Center of Mass Estimation for Balance Evaluation Using Convolutional Neural Networks

Wenchuan Wei
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA, USA
w8wei@eng.ucsd.edu

Sujit Dey
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA, USA
sdey@ucsd.edu

***Abstract*—The Center of Mass (CoM) position of the human body is an important indicator when evaluating a person's balance ability. Traditionally the CoM position is measured using laboratory-grade devices like a force plate, which is expensive and inconvenient for home use. In this paper, we propose a deep learning-based framework that uses a single depth camera to estimate the CoM position of a human subject. The proposed framework takes the depth image captured by the depth camera as input, and uses supervised learning to estimate the subject's horizontal CoM position. The model is trained and tested on data collected from multiple subjects in various postures. Evaluation results demonstrate the high accuracy of the proposed approach in estimating the CoM of existing subjects or a new subject. Compared with existing CoM estimation techniques, the proposed framework is easy to set up and does not need any subject identification process, which makes it convenient for home use. The proposed framework can be used as a portable and low-cost tool for CoM measurements and can enable automated balance evaluation at home.**

Keywords—Center of Mass, Center of Pressure, Convolutional Neural Networks, Balance Evaluation, Deep Learning

I. INTRODUCTION

The Center of Mass (CoM) position of the human body is an important indicator when evaluating the balance ability. For the 3D position of human's CoM, the horizontal CoM (i.e., the projection of CoM on the ground) is even more important and often used to evaluate the progress of rehabilitation programs, predict fall risk for people with mobility problems [1, 2], etc. For example, the static body sway (i.e., the range of horizontal CoM in static upright posture) is a clinically relevant activity parameter to assess postural balance across a wide spectrum of patient populations [3]. Since the CoM position of the human body cannot be directly measured, the Center of Pressure (CoP) of the ground reaction force is measured instead. According to Newton's second law, the CoP should coincide with the horizontal CoM position when the subject is in a stable posture. Traditionally, a laboratory-grade force plate is used to measure the CoP. However, due to its high cost and complicated setup procedure, the force plate is primarily limited to laboratory use. The Wii Balance Board (WBB) is a device designed by Nintendo for balance-related games. It can calculate the CoP from the vertical ground reaction forces measured by four pressure sensors placed at its four corners. Bartlett et al. have validated that the error of CoP measurements by the WBB is within 5 *mm* [4]. Because of its low cost, portability, and high accuracy in CoP measurement, the WBB has been increasingly used as a replacement of the force plate in many studies [5, 6].

However, the method of using a force plate or WBB to measure the CoP and taking it as the horizontal CoM works only when the force plate or WBB is placed on a horizontal and firm plane, which limits its application. In balance evaluation, we often need to test the subject's balance ability on different surface types (e.g., the incline ramp, or the foam). Based on the fact that the CoM position of the human body is determined by some body parameters (e.g., body shape and density) and pose, people have proposed to use body parameters and pose to estimate the CoM position. Initially, the kinematic method is proposed by Winter [7] to estimate the CoM of the whole body as the weighted sum of the CoM of body segments. The weight of each segment is taken from previous anthropometric studies and therefore not personalized for each subject. To enable subject-specific CoM estimation, Chen et al. propose to measure the size of each body segment with a measuring tape and use an optimization method to estimate the density of each segment [8]. However, modeling the human body as geometrical segments (e.g., frustum) is not accurate. Later, Cotton et al. propose the Statically Equivalent Serial Chain (SESC) model for CoM estimation [9]. The body parameters of each subject are estimated from an identification/calibration process, for which the subject needs to perform multiple static postures. To achieve identification-free CoM estimation, Kaichi et al. recently propose a voxel reconstruction approach [10], where five cameras are used to reconstruct the subject's 3D body and the CoM is estimated by assigning weights to all body parts. However, the five cameras need to be carefully calibrated, which limits its application for home use.

With the rapid development of computer vision technologies in recent years, more and more vision-based models have been proposed to learn and predict some human-related activities from images or videos. Kahou et al. propose to recognize the facial expression of a human from a video sequence [11]. Alarrai et al. develop a fall prediction framework for the elderly using a depth camera [12]. Inspired by the vision-based techniques, we propose to learn the body parameters (size, density, etc.) from a depth image of the subject and use deep learning to estimate the horizontal CoM position of the subject. We have selected the depth camera instead of a RGB camera because the depth map captured by the depth camera provides more information about the subject's pose in the depth direction, which is essential in CoM estimation. Besides, depth cameras are color and texture

