

# 体格を考慮した歩き方からの手荷物の主観的重さ推定の検討

水野雅也<sup>†</sup> 藤田倫弘<sup>††</sup> 川西康友<sup>††</sup> 出口大輔<sup>†</sup> 村瀬 洋<sup>†</sup>

<sup>†</sup> 名古屋大学 大学院 情報学研究科 〒464-8601 愛知県名古屋市千種区不老町  
<sup>††</sup> 理化学研究所ガーディアンロボットプロジェクト 〒619-0088 京都府相楽郡精華町光台 2-2-2  
E-mail: <sup>†</sup>mizunom@vislab.is.i.nagoys-u.ac.jp, <sup>††</sup>{tomohiro.fujita,yasutomo.kawanishi}@riken.jp,  
<sup>†††</sup>{ddeguchi,murase}@nagoya-u.jp

**あらまし** 本研究では、ロボットによる支援の1つである重い手荷物の運搬支援の実現のため、支援すべき人に適切なタイミングで支援する際の判断基準としてその人が手荷物をどのくらい重く感じているか（主観的重さ）を推定することを目的とする。手荷物の重さが所持者の歩行時の姿勢に影響を与えることに着目し、歩行の様子から手荷物の主観的重さを推定する手法を提案する。歩行時の姿勢は、人の体格や手荷物の種類によって変化する。そこで、主観的重さ推定モデルの学習と同時に、人の体重推定と手荷物種類の認識もサブタスクとして同時に学習する。このとき、入力として3D骨格系列に加え、人物シルエット画像系列も用いることで体格の情報を得る。一方、関節点の速度も重要な情報であるものの、歩行の様子を表現する3D人物骨格系列をサンプリングにより固定長化して入力する一般的なGCNベースの認識手法を利用すると関節点の速度の情報が失われてしまう。そこで、3D人物骨格系列に対し、各関節点の速度を付加した特徴量を用いることで、速度の情報を保ったまま固定長化する。実験では、サブタスクを利用しない手法や骨格座標・骨格速度系列を利用しない手法と比較して高精度な結果を示した。

**キーワード** 主観的重さ推定, 人物骨格系列, グラフ畳み込み

## 1. はじめに

近年、見回りロボットが様々な場所で利用されるようになってきた[1],[2]。人と共存するロボットには、見回りだけでなく、観測した周囲環境の情報をもとに自主的に人を支援する能力が求められる。本報告では、このような支援の一つである、手荷物を重そうに運ぶ人を自主的に支援することに着目する。

手荷物を運ぶロボットは広く開発が行われている[3],[4]。そのため、本研究ではロボットによる自主的な支援のために、重い手荷物を運ぶ支援を行うか否かを適切に判断する機能に着目する。人が支援を必要としていないにもかかわらずロボットが支援をしてしまうと、その人がロボットの存在を煩わしく感じる恐れがある。人がロボットに煩わしさを感じるようでは、人とロボットが信頼関係を築けず、ロボットが人に受け入れられなくなってしまうことが考えられる。これを避けるためには、ロボットが支援を必要としている人を見極め、適切なタイミングで支援する必要がある。

上述の見極めのためには、人が手荷物をどの程度重く感じているかを推定する必要がある。本研究では、「手荷物をどの程度重く感じているか」を主観的重さと定義する。主観的重さが大きければ、人がその手荷物を重く感じているといえる。そのため、主観的重さが推定できればその人にロボットによる支援が必要かどうかを判断できる。以上より本研究では手荷物の主観的重さの推定を目的とする。

手荷物の実際の重さによる負荷が人の歩行に与える影響を調査した論文[5]では、手荷物の実際の重さが大きくなると、骨

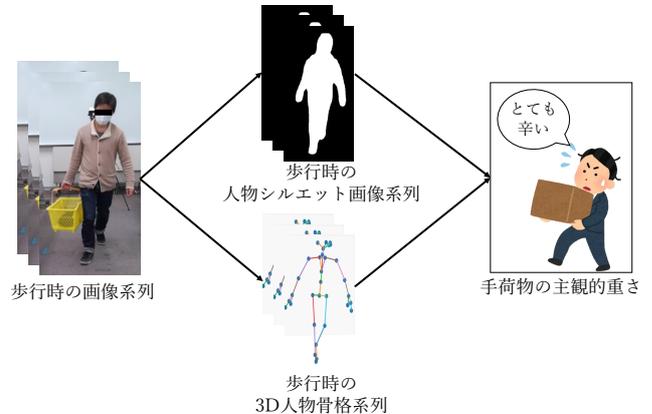


図1: 歩き方の特徴からの主観的重さ推定

盤が大きく傾き、これが歩行に影響を与えることが示されている。また、手荷物の実際の重さが大きければ主観的重さも大きくなることから、主観的重さは実際の重さの影響を受ける。このことから、主観的重さも歩き方に影響を与えるといえる。以上より、本研究では歩き方に着目して手荷物の主観的重さを推定する。しかし、歩き方は手荷物の主観的重さだけでなく所持者の体重や手荷物の種類にも影響を受ける。論文[5]では、手荷物の物理的重さが所持者の体重の15%を超えると歩行効率が低下することが示されている。そのため、主観的重さ推定では人の体重や手荷物種類を考慮する必要がある。

