Image-Based Stability Quantification

Jesse Scott, *Member, IEEE*, John Challis, Robert T. Collins, *Senior Member, IEEE*, and Yanxi Liu, *Senior Member, IEEE*

***Abstract*—Quantitative evaluation of human stability using foot pressure/force measurement hardware and motion capture (mocap) technology is expensive, time consuming, and restricted to the laboratory. We propose a novel image-based method to estimate three key components for stability computation: Center of Mass (CoM), Base of Support (BoS), and Center of Pressure (CoP). Furthermore, we quantitatively validate our image-based methods for computing two classic stability measures, CoMtoCoP and CoMtoBoS distances, against values generated directly from laboratory-based sensor output (ground truth) using a publicly available, multi-modality (mocap, foot pressure, two-view videos), ten-subject human motion dataset. Using Leave One Subject Out (LOSO) cross-validation, experimental results show: 1) our image-based CoM estimation method (CoMNet) consistently outperforms state-of-the-art inertial sensor-based CoM estimation techniques; 2) stability computed by our image-based method combined with insole foot pressure sensor data produces consistent, strong, and statistically significant correlation with ground truth stability measures (CoMtoCoP $r = 0.79$ $p < 0.001$, CoMtoBoS $r = 0.75$ $p < 0.001$); 3) our fully image-based estimation of stability produces consistent, positive, and statistically significant correlation on the two stability metrics (CoMtoCoP $r = 0.31$ $p < 0.001$, CoMtoBoS $r = 0.22$ $p < 0.043$). Our study provides promising quantitative evidence for the feasibility of image-based stability evaluation in natural environments.**

***Index Terms*—Image-based, stability, base of support, center of mass, center of pressure, deep learning.**

I. Introduction

FALLS in the elderly are an important worldwide health problem [1], and their frequency increases with age [2]. Therefore, frequent and accurate monitoring of human motion stability, especially for the elderly, becomes more and more necessary [3], [4], [5]. Three essential and commonly used component measures for human stability assessment are: Base of Support (BoS), Center of Pressure (CoP), and Center of Mass (CoM) [6], [7] (Fig. 1). Accurate estimation of these key components is currently expensive and time-consuming, involving foot pressure/force plates, motion capture hardware/software, and tedious post-processing of error-prone sensor data [8], [9]. For these reasons, stability measurement is usually restricted to a laboratory environment. A fully or partially image-based method for stability monitoring would be an attractive alternative for deployment in rehabilitation or elder care facilities where unencumbered long-term monitoring could have significant clinical value and allow for preventative or timely corrective interventions to reduce falls.

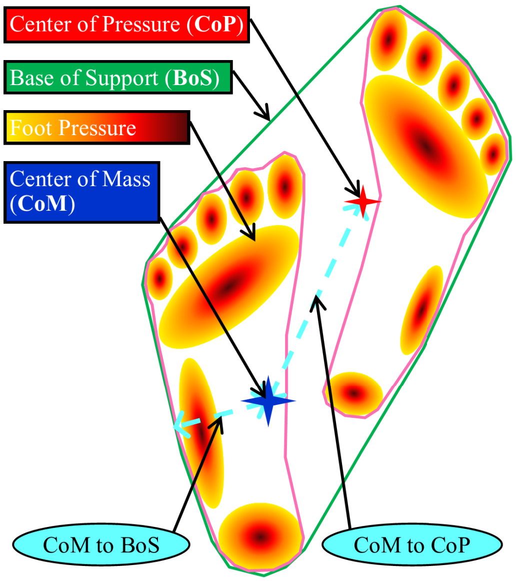


Fig. 1. Stability components and two stability metrics (CoMtoCoP and CoMtoBoS) relative to localized foot pressure with CoP (red star), CoM (blue star), BoS (green border), foot pressure (yellow/red/brown gradations), and stability metrics (cyan lines) [10].

In recent years, human pose extraction from images and video has become an active research area in computer vision and machine learning [11]. However, little work has been done in image-based mapping of body kinematics (pose) to dynamics (foot pressure/force). In an initial study [10], we demonstrated the feasibility of predicting *foot pressure* from images of human pose. We take a step further in this work to explore the feasibility of predicting BoS, CoP, and CoM measures from visual input, and of using these image-based estimates to compute two classic human stability metrics: CoMtoCoP and CoMtoBoS (Fig. 1, Table I).

Using Taiji (a.k.a. Tai Chi) performance data provided by the PSU-TMM100 dataset [10], this work makes the following main contributions:

Manuscript received 13 July 2022; revised 20 October 2022; accepted 9 November 2022. Date of publication 1 December 2022; date of current version 1 February 2023. This research is supported in part by NSF grant IIS-1218729, NSF grant IIS-1909315, and the Penn State College of Engineering Dean's office. *(Corresponding authors: Jesse Scott; Yanxi Liu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the IRB of Penn State University under Application No. IRB STUDY8085. PSU-TMM100 was collected through Penn State University IRB STUDY8085, originally approved on March 19, 2018. The dataset was previously published in [10].

Jesse Scott, Robert T. Collins, and Yanxi Liu are with the School of Electrical Engineering and Computer Science, Penn State University, University Park, PA 16802 USA (e-mail: jus121@psu.edu; rtc12@psu.edu; yul11@psu.edu).

John Challis is with the Biomechanics Laboratory, Kinesiology Department, Penn State University, University Park, PA 16802 USA (e-mail: jhc10@psu.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2022.3226191, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2022.3226191

TABLE I

SELECTED BIOMECHANICAL STABILITY METRICS DETERMINED FROM *CoM*, *CoP*, AND *BoS*

Name	Equation	
CoMtoCoP [12], [13]	$\lVert CoM - CoP \rVert_2$	(1)
CoMtoBoS [14]	$\lVert CoM - BoS_{nearest} \rVert_2$ $\begin{cases} positive & \text{if } CoM \text{ is inside BoS} \\ negative & \text{if } CoM \text{ is outside BoS} \end{cases}$	(2)

1) developing and validating an image-based machine learning algorithm for CoM estimation from image data;

2) assessing two stability metrics (Table I) with a thorough comparison using component values CoP, BoS, and CoM obtained from either image-based or sensor-based (ground truth) measurements (2^3 = 8 combinations evaluated);

3) finding that a fully image-based approach (eliminating the need for foot pressure sensors and motion capture) produces stability estimates that are positively correlated with ground truth (CoMtoCoP $r = 0.31$ $p < 0.001$, CoMtoBoS $r = 0.22$ $p < 0.043$); and

4) finding that insole foot pressure data combined with image-based foot localization and CoM prediction (eliminating need for motion capture hardware) produces stability estimates that are strongly correlated with ground truth estimates (CoMtoCoP $r = 0.79$ $p < 0.001$, CoMtoBoS $r = 0.75$ $p < 0.001$).

The paper is organized as follows: Section II covers background information on stability components and metrics, image-based estimation of dynamics, and the PSU-TMM100 dataset that provides ground truth sensor measurements in this research. Section III covers calculation of the stability components from ground truth data and image-based data, while Section IV covers the stability metric calculations. Section V quantifies and visualizes image-based estimates for CoM, CoP, BoS, CoMtoCoP, and CoMtoBoS, and compares them with sensor-based ground truth estimates. Section VI summarizes the results.

II. BACKGROUND

In a review of video-based measurement for human movement science, Seethapathi et al. [15] indicate that improving kinematic accuracy and estimating dynamics (contact forces) are the key to practical use of computer vision as a tool in biomechanics. Upright human body stability is often investigated by examining relative motion of the CoM compared to the BoS or CoP, which requires measurement of pose and contact forces. Currently, no research exists that uses standard RGB video cameras to automatically determine human body stability during complex *actions*. Our approach is novel in being the first to use pose and ground force dynamics computed solely from video for stability analysis.

A. Balance and Stability

Balance and stability are terms often used interchangeably to describe how well an individual is able to keep from falling. In kinesiology, **balance** describes maintaining static position without significant movement; e.g., balancing on one foot [6]. **Stability** describes continuing dynamic movement of the body while preventing an uncontrolled fall or unplanned movement [16]. Humans have a natural physiological ability to sense their own balance and maintain stability [17], but there is a difference between perception and physical ability that is not easily determined [18]. Computational evaluation of quantified stability uses specialized equipment like force plates to capture 3D foot forces and motion capture technology to measure body movements [6], constraining research to a laboratory setting and limiting its ecological validity.

