8 《実験2》提案手法の写真撮影の失敗例。物体選択画面でドーナツを選択しなかったため(VI-12撮影)

本実験で用いた提案システムはYOLOを使う実装であり、全方位カメラはRicoh Theta Z1を用いた。実験参加者の基本情報を表3に記す。参加者は全員20代の大学生と大学院生であり、性別は男性8名、女性2名、年齢構成は23歳が3名、22歳が4名、21歳が3名であった。各参加者にはスマートフォン(iPhone)と提案システムで10枚ずつ写真を撮影してもらい、別の参加者に写真の質を10段階で評価してもらった。提案システムによって生成された写真の評価がスマートフォンで撮影されたものに近いほど、提案手法で出力した写真の質が高いとみなすことができる。また、提案システムによって出力された画像とスマートフォンで撮影された画像がどれくらい似ているかを5段階で評価してもらった。代表的な成功例を表4に、失敗例を表5に示す。

実験後、参加者にアンケートに答えてもらった。表6にアンケートの回答を示す。以下に回答の概略を記す。

- Q1の「物体認識の精度は十分ですか?」という質問(5段階:5が十分,1が不十分)という質問に対しては、4が5名、3が3名、2が2名であった。
- Q2の「切り出し結果に満足できましたか?」という質問(5段階:5が満足,1が不満)という質問に

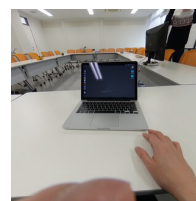
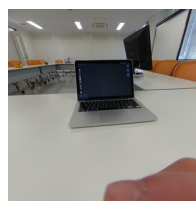
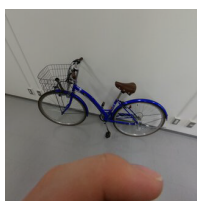
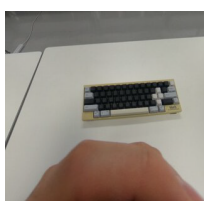
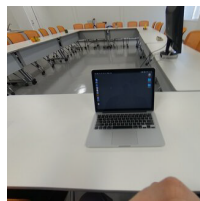


(a) VFN 有り



(b) VFN 無し

図10 《実験2》VFNの有無



(a) ノート PC (VI-10 撮影) (b) キーボード (VI-10 撮影) (c) 自転車 (VI-10 撮影) (d) ノート PC (VI-12 撮影) (e) ノート PC (VI-16 撮影)

図9 《実験2》提案手法の写真撮影時に指が入った例

表3 実験3の参加者の基本情報

	S-1	S-2	S-3	S-4	S-5	S-6	S-7	S-8	S-9	S-10
年齢	23	23	22	22	22	21	21	21	22	23
性別	女性	男性	男性	男性	男性	女性	男性	男性	男性	男性

表4 《実験3》成功したケースの代表例




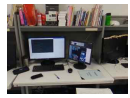

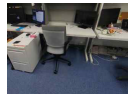


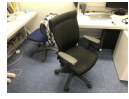



iPhone		VisPhoto		類似度
写真	質	写真	質	
	10		10	4
	9		8	5
	9		8	5
	9		9	5
	10		10	5
	5		2	4

表5 《実験3》失敗したケースの代表例













iPhone		VisPhoto		類似度
写真	質	写真	質	
	5		3	1
	9		2	1
	8		7	1
	6		5	2
	9		7	3
	9		8	1

表6 《実験3》晴眼者対象のアンケートへの回答

	S-1	S-2	S-3	S-4	S-5	S-6	S-7	S-8	S-9	S-10
Q1. 物体認識の精度は十分ですか？ (5段階：十分(5)～不十分(1))	4	4	3	4	4	3	2	4	2	3
Q2. 切り出し結果に満足できましたか？ (5段階：満足(5)～不満(1))	4	4	4	4	3	4	2	4	4	3
Q3. 切り取られた画像と全方位画像全体のどちらを見るのがいいと思いますか？ (5段階：満足(5)～不満(1))	切り出し画像	切り出し画像	切り出し画像	切り出し画像	切り出し画像	切り出し画像	どちらもいえない	切り出し画像	切り出し画像	全方位画像

