

身体動揺の時系列深度画像における時間手掛りを用いた手荷物の軽重認識

山口 優太 **, 中山 晴貴 **, 神谷 卓也 **, 西山 正志 **, 岩井 儀雄 **, 櫛田 大輔 **

Baggage weight classification by extracting temporal features of body sway acquired from depth image sequences

Yuta YAMAGUCHI, Haruki NAKAYAMA, Takuya KAMITANI, Masashi NISHIYAMA, Yoshio IWAI and Daisuke KUSHIDA

We propose a method for classifying the weight of baggage carried by a person in an upright posture by finding temporal cues of body sway from depth image sequences. When a standing person is viewed from an overhead depth camera, body sway, which is a slight movement that naturally occurred in the human body, is observed. We consider body sway as discriminative cues for baggage weight classification because it varies depending on the weight of baggage carried by a standing person. To find the cues of body sway from depth image sequences, we can use the existing feature extraction¹⁾. However, the accuracy of baggage weight classification is reduced if the existing feature extraction is simply performed. The existing feature extraction causes this problem by seeing both the motion and the shape representing each person's cue to identify people. We consider that the shape of a person does not change even if the weight of the baggage changes. To this end, we design a novel feature extraction that suppresses spatial cues of the shape of a person and emphasizes temporal cues of the motion using the head region's center position. The experimental results show that our feature extraction improves baggage weight classification accuracy compared to the existing feature extraction.

Key words: Baggage weight classification, body sway, depth camera, feature extraction, temporal cue.

1. はじめに

カメラから撮影された画像の中に存在する人物を対象とし、様々な属性を認識する技術の開発が進んでいる。人物の属性として、性別、年代、服装を認識する手法^{2,3,4)}が提案されている。近年では、手荷物を所持しているかどうかの属性を判別する手法^{5,6)}が注目されつつある。ここで述べる手荷物とは、人物が片手で所持するケースや袋を指し、その中に物体が入っているとす。手荷物の所持の有無だけでなく、さらに詳細な人物の属性を収集するため、所持する手荷物の重さを認識する手法⁷⁾が提案されている。この既存手法では、歩く人物を対象とし、手荷物の重さを、時系列の画像から認識している。ただし、人物は常に歩いている訳ではない。例えば、エレベータの到着を待つ時や、ショッピングの清算を待つ時に、人物は歩くことを止め、その場に立つ。

本論文では、立ち止まる人物を対象とし、片手で所持する手荷物の重さを、時系列の画像から認識する手法について議論する。この手荷物の軽重認識を実現できれば、様々な応用に繋がると考えられる。例えば、エレベータ前で、不自然に重い手荷物を所持し、立ち止まっている不審人物を検出するセキュリティ応用が考えられる。他にも、ショッピングモールのレジ前で、重い手荷物を持ち、並んでいる人物を手助けする代行応用が考えられる。

手荷物の軽重認識のために、立ち止まる人物から、手がかりとして何が捉えられるかについて考える。立ち止まる人物は、身体の動きを完全に静止させることはできず、前後左右の様々な方向に微小に揺れ動いている。この身体の微小な動きは、身体動揺⁸⁾と呼ばれ、人物が意識的に身体を静止しようとして

も自然に発生する。本論文では、手荷物の重さを認識するために、身体動揺を表す手掛りを、時系列の画像から捉えることを狙う。なお、人物は、直立姿勢を維持し、片手で手荷物を所持するとし、深度カメラから撮影される条件で議論を進める。立ち止まる人物を撮影した場合、時系列の深度画像から、身体動揺を表す空間手掛りと時間手掛りとを捉えることができる。空間手掛りとは、ある時刻における人物の形状である。時間手掛りとは、人物の形状が、時間が経過するとともに、連続的に微小移動する様子である。

身体動揺を表す空間手掛りと時間手掛りについて、手荷物の軽重認識にどう有効であるかを考える。まずその前に、身体動揺に関する既存の特徴抽出¹⁾の考え方について述べる。この特徴抽出は、人物対応付けを目的とし、身体動揺を表す空間手掛りと時間手掛りの両方を捉える設計となっている。人物対応付けにおける空間手掛りでは、形状について、人物間の差を捉えている。また時間手掛りでは、形状の時間方向の微小移動について、人物間の差を捉えている。既存の特徴抽出を、そのまま軽重認識に用いる場合、荷物の重さに関する差を考慮していないため、高い精度を期待できないと考えられる。特に、人物対応付けでは重要であった空間手掛りは、手荷物の軽重認識において、有用な特徴ではない可能性があると考えられる。その理由として、手荷物の重さが変化したとしても、個人ごとの形状は変わらないことが挙げられる。一方、時間手掛りでは、手荷物の軽重認識において重要であると考えられる。その理由として、手荷物の重さに応じて身体の3次元的な重心位置が変わるため、微小移動の様子に変化が生じるからと考えられる。手荷物の軽重認識の精度を高めるためには、時系列の深度画像から捉えられる空間手掛りを抑止する必要がある、時間手掛りのみを捉える必要があると想定される。

そこで本論文では、片手で手荷物を所持し立ち止まる人物を対象とし、時系列の深度画像から、身体動揺を表す時間手掛り

* 原稿受付 令和3年5月11日

* 掲載決定 令和3年9月15日

** 鳥取大学大学院工学研究科(鳥取市湖山町南4丁目101)

を捉える特徴抽出を行い、手荷物の重さを認識する手法について述べる。実験結果より、提案手法は、空間手掛りと時間手掛りの両方を捉える既存の特徴抽出¹⁾を用いた認識手法と比べて、軽重認識の精度を向上させることを確認した。また、深層学習を用いる手法^{9, 10, 11)}、および、時系列画像に対する特徴抽出^{12, 13, 14)}と識別器とを組み合わせる手法と比べても、軽重認識の精度を高めることを確認した。本論文は、我々の先行研究¹⁵⁾を発展させたものであり、時間手掛りを捉える特徴抽出を新たに開発し、実験および考察を拡充させたものである。以下では、2.で関連研究を述べ、3.で提案手法について述べ、4.で手荷物の軽重認識の精度について述べる。最後に5.でまとめる。

2. 関連研究

2.1 手荷物の重さに関する知見

身体運動学の分野において、手荷物の重さに関する分析研究^{16, 17, 18)}が実施されている。これらの文献では、歩行中の人物を対象とし、その人物が所持する手荷物の重さが変化した場合、動きに差異が生じるかどうかをモーションキャプチャを用いて分析している。歩行する人物が手荷物を片手で所持する時、もう一方の手の動きが、手荷物の重さに応じて大きくなっていくことや、骨盤の動きが制限されていくことが報告されている。児童や女性など特定の人物カテゴリに属する歩行者について、手荷物の重さに関する分析^{19, 20)}が実施されており、同様の結果が報告されている。

理学療法分野では、立ち止まる人物の身体動揺が、手荷物の重さによって、どのように変化するかについて、圧力センサを用いて分析した研究が実施されている。文献²¹⁾では、本人の背筋力限界に合わせた重い手荷物を保持した場合、身体動揺の高周波成分に変化が生じることが報告されている。具体的な周波数帯は明記されていないものの、背筋力限界の50%の負荷により、人間の姿勢制御が大きく変化し、高周波帯の揺れが増加すると述べられている。人間工学の分野では、人物が所持する手荷物の重さと、人物が履いている靴の違いにより、身体動揺が変化することが、文献²²⁾で報告されている。ただし、立ち止まる人物が所持する手荷物の重さを、自動で認識する手法については、これらの文献では議論されていなかった。このため、本論文では、手荷物の軽重認識の手法を構築し、その精度を評価する。

2.2 カメラを用いた身体動揺の観測

近年、非接触センサであるカメラを用いて、人物の時系列画像から身体動揺を観測する試みが行われている。文献²³⁾では、転倒防止のリスクアセスメントを目的とし、複数カメラを用いて身体の重心位置を推定し、その時間方向の微小移動を計測している。文献^{24, 25)}では、身体バランスの臨床調査を目的とし、深度カメラを用いて身体の関節位置を推定し、その時間方向の微小移動を計測している。文献²⁶⁾では、ヒューマンインタフェースに用いるアバタ映像の自動生成を目的とし、正面に取り付けたRGBカメラから撮影された映像を用いて、揺れの中心を推定している。これらの先行研究の成果からも分かるように、立ち止まる人物から、身体動揺を、カメラを用いて非接触で観測することは可能であると言える。本論文では、カメラから身体動揺を観測した後の新たな応用の一つとして、手荷物の重さを、身体動揺の時系列画像から認識する手法の開発に取り組む。

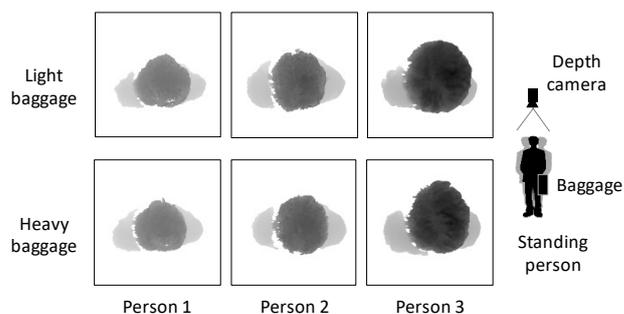


Fig. 1 Examples of depth images of standing people carrying baggage acquired from the overhead depth camera. Black color represents near regions from the camera, and white color represents far regions.

