Estimation of Center of Mass for Sports Scene Using Weighted Visual Hull

Tomoya Kaichi[1] Shohei Mori[1] Hideo Saito[1] Kosuke Takahashi[2]
Dan Mikami[2] Mariko Isogawa[2] Hideaki Kimata[2]
[1]Keio University, Yokohama, Japan [2]NTT Corporation, Tokyo, Japan
{kaichi, mori, saito}@hvrl.ics.keio.ac.jp
{takahashi.kosuke, mikami.dan, isogawa.mariko, kimata.hideaki}@lab.ntt.co.jp

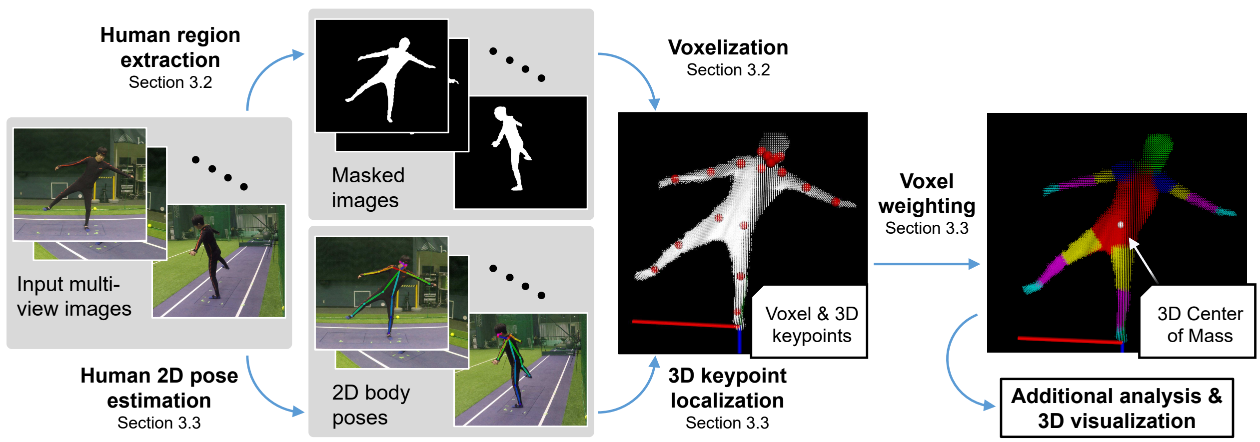


Figure 1: Overview of the proposed CoM estimation using visual hull and body parts dependent voxel-wise weighting.

Abstract

This paper presents a method to estimate the 3D position of a center of mass (CoM) of a human body from a set of multi-view images. As a well-known fact, in sports, collections of CoM are important for analyzing the athletes' performance. Most conventional studies in CoM estimation require installing a measuring system (e.g., a force plate or optical motion capture system) or attaching sensors to the athlete. While such systems reliably estimate CoM, casual settings are preferable for simplifying preparations. To address this issue, the proposed method takes a vision-based approach that does not require specialized hardware and wearable devices. Our method calculates subject's CoM using voxels with body parts dependent weighting. This individual voxel reconstruction and voxel-wise weighting reflects the differences in each body shape, and are expected to contribute to higher performance in analysis. The results using real data demonstrated the performance of the proposed method were compared to force plate data, and provided a 3D CoM visualization in a dynamic scene.

1. Introduction

The quantitative data obtained by sports motion analyses are used to improve the performance of athletes [1, 13]. In many sports, superior balance ability, the ability to perform a task while maintaining or regaining a stable position [21] is necessary not only to achieve high level competitive levels [10] but also to avoid injury [8]. Traditionally, the feasible movements for the control of balance are described in a single-dimensional space related to the horizontal position of the body center of mass (CoM) [17]. Therefore, accurate quantification of an athlete's dynamic CoM in the field plays an important role in many sports applications. Indeed, literatures [11, 20] show that the CoM affects the performance of the athletes in some sports.

