

事例ベースの姿勢正規化による白杖利用者認識に向けた検討

○西田尚樹 †, 川西康友 †, 出口大輔 †, 井手一郎 †, 村瀬 洋 †, 朴 君 ‡

○ Naoki NISHIDA † Yasutomo KAWANISHI †, Daisuke DEGUCHI †, Ichiro IDE †,
Hiroshi MURASE †, Jun PIAO ‡

†: 名古屋大学, nishidan@murase.is.i.nagoya-u.ac.jp

‡: 日本電気株式会社, j-piao@cw.jp.nec.com

<要約>本研究では、視覚障害者を支援するために、人物の姿勢変化から歩行者系列が白杖利用者か否か分類することを目的とする。しかし、人物の向きによって画像上の人物姿勢は大きく異なるため、分類しなければならぬ姿勢パターンが多くなり、分類が困難となる。そこで本発表では、事例ベースで様々な向きの姿勢をある1つの向きの姿勢に正規化する手法を提案する。分類する姿勢のパターンが限定されることにより、高精度の分類が可能となる。評価実験において、提案手法により認識率が5%向上することを確認した。

<キーワード>白杖利用者, 姿勢推定, 事例ベース姿勢正規化

1 はじめに

現在、我が国では高齢化に伴って身体障害者が増加しており、そのような人々を支援する様々な設備が設置されている。特に必要とされているのが、視覚障害者への支援設備であり、視覚障害者を誘導するための点字ブロックは街や公共施設の至るところで見ることができる。しかし、そのような状況でも視覚障害者が駅ホームから線路へ転落するなどの事故が後を絶たない。近年、このような事故を避けるため、駅にはホームドアなどの転落防止設備が新たに設置されつつあるが、ホームドアはまだ十分に普及しておらず、確実な事故防止には人手による支援が不可欠である。人手による支援のためには、支援対象者の所在を把握する必要があるが、そのために駅構内等に設置されている防犯カメラ映像を利用する方法が注目されている。例として、谷川らは防犯カメラ映像内の車椅子利用者を自動的に認識し、その所在を把握する手法 [1] を提案している。このような技術をもとに、映像内の視覚障害者の所在を自動的に把握し、駅員等へ通知することで、視覚障害者への早期支援をするサービスへの期待が高まっている。

視覚障害者への支援は、一般の歩行者には不要であることが多いため、視覚障害者のみへのサービス提供

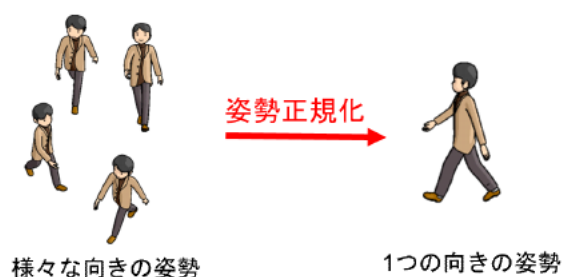


図 1 姿勢正規化

を実現するためには、一般の歩行者と視覚障害者を見分ける必要がある。視覚障害者は視覚の補助として、障害物を探索するため、また、自らが視覚障害者であることを周囲に知らせるために白杖を携帯しており、この白杖を手がかりとして視覚障害者を発見することが考えられる。そこで本研究では、映像内の白杖を保持する視覚障害者（白杖利用者）を認識することを目的とする。

白杖利用者の認識手法としては、我々が以前提案した手法 [5] のように、YOLO [4] など画像中の歩行者周辺で物体検出を行ない、白杖を検出することが考えられる。しかし、これらの手法は物体の見えを基に検出しているため、白い傘など、白杖と似た見えの物体を誤検出する可能性がある。

