VisPhoto: 全方位カメラを用いた視覚障碍者のための写真撮影支援システム

岩村 雅一[1,a)]　平林 直樹[1,b)]　程 征[1,c)]　南谷 和範[2,d)]　黄瀬 浩一[1,e)]

概要：多くの視覚障碍者が写真を撮影したいと望んでいることが知られている．しかし撮影する際に，カメラを被写体に向けることが難しいという問題に直面する．本論文は，VisPhoto という新しい写真撮影支援システムを提案することで，この問題を解決する．既存手法と異なり，提案システムは全方位画像のポストプロダクションで写真を生成する．つまり，シャッターボタンを押すと提案システムは全方向画像を撮影する．ポストプロダクションでは，画像から検出された物体に対し，ユーザーはそれを写真に含めるか含めないかをを選択する．最後に，提案システムはユーザーが選択した物体を切り取り，同時に写真の美しさを考慮した画像を出力する．ユーザー調査には視覚障碍者と晴眼者が参加し，提案システムの利点，欠点，有用性，および生成された写真の品質について評価してもらった．

1. はじめに

多くの視覚障碍者[*1]が，晴眼者と同じように写真を撮りたいと望んでいることが知られている．写真を通して「印象に残る場面を画像で残したい」,「体験を他人と共有して感動を分かち合いたい」という需要は障碍の有無に関わらず晴眼者と同じである [1], [2], [3].

彼らのそうした要求があるにも関わらず，視覚障碍者が満足できる写真を撮影するにはいくつかの課題を克服する必要がある．Harada らはフレーム，照明およびフォーカスの問題を論じた [1]．Balata らは構図に焦点を当てた [4]．しかし，視覚障碍者が写真を撮影するとき，そこには根本的な問題が存在する．それは被写体が見えないため，写したいもの全体をフレームに入れることが難しいということだ．この問題点を解決するために，リアルタイムで処理が可能なアプリケーションが開発された． [2], [4], [5], [6], [7] は，音や振動を使ってユーザーにカメラの向きを指示する補助ツールである．iOS では VoiceOver を用いることで撮影時にフレームに入った人数を伝えることが可能である．しかし，これらのアプローチでは，カメラを被写体に向けるのに時間がかかってしまう．したがって，被写体が静止しているか，動きが遅い場合にのみ使われる．

本論文では，上記の問題を解決するために全方位カメラを用いて，VisPhoto と呼ばれる新しい写真撮影手法を提案する．提案システムでは，全方位カメラを用いることで撮影者の周囲の景色すべてを一度に撮影でき，カメラを被写体に向けるために時間をかける必要がない．全方位画像を撮影した後，ユーザーは後日特定の領域を切り出すことができる．つまり，切り出し処理は通常の写真を撮影することに相当する．言い換えれば，提案システムは全方位画像にポストプロダクション[*2]を行うことで画像を生成する．

提案システムの利点は次の通りである．

(1) 上記のように，ユーザーはカメラを被写体に向けることが不要である．
(2) カメラの回転は自動的に修正できる．したがって，ユーザーが撮影するときにカメラの角度を気にせずに使用できる．
(3) 原則，ユーザーは動いている被写体も撮影できる．一方，従来のアプローチでは容易ではない [2], [4], [5], [6], [7].
(4) 現在主流の手法と異なり，リアルタイムのフィードバックは必要ない．したがって，切り出しの前に計算が重い処理も可能である．本論文では，我々は全方位画像に対し物体検出を行った．物体検出は適切なデバイスで実行された場合は時間がかからないが，スマートフォンなどのポータブルデバイスでは時間がかかる．

[1] 大阪府立大学
[2] 大学入試センター
[a)] masa@cs.osakafu-u.ac.jp
[b)] hirabayashi@m.cs.osakafu-u.ac.jp
[c)] zheng386@m.cs.osakafu-u.ac.jp
[d)] minatani@rd.dnc.ac.jp
[e)] kise@cs.osakafu-u.ac.jp

[*1] [1] に倣い，本論文では「視覚障碍者」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．

[*2] ポストプロダクションは，生の写真を撮影した後に行われる処理すべての段階を指す [8].

