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—

井上 慶彦[1,a)]　岩村 雅一[1,b)]　黄瀬 浩一[1,c)]

概要：視覚障害者は，晴眼者（視覚に障害を持たない人）と比べて視覚から得られる情報が限られている．それを補うために，スマートフォンのカメラを用いて物体情報を取得できるような補助アプリもあるが，そもそも知りたい物体の位置を把握すること自体が困難である場合が少なくない．全方位カメラは一度に全方向の情報を取得できることから，対象物にカメラを向ける必要がなく，視覚障害者支援に適していると考えられる．全方位カメラにより周辺画像を取得し，得られた画像に対して物体検出及び認識を行い，音声情報などで視覚障害者にどこに何があるのかを伝えることができるようなシステムが実現できれば，汎用性の高い視覚障害者支援システムとなる．このようなシステムを実現するためには，まず全方位画像に対する正確な物体検出及び認識と，動画において前後フレームで物体情報を保持するためのトラッキングを行う必要がある．全方位カメラから得られる画像は，通常のカメラから得られる画像とは異なった投影方法なので，通常の画像を対象とした物体検出アルゴリズムやトラッキング手法をそのまま適用することは難しい．本稿では物体検出とトラッキングを別々の投影方法で行い，結果を統合することで，全方位画像に対して高精度な物体検出とトラッキングを実現する手法を提案する．

キーワード：全方位カメラ，物体検出，物体認識，トラッキング，視覚障害者支援

1. はじめに

視覚障害者は，晴眼者と比べて視覚から得られる情報が限られている．視覚情報を補うために，物体情報を取得し音声で伝えるツールとして，Seeing AI [*1]や，TapTapSee [*2]などのスマートフォンのカメラを用いた物体認識アプリも存在し，そのようなアプリの精度を上げるための研究 [1] も行われているが，対象物体をカメラに写る範囲に収めなければならないという問題点がある．視覚障害者は，必要とする文字や物体の位置の把握が困難である場合が少なくないので，この問題点を解決する必要がある．この問題は，広範囲を一度に撮影できる全方位カメラを用いることで解決できると考えられる．一度に全方位の状況を取得可能な全方位カメラは，撮影したい物体がどこにあるのかを事前に把握しておく必要がないので，視覚障害者支援に有用であると考えられる．

目標とする視覚障害者支援システムは，全方位カメラを用いて周辺の物体を検出，認識し，その情報を視覚障害者にリアルタイムで伝達することで，視覚障害者が外出時などにその場所の周辺状況を把握することが可能となるシステムである．これを実現するには，まず全方位カメラから得られた画像に対して，正確な物体認識を行う必要がある．ただし，一般に行われるように，動画像の毎フレームに対して静止画に対する物体検出手法を用いる場合だと，毎フレームで独立に物体検出が行われてしまい，同じ物体であるのに毎フレームで新規物体として認識されてしまう問題がある．そこで，画像に対しての物体検出にトラッキングを併用することによって，前後フレームに関連性を持たせ，検出結果を保存することを考える．この処理によって，同一物体が毎フレームで新規物体と認識されてしまう問題点が解決される．したがって，目標のシステムを実現するためには，全方位カメラから得られた画像に対して正確な物体検出，認識及びトラッキングを組み合わせる必要がある．

全方位カメラでの物体検出及びトラッキングの手法としては，Markovic らの手法 [2] がある．これは動いている物体を検出し，トラッキングを行う手法だが，その物体が何なのかという物体認識は行われないため，目標のシステムには用いることができない．物体認識とトラッキングを組み合わせた手法としては，Detect to Track and Track to

[1] 大阪府立大学
Osaka Prefecture University
[a)] yoshihiko@m.cs.osakafu-u.ac.jp
[b)] masa@cs.osakafu-u.ac.jp
[c)] kise@cs.osakafu-u.ac.jp
[*1] https://www.microsoft.com/en-us/seeing-ai/
[*2] http://otoiro.net/?p=680