これに対し、白杖利用者は、障害物を探索するために白杖を左右に振る特有の動作を行なうので、この行動に着目して認識することが考えられる。人物の行動を認識する手法としては、人物の姿勢推定に基づいた手法が多く研究されている [2], [3]。これらの手法では、画像から検出された人物骨格の時間変化から、人物の行動を識別している。本発表では、白杖利用者の認識に向けた検討として、推定した人物姿勢系列の変化のみから、行動の種類ではなく、その歩行者が白杖利用者か否かを分類する手法を提案する。

人物の姿勢を推定する代表的な手法として Convolutional Pose Machine [6] や OpenPose [7] などがある。これらの手法は、Convolutional Neural Network (CNN) を用いて人体を構成する各関節点の座標を推定する。これらの手法により得られる関節点の座標系列を入力として、その時系列変化のパターンを分類することで、歩行者系列が白杖利用者か否かを判定できると考えられる。時系列変化の特徴を捉える手法としては、文章や映像といった連続した系列データの認識を目的とした Recurrent Neural Network (RNN) が挙げられる。特に、RNN の一種である Long Short-Term Memory (LSTM) [8] は長時間の系列に対応しており、これを用いた手法 [9] は行動認識の分野で高精度を実現している。

しかし、OpenPoseなどで得られる姿勢は画像中の2次元座標で構成されるため、人物の向きによってそれらの位置関係は大きく変動する。そのため、あらゆる向きの姿勢系列を分類できるようにする必要がある。この問題に対処するため、竹林らは3次元姿勢データを利用した手法 [10] を提案している。この手法では、推定した3次元姿勢を様々な方向から仮想的に観測したものを2次元平面に投影し、それらを学習データとすることであらゆる向きの姿勢の特徴の違いを認識器に学習させている。これにより、人物の転倒動作の認識において、77%の認識率を達成している。しかし、本研究で見分ける対象である白杖利用者と一般歩行者はどちらも歩行しており、その動作の大部分は似ている。そのため、転倒動作と比べて分類は困難である。白杖利用者の認識のように詳細な分類問題においては、分類器に入力される姿勢系列は分類したい事象以外の変動、特に見え方に大きな影響を与える向きによる違い

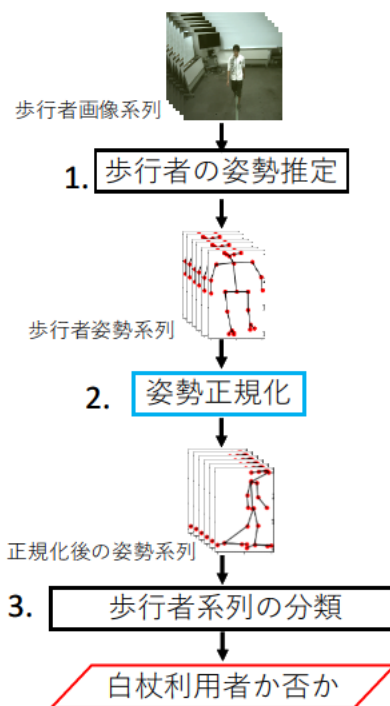


図2 提案手法の処理手順

は存在しないことが望ましい。そのような理想的な状況では、分類器は向きの違いを考慮せず、白杖利用者と一般歩行者の分類にのみ特化することができる。向きによる違いを排除する方法としては、1つの向きの姿勢系列のみを用いて分類器を学習することが挙げられる。しかし、この手法では学習した向き以外の姿勢系列を分類できないという問題がある。

そこで本研究では、図1に示すように、あらゆる向きの姿勢をある1つの向きに変換（姿勢正規化）し、変換された姿勢に対して分類を行なう。これにより、限られた向きの姿勢系列の分類器のみで、あらゆる向きの姿勢に対応することができると考えられる。姿勢正規化は、ある1つの向きの姿勢とそれを同時に様々な別方向から見た姿勢の組を用意し、この組の対応関係を用いて入力された姿勢を変換することで行なう。本発表では、このような事例ベースの姿勢正規化法を導入することにより、様々な向きの歩行者系列に対して、白杖利用者か否かを精度よく分類する手法を提案する。

