Extracting features of body sway for baggage weight classification

Yuta Yamaguchi, Takuya Kamitani, Masashi Nishiyama, Yoshio Iwai, Daisuke Kushida
Graduate School of Sustainability Science, Tottori University, Japan

***Abstract*—We propose a method for extracting features of body sway to classify the weight of baggage carried by a person as heavy or light by using a video sequence of depth images. The slight swaying movement naturally occurring in the body can be measured using a video sequence of a standing person acquired using an overhead camera. We hypothesize that body sway varies with the weight of the baggage carried by a person. Our method extracts a temporal feature of body sway used to discriminate between people carrying light and heavy baggage using a frequency analysis technique. We collected a dataset of video sequences from 40 participants for an experiment to test the proposed method to classify the weight of baggage. The results confirmed that feature extraction using body sway can help classify baggage as light or heavy with an accuracy of 95%.**

***Index Terms*—Baggage weight classification, body sway, overhead camera, feature extraction**

I. Introduction

Personal attribute classification using video sequences acquired from surveillance cameras has been widely researched. For instance, several methods [1], [2] have been designed for classifying the personal attributes of gender, age, and clothes of people. A few methods [3], [4] have been proposed to detect baggage being carried by people to collect specific personal attributes, but consider only the apparent type of baggage. In this paper, we propose a method to classify baggage by weight as heavy or light based on video sequences of people as they are carrying it. The classification of baggage weight has a variety of applications. For instance, it can be used to identify people carrying unnaturally heavy baggage at major transportation hubs, and to identify and help older people with heavy baggage through support robots.

We consider the kinds of cues available to classify the weight of baggage using a video sequence of the person carrying it. A spatial and a temporal cue are observable in such videos. The spatial cue is represented by the appearance of the person each time, and the temporal cue by the continuous change in his/her movements. When comparing people carrying heavy and light baggage, we do not observe a sufficient difference in the spatial cue because the weight of the baggage does not change her/his appearance. By contrast, we think that a significant difference in the temporal cue is observable. We hypothesize that this difference derives from changes in the center of gravity of the person carrying the baggage. The weight of baggage changes the center of gravity, causing the person to move the body to maintain pose.

In the field of ergonomics [5], [6], researchers have investigated whether there is a difference in movement when a walking person is carrying heavy or light baggage. They have reported that when a walking person carries baggage in one hand, the movement of the other hand increases with the weight of the baggage. Similar analytical studies have been performed on people in various categories [7], [8]. The relevant studies have used a motion capture system. By contrast, one study [9] employed video sequences instead of a motion capture system. These analytical studies assume that the person is walking, but this is not always the case. For example, while waiting for a security gate to open, an elevator to arrive, or a traffic light to change, a person stops walking and stands still. It is difficult in this case to directly apply knowledge gleaned from current studies.

We focus on ways of observing differences in movements related to the weight of baggage being carried by a person who is standing. People cannot completely stop their bodies from moving in such cases, and they slightly sway in different directions. This slight movement is called body sway [10], and occurs naturally even though they might consciously try to stop it. Body sway has significant attention as a temporal cue for a standing person. Researchers [11]–[13] have measured the center of gravity of swaying bodies using a force plate embedded on the floor. Kamitani et al. [14] used video sequences acquired from an RGB camera to extract features of body sway for identifying individuals. However, this method is not intended for personal attribute classification in case of a person carrying baggage. Moreover, subjects may object to the use of RGB cameras on privacy grounds because of the ease of identifying a person from images recorded using an RGB camera compared with using depth cameras.

To classify baggage being carried by a person standing as light or heavy, we propose here a method for extracting a feature of the person's body sway using a video sequence acquired from a depth camera. Our method estimates the person's head region from the depth images, and measures the movement caused by the body sway in each frame. We apply a frequency analysis technique to temporal changes in these movements and extract discriminative features to classify baggage according to weight. In experiments, our method obtained an accuracy of 95.0% in baggage weight classification on a dataset collected from 40 participants. To the best of our knowledge, this is the first study in video analysis that focuses on the use of body sway to classify baggage according to weight.

