

解説

全方位カメラを用いた視覚障害者の写真撮影支援[†]

岩村 雅一^{*1}・平林 直樹^{*1}・程 征^{*1}・南谷 和範^{*2}・黄瀬 浩一^{*1}

1. はじめに

視覚障害者も写真を撮る，ということをご存知だろうか？ 2011年に発表された調査[1]では，全盲，弱視ともに約70%の人が最近写真を撮ったと報告している．写真を撮る目的・対象は，晴眼者（目が見える人）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，「インスタ映えしたい」といった欲求もあるようだ．その他，[2–5]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，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，晴眼者とは別の目的の写真撮影もあるという．

では，視覚障害者はどのように写真を撮るのだろうか？ 弱視であれば，限られた視力で撮影対象にカメラを向けることもできるかもしれない．しかし全盲ではそうもいかない．全盲の人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．ただ，残念ながら，対象は人だけである．人だけでなく，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「もうちょっと右」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者に指示するものである[1,6–9]．この方法の難点は，利用者が指示通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．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ったら，がっかりである．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だが，全盲の人には難しいのである．

そこで我々は，利用者が撮影対象にカメラを向けなくてもいい撮影方法として，全方位カメラを使う方法を提案している[10,11]^{1,2}．全方位カメラ（全天球カメラとも言う）とは，文字通り，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．2020年現在，様々な民生品が販売されており，安い



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

ものは2万円台で購入できる．つまり，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無い．これが提案手法の基本的なアイデアである．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，歪んでいる．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，歪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．実際，全方位カメラで撮影した動画から通常の動画への変換が試みられている[12–16]．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

[†] Photo Taking Assistance with Omni-Directional Camera for People with Visual Impairment
Masakazu IWAMURA, Naoki HIRABAYASHI, Zheng CHENG,
Kazunori MINATANI, and Koichi KISE

^{*1} 大阪府立大学大学院工学研究科
Graduate School of Engineering, Osaka Prefecture University

^{*2} 大学入試センター
National Entrance Examination Center

1 CHI2020 Extended Abstract [10] のビデオプレビュー：<https://youtu.be/hOZB-Anhs18>

2 インタラクシオン 2020 [11] でのデモの中継ビデオ：<https://youtu.be/q1njuPeysE?t=5637>

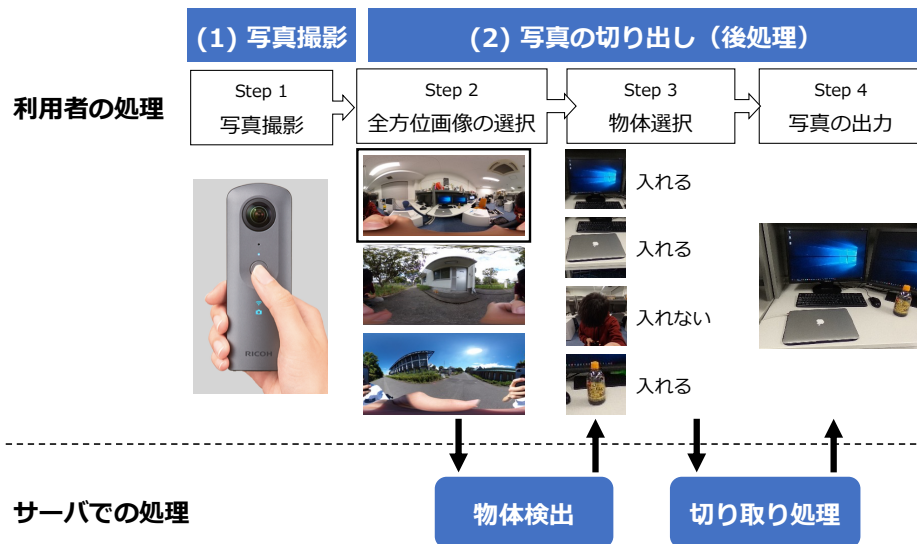


図3 提案システムの概要

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。[12, 13]も物体情報を考慮に入れて切り出すが、情報量の大きさを基準に切り出すものであり、利用者の希望を反映する提案手法とは異なる。