対しては、4が7名、3が2名、2が1名であった。

• Q3の「切り取られた画像と全方位画像全体のどちらを見るのがいいと思いますか？」という質問には、8名が切り出し画像、1名が全方位画像、1名が両者に利点、欠点があり、どちらもいえないと回答した。

以上より、提案システムによって切り出された画像は常に晴眼者がスマートフォンで撮影した画像と合致するわけではないが、多くの場合その質は許容できる範囲のものであることが示された。

### 5. 現状の制約と改善点

実験で用いたプロトタイプは、本論文で提案する視

覚障害者向けの写真撮影方式の有効性を確認する上で必要最低限のものである。ここでは現状の制約と改善点を論じる。

#### ・物体検出

実験1と3で用いた物体検出手法（YOLOを使う実装）は、あらかじめ学習された80カテゴリーの物体が検出できる。しかし、この80カテゴリーで利用者が写真に写したい対象を網羅するのは困難である。たとえ80カテゴリーに含まれる物体であっても、写真に写っている物体を全て検出できるわけではなく、誤検出や検出漏れ、カテゴリーの間違いといった問題も生じる。検出誤りが生じ得るのは実験2で用いたGoogle Vision APIも同じである。このような物体検出手法に

起因する問題は、将来的により良い物体検出手法に置き換えることで緩和できると期待している。

#### ・物体選択

実験1のQ16「VisPhotoにどんな機能があれば便利ですか?」という質問に対して「物体が何も検出されなくても特定方向(例えば、1時方向)を切り出す機能が欲しい」という回答があった。提案手法は検出された物体を写真に含める(若しくは含めない)手法であるため、検出されない物体は写真に含めることができない。もし具体的に撮影したい対象と方向が決まっているのであれば、2.1で述べた既存の写真撮影手法が適している場合も考えられる。

また、同一カテゴリーの物体が複数検出された場合、それらを区別して利用者に伝えるのが現状では難しい。後者については、質問応答システム(例えば、[38])を導入するなど、物体の外見を対話的に伝えられるシステムにするのが望ましい。

#### ・写真切り出し

3.3.2で述べた写真を切り出すアルゴリズムは、利用者が望んだ物体が切り出した長方形領域に必ず含まれることを保証するものでないことに注意が必要である。例えば、利用者が全方位画像中の離れた位置にある二つの物体を写真に収めたいと希望した場合、実際に切り取られるのは二つの物体の間になるため、どちらの物体も含まれていない場合があり得る。また、写真に含めたくない物体が含まれたい物体に囲まれている場合には、含めたくない物体も否応なく出力する写真に含まれるのを避けられない。前者に対しては、切り出された写真に再度物体検出を適用して、利用者の希望に沿っていない場合は警告するといった対処方法も考えられる。

#### ・撮影対象が分かるのか

全盲の視覚障害者の場合、写真撮影時に「そもそも何を写しているのか分からない」という問題が生じる可能性がある。このような問題が生じるか否かは、視覚障害者がどのような目的で写真を撮影するかと密接な関係にある。思い出に残すことやSNS等で他者と共有するための写真撮影では、利用者が写真撮影という行為を能動的に行っていることから、撮影対象が撮影前に決まっていることが多いと考えられる。一方、視覚障害者を支援するスマートフォンアプリで情報を取得する場合は、撮影対象が分からない場合も考えられる。更にいえば、利用者が潜在的な撮影対象にそもそも気付いていない場合も考えられる。この場合は、周囲を

常に撮影し続け、必要な情報を利用者に伝える「受動的な写真撮影」[21],[39]が有効な場合も考えられる。

## 6. む す び

本論文では、全盲の視覚障害者が手軽に写真を撮影できるシステムを提案した。全盲の視覚障害者はカメラを被写体に向けるのが容易でないことから、このシステムでは全方位カメラを用いて利用者の写真を撮影する。そして、後から物体検出結果を参照しながら、利用者の希望に沿った写真を切り抜く。これら二つの処理が通常の「写真撮影」に相当する。提案手法を実現したプロトタイプを作成し、3種類の評価実験を行った。一つ目は、全盲の視覚障害者8名に参加してもらい、提案システムの有用性や操作性等を検証した。二つ目は、視覚障害者10名(全盲7名、弱視3名)に参加してもらい、提案手法を用いることで視覚障害者が独力で写真を撮影できることを確認した。三つ目は、晴眼者10名に参加してもらい、提案手法で出力した写真の質を検証した。いずれも肯定的な評価を得た。5で議論したように、現在のプロトタイプには様々な課題があるものの、限定された条件であっても、視覚障害者が独力で写真を撮れるという、それを補って余りある利点があることが評価実験で確認できた。