3. 身体動揺の時系列画像から手荷物の重さを認識する手法

3.1 概要

片手で手荷物を所持し立ち止まる人物から、身体動揺を観測するために、深度カメラをどこに設置すればよいかについて考える。このためにまず、身体動揺の時間手掛りである形状の微小移動に着目し、その移動量の大きさについて考える。観測される移動量の大きさは、頭や足などの身体部位によって異なる。具体的には、床と接触している足に近づくほど移動量は小さくなり、頭に近づくほど移動量は大きくなる。本論文では、頭の移動量を安定に捉えるため、人物の頭上に設置された深度カメラを採用する。このカメラの設置条件には、視野内に複数の人物が立つ場合において、相互遮蔽の影響を受けない副次的な利点もある。頭上に設置された深度カメラを用いることで、手荷物を持つ人物を撮影した場合の例を Fig. 1 に示す。画像中の主な面積を人物の頭が占めており、左右の肩の一部が見えていることが分かる。

深度カメラから観測された人物の頭を、計算機上でどのように表すかについて議論する。一般的に深度カメラからは、3次元の点群を得ることができる。点群の深度値を用いることで、画像中の頭を簡単に見つけられる利点がある。ただし、頭の表面から得られた点群には、毛髪の影響により、深度値にノイズが含まれることが多い。身体動揺の移動量は非常に小さいため、点群をそのまま用いた場合、その移動量を安定に捉えることは難しいと考えられる。そこで本論文では、頭の形状を、3次元の点群で表すのではなく、2次元のシルエットで表すことで、深度値のノイズの影響を抑えることを考える。各時刻の深度画像において、頭領域のみを指すシルエット画像を生成した例を Fig. 2 に示す。なお、シルエット画像の生成の詳細は3.2で述べる。

頭領域のシルエット画像を対象とする場合、見かけの大きさが、人物間で変化することを考慮しなければならない。先に示した Fig. 2 では、身長異なる3名の人物を撮影対象としている。図中の左側に、身長が低い人物のシルエット画像を示し、右側に、身長が高い人物のシルエット画像を示している。また図中の上側では軽い手荷物を持つ場合、下側では重い手荷物を持つ場合を撮影条件としている。頭領域の見かけの大きさは、人物間の身長差による影響を大きく受けており、手荷物の重さによる影響をほぼ受けないことが分かる。よって身長差の影響は、手荷物の軽重認識において有用ではないと考えられる。そこで本論文では、頭領域の見かけの大きさについて、人物間の

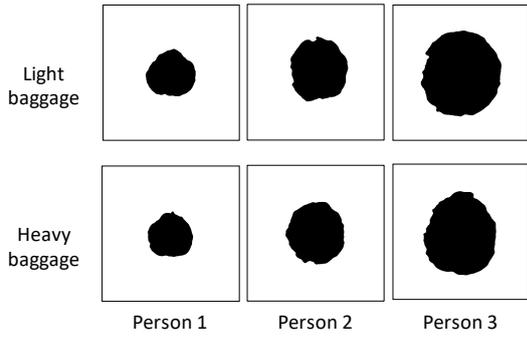


Fig. 2 Examples of silhouette images of head regions generated from the depth images. On the left is a silhouette image of a short person, and on the right is a silhouette image of a tall person.

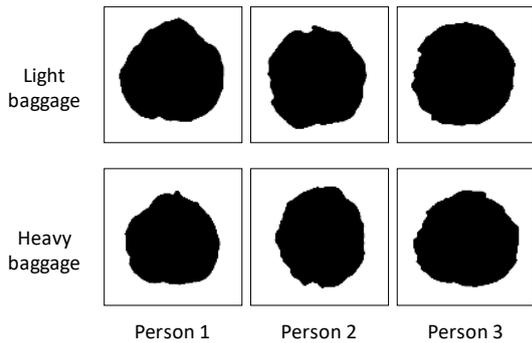


Fig. 3 Examples of silhouette images with the normalization of the apparent size of the head regions.

身長差の影響を抑えるため、シルエット画像に対して頭領域の正規化を適用する。正規化が適用された頭領域のシルエット画像の例を Fig. 3 に示す。なお、正規化の詳細は3.3で述べる。

正規化された頭領域のシルエット画像から、身体動揺の時間手掛りを表す移動量を、どのように捉えるかについて議論する。ここでは、時系列の複数枚のシルエット画像が、同一人物から獲得されているとする。まず既存の特徴抽出¹⁾のように、シルエット画像間で形状差分を求めることが考えられる。ただし、先に述べた見かけの大きさの正規化を適用したとしても、頭領域のシルエット画像には、人物毎の輪郭形状の影響が含まれている。このため、シルエット画像間の形状差分を各時刻で求めた場合、空間手掛りである形状の影響を抑えることはできない。そこで本論文では、形状差分ではなく、頭領域の中心位置の差分を用いることで、時間手掛りを表す特徴量を抽出することを考える。各時刻で獲得された頭領域の中心位置と、全時刻から平均化された頭領域の中心位置との間で距離を算出し、その距離の時系列変化を表す信号を周波数解析する。なお、特徴抽出の詳細は3.4で述べる。

身体動揺を用いて手荷物の軽重認識を行う提案手法の流れを Fig. 4 に示す。ここまで述べてきた特徴抽出を行い、得られた特徴量を識別器に入力し、手荷物の重さを認識する。識別処理については3.5で詳細を述べる。次節より、提案手法の具体的な処理について説明していく。

3.2 頭領域のシルエット画像の生成

獲得された深度画像から、頭領域のシルエット画像を生成する処理の詳細について述べる。まず、深度画像中の各画素に対

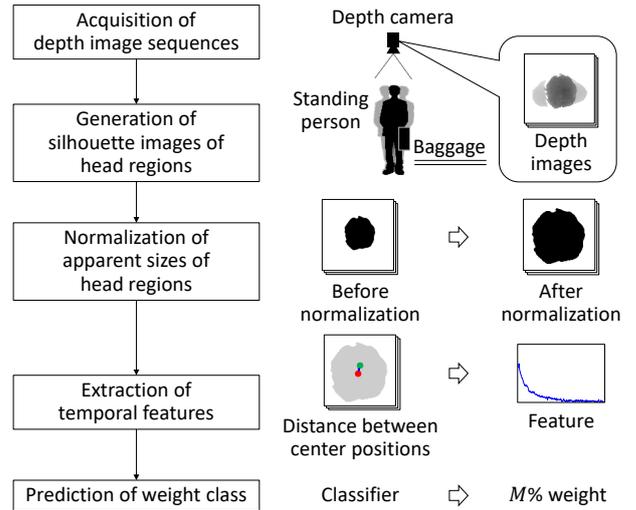


Fig. 4 Overview of our method for baggage weight classification.

