Subjective Baggage-Weight Estimation from Gait: Can You Estimate How Heavy the Person Feels?

Masaya Mizuno[1], Yasutomo Kawanishi[2,1], Tomohiro Fujita[2], Daisuke Deguchi[1] and Hiroshi Murase[1]

[1]Graduate School of Informatics, Nagoya University, Chikusa-ku, Nagoya, Japan

[2]RIKEN Guardian Robot Project, Souraku-gun, Kyoto, Japan

Keywords: Subjective Baggage-Weight, G2SW (Gait to Subjective Weight), Graph Convolution, Multi-Task Learning.

Abstract: We propose a new computer vision problem of subjective baggage-weight estimation by defining the term *subjective weight* as how heavy the person feels. We propose a method named G2SW (Gait to Subjective Weight), which is based on the assumption that cues of the subjective weight appear in the human gait, described by a 3D skeleton sequence. The method uses 3D locations and velocities of body joints as input and estimates subjective weight using a Graph Convolutional Network. It also estimates human body weight as a sub-task based on the assumption that the strength of a person depends on body weight. For the evaluation, we built a dataset for subjective baggage-weight estimation, consisting of 3D skeleton sequences with subjective weight annotations. We confirmed that the subjective weight could be estimated from a human gait and also confirmed that the sub-task of body weight estimation pulls up the performance of the subjective weight estimation.

1 INTRODUCTION

Robots have been widely developed for various applications. Especially, in daily environments, various kinds of human support robots have been proposed (Yamamoto et al., 2019; Yuguchi et al., 2022). Such a robot that works in our living space should have a function of environmental recognition but also provide proactive support. In this study, we focus on such a support provided by a robot to a person carrying heavy baggage.

Robots that can carry the baggage should be developed to support a person carrying heavy baggage. Still, it is also important to develop a function that determines whether to support a person. If the robot tries to support a person who does not need the support, the person may get irritated with the robot. In that case, the person will not accept the robots. To avoid this situation, we focus on the functions that determine the needs of the support and provide the support at an appropriate time.

To make such decisions, the robots need to estimate how heavy the person feels. In this study, we define *subjective weight* as how heavy a person feels. The greater the subjective weight, the more difficult a person to think carrying baggage. To quantify the subjective weight, we employ the New Borg Scale (Gunnar, 1982), which is a measure of subjective load during exercise. By estimating the subjective weight, the

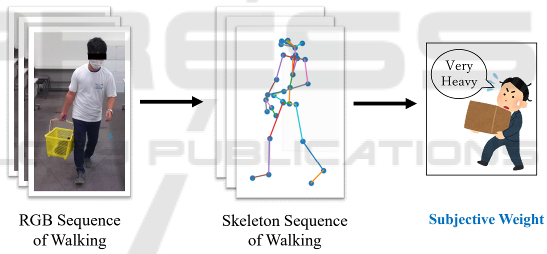

Figure 1: Estimation of the subjective baggage-weights from human gait.

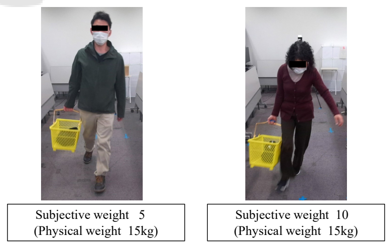

Figure 2: Example of the gait with the baggage of same physical weight.

robot will decides whether the person needs support. To make the problem setting clearer and simpler, we assume a situation where one person is walking with one piece of baggage as shown in Fig. 1.

Since the subjective weight is related to the actual weight of baggage (which is called *physical weight*), physical weight can be used as a clue for subjective

Mizuno, M., Kawanishi, Y., Fujita, T., Deguchi, D. and Murase, H.
Subjective Baggage-Weight Estimation from Gait: Can You Estimate How Heavy the Person Feels?.
DOI: 10.5220/0011668900003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 5: VISAPP*, pages 567-574
ISBN: 978-989-758-634-7; ISSN: 2184-4321

weight estimation. However, even if the baggage is the same, the physical weight varies depending on what is inside. Therefore, it is difficult to estimate the physical weight from the appearance of the baggage itself. Additionally, if the physical weight is even the same, the subjective weight varies from person to person. For example, a piece of baggage of the same physical weight may be felt heavy by a physically weak person like a child while it may be felt light by a physically strong person like a muscular person.

