Single Camera 3D Human Pose Estimation for Tele-rehabilitation

Oky Dicky Ardiansyah Prima, Takashi Imabuchi,
Yuta Ono, Yoshitoshi Murata, Hisayoshi Ito
Graduate School of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
e-mail: prima@iwate-pu.ac.jp, {g236o001, g231q005}@s.iwate-pu.ac.jp,
{y-murata, hito}@iwate-pu.ac.jp

Yukihide Nishimura
Department of Rehabilitation Medicine
Iwate Medical University
Morioka, Japan
e-mail: ynishi@iwate-med.ac.jp

Abstract—**The need of using advanced remote devices to promote effective self-management of rehabilitation has rapidly grown in developed countries. The widely spread camera-equipped mobile devices and Internet of Things (IoT) have been expected to deliver professional services by connecting clinician to client for assessment and consultation. This study proposes an IoT-based Tele-Rehabilitation (TR) framework using a single camera to observe the body joints of the client in three-dimensional (3D) space on performing Activities of Daily Living (ADL). Our experiments show that the proposed framework is capable to measure joint and orientation angles of elbow and knee comparable with the measurements using the Kinect. A waterproof camera was used to show that the proposed system can be extended to do the joint measurements during aquatic therapy and fitness pools.**

Keywords-rehabilitation; 3D human pose; ADL; range of motion; IoT.

I. Introduction

The growing number of elderly people and the decreasing number of healthcare professionals led to the implementation of Tele-Rehabilitation (TR) initiatives in healthcare regular practice [1]. TR is the provision of rehabilitation services from a distance using communication technologies. It is a developing field of telehealth to increase accessibility and enhancing continuity of care to individuals who are in geographically remote regions. TR enables clinicians to optimize the timing, intensity and duration of therapy effectively. The use of videoconferencing and virtual reality systems allows clinicians to interact with patients in real-time. Therefore, TR is expected to reduce the potential time and cost of rehabilitation services, especially for individuals who have economically disadvantaged. The future of TR is promising as a wide range of services to suit the needs of the individual.

Efforts have been made to conduct TR for physical therapy using low-cost depth sensors, such as Microsoft Kinect. There are two versions of Kinect: Kinect V1 and V2. Hereinafter, we simply referred both versions as Kinect. The Software Development Kit (SDK) enables developers to access body joints positions and orientations. Previous studies have shown that these devices show performance adequate for a range of healthcare imaging applications [2][3]. However, there are some concerns with occlusions and noises during tracking the joints [4]. Data smoothing techniques, such as Kalman filtering are required to minimize the problems. Moreover, a study of validity and reliability of the Kinect on measuring joint angles shows that 95% Limits of Agreement (LOA) between the Kinect and the goniometer exceeded ±5° which suggests a concern of using the Kinect in rehabilitation [5].

Despite the issues of the Kinect, many attempts have been made to use this device to support rehabilitation virtually. The Kinect can be used as a natural user input interface to the Virtual Reality (VR) system on carrying out physical rehabilitation therapies [6]. This system will enable the users to get an audiovisual feedback in real-time whether they are properly doing the specific therapy or not. The users can record their body joint movements during exercise and sent them to clinicians to get more advices on maintaining or improving the rehabilitation stages.

The Kinect is known to have some limitations on measuring body joints if some parts of the body are occluded. The Kinect needs to capture the full body of the user to measure the joint locations properly. Therefore, measuring a target user performing Activities of Daily Living (ADL) tasks where the lower limbs are hidden, such as dining and sleeping would be difficult to achieve.

The needs of TR are not limited to in-room rehabilitation programs. The basic ADL includes a functional mobility to move from one place to another while performing tasks, such as the ability to walk, get in and out of bed, and get into and out of a chair. Since the newest Kinect (Kinect V2) is only capable to measure targets up to 6m and inside 70° by 60° Field of View (FOV) from the its sensor, area measurements of these tasks are limited.