以降では、我々が提案する手法 VisPhoto の概要を紹介し、全盲の視覚障害者 8 名と晴眼者 10 名による提案手法の評価結果の概要を記す。

2. 提案手法 VisPhoto

図3に提案手法の概要を示す。提案手法は、(1) 全方位カメラでの写真撮影と(2) 写真を切り出す後処理の2つの処理に大別できる。(1)は図中の Step 1, (2)は Step 2~4 から成る。以下、順を追って説明する。

Step 1: 全方位カメラでの写真撮影

Step 1では、全方位カメラ上で動く専用の写真撮影アプリを起動した後、利用者が写真を撮りたいときに全方位カメラのシャッターボタンを押すことでカメラを中心とした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。そのため、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、Haradaら [17] 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。具体的には、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。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。これは何らかの音声情報を残すことを利用者に促すためである。撮影された全方位画像と録音された音声は、全方位カメラがインターネットに接続された際に自動的にサーバにアップロードされる。なお、撮影された全方位画像は自動的に天頂補正される。

前述のアプリの実装面は以下の通りである。全方位カメラとしては、2つの魚眼レンズを持つ Ricoh Theta V または Z1 を用いる。これらのカメラでは Android OS が動作しており、Android アプリを作成して実行することで、動作をカスタマイズできる。そこで、全方位画像の撮影と音声の録音、インターネットへのアップロード機能を供えたアプリを作成した。全盲の視覚障害者が使用することを想定して、アプリの状態が音声で理解できるようになっている。例えば、アプリの起動時には“welcome to VisPhoto”，録音開始時に“start recording”，録音終了時には“stop recording”，写真撮影の準備ができた時には“ready”，アプリ終了時には“VisPhoto is shutting down”等である。しかし、残念ながら全方位カメラ本体は視覚障害者が使用することが想定されておらず、カメラの状態は LED の点灯と点滅で知らせる仕様である (Z1 はディスプレイでの表示もある)。そのため、VisPhoto 専用アプリを起動するまでは完全な無音であるというのが現時点での欠点である。回避策としては、全盲の利用者に対しては「電源ボタンを2秒押し、その後16秒待った後に…」のように覚えてもらう他ない。この改善は今後の課題である。

Step 2: 全方位画像の選択

後処理である Step 2~4 では、利用者の希望を反映して、利用者が望む写真を切り出す。この操作をウェブインタフェースを用いて行う。利用者として全盲の視覚障害者を想定しているので、スクリーンリーダーを使用して理解してもらえるように、ウェブページの主要な情報はテキストで出力する。見出し語に HTML の h4 タグや h5 タグを付与して、項目間を素早く移動できるように工夫している。

Step 2では、サーバにアップロードされた写真(全方位画像)の中から、切り出し処理したい写真を選択する。図4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。写真が見えなくても写真が選択できる



図4 Step 2 (写真選択) のウェブインタフェース

ように、写真以外の情報として、Step 1 で録音した音声、撮影日時、それに加えて画像に写るもの（物体）のリストを写真と共に表示する。物体リストについての詳細は、次項で述べる。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（GPS 情報）を同時に表示するのが望ましいが、Ricoh Theta は GPS 情報を取得しないため、実現できていない。

Step 3: 物体選択

Step 3 では、Step 2 で選択した写真に物体検出技術を適用し、検出された物体の一覧を利用者に提示して、利用者に各物体を写真に含めたいか、含めたくないか、あるいはどちらでも良いかを入力してもらう。物体検出とは、あらかじめ登録（学習）してあるカテゴリの物体を検出して、その位置を出力するコンピュータビジョン分野の主要なタスクである（[18, 19] 等）。物体の位置は物体の外接方形（Bounding Box）で表すこ