今後の課題は以下のとおりである。重要な課題の一つは、物体検出処理の改善である。提案手法では、物体が正しく検出されることを前提としているため、提案手法の有効性は物体検出の性能に依るところが大きい。ただし、物体検出手法は最新のものをプラグインすれば良いので、常に最新の手法を利用でき、更に複数の手法を組み合わせることも原理的には可能である。また、全方位画像に大きく写る物体から二つ以上の物体が検出されてしまう問題も解決が必要である。提案システムで使用しているRicoh Thetaが無音のために視覚障害者にとって使いにくいという問題も改善が望まれる。これについては、既にバンダーに改善を要求済みである。ユーザインタフェースの洗練も必要であろう。現在のプロトタイプのウェブユーザインタフェースでは写真を必ず表示するようになっているが、全盲の利用者が対象であれば写真を表示する必要はないだろうから、写真を表示しないモードを用意するのが望ましい。また、提案手法は撮影した写真をサーバにアップロードすることを前提としているため、プライバシーへの配慮の必要もあるかもしれない。Google等のサービスでも写真をサーバにアップロードしてい

るため、それと同様という考えもあるが、これを気にする利用者も少なからずいそうである。特に全方位カメラを利用する場合には意図せず写る範囲が通常のカメラより広いことも問題となり得る。