When we humans see someone walking with baggage, we can probably guess how heavy the person feels to carry the baggage from the walking behavior, that is, human gait. Hyung et al. (Hyung et al., 2016) reported the relationship between the physical weight of baggage and human gait, in which the pelvic tilt increases when the physical weight of baggage increases. Additionally, the paper also shows that the relationship between pelvic tilt and physical weight varies from person to person, which depends on his/her body weight. As shown in Fig. 2, even when the physical baggage-weights are the same, the human gait are different when the subjective weight is different. In this study, we propose a method that estimates the subjective weights by focusing on a human gait.

A human gait can be found in temporal variations of human skeleton sequences (Kato et al., 2017). The temporal variations of the human skeleton are often represented by the skeleton sequence, which has been used in gait recognition, action recognition, etc. (Teepe et al., 2021; Liu et al., 2016; Yan et al., 2018; Liu et al., 2020; Nishida et al., 2020; Temuroglu et al., 2020). These methods use graph representations for human skeleton sequences to realize the recognition tasks. In this study, we use 3D human skeleton sequences during walking as a representation of human gait.

Based on the above, we propose G2SW (Gait to Subjective Weight) that is a method for estimating the subjective weight from the human gait represented by a skeleton sequence (Fig. 1). We modify a graph-based action recognition method to estimate the subjective weight estimation.

We focus on the fact that the gait is the repetition of two steps (one walking cycle) and regard 3D body joint locations in one walking cycle to represent the gait. While graph-based action recognition methods usually accept a fixed length input, we need to normalize the length of walking cycles to a fixed length, by resampling frames in a walking cycle. However, this result in the loss of information on gait speed, since all gait cycles have the same length. For this problem, we introduce velocities as additional feature for each body joints. This locations-and-velocities representation retains velocity information but has a fixed length.

As noted above, the subjective baggage-weight is affected by body weight of the person. To take this into account in the estimation, the proposed method simultaneously estimates the body-weight of the person as a sub-task in the training phase. By using the sub-task, the network is trained to consider body weight in the subjective baggage-weight estimation.

The main contributions of this work are summarized as follows:

- We propose a new computer vision problem of subjective baggage-weight estimation of a piece of baggage, by defining subjective baggage-weight as how heavy a person feels, quantified by the New Borg Scale (Gunnar, 1982).
- We propose G2SW, an estimation method of subjective baggage-weight from the human gait. The method uses velocity information as additional feature, and body-weight estimation is added as a sub-task to focus on the difference of persons.
- We built a novel dataset of human skeleton sequences with subjective baggage-weight annotations.

The following section 2 presents the related work in this literature. Section 3 presents the proposed method that estimates the subjective weights. Then, section 4 presents the experimental evaluation. Finally, section 5 summarizes and discusses future issues.

2 RELATED WORK

2.1 Baggage Weight Estimation

Yamaguchi et al. (Yamaguchi et al., 2020) have proposed a method to estimate baggage weight from body sway. Body sway is the slight swaying of a person's body even when standing upright and stationary. The method estimates the baggage weight by focusing on the characteristic that the heavier the weight, the greater the body sway. Because this method requires observing a stationary standing person from a bird's-eye view, the method is not directly applicable to a robotic application.

Oji et al. (Oji et al., 2018) have proposed a weight estimation method from lifting motion. The method estimates the weight of an object from a hand motion by focusing on the fact that the hand motion changes depending on the weight of the object when lifting an object. However, it requires object lifting motion, its applicable situation is limited.

2.2 Action Recognition by a Body Skeleton Sequence

Long Short-Term Memory(LSTM), which can capture temporal information, is often used in action recognition from a skeleton sequence (Liu et al., 2016; Liu et al., 2017; Ullah et al., 2018; Majd and Safabakhsh, 2020).