Motion capture systems and force plates are often used for measuring the CoM of the human body [4, 15]. However, these systems are large and expensive, and designed only for special environments such as lab and studios. Gonzalez *et al.* proposed a method to estimate CoM using a portable Wii balance board and Kinect to make the whole system affordable [7]. On the other hand, the applications are limited to indoor scenarios because of the poor outdoor performance of Kinect sensors. The recent active developments in wearable motion capture systems have revealed that wearable devices can be used for CoM estimation [16]. While the wearable systems are capable of estimating positions of body joints even outdoors, many sports communities forbid wearing electronic devices during games.

From this background, we propose a method to measure CoM of athletes during sports games (Figure 1). This method, for example, enables data analysts to compare movements during practice and matches; in addition, they will be able to analyze the other team's players. To achieve this, our method must meet the following three conditions.

- The method is capable of outdoor CoM estimation
- without wearable devices,
- and has ability to reflect athlete's figure to CoM without prior personalization.

To satisfy these conditions, we propose to estimate CoM using multi-view RGB images only. First, we reconstruct a 3D model of the subject's body using multi-view RGB images. The 3D model is divided into nine body parts, and weights depending on the body parts are assigned to each part. Then, the weighted average of the parts calculates the whole body CoM. Since the proposed method uses only RGB images, wearable devices are not necessary and outdoor CoM estimation is achieved. Moreover, 3D shape reconstruction of subjects' body handles the differences in individual's figure when calculating CoM.

We evaluated the accuracy of the proposed method via an experiment using three people and four static postures. Also, to replicate an actual sports scene, we estimated the transition of the CoM for swinging a baseball bat, thus confirming that we could obtain reasonable results.

2. Related work

CoM trajectories, which describes the player's balance ability, plays a key role to in improving athletic performance as well as preventing sport-related injuries [10, 8, 9]. Different types of devices, such as force plates, motion capture systems, depth sensors, and wearable devices have been used to estimate CoM depending on the environment.

The force plate approaches measure ground reaction forces to calculate CoM motion based upon Newton's Second Law, which states that the net external force acting upon a body is equal to its mass multiplied by its acceleration. The motion capture approaches use multiple markers on the body to track and measure the position of body segments, incorporating an anthropometric model to calculate segmental center of mass positions. Saini *et al.* compared the accuracy of these two simple methods and confirmed that they could accurately estimate CoM when subjects moved slowly [18]. Carpentier *et al.* proposed a method to estimate CoM by combining the data from force plates and motion capture systems [4]. They reduced sensor noise by using data fusion based on complementary filtering. However, force plates can only be used when the body is touching the ground. In addition, it is difficult to move force plates and motion capture systems and these devices limit the range of movement of the subject.

To relax the restrictions on the measuring environments, González *et al.* proposed to use Kinect and Wii balance board together [7, 6]. They reported that the method could estimate CoM with accuracy close to that of a Vicon motion capture system by personalizing each part of the human body beforehand [7]. However, the measurement accuracy of Kinect decreases outdoors. Besides, just as with motion capture, Kinect acquires only the skeleton of the human body, and volumetric properties, therefore, are ignored in the CoM estimation without the pressure sensor.

Wearable motion capture systems can perform CoM measurements with fewer restrictions regarding environments [16]. Najafi *et al.* estimated the trajectory of CoM during a golf club swing by using wearable sensors; they showed that wearable technologies based on inertial sensors are a viable option for assessing dynamic postural control in complex tasks. This method requires the player to wear sensors. However, as most sports forbid wearing electronic equipment during games, wearable sensors cannot be used to estimate players' CoM during actual matches.

Consequently, the proposed framework is the first attempt towards an end-to-end automated process for CoM trajectory estimation considering volumetric properties of the measured athlete using image inputs only, although skelton-based approaches using manually selected joint locations in images have been used for CoM estimation [5].

3. Method

3.1. Overview

The proposed method estimates the CoM of a single person in a process. We place N calibrated cameras ($N \geq 2$) so as to record the target. A schematic of the proposed process is shown in Figure 1. The input is only RGB images taken from multiple viewpoints. Those images are used for 3D reconstruction of body shape and 3D kinematic structure estimation of the human body. Based on the joint positions obtained via the estimation of body structure, the human body model is segmented into nine parts. Then, CoM is obtained by assigning a weight to each part of the human body, determined in advance.

