

広域観測カメラによる視覚障がい者検出

中泉 安貴[†] 宍戸 英彦[‡] 北原 格[‡] 亀田 能成[‡]

[†]筑波大学 〒305-8574 茨城県つくば市天王台 1-1-1

E-mail: [†] nakaizumi.yasutaka@image.iit.tsukuba.ac.jp, [†] shishido@ccs.tsukuba.ac.jp,

[‡] {kitahara, kameda}@iit.tsukuba.ac.jp

あらまし 視覚障がい者にとって外出は大きな危険を伴う行為である。本研究では、視覚障がい者の外出支援として、社会全体が視覚障がい者の外出を見守る社会の実現を最終的な目標とする。この目標の実現のため、広域を観測できる高解像度のカメラを用いて、白杖を使用している視覚障がい者の検出を行う。まず、高解像度で撮影した映像に対し、映像の分割処理を施す。白杖を持った視覚障がい者を検出するために、ディープラーニングの物体検出アルゴリズムの一つである YOLOv3 を用いて、白杖を持っている視覚障がい者、歩行者（健常者）、及び閉じた傘を持っている歩行者をクラス分けして学習をさせる。最後に、YOLOv3 の検出結果を元に、白杖を使用している視覚障がい者の広域内での位置を特定する。

キーワード 視覚障がい者、白杖使用者、YOLOv3、高解像度映像、3次元位置特定

1. はじめに

視覚障がい者にとって、外出をし、道路を歩くことは多くの危険を伴う行為である。厚生労働省調査によると、日本における視覚障がい者数は 31.2 万人である[1]。日本盲人連合会の調査[2]によると、週に 2,3 日以上外出する人は 89.3%である。道路上で危険を感じたことがある人の割合は 87.8%である。これらのことから、視覚障がい者の外出及び目的地までの移動に対して、何らかの社会的な支援を準備しておくことが望ましい。この仕組みは、視覚障がい者への負担なしで実現されるべきである。

本研究では、視覚障がい者の外出を支援するため、視覚障がい者の外出を社会全体で見守る社会の実現を最終的な目標とする。公共の場において、視覚障がい者を見つけておくことが重要である。

この目標の実現のため、広域を観測できる高解像度のカメラを用いて、白杖を使用している視覚障がい者の検出を行う。広域を観測できる高解像度カメラを用いて、白杖を使用している視覚障がい者の映像を撮影する。高解像度の映像を用いた視覚障がい者の検出を行うために高解像度で撮影した映像に対し、映像の細分化を行う。細分化した映像に対し、画像認識アルゴリズムの一つである YOLOv3 を用いて、路上にいる白杖を使用している視覚障がい者の検出を試みる。YOLOv3 には、白杖を使用している視覚障がい者と歩行者（健常者）、さらに白杖を使用している視覚障がい

者と誤検出する可能性を考え、閉じた傘を持った歩行者を学習させる。また、YOLOv3 で検出した情報を用いて、視覚障がい者の現実での位置を、視覚障がい者の足元の位置から特定する手法を提案する。

駅のホームや交差点、ショッピングモールなどの公共の場の多くは平面と捉えられることを考慮して、本稿では実験対象とする広域は平面と仮定する。

2. 関連研究

白杖を使用している視覚障がい者を観測カメラから検出する研究として、真田ら[4]はフレーム間差分画像とエッジ検出を用いた白杖の検出を試みている。この研究では、白杖と背景のコントラストが大きい場合に有効であるが、コントラストが小さい時に白杖のエッジ検出に失敗するという結果となっている。

YOLOv3[3]は、犬の種別などの、形状は似ているが細部が異なる物体の識別も可能である。そのため、白杖を使用している視覚障がい者と歩行者（健常者）の識別も可能であると考えられる。検出精度に対して重要となるのが、画像中に占める物体の大きさである。検出精度が担保される物体の大きさを調査するため、YOLOv3 の論文[3]で使用されている、PASCAL VOC のデータセット[6]を解析する。PASCAL VOC のデータセット中の、人物が写っている画像について解析したものを図 1 に示す。人物が写っている画像 4145 枚に対し、画像中に占める縦方向の人物の割合を、各人に付けられた外接矩形を元に調べる。これによると、PASCAL VOC のデータセットでは、画像中に占める縦