(5) 画像を切り出す際に美しさを評価するアルゴリズムを用いることで，生成した画像は美しいことが期待される．

一方，写真の生成はポストプロダクションまで時間が経つため、ユーザーは写真を撮影した際の被写体と状況を覚えておく必要がある．これはユーザーにとって負担が増加するため，欠点と見なすことになる．この問題点を補うために，[1], [9] のようなボイスメモの録音機能を実装した．録音された音声は，全方位画像が撮影された日付，時刻とともにポストプロダクションのときに提供する．

提案されたシステムの利点，欠点，有用性を評価するために，視覚障碍者 8 人を対象にユーザー調査を実施した．2 時間以内の説明を受けた後，自宅や屋外で数日かけて提案システムを評価してもらった．また，提案システムによって生成された写真の品質を評価するため，10 人の晴眼者を対象にユーザー調査を実施した．提案システムとスマートフォン（iPhone）を使用し，同じ被写体の写真を撮影してもらった．撮影した写真の品質は，別の参加者によって評価された．

2. 関連研究

視覚障碍者が写真を撮影する際に最も使用している補助アプリケーションはおそらく iOS の VoiceOver である．VoiceOver は撮影する前にフレーム内に含まれる人物の数を教えてくれる．同様に，既存研究における現在の主流のアプローチは，視覚障碍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ことである [2], [4], [5], [6], [7]．[4] の表 1 は，feedback modalities と aiming assistance directions について，代表的な方法の違いをまとめている．EasySnap [2], [5] は，被写体にカメラを向ける際に役立つ iPhone アプリケーションである．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ユーザーに提示する．PortraitFramer [2] は EasySnap に似ているが，複数の人物の顔が含まれる場合に使用される．これはフレーム内に含まれる顔の数やサイズと位置を提示する．Vazquez ら [7] は，カメラロールを自動的に修正する機能を提案した．Balata ら [4] は，被写体にカメラを向ける際に中央および黄金比の構図に最適化するスマートフォンアプリケーションを提案した．

2.1 写真の管理，閲覧，共有の補助

Frohlich ら [10], [11] は写真と音声録音を組み合わせる価値を探った．Frohlich らは周囲の環境音が多数のユーザーにとって価値があることを発見した．視覚障碍者向けの研究ではないが，示唆に富んだ研究である．

上記の EasySnap [2] は，アクセス可能なフォトアルバムである．写真が撮影された場所と認識エンジンによって認識された結果が各写真に関連付けられているため，視覚障碍者は写真を確認および共有できる．Harada ら [1] 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 [9], [12] は全盲の人を対象に，写真ライブラリを整理して閲覧できる VisSnap という iPhone アプリケーションを提案した．VizSnap は，ユーザーが被写体にカメラを向ける際，周囲の環境音を録音し，撮影日時および位置情報も記録する．

2.2 写真に関連したサービス

VizWiz [13]，Be My Eyes [14]，TapTapSee [15] などのアプリケーションは，視覚障碍者が写真から情報を取得する際に役立つ．そのため，ユーザーは欲しい情報がある物体の写真を撮る必要がある．しかし，視覚障碍者にとって写真を撮ることは簡単ではない．Gurari ら [16] は、VizWiz の視覚的な質問に ∼28%は回答できないとみなされることを明らかにした．それらの一部は写真の品質によるが，視覚障碍者が写真を撮ることが難しいことを示している．提案システムは晴眼者と共有できる写真を生成することを目的としているが，支援技術にも使用できる．

他にもいくつかのサービスが挙げられる．OrCam MyEye 2 [17] は，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 [18] は，ユーザー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．

2.3 ライトフィールドカメラ

提案システムのプロセスは，ライトフィールドカメラ [19] のプロセスに似ており，撮影した後にユーザーが望んだ画像を生成する．代表的な製品として Lytro [20] は，2018 年にサービスを停止した．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” を保存できるので，後で写真の焦点を合わせることができる．例えば，人と犬が含まれている写真があるとする．ユーザーは人物に焦点が合っている画像，犬に焦点が合っている画像，および両方に焦点が合っていない画像を生成できる．