B. Quantification and Metrics

A comprehensive review by Bruijn et al. [19] breaks stability metrics into three categories: 1) ability to recover from small perturbations, derived from dynamical systems theory and biomechanics, 2) ability to recover from larger perturbations, and 3) determining the maximum controllable perturbation.

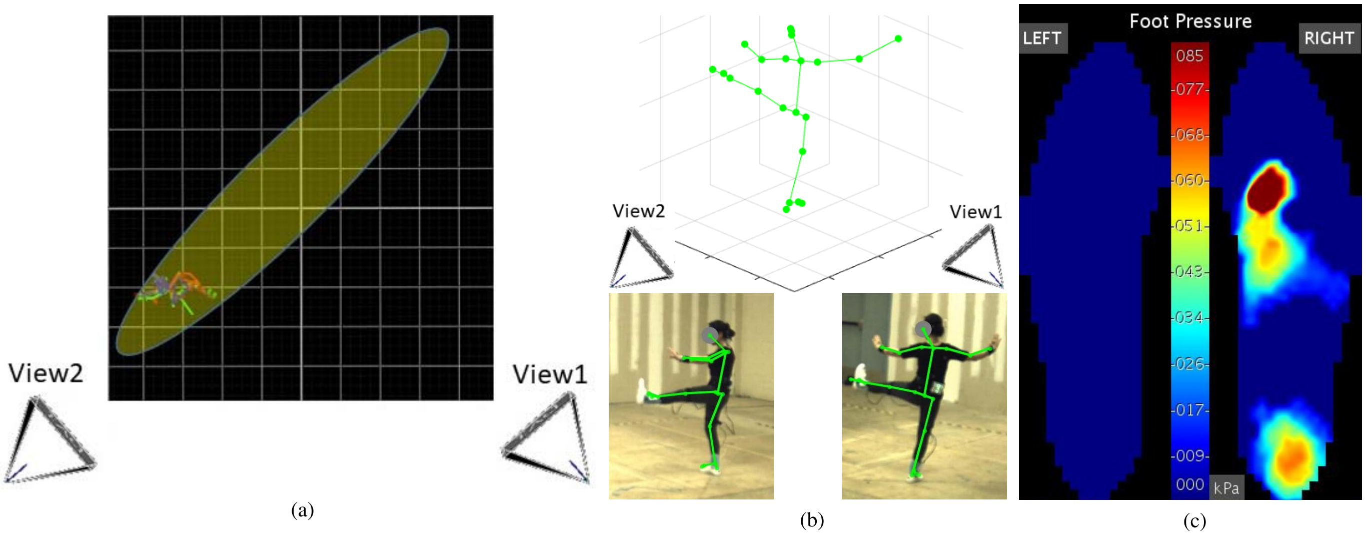
The size of the BoS determines the tolerable condition during gait termination [20] and unexpected perturbation recovery in upright stance [21]. King et al. identify a decrease in the size of the functional BoS with increasing age [22]. Given that the BoS is a determinant of upright stance balance and gait stability, its quantification is an important feature during the analysis of human movement. BoS boundaries are established in [23] by subjects swaying in a circular fashion, defining the boundary by the maximum CoP positions. Force plate and motion analysis data are used to determine a BoS of subjects walking in [24], but these testing conditions limit data collection to a laboratory. Body segment inertial properties and motion analysis data are used in [25] to generate estimates of the CoP motion during gait, while CoP motion is determined for sidestep movements by exploiting convolutional neural network (CNN) models in [26].

Previous work has reported using the center of the hip joints as an approximation for CoM; e.g., [27]. More recently, Chebel et al. [28] present a state-of-the-art neural network for 3D CoM estimation using two subject height measurements (head and hip) and 11 inertial sensors measuring joint angles as input while subjects either squat in place or walk. They report RMSE errors in a componentwise format that works out to total 3D mean error of 18.1 mm for a full body model tested on new subjects. In comparison, our CoMNet (Section III-B.1), a neural network predicting 3D CoM trained on image-based poses only and tested on unseen subjects, has a mean error of 17.6 mm.

C. Image-Based Dynamics Estimation

Previous work in computer vision and graphics has explored estimation of ground contact forces from video and pose [29], [30], [31], [32], [33], but these estimates tend to be simple force vectors rather than the full foot pressure maps estimated in our work.

Using an RGB-D camera to record objects with known geometric and physical properties being manipulated by hands, [34] and [35] estimate the distribution of forces among the fingers in contact from vision-derived first and second order object kinematics using a network learned from hours of hand-object interactions. This approach shows progress in



(a) (b) (c)

Fig. 2. PSU-TMM100 data collection. **(a)**: Top-down view of motion capture environment. **(b)**: 3D pose computed from two video camera views. **(c)**: Foot pressure recorded synchronously using insole sensors.

estimating dynamics from video, but it is constrained to hands, requires sensing depth in addition to video, and is not ground truth validated.

D. PSU-TMM100 Dataset

Ground truth (GT) data for training and evaluation in this work is provided by simultaneously recorded video, motion capture, and foot pressure sensor data (Fig. 2) from the PSU-TMM100 dataset [10], where subjects perform approximately five-minute-long Taiji sequences by moving continuously through a set of complex body poses with a large range of joint articulations and limb orientations. PSU-TMM100 is the *only* available dataset that includes synchronized, sensor-measured recordings of these three modalities, making it a unique and valuable resource for learning to predict stability from imagery. The dataset was collected using IRB-approved protocols (Study8085, initial approval 03-19-2018) with informed consent from all subjects. PSU-TMM100 demographics are ten subjects (five male and five female) with a wide range of experience performing Taiji (4–40 years, $\mu = 13$, $\sigma = 12$) and an average of ten performances (75,775–158,875 frames, $\mu = 131,535$, $\sigma = 26.749$) sampled at 50 Hz from each subject. As Taiji is a slow activity, all experiments use a sub-sampling of data to 5 Hz, reducing the computational resources needed for extensive training and testing on 5-minute motion sequences. Subjects have a broad range of mass (52.5–77.11 kg, $\mu = 63.70$, $\sigma = 6.95$) and height (1.54–1.80 m, $\mu = 1.66$, $\sigma = 0.08$). Four performances (takes) of Subject 2 (Takes 7, 9, 10, and 11) contain corrupt foot pressure data due to an insole sensor malfunction during recording. These outlier takes were discarded from the dataset prior to performing any evaluations reported in this paper.

1) Motion Capture: Ground Truth (GT) 3D pose in the dataset is provided by a Vicon motion capture system. Fig. 3c shows the 21 GT joints whose kinematics are generated by the Vicon Plug-in Gait (PiG) model, which is based on the Conventional Gait Model (CGM) [36], [37] originating from generic body segment inertial properties originally derived by Dempster from cadaver data [18], [38]. The PiG model also generates the GT CoM. A study comparing CoM location estimated by the Dempster parameters versus a more accurate reaction board method indicates a difference of 1 % or less expressed as a percentage of subject height (Fig. 5 of [39]), which for this dataset is 16.8 mm.

2) Video Pose: Two HD video camera views spatiotemporally synchronized with the mocap system provide the data for estimating image-based pose (Fig. 2b). Four body joint configurations, OpenPose (OP), Mocap (GT), BioPose (BP), and HybridPose (HP), are used in this study (Fig. 3).

We use OpenPose, an open-source 2D human pose estimator [40], [41] to predict 2D body joint locations, and two-view triangulation [42] to reconstruct those 25 3D joint estimates (Fig. 3a). Triangulation of two views requires synchronized and calibrated cameras. While estimation of 3D pose from a single camera view is desirable, state of the art in that area is not yet mature, suffering from lower joint detection rates, decreased joint position accuracy, and inaccurate estimation of 3D body orientation with respect to gravity [43].

There are 12 joints in common between GT (Fig. 3c) and OP, and we train the BioPose correction network from [10], [44] to predict those 12 common joints (OP 1-12), improving their 3D biomechanical accuracy and generating BioPose (BP) joints (Fig. 3d). Lastly, HybridPose (HP) (Fig. 3b) is constructed by combining BioPose joints (BP 1-12) with the 13 remaining non-overlapping OpenPose joints (OP 13-25).

3) Insole Pressure Measurement: This research uses insole pressure data spatiotemporally synchronized with the video and motion capture data as the GT foot pressure (Fig. 2c) for training and testing the dynamics estimation networks. Insole sensors accurately measure foot pressure normal to the sensing plane, but with slower response times than force plates [45], although still fast enough for human movement [46].