方向の人物の大きさが 10% 以下のデータに関しては枚数が少ないことがわかる、10%～100%の画像は多く存在する。

このことは、YOLOv3 が小さく写っている物体の認識率が悪くなることを直接的に意味しないが、少なくとも小さく写る物体を想定していないと言えよう。

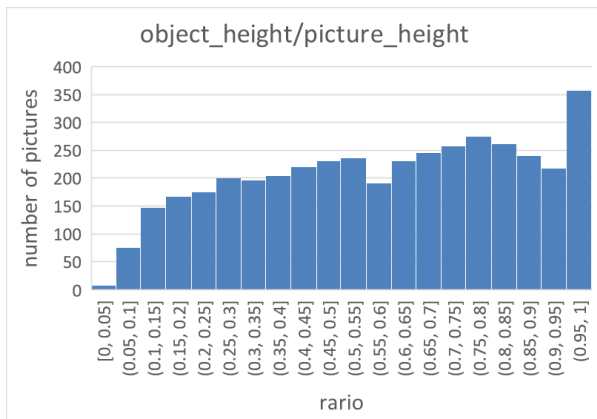


図 1 PASCAL VOC データセット解析結果

高解像度画像に対する YOLOv3 の利用方法として、C. Liu ら[5]は、ACF を用いて、潜在領域を抽出し、元画像から一部を切り抜いた画像に対して、YOLOv3 で検出を行う手法を提案している。画像の一部を切り抜く工程を入れることで、検出速度が遅くなった代わりに、YOLOv3 に高解像度画像を直接入力するより検出精度が上がったとしている。

3. 提案手法

3.1. 高解像度映像分割, 検出

本研究では、横 3840 ピクセル、縦 2160 ピクセルの高解像度な映像を使用する。その中で歩行者が占める縦方向のピクセル数は、約 50 ピクセル～280 ピクセルである。これは映像縦方向中の 2.3%～13%を占める。このような小さい物体の検出は、映像を元のサイズのまま YOLOv3 へ入力したのでは困難である。YOLOv3 で検出パフォーマンスを確保するため、映像を分割して入力とする。処理の流れを図 2 に示す。

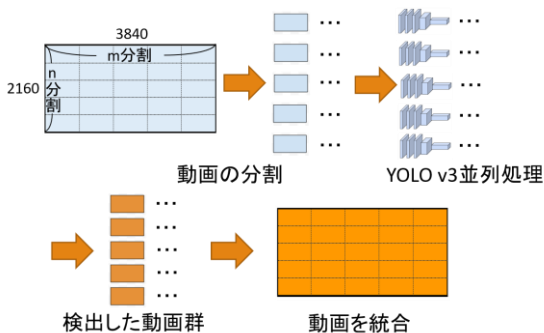


図 2 映像分割処理の流れ

映像を $m \times n$ に分割する。n は、歩行者の画像中に占める縦方向の大きさが、10%を超えるように設定する。また、YOLO3 の入力層は 416×416 ピクセルであるため、分割映像の大きさを、このサイズに近づける。

映像分割後、分割した映像を YOLOv3 の検出器にかける。映像の各フレームで検出された座標を、元映像の座標に当てはめる。

3.2. 検出クラス

白杖を使用している視覚障がい者、歩行者（健常者）、閉じた傘を持っている歩行者の 3 クラスを検出させる。本来白杖を使用している視覚障がい者と歩行者（健常者）の 2 クラスを検出出来れば十分であると考えらえるが、閉じた傘を持っている歩行者は白杖を使用している視覚障がい者と誤って検出する可能性がある。そのため、検出するクラスに追加した。

視覚障がい者の足元の位置を、YOLOv3 で検出された外接矩形の、底辺の中央の画像座標（図 3）とする。



図 3 足元の位置

3.3. 二次元座標から現実の位置特定

実際の支援を視野に入れた時、視覚障がい者の現実の位置を特定することは必須である。本研究では、実験環境を平面と仮定して、白杖を使用している視覚障がい者の位置特定にホモグラフィ変換を用いる。

YOLOv3 で検出された外接矩形の、白杖を使用している視覚障がい者の足元の位置（図 3）を、ホモグラフィ変換で広域での平面座標に変換する。

4. 実験

4.1. YOLOv3 での学習

白杖を使用している視覚障がい者を撮影した公開データセットがないため、我々は独自にデータセットを構築した。

学習データの撮影は、晴天の日の午後に撮影を行った。白杖を使用している視覚障がい者を、様々な角度から、白杖が写るように撮影を行った。撮影枚数の総数は 3251 枚である。サンプル数は、白杖を使用している視覚障がい者用 2562 サンプル、歩行者（健常者）用