2 姿勢正規化を用いた白杖利用者の分類法

提案手法の処理手順を図2に示す。提案手法では、まず歩行者画像系列を入力として、1) 各フレームにお

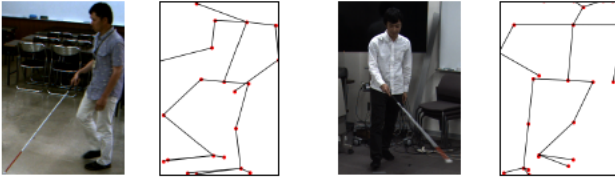


図 3 推定した姿勢の例

る歩行者の姿勢を推定する．その後，提案手法を用いて 2) 各姿勢をある一定の向きに正規化し，3) 正規化後の姿勢系列を分類する．以下で，各処理の詳細について述べる．

2.1 歩行者の姿勢推定

人物の姿勢を手首や鼻などの関節点の 2 次元座標により定義する．関節点の種類数を J とすると，人物の姿勢は $\mathbf{p} \in \mathbb{R}^{2J}$ で表現できる．歩行者の姿勢推定では，人物追跡により得られた幅 w ，高さ h のカラー画像 N フレームからなる入力の歩行者画像系列 $\mathbf{I} = (\mathbf{I}_1, \dots, \mathbf{I}_n, \dots, \mathbf{I}_N)$ の各画像 $\mathbf{I}_n \in \mathbb{R}^{3wh}$ に対して，この $\mathbf{p}_n \in \mathbb{R}^{2J}$ を求める．

\mathbf{p}_n の推定には，OpenPose[7] を用いる．OpenPose は単一の画像 \mathbf{I}_n の特徴から，関節点の座標を示すヒートマップと関節点間の繋がりを示すベクトル場のマップを関節点ごとに推定する．これらと画像特徴を入力として，再度 2 つのマップの推定を行なうという処理を繰り返して推定の最適化を行い，各関節の組み合わせを求めると全関節点の 2 次元座標 \mathbf{p}_n とその信頼度が出力される．

提案手法では，これらの関節点の信頼度の合計がしきい値以下である姿勢を除いた上で，系列の先頭から順に Q フレームだけ取り出して次の処理（姿勢正規化）の入力となる姿勢系列 $\mathbf{S} = \{\mathbf{p}_1, \dots, \mathbf{p}_q, \dots, \mathbf{p}_Q\}$ を構成する．また，推定した各関節点の座標値 (x_q^j, y_q^j) に対しては，以下の式によりその範囲をそれぞれ正規化（座標値正規化）する．

$$x_q^j = \frac{x_q^j - \min_i x_q^i}{\max_i x_q^i - \min_i x_q^i} \theta_x \quad (1)$$

$$y_q^j = \frac{y_q^j - \min_i y_q^i}{\max_i y_q^i - \min_i y_q^i} \theta_y \quad (2)$$

ここで， θ_x, θ_y はパラメータとなる定数である．次節で説明する姿勢正規化では，座標値正規化された関節点により構成される姿勢を，ある 1 つの向きの姿勢へと変換する．図 3 に推定した姿勢の例を示す．