3.2. 3D reconstruction of the human body

The 3D reconstruction of the human body is performed using Laulentini's method [12]. We extract the subject's

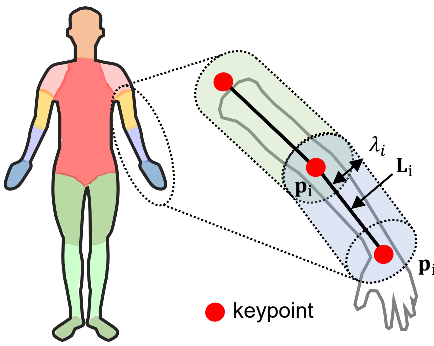


Figure 2: Variables in a segmented part.

2D silhouette from the input images (e.g., using [19]) and reproject the silhouettes to a 3D world. In case that the subject holds tools, we shall have a choice whether if we include and exclude the tools from the following CoM calculation. The common parts of the reprojected silhouette are the 3D shape of the body $\mathbf{V}$. $\mathbf{V}(\ni \mathbf{v}_j)$ denotes a set of voxels, where each voxel element $\mathbf{v}_j$ contains 3D positional information. In order to retrieve precise 3D model, cameras should be placed so that the common part is as small as possible. The arrangement of cameras depends on the posture of the subject and the number of cameras N. By reconstructing the 3D shape of the subject's body, it is possible to reflect any individual's unique figure.

3.3. Human kinematic structure estimation

CoM is the unique position at which the weighted position vectors of all the parts of a system sum up to zero. Because each body part has a different density [2], it is expected that assigning the appropriate weight to each body part will lead to more accurate CoM estimation. As shown in Figure 2, the 3D model reconstructed in Section 3.2 is divided into nine parts, head, body, shoulder, back arm, forearm, hand, thigh, calf, and toe.

We apply the method of Cao *et al*. [3] to the input images to obtain 18 keypoints, which represent the joints and face of an individual on the 2D image. By applying the direct linear transform to each 2D keypoint $\mathbf{q}$ to triangulate them, we obtain the 3D position $\mathbf{p}$ of each $\mathbf{q}$.

As shown in Figure 2, the 3D model $\mathbf{V}$ is segmented into $\mathbf{V}_i$ ($0 \leq i < 9$) based on the distance between line segments $\mathbf{L}_i$ connecting adjacent keypoints $\mathbf{p}$ and each voxel $\mathbf{v}_j$. Algorithm 1 shows the segmentation procedure. A voxel $\mathbf{v}_j$ which exists within a distance λ_i from $\mathbf{L}_i$ is classified as $\mathbf{V}_i$. A voxel $\mathbf{v}_j$ located in the common area of two or more body parts is classified as the part with the smaller distance. All voxels $\mathbf{v}_j$ that are not classified as any body part are removed. We weight the segmented model based on the weight of each part of the human body as reported by Lava

Algorithm 1: Proposed segmentation procedures

$\mathbf{V}_i$: A part of 3D human body model $\mathbf{V}$
$\mathbf{v}_j$: A voxel costituting the 3D model $\mathbf{V}$
$\mathbf{p}_i$, $\mathbf{p}'_i$: Keypoints that divide $\mathbf{V}$ into $\mathbf{V}_i$
$\mathbf{L}_i(\mathbf{p}_i, \mathbf{p}'_i)$: Line segment between $\mathbf{p}_i$ and $\mathbf{p}'_i$

```
foreach v_j do
    foreach L_i(p_i, p'_i) do
        D_i ← CalcDistance(v_j, L_i)
    end
    if Min(D_i) < λ_i then
        stock v_j to V_i
    else
        remove v_j
    end
end
```

[14]. The overall CoM of the human body $\mathbf{C}$ is computed via Eq. (1), which represents their weighted average:

$$\mathbf{C} = \frac{1}{M} \sum_{i=1}^{M} w_i \mathbf{v}_i \tag{1}$$

where M denotes the total number of voxels and w_i represents the weight assigned to $\mathbf{V}_i$.