4891 サンプル，閉じた傘を持っている歩行者用 1414 サンプルである．学習データセットの一例を図 4 に示す．緑が白杖を使用している視覚障がい者，黄色が歩行者（健常者），赤が閉じた傘を持っている歩行者である．

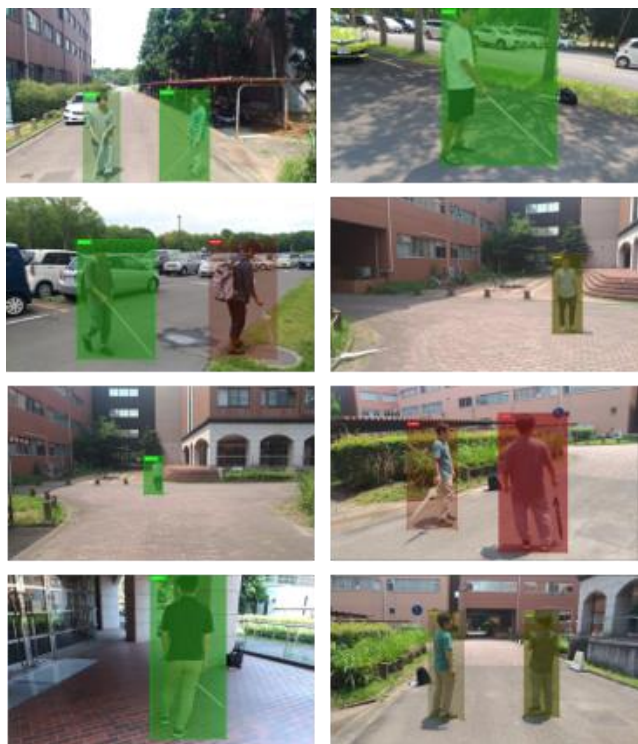


図 4 学習データセットの一例

学習最初の重みファイルは，学習を初期状態から始めることが可能である darknet53.conv.74 を使用した．iteration と loss の値，mAP（mean Average Precision）の推移を図 5 に示す．learning_rate は 0.01，iteration は 23500 である．その中で mAP が一番高い，iteration が 18000 の重みファイルをモデルとして採用した．

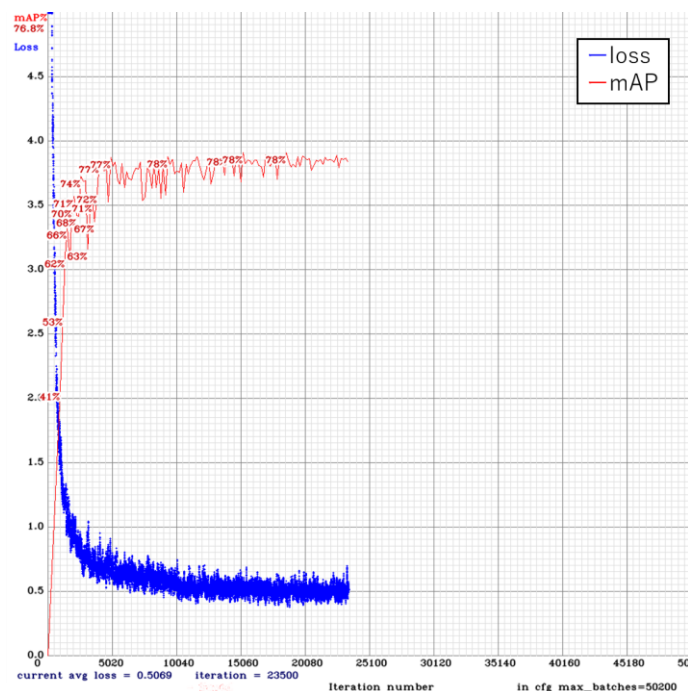


図 5 学習結果

4.2. 広域での検出実験

筑波大学構内の広場を，対象とする広域として撮影し，視覚障がい者の検出実験を行う．この広場は，広い平面を確保することができる(図 6[7])という特徴があり，研究条件に合致している．

映像の撮影は，ASUS 製のスマートフォン Zenfone3 を用いて行った．カメラの焦点距離（画角）は，35mm 判換算で 23mm である．映像の解像度は 3840×2160 ピクセルである．学習と条件をそろえるため，晴天の日を選んで撮影を行った．