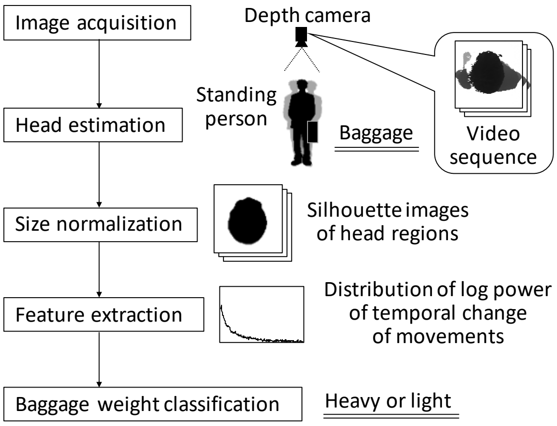

Fig. 1. Overview of our method.

II. Feature extraction using body sway

A. Overview

We consider a suitable setting of a depth camera to observe the body sway of a standing person. We set an overhead camera on a ceiling to avoid occlusion, whereby a person is hidden by other objects or people. The field of view of the depth camera contained the head and shoulders of the standing person and the floor as background. Our method aims to stably extract a feature of body sway from the movement of the head region, which is closest to the camera.

Figure 1 provides an overview of our method. We assume that a person stands below a depth camera and maintains his/her pose while waiting for a security gate to open. We acquire a video sequence of the standing person using the overhead camera. Our method estimates the head region from each frame of the video sequence and applies the size normalization of this region to represent the silhouette of the head. Finally, we extract a feature for baggage weight classification using the magnitude of body sway computed from a video sequence of normalized silhouette images. The details of head estimation are provided in II-B, those of size normalization in II-C, and those of feature extraction in II-D.

B. Estimating head region

We describe a method for estimating the head region of a standing person from depth images in a video sequence. First, we separate a region occupied by a person from a background region in the image. Because we can assume that the background floor does not move as the video is recorded, we set a fixed threshold for depth values and removed the background region. The range of the depth values corresponding to the person region was set to $[0, D_g - D_m]$, where D_g is the distance between the camera and the floor, and D_m is the margin.

Second, we estimate the head region using the person region obtained from the fixed threshold mentioned above. As described in Section II-A, the person region acquired from an overhead camera includes the head and shoulders of each person. Note that it is difficult to separate the head region from a shoulder region at a fixed depth value because people's heights vary. We thus need to set an adaptive threshold of depth values for each person. To this end, we use the histogram of depth values in the person region. In the camera setting, the head was closer to the depth camera than the shoulders. We could thus assume that the histogram of depth values in the person region had two peaks of distribution. Figure 2 shows an example of the histogram of depth values. We apply Otsu's method [15] to adaptively set the threshold according to the shape of the distributions. Finally, we set the head region using pixels with depth values smaller than the adaptive threshold. Our method outputs a video sequence of silhouette images in which the pixel value corresponding to the head is one and the other values are zero.

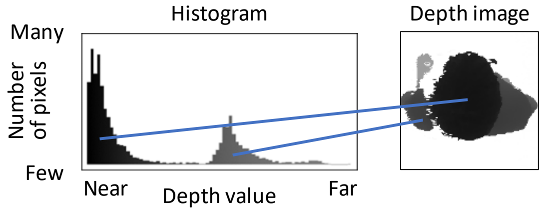

Fig. 2. Example of the histogram of the depth values in the person region.

C. Size normalization

The apparent size of the head region observed from the overhead camera depends on the height of the person. Figure 3(a) shows examples of the head regions of three persons. We see that the apparent size of the tall person is large, and vice versa. We believe that differences in the spatial cue caused by height may reduce the accuracy of baggage weight classification. To reduce the influence of these differences, we apply size normalization to the head regions in the silhouette images. We set the reference values H and W for the height and width of the head regions, respectively. Our method applies a rescaling technique so that the height and width of the head of each silhouette image are the same as the corresponding reference values. Figure 3(b) shows examples of silhouette images with the size normalization of the head regions.

D. Designing temporal feature

We extract a feature representing the temporal cue for baggage weight classification from a video sequence of normalized silhouette images. We expand an available method [14] for normalized silhouette images. The reference time at which the head is at the center is used because the head moves forward, backward, left, and right due to body sway. We determine the reference time by searching in the temporal direction for the frames most similar to each other in the video sequence of silhouette images of the head region. We compute the difference between the silhouette image at each time and the silhouette image at the reference time to obtain the magnitude of movement representing the magnitude of body sway. Finally, we apply power spectral density (PSD) to the temporal

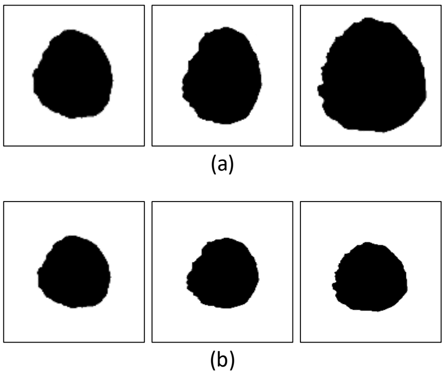

Fig. 3. Examples of silhouette images with the size normalization of the head regions.