ライトフィールドカメラと提案システムを比較すると，ライトフィールドカメラは全方向カメラに似ており、リフォーカスされた画像は切り出した画像に似ている。

3. 提案システム

3.1 概要

提案システムの概要を図 1 に示す．写真生成の過程は四段階に分けることができる: 写真撮影とボイスメモの録音，写真選択，物体選択，そして結果の表示だ．Step 1 でユーザーは全方位カメラで写真を撮影し，その後ボイスメ

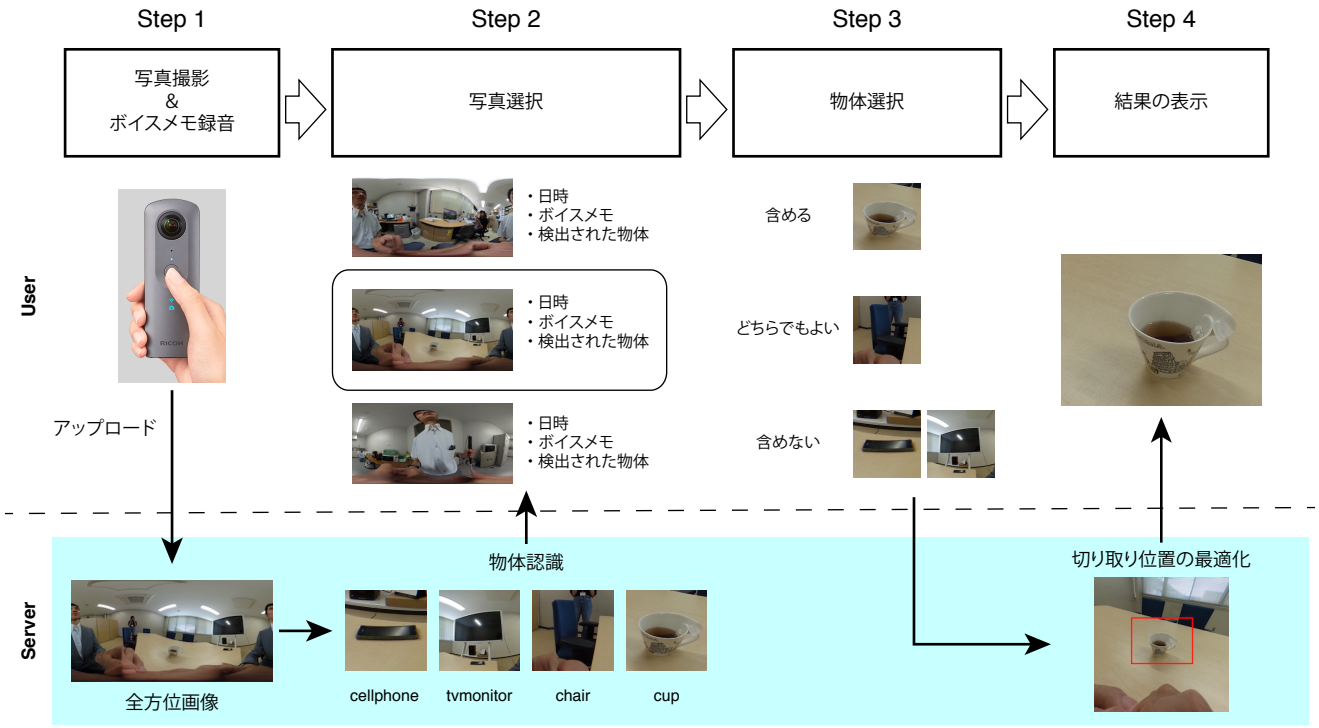
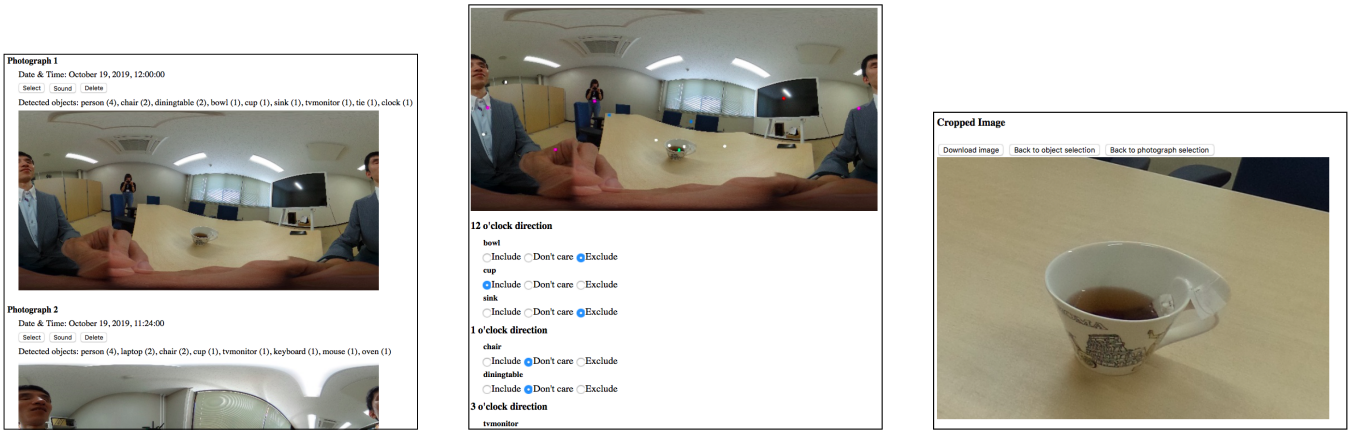