III. Stability Components

A. Ground Truth (GT) CoM, BoS, and CoP

1) Center of Mass (CoM): The CoM is the 3D point about which the mass of a body is evenly distributed [47]. The 3D CoM can be calculated for static and rigid objects, but the

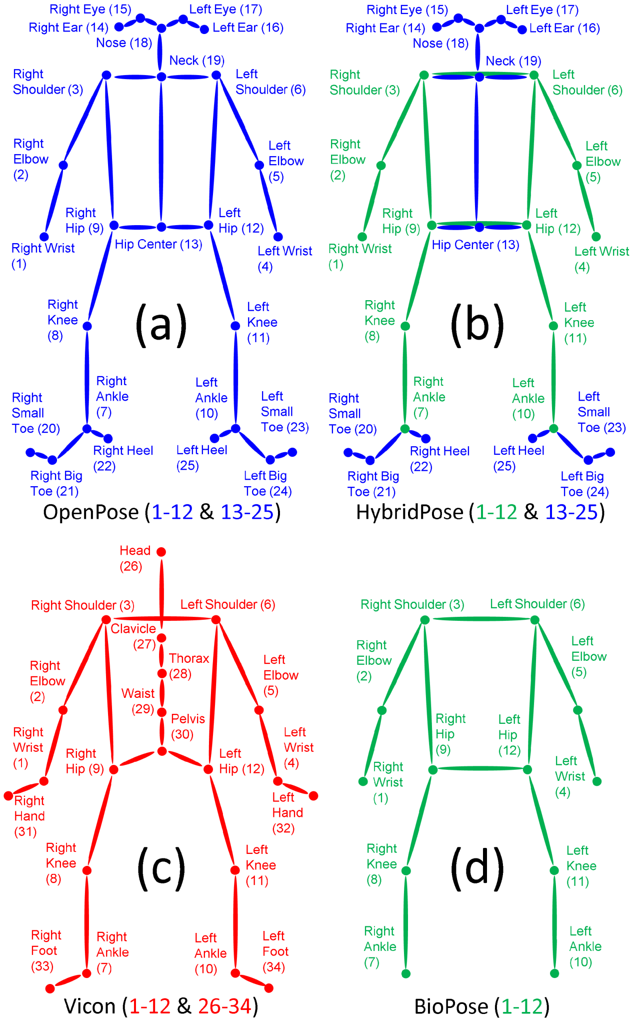


Fig. 3. Comparison of body joints. (a): OpenPose (OP) [40], [41]. (b): HybridPose (HP) joints = $BP \cup OP$(13-25). (c): Ground Truth from Vicon motion capture (GT). (d): BioPose (BP) = $GT \cap OP$. BP joints are common to all joint sets. HP is the 12 BP joints plus the remaining 13 OP joints.

human body is much more complex with varying human tissue masses, body shape, and articulated body pose. Ground truth 3D CoM is calculated by Vicon PiG and is available directly from the motion capture portion of the dataset. Vicon PiG lower body model has been medically validated [48], [49], with [50] providing a thorough evaluation of PiG (and CGM) establishing its widespread use as well as the model's strengths and weaknesses. For the purposes of this study, we treat PiG-modeled joints and calculated CoM as ground truth, following the precedent set by many biomechanical research laboratories and commercial applications [51]. Specifically, the CoM is a 3D position; when projected onto the floor plane, it is referred to as the 2D CoM.

2) Base of Support (BoS): BoS is the convex hull that includes every point of contact that the subject makes with the supporting surface, including body parts (feet or hands) or support devices (crutches or walker) [52]. Ground truth BoS is calculated from insole foot pressure maps after the feet are spatiotemporally localized using the mocap position of the ankles and toes to determine both location and orientation. This localized pressure map is used to create a binary mask of pressures above a minimum threshold (multiple thresholds are evaluated in Fig. 5 and 6) from which a convex hull is calculated (Fig. 8).

3) Center of Pressure (CoP): The CoP is the point at which the ground reaction force vector intercepts the supporting surface, calculated as the weighted sum of all forces acting between a physical object and its supporting surface [53]. CoP is calculated as a spatially weighted mean of all foot pressure samples in the XY plane of the floor using the same localized pressure map used in the calculation of BoS (Fig. 8).

B. Image-Based CoM, BoS, and CoP

Input for our image-based CoM, CoP, and BoS computation begins with triangulated 3D poses calculated from two camera viewpoints (Fig. 2b). All experiments use the same Leave One Subject Out (LOSO) data segmentation for cross-validation, ensuring that the subject being evaluated has not been used in training.

1) CoM Prediction: We use a two-layer fully connected neural network called CoMNet to predict the CoM on a per-frame basis. CoMNet is trained to take 3D pose data and regress a 3D CoM location relative to the hip center. While CoMNet uses joint locations, it does not require the joint velocities/accelerations, subject measurements (height or weight), or the Dempster tables [38] to predict a CoM.

CoMNet training is completed on a Nvidia Quadro K4000 with an RMSE loss function and an Adam optimizer. The network is empirically optimized to have 3072 wide fully connected input and hidden layers using batch normalization, a rectified linear unit, and 50 % dropout regularization. CoMNet training takes approximately 2 hours for each of the 10 LOSO cross-validations. It takes 25 epochs with an initial learning rate of $5e-4$ and a piece-wise learning rate drop factor of 0.25 every 5 epochs. The CoMNet network and training weights will be available upon request following publication.

2) CoP and BoS Prediction: We use the PressNet-Simple 3D (PNS3) network from [10] for image-based foot pressure predictions. While OpenPose joint data are shown in [10] to be the best input for predicting foot pressure, we evaluate motion capture, HybridPose, and OpenPose data for foot localization for all takes of PSU-TMM100. The calculation of CoP and BoS follows the same calculation steps as the ground truth process but replaces sensor inputs with image-based data. CoP and BoS are used for comparing each image-based configuration against ground truth motion capture and insole pressure data. CoP accuracy is evaluated using Euclidean distance between predicted and ground truth locations. BoS is evaluated using the Intersection over Union (IoU) metric, also known as the Jaccard Index [54].

IV. Stability Metrics

After a thorough review of the human balance and stability literature; e.g., [7], [12], [14], [19], [55], [56], two well-established stability metrics were selected for evaluation in this paper: CoMtoCoP and CoMtoBoS (Table I). These two metrics can be calculated from the available data modalities and are well suited for a non-repetitive performance like Taiji that focuses on maintaining biomechanical stability. Both metrics are easily understandable and collectively use all three stability components CoP, CoM, and BoS. A more extensive set of experiments that include additional stability metrics xCoMtoBoS, CoMvtoBoS, and TTC (time to contact) can be found in the first author's Ph.D. thesis [57].

A. CoMtoCoP

The Euclidean distance between a subject's 2D CoM and CoP measures the spatial difference between ground reaction force and gravitational force (Table I, Equation 1) [12], [13]. Conceptually, the further apart these two points are, the greater the potential for instability [7]. While keeping the two points close together may seem advantageous, in dynamic tasks, trained athletes can tolerate greater excursion compared to those not trained [59], as can the young compared with the old [60]. Therefore, subjects who are better at maintaining their stability (perceptually and physically) can allow this distance to become large while still being able to avoid instability. CoMtoCoP is a nonnegative distance measurement typically reported in millimeters, with values normally near zero. A larger variance during a performance indicates subjects with better stability control.

B. CoMtoBoS

The Euclidean distance from the 2D CoM to the border of the BoS quantifies both the magnitude and condition of mechanical imbalance (Table I, Equation 2) [14]. CoMtoBoS magnitude is the distance from the CoM to the nearest point on the BoS boundary; CoMtoBoS is positive if CoM is inside the BoS and negative otherwise. Negative values indicate imbalance/instability that requires intervention to prevent an eventual fall while positive values indicate mechanical balance and stability [14]. There is an inherent maximum positive distance and no limit in the negative direction, but small positive values indicate better stability control [7].