Fig. 4. Baggage carried by participants when acquiring video sequences. (a) Light baggage weighing 1.7 kg and (b) heavy baggage weighing 11.7 kg.

change in movements to determine an N-dimensional feature corresponding to the components of frequency.

III. Experiments

A. Dataset

To confirm the effectiveness of our method, we collected a dataset of video sequences of people carrying baggage while standing. We prepared two attache cases that looked identical. Each weighed 1.7 kg. When nothing was in the attache case, it was regarded as light baggage. On the contrary, when a weight of 10 kg was put in the attache case, it was regarded as heavy baggage. We placed the weight at the center of the case. The interior of the attache case was filled with cushioning material to prevent the weight from moving. Figure 4(a) shows the light baggage and (b) the heavy baggage. They weighed 1.7 kg and 11.7 kg, respectively. Figure 5 shows examples of two people carrying the cases. We see that they appeared nearly identical in both cases regardless of the weight of the cases.

Forty participants (39 males and one female, with an average age of 22.1 ± 2.1 years and an average height of 171.2 ± 5.4 cm) participated in the study. Figure 6 shows the settings for the depth camera and the participants. We calibrated the depth camera so that the optical axis was perpendicular to the floor. Its resolution was set to 512×424 pixels. The sampling frequency was 30 Hz. The parameters described in Section II-B were set to $D_g = 2.4$ m, and $D_m = 0.5$ m. Before shooting a video sequence, we asked each participant to align the arch of his/her foot with markers on the floor and hold the attache case in the right hand. The time taken to acquire

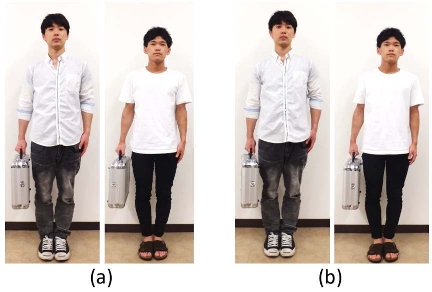

Fig. 5. Examples of people carrying attache cases. (a) Light baggage and (b) heavy baggage.

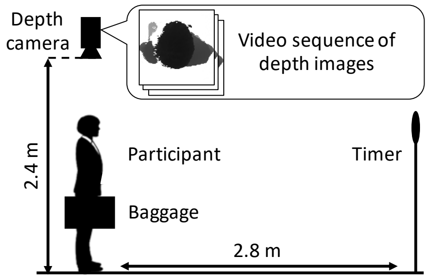

Fig. 6. The settings of the depth camera and the participant.

one video sequence of depth images was set to 180 seconds. From the beginning to the end of the acquisition, we asked the participants to maintain an upright posture (Romberg posture) and look at the clock placed in front of them. To reduce fatigue on the arm holding the case, each participant first carried light baggage for a video sequence and the heavy baggage for another sequence. We acquired 40 video sequences for light baggage and 40 for heavy baggage.

B. Accuracy of baggage weight classification

We evaluated the accuracy of baggage weight classification. The parameters of the size normalization described in Section II-C were set to $H = 100$ and $W = 100$ pixels. The dimensionality of the feature used in our method was $N = 128$. We used the linear support vector machine (SVM), and computed the rate of correct classification of baggage as light and heavy using leave-one-participant-out evaluations.

The accuracy of our method for baggage weight classification was 95.0%. We compared its performance with that of a method based on long short-term memory (LSTM) [16] using temporal changes in movements in the training samples. This temporal change represents the one-dimensional signal before extracting the PSD feature. The accuracy of LSTM was 51.4%. This confirms that our feature extraction is superior to that of the conventional deep learning technique designed for time series analysis.