図 1: 提案システムの概要.



(a) 写真選択 (Step 2 of Fig. 1). (b) 物体選択 (Step 3 of Fig. 1). (c) 結果の表示 (Step 4 of Fig. 1).

図 2: ウェブインタフェースのスクリーンショット.

モを録音する．Step 2〜Step 4 で，ユーザーはウェブインタフェースを用いて目的の画像を生成する．Step 2 〜Step 4 のウェブインタフェースのスクリーンーンショットを図 2 に示す．

Step 1 でシャッターボタンが押されたとき，提案システムは全方位画像を撮影する．また同時に音声録音が始まり，シャッターボタンが離されるまで録音が続く．ボタンが 10 秒未満で離されたとき，録音は 10 秒間続く．つまり録音は少なくとも 10 秒間は続く．この設計には二つ理由がある．一つは撮影時の被写体や撮影環境を思い出せるようにユーザーが何かを話すことを促すためである．もう一つは Android の音声録音 API のエラーを防ぐためである．撮影された全方位画像とボイスメモはインターネット接続時に自動的にサーバーにアップロードされる．

Step 2(図 2(a) を参照) で，サーバーに保存された全方位画像からユーザーは一つを選択する．画像の補足情報として撮影した日時，ボイスメモ，検出された物体のリストが与えられる．GPS 情報は全方位カメラが記録しないため与えられない．

Step 3(図 2(b) を参照) で，選択された写真中に検出された物体の中から，ユーザーはその物体を写真に含めるか含めないかを選択する．ウェブインタフェースは検出された物体のリストをクロックポジションを用いて並べる．各オブジェクトについて「含める」，「どちらでもよい」，「含

めない」のラジオボタンが表示される．

Step 4(図 2(c) を参照) で，提案システムはユーザーの選択結果と写真の美しさを考慮した切り取り領域を写真として出力する．写真はダウンロード可能で，ファイル名には選択された物体の名前が含まれる．したがってユーザーはどんな物体が写真に含まれるかを後で知ることができる．

3.2 実装

全方位カメラとして，我々は Ricoh Theta シリーズを使用した．Theta V と Theta Z1 はプラグインをインストールすることで新しい機能を追加することができる [21]．Theta の OS は Android であるため，プラグインは Android アプリとして開発することができる．我々は VisPhoto プラグインを Android アプリとして実装した．視覚障碍者が使用するため，アプリは全て音声ガイドが付いている．たとえば，VisPhoto プラグインが起動した直後に “welcome to VisPhoto” と鳴り，同様に “start recording”，“stop recording”，“ready”，“VisPhoto is shutting down” という音声が鳴る．しかし Theta 自体は音声ガイドに対応しておらず，本体のステータスは LED ライトによって示される．そのため，ユーザーは何の音声フィードバックもなく VisPhoto を起動しなければならない．これが現在の実装の欠点であり，製造元に改善を要請する必要がある．