応する3次元点について、人物領域または背景領域に属するかを判別する。背景である床面は、撮影中に動かないと仮定できるため、3次元点の深度値に対する固定閾値を設けることで除去する。次に、身長に基づく深度値の閾値を用いて、人物領域の中から、頭領域の画素のみを判別する。人物毎に身長は異なるため、深度値の閾値を適応的に設定する必要がある。このため本論文では、人物領域における深度値ヒストグラムを用いる。頭上に深度カメラを設置した場合、頭はカメラに近くなり、肩はカメラから遠くなるため、深度値ヒストグラムは二つのピークを持つと想定される。観測された深度値ヒストグラムの例を Fig. 5 に示す。図中 (a) では身長が低い人物の深度値ヒストグラムの例を、図中 (b) では身長が高い人物の深度値ヒストグラムの例を表す。人物の身長が異なると、深度値ヒストグラムはそれぞれ異なることが分かる。各ヒストグラムに応じて閾値を適応的に設定するため、本論文では大津の手法²⁷⁾を適用する。適応的に設定された閾値よりも、深度値が小さい画素の集合を頭領域とする。最終的に提案手法では、頭に対応する画素値を1、それ以外を0とするシルエット画像を生成する。

適応的に閾値を設定する利点について述べる。髪型の影響により、身長が変化したとしても、閾値を安定に決めることができる点が挙げられる。また、距離センサで発生する奥行方向のノイズに対して、閾値を頑健に決めることができる点が挙げられる。

3.3 頭領域の見かけの大きさの正規化

頭領域の見かけの大きさの影響を抑止するため、シルエット画像に対し空間方向の正規化を適用する。この正規化のために、文献²⁸⁾で述べられている手法を用いる。以下、その手法の流れを説明する。生成されたシルエット画像において、頭領域を囲むように、縦幅が H' 、横幅が W' の範囲を設定する。その際、各時刻において、頭領域が範囲外にはみ出さないように、マージンを設ける。このマージンは、時系列に生成されたシルエット画像の中で、頭領域の移動量の最大値を基に決定される。なお W' と H' は、それぞれの時系列シルエット画像で設定され、一つの時系列シルエット画像の中では固定の値とする。設定された範囲を、縦幅が H 、横幅が W の画像へ変換する。正規化前のシルエット画像の画素位置を \mathbf{x}' で表し、正規

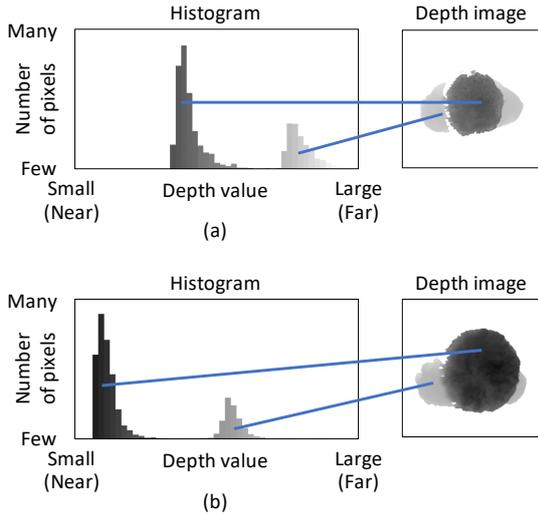


Fig. 5 Examples of the histograms of the depth values in the whole body regions of people.

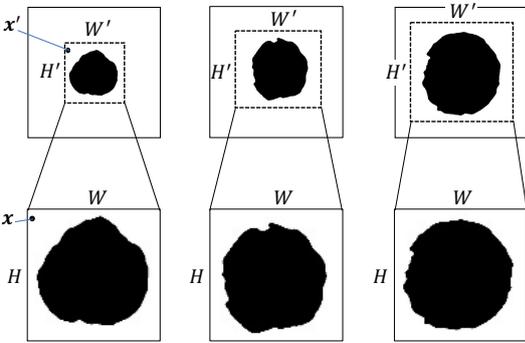


Fig. 6 Examples of the normalization of the apparent size of the head regions in the silhouette images.

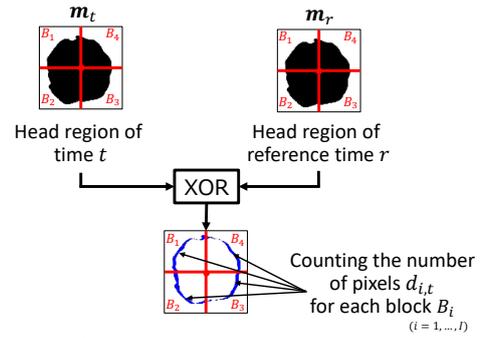
化後のシルエット画像の画素位置を \boldsymbol{x} で表すと、この画像変換は $\boldsymbol{x} = \boldsymbol{S}\boldsymbol{x}'$ となる。なお、 $\boldsymbol{S} = \text{diag}(W/W', H/H')$ とする。シルエット画像において、頭領域の見かけの大きさが正規化される例を Fig. 6 に示す。

3.4 時間手掛りを捉えるための特徴抽出

手荷物の軽重認識で用いる時間手掛りを、時系列に獲得されたシルエット画像から、特徴量として抽出する処理の詳細を説明する。提案する特徴抽出を説明する前に、既存の特徴抽出¹⁾を先に説明する。この既存の特徴抽出では、Local movement (LM) 特徴量が算出される。LM 特徴量は、1.でも述べたように、人物対応付けを目的とし、身体動揺を表す空間手掛りと時間手掛りとの両方を捉える設計となっている。空間手掛りを捉えるために、LM 特徴量では、頭領域を、放射状に I 個のブロック B_i ($i = 1, \dots, I$) に分割する。次に、時間手掛りを捉えるために、各ブロックにおいて、式 (1) を用いて移動量を算出する。

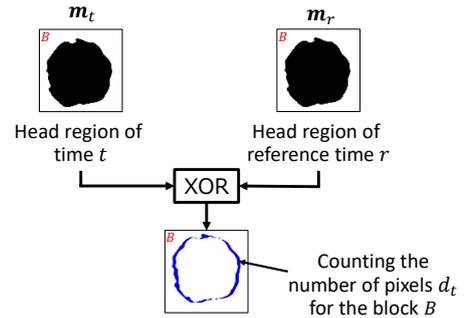
$$d_{i,t} = \sum_{\boldsymbol{x} \in B_i} |m_t(\boldsymbol{x}) - m_r(\boldsymbol{x})| \quad (1)$$

ここで、 \boldsymbol{x} は各ブロック中の画素位置を表し、 $||$ は絶対値を表す。 $m_t(\boldsymbol{x})$ は、時刻 t のシルエット画像 \boldsymbol{m}_t における位置 \boldsymbol{x} での画素値を表す。 $m_r(\boldsymbol{x})$ は、参照時刻のシルエット画像 \boldsymbol{m}_r における位置 \boldsymbol{x} での画素値を表す。なお参照時刻 r は、



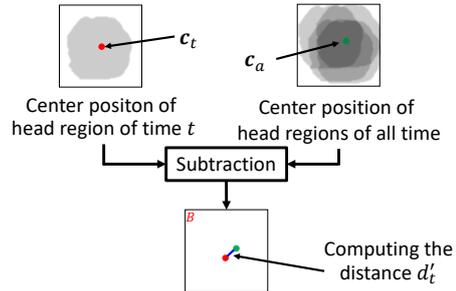
Distance between head regions

(a) LM feature



Distance between head regions

(b) ILM feature



Distance between center positions

(c) CM feature

Fig. 7 Computation of temporal changes from silhouette images using the existing methods and our method.

全時刻のシルエット画像と最も類似するシルエット画像が獲得された時刻とする。LM 特徴量における移動量の算出の概要を Fig. 7(a) に示す。各時刻 t ($t = 1, \dots, T$) において、移動量 $d_{i,t}$ を算出し時系列信号を得る。最後に、移動量の時系列信号から Power spectral density (PSD) を求め LM 特徴量を抽出する。PSD を Welch 法で求めるため、時系列信号を短時間セグメントにオーバーラップありで分割し、各セグメントで離散フーリエ変換を適用した後、それらの出力をセグメント間で平均する。これにより、各周波数の対数パワースペクトル密度を安定に推定できる効果がある。

既存の LM 特徴量では、人物対応付けを目的としているため、シルエット画像をブロックに分割することで、人物間の形状差を空間手掛りとして捉えていた。ただし、手荷物の軽重認識において、空間手掛りである人物間の形状差は、精度

低下を招く要因になると考えられる。そこで、既存の LM 特徴量に対して、軽重認識を目的とした改良を加えることを考える。ここでは、ブロック分割を適用せず、シルエット画像の全体を表すブロック B から移動量を算出する。この特徴量を Improved-LM (ILM) 特徴量と呼ぶ。ILM 特徴量では、身体動揺から観測される移動量を、式 (2) を用いて算出する。

$$d_t = \sum_{\mathbf{x} \in B} |m_t(\mathbf{x}) - m_r(\mathbf{x})| \quad (2)$$

ILM 特徴量における移動量の算出の概要を Fig. 7(b) に示す。各時刻 t において移動量 d_t を算出し、時系列信号を得る。最後に、移動量の時系列信号から PSD を求めることで ILM 特徴量を抽出する。