表 1　提案手法で用いる図法のまとめ

	物体検出	トラッキング
正距円筒図法	X	◯
キューブマップ	△(境目X)	X
8面キューブマップ	◯	X
提案手法	◯	◯

Detect [3] がある．しかしこの手法は，全方位カメラから得られる画像は歪むため，歪みが大きい部分では本来の認識性能を期待できない．Detect to Track and Track to Detect に限らず，通常の単眼カメラを用いて行う物体検出手法やトラッキング手法は，そのまま全方位カメラから得られた画像に適用することが難しい．これは全方位カメラから得られる画像の投影方法が，正距円筒図法と呼ばれる特殊な投影方法であることが原因である．正距円筒図法は 360 度全方位を一枚の画像に無理矢理収めているため，画像に歪みが生じてしまう．一般的に用いられている物体認識手法は，その歪みに対処できないため，検出や認識をうまく行うことができない．

本稿では，この問題を解決するために全方位画像の投影方法を変換し，物体検出やトラッキングを扱う手法を提案する．提案手法で用いる投影方法をまとめると，表 1 のようになる．物体検出にはキューブマップを用いる．これは空間を立方体の各面に投影する図法で，各面については通常のカメラで撮影した画像とほとんど変わらない画像を得ることができる．しかしキューブマップは空間を別々の 6 枚の画像に分割してしまうため，各面の境目に跨った物体をうまく認識できないという問題点がある．そこで，キューブマップの水平方向の面を増やした 8 面キューブマップを用いて物体認識を行う．また，キューブマップのみで物体認識とトラッキングの両方の処理をすると，物体が違う面に移動した際にトラッキングが途絶えてしまうことがあるので，トラッキングは 8 面キューブマップで物体認識を行った結果をもとに，元の正距円筒図法の画像上で行うことにする．

以上をまとめると，本稿では，全方位カメラを用いた物体検出，認識を行う際の投影方法の問題を解決した上で，時系列情報を保持するために，物体検出とトラッキングを組み合わせることを考える．高精度な検出とトラッキングを行うために，物体検出とトラッキングをそれぞれ別の投影法で行い，統合する手法を提案する．

2. 関連研究

2.1 Markovic らの手法

全方位カメラに対する物体検出およびトラッキングの手法の Markovic らの手法 [2] について述べる．Markovic らの手法は，移動するロボットに備え付けられた全方位カメラからの入力動画に対して，動画内で動いている物体を検出し，単位球上でトラッキングする手法となっている．物体検出は，画像内のオプティカルフローを計算したのち，終点ベクトルの距離を解析的に計算し，単位球上で動的か静的か区別されたフローベクトルを用いることで実行される．トラッキングは，単位球上のベイズ推定問題として提起され，フォンミーゼスフィッシャー分布に基づく解が利用される．移動するロボットに全方位カメラを搭載することは，周囲のシーンに関する全ての情報が一枚の画像フレームに格納されるため有用である．しかし，Markovic らの手法は，動いた物体のみを検出し追跡する手法のため，その物体が何であるかという認識は行われていない．現在研究している視覚障害者支援システムでは，周辺に何があるかを知るために，まず物体認識を行うことが重要であるため，Markovic らの手法を用いることはできない．

2.2 Detect to Track and Track to Detect

物体認識とトラッキングを組み合わせた手法 Detect to Track and Track to Detect [3] について述べる．精度向上のために年々複雑になっていきている検出とトラッキングの手法を，検出とトラッキングを共同で行う畳み込みニューラルネットワークのアーキテクチャによって単純化と精度向上を達成した手法である．フレームベースで検出とトラッキングを同時に行えるように畳み込みニューラルネットワークのアーキテクチャを設定し，トラッキング中の畳み込みニューラルネットワークを支援するために，時間経過とともにオブジェクトの共起を表す相関特徴量の導入した．また，フレームレベルの検出をフレーム間の追跡に基づいて行うことで動画における高精度を達成した．この手法は通常の単眼カメラから得られた動画像が対象となっているので，そのまま全方位カメラから得られた動画像に適用することはできない．