invariant and work in low light conditions [13]. Fig. 1 shows the architecture of our proposed CoM estimation framework, which is built using Convolutional Neural Networks (CNN) and trained using data collected from multiple subjects in various postures. A WBB is used to measure the ground-truth CoP which is equivalent to the horizontal CoM in stable postures. Please note that this paper discusses the CoM estimation for stable postures. We will extend the proposed technique to estimate the CoM position for unstable postures in our future work. The WBB is used only for collecting the ground-truth CoP in the training process. Once the model is trained, only the depth camera is needed to estimate the subject-specific CoM position for home or clinical use. The depth camera (e.g., Microsoft Kinect) is anyway necessary in most automated training systems for its ability in skeleton tracking and motion capture [14, 15]. By using the CoM estimation model proposed in this paper, the CoM position can also be tracked without any extra device. Evaluation results demonstrate the high accuracy of the proposed method in estimating the CoM of existing subjects or a new subject. By using a single depth camera that does not need complicated calibration or subject identification, the proposed framework can be used as a portable and low-cost tool for CoM measurements and therefore enable automated balance evaluation at home.

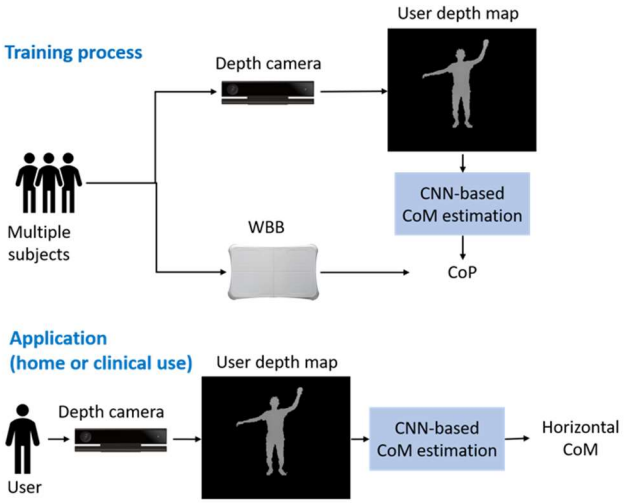

Fig. 1. The training and application process of the proposed CoM estimation framework.

The remainder of this paper is organized as follows: Section II reviews related work about CoM estimation in more details. Section III introduces the devices used in the proposed framework. In Section IV, we discuss the details in the proposed CoM estimation approach. Section V presents the experimental results. Section VI concludes the paper and discusses future work.

II. Related Work on CoM Estimation

While we have briefly discussed the related work on CoM estimation in Section I, we next explain the most relevant techniques in more details, pointing out their disadvantages and the need and differentiation of our proposed technique.

Winter's method: Traditionally, people use the kinematic method proposed by Winter [7] to estimate the CoM of a human. The human body is represented by 16 segments and the position of each segment is tracked by a marker-based motion capture system. The CoM of the whole body is calculated as

$$CoM = \frac{1}{M}\sum_{i=1}^{16} m_i \cdot CoM_i \quad (1)$$

where m_i and CoM_i are the mass and CoM of the i-th segment. The information needed for this calculation is taken from previous anthropometric studies. However, such information may differ in subjects of different age, sex, and fitness level, etc. Therefore, this method is not able to provide subject-specific CoM estimation.

The optimization-based method: To achieve subject-specific CoM estimation, Chen et al. propose to model the human body as some geometric shapes (e.g., modeling the neck as a frustum) and measure the proximal and distal circumference lengths for each segment with a measuring tape [8]. They use the Vicon motion capture system [16] to track the subject's kinematic data during static postures and force plates are used to measure the CoP as the ground-truth horizontal CoM. Then the body parameters of a subject are calculated using an optimization-based method and used for the estimation of CoM. This method requires the measurements of body size for each subject, which is inconvenient. Moreover, modeling the body segments as geometrical shapes (e.g., frustum) is not accurate.

The SESC model: Cotton et al. propose the SESC model which translates the human's mass distribution to the geometry of a linked chain [9]. The subject-specific SESC parameters are obtained in an identification phase, for which the subject should perform 14 static postures. The posture is tracked by the motion capture system Vicon [16] and the horizontal CoM position is measured by a force plate. Gonzalez et al. propose that the estimation error of this method can be further reduced by assuming the bilateral symmetry of the human body and using an identification phase with 40 static postures [17]. They have also conducted comprehensive study using low-cost sensors Kinect and WBB that can be easily set up inside a patient's home. The estimation errors using Kinect and WBB have been shown to be comparable to those obtained using high-end equipments. Later, Conzalez et al. propose to use a Kalman filter and visual feedback in the identification phase to further reduce the estimation error [18]. However, the identification phase required by the SESC method (about 8 minutes) needs to be conducted each time when a new subject comes or the mass distribution of an existing subject has changed, which limits its application.