検出の際には，映像を 5×5 に分割する．それぞれの映像は 768×432 ピクセルとなり，映像中の人物の占める縦方向の大きさは約 11%～65%となる．2 節での検証から，画像中に占める歩行者の縦方向の割合が 10%を超えていれば，YOLOv3 で検出することが可能であると考えられる．よって，11%～65%は，YOLOv3 で検出できる大きさを満たしていると考えられる．これ以上分割数を上げると，映像の境界が多くなり継続的な検出が困難となると考え，5×5 分割とする．

実際に撮影した映像の 1 フレーム切り出したものを図 7 に示す．このフレームについて，検出を行った結果を図 8 に示す．画像左中央付近に白杖を使用している視覚障がい者が写っている．画像全体で，白杖を使用している視覚障がい者と，歩行者（健常者）が正しく認識されている．なお，検出結果は，青が“white Cane”（白杖を使用している視覚障がい者）であり，ピンクが“person”（歩行者（健常者）），黄緑が“umbrella”（閉じた傘を持っている歩行者）である．



図 6 広場の大きさ



図 7 入力フレーム

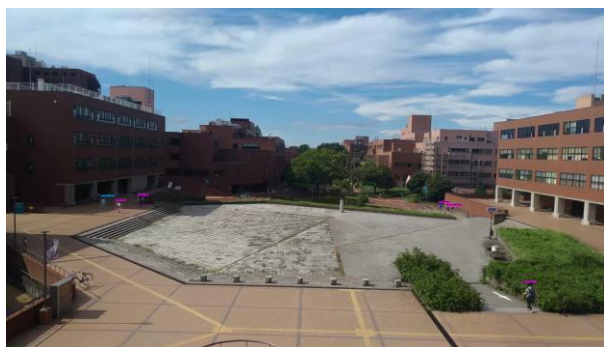


図 8 検出結果

図 8 中の白杖を使用している視覚障がい者について、部分拡大したものを図 9(a)に、図 7 のフレーム以外で

の、白杖を使用している視覚障がい者の検出結果について、部分拡大した画像を図 9(b),(c),(d)に示す。本手法により、白杖を使用している視覚障がい者の検出が可能であることが確認できている

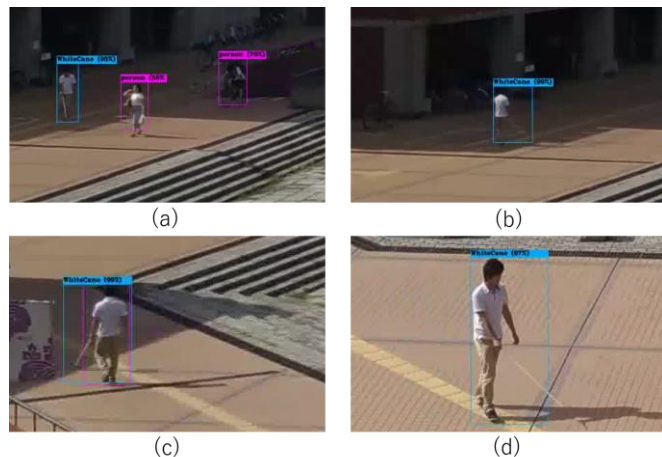


図 9 視覚障がい者の検出結果

正しく検出できなかった例を図 10 に示す。体と白杖が重なった場合(図 10(a))や、白杖が陰に入り込み、黒く映った場合(図 10(b)), 障害物に白杖が隠れてしまった場合(図 10(c))などに検出ができていない。また、白杖を傘と誤検出している例(図 10(d))もある。

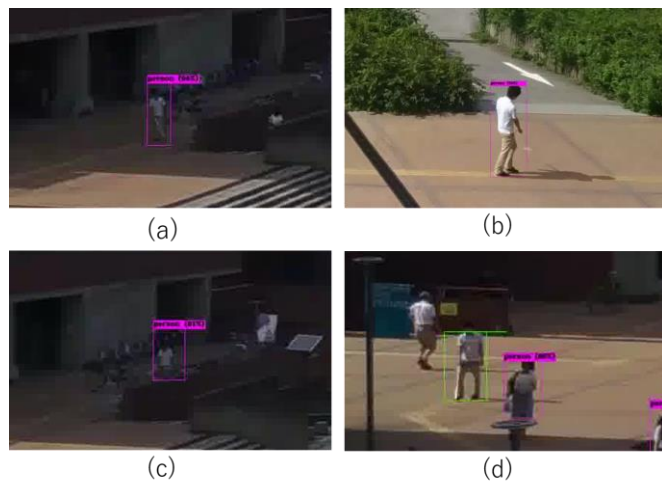


図 10 視覚障がい者の誤検出

検出が失敗する原因として、学習データ不足が考えられる。学習データの撮影環境は、主に晴天を選んでいたなどの理由から、白杖がはっきりと映り込んでいることが多い。白杖がはっきりと映っていない場合には検出が成功しにくい。また、誤検出の可能性を考えて傘を持った歩行者を学習させたが、結果的に白杖に