3. 提案手法

本稿では，全方位カメラから得られた動画像に対して，高精度な物体検出及びトラッキングを実現する方法を提案する．提案手法は図 1 のように，物体検出とトラッキングをそれぞれ別の投影方法の画像を用いて行うことで，高精度な物体検出とトラッキングを実現する手法である．まず入力画像を後述の 8 面キューブマップに変換し，その画像に対して物体検出，認識を行い，その結果を統合したのち，元の入力画像上でトラッキングを行うという手法になっている．以下，3.1 項で物体検出における投影方法，3.2 項で物体領域の決定方法，3.3 項でトラッキングについて説明する．

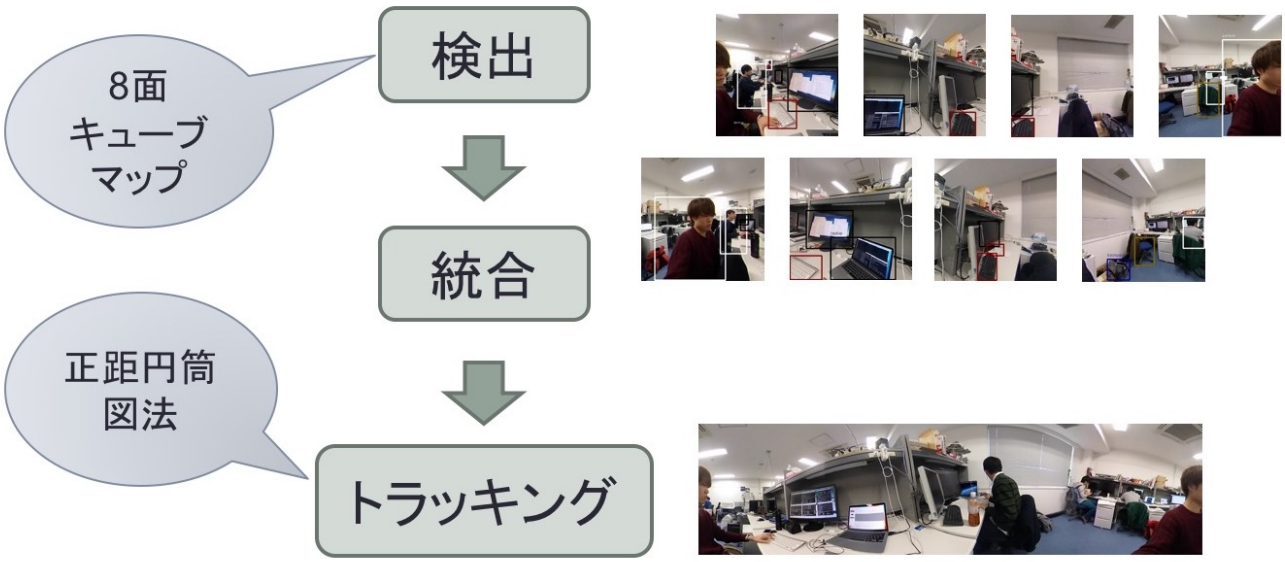

図 1 提案手法の流れ

3.1 投影方法

多くの全方位カメラは図 2 のような正距円筒図法と呼ばれる投影方法で全方位画像を取得するが，これは 360 度全方位を一枚の画像に収めるため，画像に歪みが生じてしまう．物体検出を行う際，一般的な物体検出アルゴリズムはこの歪みを考慮していない．そこで，正確な物体検出を行うために，画像の投影方法を変換することを考える．今回物体検出に用いた投影方法はキューブマップと呼ばれるものである．キューブマップは図 3 のように，撮影点を中心として周りの空間を立方体の 6 面それぞれに投影する方法であり，それぞれの面については，通常の単眼カメラで撮影した場合とほぼ同じような，歪みの少ない画像を得ることができる．正距円筒図法からキューブマップへの変換は，「正距円筒画像を立方体表面にマッピングする際の画像変形について」 [*3]を参考にして実装を行った．なお，上下の面は得られる情報が少ないと考え，本提案手法では無視することとする．

しかし，キューブマップは，正距円筒図法と違い，画像が分割されてしまうという問題点がある．画像が分割されてしまうことにより，面と面の境目の物体がうまく検出されない場合がある．そこで，図 4 のように，投影する角度を 45 度ずらしたキューブマップ画像を別に用意し，水平方向の計 8 枚の画像を用いる．便宜上この 8 枚の画像を以後 8 面キューブマップと呼ぶこととする．8 面キューブマップは，通常のキューブマップで画像の境目になってしまっている部分を中心とした 4 枚の画像を用意するため，隣り合った画像間には図 5 のようにオーバーラップが生じるが，面と面の境目で物体検出がうまくいかない問題点は解決される．