4. Experiments

This section provides two performance evaluations of the proposed method using real data. First, we compare three methods to show that the proposed method outperforms in the accuracy in terms of center of posture (CoP) error metric [22]. Second, we present a 3D visualization of different performances to demonstrate that the proposed method has ability to provide meaningful 3D data for sports performance analysis.

4.1. Evaluating CoP accuracy

4.1.1 Setups

As shown in Figure 3, a force plate (TF-6090) and five cameras (GoPro, 30fps, 1920×1080 resolutions) are utilized in this evaluation. The intrinsic parameters of these cameras are estimated by Zhang's method [23] beforehand. These cameras are set so as to surround the force plate at $0°, 45°, 100°, 260°$, and $300°$ respectively, where $0°$ represents a face-on view of the subject, capturing the subject standing on the force plate. Three subjects (two male and one female) each stood on the force plate in four static postures: upright standing, single-leg standing, squatting, and bending forward. We extracted human region using a semi-

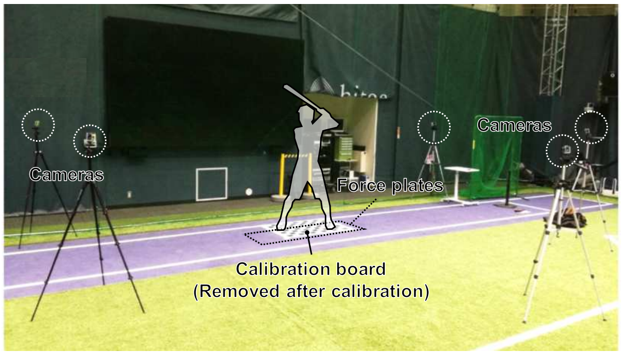


Figure 3: Experimental setup.

automated manner implemented on GIMP2 [1] in this experiment, to confirm pixel precise mask.

To demonstrate the performance of the proposed method, we compared it with the following two methods:

Uniform: Voxels with a uniform weight This method estimates the CoM as the center of the reconstructed 3D model in which all parts are assigned a uniform weight. The CoM is computed by Eq. (1) with all $w_i = 1$.

Articulated: Articulated joints model This method estimates the CoM as the center of the weighted articulated joint model. The CoM is computed by

$$C = \frac{1}{M'} \sum_{i=1}^{M'} w_i \boldsymbol{j}_i, \tag{2}$$

where $\boldsymbol{j}_i$ denotes the 3D positions of the mid-points of each joint and M' represents the number of midpoints. The 3D joint positions are computed by triangulation with the 2D joints detected by [3]

In this evaluations, the CoM estimation error of each method is evaluated as the Euclidean distance of the 2D coordinates of the center of pressure, $\boldsymbol{q}$, which represents the vertical projection of the estimated CoM as follows:

$$E_{COP} = |\boldsymbol{q} - \boldsymbol{q}_f|, \tag{3}$$

where $\boldsymbol{q}_f$ denotes the CoP estimated from a force plate. Note that the CoP estimated from a force plate is not always a completely accurate, but we consider a force plate provides a reference measurement for comparison with the proposed approach, since a force plate is commonly used for measuring the CoP in the practical use of sports performance measurement.

[1] https://www.gimp.org/

4.1.2 Results

Figure 4 shows input images from one view (first row), the reconstructed 3D model showing joint positions (second row), the labeled 3D model based on the joint positions (third row), and the 3D model with the estimated CoM (fourth row). The results in the second and third row show that the estimated 3D joint positions are sufficient to assign each voxel to the appropriate body parts.

Figure 5 shows the average estimation errors of each method. From these results, we observe that the proposed method outperforms the other methods and robustly estimates the CoM with errors of around 10 mm for CoP in all postures. In particular, the Uniform and Articulated methods show degraded performance in the cases for which there is some weight bias, such as squatting or bending forward. However, the proposed method estimates CoM robustly in these cases. In the case of standing upright, the estimation precision of all methods is similar due to the absence of weight bias among each body part.