**謝辞** 本研究は電気通信普及財団 研究調査助成と JSPS 科研費 JP20H04212 の補助を受けて実施した。

## 文 献

- [1] C. Jayant, H. Ji, S. White, and J.P. Bigham, "Supporting blind photography," Proc. 13th international ACM SIGACCESS Conf. Computers and accessibility, 2011.
- [2] S. Harada, D. Sato, D.W. Adams, S. Kurniawan, H. Takagi, and C. Asakawa, "Accessible photo album: enhancing the photo sharing experience for people with visual impairment," Proc. SIGCHI Conf. Human Factors in Computing Systems, 2013.
- [3] V. Voykinska, S. Azenkot, S. Wu, and G. Leshed, "How blind people interact with visual content on social networking services," Proc. 19th ACM Conf. Computer-Supported Cooperative Work & Social Computing, 2016.
- [4] J.P. Bigham, S. White, T. Yeh, C. Jayant, H. Ji, G. Little, A. Miller, R.C. Miller, R. Miller, A. Tatarowicz, and B. White, "VizWiz: nearly real-time answers to visual questions," Proc. 23rd annual ACM Symp. User interface software and technology, 2010.
- [5] "Be My Eyes," <https://www.bemyeyes.com/>
- [6] "Envision AI," <https://apps.apple.com/jp/app/envision-ai/id1268632314>
- [7] "TapTapSee," <https://taptapseeapp.com/>
- [8] H. Kacorri, K.M. Kitani, J.P. Bigham, and C. Asakawa, "People with visual impairment training personal object recognizers: Feasibility and challenges," Proc. CHI Conf. Human Factors in Computing Systems, 2017.
- [9] J. Balata, Z. Mikovec, and L. Neoproud, "BlindCamera: Central and golden-ratio composition for blind photographers," Proc. Multimedia, Interaction, Design and Innovation on ZZZ, 2015.
- [10] S. White, H. Ji, and J.P. Bigham, "EasySnap: real-time audio feedback for blind photography," Adjunct proceedings of the 23rd annual ACM Symp. User interface software and technology - UIST '10, 2010.
- [11] M. Vázquez and A. Steinfeld, "Helping visually impaired users properly aim a camera," Proc. 14th international ACM SIGACCESS Conf. Computers and accessibility, 2012.
- [12] M. Vázquez and A. Steinfeld, "An assisted photography framework to help visually impaired users properly aim a camera," ACM Trans. Computer-Human Interaction, vol.21, no.5, pp.1–29, 2014.
- [13] M. Iwamura, N. Hirabayashi, Z. Cheng, K. Minatani, and K. Kise, "VisPhoto: Photography for people with visual impairment as post-production of omni-directional camera image," Proc. ACM CHI Conf. Human Factors in Computing Systems Extended Abstracts Proceedings, 2020.
- [14] 岩村雅一, 平林直樹, 程 征, 南谷和範, 黄瀬浩一, "VisPhoto: 全方位カメラを用いた視覚障害者のための写真撮影支援システム," インタラクシオン 2020, 2020.
- [15] 岩村雅一, 平林直樹, 程 征, 南谷和範, 黄瀬浩一, "[解説] 全方位カメラを用いた視覚障害者の写真撮影支援," 日本知能情報ファジィ学会学会誌, June 2020.
- [16] D. Frohlich and E. Tallyn, "Audiophotography: practice and prospects," Proc. CHI '99 extended abstracts on Human factors in computing systems, 1999.
- [17] D.M. Frohlich, Audiophotography, Springer Netherlands, 2004.
- [18] D. Adams, L. Morales, and S. Kurniawan, "A qualitative study to support a blind photography mobile application," Proc. 6th Int. Conf. Pervasive Technologies Related to Assistive Environments, 2013.
- [19] D. Adams, S. Kurniawan, C. Herrera, V. Kang, and N. Friedman, "Blind photographers and VizSnap," Proc. 18th Int. ACM SIGACCESS Conf. Computers and Accessibility, 2016.
- [20] D. Gurari, Q. Li, A.J. Stangl, A. Guo, C. Lin, K. Grauman, J. Luo, and J.P. Bigham, "VizWiz grand challenge: Answering visual questions from blind people," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018.
- [21] M. Iwamura, Y. Inoue, K. Minatani, and K. Kise, "Suitable camera and rotation navigation for people with visual impairment on looking for something using object detection technique," Proc. 17th Int. Conf. Computers Helping People with Special Needs, Part I, pp.495–509, 2020.
- [22] "OrCam MyEye 2," <https://www.orcam.com/en/myeye2/>
- [23] P. Meijer, "The voice - new frontiers in sensory substitution," <https://www.seeingwithsound.com/>
- [24] Y.-C. Su, D. Jayaraman, and K. Grauman, "Pano2Vid: Automatic cinematography for watching 360° videos," Proc. 13th Asian Conf. Comput. Vis., 2016.
- [25] Y.-C. Su and K. Grauman, "Making 360° video watchable in 2d: Learning videography for click free viewing," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017.
- [26] H.-N. Hu, Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun, "Deep 360 pilot: Learning a deep agent for piloting through 360° sports videos," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017.
- [27] Y.-C. Lin, Y.-J. Chang, H.-N. Hu, H.-T. Cheng, C.-W. Huang, and M. Sun, "Tell me where to look: Investigating ways for assisting focus in 360° video," Proc. CHI Conf. Human Factors in Computing Systems, 2017.
- [28] W.-S. Lai, Y. Huang, N. Joshi, C. Buehler, M.-H. Yang, and S.B. Kang, "Semantic-driven generation of hyperlapse from 360 degree video," IEEE Trans. Vis. Comput. Graphics, vol.24, no.9, pp.2610–2621, Sept. 2018.
- [29] 原 崇之, 柿沼明宏, "球面上の重要度分布予測に基づく全天球画像の初期視点生成," 信学論 (D), vol.J101-D, no.9, pp.1237–249, June 2018.
- [30] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Technical report, Stanford University Computer Science, 2005. <http://graphics.stanford.edu/papers/lfcamera/>
- [31] "Lytro — wikipedia," <https://en.wikipedia.org/wiki/Lytro>
- [32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in Neural Information Processing Systems 28, pp.91–99, 2015.
- [33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only

look once: Unified, real-time object detection,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016.

[34] Trieu, “Github - thtrieu/darkflow: Translate darknet to tensorflow,” <https://github.com/thtrieu/darkflow>

[35] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C.L. Zitnick, and P. Dollár, “Microsoft coco: Common objects in context,” arXiv preprint arXiv:1405.0312, 2014. <https://arxiv.org/abs/1405.0312>

[36] Y.-L. Chen, J. Klopp, M. Sun, S.-Y. Chien, and K.-L. Ma, “Learning to compose with professional photographs on the web,” Proc. ACM on Multimedia Conference, 2017.