サーバーに画像がアップロードされるとすぐに物体検出が行われる．我々は *you only look once* (YOLO) version 1 [22] の tensorflow 実装 [23] を用いた．ここで起きる問題として撮影された全方位画像は正距円筒図法であるということだ．YOLO は入力として一般的なデジタルカメラで扱われる透視投影画像を想定している．この問題を解決するため，我々は正距円筒画像を八個の透視投影画像に変換し，それらに対して YOLO を適用した．ここで画像のサイズと数は注意深く決定する必要がある．透視投影画像のサイズが大きすぎると画像の周囲は歪んでしまう．逆に画像のサイズが小さいと大きな物体は検出できない．また透視投影画像の数が多いと処理時間は増え，逆に少ないと画像全体をカバーすることができない．八個の透視投影画像から検出された物体の情報は元の正距円筒画像に投影され，併合される．その後，検出された物体はクロックポジションを用いて並べられる．

「どの物体を写真に含めるか含めないか」というユーザーの選択結果が与えられると，希望した画像が全方位画像から切り出される．この処理はまず「含める」と選択された物体の中心位置の平均を中央とする透視投影画像を生成することから始まる．つまり，新しく生成される透視投影画像の中心の座標 $(x^{\mathrm{C}}, y^{\mathrm{C}})$ は，

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

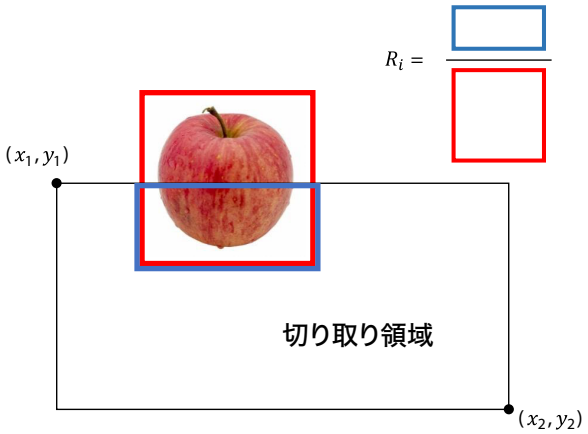


図 3: R_i の定義．赤い矩形は物体 (リンゴ) のバウンディングボックスを表す．青い矩形は物体の領域と切り取り領域の共通部分を表す．そして，R_i は赤い矩形の青い矩形に対する面積比として定義される．

で与えられる．ここで $(x_i^{\mathrm{C}}, y_i^{\mathrm{C}})$ 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．さらに，我々は長方形領域 θ のエネルギー関数 $E(\theta)$ を定義し，それを最小化するような切り取り領域を求める．長方形領域 θ は $\theta = (x_1, y_1, x_2, y_2)$ で定義される．ここで (x_1, y_1) と (x_2, y_2) は領域の左上と右下の座標としてそれぞれ定義される．I_{in} と同様に I_{out} は「含めない」と選択された物体の集合を表す．図 3 に示す通り，R_i は i 番目の物体と切り取り領域の共通部分の面積比とする．

そしてエネルギー関数 $E(\theta)$ を以下で与える．

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

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

ここで $E_{\mathrm{aes}}(\theta)$ は view finding network (VFN) [24], [25] 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，これが画像のもっともよい切り取り領域を見つける．より大きなスコアはより美しい写真である．式 (2) は Nelder-Mead 法により最小化され，それによって目的関数の最小値を求める．

4. ユーザー調査

4.1 視覚障碍者のユーザー調査

システムの有用性，利点，欠点を検証するため，我々は八人の視覚障碍者を対象にユーザー調査を行った．二時間以内の説明を行った後，ユーザーは提案システムを自宅や屋外で数日間使用し，評価した．

我々は二つのアンケート調査を準備した．一つは提案システム使用前に行い，もう一つは提案システム使用後に行った．前者はインタビュー形式で行い，後者はフォーム

を埋めてメールで送信してもらった．参加した視覚障碍者は全員ほぼ全盲で点字使用者であった．

視覚障碍者を対象としたユーザー調査では，提案システムの有用性は肯定的に評価された．

4.2 晴眼者のユーザー調査

我々は提案システムによって生成された画像の質を評価するため，十人の晴眼者に対してユーザー調査を行った．参加者はいくつかの被写体をスマートフォン (iPhone) と提案システムで撮影した．これらの写真の質は別の参加者によって評価してもらった．