V. Results

A. CoM Prediction

Fig. 4 evaluates CoM location estimates produced by various configurations of CoMNet against Vicon PiG CoM estimates (GT) provided with the dataset. HybridPose CoMNet, that is, CoMNet trained to take HybridPose 3D joint estimates as input, outperforms CoMNet trained on either BioPose or OpenPose joints. HybridPose CoMNet is thus the best performing variant using purely image-based inputs, with mean (+/- std) location error of 17.6 (6.1) mm. Noting that GT CoM locations provided by PiG are computed by a segmental method using Vicon Mocap joints and Dempster table parameters, two additional baseline methods are evaluated. "Dempster" is the Dempster segmental method applied to image-estimated HybridPose 3D joints. The larger mean error of 27.5 (13.0) mm indicates that CoMNet is compensating for differences between 3D joints estimated by HybridPose and Mocap. A second baseline, "Mocap," is our CoMNet trained using GT Mocap joint data as input. The mean error of 12.5 mm establishes a practical limit on CoMNet accuracy when input joints are as accurate as possible.

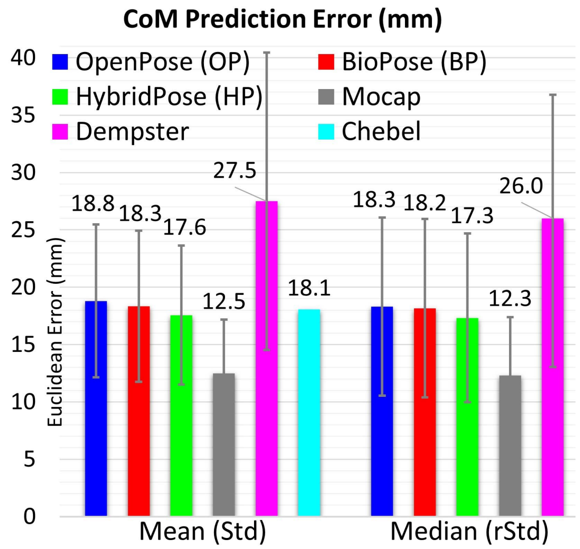


Fig. 4. 3D CoM prediction error (mm) of CoMNet when input pose comes from: OP, BP, HP (best), and Mocap (practical limit) as well as Dempster [38] applied to HP joints and Chebel et al. [28]. Statistics provided: mean (Std) and median (rStd) as compared to GT CoM derived from Vicon PiG (Section II-D.1). Results are based on poses when all body joints are detected. Robust standard deviation (rStd) = 1.4826 times median absolute deviation (MAD) [58].

ComNet HybridPose outperforms BioPose input joints, indicating that useful information is learned by CoMNet when the additional 13 OpenPose joints are combined with BioPose joints. Additionally, all image-based configurations produce similar and consistent results. Using only image-based pose input, CoMNet establishes a state of the art better than the mean Euclidean error of 18.1 mm achieved by Chebel et al. [28] that requires subject measurements and inertial sensors.

B. CoP

Fig. 5 shows results of the PNS3 network architecture [10] on all valid performances in the dataset for overall mean/median (black solid/dashed) and per-subject mean (colors) accuracy. We compare ground truth foot localization with HybridPose and OpenPose localization to quantify the performance of image-based localization. All three foot localization plots show peak performance between 10 kPa [61] and 15 kPa [62] (indicated by gray vertical lines), which are commonly used threshold and peak accuracies, respectively. There are three key observations:

1) HybridPose localization provides the best fully image-based CoP results due to improved ankle accuracy from the BioPose network, with 51.3/48.0 mm (mean/median) error being a small increase from the mocap localization error of 43.5/41.6 mm.
2) HybridPose does not uniformly improve CoP results over OpenPose as Subjects 7 and 8 (light and dark blue plots in Fig. 5) are better with OpenPose.
3) The all-performances results are similar to the one-take-per-subject results reported in [10].

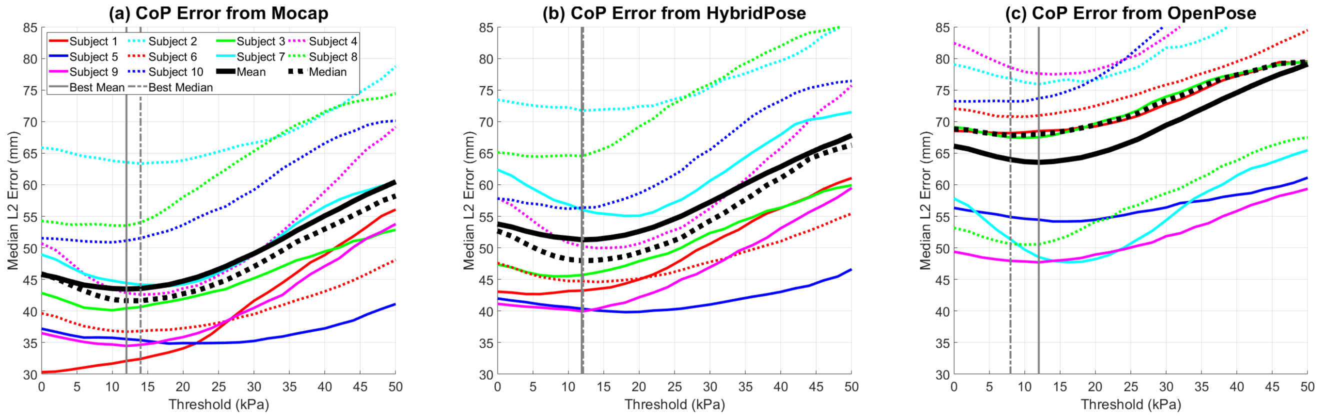


Fig. 5. CoP ℓ^2 error (mm) relative to sensor-based GT (lower better). All results use PNS3 [10] predicted pressure distribution maps and foot localization from **(a)** Mocap, **(b)** HybridPose, or **(c)** OpenPose, respectively. BioPose localization is excluded due to a lack of required joint locations, toes and heels (Fig. 3d). HybridPose input (b) provides the best image-based result. The x-axis shows increasing thresholds (kPa) where pressures below the threshold are set to zero.

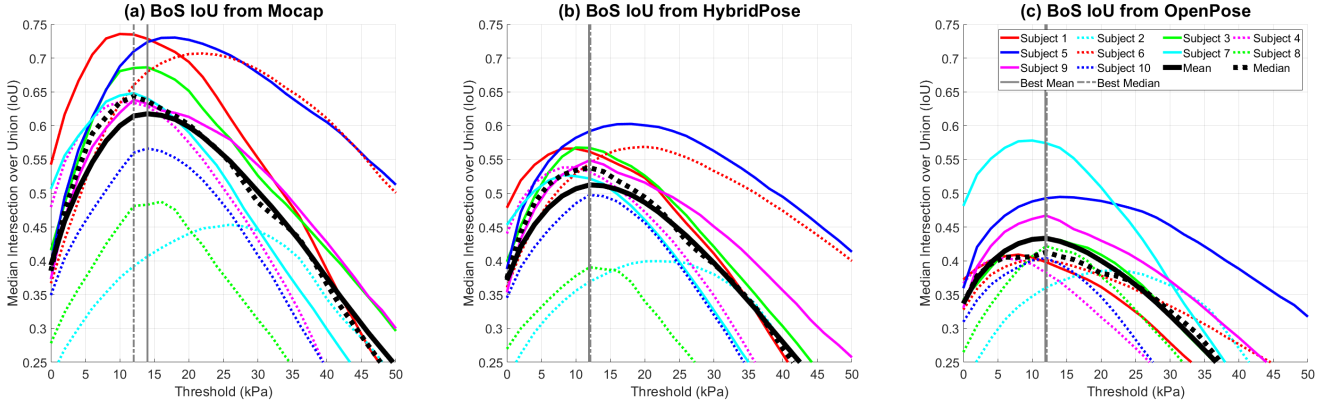


Fig. 6. BoS accuracy using IoU relative to sensor-based GT (higher better). All results use PNS3 [10] predicted pressure distribution maps and foot localization from **(a)** Mocap, **(b)** HybridPose, or **(c)** OpenPose, respectively. BioPose localization is excluded due to a lack of required joint locations, toes and heels (Fig. 3d). HybridPose input (b) provides the best image-based result. The x-axis shows increasing thresholds (kPa) where pressures below the threshold are set to zero.

C. BoS

BoS was also evaluated on all valid performances to determine how different foot localization methods affect IoU accuracy (Fig. 6). Foot localization accuracy affects IoU of BoS more than CoP error as foot pressure pixels of small magnitude can cause large changes in the size and shape of the BoS while having little change on CoP. There are three key observations:

1) HybridPose localization provides the best (higher is better) image-based IoU results (improved ankle accuracy from BioPose network) with 51.24/53.78% (mean/median); a small decrease from the mocap localization IoU of 61.76/64.32%.
2) HybridPose does not uniformly improve IoU results over OpenPose as Subjects 7 and 8 (light and dark blue plots in Fig. 6) are better with OpenPose.
3) The all-performances results of each subject are similar to the one-take-per-subject results in [10], suggesting those one-take results were statistically representative.