The precision of all methods is greater in the case of single-leg standing than for squatting or bending forward. This is due to self-occlusion, which affects the precision of the reconstructed 3D model and the estimation of joint positions. Such self-occlusion is greater in the cases of squatting and bending forward than for single-leg standing.

4.2. CoM estimation for dynamic motion

4.2.1 Setups

Compared with the 2D CoP estimation approaches utilizing a force plate, the vision-based approach including the proposed method can estimate 3D positions of CoM, which is a significant advantage for analyzing a player's performance in a sports scene. In particular, it is known that CoP estimated with a force plate does not match the CoM projections when the subject is in motion, such as walking, running, and so on. Here we demonstrate that the proposed method can estimate the 3D positions of CoM in such a challenging situation.

In this evaluation, the four high-speed cameras (HAS-U2, 200fps, 1280$\times$1024 resolutions) are set so as to surround the subject, a professional baseball player. The subject swings a bat two times assuming (a) an inside pitch and (b) an outside pitch. We extracted the subject's regions in the same manner in the previous experiment (i.e., the bat held by the subject is excluded by masking it out).

4.2.2 Results

Figure 6 illustrates the 3D trajectory of the estimated CoM. The red and blue trajectories correspond to the case of (a) inside and (b) outside pitch, respectively. The subject is a left-handed batter and assumes that a ball is coming from

