APS: A 3D Human Body Posture Set as a Baseline for Posture Guidance

Hesham Elsayed
elsayed@tk.tu-darmstadt.de
TU Darmstadt
Darmstadt, Germany
Honda Research Institute Europe
Offenbach, Germany

Martin Weigel
martin.weigel@honda-ri.de
Honda Research Institute Europe
Offenbach, Germany

Julius von Willich
willich@tk.tu-darmstadt.de
TU Darmstadt
Darmstadt, Germany

Markus Funk
funk@tk.tu-darmstadt.de
TU Darmstadt
Darmstadt, Germany

Max Mühlhäuser
max@informatik.tu-darmstadt.de
TU Darmstadt
Darmstadt, Germany

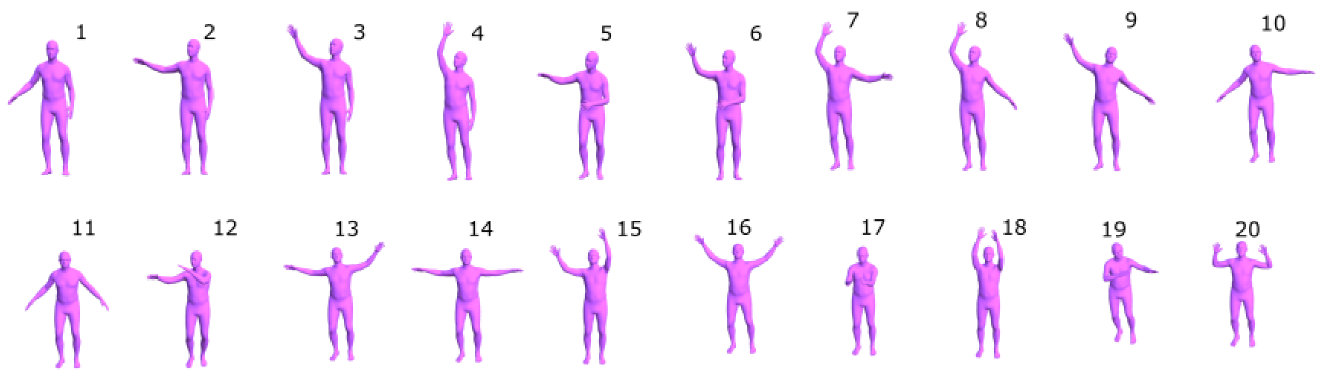

Figure 1: Body postures 1-20 in the APS Posture Set

ABSTRACT

Human body postures are an important input modality for motion guidance and other application domains in HCI, e.g. games, character animations, and interaction with public displays. However, for training and guidance of body postures prior research had to define their own whole body gesture sets. Hence, the interaction designs and evaluation results are difficult to compare, due to a lack of a standardized posture set. In this work, we contribute APS (APS Posture Set), a novel posture set including 40 body postures. It is based on prior research, sports, and body language. For each identified posture, we collected 3D posture data using a Microsoft Kinect. We make the skeleton data, 3D mesh objects and SMPL data available for future research. Taken together, APS can be used to facilitate design of interfaces that use body gestures and as a reference set for future user studies and system evaluations.

CCS CONCEPTS

• **Human-centered computing** → *Human computer interaction (HCI)*; *Interaction design.*

KEYWORDS

human body postures; 3D dataset; whole body input

ACM Reference Format:
Hesham Elsayed, Martin Weigel, Julius von Willich, Markus Funk, and Max Mühlhäuser. 2019. APS: A 3D Human Body Posture Set as a Baseline for Posture Guidance. In *The 12th PErvasive Technologies Related to Assistive Environments Conference (PETRA '19), June 5–7, 2019, Rhodes, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3316782.3324012

PETRA '19, June 5–7, 2019, Rhodes, Greece

ACM ISBN 978-1-4503-6232-0/19/06…$15.00
https://doi.org/10.1145/3316782.3324012

Figure 2: Body postures 21-40 in the APS Posture Set.

1 INTRODUCTION

With the emergence of posture-sensing technologies human body postures have the potential to replace and augment existing input modalities, such as touch and speech. This paradigm shift is enabled by advances in sensing technologies, which make tracking of user's body motion easier and cheaper. For instance, body postures can be sensed using a sequence of RGB images [4], depth images [20] (e.g., Microsoft Kinect), and IMU data [13, 26].