D. Stability Metrics

To evaluate image-based estimation of stability metrics, CoMtoCoP and CoMtoBoS are calculated from combinations of ground truth (GT) and image-based estimates (IM) over three data channels (foot pressure, foot localization, and CoM) for eight combinations total. The image-based estimates used are: 1) PNS3 with OpenPose for foot pressure (shown in [10] to be the state of the art); 2) HybridPose for foot localization (shown in Fig. 5b and 6b to produce the best CoP and BoS results); and 3) HybridPose CoMNet for CoM (shown in Fig. 6b to provide the most accurate CoM estimate).

Stability metric values computed from each combination are compared to fully ground truth estimates to determine correlation coefficient (r-value) and statistical significance (p-value). The mean and std of r-values across all ten Leave One Subject Out (LOSO) experiments are reported in Table II. Using sensor-based pressure measurements (GT) with image-based inputs (IM) for localization and CoM (Table II blue) produces correlation r-values of 0.79 and 0.75 for both CoMtoCoP and CoMtoBoS, respectively. Using image-based localization and CoM eliminates the need for motion capture hardware. Switching to image-based foot pressure; i.e., fully image-based stability (Table II green), yields reduced but still positive correlation coefficients of 0.31 and 0.22, respectively.

From Table II results, it is observed that image-based foot pressure prediction has the largest effect on the r-values with 1.00 to 0.39 and 1.00 to 0.32 decreases relative to GT input for CoMtoCoP and CoMtoBoS, respectively. Conversely, image-based foot localization effects on r-values are relatively small

TABLE II

COMBINATORIAL STUDY OF CORRELATION COEFFICIENT (R-VALUE) WITH MEAN ABSOLUTE ERROR (MAE) AND STANDARD DEVIATION (STD) OF DISTANCE FROM GT CALCULATIONS FOR BOTH COMTOCOP AND COMTOBOS COMPARED TO ALL GROUND TRUTH IN MM. COP AND BOS ARE DIRECTLY COMPUTED BY COMBINING PRESSURE AND LOCALIZATION. INPUT COMBINATION ORDER: FOOT PRESSURE - FOOT LOCALIZATION - CENTER OF MASS. DATA SOURCES ARE GROUND TRUTH (GT) OR IMAGE-BASED PREDICTIONS (IM). VALUES ARE THE MEAN FOR ALL TEN LOSO EXPERIMENTS. KEY COMBINATIONS ARE ALL GROUND TRUTH, ONLY GT FOOT PRESSURE, AND FULLY IMAGE-BASED CORRESPONDING TO FIG. 7. ONLY COMPLETE PERFORMANCES ARE INCLUDED AND ALL RESULTS ARE P <= 0.001 EXCEPT P < 0.05(*) AND P > 0.05(+)

Combinatorial Study of Ground Truth and Image-based Inputs using r-value (Std) & MAE (Std) in mm								
Pressure-Location-CoM	GT-GT-GT	GT-GT-IM	GT-IM-GT	GT-IM-IM	IM-GT-GT	IM-GT-IM	IM-IM-GT	IM-IM-IM
CoMtoCoP								
r-value (Std)	1.00 (0.00)	0.88 (0.20)	0.88 (0.08)	0.79 (0.18)	0.39 (0.09)	0.34 (0.13)	0.35 (0.09)	0.31 (0.10)
MAE (Std)	0.00 (0.00)	10.14 (23.75)	15.90 (22.88)	18.12 (34.70)	37.37 (51.47)	40.00 (61.21)	40.28 (54.77)	41.92 (62.81)
CoMtoBoS								
r-value (Std)	1.00 (0.00)	0.83 (0.24)	0.86 (0.08)	0.75 (0.21)	0.32 (0.15)*	0.25 (0.14)+	0.27 (0.12)*	0.22 (0.12)*
MAE (Std)	0.00 (0.00)	9.12 (22.11)	11.82 (16.98)	14.76 (28.95)	25.47 (35.35)	28.40 (45.10)	28.95 (38.54)	31.07 (46.44)

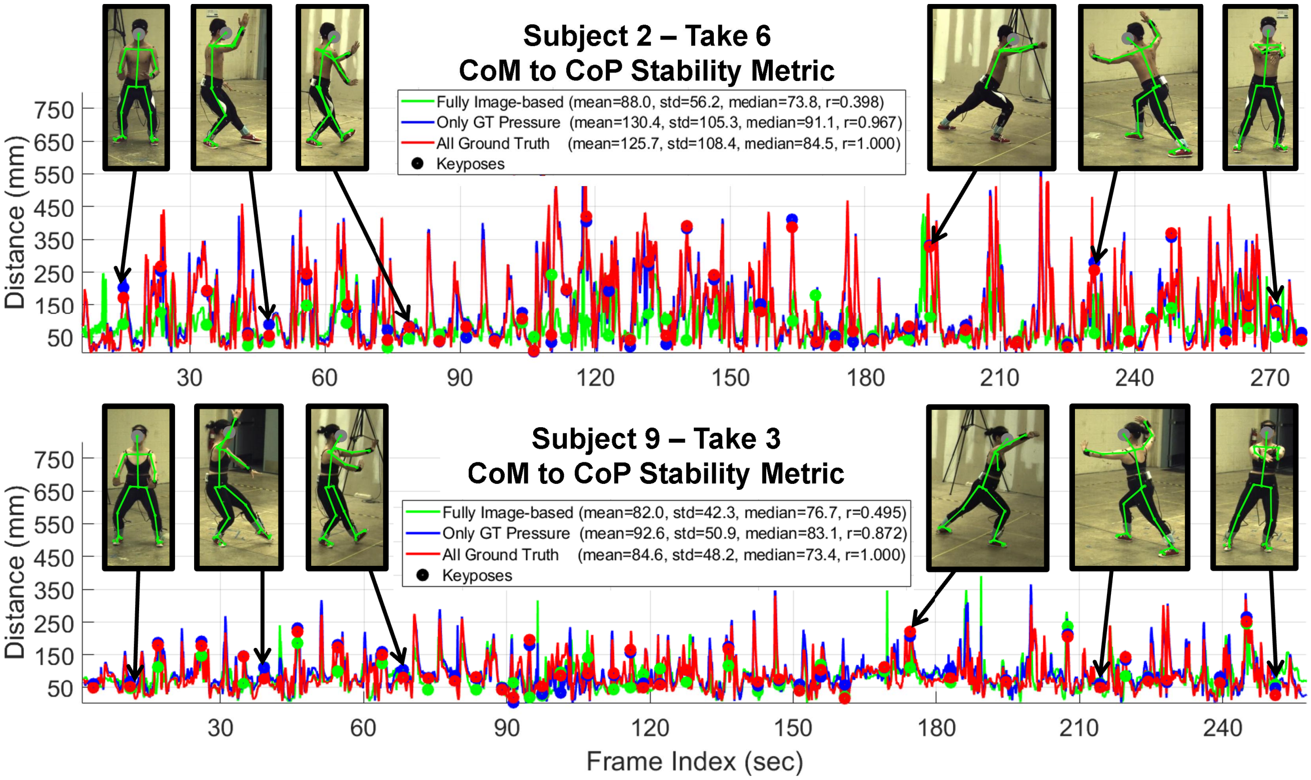


Fig. 7. Examples of CoMtoCoP results highlighting similar trends of all three combinations presented: fully ground truth (red), ground truth foot pressure with all other inputs image-based (blue), and fully image-based (green). Based on CoMtoCoP r-value: Subject2 - Take6 (top) is the best for Only GT Pressure and Subject9 - Take3 (bottom) is the best for Fully Image-based. Plots include image call-outs of key poses with video joint overlay, mean, standard deviation, median, and r-value for each combination. Plot colors are related to highlighted columns of the comprehensive results in Table II. The red line heavily occludes blue and green because of very strong correlation.

with 1.00 to 0.88 and 1.00 to 0.86 decreases, respectively, while image-based CoM effects are also small with 1.00 to 0.88 and 1.00 to 0.83 decreases, respectively. These results indicate that image-based foot pressure estimation has the largest room for improvement at approximately five times the r-value effect of image-based localization or CoM estimation.