In recent years, the Graph Convolutional Network (GCN), which consists of the graph convolution layers, has become the mainstream of action recognition. It regards a skeleton as a graph. Generally, each body joint and each limb are represented as a vertex and an edge in a graph, respectively. ST-GCN (Yan et al., 2018) is a method of action recognition from a skeleton sequence that considers the skeleton sequence as a temporally-connected graph. This method extracts spatial features by applying graph convolution for each frame, followed by temporal convolution for each temporal sequence of a body joint to extract temporal features. By the structure, the method can consider the skeleton structure and motion in the action recognition task.

There are several methods that extend the ST-GCN. One of the extensions is the multi-scale direction, represented by MS-G3D (Liu et al., 2020). The main component of the method is the G3D module, which is a graph convolution version of the I3D module (Carreira and Zisserman, 2017). The module consists of graph convolutions over a spatio-temporal graph corresponding to a skeleton sequence. The method further extends the module to multi-scale using multiple graphs of different multi-hop connections. The multi-hop connections of the graphs allow us to directly connect body joints that are skeletally distant from each other but are important for the recognition task.

3 PROPOSED METHOD

3.1 Overview

This paper proposes a method for estimating the subjective baggage-weight from a human gait, named G2SW (Gait to Subjective Weight). In this study, we use a 3D human skeleton sequence as a representation of human gait. A 3D human skeleton sequence is a set of (X,Y,Z) coordinates of joint locations in the world coordinate system. Here, $(X_t^j, Y_t^j, Z_t^j)^\top$ denotes the location of the j-th body joint in t-th frame. Additionally, noting that walking is a repetition of two steps, we define two steps as one cycle of walking and use the 3D human skeleton sequence $\mathcal{S}_i$ for one cycle of walking as input.

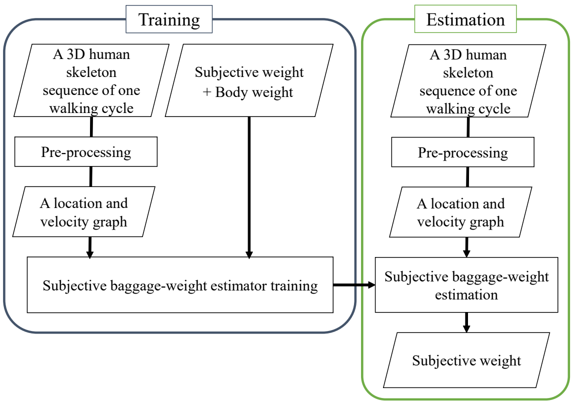

Figure 3: The training and estimation processes of the proposed method.

In this section, we describe the detail of G2SW, a regression-based method for estimating the subjective weight based on the human gait represented by a 3D human skeleton sequence of one cycle of walking. Figure 3 shows the flowchart of the training and estimation steps of the proposed method. As a pre-processing, the i-th 3D human skeleton sequence $\mathcal{S}_i$ is converted into a location and velocity graph $\widehat{\mathcal{S}_i}$. Then the location and velocity graph is input to the subjective weight estimator (G2SW) to estimate the subjective weight. For training the G2SW, we employ multi-task learning where body weight estimation is a sub-task.

In the following, first, we define the subjective weight in section 3.2. Then, the pre-processing for the input is explained in section 3.3. The network architecture and its multi-task learning are explained in section 3.4.

3.2 Definition

In this study, we define subjective weight as how heavy a person feels. We employ the New Borg Scale (Gunnar, 1982) to quantify the subjective weight. Originally, the New Borg Scale quantifies how hard the activity is as shown in Table 5. We use the scale to quantify how heavy a person feels, and the proposed method G2SW estimates the value of the New Borg Scale. In the new Borg Scale, there are scale values that do not have subjective descriptions. In the subjective weight assessment, it is possible to choose these values if participant feels that subjective weight exists between these subjective descriptions.