Voxel reconstruction method: To avoid complicated identification phase on each subject, Kaichi et al. propose the voxel reconstruction approach [10], where five cameras are used to capture multiple views of the human body. Then they use a 3D reconstruction approach to reconstruct the subject's body and further segment the body into 9 parts. The CoM of the whole body is estimated by assigning weights to different body parts. The weights of each part are from previous anthropometric studies. As discussed earlier, the difference of body size and density in different subjects are the main challenges in the subject-specific CoM estimation. By reconstructing the 3D body, this method solves the problem of difference in body shapes. However, it still fails to consider the difference of body density since it uses the density information from previous studies. Moreover, it uses five

cameras that need to be carefully calibrated for 3D construction, which is not convenient for home use. In comparison, our proposed framework uses a single depth camera and does not need any complicated calibration or subject identification process. The body parameters of each subject can be learned through supervised learning and used to estimate the subject-specific CoM.

III. Devices: Kinect and Wii Balance Board

A. Depth Camera in Kinect

Kinect is a motion capture sensor that consists of a RGB camera and a depth camera [19]. The depth map (424 by 521 pixels) captured by its depth camera represents the distance of each pixel from the sensor. Based on the depth map, we use the method proposed in [13] to remove the background to obtain the user depth map, and extract the user skeleton that is composed of 25 joints. Fig. 2 shows an example of the original depth map and the user depth map with the skeleton overlay.

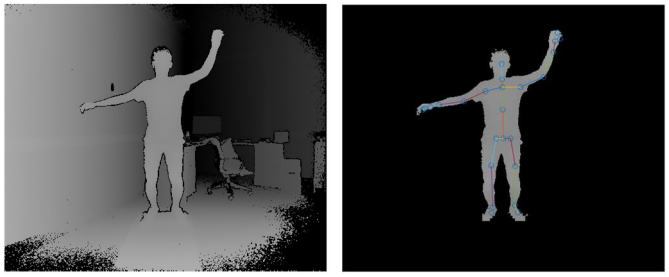

Fig. 2. Depth map captured by the depth camera of Kinect. Left: full depth map. Right: user depth map and the skeleton overlay.

B. Wii Balance Board

The Wii balance board (WBB) consists of four pressure sensors located at the four corners of the board (see Fig. 3). When a subject stands on the board, the four pressure sensors measure the vertical force applied to the four corners and the Center of Pressure (CoP) can be calculated. Fig. 3 shows the WBB and the coordinate system we use in this paper. Suppose that the board is placed on a horizontal ground. We define the x and y axis as the length and width direction of the board, and the z axis in the upright direction. The origin is located at the center of the board. Given the forces/pressures that the four sensors measure, the subject's CoP (x, y) coordinate can be calculated as

$$x = \frac{L}{2} \times \frac{(P_2 + P_4) - (P_1 + P_3)}{P_1 + P_2 + P_3 + P_4} \tag{2}$$

$$y = \frac{W}{2} \times \frac{(P_1 + P_2) - (P_3 + P_4)}{P_1 + P_2 + P_3 + P_4} \tag{3}$$

where L and W are the length and width of the board, and P_i is the force measured by the i-th pressure sensor. Note that all postures discussed in this paper are stable postures, for which the CoP position measured by the WBB is equivalent to the horizontal CoM position of the subject. Therefore, we will use horizontal CoM position to refer to the CoP position measured by the WBB in the rest of this paper.

Several studies have explored the accuracy of the WBB in CoP measurement and found that the CoP location obtained by a WBB and a laboratory-grade force plate are fairly similar, with an offset difference smaller than 5 *mm* [4, 20]. Besides, the other advantages of the WBB such as low cost and portability makes it a good tool for CoP measurements for clinical or home use. Therefore, we use the horizontal CoM measurements provided by the WBB as the ground truth to train our model.

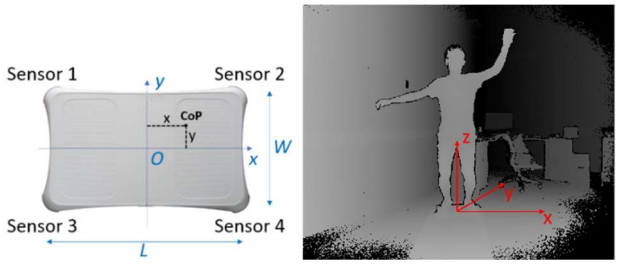

Fig. 3. Left: the WBB and its four pressure sensors. Right: the 3D coordinate system shown in the depth image.