Table II also reports the Mean Absolute Error (MAE) for each of the eight combinations of stability estimates relative to GT. MAE consistently increases when the stability metrics use more image-based input data, while standard deviation increases primarily when image-based foot pressure is included. The only minor difference between the two metrics is that CoMtoBoS has both lower r-values and MAE across most combinations when compared to CoMtoCoP. Since both BoS and CoP derive from foot pressure, it is expected that both metrics would have generally similar behavior (Fig. 5 and 6). Additionally, the overall lower values are expected since CoMtoBoS has a data range that includes negative values, unlike CoMtoCoP which has to be ≥ 0.

Fig. 7 visualizes CoMtoCoP stability metric results for two performances by plotting a fully ground truth result computed using motion capture and insole sensors compared to two combinations that use some or all image-based input data: 1) image-based localization and CoM prediction with GT foot pressure, which eliminates the need for motion capture sensor requirements and 2) fully image-based predictions that eliminate both motion capture and foot pressure sensor requirements. As compared to the red

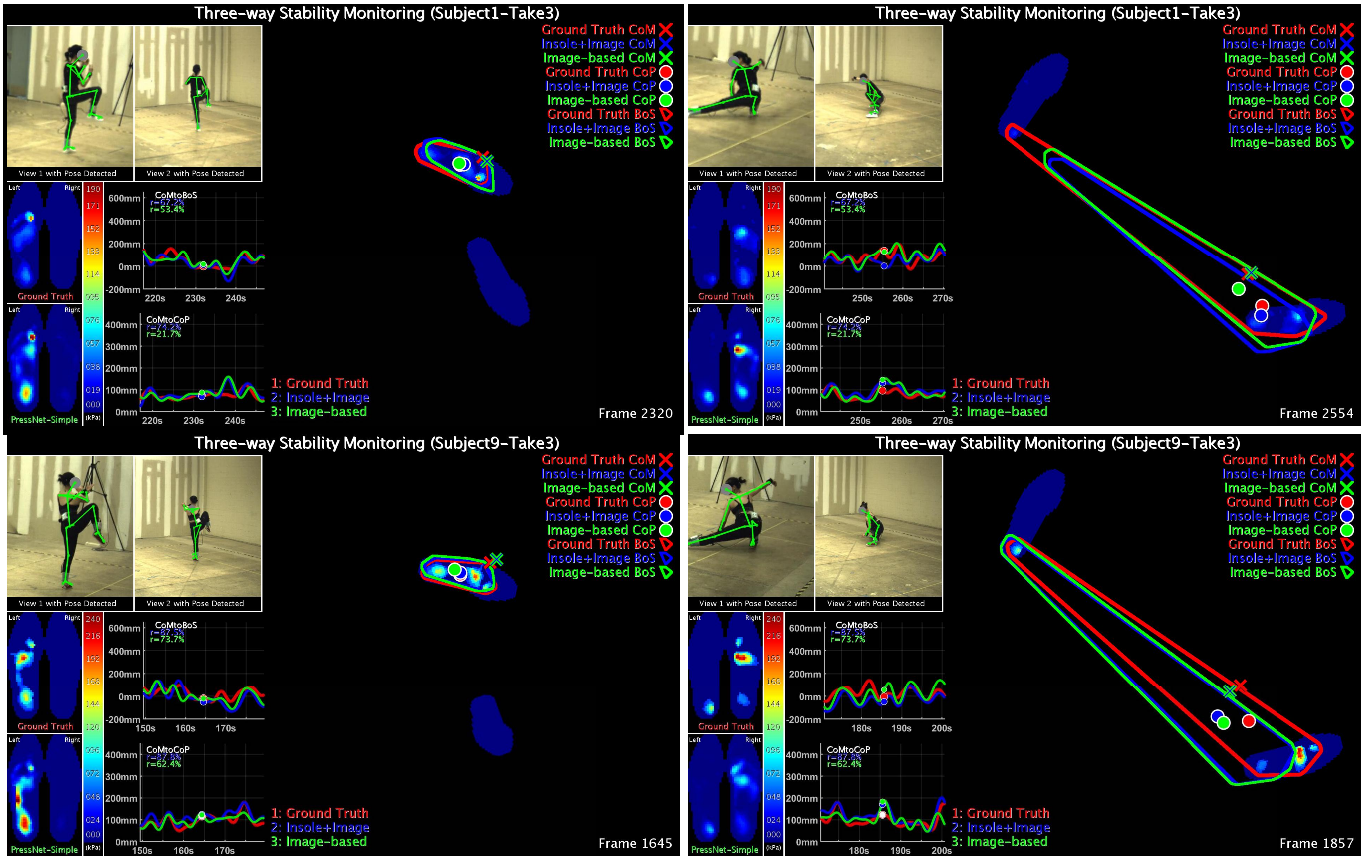


Fig. 8. Image-based stability results compared to GT of Subject 1 - Take 3 (Top) and Subject 9 - Take 3 (Bottom) representing the best performance (based on r-value) for CoMtoCoP and CoMtoBoS, respectively. Each frame includes input images, motion capture, and foot pressure plus output stability components and metrics. Samples include single foot leg lift (Left) and double foot lunge (Right) poses selected from full performances. Results color map: Ground Truth (red), Image+Insole (blue), and Image-based (green).

curve showing (All Ground Truth) results, Subject 2 - Take 6 represents the best r-value results for the blue curve (Only GT Foot Pressure) while Subject 9 - Take 3 represents the best r-value results for the green curve (Fully Image-based). Each plot includes six keypose images with detected joint overlay. Both plots show strong overlap between the blue and red curves due to their strong correlation ($r = 0.97$ and 0.87), while the green curves exhibit only partial overlap with the red ground truth curves, reflecting only moderate correlation ($r = 0.40$ and 0.50).

Fig. 8 focuses on the qualitative results of calculating imaged-based stability components (CoP, BoS, and CoM) and stability metrics (CoMtoCoP and CoMtoBoS). The frames show two Taiji poses from performances by Subject 1 - Take 3 (top) and Subject 9 - Take 3 (bottom), representing the best r-value results (0.48 and 0.50, respectively) when using Fully Image-based estimation with PNS3 (OpenPose) foot pressure prediction, HybridPose for foot localization, and CoMNet from HybridPose for CoM prediction (green in Table II). These information-rich frames (Fig. 8) facilitate at a glance a qualitative comparison of estimated components CoM, CoP, BoS and stability measures CoMtoCoP and CoMtoBoS computed from either all ground truth values (red), using ground truth insole pressure but otherwise image-based estimates (blue), or fully image-based estimation (green).

There are two key takeaways from this analysis. First, a fully image-based approach (eliminating the need for foot pressure sensors and motion capture) produces stability estimates that are positively correlated with GT (CoMtoCoP $r = 0.31$ $p < 0.001$, CoMtoBoS $r = 0.22$ $p < 0.043$). Second, a hybrid approach using insole foot pressure sensor data combined with image-based foot localization and CoM prediction (eliminating need for motion capture hardware) produces stability estimates that are strongly correlated with GT estimates (CoMtoCoP $r = 0.79$ $p < 0.001$, CoMtoBoS $r = 0.75$ $p < 0.001$).

E. Computational Costs

For each sampled time instance, all data processing and analyses are performed in under 2 seconds using an 8 core PC with 64 GB of RAM, without optimizing for speed of processing. Of this time, over 1 second is used to estimate the 3D image-based pose while the remaining time is used for foot pressure and CoM estimation combined with stability calculations.