In Human-computer Interaction (HCI) literature, a common use-case for body-based interaction is motion guidance: for full-body motion training [2, 12, 25]; learning physical activities, such as dancing [6], martial arts [7] and sports [22]; gesture guidance [21] and learning [8]; interaction with public displays [1, 27]; spatial guidance [10, 11]; and physiotherapy [23]. With the multitude of proposed systems and methods, varying across body parts, modalities used, technical setups and evaluation metrics, drawing conclusions between the various approaches becomes difficult. For instance, while YouMove introduced by Anderson et al. [2] and Physio@Home introduced by Tang et al. [23] both employ augmented reality for guiding motion, they use different visualizations for augmented feedback, capture the user from different views and train/guide different body parts (full-body movements in YouMove and shoulder movements in Physio@Home). Although motion and posture guidance systems in HCI have often been used in sports and physical activities, the interaction techniques are directly applicable in assistive environments, e.g for training workers manufacturing tasks [5].Drawing conclusions between different approaches and interaction techniques is currently not possible due to the different postures used.

In this work, we introduce *APS (APS Posture Set)*, a 3D human body posture dataset. APS is intended as a baseline for posture training and system evaluations. The benefits of a predefined standard benchmark for evaluation have been established in the HCI community. For instance, MacKenzie and Soukoreff [17] introduce a standard phrase set for evaluating text entry techniques. Funk et al. [9] introduce a standardized task to support interactive augmented reality tasks for assembly instructions. Thus, we introduce a compact set of postures to enable development and evaluation of approaches for posture guidance and training.

2 METHODOLOGY

While constructing our baseline posture set, we referred to prior work utilizing full-body and mid-air gestures for interaction. Karam and Schraefel [14] introduced a taxonomy of gesture styles identifying five different classes of gestures, namely: gesticulation, manipulations, semaphores, deictic and language gestures. We analyze static semaphoric gestures as these are person-centric gestures independent of spatial and object-specific information. As the mapping between this type of gesture and meaning is often unrelated and needs to be learned, we extracted postures from interactions with public displays, aircraft marshalling, the semaphore flag signaling systems and warm-up exercises. In cases where several postures were visually indistinguishable from each other, only one posture was kept, e.g posture 4 in APS corresponds to postures for the letter 'D' or number '4' in the semaphore flag signalling system and the Wing-walker/guide signal in aircraft marshalling [3]. Similarly, for postures that were the result of a reflection across the sagittal plane, only one posture was kept, e.g the letters 'W' and 'O' in the semaphore flag signalling system. Figures 1 and 2 depict the single postures in APS.

For visualizing postures we use Posebits [18]. Posebits are conditions defined on geometric relationships between body parts. These conditions can be easily defined or generated, e.g given a posture an example posebit is "is the right hand above shoulder level?". Structuring a set of postures based on posebits results in a binary tree, with each node containing the set of postures constrained by the conditions of nodes higher in the tree. Similar to decision trees, the posebit to use for splitting at a certain node is chosen to maximize information gain. Figure 3 shows the postures in APS structured according to the posebit "is a hand above shoulder level?".

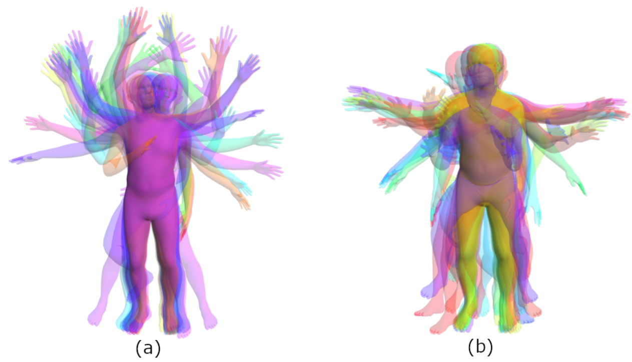

Figure 3: Overlapping postures showing (a) all postures with at least one hand above shoulder level and (b) all postures with no hands above shoulder level in the APS dataset

Using posebits supports straightforward generation of meaningful structure from a set of postures. Additionally, due to its coarse- to fine-grained nature, extracting prototypical postures becomes uncomplicated.