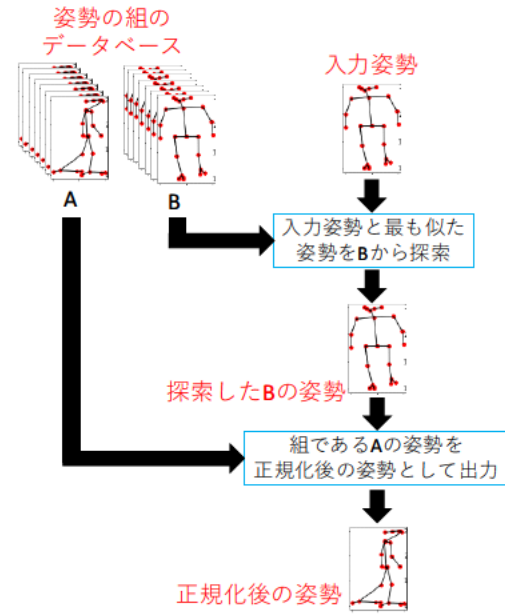


図 4 姿勢正規化の処理手順

2.2 姿勢正規化

提案手法である姿勢正規化では，姿勢推定で求められた姿勢系列 \mathbf{S} をある一定の向きの姿勢系列 $\mathbf{S}' = \{\mathbf{p}'_1, \dots, \mathbf{p}'_q, \dots, \mathbf{p}'_Q\}$ に変換する．以下で，図 4 に示す姿勢正規化の処理手順を説明する．

本研究では，姿勢正規化の手法として，各姿勢 \mathbf{p}_q に対し，あらかじめ用意した姿勢集合の中から変換先として最も適した姿勢を選択し，その姿勢を正規化後の姿勢として出力するという事例ベースの手法をとる．そのため，変換のための姿勢の組のデータベース（姿勢の組の DB）を構築する段階と，これを用いて姿勢正規化を行なう段階に分かれる．姿勢の組の DB 構築段階では，まず変換先となる M 個の基準姿勢で構成された姿勢集合 A とそれらの基準姿勢を D 種類の方向から見た姿勢集合 B を用意する．

$$A = (\mathbf{a}_1, \dots, \mathbf{a}_m, \dots, \mathbf{a}_M) \quad (3)$$

$$B = (\mathbf{b}_{11}, \dots, \mathbf{b}_{1M}, \dots, \mathbf{b}_{d1}, \dots, \mathbf{b}_{dM}) \quad (4)$$

これらから以下のような姿勢の組を作り，姿勢の組の DB を構築する．

$$E = \{(\mathbf{a}_m, \mathbf{b}_{dm}) | \forall m, d, \mathbf{a}_m \in A, \mathbf{b}_{dm} \in B\} \quad (5)$$

姿勢の組の例を図 5 に示す．ここで基準姿勢とは，カ

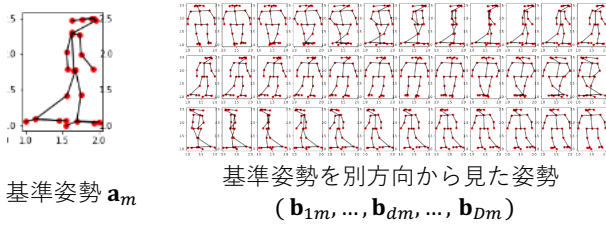


図 5 3次元姿勢の回転例

メラに対してある一定の方向に向いて歩行したときの姿勢である。データベース構築に用いる姿勢は、複数台のキャリブレーション済みカメラを用いて人物の各関節の3次元座標により構成される3次元姿勢を推定し、その姿勢を D 方向から仮想的に観測して作成する。

正規化段階では、入力姿勢 p_q と最も似た姿勢 $b_{d'm'} \in B$ を探索する。似た姿勢の基準として姿勢間の Euclidean 距離を用いることで、 $b_{d'm'}$ は以下の式で求められる。

$$b_{d'm'} = \operatorname{argmin}_{b_{dm}} (\|p_q - b_{dm}\|_2^2) \quad (6)$$

最後に、探索された $b_{d'm'}$ と組である $a_{m'}$ を正規化後の姿勢 $p'_q \in \{a_{m'} | (a_{m'}, b_{d'm'}) \in E\}$ として出力する。

以上の手順により、人物の姿勢をある一定の向きに正規化するが、本発表では変換先となる基準姿勢の向きは真横とした。これは、白杖利用者特有の動作を考慮したためである。白杖利用者が白杖を左右に振る動作には個人差があるが、大半の白杖利用者は共通して白杖を持った手を前方に突き出す。一般の歩行者が腕を前後に振り歩行するのに対し、白杖利用者の腕は前方に突き出されたまま固定されている。この姿勢の特徴は、真横を向いている姿勢に顕著に表れるため、基準姿勢の向きを右向きとした。姿勢正規化により、人物姿勢を変換した例を図 6 に示す。