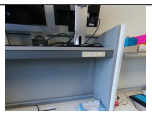
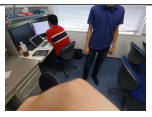
提案システムによって生成された写真の質がスマートフォンで撮影されたものに近ければ成功したとみなすことができる．逆に近くなければ失敗だとみなす．代表的な成功例と失敗例は表 1，表 2 に示す．表中で画像の隣にある「スコア」は別の参加者によって評価された写真の質を示す．「類似度」は提案システムによって出力された画像とスマートフォンで撮影された画像がどれくらい似ているかを五段階で評価してもらったものを表す．より大きな値が良い質を意味している．これらの評価より，提案システムによって切り出された画像は常に晴眼者がスマートフォンで撮影した画像と合致するわけではないが，多くの場合その質は許容できる範囲のものであることが明らかになった．

アンケートの回答について，物体検出の精度は完全に満足できるものではなかったが，許容の範囲内であった．ここで提案システムは特定の物体検出手法に頼っていいるわけではないことを述べておく．仮により良い高精度な物体検出手法が現れたとすれば我々はそれを提案システムに導入することが可能である．

表 1: 成功したケースの代表例．

iPhone		VisPhoto		類似度
画像	スコア	画像	スコア	
	10		10	5
	10		9	5
	8		10	5
	10		10	5
	10		10	5

表 2: 失敗したケースの代表例．

iPhone		VisPhoto		類似度
画像	スコア	画像	スコア	
	9		7	4
	7		5	1
	10		8	2
	10		2	1
	10		4	2

5. 結論

本稿では，VisPhoto と呼ばれる全方位カメラを用いた画期的な写真撮影支援システムを提案した．撮影時にリアルタイムなフィードバックでユーザーによって被写体にレンズを向けられるようにする先行研究と異なり，提案システムでは全方位カメラ使用時にリアルタイムなフィードバックを必要としない．これは全方位カメラはユーザーの周囲をすべて一度に撮影することができ，後から良い領域を切り取ること可能だからだ．良い切り取り領域を選択するために，我々は検出された物体の情報と切り取られた領域の美しさを用いた．提案システムは視覚障碍者と晴眼者によって評価された．視覚障碍者対象のユーザー調査において，提案システムは肯定的に評価された．また，晴眼者対象のユーザー調査において，システムの切り取り結果は必ずしもスマートフォンで撮影された画像と一致するわけではないが，ほとんどの場合画像の質は許容の範囲内であることが明らかになった．

今後の課題を以下に述べる．Harada et al. [1] と Adams et al. [9] はユーザーに周囲の音を録音するように促している．今回我々は写真を撮影した環境と被写体をしっかりと思い出せるよう，ユーザーに何かを話すようにお願いした．将来的には周囲の音を記録することも視野に入れている．

それに加えて，提案システムは現在画像のぼかしや明暗を考慮していない．将来的には，これらに関する機能も追加する予定である．

参考文献

[1] Harada, S., Sato, D., Adams, D. W., Kurniawan, S., Takagi, H. and Asakawa, C.: Accessible photo album: enhancing the photo sharing experience for people with visual impairment, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, ACM Press, (online), DOI: 10.1145/2470654.2481292 (2013).

[2] Jayant, C., Ji, H., White, S. and Bigham, J. P.: Supporting blind photography, *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '11*, ACM Press, (online), DOI: 10.1145/2049536.2049573 (2011).

[3] Voykinska, V., Azenkot, S., Wu, S. and Leshed, G.: How Blind People Interact with Visual Content on Social Networking Services, *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, ACM Press, (online), DOI: 10.1145/2818048.2820013 (2016).

[4] Balata, J., Mikovec, Z. and Neoproud, L.: BlindCamera: Central and Golden-ratio Composition for Blind Photographers, *Proceedings of the Mulitimedia, Interaction, Design and Innnovation on ZZZ - MIDI '15*, ACM Press, (online), DOI: 10.1145/2814464.2814472 (2015).