F. Stability Trends Analysis

Based on the stability analysis completed at 5 Hz sampling rate (Section II-D), low frequency content is modeled using a zero-lag, low-pass filter (0.2 Hz). Fig. 7 stability metric data are low-pass filtered to generate low frequency stability trends (Fig. 9). The computed “Only GT Pressure” and “Fully Image-based” curves for Subject 2 (r-values of 0.98 and 0.67 compared to low-pass filtered ground truth) and Subject 9 (r-values of 0.88 and 0.62) illustrate similar trends;

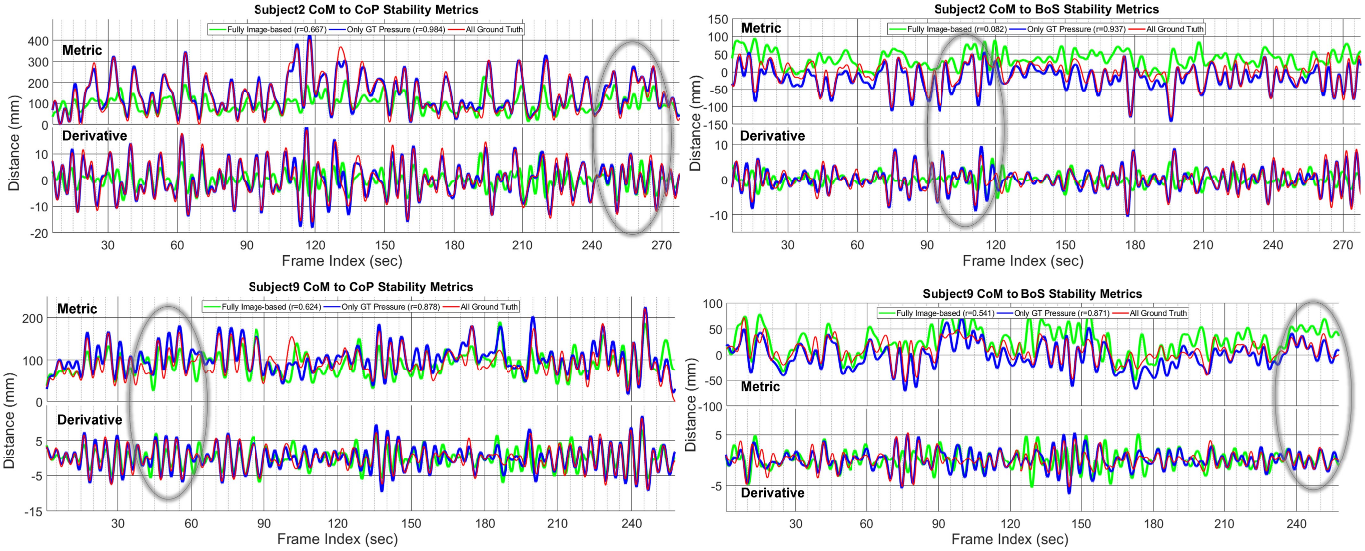


Fig. 9. Smoothed CoMtoCoP (left) and CoMtoBoS (right) stability and their derivative curves from Fig. 7 highlighting similar trends between ground truth (red), image-based with insole sensor only (blue) of Subject 2 (top), and fully image-based (green) of Subject 9 (bottom). Gray circles highlight when derivatives are well correlated.

i.e., upward/downward-sloping curves, indicating relative consistency with GT stability measures.

VI. Conclusion

This work demonstrates that image-based stability quantification is computationally feasible (Fig. 8). Using 2D pose extracted from two RGB cameras, 3D pose is triangulated and used to predict foot pressure and to compute CoM. The predicted foot pressure is further combined with image-based foot localization to calculate BoS and CoP. These three stability components (CoM, CoP, and BoS) are combined to calculate image-based predictions of stability metrics CoMtoCoP and CoMtoBoS, which are quantitatively shown in Section V-D to have significant positive correlation with the GT stability metric values. Stability metrics computed from image-based pose combined with pressure sensors produce strong correlations of 0.79 and 0.75, respectively. Fully image-based stability estimates have lower yet positive correlations of 0.31 and 0.22 with ground truth, respectively, indicating the potential feasibility for fully image-based stability estimation given future improvements to image-based foot pressure prediction.

CoMNet predicts image-based 3D CoM from 3D pose with a mean Euclidean error of 17.56 mm, outperforming the state-of-the-art method using body-worn inertial sensors [28], and predicting an error nearly as low as the expected error in ground truth motion capture calculations [39] while using only image-based data. Additionally, the work originally published in [10] reporting CoP and BoS results for one-take-per-subject sub-sampling is validated here for all valid dataset performances (Fig. 5 and 6), confirming the sub-sampling in [10] was a representative cross-section.

Computing quantified stability measures exclusively from imagery substantially reduces the need for expensive, physically encumbering equipment that constrains data collection to laboratory environments. The presented methods therefore may enable smart health interventions in real-world conditions based on timely image-based evaluation of human stability.

References

[1] F. Englander, T. J. Hodson, and R. A. Terregrossa, “Economic dimensions of slip and fall injuries,” *J. Forensic Sci.*, vol. 41, no. 5, pp. 733–746, Sep. 1996.

[2] American Geriatrics Society, British Geriatrics Society and American Academy of Orthopaedic Surgeons Panel on Falls Prevention, “Guideline for the prevention of falls in older persons,” *J. Amer. Geriatrics Soc.*, vol. 49, no. 5, pp. 664–672, May 2001.

[3] J. Parkkari et al., “Majority of hip fractures occur as a result of a fall and impact on the greater trochanter of the femur: A prospective controlled hip fracture study with 206 consecutive patients,” *Calcified Tissue Int.*, vol. 65, no. 3, pp. 183–187, Sep. 1999.

[4] D. A. Sterling, J. A. O’Connor, and J. Bonadies, “Geriatric falls: Injury severity is high and disproportionate to mechanism,” *J. Trauma Acute Care Surg.*, vol. 50, no. 1, pp. 116–119, 2001.

[5] *Web-Based Injury Statistics Query and Reporitng System (Wisqars)*. accessed: Sep. 13, 2015. [Online]. Available: http://www.cdc.gov/injury/wisqars/

[6] D. Winter, *A.B.C. (Anatomy, Biomechanics Control) Balance During Standing and Walking*. Waterloo, ON, Canada: Waterloo Biomechanics, 1995.

[7] A. L. Hof, M. G. J. Gazendam, and W. E. Sinke, “The condition for dynamic stability,” *J. Biomech.*, vol. 38, no. 1, pp. 1–8, 2005.

[8] J. H. Challis, “The variability in running gait caused by force plate targeting,” *J. Appl. Biomechanics*, vol. 17, no. 1, pp. 77–83, Feb. 2001.

[9] M. Whittle, *Gait Analysis: An Introduction*, 4th ed. Oxford, U.K.: Butterworth Heinemann, 2007.

[10] J. Scott, B. Ravichandran, C. Funk, R. T. Collins, and Y. Liu, “From image to stability: Learning dynamics from human pose,” in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 12368. Cham, Switzerland: Springer, Nov. 2020, pp. 536–554.

[11] C. Zheng et al., “Deep learning-based human pose estimation: A survey,” 2020, *arXiv:2012.13392*.

[12] Y. Jian, D. Winter, M. Ishac, and L. Gilchrist, “Trajectory of the body COG and COP during initiation and termination of gait,” *Gait Posture*, vol. 1, no. 1, pp. 9–22, Mar. 1993.

[13] H. Chaudhry, B. Bukiet, Z. Ji, and T. Findley, “Measurement of balance in computer posturography: Comparison of methods—A brief review,” *J. Bodywork Movement Therapies*, vol. 15, no. 1, pp. 82–91, Jan. 2011.

[14] V. Lugade, V. Lin, and L.-S. Chou, “Center of mass and base of support interaction during gait,” *Gait Posture*, vol. 33, no. 3, pp. 406–411, 2011.

[15] N. Seethapathi, S. Wang, R. Saluja, G. Blohm, and K. P. Kording, “Movement science needs different pose tracking algorithms,” 2019, *arXiv:1907.10226*.

[16] M. P. Murray, A. Seireg, and R. C. Scholz, “Center of gravity, center of pressure, and supportive forces during human activities,” *J. Appl. Physiol.*, vol. 23, no. 6, pp. 831–838, Dec. 1967.

[17] L. Assländer, G. Hettich, and T. Mergner, "Visual contribution to human standing balance during support surface tilts," *Human Movement Sci.*, vol. 41, pp. 147–164, Jun. 2015.

[18] D. A. Winter, *Biomechanics Motor Control Human Movement*. Chichester, U.K.: Wiley, Sep. 2009.

[19] S. M. Bruijn, O. G. Meijer, P. J. Beek, and J. H. Van Dieën, "Assessing the stability of human locomotion: A review of current measures," *J. Roy. Soc. Interface*, vol. 10, no. 83, Jun. 2013, Art. no. 20120999.

[20] Y.-C. Pai and J. Patton, "Center of mass velocity-position predictions for balance control," *J. Biomech.*, vol. 30, no. 4, pp. 347–354, Apr. 1997.

[21] A. L. Hof and C. Curtze, "A stricter condition for standing balance after unexpected perturbations," *J. Biomechanics*, vol. 49, no. 4, pp. 580–585, Feb. 2016.