ここまで述べてきた ILM 特徴量では、移動量を算出する際、シルエット画像間における画素値の差分を用いていた。シルエット画像には、3.1 で述べたように頭領域の輪郭形状が残るため、ILM 特徴量は、空間手掛りである形状の影響をなお受けていると考えられる。そこで本論文では、頭領域そのものではなく、頭領域の中心位置を用いることで、空間手掛りの影響をさらに抑えることを狙う。この中心位置を用いる特徴量を Center movement (CM) 特徴量と呼ぶ。CM 特徴量では、身体動揺から観測される移動量を、式 (3) を用いて算出する。

$$d'_t = \|\mathbf{c}_t - \mathbf{c}_a\|_2 \quad (3)$$

ここで、 $\|\cdot\|_2$ は L2 ノルムを表す。 \mathbf{c}_t は、各時刻 t において $m_t(\mathbf{x}) = 1$ を満たす全ての画素位置 $\mathbf{x} \in B$ から求めた平均位置を表す。 \mathbf{c}_a は、全時刻において $m_t(\mathbf{x}) = 1$ を満たす全ての $\mathbf{x} \in B$ から求めた平均位置を表す。CM 特徴量における移動量の算出の概要を Fig. 7(c) に示す。各時刻 t において移動量 d'_t を算出することで、時系列信号を得る。最後に、移動量の時系列信号から PSD を求めることで CM 特徴量を抽出する。

3.5 識別処理

本論文では、ここまで述べてきた特徴量を、識別器の入力とすることで、軽重認識を行う。手荷物の重さクラスを設定し、その分類問題として解くことを考える。以下では、手荷物の重さを、どのようにクラスとして設定するかについて述べる。単純に考えた場合、計量された重さの値そのものを離散化することで、重さクラスを設定することが考えられる。ただし、人物間で筋肉量が異なるため、所持できる限界の手荷物の重さが異なる。よって本論文では、人物の体重に対する手荷物の重さの割合を指標として、重さクラスを設定する。この体重に対する割合を用いる指標は、厚生労働省通達²⁹⁾や、理学療法分野の分析研究^{22,30)}で用いられている。手荷物を所持していない時の体重を計測し、その $M\%$ の重量を、手荷物の重さとする。離散的に M を設定することで、手荷物の重さクラスを設定する。なお分類問題を解くために様々な識別器が存在しているが、本論文では、線形 Support vector machine (SVM)³¹⁾、Random forests (RF)³²⁾、および、Gradient boosted decision trees (GBDT)³³⁾を用いる。提案する CM 特徴量と、相性のよい識別器を実験的に定める。

4. 実験

4.1 評価データセット

提案手法の有効性を確認するため、片手で手荷物を所持し立ち止まる人物について、時系列の深度画像を撮影した。手荷物

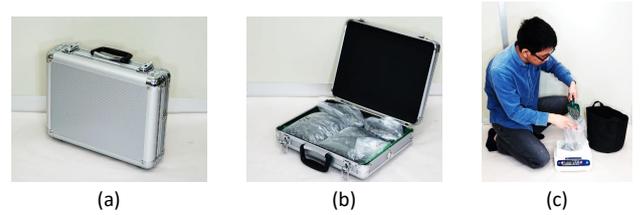


Fig. 8 Baggage (attache case) carried by the participants in our experiments.

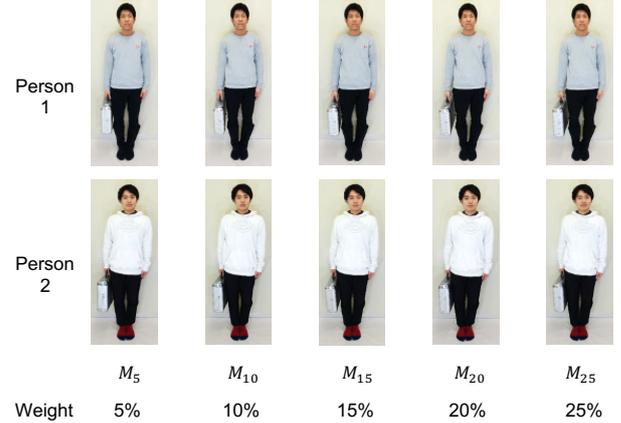


Fig. 9 Examples of the participants carrying an attache case when changing the weight class of baggage.

の軽重認識に用いる重さクラスとして、3.5 で述べたパラメータ M を変化させた以下の 5 個を設定した。

- M_5 : 各人物の体重 5% の重さ
- M_{10} : 各人物の体重 10% の重さ
- M_{15} : 各人物の体重 15% の重さ
- M_{20} : 各人物の体重 20% の重さ
- M_{25} : 各人物の体重 25% の重さ

重さクラス M_{10} では、例えば体重 66 kg の人物は 6.6 kg の手荷物を所持し、体重 54 kg の人物は 5.4 kg の手荷物を所持した。なお、体重 30% の重さとした場合、一部の人物は手荷物を所持し続けることができなかった。

本論文では、手荷物として、Fig. 8(a) のアタッシュケースを準備した。そのケースの中に物体を入れ、手荷物の重さを調整した。具体的には、大磯砂利を袋詰めにした物体で調整した。アタッシュケースに重りを詰めた例を Fig. 8(b) に示す。重りが内部で移動しないよう、ケース中を緩衝材で固定した。ケースに入れる物体の重さを調整する様子を Fig. 8(c) に示す。重さを調整した手荷物を人物が片手で持った場合の例を Fig. 9 に示す。手荷物の重さクラスが変わったとしても、同じ人物であれば見え方が変わらないことが分かる。

実験協力者は 15 名 (男性 15 名、平均身長 168.4 ± 3.7 cm、平均体重 60.6 ± 6.7 kg、平均年齢 22.6 ± 1.5 歳) とした。なお体重を計測するために、撮影直前に体重計 (タニタ体組成計 BC-314) を用いて測定した。実験協力者が所持した手荷物の重さの平均を Table 1 に示す。実験協力者が立つ位置と深度カメラの設置位置との関係を Fig. 10 に示す。深度カメラとして Microsoft 社の Kinect v2 を用いた。カメラ光軸が床面と垂直となるようにキャリブレーションを行った。深度画像のサイズを

Table 1 Average weight of baggage carried by the participants.

Weight class label	Percentage of body weight	Baggage weight
M_5	5%	3.0 ± 0.3 kg
M_{10}	10%	6.1 ± 0.7 kg
M_{15}	15%	9.1 ± 1.0 kg
M_{20}	20%	12.1 ± 1.3 kg
M_{25}	25%	15.2 ± 1.7 kg

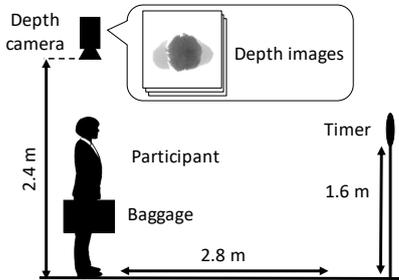


Fig. 10 Experimental setting of the depth camera and the participant.

512×424 画素とし、サンプリング周波数を 30 Hz とした。カメラから床面までの距離を 2.4 m とした。撮影する前に、床面のマーカに土踏まずを合わせるよう実験協力者へ指示した。また、手荷物を右手で持つように指示した。その際、手荷物が脚に密着しないこととした。各人物について、時系列の深度画像を撮影する際の時間長を 160 秒とした。撮影開始から終了まで、直立姿勢（ロンバルク肢位）を維持することとし、正面に設置された時計を見ることとした。時計の配置位置は、実験協力者から 2.8 m 先とした。時計の下端が、床から 1.6 m の高さとなるよう設置した。時計は丸形で、その直径は 33 cm であった。手荷物を所持した時の疲労の影響を少なくするため、軽い手荷物から順番に撮影を行った。各重さクラス M_5 , M_{10} , M_{15} , M_{20} , M_{25} において、各実験協力者の時系列画像を 1 回撮影した。時系列の深度画像の合計個数を $15 \times 5 = 75$ 個とした。

4.2 提案手法における CM 特徴量の効果の確認

4.2.1 実験条件

本論文にて新しく提案した CM 特徴量が、軽重認識において有効であるかどうかを評価した。比較する特徴量として 3.4 で述べた LM 特徴量、ILM 特徴量、および、CM 特徴量を用いた。特徴量の有効性は、どの識別器を用いるかで変わると考えられるため、特徴量と識別器との組み合わせを変えた場合の認識精度を比較した。識別器として 3.5 で述べた SVM, RF, および、GBDT を用いた。なお、頭領域の見かけの大きさの正規化の効果については 4.3 で評価し、シルエット画像として頭領域を用いた場合の効果については 4.4 で評価する。