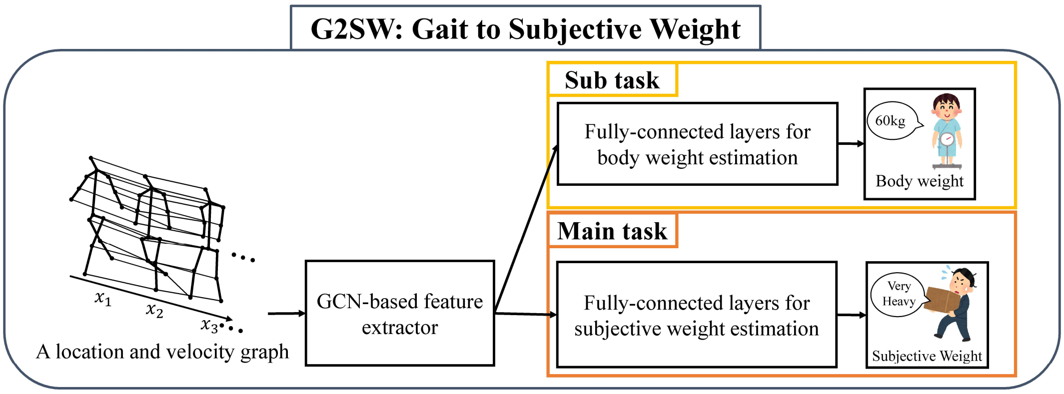

Figure 4: The network architecture of the proposed G2SW.

Scale	Description	Scale	Description
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

Figure 5: New Borg Scale.

3.3 Pre-Processing

In the proposed method, a 3D human skeleton sequence of one walking cycle is assumed to be cropped beforehand based on the frame where the positions of the left and right legs are most distant.

Since the 3D human skeleton sequence of one walking cycle captured by a sensor is described in the world coordinate, it varies according to the location and orientation of the person. The recognition should be robust to the location and orientation. Also, the lengths of walking cycles are different. For accurate recognition, these variations should be normalized.

When simply sampling frames of a fixed length from a walking cycle, walking speed information will be ignored. Therefore, we enhance the input skeleton sequences by adding the velocity information of each body joint.

Therefore, the pre-processing consists of i) location and orientation normalization, ii) velocity calculation, and iii) frame sampling for the fixed length. After the pre-processing, an input 3D skeleton sequence will be normalized and enhanced. We named the output of the pre-processing as *location and velocity graph*.

I) Location and Orientation Normalization. First, the location of the skeletons of each frame is normalized by aligning the input skeletons so that the position of the pelvis in each frame becomes the origin $(0,0,0)^\top$. Also, the orientation is normalized by rotating so that the locations of both hips are on the X-Z plane where the horizontal plane is X-Y. By this process, the locations of the j-th body joint in t-th frame will be $(\overline{X}_t^j, \overline{Y}_t^j, \overline{Z}_t^j)^\top$.

II) Velocity Calculation. The velocity here is defined as the difference between the locations of each body joint in adjacent frames.

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

The calculated velocity of each body joint is appended to the corresponding body joint so that the j-th body joint in the t-th frame has a 6-dimensional feature $(\overline{X}_t^j, \overline{Y}_t^j, \overline{Z}_t^j, \dot{X}_t^j, \dot{Y}_t^j, \dot{Z}_t^j)^\top$.

III) Frame Sampling for the Fixed Length. Since the length of one walking cycle is different among 3D human skeleton sequences, the length should be fixed to input them into a graph convolutional network. M frames are sampled from the original sequence at approximately equivalent intervals by interpolating from adjacent frames.

3.4 The Proposed G2SW and Its Multi-Task Training

In the proposed G2SW, subjective weight is estimated from the location and velocity graph $\overline{S}_i$, which is preprocessed output of the i-th 3D skeleton sequence of one walking cycle S_i.