人の歩行時の姿勢を表現した特徴として、骨格の時間変化を表現できる人物骨格系列がある [6]。人物骨格系列は背景や服装の違いに頑健であるという特徴がある。そのため、本研究では歩き方を歩行時の人物骨格系列で表現する。また、歩行は2歩分の繰り返しであることに着目し、2歩分を1周期として歩行1周期分の人物骨格系列を歩き方の特徴として利用する。しかし、歩き方の特徴として人物骨格系列のみを利用すると、人の体格といった情報が失われてしまう。本節で述べたように、手荷物の主観的重さ推定では人の体格の情報が必要である。そのため、本研究では人の体格も保持した歩き方の特徴として、歩行1周期分の人物シルエット画像も利用する。人物シルエット画像は人物の領域のみをマスクした画像であり、歩容認識で広く利用されている [7], [8]。人物シルエット画像は歩き方に関係のない情報である背景や服の色、照明の影響を排除できる。

人物骨格系列は行動認識等で注目されている [9]~[12]。これらの行動認識手法では、行動認識に重要な関節間の関係性を考慮できる GCN ベースの手法が採用されている。そのため、手荷物の主観的重さの推定にも歩行時の人物骨格系列を入力とする行動認識手法を応用できる。一方で、GCN ベースの行動認識手法では入力を固定長にする必要がある。系列の固定長化の方法は様々であるが、歩行時の姿勢の違いを特に注目したいため、系列の対応するフレーム間で歩行段階が一致していることが望ましい。そのため、本研究では系列から等間隔にフレームを抜き出すことで固定長化するフレームサンプリングにより系列の固定長化を行う。しかし、フレームサンプリングによる系列の固定長化では、歩き方の特徴として重要な関節点の速度の情報が失われてしまう。

これらの問題に対し、本研究では、主観的重さ推定と同時にサブタスクとして人の体重推定と手荷物種類認識も学習する手法を提案する。主観的重さ推定以外に歩き方に影響を与える人の体重と手荷物種類を同時に推定することで、人の体重と手荷物種類を考慮して主観的重さを推定できる。また、歩行1周期分の人物骨格系列に含まれる各関節点に速度の情報を追加した後固定長とすることで、骨格座標・骨格速度系列に変換して利用する。骨格座標・骨格速度系列によって、歩行1周期分の系列が固定長であっても元の速度情報を残したまま表現できる。

以下、2節で関連研究について述べる。次に、3節で手荷物の主観的重さの推定手法について述べる。さらに、4節で実験とその評価について述べる。最後に、5節でまとめと今後の課題について述べる。

## 2. 関連研究

### 2.1 重さ推定の研究

手荷物の重さが変化しても、手荷物のそのものの見た目は変化しない場合が多く、手荷物の外観からその重さを推定することは難しい。そのため、手荷物を所持した人に生じる体の変化に着目して重さを推定する方法が研究されている。手荷物を所持した人に生じる変化を捉えて手荷物の重さを推定する研究に、身体動揺を用いる研究 [13] がある。身体動揺とは、直立して静止していても体に生じるわずかな揺れである。この研究 [13] で

は、手荷物の重さが重いほど大きくなるという身体動揺の特徴に着目して、手荷物の重さを推定する。この研究 [13] では、身体動揺を取得するために静止している人を頭上真上から撮影する必要がある、今回想定しているようなロボットの視点からは観測が難しい。

他の体の変化に着目した物体の重さ推定手法として、持ち上げ動作を利用する研究 [14] がある。この研究では、物体の重さによって持ち上げる時の手の動きが変化することに着目し、この時の手の動きから物体の重さを推定する。しかし、実用の際には物体を持ち上げる瞬間を撮影する必要がある、利用できる状況が限定される。

### 2.2 人物骨格系列を利用した認識研究

人の行動を認識する場合、体の関節点同士の位置関係やそれらの時間変化を捉えることが重要である。そのため、骨格系列を利用する行動認識タスクにおいては、骨格の関節点同士の位置関係や骨格の時間的な変化の特徴を捉えて認識するような研究がなされてきた。