[*3] http://fmskatsuhiko.web.fc2.com/spherecube.html

3.2 物体領域の決定

8 面キューブマップは，隣り合うそれぞれの画像にオーバーラップが生じているため，8 面キューブマップの各面で得られた物体検出結果には重複が生じている場合がある．そこで，物体検出の結果を Non-Maximum Suppression を用いて統合する．Non-Maximum Suppression とは，同じクラスとして分類されたバウンディングボックスの重なりを無くすためのアルゴリズムである．重なり合ったバウンディングボックスについて，Intersection over Union(IoU) の値を基準に統合する手法である．IoU とは，画像の重なりの割合を表す値で，この値が大きいほど画像が大幅に重なっているということである．この IoU 値に閾値を設定し，その閾値以上の IoU 値となったバウンディングボックスの組は一つに統合する．今回は，8 面キューブマップを用いて得られた物体検出の結果を全て，一度元の正距円筒図法の画像上の座標に変換し，そこで Non-Maximum Suppression を用いることにより，同一物体について複数のバウンディングボックスが検出されることを防いでいる．

3.3 トラッキング

8 面キューブマップは，物体検出には適しているがトラッキングを行うには不適当である．これは全方位画像が分割されてしまうことが原因である．全方位を一枚の画像内に収める正距円筒図法とは違い，キューブマップは全方位を水平方向に 4 枚，8 面キューブマップでは水平方向に 8 枚に分割することになる．分割された画像はそれぞれ別の画像として処理されるため，隣り合う画像間を移動する物体などに対してトラッキングを行いにくい．例えば，ある時間に正面にいた物体が，その後左右どちらかの面に移動した際，それを同一物体として認識することができない．そ