3 APS DATA COLLECTION

APS consists of skeleton data, 3D mesh objects of the postures and the parameters required to generate 3D body models in all postures. The 40 postures were performed by a single user. The 3D body models are approximated from the skeleton data recorded using a Microsoft Kinect Version 2, and can be used when using a different motion tracking setup than the Microsoft Kinect.

The Kinect provides skeleton tracking of 25 joints in 3D updated at 30 Hz. The Kinect skeleton includes sufficient joints of major body parts—such as the wrists, elbows and knees—to represent all postures in APS. We manually pick a frame where the posture is correctly performed and the joints are detected, from which the 3D skeleton data are read and saved as JSON. The bones are structured as a tree according to the hierarchy of the Kinect skeleton.

In addition to the Kinect data, we represent postures as statistical 3D body models in the SMPL format [16]. SMPL decomposes the 3D body representation into shape and pose parameters. The person-specific shape parameters are a set of coefficients representing the person's body shape according to the principal components responsible for the greatest variance in shape learned by the model. The pose parameters are the relative orientations of the joints across the kinematic chain.

In comparison with the Kinect skeleton data, SMPL support: (1) interfaces that depend on the user's body shape, e.g by projecting information on body parts, (2) consistent placement of wearable devices across users and (3) realistic visualizations in augmented- and virtual environments. The SMPL model uses a skeleton with 23 joints, where the exact joint locations depend on the body shape. The joint information delivered by the Kinect is used to fit a SMPL model by minimizing joint position error and taking into consideration posture constraints (as in Varol et al. [24]). To this end, the joints used are the wrists, elbows, shoulders, hips, knees, and ankles which are present in both the kinect and SMPL skeleton rigs.

4 THE APS POSTURE SET

We introduce the APS posture set consisting of 40 distinct postures as baseline postures for evaluating posture recognition, posture training, and motion guidance. Our 40 postures are depicted in Figure 1 and Figure 2.

Our posture set comes with three different attributes per posture: (1) posture name (2) posture reference image (3) posture reference skeleton data in the SMPL format (as described in section 3).

(1) The posture name is there to distinguish the postures from each other and naming them in instructions.

(2) As depicted, we rendered our reference postures using a pink avatar. These pictures of reference postures can be used as visual baseline descriptions of the postures.

(3) The posture skeleton data can be used by researchers to computationally compare an actual posture to a baseline posture and calculate a difference from two positions. This can be used for example when calculating visual, auditory, or tactile instructions for giving feedback according to a posture.

We want to solve the problem of current systems, which are not comparable due to differences in user study setups, when teaching users how to learn postures and therefore we make the APS posture set available for download for the research community[1]. We intend that by offering a compact set of standard postures, the scientific community can benefit from increased comparability of existing and future posture and motion guidance systems.

5 APPLICABILITY

We envision our APS posture set to be useful in a variety of domains, such as industrial assistance systems, general evaluation of posture guidance systems or for interaction using full-body gestures.

Industrial Assistance Systems

Industrial assistance systems are increasingly incorporating sensing technology in the workplace, offering many possibilities for natural body-based input. Posture guidance systems, typically using wearable displays for output, can be used for training in assistive environments [5]. The postures in APS

[1]Download the APS posture set here: http://makufunk.de/download/APS.zip

are not directed towards a particular scenario in the industry, but are rather generic covering a variety of possible postures assumable by the human body. For more specific scenarios, such as teaching particular assembly tasks or ensuring proper posture while loading/unloading packages from transport vehicles, APS can be used as a valuable baseline for evaluating interaction techniques before developing approaches that target such scenarios specifically.

Evaluation of Posture Guidance

One of the primary usages of APS is the comparison of different posture guidance applications and their parameters. To this end, we suggest using posture completion time to evaluate various guidance techniques. The time to complete a posture is measured from the start of the guidance, i.e. as soon as the user receives information on the posture to be performed, until the correct posture is performed. In cases where the user is unable to perform a particular posture, an error measure such as the 3D joint position error of the aligned skeletons could be appropriate. To be able to associate completion times of postures across user studies, the user starts each trial standing upright with arms relaxed to the side.

Full-body Interaction

Postures in APS are mostly based on the semaphore flag and aircraft marshalling signalling systems, which were designed for communication over distance. The postures are therefore clearly visually distinguishable from each other and are suitable as full-body gestures for interaction, e.g with public displays or smart devices.