精度評価にあたり、設定した特徴抽出のパラメータを列挙する。3.3 で述べた見かけの大きさの正規化において、縦幅を $H = 100$ 画素、横幅を $W = 100$ 画素とした。LM 特徴量に用いるブロック分割数は、文献¹⁾で述べられていた 25 を用いた。次に、識別器のパラメータを列挙する。線形 SVM では正規化項の係数を 0.1, 1.0 または 10.0 とし、RF では森を構成する決定木の数を 100, 200 または 300 とし、GBDT では弱識別器の木の本数を 100, 200 または 300 とした。特に RF と GBDT はランダムサンプリングによる木の選び方の影響が大きいので、同じパラメータでの試行回数を 10 回とした。

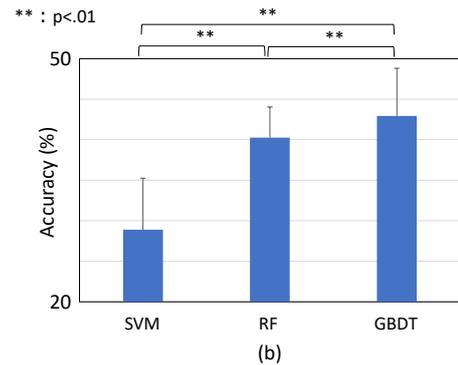
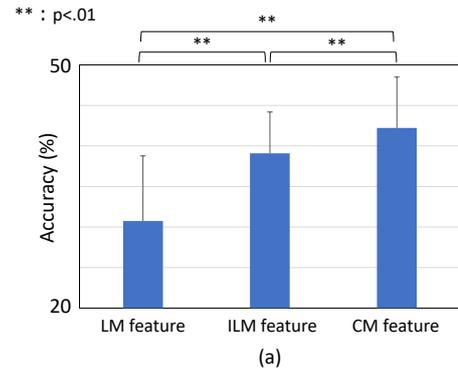


Fig. 11 Accuracy of baggage weight classification. In (a), we evaluated the accuracy when using the LM, ILM, and CM features. In (b), we evaluated it when using the SVM, RF, and GBDT classifiers.

4.2.2 実験結果

手荷物の軽重認識についての実験結果を述べる。認識精度の評価指標として、重さクラスが正しく出力されたかどうかを表す 1 位正解率を用いた。精度を求める際、Leave-one-participant-out を適用した。具体的には、実験協力者 1 名から獲得した頭領域の時系列シルエット画像をテストに用い、他 14 名の時系列シルエット画像を識別器の訓練に用いた。この訓練とテストとを、実験協力者の人数分だけ繰り返した。

Fig. 11(a) にて、LM 特徴量、ILM 特徴量、または、CM 特徴量と、識別器とを組み合わせた時の認識精度を示す。図中の棒グラフでは、特徴量を固定し、識別器とそのパラメータとの全組み合わせで試行した時の平均精度を用いた。また、棒グラフに付属するバーは標準偏差を表す。LM 特徴量を用いた場合は $30.7 \pm 8.0\%$ 、ILM 特徴量を用いた場合は $39.1 \pm 5.1\%$ 、CM 特徴量を用いた場合は $42.2 \pm 6.3\%$ であった。認識精度の間に有意差があるかどうかを確認するため、ウィルコクソンの符号付順位検定、および、ボンフェローニ補正を適用した。有意水準は $p < 0.01$ とした。その結果、LM 特徴量と ILM 特徴量との間で、LM 特徴量と CM 特徴量との間で、および、ILM 特徴量と CM 特徴量との間で、精度に有意差が見られた。提案した CM 特徴量は、既存の LM 特徴量、および、その改良である ILM 特徴量と比べて、認識精度が高いと言える。以上より、空間手掛りを抑止し、時間手掛りを捉えるよう設計された CM 特徴量は、軽重認識の精度を高める効果があることが分かった。

次に、Fig. 11(b) にて、識別器として SVM, RF, または、GBDT と、各特徴量とを組み合わせた時の認識精度を示す。こ

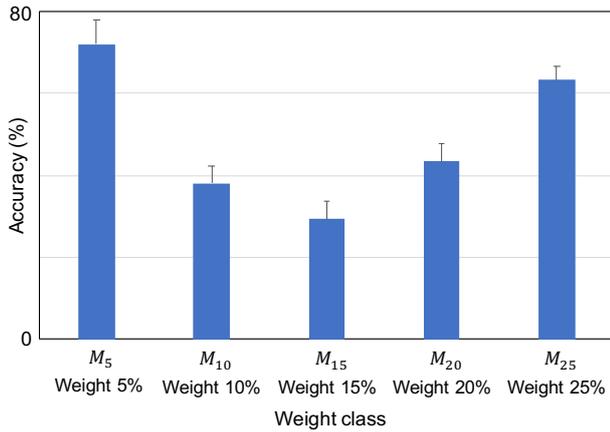


Fig. 12 Accuracy of each weight class of baggage. We used our method that incorporates the CM feature into the GBDT.

の棒グラフでは、識別器を固定し、そのパラメータと特徴量との組み合わせを全通りで試行した時の平均精度と標準偏差を用いた。SVMを用いた場合は $28.9 \pm 6.4\%$ 、RFを用いた場合は $40.2 \pm 3.8\%$ 、GBDTを用いた場合は $42.9 \pm 5.9\%$ であった。認識精度の間に有意差があるかどうかを確認するため、ウィルコクソンの符号付順位和検定、および、ボンフェローニ補正を適用した。その結果、SVMとRFとの間、SVMとGBDTとの間、および、RFとGBDTとの間に有意差が見られた。この検定結果より、手荷物の軽重認識において、GBDTが、SVMとRFと比べて、認識精度が安定になることが分かった。

ここまで述べた評価結果より、提案したCM特徴量とGBDTとを組ませた手法が、手荷物の軽重認識の精度を高めると言える。実際にGBDTの木の数を300とした時、CM特徴量と組み合わせた認識手法の精度が $49.2 \pm 2.1\%$ と最も高かった。さらに評価を進めるため、手荷物の重さクラス毎の認識精度を評価した結果を Fig. 12 に示す。重さクラス M_5 と M_{25} の認識精度は、他の重さクラスの精度と比較すると、高いことが分かった。一方で、 M_{10} 、 M_{15} 、 M_{20} の認識精度は低いことが分かった。中でも M_{15} の精度は、チャンスレートの20%の精度を上回っているものの、認識精度が大きく低下していた。

それぞれの重さクラスにおいて、CM特徴量を人物間で平均した結果を Fig. 13 に示す。PSDの値は、2 Hzを中心とした帯域において、 M_5 から M_{25} になるにつれ大きくなるのが分かる。ただしこの結果は、人物間で特徴量を平均したものであり、実際には個人間で特徴量のばらつきが大きかった。個人間のばらつきは、手荷物の軽重認識において、クラス内変動となるため、誤認識を引き起こす原因になると考えられる。次に、CM特徴量とGBDTとを用いた認識手法において、精度が最も高かった時の混合行列を Table 2 に示す。正解が M_5 の時、誤認識は他クラスで僅かに発生した。正解が M_{10} の時、誤認識は M_5 で多く発生した。正解が M_{15} の時、および、正解が M_{20} の時、誤認識はどの重さクラスでも似た割合で発生した。正解が M_{25} の時、誤認識は M_{20} に集まった。これらの結果より、重さクラス M_{10} や M_{15} や M_{20} のCM特徴量は、他の重さクラスの特徴量と類似することが多いと考えられる。提案したCM特徴量の限界であるため、中間の重さクラスでも安定に認識できる特徴量の開発が今後は必要である。

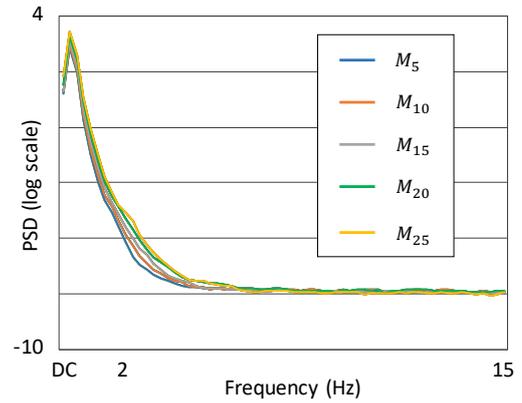


Fig. 13 CM features averaged between the participants for each weight class.