With the recent enhanced techniques utilizing Artificial Intelligent (AI), the state-of-the-art computer vision has enabled to measure body joints in three-dimensional (3D) using a single camera. This measurement takes two steps: joint localization in two-dimensional (2D) image coordinate and 3D coordinate estimation for each 2D joint. Depends on the method to localize joints, method based on bottom-up human pose estimation, such as “OpenPose” enables joint localization without capturing the full body beforehand [7]. Once this method localizes some joints, part affinity fields between joints are calculated to connect corresponding joints to estimate 2D human pose. Many studies have investigated the problem of inferring 3D joints from their 2D projections. These studies involve traditional 2D to 3D methods, which define the bone lengths and use a binary decision tree to estimate the 3D joints [8] or deep-net-based 2D to 3D methods

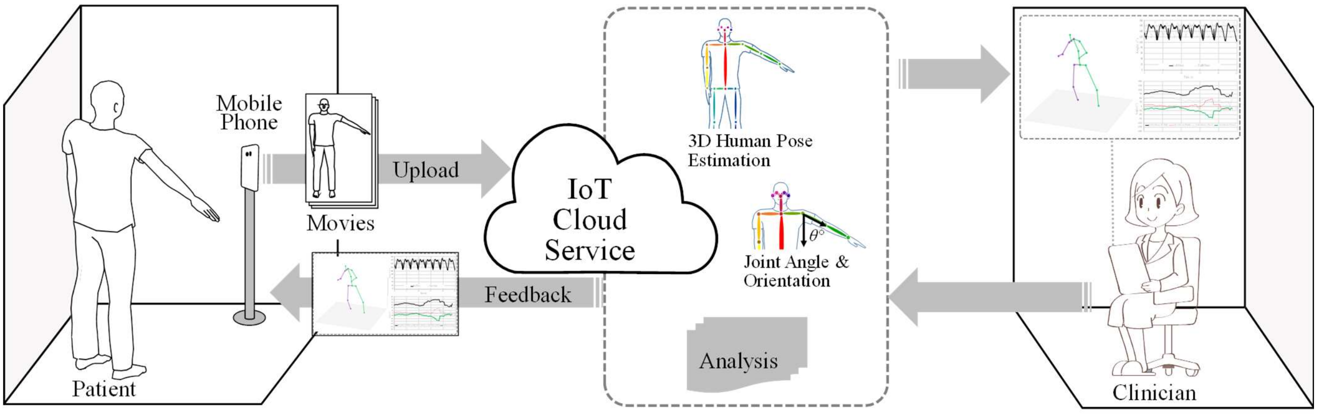


Figure 1. The proposed TR framework in this study.

which estimate 3D joints with Deep Neural Networks (DNN) [9]. The use of DNN may produce a high calculation costs to be performed on dedicated Graphics Processing Units (GPUs). However, this calculation can be performed in a server where the client can send the recorded movie to and get the resulted 3D joints using the Internet of Things (IoT) infrastructures.

This study proposes a framework to enable the use of a single camera 3D human pose estimation to estimate 3D joints and orientation of limbs which can be used in a broad range of TR services. Preliminary evaluations are made to investigate the accuracy of the estimated joint and orientation angles compared to the Kinect and to reveal its ability to handle joint occlusions on estimating the 3D joints. An attempt to estimate 3D joints for a user performing fitness pool is provided to open the further development in the future.

This paper is organized as follows. Section II describes related works on TR. Section III describes methods to measure joint and orientation angles of elbow and knee, and to perform our experiments. Section IV shows the accuracy of joint and orientation angles of elbow and knee measured in this study for a subject performing several tasks. Finally, Section V concludes the achievements and discusses the future prospective of this study.