IV. Center of Mass Estimation From a Depth Image

A. Input and Output of the Model

As discussed in Section I, we would like to develop a model that takes the full depth map captured by a depth camera as input and estimates the horizontal CoM position of the subject. To train the model, we have collected the depth maps and the ground-truth CoM positions measured by the WBB from multiple subject with different stable postures (see Fig. 1). (More details about the data collection process will be introduced in Section V-A). The depth map and the ground-truth CoM position of a subject in a frame constitute a data sample: (Depth map, CoM_x, CoM_y). In the training process, we use all collected samples to train and validate the CoM estimation model.

Details about the CoM estimation model is shown in Fig. 4. We use the algorithms proposed in [13] to extract the user depth map and the user skeleton with 25 body joints from the full depth map. To help the model distinguish between different body parts (since different body parts may have different densities), we propose to provide the model with information about the joint positions, which are represented by joint heatmaps. The heatmap of a joint has the same size of the depth map and each pixel in the heatmap represents the probability of the joint located at this position (see Fig. 4). In our proposed framework, we use 2-D Gaussian distribution to calculate the probability. The two dimensions of the Gaussian distribution are assumed to be independent and have equal standard deviation σ_0. Following the coordinate system shown in Fig. 3, the width and height direction of the depth image are the x and z axis. For a joint i ($1 \le i \le 25$), we use the algorithm proposed in [13] to obtain the 2-D position of this joint on the depth map as (x_i, z_i). Its heatmap $H_i(x, z)$ is calculated as

$$H_i(x, z) = \frac{1}{2\pi\sigma_0^2} e^{-\frac{(x - x_i)^2 + (z - z_i)^2}{2\sigma_0^2}}. \tag{4}$$

Fig. 4 shows the heatmaps of two joints (spine_mid and right_foot) as an example. Then a CNN-based model takes the user depth map and the joint heatmaps as input and estimate the CoM position of the subject, which is the output of the model. As discussed in Section III-B, the horizontal CoM coordinates measured by the WBB are continuous values (x, y), which make

the CoM estimation a regression problem. However, Tompson et al. have shown in the problem of pose estimation that direct regression of pose coordinates from images is a highly non-linear problem and may be difficult to learn the mapping [21]. To improve learning, we propose to quantify the CoM coordinates into discrete values. The CoM coordinates are discretized uniformly in the x and y direction, resulting in N_x x N_y classes. By discretizing the continuous CoM coordinates, we cast the highly non-linear problem of direct CoM coordinate regression to a more manageable form of prediction in a discretized space.

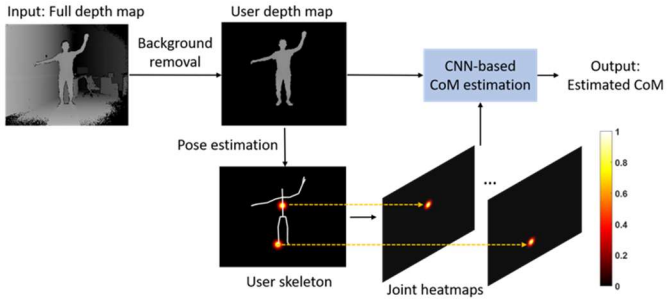

Fig. 4. The proposed CoM estimation Framework.

B. Data Augmentation

In deep learning, an effective way to improve learning and reduce overfitting is increasing the amount and diversity of training data, which is called data augmentation. Traditional data augmentation methods used in computer vision include flipping, rotating, translating the image, and adding random noise to the image. The above techniques work great in image classification problems as they will not change the image categories. However, the CoM position of the subject may be different by using these operations. For example, a subject who is leaning to the left has a positive x value in his CoM position. If we flip his depth image, he would be leaning to the right and the x value of his CoM will be different. Therefore, the traditional data augmentation techniques cannot be directly applied in our dataset. To solve this problem, we propose to train two different models for the x and y component of the CoM separately. Different data augmentation methods are applied to the two components as follows.

***x* component of the CoM:** 1) Adding a random depth value to the user body in the user depth map. This operation is identical to shifting the user body in the y direction and will not change the x value of the CoM. 2) Randomly shifting the user body in the user depth map in z direction. This operation will also not change the x value of the CoM.

***y* component of the CoM:** 1) Randomly shifting the user body in the user depth map in the x direction. 2) Randomly shifting the user body in the user depth map in z direction. Both operations will not change the y value of the CoM.