Table 2 Confusion matrix of baggage weight classification among M_5 , M_{10} , M_{15} , M_{20} , and M_{25} weight classes.

True weight class	Predict weight class				
	M_5	M_{10}	M_{15}	M_{20}	M_{25}
M_5	11	1	2	1	0
M_{10}	4	6	2	2	1
M_{15}	3	3	5	2	2
M_{20}	2	2	1	7	3
M_{25}	1	0	0	4	10

4.3 見かけの大きさの正規化の効果

シルエット画像中の頭領域の見かけの大きさについて、3.3で述べた正規化を適用することの効果について評価した。提案した特徴抽出における正規化では、頭領域の縦幅と横幅とを、一定の大きさ $W = 100, H = 100$ 画素となるように、 $\mathbf{S} = \text{diag}(W/W', H/H')$ を用いて画像変換を行っていた。他の処理として、正規化前の頭領域のアスペクト比を固定したまま、画像変換を行うことが挙げられる。この実験では、頭領域の縦幅のみを揃えるために $\mathbf{S} = \text{diag}(H/H', H/H')$ を用いて画像変換を行う場合、および、頭領域の横幅のみを揃えるために $\mathbf{S} = \text{diag}(W/W', W/W')$ を用いて画像変換を行う場合について評価した。さらに、正規化を適用しない場合についても評価した。認識手法として、CM特徴量とGBDTとを組み合わせたものを用いた。その他の実験条件は4.2.2と同じとした。

シルエット画像における頭領域の見かけの大きさについて、正規化の処理を変更した場合の認識精度を以下に示す。提案手法の正規化である縦幅と横幅との両方を揃える場合は $49.2 \pm 2.1\%$ であった。一方、縦幅のみを揃える場合は $39.2 \pm 2.2\%$ であり、横幅のみを揃える場合は $39.5 \pm 1.5\%$ であった。なお、正規化を適用しない場合は $37.7 \pm 1.2\%$ であった。提案手法で用いた正規化は、他の処理と比較して、高い認識精度を得ることができた。頭領域の縦幅と横幅の両方を一定に揃える正規化により、精度が高まる理由を考察する。3.1で述べたように、手荷物の軽重認識において、人物間の形状差は、手掛かりとして有用である可能性は低いと考えられる。このため、軽重認識の特徴量を抽出する際、人物間の頭の形状差を極力抑えた方がよいと本論文では想定した。縦幅と横幅を一定に揃える正規化により、人物間の形状差を抑える効果が生じて、認識精度が向上したと考えられる。以上より、CM特徴量を用いた手荷物の軽重認識において、縦幅と横幅とを揃える正規化

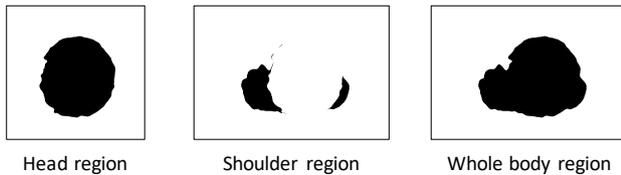


Fig. 14 Examples of the head region, shoulder region, and whole body region.

を用いることは効果があると考えられる。

4.4 頭領域のシルエット画像の効果

提案した特徴抽出では頭領域のシルエット画像を用いたが、頭上に設置されたカメラから獲得される深度画像の中には、頭以外の身体部位も観測される。そこで、他の身体部位を指す領域を用いた場合の認識精度を評価した。具体的には、深度画像中の人物について、肩のみを指す領域、または、全身を指す領域を用いて、時系列のシルエット画像を生成した。生成されたシルエット画像の例を Fig. 14 に示す。なお、肩領域と全身領域とでは、横幅が縦幅と比べて大きくなるため、シルエット画像のアスペクト比を図中のように横長とした。各領域において、時系列のシルエット画像から CM 特徴量を算出し GBDT を適用することで、軽重認識の精度を算出した。その他の実験条件は 4.2.2 と同じとした。

頭領域、肩領域、または、全身領域のシルエット画像を用いた場合の認識精度を以下に示す。頭領域を用いた場合は $49.2 \pm 2.1\%$ であった。肩領域を用いた場合は $32.0 \pm 6.5\%$ であった。また、全身領域を用いた場合は $31.6 \pm 1.5\%$ であった。頭領域のみを用いて特徴抽出を行った場合は、全身領域および肩領域から特徴抽出を行った場合と比較して、認識精度が高かった。以上の結果より、手荷物の軽重認識において、頭領域を用いることは有効であると考えられる。

4.5 深層学習を用いた手法、および、いくつかの既存の特徴抽出と識別器とを組合わせた手法との比較

深層学習を用いた手法、および、文献¹⁾以外の既存の特徴抽出と識別器とを組合わせた手法について、手荷物の軽重認識の精度を評価した。その結果を Table 3 に示す。表中では、提案した CM 特徴量と GBDT とを組合わせた手法を Ours とする。比較した各手法について、詳細を以下で述べる。

まず、時系列信号を扱う深層学習を用いることで、手荷物の軽重認識を行う手法を設計した。具体的には、Long short-term memory (LSTM)⁹⁾ および Gated recurrent unit (GRU)¹⁰⁾ を用いた。まず、3.4 で述べた各時刻における中心位置の移動量 d_t を、1 人あたり 160 秒間で時系列に並べた信号を準備した。この時系列信号から、10 秒間の短信号をランダムに生成した。各重さクラスにつき生成された 4200 個の短信号を訓練サンプルとした。訓練時には、隠れ状態の次元数を 128 とした。損失項として多クラス交差エントロピー、最適化アルゴリズムとして RSMPProp を用いた。Table 3 で示したように、LSTM を用いた場合の認識精度は $22.5 \pm 1.8\%$ であり、GRU を用いた場合の認識精度は $18.0 \pm 2.5\%$ であった。次に、時系列画像を扱う深層学習を用いることで、手荷物の軽重認識を行う手法を設計した。具体的には、三次元畳み込み層を用いる C3D¹¹⁾ を用いた。ネットワーク構造として、三次元畳み込み層 4 個と三次元プーリング層 4 個を用いた。損失項として多クラス交差エントロピー、最適化アルゴリズムとして RSMPProp を用いた。一人あ

たり 160 秒の時系列シルエット画像から、10 秒間の短映像をランダムに生成した。各重さクラスにつき生成された 4200 個の短映像を訓練サンプルとした。なお、各時刻のシルエット画像のサイズを 100×100 画素とした。Table 3 で示したように、C3D を用いた場合の認識精度は $27.7 \pm 2.5\%$ であった。これらの結果より、提案手法の認識精度は、深層学習を用いる LSTM, GRU, および、C3D の精度と比較して、高いことが分かった。

次に、時系列画像から特徴量を抽出し、識別器と組合わせることで、手荷物の軽重認識を行う手法を設計した。具体的には、特徴量として、Spatio-temporal histogram of oriented gradients (STHOG)¹²⁾, Gait energy image (GEI)¹³⁾, および、Dynamic image (DI)¹⁴⁾ を用いた。特徴量を算出する際、ここまでの実験条件と同様に、1 人あたり 160 秒間の時系列シルエット画像を用いた。各時刻のシルエット画像のサイズを 100×100 画素とした。識別器として、4.2.2 で述べた GBDT を用いた。以下で、それぞれの特徴量のパラメータについて詳細を述べる。STHOG では、セルサイズを $25(\text{画素}) \times 25(\text{画素}) \times 300(\text{フレーム})$, ブロックサイズを $2(\text{セル}) \times 2(\text{セル}) \times 2(\text{セル})$, ウィンドウサイズを $2(\text{ブロック}) \times 2(\text{ブロック}) \times 8(\text{ブロック})(100(\text{画素}) \times 100(\text{画素}) \times 4800(\text{フレーム}))$ とし、特徴量を抽出した。なおパラメータ記述の仕様は文献^{34,35)} に従った。GEI では、160 秒間の時系列シルエット画像において、各画素で時間方向に画素値を平均し 1 枚の平均画像を求めることで、特徴量を抽出した。DI では、160 秒間の時系列シルエット画像に対して、Rank SVM³⁶⁾ を適用することで、特徴量を抽出した。各特徴量を用いて得られた認識精度を以下で述べる。Table 3 で示したように、STHOG の場合は $37.6 \pm 3.1\%$ であり、GEI の場合は $24.7 \pm 2.8\%$ であり、DI の場合は $27.6 \pm 5.2\%$ であった。これらの結果より、提案した CM 特徴量と識別器とを組合わせる手法の認識精度は、時系列画像から抽出された特徴量と識別器とを組合わせる手法の認識精度と比較して、高いことが分かった。