The architecture of the proposed G2SW is shown in Fig. 4. In the proposed G2SW, a feature representation is calculated using a GCN-based feature extractor f as

$$\mathbf{p}_i = f(\overline{S}_i; \theta_f, A), \quad (2)$$

where A denotes an adjacent matrix that defines the adjacency of human body joints. This function f consists of multiple graph convolution layers. In this study, as the feature extraction function f, the two consecutive blocks of the MS-G3D module (Liu et al., 2020) are used. Here, parameters in the network are represented by θ_f. After the MS-G3D blocks, the graph-shaped output is reshaped to a 1-dimensional vector $\mathbf{p}_i$.

Then, subjective weight is calculated using fully-connected layers g. At that time, body weight is also calculated using fully-connected layers h, simultaneously.

$$w_i^s = g(\mathbf{p}_i; \theta_g), \tag{3}$$

$$w_i^b = h(\mathbf{p}_i; \theta_h). \tag{4}$$

These two functions g and h consist of four fully-connected layers whose parameters are θ_g and θ_h, respectively. Leaky ReLU (Maas et al., 2013) is used as the activation function for the hidden layers.

Given a batch of $\overline{\mathcal{S}_i}$ and corresponding ground truth of subjective weight and body weight $(\widehat{w}_i^s, \widehat{w}_i^b)$, the network is trained in multi-task learning manner. The parameters θ_f, θ_g, and θ_h are updated using back-propagation to minimize the mean squared error of the loss L consists of subjective weight loss L_s and body weight loss L_b.

$$L = \lambda_s L_s + \lambda_b L_b, \tag{5}$$

$$L_s = \sum_i (w_i^s - \widehat{w}_i^s)^2, \tag{6}$$

$$L_b = \sum_i (w_i^b - \widehat{w}_i^b)^2, \tag{7}$$

where λ_s and λ_b are the weight of the loss. Here, the ranges of subjective weight and body weight are normalized to be a similar scale.

4 EVALUATION

4.1 Dataset

Because there are no publicly available datasets that consist of 3D skeleton sequences with annotations of the subjective baggage weights, we originally captured a dataset for the evaluation. This section describes the details of our dataset.

In this study, we assume a situation where one person is walking with a piece of baggage. The 3D human skeleton sequences were collected by observing each participant walking with a piece of baggage using a Microsoft Azure Kinect sensor installed from a height of 2 m. The frame rate was 30 fps. Fig. 6 shows captured images and the 3D human skeleton sequences of each type of baggage.

Since subjective weight may be affected by body size and gender, the set of participants should not have biases in body size and gender. We employed 30 participants (15 males and 15 females) of diverse heights and weights for the dataset. Figure 7 shows the distribution of the participants' heights and weights.

We prepared five types of baggage, consisting of a handbag, shoulder bag, backpack, cardboard box, and shopping basket , and we prepared six variations of the contents weight of the baggage consisting of 0 kg, 5 kg, 7.5 kg, 10 kg, 12.5 kg, and 15 kg. The subjective weights were annotated by a questionnaire survey to the participants themselves. In the questionnaire, participants scored how hard they felt after walking with each baggage according to the New Borg Scale (Gunnar, 1982) (Table 5). To prevent the participants from knowing the actual value of the physical weight, the contents of the baggage were hidden from them.

In a session, a participant walked with a piece of prepared baggage, and a short break was inserted after each session to avoid the effect of the previous session. In this experiment, 30 patterns (five baggage types × six weights) of 3D human skeleton sequences were captured for each subject.

All the participants consented to the use and disclosure of their captured data for research purposes. It should be noted that the Ethics Committee at Nagoya University has approved this experiment.

4.2 Evaluation Protocol and Metrics

In this experiment, we performed 5-fold cross-validation that splits 5 people for evaluation and the rest of 30 people for training from the dataset.

Because the total number of pre-processed 3D human skeleton sequences in the dataset was only 24,015 walking cycles, data augmentation was applied. From an input 3D skeleton sequence of one walking cycle, three frames are randomly dropped. We performed this ten times for each walking cycle, thus increasing the data volume to 240,150 walking cycles. In the experiment, the frame length of a location and velocity graph is set to $M = 50$ after this data augmentation.