時間的な変化に着目して人体骨格系列から行動認識を行う手法として、Long Short-Term Memory (LSTM) を利用するものがある [9]。LSTM を用いることで行動認識において重要な時間的特徴を考慮して認識できる。また、加藤ら [6] は、LSTM を用いて人物骨格系列から歩行に動きの属性を付与する手法を提案し、歩行の様子を表現するオノマトペを歩行時の人物骨格系列から推定している。

近年では骨格の時系列的な特徴だけでなく、空間的な特徴も重要視した手法が主流になっている。空間的な特徴を利用するため、人体の構造をグラフで表現し、そのグラフを GCN [15] に入力して特徴抽出し、その特徴から行動を認識する。骨格系列のグラフを利用する行動認識研究として ST-GCN [10] がある。この手法は人物骨格系列の各フレームを骨格構造に従ったグラフで表現し、GCN に入力して時系列特徴を抽出した後、各関節点に対して時間方向の畳み込みを行って特徴を抽出する。この手法により、姿勢の空間的・時間的変化を考慮した認識を実現している。また、ST-GCN を発展させた手法として MS-G3D [11] がある。この手法では、人物骨格に従ったグラフ構造の各関節点からのホップ数ごとに畳み込みフィルタを用意し、それらを用いて畳み込みを行う。この手法を用いることで、骨格構造的に離れた関節点同士の特徴をうまく抽出できるようになり、離れた関節点同士の関係に特徴を有する行動の認識に効果的である。ST-GCN を発展させた別の手法として STA-GCN [12] がある。この手法では、フレームごとに関節の重要度を表す Attention node と、動作ごとに関節間の関係性を表す Attention edge を獲得し、これによって入力ごとに異なるグラフ構造を生成する。この手法により、動作ごとに異なる関節間の関係性を考慮できる。

### 2.3 歩容認識手法

本研究では歩容認識手法を一部参考にしてしている。そのため、歩容認識の関連研究について概観する。歩容認識とは、人の歩行の様子からその人が誰かを認識する生体認識手法の一つである。歩容認識には、歩きの仕組みを考慮したうえで認識するモ

データベース手法と、歩きの仕組みを考慮せずに認識するモデルフリー手法に分かれている。

モデルフリー手法では主に画像から人物領域のみをマスクした人物シルエット画像を利用した認識が行われている [7], [8]. shiraga らの手法では、歩行 1 周期分の人物シルエット画像を平均することで得られる Gait Energy Images (GEI) から、CNN を用いて歩容認識をしている。しかし、人物シルエット画像は服装やカメラ視点の影響を大きく受けるため、人物シルエット画像のみを利用する手法ではこれらの影響を排除することは難しい。

データベース手法では、近年 3D 人物骨格系列を利用した手法が提案されている。2.2 項で説明した ST-GCN の考え方を元にして歩容認識を行う手法が提案されている [16]。3D 人物骨格系列を利用することで、服装やカメラ視点の影響を排除して歩容認識ができる。しかし、実環境では 3D 人物骨格推定手法で推定した骨格系列を利用するため、精度は人物骨格推定手法に依存してしまう。

### 3. 歩き方の特徴に着目した手荷物の主観的重さ推定

#### 3.1 提案手法の概要

本研究では、歩き方の特徴として歩行 1 周期分の 3D 人物骨格系列と人物シルエット画像系列を利用する。3D 人物骨格系列はカメラ座標系における関節点の  $(X, Y, Z)$  座標の集合である。このとき、左足が前かつ足の開きが極大となるフレームを基準とし、それらのフレームの間の系列を歩行 1 周期分として切り出す。以降では、歩行 1 周期分の 3D 人物骨格系列により表現された歩行時の姿勢から、手荷物の主観的重さを回帰で推定する手法について述べる。図 2 に提案手法の処理手順を示す。提案手法ではまず、歩行 1 周期分の 3D 人物骨格系列  $S_i$  と人物シルエット画像系列  $U_i$  を入力とする。次に、歩行 1 周期分の 3D 人物骨格系列  $S_i$  に対して前処理を行い、骨格座標・骨格速度系列  $\hat{S}_i$  に変換する。また、歩行 1 周期分の人物シルエット画像系列  $U_i$  に対しても前処理を行い、平均シルエット画像  $\hat{u}_i$  に変換する。最後に、 $\hat{S}_i$  と  $\hat{u}_i$  を手荷物の主観的重さ推定器  $f$  に入力し、手荷物の主観的重さの推定結果を得る。手荷物の主観的重さ推定器の学習はマルチタスク学習により行う。マルチタスク学習では、手荷物の主観的重さ推定をメインタスクとして、手荷物の種類  $k$  の認識をサブタスク 1、手荷物を所持している人の体重推定をサブタスク 2 として学習する。

以下ではまず、手荷物の主観的重さの定義について述べる。次に、歩行 1 周期分の 3D 人物骨格系列  $S_i$  と人物シルエット画像系列  $U_i$  に対して行う前処理について述べる。最後に、提案手法のマルチタスク学習について述べる。