[5] White, S., Ji, H. and Bigham, J. P.: EasySnap: real-time audio feedback for blind photography, *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*, ACM Press, (online), DOI: 10.1145/1866218.1866244 (2010).

[6] Vázquez, M. and Steinfeld, A.: Helping visually impaired users properly aim a camera, *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*, ACM Press, (online), DOI: 10.1145/2384916.2384934 (2012).

[7] Vázquez, M. and Steinfeld, A.: An Assisted Photography Framework to Help Visually Impaired Users Properly Aim a Camera, *ACM Transactions on Computer-Human Interaction*, Vol. 21, No. 5, pp. 1–29 (online), DOI: 10.1145/2651380 (2014).

[8] Wikipedia: Post-production — Wikipedia, , available from ⟨https://en.wikipedia.org/wiki/Post-production⟩ (accessed Sept. 19, 2019).

[9] Adams, D., Kurniawan, S., Herrera, C., Kang, V. and Friedman, N.: Blind Photographers and VizSnap, *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '16*, ACM Press, (online), DOI: 10.1145/2982142.2982169 (2016).

[10] Frohlich, D. and Tallyn, E.: Audiophotography: practice and prospects, *CHI '99 extended abstracts on Human factors in computing systems - CHI '99*, ACM Press, (online), DOI: 10.1145/632716.632897 (1999).

[11] Frohlich, D. M.: *Audiophotography*, Springer Netherlands (2004).

[12] Adams, D., Morales, L. and Kurniawan, S.: A qualitative study to support a blind photography mobile application, *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '13*, ACM Press, (online), DOI: 10.1145/2504335.2504360 (2013).

[13] Bigham, J. P., White, S., Yeh, T., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A. and White, B.: VizWiz: nearly real-time answers to visual questions, *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*, ACM Press, (online), DOI: 10.1145/1866029.1866080 (2010).

[14] be my eyes: Be My Eyes - Bringing sight to blind and low-vision people, , available from ⟨https://www.bemyeyes.com/⟩ (accessed Sept. 20, 2019).

[15] TapTapSee: TapTapSee - Blind and Visually Impaired Assistive Technology - powered by CloudSight.ai Image Recognition API, , available from ⟨https://taptapseeapp.com/⟩ (accessed Sept. 20, 2019).

[16] Gurari, D., Li, Q., Stangl, A. J., Guo, A., Lin, C., Grauman, K., Luo, J. and Bigham, J. P.: VizWiz Grand Challenge: Answering Visual Questions from Blind People, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (online), DOI: 10.1109/cvpr.2018.00380 (2018).

[17] OrCam: OrCam MyEye 2, , available from ⟨https://www.orcam.com/en/myeye2/⟩ (accessed Sept. 18, 2019).

[18] Meijer, P.: The vOICe - New Frontiers in Sensory Substitution, , available from ⟨https://www.seeingwithsound.com/⟩ (accessed Sept. 18, 2019).

[19] Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M. and Hanrahan, P.: Light Field Photography with a Hand-Held Plenoptic Camera, Technical report, Stanford University Computer Science (2005).

[20] Wikipedia: Lytro — Wikipedia, , available from ⟨https://en.wikipedia.org/wiki/Lytro⟩ (accessed Sept. 20, 2019).

[21] Company, R.: RICOH THETA Plug-in Store, , available from ⟨https://pluginstore.theta360.com/⟩ (accessed Sept. 20, 2019).

[22] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (online), DOI: 10.1109/cvpr.2016.91 (2016).

[23] Trieu: GitHub - thtrieu/darkflow: Translate darknet to tensorflow., , available from ⟨https://github.com/thtrieu/darkflow⟩ (accessed Sept. 20, 2019).

[24] Chen, Y.-L., Klopp, J., Sun, M., Chien, S.-Y. and Ma, K.-L.: Learning to Compose with Professional Photographs on the Web, *Proceedings of the 2017 ACM on Multimedia Conference - MM '17*, ACM Press, (online), DOI: 10.1145/3123266.3123274 (2017).

[25] Chen, Y.: View Finding Network, , available from ⟨https://modeldepot.io/yilingchen/view-finding-network⟩ (accessed Sept. 18, 2019).