[37] Y. Chen, “View finding network,” <https://github.com/yiling-chen/view-finding-network>

[38] 高嶋慶伍, 河合隆哲, 岩村雅一, 南谷和範, 黄瀬浩一, “視覚障害者の周囲の視覚情報取得における QA システムの有用性に関する調査,” ヒューマンインタフェース学会 アクセシブル・インタフェース研究会 SIGACI-27, March 2021.

[39] 河合隆哲, 高嶋慶伍, 岩村雅一, 南谷和範, 黄瀬浩一, “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,” ヒューマンインタフェース学会 アクセシブル・インタフェース研究会 SIGACI-27, March 2021.

## 付 録

### 1. 実験 1 の詳細

表 A・1 に視覚障害者を対象としたアンケートの回答を記す。自由記述の回答は表 A・2～A・8 に記す。

表 A・1 《実験 1》視覚障害者対象のアンケートへの回答

(a) VisPhoto 使用前: 写真への需要

	VI-1	VI-2	VI-3	VI-4	VI-5	VI-6	VI-7	VI-8
Q1. どれくらい写真を撮りたいですか？ (5 段階: 撮りたい (5)～撮りたくない (1))	5	4	5	2	4	3	4	5
Q2. 実際写真を撮る機会はありますか？ (5 段階: 機会がある (5)～機会がない (1))	2	1	1	4	4	1	2	4
Q3. どれくらいの頻度で写真を撮りますか？ (1: 毎日, 2: 週に 3～4 回, 3: 週に 1 回, 4: 月に 1 回, 5: 年に数回.)	5	5	5	2	3	5	4	3 か 4
Q4. 写真を撮る目的は何ですか？	表 A・2 に記載.							
Q5. うまく写真を撮れますか？ (5 段階: 撮れる (5)～撮れない (1))	1	2	2	3	3	2	2	4
Q6. 写真を撮るときに使用するデバイスは何ですか？	iPhone	iPhone	iPhone	iPhone	iPhone	iPhone	iPhone Theta	iPhone
Q7. 現在使っている写真撮影機器に満足していますか？ (5 段階: 満足 (5)～不満 (1))	2	3	2	2	3	1	2	2
Q8. 写真を撮るときどんな失敗をしますか？	表 A・3 に記載.							
Q9. 何の写真をよく撮りますか？	表 A・4 に記載.							
Q10. うまく撮影するための工夫や対策があれば教えてください.	表 A・5 に記載.							
Q11. 撮影した写真はどのように整理していますか？	表 A・6 に記載.							
Q12. 写真に写るものの名前がファイル名に入っていたら便利ですか？	はい	はい	はい	はい	はい	はい	はい	はい

(b) VisPhoto 使用后: VisPhoto の有用性

	VI-1	VI-2	VI-3	VI-4	VI-5	VI-6	VI-7	VI-8
Q13. VisPhoto の使い方にはすぐに理解できましたか？ (5 段階: 早い (5)～遅い (1))	5	5	5	5	4	5	5	4
Q14. 実際に VisPhoto を使いたいと思いますか？ (5 段階: 思う (5)～思わない (1))	5	5	5	5	5	3	3	4
Q15. VisPhoto があれば、もっと写真を撮りたいと思いますか？ (5 段階: 思う (5)～思わない (1))	5	5	4	5	5	4	3	3
Q16. VisPhoto にどんな機能があれば便利ですか？	表 A・7 に記載.							
Q17. 写真撮影時に撮りたかったものを覚えていますか？ (5 段階: 覚えている (5)～覚えていない (1))	4	4	1	5	4	5	4	5
Q18. 写真撮影時に撮りたかったものをどれだけの期間覚えていられると思いますか？ (1: 半日未満, 2: 1 日, 3: 2～3 日, 4: 1 週間, 5: 1 か月.)	5	3	4	4	4	5	3	3
Q19. どうすれば写真の内容を長く覚えておけると思いますか？	表 A・8 に記載.							