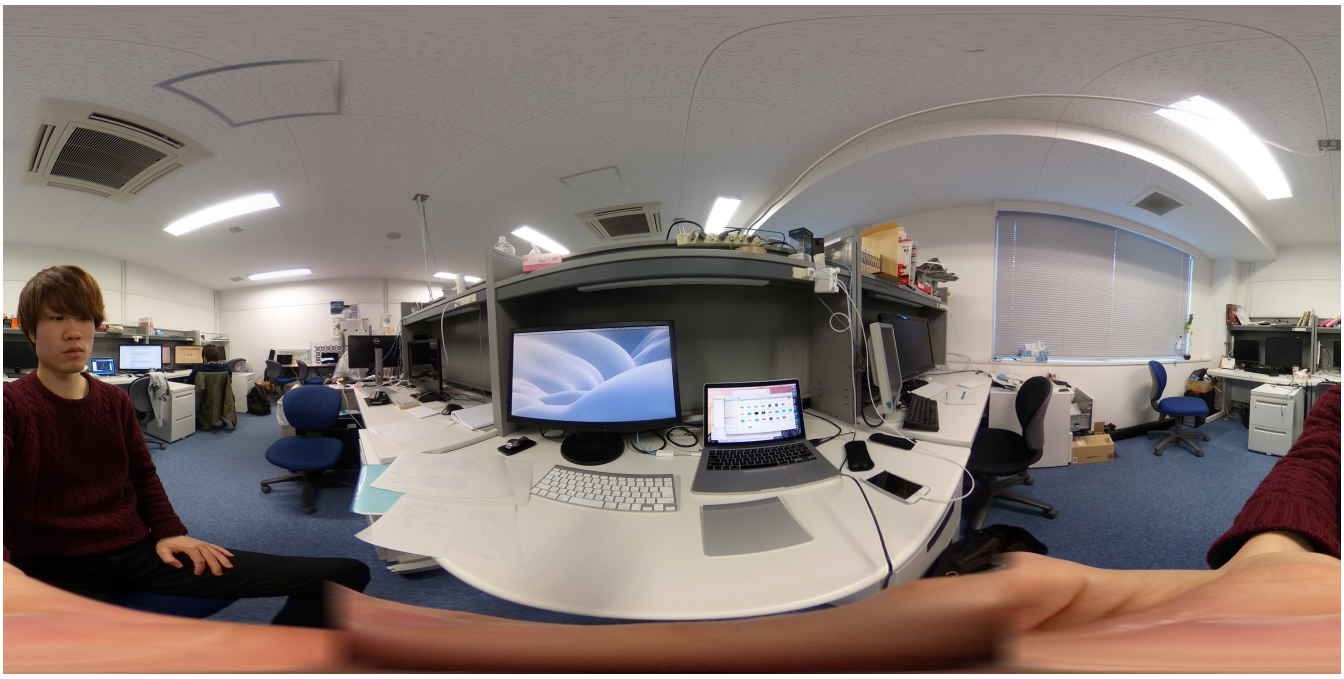

図 2 正距円筒図法

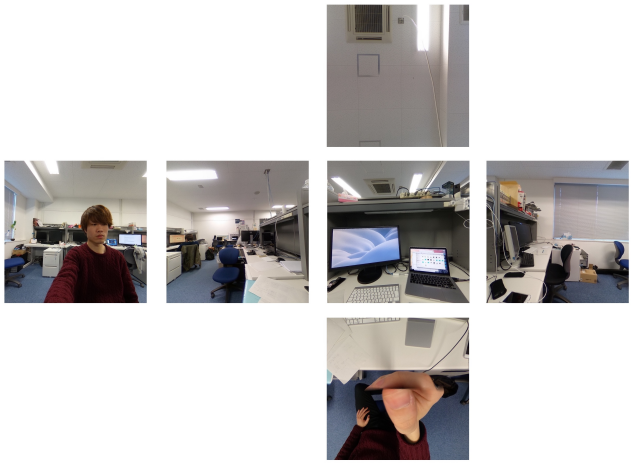

図 3 キューブマップ

図 4 通常のキューブマップ (左) と 45 度ずらしたキューブマップ (右) のイメージ図

こで，トラッキングは正距円筒図法の画像上で行う．正距円筒図法の画像は物体検出には適していないが，全方位の状況を一枚の画像で表示し，常に物体を画像内に表示させ

図 5 8 面キューブマップのイメージ図

ることができるため，トラッキングを行うには適した投影方法であると言える．

4. 実験

今回，提案手法を用いて実験を行ったのでその条件と結果について述べる．

4.1 実験条件

全方位カメラには RICOH Theta V [*4] を用いた．RICOH

[*4] https://theta360.com/ja/about/theta/v.html

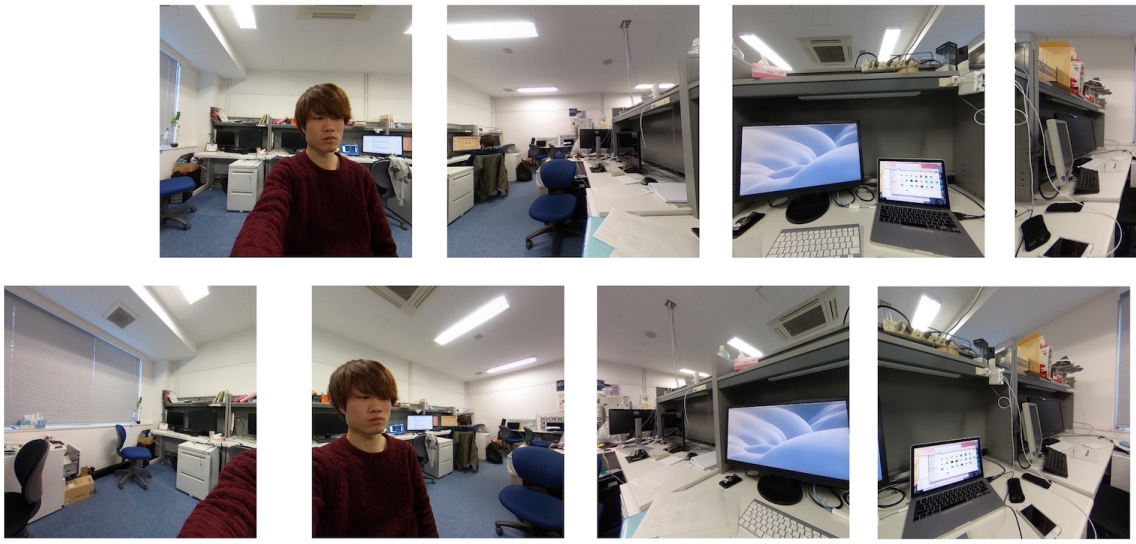

図 6 8 面キューブマップ

Theta V は民生用全方位カメラの代表的な機種の一つであり，前後に搭載された 2 つの魚眼レンズから全方位画像を得ることができる．取得画像の投影方法は正距円筒図法で，解像度は 3840×1920，もしくは 1920×960 となっている．今回の実験では，実行速度向上のため，解像度は 1920×960 としている．RICOH Theta V には静止画撮影モード，動画撮影モード，ライブストリーミングモードがあるが，今回はリアルタイムに情報を取得するために，ライブストリーミングモードにより全方位画像を取得する．

物体検出，認識の手法には YOLOv2 [4] を用いた．これは Convolutional Neural Network を用いた物体検出アルゴリズムであり，入力画像をバウンディングボックスに分割してクラス分類を行うことで高速高精度な検出を可能にしている．また，COCO detection dataset と ImageNet classification dataset を統合して学習させることにより，9000 以上の物体カテゴリを検出することができる．