II. RELATED WORK

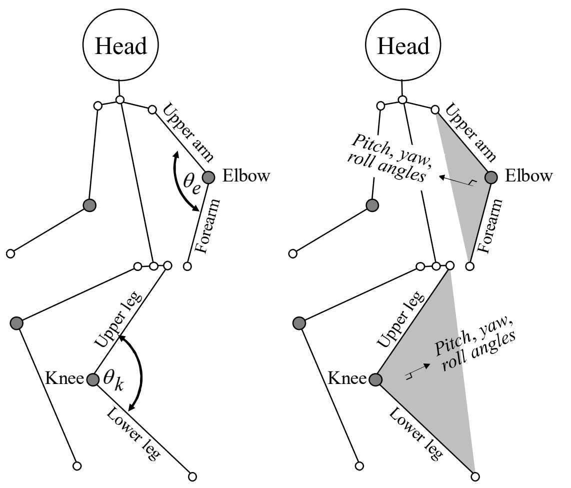
A single camera has a prospect to be a useful tool to assess a certain Range of Motion (ROM). DrGoniometer, an iPhone app, has enabled the manual measurement of patient articulation angles and storage of all related information to build up historical data for each articulation and movement [10]. For shoulder proprioception assessment, Mitchell et al. (2014) assessed the validity of DrGoniometer by measuring ROM of participants on performing active shoulder external rotation [11], where DrGoniometer was found to be comparable to the standard goniometer. For elbow proprioception assessment, Ferriero et al. (2011) found that DrGoniometer is reliable for elbow joint goniometry on measuring passive ROM of elbow [12].

Enhanced computer vision techniques enable the automated measurement of head and body pose. For cervical spine proprioception assessment, the Perspective-n-Point camera pose determination (PnP problem) [13] from a single camera can be used to measure head repositioning accuracy to diagnose people with neck disorders. The PnP problem estimates the relative position between camera and head posture from predefined 3D facial feature points (3D landmarks) and their corresponding points in camera coordinates. Head pose estimation based on the PnP problem has been implemented in some open source software libraries, such as OpenCV [14] and Dlib [15]. There are many popular photographic apps on Android and iPhone utilizing this technique to modify facial shapes. However, to our knowledge, there are no reports on the validity of the head pose measured by the PnP problem for a rehabilitation purpose. For ROM assessment, limb joints in 3D can be estimated from a targeted body captured by a single camera. The DNN have boosted the accuracy of the detection of joint locations from an image. OpenPose has enabled to localize multiple human body joint in real time by implementing the Part Affinity Fields (PAFs) to encode the location and orientation of limbs without capturing the full body [7]. Martinez et al. (2017) proposed a relatively simple deep feedforward network over the Human3.6M [16], the largest publicly available 3D human pose dataset containing 3.6 million human poses and corresponding images to estimate 3D joints from their 2D projections. This work has been made available as open source software, namely "3D-pose-baseline [17]."

The 3D joint estimation from a single camera is hard to be run on a mobile device because of its extensive calculation that has to be performed on GPUs. However, with the spread of IoT infrastructures, this calculation can be done in the cloud server. Using services over the Internet, clinicians can conduct physical TR with patients at their homes. This new service will provide not only a support for assistance with exercise but also smart data to maintain or improve the rehabilitation stages.

III. METHODS

The proposed TR framework was developed based-on OpenPose and 3D-pose-baseline. The framework processes user's videos and estimates the 3D human pose from the



(a) Joint angles (b) Joint orientation angles

Figure 2. Measurements of joint and orientation angles (θ_e, θ_k, *pitch*, *roll*, *yaw*) of elbow and knee in this study.

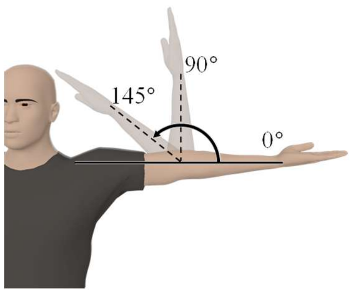


Figure 3. Elbow joint angle during side flexion.

videos. Body joint and orientation angles are measured, and these data are presented as time series plots concurrently with the 3D plot of the body posture. Here, the joint orientation angles consist of pitch, yaw, and roll angles. At this moment, only joint and orientation angles of elbow and knee are measured and visualized by the proposed framework. However, other joints can be measured with the same manners. Python programming language was used to process the calculations.

Our proposed framework, as shown in Figure 1, is based on the client-server model. Using mobile phones, patients can record their movements on performing ADL tasks and send the recorded videos to the IoT cloud service. The IoT will measure the body joint and orientation shortly after it detects a new incoming video. The measurement results are sent to clinicians where they can monitor the rehabilitation stages and give necessary advices. All results are stored in the database which can be accessed by the patients from their mobile devices.