6 DISCUSSION & FUTURE WORK

Human body interactions are rich in information and gestures. Body posture capture only a small—but important—part of this information. Our proposed dataset does not include *face and hand gestures*. Prior work has shown that capturing such aspects is possible with existing technologies. For example using articulated hand models [19] or face models [15]. However, we excluded these gestures, since they seem to be orthogonal to our proposed whole-body postures. They can be added in the future to further increase the expressiveness of our dataset and to enable novel interaction techniques.

The focus of APS is on static postures. These can be seen as keyframes of *body movements* (e.g., the holding postures of a yoga workout). However, some applications might require more detail about the movement, e.g speed and expression. We omitted such details in this dataset, since they would drastically increase its complexity. We hope to extend APS by augmenting it with existing movement notations (e.g. Labannotation) in the future to describe the transition between two postures.

7 CONCLUSION

In this work, we presented APS, a novel 3D human posture set. It is designed to be used in user studies and system evaluations of posture guidance applications. We clustered 40 postures in APS according to posebits [18] to extract meaningful structures resembling their appearance. We contributed the Kinect skeleton data, 3D meshes, and SMPL models of all APS postures to enable researchers working on the same posture data (see supplementary material). We discuss the usage of this data along with common measurement metrices used in HCI. We are confident that such a dataset is an important step to increase the comparability of different motion guidance systems and their training techniques.

REFERENCES

[1] Christopher Ackad, Andrew Clayphan, Martin Tomitsch, and Judy Kay. 2015. An In-the-wild Study of Learning Mid-air Gestures to Browse Hierarchical Information at a Large Interactive Public Display. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 1227–1238. https://doi.org/10.1145/2750858.2807532

[2] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: Enhancing Movement Training with an Augmented Reality Mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 311–320. https://doi.org/10.1145/2501988.2502045

[3] CAA Civil Aviation Authority. 1997. Visual Aids Handbook A compendium of Visual Aids intended for the guidance of Pilots and Personnel engaged in the handling of aircraft.

[4] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter V Gehler, Javier Romero, and Michael J Black. 2016. Keep it {SMPL:} Automatic Estimation of 3D Human Pose and Shape from a Single Image. *CoRR* abs/1607.08128 (2016). arXiv:1607.08128 http://arxiv.org/abs/1607.08128

[5] Sebastian Büttner, Henrik Mucha, Markus Funk, Thomas Kosch, Mario Aehnelt, Sebastian Robert, and Carsten Röcker. 2017. The Design Space of Augmented and Virtual Reality Applications for Assistive Environments in Manufacturing: A Visual Approach. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17)*. ACM, New York, NY, USA, 433–440. https://doi.org/10.1145/3056540.3076193

[6] Jacky C. P. Chan, Howard Leung, Jeff K. T. Tang, and Taku Komura. 2011. A Virtual Reality Dance Training System Using Motion Capture Technology. *IEEE Trans. Learn. Technol.* 4, 2 (April 2011), 187–195. https://doi.org/10.1109/TLT.2010.27

[7] Philo Tan Chua, Rebecca Crivella, Bo Daly, Ning Hu, Russ Schaaf, David Ventura, Todd Camill, Jessica Hodgins, and Randy Pausch. 2003. Training for Physical Tasks in Virtual Environments: Tai Chi. In *Proceedings of the IEEE Virtual Reality 2003 (VR '03)*. IEEE Computer Society, Washington, DC, USA, 87—-. http://dl.acm.org/citation.cfm?id=832289.835983

[8] Dustin Freeman, Hrvoje Benko, Meredith Ringel Morris, and Daniel Wigdor. 2009. ShadowGuides: Visualizations for In-situ Learning of Multi-touch and Whole-hand Gestures. In *Proceedings of the ACM*

International Conference on Interactive Tabletops and Surfaces (ITS '09). ACM, New York, NY, USA, 165–172. https://doi.org/10.1145/1731903.1731935

[9] Markus Funk, Thomas Kosch, Scott W. Greenwald, and Albrecht Schmidt. 2015. A Benchmark for Interactive Augmented Reality Instructions for Assembly Tasks. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15)*. ACM, New York, NY, USA, 253–257. https://doi.org/10.1145/2836041.2836067