We evaluated the G2SW performance for the subjective weight estimation for each type of baggage. As an evaluation metric, we employ the mean absolute error (MAE) of the estimation results.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |w_i^s - \widehat{w}_i^s|, \tag{8}$$

where N represents the number of the 3D skeleton sequences.

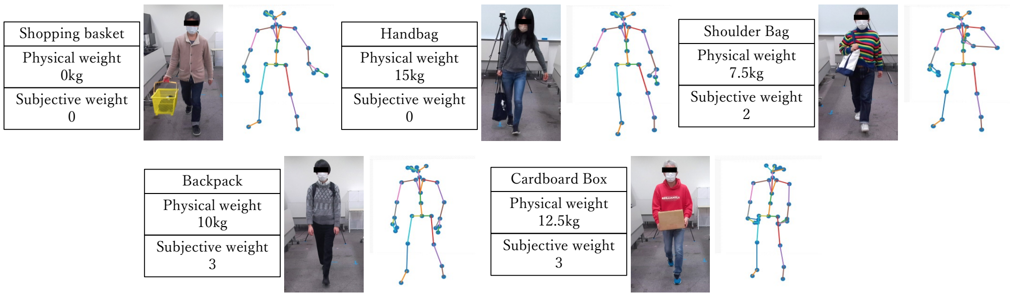

Figure 6: Examples of captured images and 3D human skeletons of each type of baggage.

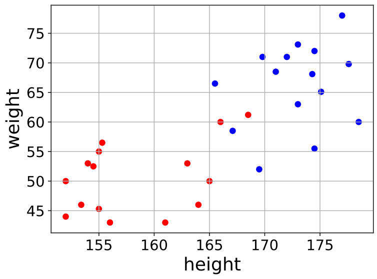

Figure 7: The distribution of the participants' heights (cm) and weights (kg) (blue:male, red:female).

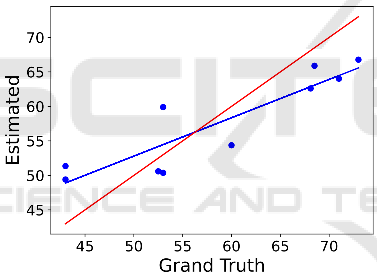

Figure 8: A plot of body weight Estimation (blue points: estimation plot of each person, blue line: linear fitting result of estimation plot, red line: grand truth (the unit of body weight is kg)).

In terms of the application for deciding whether to support a person, we also evaluated the performance of estimation within a tolerance error threshold, named Tolerance Accuracy (TA).

$$\mathrm{TA}_\tau = 100\frac{NW_\tau}{N}, \tag{9}$$

where τ is the tolerance error threshold, and NW_τ represents the number of data within the estimation error τ.

4.3 Preliminary Experiment: Body Weight Estimation

In this study, body weight estimation was performed as a sub-task of G2SW. However, if the accuracy of weight estimation from skeleton features is low, it is

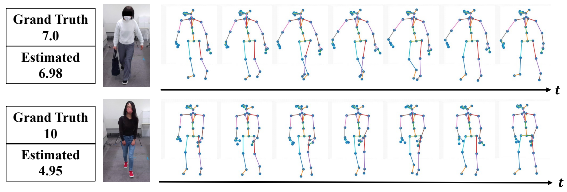

Figure 9: Example of estimation results (skeleton sequence represents gait).

inappropriate to use body weight estimation as a subtask. Therefore, we confirmed whether it is possible to estimate the body weight from skeleton features. Figure 8 shows the plots of the mean of each person's weight estimate in the test data and the linear fitting result. From this, it can be seen that the weight estimates are correlated with the true values. From the above finding, we can say that the estimation of the body weight can be estimated from a 3D skeleton sequence of one walking cycle.

4.4 Main Experiment: Subjective Baggage-Weight Estimation

Table 1 shows the mean absolute errors of the subjective weight estimation and Tolerance Accuracy of $\tau = 1, 2$, and 3. Figure 9 shows an example of the estimation results. From Table 1, we confirmed that G2SW could estimate subjective weights with the mean absolute error of 1.33 in New Borg Scale as the average of the entire baggage. And, G2SW could estimate 50.6% in Tolerance Accuracy with $\tau = 1(TA_1)$, 74.7% in Tolerance Accuracy with $\tau = 2(TA_2)$, and 88.4% in Tolerance Accuracy with $\tau = 3(TA_3)$ as the average of the entire baggage.