As described in Section II-C, our method normalized the apparent size of the head regions in the silhouette images. We compared the effect of this size normalization. The accuracy of our method without normalization was 77.5% whereas that with normalization was 95.0%. This confirms that the normal-

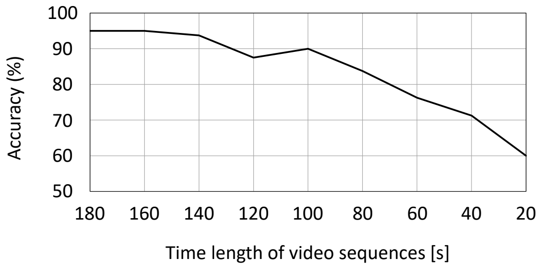

Fig. 7. Accuracy of baggage weight classification when changing the length of video sequences.

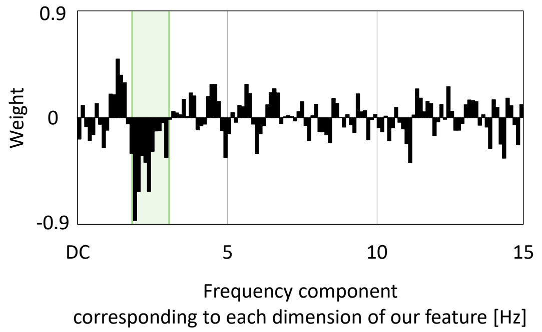

Fig. 8. The weight vector of the linear SVM for baggage weight classification. The positive weight helps to identify light baggage, and the negative weight helps to identify heavy baggage. The light green area corresponds to the frequency band from 2.0 Hz to 3.0 Hz.

ized apparent size is effective in improving the performance of the method at baggage weight classification.

Figure 7 shows the accuracy of baggage weight classification when reducing the length of the video sequences from 180 seconds. When it was reduced to 100 or 60 seconds, the performance reduced by 7.5 and 18.7 percentage points, respectively. We believe that the degradation in performance is related to the periodicity of body sway. We should conduct further examinations to reduce the length of video needed for correct classification in future work.

C. *Visualizing weight of SVM*

To investigate the important components of our features for baggage weight classification, we visualize the weight vector of the linear SVM in Figure 8. Each dimension of our feature corresponded to the frequency component from DC to 15.0 Hz. We see that large negative weights appeared in the frequency components from 1.8 Hz to 3.0 Hz. In particular, a remarkable difference between heavy and light baggage was observed in the frequency of 2.1 Hz. We think that these components are informative features because they increase when the person is carrying heavier baggage.

IV. Conclusions

We proposed a method for extracting the feature of body sway from a person standing while carrying a bag by using a depth camera to classify it as heavy or light. We observed temporal changes in the movement of the head region and computed temporal features by using a frequency analysis technique. We collected a dataset of video sequences of 40 participants and confirmed that the accuracy of baggage weight classification was high, at 95.0%. Furthermore, we showed that there was a significant difference in the frequency components of waves of body sway from 2.0 Hz to 3.0 Hz, and this was related to the weight of the baggage. In future work, we will further evaluate our method on datasets of various types featuring people carrying different weights. We will also explore the use of the regression technique instead of the classification technique. **Acknowledgment** This work was partially supported by JSPS KAKENHI under JP18H04114.

References

[1] 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*, 2017, pp. 1131–1140.

[2] 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.

[3] D. Damen and D. Hogg, "Detecting carried objects in short video sequences," in *Proceedings of European Conference on Computer Vision*, 2008, pp. 154–167.

[4] 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*, 2012, pp. 398–403.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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*, 2005, pp. 15–22.

[10] 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.

[11] T. E. Prieto, J. B. Myklebust, R. G. Hoffmann, E. G. Lovett, and B. M. Myklebust, "Measures of postural steadiness: differences between healthy young and elderly adults," *IEEE Transactions on biomedical engineering*, vol. 43, no. 9, pp. 956–966, 1996.

[12] G. L. Cavalheiro, M. F. Almeida, A. A. Pereira, and A. O. Andrade, "Study of age-related changes in postural control during quiet standing through linear discriminant analysis," *BioMedical Engineering OnLine*, vol. 8, no. 35, 2009.

[13] Y. Yu, H. C. Chung, L. Hemingway, and T. A. Stoffregen, "Standing body sway in women with and without morning sickness in pregnancy," *Gait & posture*, vol. 37, no. 1, pp. 103–107, 2013.

[14] 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.

[15] 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.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.