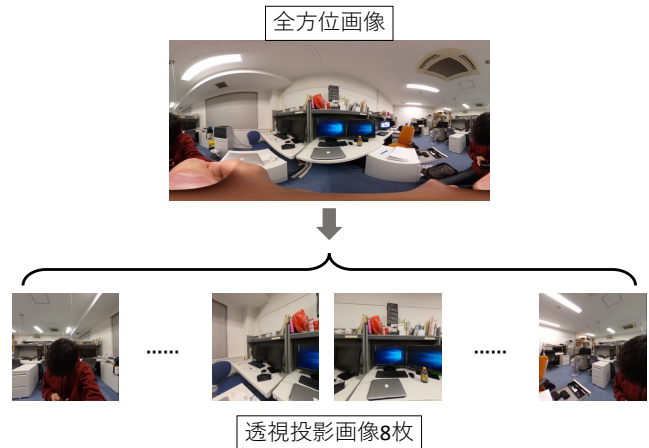


図5 全方位画像を透視投影画像に変換してから物体を検出する。

とが多い。提案手法の現在の実装では、深層ニューラルネットワークに基づく物体検出手法の一つである You Only Look Once (YOLO) Version 1 [18] の TensorFlow 実装 [20] を用いた。このニューラルネットワークは、80 カテゴリの物体カテゴリから成る COCO dataset [21] で事前に学習した。YOLO 等の物体検出手法は普通のデジタルカメラで撮影した画像（透視投影画像）を対象にしており、歪んだ全方位画像にそのまま適用できない。そのため、提案手法では図5のように全方位画像を8枚の透視投影画像に変換してから YOLO を適用する。8枚の透視投影画像の中心は全方位画像の赤道上に45度ずつに並んでおり、各画像は画角が上下左右90度ずつになる大きさである。各透視投影画像で検出された物体の情報は、元の全方位画像に投影される。現在の実装では全方位画像に大きく写り、複数の透視投影画像で検出される物体が2つ以上検出されてしまう問題があり、この解決は今後の課題である。

図6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例えて表す方法で、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、画像の横方向を見たとき、中央が12時、両端が6時に相当する。つまり、横方向に画像を4等分したとき、分割した画像の境界（両端を含む）を左から見ていくと、6時（左端）、9時、12時（中央）、3時、6時（右端）になる。図6の例では、1時の方向に bench、3時の方向に person、5時と11時の方向にそれぞれ car が検出されている。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3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。

Step 4: 写真の出力

Step 4 では、Step 3 で利用者が入力した希望に合わせて写真を切り出し、図7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、切り出した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、後から何を撮影したのかを理



図6 Step 3 (物体選択) のウェブインタフェース



図7 Step 4 (写真の出力) のウェブインタフェース

解するための工夫である。

利用者の希望に合わせて、かつ見た目が良い写真を切り出すために、以下のエネルギー関数最小化問題を定義する。この処理は透視投影画像上で行うため、その準備として、含めたい物体を全て含む透視投影画像を作成する³。そのために、全方位画像上で、含めたい画像の中心座標の重心を計算する。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、

3 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という制限がある。

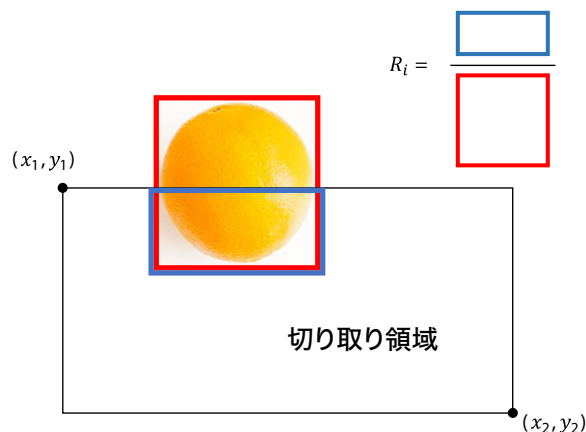


図8 R_i の定義. 赤い矩形は物体 (オレンジ) の外接方形, 青い矩形は物体の領域と切り取り領域の共通部分を表す. このとき, R_i は赤い矩形の青い矩形に対する面積比として定義される

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

で与えられる。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。これで透視投影画像が準備できたので、これ以降「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 $\theta = (x_1, y_1, x_2, y_2)$ で定義する。ここで (x_1, y_1) と (x_2, y_2) は領域の左上と右下の座標である。また, 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) [22, 23] の出力でスコアである。VFN は写真の美しさを評価する deep neural network 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8に示すように、 i 番目の物体と切り取り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 θ を求める。

3. 評価

提案手法の評価のために、2種類の調査を行なった。1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。2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名に参加してもらった。

3.1 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。実験参加者の基本情報を表1に記す。実験参加者には最初に写真撮影に関する項目をインタビュー形式で質問し、回答を得た。そして、提案手法の使い方

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
26	女性	生まれつき弱視. 7 歳から病患で失明.
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.
33	女性	6 歳まで健常. 6 歳から弱視. 13 歳で失明.
28	男性	生まれつき左目は盲, 右目は弱視. 7 歳で失明.
29	男性	2 歳まで健常. 2 歳で失明.
23	女性	2 歳まで健常. 2 歳で失明.
47	男性	生まれつき弱視. 6 歳頃失明.
28	女性	生まれつき右目は盲, 左目は弱視. 25 歳で失明.