4.5 Discussion

Through the experiments, we confirmed that G2SW can estimate subjective weight.

Table 1: G2SW's Evaluations for subjective baggage-weight (subjective weight: 0–10).

Type of Baggage	MSE↓	TA_1 ↑	TA_2 ↑	TA_3 ↑
Handbag	1.42	46.6%	71.1%	87.0%
Shoulder bag	1.09	57.8%	81.6%	93.6%
Backpack	1.38	49.5%	73.8%	87.1%
Cardboard box	1.44	47.0%	72.5%	86.7%
Shopping basket	1.32	52.2%	74.8%	88.0%
Average	1.33	50.6%	74.7%	88.4%

Table 2: Comparison of subjective weight estimation accuracy with and without velocity information.

	MSE↓	TA_1 ↑	TA_2 ↑	TA_3 ↑
with velocity	**1.23**	**50.6%**	**74.7%**	**88.4%**
without velocity	1.42	47.8%	72.2%	86.6%

4.5.1 Difference Among Baggage Types

Table 1 confirmed that the subjective weight estimation for carrying a shoulder bag is more accurate than that for carrying other baggage. This is because the degree of postural change caused by subjective weight tends to be larger when carrying a shoulder bag than other baggages.

A possible cause of greater postural change due to subjective weight is gait stability. The more unstable the gait, the more likely the posture changes by external factors such as the weight of baggage. A handbag, shoulder bag, and shopping basket are held with only one shoulder or one hand, making gait unstable, while backpacks and cardboard boxes are held with both shoulders or hands, making walking more stable relatively.

Another cause of greater postural change due to subjective weight is the distance between the baggage and the human center of gravity. When the distance is large, the human needs to change his posture more significantly to maintain balance than when the distance is shorter. Among the baggage which cause unstable gait, the shoulder bag has the furthest distance to the human center of gravity. Therefore, the human needs to change his posture larger when carrying a shoulder bag than the rest baggage.

4.5.2 Effectiveness of the Velocity Feature

In the method, we propose the location and velocity graph to preserve velocity information in the fixed sequence length of one cycle walking. To confirm the effectiveness of the additional velocity features as input, we compared our method with a method that did not uses velocity information. Table 2 shows a comparison of subjective weight estimation accuracy with and without velocity information. From the table, it was confirmed that the accuracy of subjective weight estimation was improved by using the velocity information as additional information.

4.5.3 Challenges of the Practical Use

In the proposed method G2SW, the estimation is performed on the sequence for one cycle of walking cropped from the sequence during walking; however, in reality, several walking cycles are obtained from a captured sequence of walking. Therefore, multiple estimation results are obtained for a single sequence during walking. In the future, it will be necessary to consider how to integrate the multiple estimation results obtained.

5 CONCLUSION

In this study, we proposed a new computer vision problem of the subjective baggage-weight estimation when a person is walking with a piece of baggage and established G2SW which is an estimation method for subjective weights. To quantify the subjective weight, we defined it using the New Borg Scale. Since subjective weights affect the human gait, we proposed a subjective weight estimation method from a human gait, represented by a 3D human skeleton sequence. The method uses locations and velocities of body joints as input and estimates human body weight as a sub-task based on the assumption strength of a person depends on body weight.

Future work includes a further update of the gait representation describing the motion of skeletons more effectively.

ACKNOWLEDGEMENTS

This work was supported in part by the JSPS Grant-in-Aid for Scientific Research (17H00745).

REFERENCES

Carreira, J. and Zisserman, A. (2017). Quo Vadis, action recognition? a new model and the kinetics dataset. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 4724–4733, Honolulu, HI. IEEE.

Gunnar, B. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, 14(5):377–381.