[10] Markus Funk, Sven Mayer, and Albrecht Schmidt. 2015. Using In-Situ Projection to Support Cognitively Impaired Workers at the Workplace. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*. ACM, New York, NY, USA, 185–192. https://doi.org/10.1145/2700648.2809853

[11] Sebastian Günther, Florian Müller, Markus Funk, Jan Kirchner, Niloofar Dezfuli, and Max Mühlhäuser. 2018. TactileGlove: Assistive Spatial Guidance in 3D Space Through Vibrotactile Navigation. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference (PETRA '18)*. ACM, New York, NY, USA, 273–280. https://doi.org/10.1145/3197768.3197785

[12] Thuong N Hoang, Martin Reinoso, Frank Vetere, and Egemen Tanin. 2016. Onebody: Remote Posture Guidance System Using First Person View in Virtual Environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, 25:1—-25:10. https://doi.org/10.1145/2971485.2971521

[13] Yinghao Huang, Manuel Kaufmann, Emre Aksan, Michael J Black, Otmar Hilliges, and Gerard Pons-Moll. 2018. Deep Inertial Poser: Learning to Reconstruct Human Pose from Sparse Inertial Measurements in Real Time. In *SIGGRAPH Asia 2018 Technical Papers (SIGGRAPH Asia '18)*. ACM, New York, NY, USA, 185:1—-185:15. https://doi.org/10.1145/3272127.3275108

[14] Maria Karam and M.C. Schraefel. 2005. *A taxonomy of gestures in human computer interactions*. Technical Report. Electronics and Computer Science, University of Southampton.

[15] Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. 2017. Learning a Model of Facial Shape and Expression from 4D Scans. *ACM Trans. Graph.* 36, 6, Article 194 (Nov. 2017), 17 pages. https://doi.org/10.1145/3130800.3130813

[16] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-person Linear Model. *ACM Trans. Graph.* 34, 6, Article 248 (Oct. 2015), 16 pages. https://doi.org/10.1145/2816795.2818013

[17] I Scott MacKenzie and R William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 754–755. https://doi.org/10.1145/765891.765971

[18] Gerard Pons-Moll, David J. Fleet, and Bodo Rosenhahn. 2014. Posebits for Monocular Human Pose Estimation. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*. IEEE Computer Society, Washington, DC, USA, 2345–2352. https://doi.org/10.1109/CVPR.2014.300

[19] Javier Romero, Dimitrios Tzionas, and Michael J. Black. 2017. Embodied Hands: Modeling and Capturing Hands and Bodies Together. *ACM Trans. Graph.* 36, 6, Article 245 (Nov. 2017), 17 pages. https://doi.org/10.1145/3130800.3130883

[20] Jamie Shotton, Andrew Fitzgibbon, Andrew Blake, Alex Kipman, Mark Finocchio, Bob Moore, and Toby Sharp. 2011. Real-time human pose recognition in parts from single depth images. In *CVPR 2011*. 1297–1304. https://doi.org/10.1109/CVPR.2011.5995316

[21] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: Projected Visualizations for Hand Movement Guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 179–188. https://doi.org/10.1145/2207676.2207702

[22] Daniel Spelmezan, Mareike Jacobs, Anke Hilgers, and Jan Borchers. 2009. Tactile Motion Instructions for Physical Activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 2243–2252. https://doi.org/10.1145/1518701.1519044

[23] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang. 2015. Physio@Home: Exploring Visual Guidance and Feedback Techniques for Physiotherapy Exercises. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 4123–4132. https://doi.org/10.1145/2702123.2702401

[24] Gül Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, and Cordelia Schmid. 2018. BodyNet: Volumetric Inference of 3D Human Body Shapes. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII.* 20–38. https://doi.org/10.1007/978-3-030-01234-2_2

[25] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2013. MotionMA: Motion Modelling and Analysis by Demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1309–1318. https://doi.org/10.1145/2470654.2466171

[26] T von Marcard, B Rosenhahn, M J Black, and G Pons-Moll. 2017. Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs. *Comput. Graph. Forum* 36, 2 (2017), 349–360. https://doi.org/10.1111/cgf.13131

[27] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 841–850. https://doi.org/10.1145/2470654.2470774