を説明した後, 全方位カメラを数日間貸し出し, 提案システムを自由に使用してもらった. 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.

写真撮影に関する質問への回答

以下に写真撮影に関する質問への回答の概略を記す.

- 「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: 数字が大きい方がより撮りたい) に対して, 6 人の参加者が 5 または 4 と答えた.
- 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.
- 「写真を撮る目的は何ですか?」という質問では, 彼ら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.
- 「どれくらい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, 「現在使っている写真を撮るデバイス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, 実験参加者は概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.
- 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.
- 写真撮影に失敗する原因は, 概ね既発表文献で挙げられた通りの内容であった. 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.
- 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.
- 「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.

提案手法 VisPhoto に関する質問への回答

以下に提案手法 VisPhoto に関する質問の回答の概略を記す.

- 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.
- 「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.

- 「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) の回答は 6 名が肯定的に答えた.
- 「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.
- 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, 以下の通りである. 「写真撮影時に撮りたかったものを覚えていますか?」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. 「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 ヶ月が 2 名であった.
- 「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.

以上のように, 視覚障害者を対象としたユーザ調査では, 提案システムの有用性は肯定的に評価された.

3.2 実験 2: 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. 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 (数字が大きい方が上質) 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程, 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 (数字が大きい方が似ている) で評価してもらった. 代表的な成功例と失敗例をそれぞれ表 2 と表 3 に示す. これらの評価より, 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明らかになった.

4. まとめ

本稿では, 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紹介した. このシステムは, 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, 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 2 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システムのプロトタイプを作成し, 全盲の視覚障害者 8 名と晴眼者 10 名による提案手法の評価実験を行い, その結果の概要を記した.

表2 成功したケースの代表例

















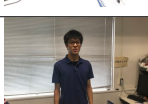



iPhone		VisPhoto		類似度
写真	質	写真	質	
	10		10	5
	10		9	5
	8		10	5
	10		10	5
	10		10	5

表3 失敗したケースの代表例

iPhone		VisPhoto		類似度
写真	質	写真	質	
	9		7	4
	7		5	1
	10		8	2
	10		2	1
	10		4	2

最終的には提案システムを公開して、一般に使ってもらえるようにしたいと考えている。そのための課題はいろいろある。重要な課題の1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、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、さら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。これらの工夫を施したい。また、全方位画像に大きく写る物体から2つ以上の物体が検出されてしまう問題も解決したい。提案システムで使用している Ricoh Theta の問題も改善が望まれる。 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のも重要な改善ポイントである。 Ricoh Theta の GPS が搭載されていないため、GPS 情報が取れない点も改善が望ましい。これらについては我々だけでは解決できないため、ベンダーへの働きかけが必要になる。ユーザインタフェースを洗練することも必要であろう。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無いだろうから、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 研究調査助成を受けて実施した。

参考文献

- [1] C. Jayant, H. Ji, S. White, and J. P. Bigham: "Supporting blind photography," *The Proc. of the 13th Int. ACM SIGACCESS Conf. on Computers and Accessibility*, 2011.
- [2] Be My Eyes: <https://www.bemyeyes.com/> [accessed Sep. 20, 2019]
- [3] Envision AI: <https://apps.apple.com/jp/app/envision-ai/id1268632314> [accessed Feb. 25, 2020]
- [4] TapTapSee: <https://taptapseeapp.com/> [accessed Sep. 20, 2019]
- [5] 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. of the 23rd Annual ACM Symp. on User Interface Software and Technology*, 2010.
- [6] J. Balata, Z. Mikovec, and L. Neoproud: "Blind-Camera: Central and Golden-ratio Composition for Blind Photographers," *Proc. of the Multimedia, Interaction, Design and Innovation*, 2015.
- [7] M. Vázquez and A. Steinfeld: "Helping visually impaired users properly aim a camera," *Proc. of the 14th Int. ACM SIGACCESS Conf. on Computers and Accessibility*, 2012.
- [8] M. Vázquez and A. Steinfeld: "An Assisted Photography Framework to Help Visually Impaired Users Properly Aim a Camera," *ACM Trans. on Computer-Human Interaction*, 2014.
- [9] S. White, H. Ji, and J. P. Bigham: "EasySnap: real-time audio feedback for blind photography," *Adjunct Proc. of the 23rd Annual ACM Symp. on User Interface Software and Technology*, 2010.
- [10] 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," *ACM CHI Conf. on Human Factors in Computing Systems (CHI2020) Extended Abstracts Proc.*, 2020.
- [11] 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," *インタラクション 2020 論文集*, pp. 874-879, 2020.
- [12] H. Hu, Y. Lin, M. Liu, H. Cheng, Y. Chang, and M. Sun: "Deep 360 Pilot: Learning a DeepAgent for Piloting through 360° SportsVideos," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2017.
- [13] W. Lai, Y. Huang, N. Joshi, C. Buehler, M. Yang, and S. B. Kang: "Semantic-Driven Generation of Hyperlapse from 360 Degree Video," *IEEE Trans. on Visualization and Computer Graphics*, 2018.