トラッキングの手法には Simple Online and Realtime Tracking(SORT) [5] を用いた．この手法は，フレーム間の予測と関連付けに焦点を当てた高速高精度なトラッキングフレームワークである．今回の実験では実装上の都合から，YOLOv2 によって “person” というラベルがついたバウンディングボックスのみをトラッキングの対象としている．

4.2 結果と考察

提案手法による出力結果中の 2 フレームを図 7 に示す．比較対象として，キューブマップ方式に変換せず，正距円筒図法の画像のみで物体検出とトラッキングを行なった際の出力結果の 1 フレームを図 8 に示す．図 8 では検出がうまく行われず，誤ったバウンディングボックスに対してトラッキングが行われてしまっている．提案手法による実験結果は，図 7 にあるように，周辺にいる 3 人の人を検出認識し，トラッキングすることに成功している．この結果より，提案手法は全方位カメラを用いた物体検出とトラッキングを行う際に有効だと言える．

現状の問題点としては，処理速度が遅いことがあげられる．物体検出に 8 面キューブマップを用いているため，処理する画像の枚数が多く，結果の表示スピードが約 1.5FPS となっている．これはトラッキングに比べて物体検出の速度が遅いことが原因として考えられるので，トラッキングと物体検出を別スレッドで並列処理することで解決できると考えている．

5. まとめと今後の展望

本稿では，視覚障害者支援システムの実現に向けた，全方位カメラを用いた物体検出とトラッキングの手法を提案した．全方位カメラから得られる正距円筒図法の画像を 8 面キューブマップに変換してから物体検出を行い，結果を統合した後，正距円筒図法の画像上でトラッキングを行った．物体検出とトラッキングを異なる投影方法の画像で行うことにより，正距円筒図法のみで検出とトラッキングを行なった場合よりも，高精度な検出とトラッキングが可能となった．