#### 3.2 手荷物の主観的重さの定義

本研究では、「手荷物を人がどの程度重く感じているか」を手荷物の主観的重さとして定義する。主観的重さは New Borg Scale [17] により定量化する。New Borg Scale は表 1 に示すように、運動のつらさを定量化した指標である。今回対象にしている「手荷物を運ぶ」という動作も運動のような肉体的な負荷

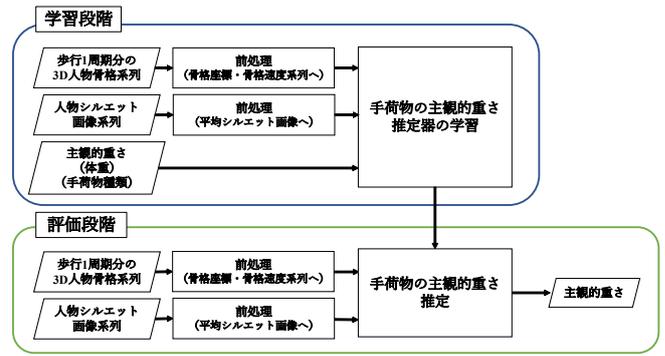


図 2: 提案手法の処理手順

| 点  | 説明              | 点   | 説明              |
|----|-----------------|-----|-----------------|
| 10 | Very, very Hard | 4   | Somewhat Hard   |
| 9  |                 | 3   | Moderate        |
| 8  |                 | 2   | Light           |
| 7  | Very Hard       | 1   | Very Light      |
| 6  |                 | 0.5 | Ver, very Light |
| 5  | Hard            | 0   | Nothing at all  |

表 1: New Borg Scale

がかかる行動であるといえるため、この指標を「人がどの程度重く感じているか」の定量化に利用する。提案手法ではこの指標の値を主観的重さとして推定する。

#### 3.3 前処理

本節では、歩行 1 周期分の 3D 人物骨格系列  $S_i$  と人物シルエット画像系列  $U_i$  に対して行う前処理について述べる。

はじめに、歩行 1 周期分の 3D 人物骨格系列  $S_i$  に対して行う前処理について述べる。まず、 $S_i$  内のすべてのフレームに対して、系列の先頭フレームの骨盤の位置が原点になる方に平行移動して人の位置の正規化をすることで  $\tilde{S}_i$  に変換する。次に、 $\tilde{S}_i$  内のすべてのフレームに対して、系列の先頭フレームの両腰の関節点の  $(X, Y)$  座標が  $Y = 0$  平面に乗るように回転させることで歩行向きの正規化を行い、 $\hat{S}_i$  に変換する。その後、 $\hat{S}_i$  に骨格速度を付与する。骨格速度とは系列  $\hat{S}_i$  の隣接するフレーム内の対応する関節点との位置の差として定義する。系列  $\hat{S}_i$  の  $t$  番目のフレームの  $j$  番目の関節点の座標を  $(\hat{X}_t^j, \hat{Y}_t^j, \hat{Z}_t^j)$  とすると、骨格速度  $(\dot{\hat{X}}_t^j, \dot{\hat{Y}}_t^j, \dot{\hat{Z}}_t^j)$  は式 (1) のように計算される。

$$(\dot{\hat{X}}_t^j, \dot{\hat{Y}}_t^j, \dot{\hat{Z}}_t^j)^\top = (\hat{X}_t^j, \hat{Y}_t^j, \hat{Z}_t^j)^\top - (\hat{X}_{t-1}^j, \hat{Y}_{t-1}^j, \hat{Z}_{t-1}^j)^\top \quad (1)$$

各関節点の骨格速度を対応する関節点に追加することで  $t$  番目のフレームの  $j$  番目の関節点を 6 次元の特徴  $(\hat{X}_t^j, \hat{Y}_t^j, \hat{Z}_t^j, \dot{\hat{X}}_t^j, \dot{\hat{Y}}_t^j, \dot{\hat{Z}}_t^j)^\top$  とする。最後に、系列からフレーム数  $M$  になるよう等間隔にフレームを抜き出すことで系列を固定長とし、骨格座標・骨格速度系列  $\hat{S}_i$  を得る。

次に、人物シルエット画像系列  $U_i$  に対して行う正規化について述べる。本研究では、一枚の画像で腕の振りなどの歩容認識に重要な特徴を表現できる平均シルエット画像を用いて推定を行う。そのため、 $U_i$  に含まれるフレームの平均をとって平