Note that the user's joint positions (and the joint heatmaps) also need to be processed in the same manner as the user body (i.e., shifting the same amount and adding the same depth value).

C. CNN-based Network Architecture

For the CoM estimation model, we propose to use convolutional neural networks, which is widely used in computer vision problems for its advantages in parameter sharing, feature extraction [22], etc. The proposed convolutional unit is shown in Fig. 5, which consists of a Convolutional (Conv) layer [23], a Batch Normalization (BN) layer [24], a Rectified Linear Unit (ReLU) layer, and a max Pooling layer.

The complete network architecture is also shown in Fig. 5. We use five Conv units to extract features from the original depth images. As discussed in Section IV-A, the joint heatmaps are also used as input of the model. The joint heatmaps are downscaled and concatenated with the activation map after the second Conv unit. We choose the activation map after the second Conv unit instead of the original user depth map because it contains higher-level features of the subject's posture. After the five Conv units, we use two Fully Connected (FC) layers, with the first one followed by a BN and a ReLU layer and the second one followed by an Argmax layer to predict the class of the discretized CoM. As discussed in Section IV-A, the continuous CoM coordinates are discretized into some classes, so the proposed model will estimate the correct class of the CoM coordinates.

The loss function of this problem is defined as the cross entropy of the ground-truth CoM class and the predicted class of the CoM as

$$Loss = -\sum_{i=1}^{N} L_i \cdot \log(S_i) \tag{5}$$

where L_i is the encoding for class i in the ground-truth CoM and S_i is the softmax output of class i in the estimated CoM. However, unlike traditional image classification problems that use one-hot encoding for the ground-truth label, we propose to use Gaussian-distributed heatmaps for the following reasons. Suppose k is the ground-truth class for a sample (e.g., the x value in a subject's ground-truth CoM is discretized as class k), its one-hot encoding is

$$L_i = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}. \tag{6}$$

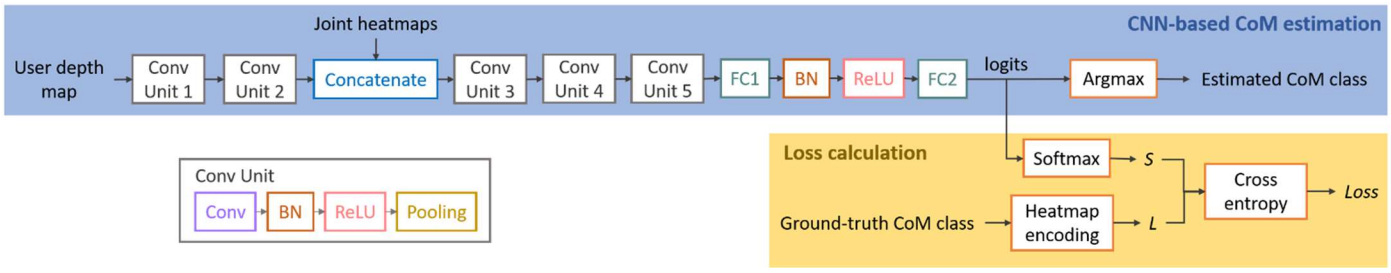

Fig. 5. The proposed network for CoM estimation.

Fig. 6 gives an example of the one-hot encoding: only the correct class k is encoded as 1 and all the other classes are encoded as 0. In image classification problems, the ground-truth label for an image is a categorical feature and all the incorrect classes ($i \neq k$) should be considered equally. Therefore, the one-hot encoding is an effective way to encode the ground-truth label. However, in the problem of CoM estimation, the ground-truth class of CoM is discretized from its continuous value, so the incorrect classes should be penalized differently based on their distance to the ground-truth/correct class. Therefore, we define a Gaussian-distributed heatmap to encode the ground-truth CoM as

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(i-k)^2}{2\sigma^2}}, \quad (7)$$

where σ is the standard deviation of the Gaussian distribution. Fig. 6 shows an example of the Gaussian heatmap. The probability of the true class k has the highest value 0.20 and all the other classes are encoded based on their distance to the true class k. The CoM heatmap represents the confidence of each class as the ground truth. In this way, the network can be trained to adjust its output to get closer to the true class in the learning process.

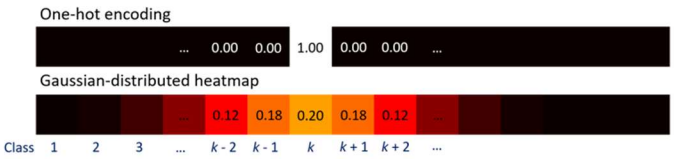

Fig. 6. One-hot encoding and Gaussian-distributed heatmap for the CoM ground-truth (k is the true class).