上記で述べた STHOG, GEI, および、DI は、時間手掛りのみでなく、空間手掛りも合わせて捉える特徴抽出の設計になっている。ここでは比較のため、STHOG について、空間手掛りを抑え時間手掛りを捉える設計に変更し、手荷物の軽重認識の精度を算出した。具体的には、STHOG のパラメータを、セルサイズを $100(\text{画素}) \times 100(\text{画素}) \times 300(\text{フレーム})$, ブロックサイズを $1(\text{セル}) \times 1(\text{セル}) \times 2(\text{セル})$, ウィンドウサイズを $1(\text{ブロック}) \times 1(\text{ブロック}) \times 8(\text{ブロック})(100(\text{画素}) \times 100(\text{画素}) \times 4800(\text{フレーム}))$ とすることで、特徴量を抽出した。この特徴量を THOG と呼ぶことにした。Table 3 で示したように、THOG を用いた場合の認識精度は $18.9 \pm 3.0\%$ であった。STHOG 特徴量について時間手掛りのみを捉える設計に変えたとしても、提案した CM 特徴量は高い精度を得ていた。

シルエット画像のサイズの影響について調査した。ここでは、Table 3 において、提案手法以外で最も精度が高かった比較手法である STHOG を対象とした。各時刻のシルエット画像のサイズを 200×200 画素とした。セルサイズを $50(\text{画素}) \times 50(\text{画素}) \times 300(\text{フレーム})$, ブロックサイズを $2(\text{セル}) \times 2(\text{セル}) \times 2(\text{セル})$, ウィンドウサイズを $2(\text{ブロック}) \times 2(\text{ブロック}) \times 8(\text{ブロック})(200(\text{画素}) \times 200(\text{画素}) \times 4800(\text{フレーム}))$ とし、特徴量を抽出した。その結果、認識精度は $31.9 \pm 2.9\%$ となり、サイズを 100×100 画素とした時の精度から 5.7 ポイント低下した。

以上の結果より、深層学習を用いた手法、および、いくつか

Table 3 We compared the accuracy of baggage weight classification between methods using deep learning techniques or several combinations of the existing feature extraction and the classifier.

Method	Accuracy
Ours	49.2±2.1%
LSTM	22.5±1.8%
GRU	18.0±2.5%
C3D	27.7±2.5%
STHOG	37.6±3.1%
GEI	24.7±2.8%
DI	27.6±5.2%
THOG	18.9±3.0%

の既存の特徴抽出と識別器とを組み合わせる手法と比較して、提案する特徴抽出と識別器とを組み合わせる手法は、手荷物の軽重認識の精度を高めることを確認した。

4.6 人が判定した時の精度

提案手法で得られた認識精度の妥当性を確認するため、手荷物の重さを、人が目視で判定した時の精度を評価した。以下では目視で判定する人を観察者と呼び、15名(男性13名、女性2名、平均年齢22.0±1.0歳)の観察者が実験に参加した。観察者に提示する刺激として、4.1で述べた深度カメラ(Kinect v2)で同時撮影された15名の実験協力者のカラー動画をを用いた。識別器と観察者との間で学習条件を可能な限り合わせるため、14名の実験協力者の動画を、訓練サンプルとした。訓練サンプルと異なる1名の実験協力者の動画を、テストサンプルとした。訓練サンプルとテストサンプルに含まれる実験協力者を、観察者ごとでランダムに変更した。観察者の学習時とテスト時の実験風景をFig. 15に示す。学習時には、観察者の手前に設置された2枚のディスプレイを利用し、テスト時には3枚のディスプレイを利用した。ディスプレイをBenQ社の24インチ(型番EW2430V)とした。以下で、観察者の学習とテストの詳細を述べる。

学習時には、訓練サンプルの動画を、重さクラスの正解ラベル付きで観察者に提示した。一つの重さクラスに属する14名の実験協力者の訓練サンプルを、1台目のディスプレイで一覧表示した。他の重さクラスの訓練サンプルと見比べることができるように、2台目のディスプレイを用いた。学習のための観察者の持ち時間を最大800秒とし、その時間内であれば、観察者はどの重さクラスの訓練サンプルを見てもよいとした。観察者は持ち時間の終了を待たずともよいとしたが、200秒以上の学習を必ず行うこととした。

テスト時には、ある重さクラスのテストサンプルの動画を提示した。この際、学習時とは別のディスプレイを用いた。観察者は、テストサンプルを見た後、どの重さクラスと判定したかについて、口頭で回答した。それぞれの重さクラスのテストサンプルを表示する順番はランダムとした。1名の観察者につき、全部で5個のテストサンプルを用いた。1個のテストサンプルにつき、回答までの観察者の持ち時間を最大160秒とした。テストサンプルを40秒以上見た後に回答するよう指示した。なおテスト時には、3台のディスプレイを用いて、学習時の訓練サンプルとテストサンプルとを見比べてもよいとした。

観察者が動画を目視することで、重さクラスを判定した時の精度は32.0±25.1%であった。これは、チャンスレートの20%より高い精度と言える。ただし、4.2.2で述べたCM特徴量と

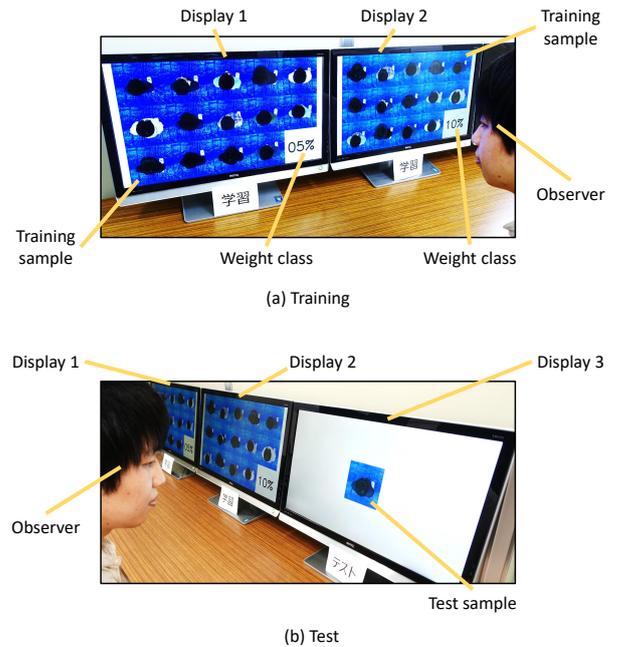


Fig. 15 Setup of the visual inspection of baggage weight classification by the observers.

GBDTとを組み合わせた認識手法の精度である49.2±2.1%と比べると、低いと言える。以上の結果より、提案手法の精度は、人が判定した時の精度を超えており、妥当な結果であると考えられる。

4.7 手荷物の重さの値そのものをクラスとする場合

手荷物の重さについて、体重の相対的な値でなく、重さの値そのものをクラスとして認識する場合を評価した。4.1で述べたアタッシュケースの中に入れる物体の重さを調整することで、1.7kgと4.7kgと8.7kgと11.7kgの手荷物をそれぞれ準備した。実験協力者は23名(男性22名、女性1名、平均身長169.0±6.2cm、平均年齢22.5±2.2歳、平均体重60.1±10.1kg)とした。各実験協力者について、各重さの手荷物を所持した場合の時系列深度画像を撮影した。手荷物を持つ腕の疲労を少なくするため、まず1.7kgの手荷物を所持した場合を撮影し、順に4.7kg、8.7kg、11.7kgの手荷物を所持した場合を撮影した。撮影と撮影の間に1時間の休憩を設けた。ある重さの時系列深度画像を撮影する際の時間長を160秒とし、撮影環境をFig. 10とした。識別器としてGBDTを用いた。以下では、LM特徴量、ILM特徴量、CM特徴量を用いた場合の認識精度を比較した。

手荷物の重さの値そのものをクラスとする場合の認識精度は、LM特徴量で37.7±1.8%、ILM特徴量で40.7±1.5%、CM特徴量で47.2±1.2%であった。4.2.2の結果と同様に、CM特徴量が、他の特徴量と比べて、高い精度を得ることができた。また、チャンスレートである25%よりも、高い精度を得ることができた。以上の結果より、提案手法は、手荷物の重さの値そのものをクラスとする場合にも、適用できる可能性があると言える。