A. Joint Angle Measurement

Joint angle was measured as a relative angle between the longitudinal axis of two adjacent segments. For the elbow joint angle, the adjacent segments are the upper arm and the forearm, respectively. Whereas, for the knee joint angle, the adjacent segments are the upper and the lower legs, respectively. Note that although the lower leg consists of a tibia and a fibula, only a single segment can be estimated from the 3D human pose estimation, as well as the Kinect's result.

Here, Figure 2(a) shows elbow and knee joint angles measured in this study. Let u and v be vectors representing two adjacent segments, the angle between u and v is equal to

$$\theta_{joints} = 180° - \frac{\bar{u} \cdot \bar{v}}{|\bar{u}||\bar{v}|}, \quad (1)$$

where *joints* represent elbow (e) and knee (k), respectively.

B. Joint Orientation

Joint orientation was measured as an orientation of a triangle plane constructed by two adjacent segments. Let

$$ax + by + cz + d = 0 \quad (2)$$

is a triangle-plane constructed by three 3D points as shown in Figure 2(b), the orientation (R) and translation (T) of this plane against pre-defined a reference triangle plane can be calculated using Singular Value Decomposition (SVD) [19]. Euler angles were calculated from R to derive the plane's orientation angles: pitch, yaw, roll angles.

C. Data Extraction and Analysis

Firstly, we measured right-hand's elbow joint and orientation angles of a subject while performing side flexion as shown in Figure 3, using a single camera and the Kinect concurrently. Occlusion is not expected to occur on this posture. The measurement results were compared to assess how the resulted measurements from the camera correspond to those from the Kinect. Secondly, we measured elbow and knee joint angles of a subject while performing the following tasks: swinging a tennis racket, running and running in place, and swimming. For swinging a tennis racket, the resulted joint angles from the camera were compared to those from the Kinect. Occlusion was expected to occur during this task. For the rest of the tasks, measurements were conducted using only a single camera which moves in accordance with the movement of the subject.

IV. RESULTS

We measured right-hand's elbow joint and orientation angles of a subject, as shown in Figure 4, while performing elbow side flexion. The subject was asked to stand upright and slowly perform this task until maximum ROM and move back to the initial pose. The minimum and maximum ROMs for the subject measured by a goniometer were 0° and 146°, respectively. During flexion, the camera measured the elbow joint angles which started from 48° and gradually increased to 122° as the peak. After reaching the peak, the angles gradually decreased to the initial angle. On the other hand, the Kinect measures with the same trend as the camera's but shows 10° and 135° for the minimum and maximum ROMs, respectively. This result indicates that the Kinect yields better absolute accuracies in joint angle measurements. For orientation, a