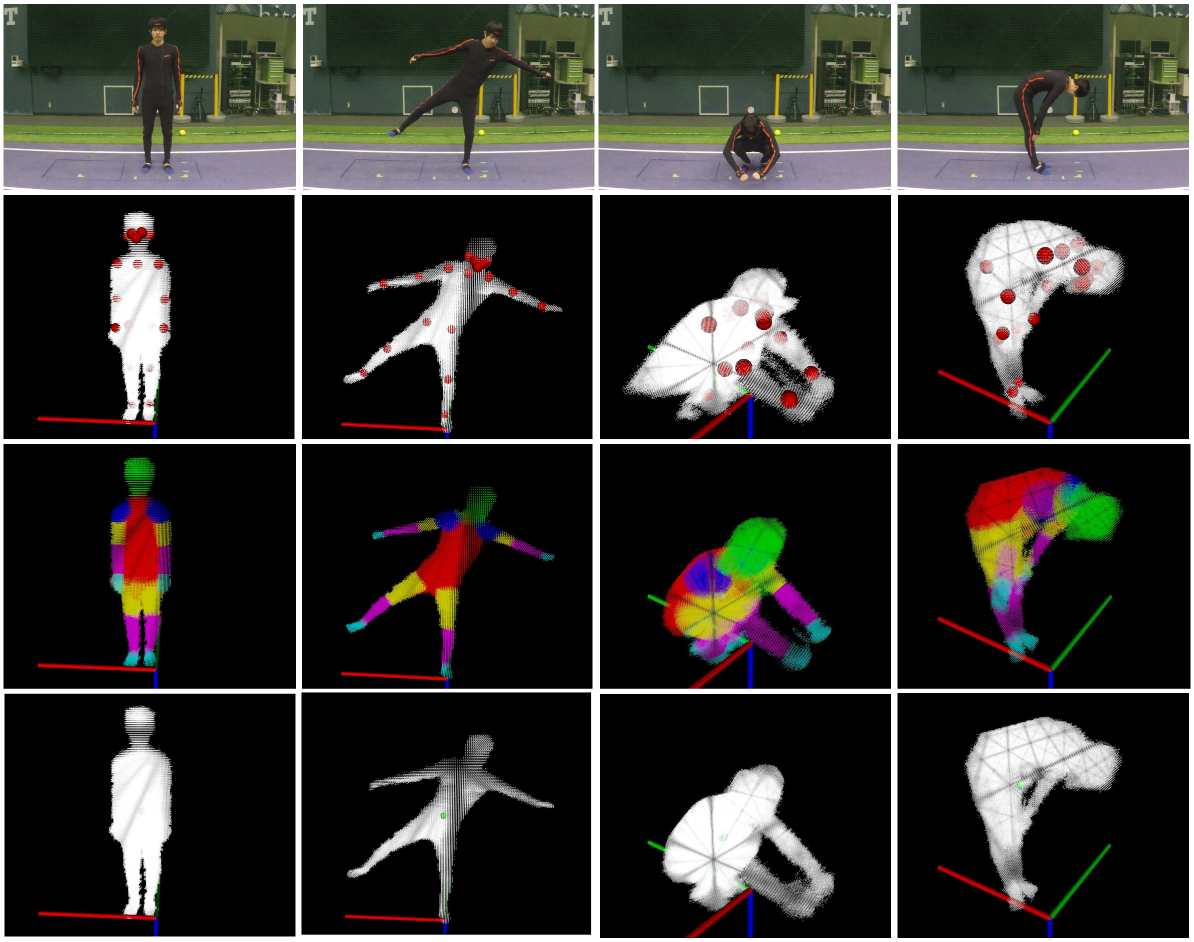


Figure 4: Experimental results on static posture: first row shows the posture of the subjects, second row shows the 3D model of human body as white dots and the keypoints as red spheres, third row shows the 3D model segmented into nine parts, and fourth row shows the estimated CoM (green sphere) in the reconstructed 3D model.

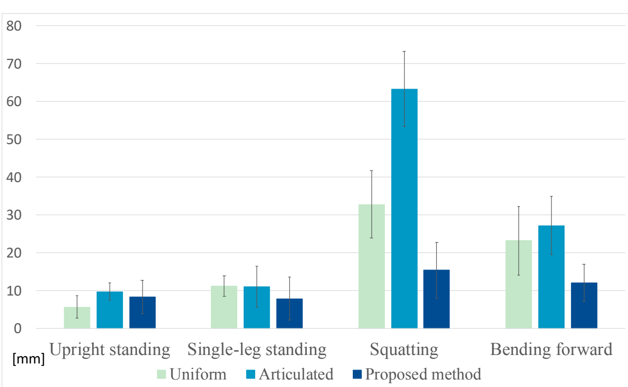


Figure 5: Error between reference value and vertically projected CoM estimated by each method.

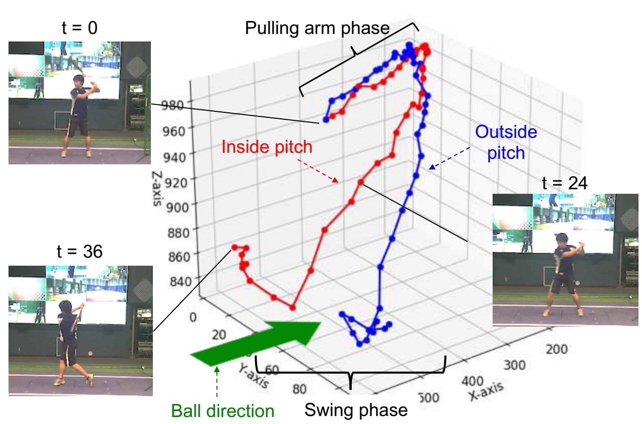
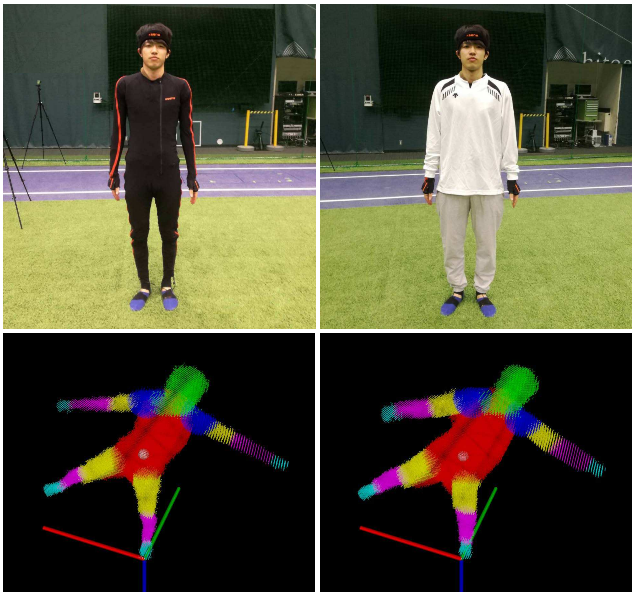


Figure 6: 3D CoM trajectories of batter swings against “inside (red)” and “outside (blue)” balls.