2.3 歩行者系列の分類

姿勢正規化によって1つの向きに変換された歩行者の姿勢系列が白杖利用者のものか否かの分類を行なう。この分類を行なうためには、2.2節で述べたような姿勢の特徴が手がかりとなる。「腕を左右に振っている」、「腕が前方に突き出されたまま固定されている」というような特徴を捉えるためには、単一の姿勢だけでなく系列全体での姿勢変化を考慮する必要がある。そのよ

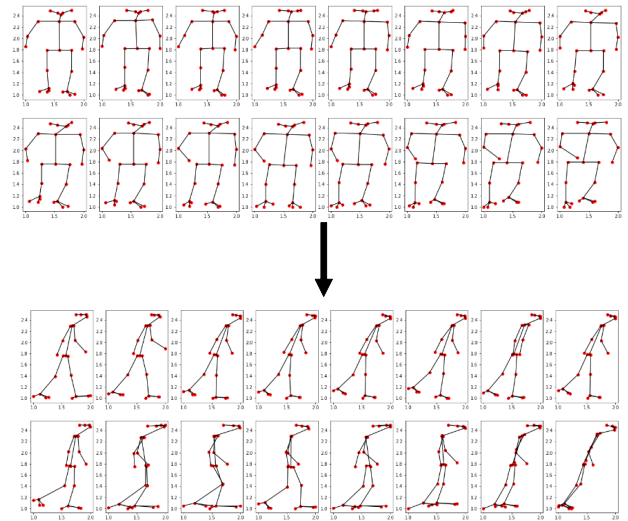


図 6 姿勢正規化の例

うな時系列変化を考慮できるニューラルネットワークとして、1節でも述べた LSTM [8] を利用する。

使用するネットワークは、3つの全結合層、1つの LSTM 層、3つの全結合層の順に並んだ構造である。ネットワークへのへの入力、姿勢正規化により変換された姿勢系列 S' を用いる。 S' をフレームの重複を許して長さ L フレームずつに分割する。これにより作成される $2JL$ 次元のベクトルを1つの入力とする。最終的に、白杖利用者と一般歩行者それぞれへの分類スコアを正規化して出力する。

3 実験および考察

本節では、提案手法の有効性を確認するために行った実験とその結果への考察について述べる。

3.1 データセット

本節では、分類器の学習および評価に用いたデータセットと姿勢正規化に用いた姿勢の組の DB について述べる。

まず学習および評価データとして、白杖を利用しない一般の歩行者と白杖利用者が歩行する様子を撮影し、データセットを構築した。撮影は屋内と屋外を含む5つの異なる場所で行った。撮影対象は、一般の歩行者役が17名、白杖利用者役が22名である。ただし、白杖利用者役のうち17名は一般の歩行者役が兼任しており、実生活で白杖を利用している人物は5名である。これらの映像の歩行者領域を人手で切り出すことで、歩行者系列を構成した。最終的に、一般の歩行者の系列

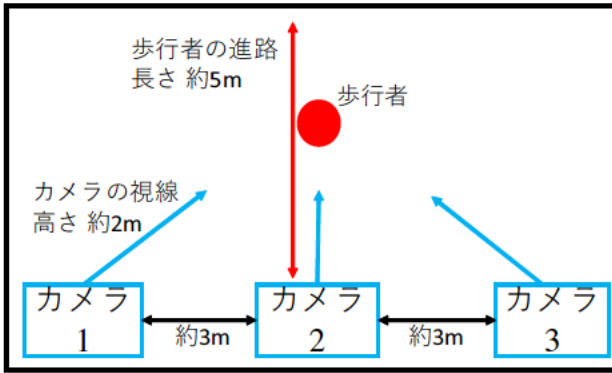


図 7 姿勢の組の DB 撮影時のカメラ配置

表 1 分類器用データセットに含まれる歩行者系列数

| 場所 | 1 | 2 | 3 | 4 | 5 | 全体 |
|--------|----|----|----|----|-----|-----|
| 白杖利用者 | 23 | 12 | 12 | 10 | 76 | 133 |
| 通常の歩行者 | 26 | 6 | 25 | 0 | 76 | 133 |
| 全系列数 | 49 | 18 | 37 | 10 | 152 | 266 |