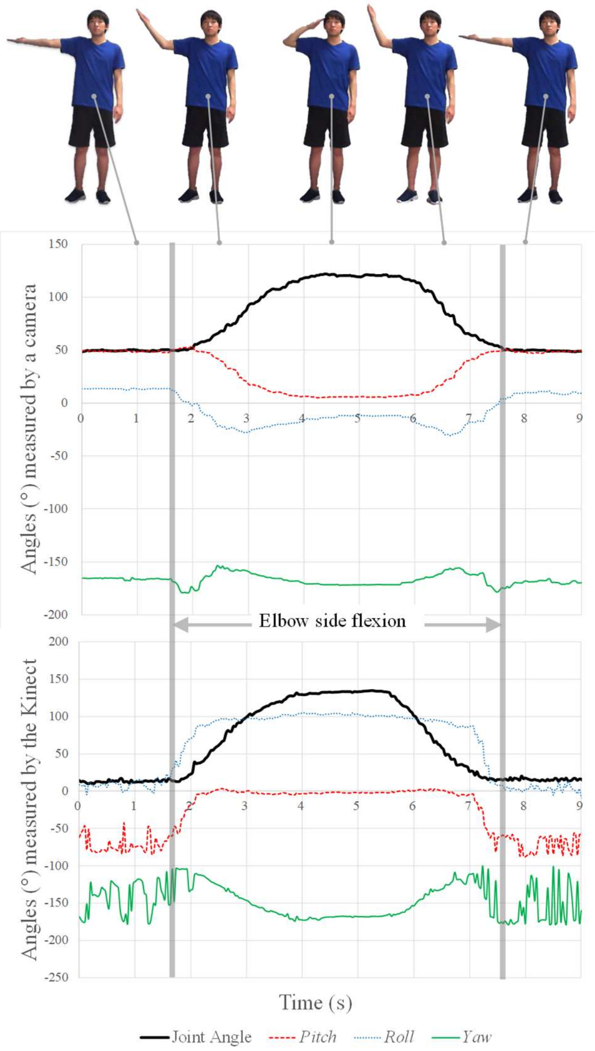


Figure 4. Right-hand's elbow joint and orientation angles of a subject while performing elbow side flexion.

great agreement was achieved for yaw angles. However, pitch and roll angles measured from the camera were inversely proportional to those from the Kinect. We considered that this problem was caused by the difference in the coordinate system representation that need to be adjusted for body joints between the proposed framework and the Kinect. As a known issue [4], noises were observed from the resulted measurements using the Kinect.

Elbow and knee joint angles of a subject while swinging a tennis racket, as shown in Figure 5, were measured by a single camera and the Kinect. Measurement scenes (P_1 - P_7) indicate occlusions of the subject's arms. Overall, the trend of the resulted joint angles from the camera agrees with those from the Kinect. However, there are gaps as occlusions occurred, such as in scenes from P_3 to P_7. The Kinect failed to measure the left-hand's elbow joint angles, which is highly occluded in these scenes. Table I shows 3D human pose estimated from the camera and the Kinect in P_4 and P_6 scenes. It is obvious

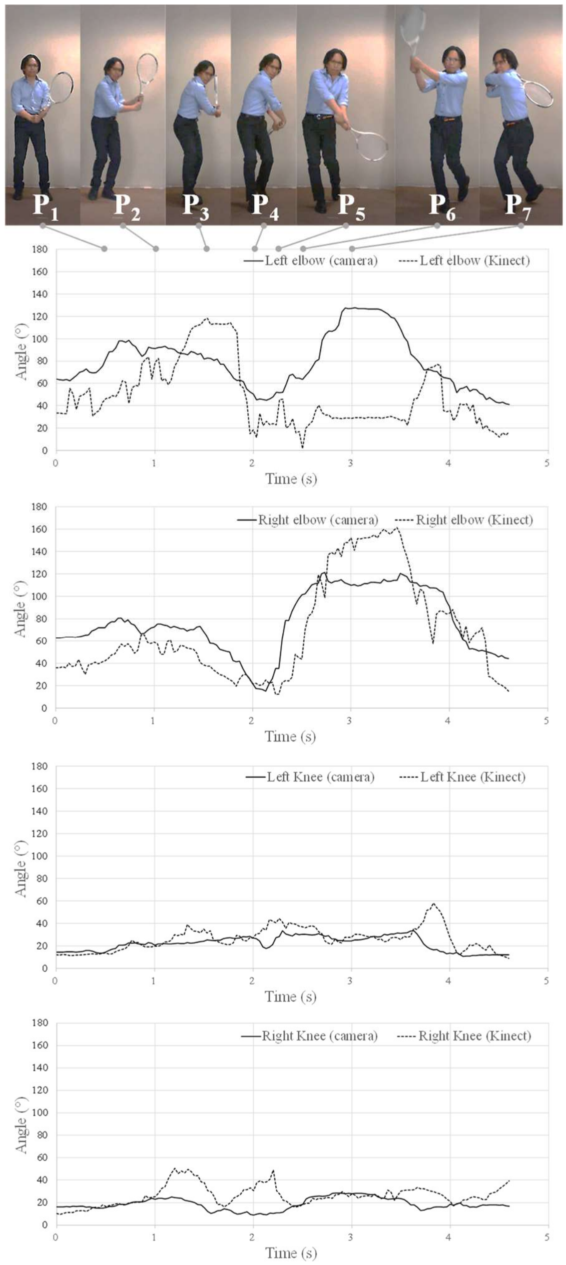
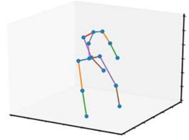
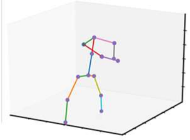
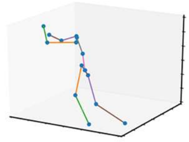
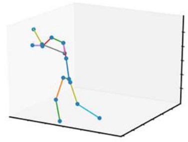


Figure 5. Measurements of elbow and knee joint angles of a subject while swinging a tennis racket.

that the camera estimated the 3D human pose better than the Kinect because the OpenPose used in this study works better on handling occlusion to determine the joint location. Hence, the joint angles measured from the camera is more reliable the those from the Kinect.