(a) Form-fitting clothes (b) Loose clothes

Figure 7: Comparison in the appearance between the subject wearing form-fitting and loose clothes.

positive to negative direction along the x-axis. Both trajectories are almost the same as the arms pull back and the gradually split in the swing phase. In the swing phase, we can see that the CoM for the outside pitch goes through the outside compared with the CoM for the inside pitch. From these results, we can conclude that the proposed method reasonably estimates the 3D trajectory of CoM in an active sports scenes without requiring the installation of devices such as force plates into the field and requiring the subject to wear any electronic devices. These visualized estimations of CoM allow analysis of the performance of a player in various situations, for example, CoM trajectories can be compared between training and a real sports game.

5. Discussions

Here, we discuss our future work of the proposed method to clarify the current limitations.

5.1. Effects of clothes

The proposed method estimates the CoM as the gravity point of a set of voxels. Therefore, one might suppose that clothes affect the performance since the subject's silhouette changes. Here, we additionally demonstrate the effects of clothes to the proposed method.

As shown in the first row of Figure 7, we utilized images from subjects wearing form-fitting and loose clothing as input for the proposed method. From the second row of Figure 7, we can see that the reconstructed 3D model with loose clothes is expanded compared with the subject wearing form-fitting clothes, even when the subject stands in the same posture. Figure 8 shows the quantitative results of such cases utilizing the same configuration introduced in Section 4.1.1. These results show that the proposed method degrades the accuracy when the subject is wearing loose clothes. Reducing the effect of loose clothing on the method's accuracy is an aim of our future works.

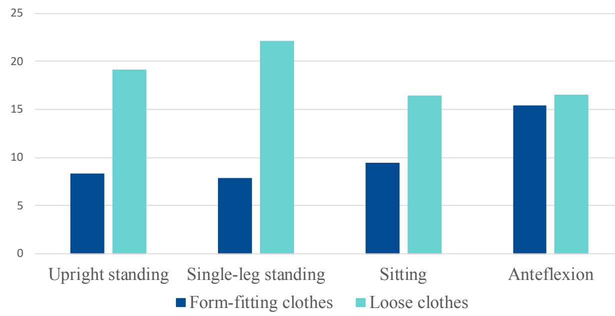


Figure 8: Comparison of CoP errors between subjects wearing form-firring and loose clothes.

5.2. Variations in segments

While we currently segment the body into nine parts, for further improvement in the accuracy, we will need more categories such as hair, clothes, shoes, tools, fat, bones, muscles, etc. In practice, we consider that fitting available anatomy models including the above-mentioned data to the subject will lead to the higher reliability.

5.3. Tools held by subjects

We need further investigation on whether if we should include tools held by subjects in CoM calculations (e.g., a bat, a racket, a golf club, etc.) from a viewpoint of sports data analysis, and how the difference appears in the CoM estimation. Note that we removed the bat to obtain the CoM of the subject only, in Figure 6, by masking the bat out in the proposed pipeline as described in Section 3.2.

5.4. Estimating multiple subjects' CoM

As the first step of our research, we currently assume one subject in a scene. If we take multiple subjects into considerations, we will definitely need an extension to separate each person in a voxel space. While the CNN-based bone estimation [3] can handle multiple persons in a single view, we need to identify persons in multiple images, which will need additional efforts.

6. Conclusion

This paper proposed a novel vision-based CoM estimation algorithm based on multi-view images for sports performance analysis. The key approach of the proposed method is to assign an appropriate weight to each voxel reconstructed in a visual hull manner. Evaluations with real data demonstrated that the proposed method can estimate the CoM with errors of about 10 mm in terms of CoP compared with the data measured with force plates in static conditions. In addition, the proposed method reasonably estimated the 3D trajectory of CoM in a dynamic scene.