V. Experimental Results

This section evaluates the performance of the proposed CoM estimation framework. First, we will introduce the data collection process. Second, we present the implementation details we use when training the model. Finally, we show the experimental results on the collected dataset.

A. Data Collection

In order to obtain a comprehensive dataset that covers as many postures as possible, we define the following postures on three body parts.

Trunk: keep it upright, or lean to the left/right/front/back with different angles.

Legs: squat with different angles, stand on one leg.

Arms: different positions of the left and right arm.

We have collected data from 21 subjects (age 23 ~ 62, 13 males, 8 females). Each subject was asked to stand on a WBB and slowly move the body to cover different postures on the three body parts, while maintaining his/her balance. The WBB recorded the CoP position, which is equivalent to the horizontal CoM position as discussed earlier. A Kinect sensor was placed in front of the subject to capture the depth images. The WBB and Kinect was synchronized to record the depth map and the horizontal CoM on the same timestamp, with a framerate of 30 frames/second. The depth map and the corresponding horizontal CoM position (CoM_x and CoM_y) in a frame constitute a data sample. From the 21 subjects, we have collected about 65,000 data samples in total.

B. Implementation Details

The depth image captured by the depth sensor of Kinect has the resolution of 424 × 512. Pixels in the depth map have the value range of [0, 1], with 0 representing the background and positive value representing the normalized depth of the pixel. For data augmentation, we use the range of [-40, 40] (pixels), [-15, 15] (pixels), and [-0.2, 0.2] (depth value) for the random shift on x, y, and z (depth) direction (see Section IV-B). (We have selected these values as the range of the random shift to make sure that the user body will be not shifted out of the depth image.) For CoM discretization, we use 4 *mm* and 2 *mm* as the precision of the uniform discretization in the x and y direction. With the size of the WBB as 431 × 236 *mm*, the discretization leads to 108 and 118 classes for the x and y component of CoM, respectively. In the heatmap of the ground-truth CoM, we use Gaussian distribution with standard deviation of 3 and 2, in the x and y direction.

The CNN-based network includes five Conv units. We use 8, 16, 32, 64, 128 channels for the Conv layer in the five Conv units respectively. The number of channels is selected empirically to extract as much features from the images as possible while not introducing too many parameters. For the BN layer, we use 0.9 as the BN momentum. The first FC layer has 246 neurons and the second FC layer has N neurons where N is the number of discretized CoM classes (N_x =108 and N_y = 118). When training the network, we use the batch size of 64 and a learning rate of 5e-4. The Adam optimizer [25] is used to minimize the cross entropy loss defined in (5).

C. CoM Estimation Results

Evaluation metrics: To evaluate the performance of the proposed model, we calculate the Root Mean Squared Error (RMSE) between the ground-truth CoM class and the estimated class as

$$RMSE = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left\|G_j - E_j\right\|^2} \quad (8)$$

where G_j and E_j are the ground-truth class and the estimated class for sample j and M is number of samples. Since the RMSE represents the average error in estimating the CoM class, we define the average error in CoM estimation as the product of the RMSE and the discretization precision (i.e., 2 *mm* and 4 *mm* for the x and y direction).

To validate the proposed CoM estimation approach, we use two different modes to train and test the proposed model. **i) Test on existing subjects.** All samples collected from the 21 subjects are randomly split into a training set (64%), a validation set (16%) and a test set (20%). The model is trained on the training set and parameters that produce the best performance (i.e., the lowest loss) on the validation set are selected as the optimal parameters. Then the model with the selected optimal parameters is applied on the test set. **ii) Test on a new subject.** In this mode, we would like to evaluate the model's performance on a new subject whose data have never been used for training and validation. Samples from 20 subjects are randomly split into a training set (80%) and

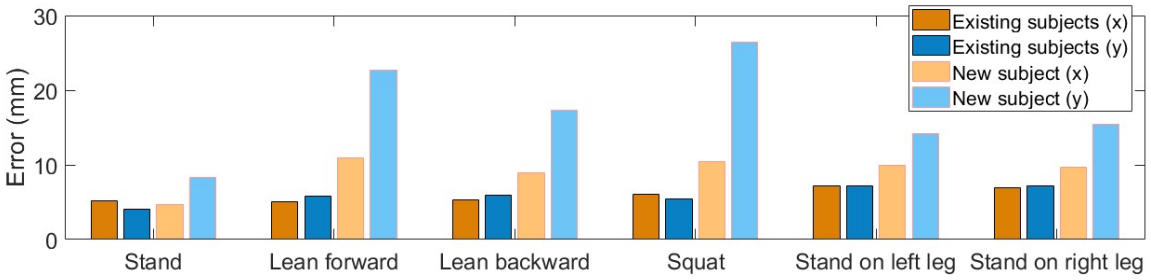

Fig. 7. The CoM estimation errors on different postures using our proposed approach.