We measured elbow and knee joint angles of a subject on performing two tasks: running and running in place, measured by a single camera. During the measurements, the right arm and the right leg of the subject, as shown in Figure 6, were occluded from time to time. Ideally, fluctuations of left elbow

TABLE I. COMPARISON OF 3D HUMAN POSE ESTIMATION BY A SINGLE CAMERA AND BY THE KINECT

Scene	Posture	Camera	Kinect
P_4			
P_6			

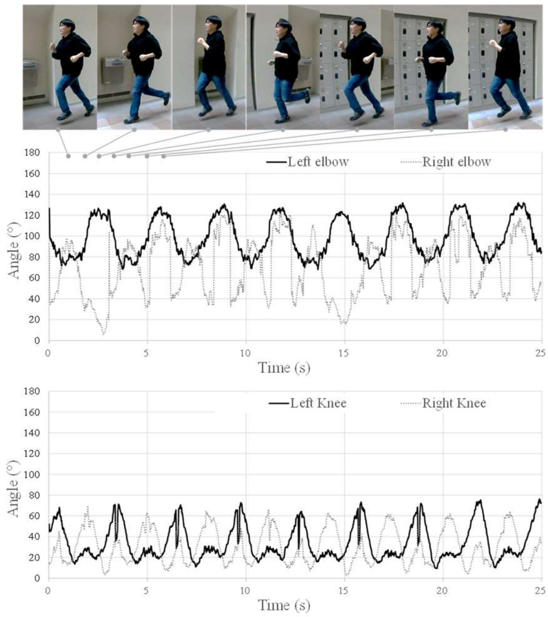
and left knee joint angles are identical with that of right elbow and right knee but half-cycle shifted in phase. On scenes where the entire right arm was mostly occluded by the body, the proposed framework estimated the posture of the right arm to be greatly extended behind the body. Thus, high values of the elbow joint angles were measured during these scenes. On the other hand, despite the right leg was sometimes partly occluded by the left leg, good results of knee joint angles were derived.

Finally, Figure 7 shows joint and orientation angles of elbow and knee of a subject entering a swimming pool before starting to swim. A waterproof camera was used to record the scene. The estimated 3D human pose was presented to show that pose estimation was not affected by the underwater light environment. Measurement of joint and orientation angles can be done in the same manner as on land. At this moment, detailed validation for the measurement results has not been carried out but we will report the analysis results in our next paper.