References

[1] S. Barris and C. Button. A review of vision-based motion analysis in sport. *Sports Medicine*, 38(12):1025–1043, 2008.

[2] R. N. Baumgartner, W. C. Chumlea, and A. F. Roche. Estimation of body composition from bioelectric impedance of body segments. *The American J. of Clinical Nutrition*, 50(2):221–226, 1989.

[3] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Computer Vision and Pattern Recognition*, 2017.

[4] J. Carpentier, M. Benallegue, N. Mansard, and J. P. Laumond. Center-of-mass estimation for a polyarticulated system in contact-a spectral approach. *IEEE Trans. on Robotics*, 32(4):801–822, 2016.

[5] R. Dawes, M. Mann, B. Weir, C. Pike, P. Golds, and M. Nicholson. Enhancing viewer engagement using biomechanical analysis of sport. In *NEM Summit*, pages 121–126, 2012.

[6] A. González, M. Hayashibe, and P. Fraisse. Estimation of the center of mass with kinect and wii balance board. In *Intelligent Robots and Systems*, 2012.

[7] A. Gonzlez, 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):16955–16971, 2014.

[8] S. A. Hale, J. Hertel, and L. C. Olmsted-Kramer. The effect of a 4-week comprehensive rehabilitation program on postural control and lower extremity function in individuals with chronic ankle instability. *J. of Orthopaedic & Sports Physical Therapy*, 37:303–311, 2007.

[9] C. Hrysomallis. Relationship between balance ability, training and sports injury risk. *Sports Medicine*, 37(6):547–556, 2007.

[10] C. Hrysomallis. Balance ability and athletic performance. *Sports Medicine*, 41(3):221–232, 2011.

[11] R. T. Imamura, A. Hreljac, R. F. Escamilla, and W. B. Edwards. A three-dimensional analysis of the center of mass for three different judo throwing techniques. *J. of Sports Science & Medicine*, 5:122–131, 2006.

[12] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.

[13] A. Lees. Technique analysis in sports: a critical review. *J. of Sports Sciences*, 20(10):813–828, 2002.

[14] P. D. Leva. Adjustments to zatsiorsky-seluyanov's segment inertia parameters. *J. of Biomechanic*, 29(9):1223–1230, 1996.

[15] A. Mapelli, M. Zago, L. Fusini, A. C. D. Galante, and C. Sforza. Validation of a protocol for the estimation of three-dimensional body center of mass kinematics in sport. *Gait & Posture*, 39(1):460–465, 2014.

[16] B. Najafi, J. Lee-Eng, J. S. Wrobel, and R. Goebel. Estimation of center of mass trajectory using wearable sensors during golf swing. *J. of Sports Science & Medicine*, 14(2):354–363, 2015.

[17] Y. C. Pai and J. Patton. Center of mass velocity-position predictions for balance control. *J. of Biomechanics*, 30(4):347–354, 1997.

[18] M. Saini, D. C. Kerrigan, M. A. Thirunarayan, and M. Duff-Raffaele. The vertical displacement of the center of mass during walking: a comparison of four measurement methods. *J. of Biomechanical Engineering*, 120(1):133–139, 1998.

[19] A. Tsitsoulis and N. G. Bourbakis. A methodology for extracting standing human bodies from single images. *IEEE Trans. on Human-Machine Systems*, 45(3):327–338.

[20] C. M. Welch, S. A. Banks, F. F. Cook, and P. Draovitch. Hitting a baseball: A biomechanical description. *J. of Orthopaedic & Sports Physical Therapy*, 22(5):193–201, 2006.

[21] D. A. Winter, A. E. Patla, and J. S. Frank. Assessment of balance control in humans. *Medical Progress Through Technology*, 16(1):31–51, 1990.

[22] V. M. Zatsiorsky and D. L. King. An algorithm for determining gravity line location from posturographic recordings. *J. of Biomechanics*, 31(2):161–164, 1997.

[23] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.