今後の展望としては，現状の問題点の解決，トラッキング精度の向上，トラッキングするラベル数の拡張，そして目標とする視覚障害者支援システムへの組み込みを考えている．

謝辞 本研究は，JSPS 科研費 17H01803 の補助による．

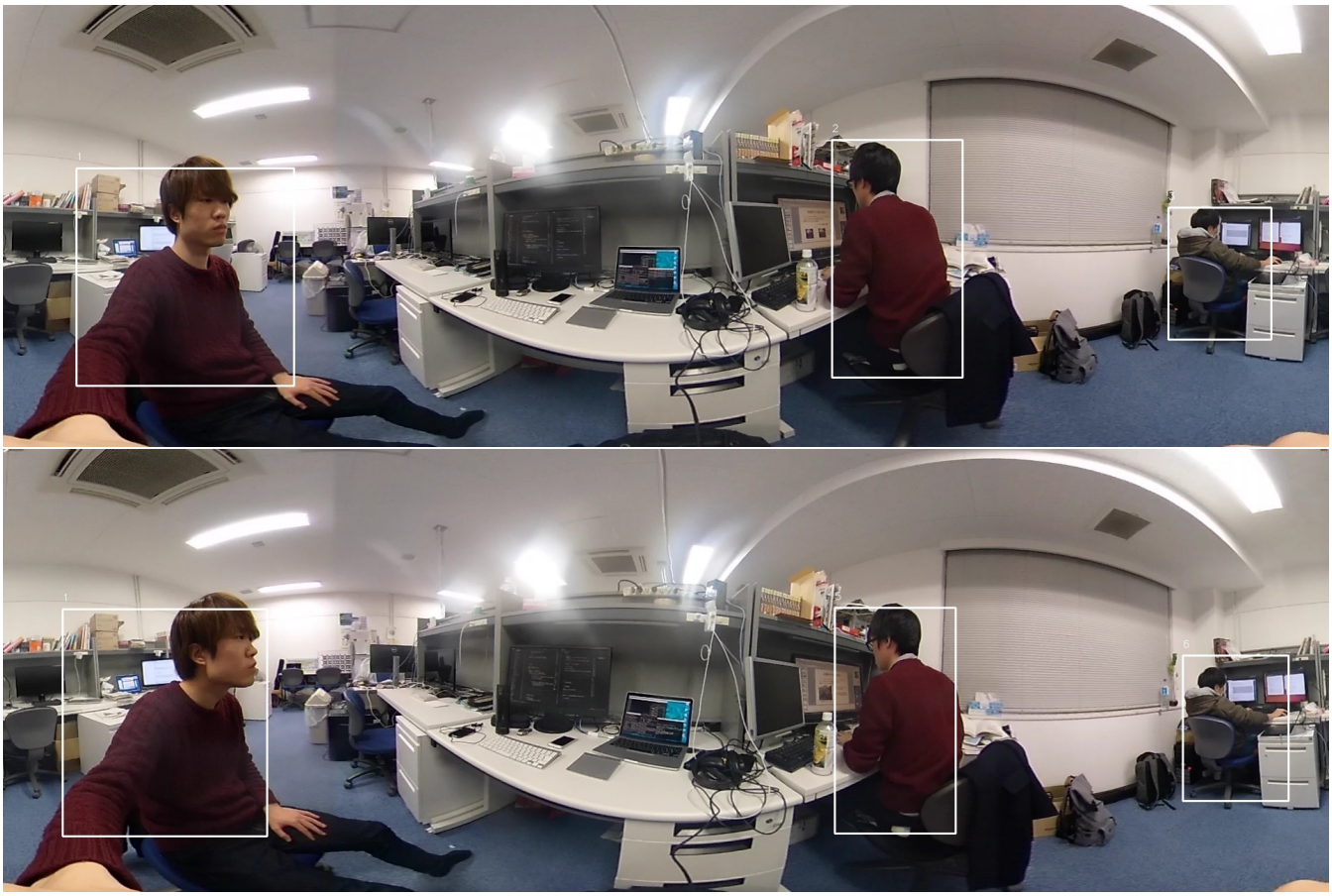

図 7　提案手法による実験の出力結果

図 8　正距円筒図法のみを用いた場合の出力結果

参考文献

[1] Kacorri, H., Kitani, K. M., Bigham, J. P. and Asakawa, C.: People with Visual Impairment Training Personal Object Recognizers: Feasibility and Challenges, *Proc. of CHI*, pp. 5839–5849 (2017).
[2] Ivan Markovic, F. C. and Petrovic, I.: Moving object detection, tracking and following using an omnidirectional camera on a mobile robot, *Proc. of ICRA* (2014).
[3] Feichtenhofer, C., Pinz, A. and Zisserman, A.: Detect to Track and Track to Detect, *Proc. of ICCV* (2017).
[4] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, *Proc. of CVPR*, pp. 6517–6525 (2017).
[5] Bewley, A., Ge, Z., Ott, L., Ramos, F. T. and Upcroft, B.: Simple online and realtime tracking, *Proc. of ICIP*, pp. 3464–3468 (2016).