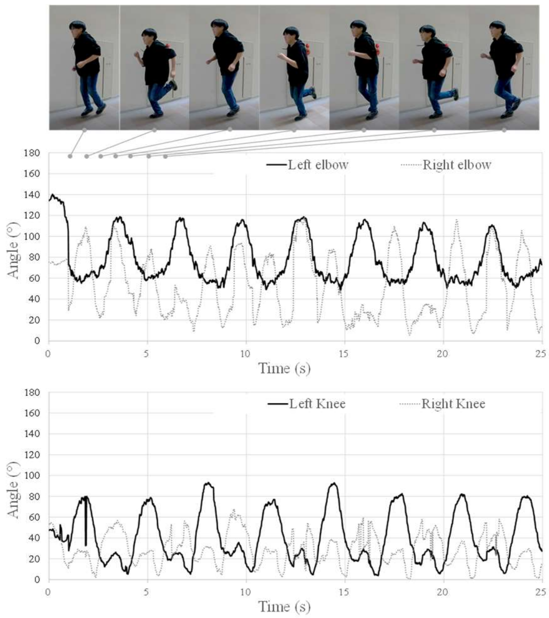
V. CONCLUSION AND FUTURE WORK

In this study, we have proposed a tele-rehabilitation framework where patients can send recorded videos during performing ADL tasks and get feedbacks to visualize their body joint and orientation angles to maintain or improve the rehabilitation stages. The proposed framework is more robust than the Kinect on handling occlusion; thus, it opens the possibility to measure body joint and orientation angles while doing ADL tasks where a part of limbs is hidden, such as dining and sleeping. The resulted measurements from the proposed framework are less sensitive to noises than those from the Kinect. Although the Kinect shows better results in term of absolute accuracies, the relative accuracy of the proposed framework against the Kinect is acceptable. The proposed framework can be extended to do the same measurements during aquatic therapy and fitness pools.

Future works include conducting experiments to measure body joint and orientation angles of a number of participants while doing ADL tasks where a part of limbs is hidden, and performing details analyses to determine validity of the proposed framework.



(a) Running



(b) Running in place

Figure 6. Measurements of elbow and knee joint angles of a subject on performing two tasks: running and running in place.

ACKNOWLEDGMENT

We acknowledge the effort from the authors of OpenPose and 3d-pose-baseline to make 3D joint measurements using a single camera possible. This work was supported by the MIC/SCOPE #181602007.

REFERENCES

[1] M. H. A Huis In't Veld, H. van Dijk, H. J. Hermens, and M. M. R. Vollenbroek-Hutten, "A systematic review of the

 ISBN: 978-1-61208-688-0

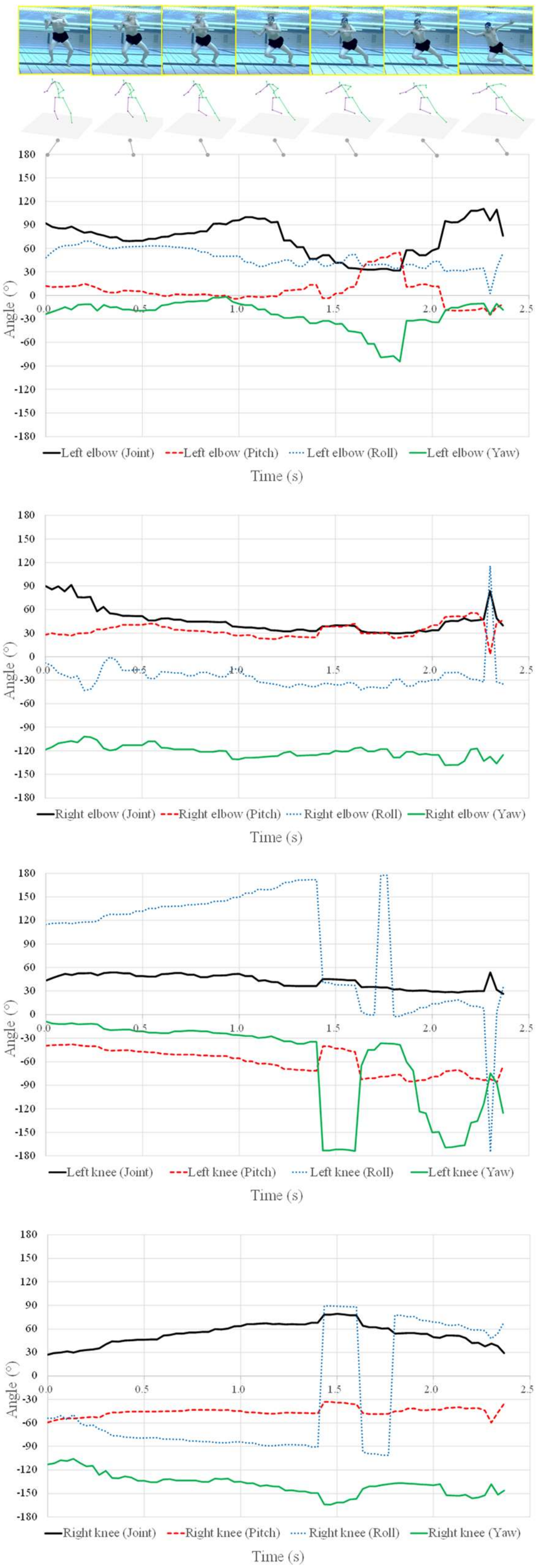


Figure 7. Measurements of elbow and knee joint and orientation angles of a subject entering a swimming pool before starting to swim.

methodology of telemedicine evaluation in patients with postural and movement disorders," Journal of Telemedicine and Telecare, vol. 12, no. 6, pp. 289-97, 2006.