[22] M. B. King, J. O. Judge, and L. Wolfson, "Functional base of support decreases with age," *J. Gerontol.*, vol. 49, no. 6, pp. M258–M263, Nov. 1994.

[23] P. Haibach, S. Slobounov, E. Slobounova, and K. Newell, "Virtual time-to-contact of postural stability boundaries as a function of support surface compliance," *Exp. Brain Res.*, vol. 177, pp. 471–482, Mar. 2006.

[24] F. Süptitz, M. M. Catalá, G.-P. Brüggemann, and K. Karamanidis, "Dynamic stability control during perturbed walking can be assessed by a reduced kinematic model across the adult female lifespan," *Human Movement Sci.*, vol. 32, no. 6, pp. 1404–1414, Dec. 2013.

[25] H. Pillet, X. Bonnet, F. Lavaste, and W. Skalli, "Evaluation of force plate-less estimation of the trajectory of the centre of pressure during gait. Comparison of two anthropometric models," *Gait Posture*, vol. 31, no. 2, pp. 147–152, Feb. 2010.

[26] W. R. Johnson, J. Alderson, D. Lloyd, and A. Mian, "Predicting athlete ground reaction forces and moments from spatio-temporal driven CNN models," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 3, pp. 689–694, Mar. 2019.

[27] K.-D. Ng, S. Mehdizadeh, A. Iaboni, A. Mansfield, A. Flint, and B. Taati, "Measuring gait variables using computer vision to assess mobility and fall risk in older adults with dementia," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–9, 2020.

[28] E. Chebel and B. Tunc, "Deep neural network approach for estimating the three-dimensional human center of mass using joint angles," *J. Biomechanics*, vol. 126, Sep. 2021, Art. no. 110648.

[29] M. A. Brubaker, L. Sigal, and D. J. Fleet, "Estimating contact dynamics," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2389–2396.

[30] M. Vondrak, L. Sigal, and O. C. Jenkins, "Physical simulation for probabilistic motion tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[31] M. A. Brubaker, D. J. Fleet, and A. Hertzmann, "Physics-based person tracking using the anthropomorphic Walker," *Int. J. Comput. Vis.*, vol. 87, nos. 1–2, pp. 140–155, 2010.

[32] X. Lv, J. Chai, and S. Xia, "Data-driven inverse dynamics for human motion," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–12, Nov. 2016.

[33] Z. Li, J. Sedlar, J. Carpentier, I. Laptev, N. Mansard, and J. Sivic, "Estimating 3D motion and forces of person-object interactions from monocular video," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8632–8641.

[34] T.-H. Pham, A. Kheddar, A. Qammaz, and A. A. Argyros, "Towards force sensing from vision: Observing hand-object interactions to infer manipulation forces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2810–2819.

[35] T.-H. Pham, N. Kyriazis, A. A. Argyros, and A. Kheddar, "Hand-object contact force estimation from markerless visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2883–2896, Dec. 2018.

[36] R. B. Davis, S. Õunpuu, D. Tyburski, and J. R. Gage, "A gait analysis data collection and reduction technique," *Human Movement Sci.*, vol. 10, no. 5, pp. 575–587, 1991.

[37] M. P. Kadaba, H. K. Ramakrishnan, and M. E. Wootten, "Measurement of lower extremity kinematics during level walking," *J. Orthopaedic Res.*, vol. 8, no. 3, pp. 383–391, 1990.

[38] W. T. Dempster, "Space requirements of the seated operator: Geometrical, kinematic, and mechanical aspects of the body with special reference to the limbs," Aerosp. Med. Res. Lab., Wright-Patterson Air Force Base, Columbus, OH, USA, WADC Tech. Rep. 55-159 (AD 87892), 1955.

[39] M. Virmavirta and J. Isolehto, "Determining the location of the body's center of mass for different groups of physically active people," *J. Biomechanics*, vol. 47, no. 8, pp. 1909–1913, Jun. 2014.

[40] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4724–4732.

[41] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1302–1310.

[42] R. Hartley, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[43] M. Kocabas, C.-H.-P. Huang, J. Tesch, L. Müller, O. Hilliges, and M. J. Black, "SPEC: Seeing people in the wild with an estimated camera," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 11035–11045.

[44] B. Ravichandran, "Biopose-3D and PressNet-KL: A path to understanding human pose stability from video," M.S. thesis, Dept. Comput. Sci. Eng., Pennsylvania State Univ., State College, PA, USA, 2020.

[45] K. J. Chesnin, L. Selby-Silverstein, and M. P. Besser, "Comparison of an in-shoe pressure measurement device to a force plate: Concurrent validity of center of pressure measurements," *Gait Posture*, vol. 12, no. 2, pp. 128–133, Oct. 2000.

[46] T. L. Chevalier, H. Hodgins, and N. Chockalingam, "Plantar pressure measurements using an in-shoe system and a pressure platform: A comparison," *Gait Posture*, vol. 31, no. 3, pp. 397–399, Mar. 2010.

[47] H. Harrison and T. Nettleton, *Principles of Engineering Mechanics*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 1994.

[48] M. C. Carson, M. E. Harrington, N. Thompson, J. J. O'Connor, and T. N. Theologis, "Kinematic analysis of a multi-segment foot model for research and clinical applications: A repeatability analysis," *J. Biomechanics*, vol. 34, no. 10, pp. 1299–1307, Oct. 2001.

[49] J. Stebbins, M. Harrington, N. Thompson, A. Zavatsky, and T. Theologis, "Repeatability of a model for measuring multi-segment foot kinematics in children," *Gait Posture*, vol. 23, no. 4, pp. 401–410, Jun. 2006.

[50] R. Baker, F. Leboeuf, J. Reay, and M. Sangeux, "The conventional gait model—Success and limitations," in *Handbook Human Motion*, B. Muller and S. I. Wolf, Eds. Cham, Switzerland: Springer, Apr. 2018, pp. 489–508.

[51] H. Kainz et al., "Reliability of four models for clinical gait analysis," *Gait Posture*, vol. 54, pp. 325–331, May 2017.

[52] M. D. Binder, N. Hirokawa, and U. Windhorst, "Base of support," in *Encyclopedia of Neuroscience*. Berlin, Germany: Springer, 2009, p. 354.

[53] M. D. Binder, N. Hirokawa, and U. Windhorst, "Center of pressure," in *Encyclopedia of Neuroscience*. Berlin, Germany: Springer, 2009, p. 604.

[54] P. Jaccard, "Distribution de la flore Alpine dans le Bassin des Dranses et dans quelques régions voisines," *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 241–272, Jan. 1901.

[55] S. M. Slobounov, E. S. Slobounova, and K. M. Newell, "Virtual time-to-collision and human postural control," *J. Motor Behav.*, vol. 29, no. 3, pp. 263–281, Sep. 1997.

[56] J. M. Haddad, J. L. Gagnon, C. J. Hasson, R. E. A. Van Emmerik, and J. Hamill, "The use of time-to-contact measures in assessing postural stability," *J. Appl. Biomechanics*, vol. 22, pp. 61–155, 2006.

[57] J. Scott, "Dynamic stability monitoring of complex human motion sequences via precision computer vision," Ph.D. dissertation, Dept. Comput. Sci. Eng., Pennsylvania State Univ., State College, PA, USA, 2022.

[58] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *J. Amer. Statist. Assoc.*, vol. 88, no. 424, pp. 1273–1283, Dec. 1993.

[59] J. P. Ambegaonkar, S. V. Caswell, J. B. Winchester, Y. Shimokochi, N. Cortes, and A. M. Caswell, "Balance comparisons between female dancers and active nondancers," *Res. Quart. Exerc. Sport*, vol. 84, no. 1, pp. 24–29, Mar. 2013.

[60] J. T. Cavanaugh, M. Shinberg, L. Ray, K. M. Shipp, M. Kuchibhatla, and M. Schenkman, "Kinematic characterization of standing reach: Comparison of younger vs. Older subjects," *Clin. Biomechanics*, vol. 14, no. 4, pp. 271–279, May 1999.

[61] N. L. W. Keijsers, N. M. Stolwijk, B. Nienhuis, and J. Duysens, "A new method to normalize plantar pressure measurements for foot size and foot progression angle," *J. Biomechanics*, vol. 42, no. 1, pp. 87–90, Jan. 2009.

[62] H. Hsiao, J. Guan, and M. Weatherly, "Accuracy and precision of two in-shoe pressure measurement systems," *Ergonomics*, vol. 45, no. 8, pp. 537–555, Jun. 2002.