a validation set (20%). Data of the 21st subject are used as the test set. We compare the results of our proposed CNN-based approach with two state-of-the-art methods: the SESC method [9] and the voxel reconstruction method [10], using the estimation errors reported in their paper. The results are shown in Table I. (For the voxel reconstruction method, the authors do not report the estimation error on each direction and only the overall error on four postures is provided, which we summarize as 8 ~ 15 *mm*.) To highlight the advantages of our proposed approach, we also summarize the requirements of each method in Table I.

TABLE I. CoM Estimation Error and Requirements of Different Methods

Method	***Error (mm)***		***Requirements***
	x	*y*	
SESC [9]	17	23	Motion capture sensor. 8-minute identification needed for each new subject.
Voxel reconstruction [10]	8 ~ 15		Five cameras. Camera calibration and synchronization needed.
Ours (on existing subjects)	6.0	9.3	Single depth camera. No calibration or identification needed.
Ours (on a new subject)	8.9	17.2	

From Table I we can see that our proposed approach achieves the lowest estimation errors when testing on existing subjects. For a new subject, the estimation errors achieved by our method are a little bit higher, but still outperform the SESC method in both *x* and *y* directions. Moreover, the SESC method requires an 8-minute identification phase for each new subject, which is not convenient for home use. For example, the body parameters learned from previous identification phase for an existing subject may be inaccurate if the subject gains or loses weight. In comparison, our proposed approach does not need any subject identification process and is much more convenient for home use. For the voxel reconstruction method [10], we attribute its good performance to the use of multiple cameras. By using a single depth camera, our proposed approach captures only the front view of the human body and the shape of the back or side of the body is ignored. The voxel reconstruction method [10] uses five cameras to capture different views and reconstruct the full 3-D user body, therefore achieving good performance on the CoM estimation. However, the complicated calibration and synchronization among the five cameras makes it only suitable for laboratory use. In comparison, our proposed approach uses a single portable and inexpensive depth camera, which is convenient for home and clinical use, while achieving comparable accuracy results.

To show the performance of our proposed approach on different postures, we classify the collected postures into six categories: standing, leaning forward, leaning backward, squatting, standing on the left leg, and standing on the right leg. For each posture type, the proposed CNN-based CoM estimation model is tested on existing subjects and a new subject separately. The estimation errors in the *x* and *y* direction are shown in Fig. 7. For existing subjects, we can see that low estimation errors (less than 10 *mm*) can be achieved in both *x* and *y* directions for all the postures. For a new subject, the estimation errors in the *y* direction (i.e., the depth direction) is higher than the errors in the *x* direction. It is because only the front view of the user body can be captured by the single camera and the shape of the back of the user body may affect the CoM position on the depth direction. (The *x* component of the CoM is less affected due to the symmetry of the human body in the *x* direction.) Comparing the results on different postures, we can find that the estimation error is higher for squatting, leaning forward, and leaning backward than the other postures, which may be due to the occlusion problem. For example, the arm may occlude some parts of the trunk when the subject is squatting. In this case, the depth information of body parts that are occluded cannot be captured by the depth camera and the precision of the CoM estimation model will be affected.

VI. Conclusions and Future Work

In this paper, we propose a deep learning-based approach to estimate the horizontal CoM position of humans. The model takes the depth map captured by a single depth camera as input and the CNN-based network is trained to estimate the horizontal CoM position from the depth map. We have collected data from multiple subjects, with their depth images captured by a Kinect camera and the horizontal CoM position recorded by a WBB. Experiments demonstrate the superiority of our proposed approach over other CoM estimation techniques. In additional to the low estimation error, our proposed CoM estimation approach does not need any subject identification process, which is convenient for home and clinical use. For future work, we would like to explore the CoM estimation for unstable postures. Besides, we plan to extend the current CoM estimation framework to a balance evaluation system, to enable quantification of the balance ability in patients with mobility problems.

Acknowledgment

This material is based upon work supported by the National Science Foundation under grant No. IIS-1522125.