- [14] Y. Lin, Y. Chang, H. Hu, H. Cheng, C. Huang, and M. Sun: "Tell Me Where to Look: Investigating Ways for Assisting Focus in 360° Video," *Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems*, 2017.
- [15] Y. Su and K. Grauman: "Making 360° Video Watchable in 2D: Learning Videography for Click Free Viewing," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2017.
- [16] Y. Su, D. Jayaraman, and K. Grauman: "Pano2Vid: Automatic Cinematography for Watching 360° Videos," *Proc. 13th Asian Conf. on Computer Vision*, 2016.
- [17] 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. of the SIGCHI Conf. on Human Factors in Computing Systems*, 2013.
- [18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: "You Only Look Once: Unified, Real-Time Object Detection," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2016.
- [19] S. Ren, K. He, R. Girshick, and Jian Sun: "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems 28*, Curran Associates, Inc., pp.91-99, 2015.
- [20] Trieu. GitHub-thtrieu/darkflow, Translate Darknet to Tensorflow: <https://github.com/thtrieu/darkflow> [accessed Sep. 20, 2019]
- [21] T. 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, 2014.
- [22] Y. Chen, View Finding Network: <https://github.com/yiling-chen/view-finding-network> [accessed Sep. 18, 2019]
- [23] Y.-L. Chen, J. Klopp, M. Sun, S.-Y. Chien, and K.-L. Ma: "Learning to Compose with Professional Photographs on the Web," *Proc. of the 2017 ACM on Multimedia Conf. (MM '17)*, 2017.
(2020年3月13日 受付)

[問い合わせ先]

〒599-8531 大阪府堺市中区学園町 1-1
大阪府立大学 大学院工学研究科 知能情報工学分野
岩村 雅一
E-mail: masa@cs.osakafu-u.ac.jp

著者紹介



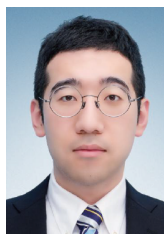
いむら まさかず
岩村 雅一 [非会員]

1998年東北大学工学部通信工学科卒業。2003年同大学大学院工学研究科 博士課程後期3年課程修了。同年同大学院工学研究科助手。2004年大阪府立大学大学院工学研究科助手。助教を経て、2011年より准教授。2010年ドイツ人工知能研究センター訪問研究員。2012年 Computer Vision Center (スペイン、バルセロナ) 訪問研究員。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年電子情報通信学会論文賞、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。2016~2018年 IAPR TC11 (Reading Systems) Vice Chair。電子情報通信学会、情報処理学会、IEEE、ACM 各会員。



ひらばやし なおき
平林 直樹 [非会員]

2020年大阪府立大学工学域電気電子系学類卒業。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。



まへはら せい
程 征 [非会員]

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



みなみ かずのり
南谷 和範 [非会員]

(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。



おがき こういち
黄瀬 浩一 [非会員]

1986年大阪大学工学部通信工学科卒。1988年同大学院博士前期課程了。同年同大学院博士後期課程入学。1990年阪府大・工・電気助手。現在、同大学院工学研究科教授。博士(工学)。2000~2001年ドイツ人工知能研究センター客員教授。2013年度より2016年度まで日本学術振興会学術システム研究センター研究員。文書画像解析、画像認識、行動解析などの研究に従事。2006年度電子情報通信学会論文賞、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, 現在、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。