均シルエット画像  $\hat{\mathbf{u}}_i$  に変換する。

### 3.4 提案手法でのマルチタスク学習

提案手法では、骨格座標・骨格速度系列  $\hat{\mathbf{S}}_i$  と平均シルエット画像  $\hat{\mathbf{u}}_i$  を入力とし、手荷物の主観的重さを推定する。提案手法の構造を図3に示す。図3内の数字はそれぞれのモジュールから出力される特徴の次元数を表す。提案手法ではまず、 $\hat{\mathbf{S}}_i$  を GCN ベースの特徴抽出器  $f_m$  に入力して骨格中間特徴  $\mathbf{m}_i$  を、 $\mathbf{m}_i$  を GCN ベースの特徴抽出器  $f_p$  に入力して骨格特徴  $\mathbf{p}_i$  を得る。 $\theta_m, \theta_p$  は学習可能な重みパラメータである。

$$\mathbf{m}_i = f_m(\hat{\mathbf{S}}_i; \theta_m) \quad (2)$$

$$\mathbf{p}_i = f_p(\mathbf{m}_i; \theta_p) \quad (3)$$

また、特徴抽出器  $f_m, f_p$  として MS-G3D [11] を利用する。次に、 $\hat{\mathbf{u}}_i$  を CNN で構成された特徴抽出器  $f_a$  に入力して外観特徴  $\mathbf{a}_i$  を得る。 $\theta_a$  は学習可能な重みパラメータである。

$$\mathbf{a}_i = f_a(\hat{\mathbf{u}}_i; \theta_a) \quad (4)$$

その後、骨格特徴  $\mathbf{p}_i$  と外観特徴  $\mathbf{a}_i$  を連結した特徴  $(\mathbf{p}_i, \mathbf{a}_i)$  から全結合層  $g_s$  を利用して主観的重さ  $w_i^s$  を推定する。この時、体重推定と手荷物種類認識も全結合層  $g_b, g_k$  を利用して同時に行う。体重推定においては骨格特徴と外観特徴を連結した特徴  $(\mathbf{p}_i, \mathbf{a}_i)$  を全結合層  $g_b$  に入力して体重  $w_i^b$  を推定し、手荷物種類認識においては骨格中間特徴と外観特徴を結合した特徴  $(\mathbf{m}_i, \mathbf{a}_i)$  を全結合層  $g_k$  に入力して手荷物種類  $k_i$  を推定する。 $\theta_{g_s}, \theta_{g_b}, \theta_{g_k}$  は学習可能な重みパラメータである。

$$w_i^s = g_s((\mathbf{p}_i, \mathbf{a}_i); \theta_{g_s}) \quad (5)$$

$$w_i^b = g_b((\mathbf{p}_i, \mathbf{a}_i); \theta_{g_b}) \quad (6)$$

$$k_i = g_k((\mathbf{m}_i, \mathbf{a}_i); \theta_{g_k}) \quad (7)$$

これらの推定器は中間層の活性化関数として Leaky ReLU [18] を利用する。

マルチタスク学習では、各推定値と真値に基づく誤差  $L$  が最小になるようにネットワークの重みを更新する。

$$L = \lambda_s L_s + \lambda_b L_b + \lambda_k L_k \quad (8)$$

ここで、 $L_s$  は主観的重さの平均二乗誤差、 $L_b$  は体重推定の平均二乗誤差、 $L_k$  は手荷物種類認識のクロスエントロピー誤差である。また、 $\lambda_s, \lambda_b, \lambda_k$  はそれぞれ  $L_s, L_b, L_k$  の重みである。

## 4. 実験

### 4.1 データセット

歩行時の人の様子を撮影し、その人の手荷物の主観的重さを真値として付与した公開データセットは存在しないため、本研究では実験用データセットの撮影を独自に行った。本節ではデータセットを構築した際の撮影条件及びその内容について述べる。

本研究では、1人の歩行者が1つの手荷物を持って歩いている状況を想定する。実験参加者に手荷物を所持して歩かせ、その様子を高さ2mの位置に設置したMicrosoft社製の Azure

Kinect センサを用いて歩行者の前後から撮影することで3D人物骨格系列を取得した。実験環境を横から見た様子を図4に示す。Azure Kinectにより取得できる3D人物骨格はカメラ座標系における関節点の  $(X, Y, Z)$  座標の集合であるが、本研究では、カメラから降ろした垂線と床面との交点を原点とし、床面が水平となる  $(Y=0)$  ように設定した世界座標系に変換して用いた。また、撮影した画像を Mask R-CNN [19] に入力することで、人物シルエット画像を得た。