と白杖利用者の系列はともに 133 系列ずつ切り出した。表 1 にデータセットの詳細を、図 8 に各場所における画像例を示す。

次に、姿勢の組の DB を構築するために、3 台のキャリブレーション済みカメラを用いて歩行者を撮影した。図 7 に撮影時のカメラ配置を示す。撮影場所は 1 箇所、撮影対象は一般の歩行者役と白杖利用者役がともに 1 名という設定で撮影した。その後、撮影した歩行者に対して OpenPose を用いて 3 次元姿勢を推定し、それらの姿勢を仮想的に観測し、2 次元平面に投影することで様々な向きの 2 次元姿勢を作成した。カメラに対して正面を向いた姿勢を 0 度とし、水平方向に 10 度刻みで反時計回りに 1 回転させた。これにより、1 つの 3 次元姿勢に対して $D = 36$ 方向から見た姿勢を作成した。このうち、90 度の向き（右向き）の姿勢を基準姿勢とすることで、3 次元姿勢 1 つにつき、基準姿勢 1 つと 36 方向から見た姿勢の組を構成した。推定した 3 次元姿勢の種類は 4,616 種類であり、最終的に $M = 4,616$ 種類の基準姿勢とそれらを回転させた $MD = 166,176$ 種類の姿勢により、姿勢の組の DB を構築した。

3.2 実験内容

実験では、姿勢正規化の有無および、正規化先の向きの違いによる歩行者系列の分類精度を比較した。分



図 8 各場所で撮影した画像例

表 2 歩行者系列の分類結果

| 場所 | 1 | 2 | 3 | 4 | 5 | 全体 |
|-------------------|------|------|------|------|------|------|
| 正規化なし | 0.82 | 0.70 | 0.55 | 0.53 | 0.64 | 0.66 |
| 0 度正規化 | 0.49 | 0.60 | 0.71 | 0.84 | 0.5 | 0.55 |
| 90 度正規化 (提案手法) | 0.80 | 0.66 | 0.70 | 0.55 | 0.71 | 0.71 |

類は歩行者系列が白杖利用者のものか否かの 2 クラス分類である。姿勢推定については、構成される姿勢系列の長さを $Q = 64$ フレーム、分割時の長さを $L = 32$ フレームとして 5 つに分割した。また、推定する関節点の種類数を $J = 25$ 、関節点の座標値の範囲はパラメータを $\theta_x = 1$ 、 $\theta_y = 1.5$ とし、水平方向を $[0, 1.0]$ 、鉛直方向を $[0, 1.5]$ とした。評価は、5 つの場所で撮影されたデータのうち、1 箇所でも撮影されたデータを評価データに、その他を学習データとした 5 分割交差検証を行った。評価指標は認識率とした。

3.3 実験結果

姿勢正規化をしていない場合、0 度（正面向き）へ正規化した場合、90 度へ正規化した場合（提案手法）の結果を表 2 に示す。実験の結果、全体の認識率において提案手法の精度が最も高く、正規化しなかった場合と比べて認識率が 0.05 向上した。また、0 度へ正規化する場合と比べても大きく精度が向上しており、提案手法の有効性を確認できた。