表 A-2 《実験 1》視覚障害者アンケート Q4「写真を撮る目的は何ですか?」の回答

VI-1	自分の記憶用と SNS.
VI-2	記録, 記念.
VI-3	記憶と思い出 (写真に何が写っているかは, 日記と照合して類推する).
VI-4	業務.
VI-5	記念.
VI-6	人に見せる (昔話のときに晴眼者に見てもらうため).
VI-7	家族に見せる. 家族と一緒にどこかに行って, 他の人に見せる.
VI-8	人に見せる.

表 A-3 《実験 1》視覚障害者アンケート Q8「写真を撮るときどんな失敗をしますか?」の回答

VI-1	距離感がうまく掴めない. ピントが合わない. 動くものはタイミングが合わない.
VI-2	被写体が写真に入らない.
VI-3	被写体が写真に入らない.
VI-4	ピントが合わない. 被写体に近過ぎたり遠過ぎたりする. 撮影した写真が明る過ぎたり暗過ぎたりする.
VI-5	写したいものの全体が写真に入っていない.
VI-6	暗過ぎて何か分からない. 指が写っている.
VI-7	写したいものの全体が写真に入っていない.
VI-8	撮影した写真が暗過ぎる. 写したいものの全体が写真に入っていない.

表 A-4 《実験 1》視覚障害者アンケート Q9「何の写真をよく撮りますか?」の回答

VI-1	食べ物, 風景.
VI-2	食べ物, 風景.
VI-3	食べ物.
VI-4	コーラのペットボトル (業務で).
VI-5	景色, ご飯, 建物.
VI-6	折り紙, 人.
VI-7	旅先の風景, 全方位カメラを使って周囲の状況を丸ごと.
VI-8	食べ物, 人, 商品 (同じものが欲しくて尋ねるとき).

表 A-5 《実験 1》視覚障害者アンケート Q10「うまく撮影するための工夫や対策があれば教えてください。」の回答

VI-1	ない.
VI-2	写真を沢山撮影して晴眼者に良いものを選んでもらう.
VI-3	ない (写真の出来を気にしないようにする).
VI-4	ない.
VI-5	人を写すときは iOS の VoiceOver の人検出機能を使う.
VI-6	写真を撮影した直後, 被写体が写真に含まれているかを晴眼者に聞く.
VI-7	人を写すときは iOS の VoiceOver の人検出機能を使う. 全方位カメラで撮影して, 見る人に任せる.
VI-8	手で被写体の位置を確認する. 人を撮るときは声を出してもらおうか, iOS の VoiceOver の人検出機能を使う.

表 A-6 《実験 1》視覚障害者アンケート Q11「撮影した写真はどのように整理していますか?」の回答

VI-1	ファイル名に場所の情報を含める.
VI-2	写真は整理できていない.
VI-3	何もしない.
VI-4	写したものをファイル名にする. グループ分けする. 撮影目的をファイル名に記号で書く.
VI-5	ファイル名にタイトルを付ける.
VI-6	ファイル名に日付を加える.
VI-7	ファイル名に日付を加える.
VI-8	写真を 3~4 枚撮って, 一番良いものを人に選んでもらい, 残りは消す. ファイル名に日付と何が写っているかを入れて, イベントごとにフォルダを作って管理する.

表 A-7 《実験 1》視覚障害者アンケート Q16「VisPhoto にどんな機能があれば便利ですか?」の回答

VI-1	写真選択ページでどれくらい物体がはっきり写っているかを教えてくれると便利. 写真に自分自身が写り込むのが気になるので, 片方の魚眼レンズだけ (つまり視野が 180 度だけ) に切り換えられると良い. 写真整理のためのアルバム機能.
VI-2	GPS 機能. 事前登録により誰が写っているかを教えてくれる機能.
VI-3	自分一人でデバイスを操作できるように, Theta に音声ガイドを付けて欲しい.
VI-4	無回答.
VI-5	シャッターを押す前に被写体の状況 (被写体との距離, 何が写っているか, 明るさが適切か, 被写体にピントが合っているか) が分かる機能.
VI-6	切り出した「写真」に結果的に何が写っているかを伝える機能 (物体の選択結果が必ずしも完全に反映されるわけではないので, 物体が何も検出されなくても, 方向を選択して (例えば「1 時方向」の) 写真を切り取る機能. 選択した対象一つだけを拡大して切り出す機能.
VI-7	選択した物体が多い場合に結果が気になるため, 切り出した「写真」に何が写っているかを伝える機能. 作成された「写真」がどの方向を切り取ったのかが分かる機能. 過去に撮影した全方位画像をアップロードして処理する機能. 屋外でも使えるように物体認識の精度を上げる. 写真の内容を説明する機能 (例えば人がどこを向いているか, 人がどれくらい離れているかなど). ユーザーインターフェースの物体選択ページで「全て含める」, 「全て含めない」, 「リセット」ボタンがあると便利. また, 切り出し処理中であることを伝えて欲しい.
VI-8	撮影したいものにレンズを向けて撮影する機能.