[2] S. Aleesandro, C. Andrea, M. Matteo, and M. T. Lorenzo, "Kinect V2 Performance Assessment in Daily-Life Gestures: Cohort Study on Healthy Subjects for a Reference Database for Automated Instrumental Evaluations on Neurological Patients," Applied Bionics and Biometrics, pp. 1-16, 2017, https://doi.org/10.1155/2017/8567084.

[3] S. H. Lee et al. , "Measurement of Shoulder Range of Motion in Patients with Adhesive Capsulitis Using a Kinect," PLOS ONE vol. 10, no. 6, pp. 1-12, 2015, doi:10.1371/journal.pone.0129398.

[4] H. M. Hondori and M. Khademi, "A Review on Technical and Clinical Impact of Microsoft Kinect on Physical Therapy and Rehabilitation," Journal of Medical Engineering, vol. 2014, pp. 1-16, 2014, http://dx.doi.org/10.1155/2014/846514.

[5] M. E. Huber, A. L. Seitz, M. Leeser, and D. Sternad, "Validity and reliability of Kinect skeleton for measuring shoulder joint angles: a feasibility study," Physiotherapy, vol. 101, no. 4, pp. 389–393, 2015.

[6] M. Pedraza-Hueso, S. Martín-Calzóna, F. J. Díaz-Pernasa, and M. Martínez-Zarzuela, "Rehabilitation Using Kinect-based Games and Virtual Reality," Procedia Computer Science, vol. 75, pp. 161 – 168, 2015.

[7] Z. Cao, T. Simon, S.E. Wei, Y. and Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," Computer Vision andPattern Recognition 2017, pp. 1302-1310, 2017.

[8] H. J. Lee and Z. Chen, "Determination of 3D Human Body Postures from a Single View," Computer Vision, Graphics and Image Processing, vol. 30, pp. 148–168, 1985.

[9] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A Simple Yet Effective Baseline for 3d Human Pose Estimation," arXiv preprint arXiv:1705.03098, pp. 1-10, 2017.

[10] DrGoniometer. http://www.drgoniometer.com/ [retrieved: January, 2019]

[11] K. Mitchell, S. B. Gutierrez, S. Sutton, S. Morton, and A. Morgenthaler, "Reliability and Validity of Goniometric iPhone Applications for the Assessment of Active Shoulder External Rotation," Physiotherapy Theory and Practice, vol. 30, no. 7, pp. 521–525, 2014.

[12] G. Ferriero, F. Sartorio, C. Foti, D. Primavera, E. Brigatti, and S. Vercelli, "Reliability of a New Application for Smartphones (DrGoniometer) for Elbow Angle Measurement.," PM&R, vol. 3, no. 12, pp. 1153–1154, 2011.

[13] L. Kneip, D. Scaramuzza, and R. Siegwart, "A Novel Parametrization of the Perspective-three-point Problem for a Direct Computation of Absolute Camera Position and Orientation," Proc. CVPR, pp. 2969–2976, 2011.

[14] OpenCV. https://opencv.org/ [retrieved: January, 2019]

[15] Dlib C++ Library. http://dlib.net/ [retrieved: January, 2019]

[16] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 7, pp. 1325-1339, 2014.

[17] 3D-pose-baseline. https://github.com/una-dinosauria/3d-pose-baseline. [retrieved: January, 2019]

[18] S. Obdržálek, G. Kurillo, J. Han, T. Abresch, and R. Bajcsy, "Real-time human pose detection and tracking for tele-rehabilitation in virtual reality," Studies in Health Technology and Informatics, vol. 173, pp. 320-324, 2012.

[19] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares Fitting of Two 3-D Point Sets," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 9, pp. 698–700, 1987.