データセット撮影は、30名の被験者(男性15名、女性15名)により実施した。また、手荷物の種類は、手提げ鞆、肩さげ鞆、リュックサック、段ボール、買い物かごとし、手荷物の内容量は、0kg, 5kg, 7.5kg, 10kg, 12.5kg, 15kgとした。内容量は重りで調整した。主観的重さの正解値は、アンケート調査により収集した。アンケートは New Borg Scale [17] (表1) を参考にして作成した指標に従い、その手荷物を所持した歩行がどの程度辛かったかについて点数を付けさせた。このとき、手荷物の実際の重さの正解値を知ることが主観的重さに影響しないように、所持する手荷物の中身は被験者に見えないようにした。さらに、前の系列の撮影で所持した手荷物の重さによる影響をとり除くため、撮影ごとに休憩を設けた。この実験により、被験者1名につき30パターンの3D人物骨格系列を撮影した。データセットの例を図5に示す。また、全実験参加者から撮影データの研究目的での利用・公開に同意を得た。

### 4.2 実験手順

本実験では、提案手法における主観的重さの推定の精度を評価する。この時、30人分のデータに対して、5人を評価に利用し、それ以外の25人を学習に利用する6分割交差検証で評価した。

データセットには歩行1周期分の3D人物骨格系列が24,015系列しか含まれないため、Gaussian ノイズを用いて3D人物骨格の各関節点にノイズを付与することでデータ拡張した。このデータ拡張により、データ数を10倍の240,150系列とした。本実験では3.3節で述べた系列の固定長化の際、固定長  $M=50$  となるように前処理を施した。

提案手法におけるサブタスクと骨格座標・骨格速度系列の有効性を確認するため、それぞれのサブタスクを抜いた場合の手法と提案手法を比較した。

実験の評価指標として、平均絶対値誤差 MAE を用いる。

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |w_i^s - \hat{w}_i^s|, \quad (9)$$

ここで、 $N$  は評価用データに含まれる歩行1周期分の3D人物骨格系列の数を表す。また、本研究の利用状況として想定している「重い手荷物を運ぶ人に対して支援を行うか否か」という観点で考えると、ある許容誤差しきい値での推定性能も評価することが望ましい。本実験では、ある許容誤差しきい値での推定性能を評価する指標を許容誤差精度 (TA) と命名し、式(10)のように定義した。

$$\text{TA}_\tau = 100 \frac{T_\tau}{N}, \quad (10)$$

ここで  $\tau$  は許容誤差しきい値を表し、 $T_\tau$  は推定誤差  $\tau$  以内の

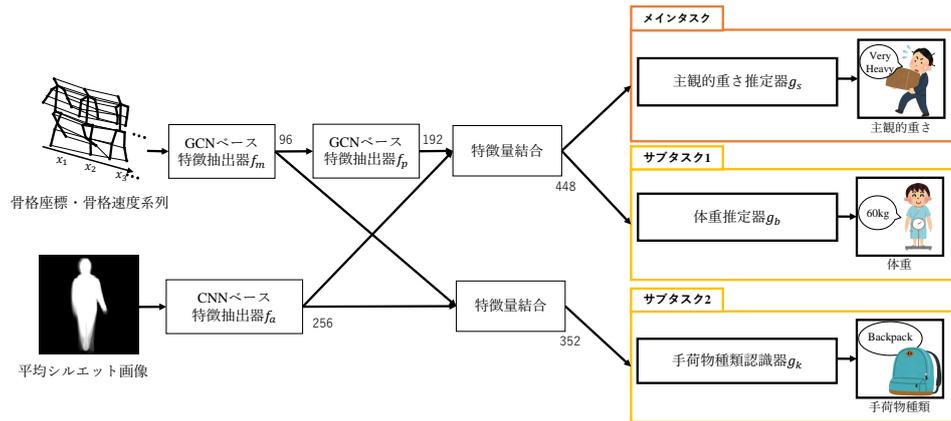


図 3: 提案手法のネットワーク図

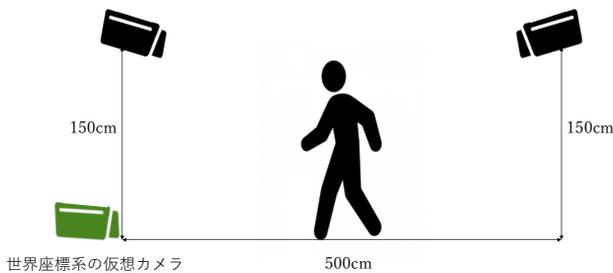


図 4: 撮影環境を横から見た図

|  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|
| 手さげ鞆<br>実際の重さ<br>13kg<br>主観的長さ<br>0    |  |  |  | 肩さげ鞆<br>実際の重さ<br>7.5kg<br>主観的長さ<br>2   |  |  |  |
| リュックザック<br>実際の重さ<br>10kg<br>主観的長さ<br>3 |  |  |  | 段ボール箱<br>実際の重さ<br>12.5kg<br>主観的長さ<br>3 |  |  |  |