認識を阻害しているとも捉えられる結果となった。

分割数に関し、本稿では 25 分割を行ったが、映像横方向に関し、YOLOv3 の入力層の 1.5 倍ほどのサイズとなっている。この入力する映像のサイズが検出精度に影響を与えているのか、分割数を変更して検証する必要がある。

本研究では、分割した映像フレームを一つずつ順に一つの YOLOv3 の検出器へ入力したが、計算機を映像の分割数分用意して、分割した映像を並列に処理させることで、リアルタイムの検出は可能となると考えられる。

4.3. 現実座標推定実験

図 8 にホモグラフィ変換を施して、上空から俯瞰した画像に変換したものを図 11 に示す。

図 11 のように現実での座標軸をとると、白杖を使用している視覚障がい者の座標は、 x 軸方向に 30.13[m], y 軸方向に 43.38[m]の位置となった。

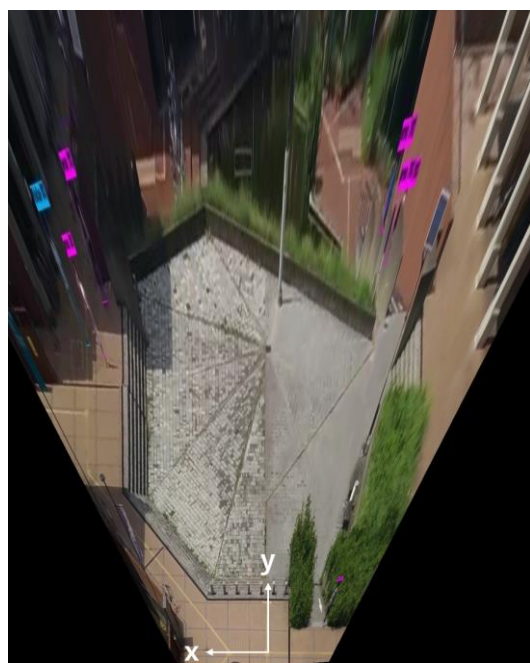


図 11 筑波大学構内の広場の俯瞰画像

5. おわりに

本研究では、白杖を持った視覚障がい者を、広域観測が可能な固定カメラから、画像認識アルゴリズムの一つである YOLOv3 を用いて検出し、その位置を求める手法を提案した。高解像度の映像から、白杖を使用している視覚障がい者を検出させるために、映像を 25 分割してそれぞれの映像フレームを検出器にかけ、分割した映像を再統合して出力とした。その結果、白杖を使用している視覚障がい者を検出することが可能と

なった。また、ホモグラフィ変換を用いて視覚障がい者の現実での位置特定が可能となった。

本研究の一部は科研費 17H01773 の助成を受けて行われた。

文 献

- [1] 厚生労働省社会・援護局障害保健福祉部，“平成 28 年生活のしづらさなどに関する調査（全国在宅障害児・者等実態調査）結果”，2018.
- [2] 福祉法人日本盲人会連合，“視覚障害者の外出時に安全を総合的に保証するシステムを確立するための研究事業”，2013.
- [3] Joseph Redmon, Ali Farhadi, “YOLOv3: An Incremental Improvement”
arXiv preprint arXiv:1804.02767, 2018.
- [4] 真田嘉寿佐，滝沢穂高，青柳まゆみ，“監視カメラ映像からの白杖使用者の検出に関する基礎的検討”，信学技報 WIT, vol. 118, no. 180, pp. 15-18, 2018.
- [5] C. Liu, Y. Guo, S. Li and F. Chang, “ACF Based Region Proposal Extraction for YOLOv3 Network Towards High-Performance Cyclist Detection in High Resolution Images”, Sensors, vol. 19, no. 12, 18 pages, 2019.
- [6] M. Everingham, S. M. A. Eslami, L. Van Gool, C.K.I. Williams, J. Winn and A. Zisserman, “The PASCAL Visual Object Classes Challenge: A Retrospective”, International Journal of Computer Vision, vol. 111, no. 1, pp. 98-136, 2015.