3.4 考察

本節では、実験結果に対する考察を述べる。全ての手法に共通して見られた傾向として、場所 4 での精度が低かった。この原因としては 2 つ考えられる。1 つ

目はこの場所での撮影データは他の場所と違い、白杖利用者役のデータしか含まれていないことである。通常の歩行者への分類精度に対して、白杖利用者への精度が低いために場所4における精度が低くなっている可能性がある。2つ目は撮影環境の違いである。場所4での撮影では、他の場所に比べてカメラの位置が高く、カメラの俯角が異なる。そのため、他の場所とは異なる姿勢パターンとなり分類精度が低下したと考えられる。

次に、正面への正規化を行った場合の結果について考察する。正面へ正規化を行った場合、場所1と場所5での精度が大きく低下し、結果として全体の精度も低下した。この2つの場所のデータに共通することとして、左右を向いた歩行者画像が多く含まれることが挙げられる。この結果から、左右を向いた姿勢を0度に正規化することは精度の低下をもたらした可能性がある。一方、場所3と4については90度への正規化を行なう場合と比べても精度が向上している。このことから、0度への正規化が90度への正規化に必ずしも常に劣るわけではなく、90度以外の向きへの正規化結果を取り入れることで、更なる精度の向上が期待できると考えられる。

4 むすび

本発表では、白杖利用者の認識に向けた検討として、人物の姿勢の変化から歩行者画像系列が白杖利用者のものか否かを分類する手法を提案した。提案手法ではあらゆる向きの姿勢を1つの向きの姿勢に正規化することで、限られたパターンの姿勢に対する分類器であらゆる向きの姿勢の分類に対応した。評価実験の結果では、姿勢正規化を取り入れることにより、歩行者画像系列の分類精度が向上し、提案手法の有効性を確認できた。また、正面への正規化を行った場合の結果から、右向き以外の向きへの正規化が有効な場面も存在することがわかった。

今後は、姿勢正規化を用いて学習画像系列を様々な向きの系列に変換することによる学習データの増量、右向き以外の向きへの姿勢正規化の取り入れ、見えの特徴との統合などについて検討する。

謝辞

本研究の一部は、JSPS 科研費 JP17H00745 の助成による。

参考文献

- [1] U. Tanikawa, Y. Kawanishi, D. Deguchi, I. Ide, H. Murase, and R. Kawai, “Wheelchair-user detection combined with parts-based tracking”, Proc. 12th Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol.5, pp.165–172, Feb. 2017.
- [2] T. M. Le, N. Inoue, and K. Shinoda, “A fine-to-coarse convolutional neural network for 3D human action recognition”, Proc. 29th British Machine Vision Conf., pp184-1–184-13, Sept. 2018.
- [3] M. Liu and J. Yuan, “Recognizing human actions as the evolution of pose estimation maps”, Proc. 2018 IEEE Conf. on Computer Vision and Pattern Recognition, pp.1159–1168, June 2018.
- [4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, real-time object detection”, Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition, pp.779–788, June 2016.
- [5] 西田尚樹, 川西康友, 出口大輔, 井手一郎, 村瀬洋, 朴君, “人物の姿勢に着目した白杖利用者検出に向けた検討”, 平成30年度電気・電子・情報関係学会東海支部連合大会, no.J5-2, Sept. 2018
- [6] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, “Convolutional pose machines”, Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition, pp. 4724–4732, June 2016.
- [7] Z. Cao, T. Simon, S.E. Wei, and Y. Sheikh, “Real-time multi-person 2D pose estimation using part affinity field”, Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition, pp.7291–7299, July 2017.

- [8] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, *Neural Computation*, vol.9, no.8, pp.1735–1780, Nov. 1997.
- [9] L. Sun, K. Jia, K. Chen, D. Y. Yeung, B. E. Shi and S. Savarese, “Lattice long short-term memory for human action recognition”, *The Proc. 2017 IEEE Int. Conf. on Computer Vision*, pp.2147–2156, Oct. 2017.
- [10] 竹林あやか, 岩田英和, 岩堀祐之, 福井真二, “行動認識による転倒シーンの自動検知”, 平成 30 年度電気・電子・情報関係学会東海支部連合大会, no.J4-1, Sept. 2018.