図 5: データ例

データ数を表す。

### 4.3 実験結果

表 2 に提案手法と提案手法から各要素を抜いた手法との主観的長さ推定結果の比較を示す。また、主観的長さ推定の例を図 6 に示す。表 2 から、人の体重推定と手荷物種類認識を入れることで精度が向上することを確認した。また、骨格座標・骨格速度系列を利用することでも精度が向上することが確認できた。

### 4.4 考察

本研究では、ロボットが重い手荷物を人に代わって運ぶ支援をするか否かの判断ができるようにするために、主観的長さの推定を目的としている。そのため、実用上は主観的長さにあるしきい値を設け、推定した主観的長さがしきい値以上であれば支援を行い、しきい値以下であれば支援を行わないとする。こ

|                          |  |  |  |                         |  |  |  |
|--------------------------|--|--|--|-------------------------|--|--|--|
| 真値<br>10.0<br>推定値<br>7.3 |  |  |  | 真値<br>4.0<br>推定値<br>4.3 |  |  |  |
| 真値<br>4.0<br>推定値<br>3.7  |  |  |  | 真値<br>6.0<br>推定値<br>5.6 |  |  |  |

図 6: 主観的長さ推定の例

の際、一般的に「かなりきつい」と感じる場合に支援が求められると考えられるため、しきい値の案として表 1 で示した主観的長さ評価指標において「Very Hard」に対応する 7 を用いることが考えられる。実用の際、最低限でも New Borg Scale において 10 に値する「Very, very Hard」と感じている人に対しては必ず支援をするべきであるため、しきい値が 7 の場合はすべての推定値の誤差が 3 以内になることが望ましい。表 2 に示したように、提案手法を用いることで許容誤差 3 の時の許容誤差精度は 87.51% であった。このことから、100% には満たないものの、87% 程度の精度で誤差 3 以内で推定できていることから、ある程度の有効性があるといえる。

また提案手法では、歩き方に影響を与える人の体重と手荷物種類を考慮するために、人の体重推定と手荷物種類認識をサブタスクとして主観的長さ推定と同時に学習した。表 2 より、これらの精度が向上したことから、提案手法におけるサブタスクの学習が有効であるといえる。

さらに、固定長の系列であっても歩き方の特徴として重要である関節点の速度を考慮した推定を行うために、3D 人物骨格系列に速度情報を付与した骨格座標・骨格速度系列を利用した。結果として推定精度が向上したことから、提案手法において骨格座標・骨格速度系列が有効であるといえる。

最後に、今回の研究に使用した歩行時の 3D 人物骨格系列は Azure Kinect SDK を用いて推定したが、推定に失敗する場面が見られた。これは、被験者が大きめの服を着用していたり、黒

表 2: 主観的長さ推定結果

| 体重推定 | 手荷物種類認識 | 骨格座標・骨格速度系列 | MAE↓        | TA <sub>1</sub> ↑ | TA <sub>2</sub> ↑ | TA <sub>3</sub> ↑ |
|------|---------|-------------|-------------|-------------------|-------------------|-------------------|
| ✓    |         | ✓           | 1.43        | 46.59%            | 71.56%            | 86.85%            |
|      | ✓       | ✓           | 1.43        | 46.22%            | 71.46%            | 86.94%            |
|      |         | ✓           | 1.43        | 46.18%            | 71.54%            | 87.26%            |
| ✓    | ✓       |             | 1.46        | 45.80%            | 70.72%            | 86.15%            |
| ✓    | ✓       | ✓           | <b>1.41</b> | <b>46.83%</b>     | <b>72.30%</b>     | <b>87.51%</b>     |

いマスクを装着していた場合などに多く見られた。実用の際は、このような推定に失敗した 3D 人物骨格系列を使用しないようにするシステムが必要であると考えられる。

## 5. むすび

自主的に人を支援するロボットが求められており、ロボットが行う支援の一つとして手荷物を重そうに持っている人を自主的に支援することが挙げられる。この支援を行うか否かの判断材料として、手荷物をどのくらい重く感じているかを表す主観的長さの推定が必要である。本研究では、手荷物の長さが人の歩行時の様子に影響を与えることに着目し、歩行の様子から手荷物の主観的長さを推定する手法を提案した。歩行時の姿勢は、人の体格や手荷物の種類によっても変化するため、主観的長さの推定の学習と同時に人の体重推定と手荷物種類の認識もサブタスクとして学習することで、これらを考慮して主観的長さを推定した。このとき、入力として 3D 人物骨格系列に加え、人物シルエット画像系列も用いることで体格の情報を考慮した。また、関節点の速度も主観的長さ推定において重要な特徴であるが、歩行の様子を表現する 3D 人物骨格系列をサンプリングにより固定長化して入力する一般的な GCN ベースの手法を利用すると関節点の速度の情報が失われてしまう。そのため、3D 人物骨格系列に対して各関節点の速度を付加した特徴量を用いることで、速度の情報を保ったまま固定長化した骨格座標・骨格速度系列を用いて推定した。