表 A-8 《実験 1》視覚障害者アンケート Q19「どうすれば写真の内容を長く覚えておけると感じますか？」の回答

VI-1	写真のファイル名を工夫する。
VI-2	写真撮影した日時、場所、写っている人、何をしている写真かという情報を利用する。
VI-3	無回答。
VI-4	無回答。
VI-5	写真にタイトルを付ける。ボイスメモ機能は非常に良いが、一つずつ聞かなければならないので、音声認識を適用して検索できるようにして欲しい。
VI-6	写真データにメモを残す。日付等から思い出す。あらかじめ自分の写真撮影パターンを把握しておく（例：朝ごはんを撮る、初めて会った人の写真を撮る等）。
VI-7	VisPhoto の音声メモがあるので、記憶できる期間は延びていると思う。それがなければ 1 日程度しか覚えていられないと思う。物体認識の情報を使う。
VI-8	写真を撮った状況を記録する。

(2020 年 11 月 13 日受付, 2021 年 3 月 8 日再受付, 4 月 14 日早期公開)



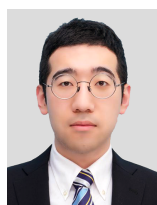
岩村 雅一 (正員)

1998 年東北大・工・通信卒。2003 年同大大学院博士課程了。同年同大大学院工学研究科助手。2004 年阪府大大学院工学研究科助手。助教を経て、2011 年より同准教授。博士 (工学)。文字・物体認識、視覚障害者支援などの研究に従事。2006 年本会論文賞、2007 年 IAPR/ICDAR Best Paper Award, 2010 年 IAPR Nakano Award, ICFHR Best Paper Award, 2011 年 IAPR/ICDAR Young Investigator Award, 2017 年 MVA Best Paper Award 各受賞。2016 年～2018 年 IAPR TC11 (Reading Systems) Vice Chair。2020 年～本会パターン認識・メディア理解研究専門委員会副委員長。電子情報通信学会、情報処理学会、IEEE、ACM 各会員。



平林 直樹

2020 年大阪府立大学工学域電気電子系学類卒。現在、大阪府立大学大学院工学研究科博士前期課程在学中。視覚障害者支援の研究に従事。



程 征

2018 年大阪府立大学工学域電気電子系学類卒。2020 年大阪府立大学大学院工学研究科博士前期課程了。在学中は視覚障害者支援の研究に従事。



南谷 和範

(独) 大学入試センター研究開発部准教授。博士 (政治学)。国立障害者リハビリテーションセンター研究所流動研究員。(独) 大学入試センター入学者選抜研究機構特任准教授等を経て現職。2017 年度ヒューマンインタフェース学会コミュニケーション支援研究賞。



黄瀬 浩一 (正員：フェロー)

1986 年阪大・工・通信卒。1988 年同大大学院博士前期課程了。同年同大大学院博士後期課程入学。1990 年阪府大・工・電気助手。現在、同大大学院工学研究科教授、並びに同大学文書解析知識科学研究所所長。博士 (工学)。2000 年～2001 年ドイツ人工知能研究センター客員教授。文書画像解析、情報検索、画像認識、行動認識などの研究に従事。2006 年度本学会論文賞、2007 年、2013 年 IAPR/ICDAR Best Paper Award, 2010 年 IAPR Nakano Award, ICFHR Best Paper Award, 2011 年 ACPR Best Paper Award, 2017 年 MVA Best Paper Award 各受賞。2016 年まで、IAPR TC11 (Reading Systems) Chair, 2013 年から、International Journal of Document Analysis and Recognition (IJ DAR) Editor-in-Chief。電子情報通信学会、情報処理学会、人工知能学会、電気学会、IEEE、ACM などの会員。