Hyung, E.-J., Lee, H.-O., and Kwon, Y.-J. (2016). Influence of load and carrying method on gait, specifically pelvic movement. *Journal of Physical Therapy Science*, 28(7):2059–2062.

Kato, H., Hirayama, T., Kawanishi, Y., Doman, K., Ide, I., Deguchi, D., and Murase, H. (2017). Toward describing human gaits by onomatopoeias. In *Proceedings of the 16th IEEE International Conference on Computer Vision Workshops*, pages 1573–1580. IEEE.

Liu, J., Shahroudy, A., Xu, D., and Wang, G. (2016). Spatio-temporal LSTM with trust gates for 3D human action recognition. In Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, *Computer Vision – ECCV 2016*, volume 9907, pages 816–833. Springer International Publishing.

Liu, J., Wang, G., Hu, P., Duan, L.-Y., and Kot, A. C. (2017). Global context-aware attention LSTM networks for 3D action recognition. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3671–3680, Honolulu, HI. IEEE.

Liu, Z., Zhang, H., Chen, Z., Wang, Z., and Ouyang, W. (2020). Disentangling and unifying graph convolutions for skeleton-based action recognition. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 140–149, Seattle, WA, USA. IEEE.

Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of the 30th International Conference on Machine Learning Workshop*, pages 1–6.

Majd, M. and Safabakhsh, R. (2020). Correlational convolutional LSTM for human action recognition. *Neurocomputing*, 396:224–229.

Nishida, N., Kawanishi, Y., Deguchi, D., Ide, I., Murase, H., and Piao, J. (2020). SOANets: Encoder-Decoder based skeleton orientation alignment network for white cane user recognition from 2D human skeleton sequence. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 435–443, Valletta, Malta. SCITEPRESS - Science and Technology Publications.

Oji, T., Makino, Y., and Shinoda, H. (2018). Weight estimation of lifted object from body motions using neural network. In Prattichizzo, D., Shinoda, H., Tan, H. Z., Ruffaldi, E., and Frisoli, A., editors, *Proceedings of the Haptics: Science, Technology, and Applications*, volume 10894, pages 3–13. Springer International Publishing, Cham.

Teepe, T., Khan, A., Gilg, J., Herzog, F., Hörmann, S., and Rigoll, G. (2021). GaitGraph: Graph convolutional network for skeleton-based gait recognition. In *Proceedings of the 2021 IEEE International Conference on Image Processing*, pages 2314–2318.

Temuroglu, O., Kawanishi, Y., Deguchi, D., Hirayama, T., Ide, I., Murase, H., Iwasaki, M., and Tsukada, A. (2020). Occlusion-aware skeleton trajectory representation for abnormal behavior detection. In Ohyama, W. and Jung, S. K., editors, *Proceedings of the 2020 International Workshop on Frontiers of Computer Vision*, volume 1212, pages 108–121, Singapore. Springer Singapore.

Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M., and Baik, S. W. (2018). Action recognition in video sequences using deep bi-directional LSTM with CNN features. *IEEE Access*, 6:1155–1166.

Yamaguchi, Y., Kamitani, T., Nishiyama, M., Iwai, Y., and Kushida, D. (2020). Extracting features of body sway for baggage weight classification. In *Proceedings of the 9th IEEE Global Conference on Consumer Electronics*, pages 345–348. IEEE.

Yamamoto, T., Terada, K., Ochiai, A., Saito, F., Asahara, Y., and Murase, K. (2019). Development of human support robot as the research platform of a domestic mobile manipulator. *ROBOMECH Journal*, 6(1):1–15.

Yan, S., Xiong, Y., and Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pages 7444–7452.

Yuguchi, A., Kawano, S., Yoshino, K., Ishi, C. T., Kawanishi, Y., Nakamura, Y., Minato, T., Saito, Y., and Minoh, M. (2022). Butsukusa: A conversational mobile robot describing its own observations and internal states. In *Proceedings of the 2022 17th ACM/IEEE International Conference on Human-Robot Interaction*, pages 1114–1118, Sapporo, Japan. IEEE.