今後の課題として、混雑した状況下への対応、複数の手荷物を所持する人への対応、一人から複数の歩行周期が得られた場合の推定結果統合方法の考案が挙げられる。

## 謝 辞

本研究の一部は JSPS 科研費 (17H00745) による。

## 文 献

- [1] Amazon Staff, “Amazon Device & Services news—September 2021”, Amazon, <https://www.aboutamazon.com/news/devices/amazon-devices-services-news-september-2021>.
- [2] 湯口 彰重, 河野 誠也, 石井 カルロス 寿憲, 吉野 幸一郎, 川西 康友, 中村 泰, 港隆 史, 齊藤 康己, 美濃 導彦, “ぶつくさ君: 自身の外界認識と内部状態を言語化するロボット”, 第 39 回日本ロボット学会学術講演会予稿集, Sep. 2021.
- [3] Piaggio Fast Forward, “gita robots,” <https://mygita.com/>.
- [4] Boston Dynamics, “Atlas,” <https://www.bostondynamics.com/atlas>.
- [5] Y. Hong, J. Li, “Influence of load and carrying methods on gait phase and ground reactions in children’s stair walking,” *Gait & Posture*, vol.22, issue.1, pp.63–68, Aug. 2005.
- [6] 加藤 大貴, 平山 高嗣, 道満 恵介, 井手 一郎, 川西 康友, 出口 大輔, 村瀬 洋, “音象徴性を利用したオノマトペによる歩容の記述,” *人工知能学会論文誌*, vol.33, no.4, pp.B-HC2.1–9, Mar. 2018.
- [7] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, Y. Yagi, “GEINet: View-invariant gait recognition using a convolutional neural network,” *Proc. of the 2016 International Conference on Biometrics*, pp.1–8, Jun. 2016.
- [8] H. Chao, Y. He, J. Zhang, J. Feng, “GaitSet: Regarding gait as a set for cross-view gait recognition,” *Proc. of the AAAI Conference on Artificial Intelligence*, pp.8126–8133, Jun. 2019.
- [9] J. Liu, A. Shahroudy, D. Xu, G. Wang, “Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition,” *Computer Vision - ECCV 2016, Part.I*, pp.816–833, Oct. 2016.
- [10] S. Yan, Y. Xiong, and D. Lin, “Spatial temporal graph convolutional networks for skeleton based action recognition,” *Proc. of the 32nd AAAI Conference on Artificial Intelligence*, pp.7444–7452, Feb. 2018.
- [11] Z. Liu, H. Zhang, Z. Chen, Z. Wang, W. Ouyang, “Disentangling and Unifying Graph Convolutions for Skeleton-Based Action Recognition,” *Proc. of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.143–152, June. 2020.
- [12] K. Shiraki, T. Hirakawa, T. Yamashita, H. Fujiyoshi, “Spatial Temporal Attention Graph Convolutional Networks with Mechanics-Stream for Skeleton-based Action Recognition” *Computer Vision - ACCV 2020, Part.V*, pp.341–357, Nov. 2020.
- [13] 山口 優太, 神谷 卓也, 西山 正志, 岩井 儀雄, “身体動揺を用いた重量物所持の認識可能性の検証”, *ビジョン技術の実用ワークショップ講演論文集*, no.IS2-A1, pp.1–5, Dec. 2019.
- [14] 仲田 仁, 田村 仁, “持ち上げ動作からの所持物体の重量推定”, 第 80 回全国大会講演論文集, vol.1, pp.435–436, Mar. 2018.
- [15] Z. Zhang, P. Cui, W. Zhu, “Deep learning on graphs: A survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol.34, pp.249–270, Mar. 2020.
- [16] T. Teepe, J. Gilg, F. Herzog, S. Hormann, G. Rigoll, “Towards a Deeper Understanding of Skeleton-Based Gait Recognition,” *Proc. of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.1569–1577, Jun. 2022.
- [17] B. Gunnar, “Psychophysical Bases of Perceived Exertion,” *Medicine & Science in Sports & Exercise*, Vol.14, No.5, pp.377–381, May. 1982.
- [18] A. L. Maas, A. Hannun, A. Y. Ng, “Rectifier Nonlinearities Improve Neural Network Acoustic Models,” *Proc. of the 30th International Conference on Machine Learning Workshop*, pp.1–6, Jun. 2013.
- [19] K. He, G. Gkioxari, P. Dollár, R. Girshick, “Mask R-CNN,” *Proc. of the 2017 IEEE International Conference on Computer Vision*, pp.2961–2969, Oct. 2017.