REFERENCE

[1] F. Wang, M. Skubic, C. Abbott, and J. M. Keller, "Body sway measurement for fall risk assessment using inexpensive webcams," *Proceedings of the IEEE Conference on Engineering in Medicine and Biology Society* (EMBC 2010), Buenos Aires, Argentina, Sep. 2010.

[2] H. G. Kang, L. Quach, W. Li, and L. A. Lipsitz, "Stiffness control of balance during dual task and prospective falls in older adults: The MOBILIZE Boston Study," *Gait & posture* 38.4 (2013): 757-763.

[3] A. K. Mishra, et al. "Examining methods to estimate static body sway from the Kinect V2. 0 skeletal data: implications for clinical rehabilitation," *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (PervasiveHealth 2017), Barcelona, Spain, May 2017.

[4] H. L. Bartlett, L. H. Ting, and J. T. Bingham, "Accuracy of force and center of pressure measures of the Wii Balance Board," *Gait & posture* 39.1 (2014): 224-228.

[5] P. Jogi, A. Zecevic, T. J. Overend, S. J. Spaulding, and J. F. Kramer, "Assessing and training standing balance in older adults: a novel approach using the 'Nintendo Wii'Balance Board," *Gait & posture* 33.2 (2011): 303-305.

[6] J. D. Holmes, M. E. Jenkins, A. M. Johnson, M. A. Hunt, and R. A. Clark, "Validity of the Nintendo Wii® balance board for the assessment of standing balance in Parkinson's disease," *Clinical Rehabilitation* 27.4 (2013): 361-366.

[7] D. A. Winter, *Biomechanics and motor control of human movement*. John Wiley & Sons, 2009.

[8] S. C. Chen, H. J. Hsieh, T. W. Lu, and C. H. Tseng, "A method for estimating subject-specific body segment inertial parameters in human movement analysis," *Gait & posture* 33.4 (2011): 695-700.

[9] S. Cotton, A. P. Murray, P. Fraisse, "Estimation of the center of mass: from humanoid robots to human beings," *IEEE/ASME Transactions on Mechatronics* 14.6 (2009): 707-712.

[10] T. Kaichi, et al. "Estimation of Center of Mass for Sports Scene Using Weighted Visual Hull," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW 2018), Salt Lake City, UT, USA, Jun. 2018.

[11] S. Ebrahimi Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal, "Recurrent neural networks for emotion recognition in video," *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (ICML 2015), Seattle, Washington, USA, Nov. 2015.

[12] R. Alazrai, Y. Mowafi, and E. Hamad, "A fall prediction methodology for elderly based on a depth camera," *Proceedings of the IEEE Conference on Engineering in Medicine and Biology Society* (EMBC 2015), Milan Italy, Nov. 2015.

[13] J. Shotton, et al. "Efficient human pose estimation from single depth images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.12 (2013): 2821-2840.

[14] W. Wei, Y. Lu, E. Rhoden, and S. Dey, "User performance evaluation and real-time guidance in cloud-based physical therapy monitoring and guidance system," *Multimedia Tools and Applications* (2017): 1-31.

[15] W. Wei, C. McElroy, and S. Dey, "Human Action Understanding and Movement Error Identification for the Treatment of Patients with Parkinson's Disease," *Proceedings of the IEEE International Conference on Healthcare Informatics* (ICHI 2018), New York City, USA, June 2018.

[16] Vicon. [Online]. Available: https://www.vicon.com/

[17] A. González, M. Hayashibe, V. Bonnet, and P. Fraisse, "Whole body center of mass estimation with portable sensors: Using the statically equivalent serial chain and a Kinect," *Sensors* 14.9 (2014): 16955-16971.

[18] A. González, P. Fraisse, and M. Hayashibe,"Adaptive interface for personalized center of mass self-identification in home rehabilitation," *IEEE Sensors Journal* 15.5 (2015): 2814-2823.

[19] Kinect. [Online]. Available: www.xbox.com/en-US/kinect

[20] A. Huurnink, D. P. Fransz, I. Kingma, and J. H. van Dieën, "Comparison of a laboratory grade force platform with a Nintendo Wii Balance Board on measurement of postural control in single-leg stance balance tasks," *Journal of biomechanics* 46.7 (2013): 1392-1395.

[21] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," *Advances in neural information processing systems* (NIPS 2014), Montreal, Canada, Dec. 2014.

[22] Y. LeCun, and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks* 3361.10 (1995): 1995.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems* (NIPS 2012), Lake Tahoe, USA, Dec. 2012. pp. 1097-1105. 2012.

[24] S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167* (2015).

[25] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*(2014).