5. まとめ

本論文では、片手で手荷物を所持し立ち止まる人物を対象とし、身体動揺の時系列深度画像を用いて、その手荷物の重さク

ラスを認識する手法について述べた。頭上に設置されたカメラで撮影された深度画像を用いることで、微小移動する頭領域の時系列シルエット画像を生成した。身体動揺を表す時間手掛かりを捉える特徴抽出を行うため、頭領域の見かけの大きさを正規化し、その領域の中心位置の時系列信号を求め、周波数解析を適用した。実験結果より、提案した CM 特徴量を用いる認識手法は、空間手掛りと時間手掛りの両方を捉える既存の LM 特徴量を用いる認識手法と比べて、手荷物の軽重認識の精度が向上することを確認した。また、深層学習を用いた手法 (LSTM, GRU, C3D)、および、その他の特徴抽出 (STHOG, GEI, DI) と識別器とを組み合わせる手法と比べても、軽重認識の精度を高めることを確認した。

人物の立ち位置が、カメラの真下からずれた場合について考察する。我々の先行研究³⁷⁾において、立ち位置が真下から前後左右のいずれかに 15 cm 移動すると、人物対応付けの精度が 97.1% から 85.6% へ低下することが分かっている。本論文の目的である手荷物の軽重認識においても、同様の精度低下が発生すると想定される。この原因として、カメラと人物との間で、相対的な位置関係が変わると、頭領域のシルエットが変形することが挙げられる。今後の課題として、人物の立ち位置変化に頑健な特徴量の開発が必要である。

さらなる今後の課題として、手荷物の種類が変わった場合や、その持ち方が変わった場合の影響調査が挙げられる。また、提案した特徴抽出の限界であった重さクラスに対する精度向上の技術開発が必要である。さらに、個人の筋力差による身体動揺への影響の調査や、高齢者など様々な年代の実験協力者を含むデータセットでの評価が必要である。

参考文献

- 1) T. Kamitani, H. Yoshimura, M. Nishiyama, and Y. Iwai. Temporal and spatial analysis of local body sway movements for the identification of people. *IEICE Transactions on Information and Systems*, Vol. 102, No. 1, pp. 165–174, 2019.
- 2) Y. Lu, A. Kumar, S. Zhai, Y. Cheng, T. Javidi, and R. Feris. Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1131–1140, 2017.
- 3) M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli. On soft biometrics. *Pattern Recognition Letters*, Vol. 68, pp. 218–230, 2015.
- 4) Y. Zhu, Y. Li, G. Mu, and G. Guo. A study on apparent age estimation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 25–31, 2015.
- 5) D. Damen and D. Hogg. Detecting carried objects in short video sequences. In *Proceedings of European Conference on Computer Vision*, pp. 154–167, 2008.
- 6) T. Senst, A. Kuhn, H. Theisel, and T. Sikora. Detecting people carrying objects utilizing lagrangian dynamics. In *Proceedings of International Conference on Advanced Video and Signal-Based Surveillance*, pp. 398–403, 2012.
- 7) M. G. Wittman, J. M. Ward, and P. J. Flynn. Visual analysis of the effects of load carriage on gait. In *Proceedings of Biometric Technology for Human Identification II*, pp. 15–22, 2005.
- 8) A. Bottaro, M. Casadio, P. G. Morasso, and V. Sanguineti. Body sway during quiet standing: Is it the residual chattering of an intermittent stabilization process? *Human Movement Science*, Vol. 24, No. 4, pp. 588–615, 2005.
- 9) S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- 10) J. Chung, C. Gülgeçre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, Vol. abs/1412.3555, pp. 1–9, 2014.
- 11) D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4489–4497, 2015.
- 12) C. Hua, Y. Makihara, and Y. Yagi. Pedestrian detection by using a spatio-temporal histogram of oriented gradients. *IEICE Transactions on Information and Systems*, Vol. 96, No. 6, pp. 1376–1386, 2013.
- 13) J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322, 2006.
- 14) B. Fernando, E. Gavves, J. Oramas, A. Ghodrati, and T. Tuytelaars. Rank pooling for action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 4, pp. 773–787, 2016.
- 15) Y. Yamaguchi, T. Kamitani, M. Nishiyama, Y. Iwai, and D. Kushida. Extracting features of body sway for baggage weight classification. In *Proceedings of the IEEE 9th Global Conference on Consumer Electronics*, pp. 345–348, 2020.
- 16) M. Lee, M. Roan, B. Smith, and T. E. Lockhart. Gait analysis to classify external load conditions using linear discriminant analysis. *Human movement science*, Vol. 28, No. 2, pp. 226–235, 2009.
- 17) B. Smith, M. Roan, and M. Lee. The effect of evenly distributed load carrying on lower body gait dynamics for normal weight and overweight subjects. *Gait & posture*, Vol. 32, No. 2, pp. 176–180, 2010.
- 18) E. J. Hyung, H. O. Lee, and Y. J. Kwono. Influence of load and carrying method on gait, specifically pelvic movement. *Journal of Physical Therapy Science*, Vol. 28, No. 7, pp. 2059–2062, 2016.
- 19) E. Orantes-Gonzalez, J. Heredia-Jimenez, and V. M. Soto-Hermoso. The effect of school trolley load on spatiotemporal gait parameters of children. *Gait & posture*, Vol. 42, No. 3, pp. 390–393, 2015.
- 20) R. Hardie, R. Haskew, J. Harris, and G. Hughes. The effects of bag style on muscle activity of the trapezius, erector spinae and latissimus dorsi during walking in female university students. *Journal of human kinetics*, Vol. 45, No. 1, pp. 39–47, 2015.
- 21) 内山応信, 出村慎一, 山次俊介, 山田孝禎, 池本幸雄, 宮口和義. 直立姿勢制御に及ぼす荷物保持の影響. 日本体育学会大会号, p. 430, 2004.
- 22) 佐藤健, 中島みづき, 村野良太, 加藤麻樹. 荷物保持の形態・重量による姿勢動揺への影響. 人間工学, Vol. 56, pp. 2G3–01, 2020.
- 23) F. Wang, M. Skubic, C. Abbott, and J. M. Keller. Body sway measurement for fall risk assessment using inexpensive webcams. In *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 2225–2229, 2010.
- 24) L. F. Yeung, K. C. Cheng, C. H. Fong, W. C. Lee, and K. Y. Tong. Evaluation of the microsoft kinect as a clinical assessment tool of body sway. *Gait & posture*, Vol. 40, No. 4, pp. 532–538, 2014.
- 25) Z. Lv, V. Penades, S. Blasco, J. Chirivella, and P. Gagliardo. Evaluation of kinect2 based balance measurement. *Neurocomputing*, Vol. 208, pp. 290–298, 2016.
- 26) M. Nishiyama, T. Miyauchi, H. Yoshimura, and Y. Iwai. Synthesizing realistic image-based avatars by body sway analysis. In *Proceedings of the Fourth International Conference on Human Agent Interaction*, pp. 155–162, 2016.
- 27) N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, Vol. 9, No. 1, pp. 62–66, 1979.
- 28) T. Kamitani, Y. Yamaguchi, S. Nakatani, M. Nishiyama, and Y. Iwai. Gender classification using video sequences of body sway recorded by overhead camera. In *Proceedings of 25th International Conference on Pattern Recognition*, pp. 9196–9202, 2021.
- 29) 厚生労働省. 職場における腰痛予防対策指針. 厚生労働省通達基発第 547 号, 1994.
- 30) 松野悟之, 明見匡人, 大道終人, 鈴木みのり, 浜岡克伺, 国宗翔. 荷物所持方法の相違が足圧中心および上半身重心の動揺に与える影響. 理学療法科学, Vol. 35, No. 6, pp. 785–789, 2020.
- 31) C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, Vol. 20, No. 3, pp. 273–297, 1995.
- 32) L. Breiman. Random forests. *Machine learning*, Vol. 45, No. 1, pp. 5–32, 2001.
- 33) J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232, 2001.
- 34) 川合諒, 榎原靖, 八木康史ほか. Sthog 特徴を用いた複数カメラ間での人物照合. 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2011, No. 10, pp. 1–8, 2011.
- 35) 柏本雄士朗, 村松大吾, 八木康史ほか. カメラ間人物照合における sthog 特徴の可視化と解析. 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2017, No. 17, pp. 1–8, 2017.
- 36) A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, Vol. 14, No. 3, pp. 199–222, 2004.
- 37) 神谷卓也, 安形俊輝, 吉村宏紀, 西山正志, 岩井儀雄. 身体動揺を用いた人物対応付けにおける立ち位置変動の影響調査. 信学技報, BioX2018-8, 第 118 巻, pp. 5–10, 2018.