Kinder-Gator: The UF Kinect Database of Child and Adult Motion

Aishat Aloba, Gianne Flores, Julia Woodward, Alex Shaw, Amanda Castonguay, Isabella Cuba, Yuzhu Dong, Eakta Jain, and Lisa Anthony

Department of CISE, University of Florida, USA

Abstract

Research has suggested that children's whole-body motions are different from those of adults. However, research on children's motions, and how these motions differ from those of adults, is limited. One possible reason for this limited research is that there are few motion capture (mocap) datasets for children, with most datasets focusing on adults instead. There are even fewer datasets that have both children's and adults' motions to allow for comparison between them. To address these problems, we present Kinder-Gator, a new dataset of ten children and ten adults performing whole-body motions in front of the Kinect v1.0. The data contains RGB and 3D joint positions for 58 motions, such as wave, walk in place, kick, and point, which have been manually labeled according to the category of the participant (child vs. adult), and the motion being performed. We believe this dataset will be useful in supporting research and applications in animation and whole-body motion recognition and interaction.

CCS Concepts

•***Computing methodologies*** → ***Motion capture;*** •***Human-centered computing*** → *User studies;*

1. Introduction

Animated movies and games increasingly include child characters. For games in particular, the motion of the virtual avatar is typically created using motion capture data [Gle08]. Motion capturing child actors is challenging, and that is probably why existing motion datasets are composed of either adult motion (performed by adults) [GS01, KCV*15] or adult actors performing "child-like" motion [VdlRBM14]. However, character animators [Ani13] have noted that children's motions (e.g., movements and poses) are different from those of adults. Thus, child motion datasets would be invaluable, but we found only one publicly available dataset [GFB12] containing both child and adult motions. Though this dataset is useful to understand differences between child and adult motion, participants performed motions as demonstrated by the researchers. Copying another's movements is different from naturally creating movement. Our dataset provides the motions of children performing actions as naturally as possible, with the goal of supporting realistic avatars and accurate recognition.

We introduce Kinder-Gator, a novel Kinect motion dataset of child and adult motion. The dataset contains 58 motions, such as wave your hand, walk in place, kick a ball, point at the camera, etc., from ten children (ages 5 to 9) and ten adults, tracked using a Microsoft Kinect 1.0. The Kinect tracks 3D positions (x:horizontal, y:vertical, z:depth) of twenty joints such as the hip, spine, elbows, and knees, along with their respective timestamps.

The motions in the Kinder-Gator dataset have been performed naturally and are manually labeled according to the category of the participant (child and adult) and the motion being performed (one of 58 motions). Analysis of a subset of the dataset made available to Jain et al. [JAA*16], including 6 actions, showed that naïve participants can perceive differences between the motion of a child and that of an adult. These results are an example of how our dataset can facilitate research into understanding the differences between child and adult motions. Dong et al. [DPR*17, DAAJ18] also applied style translation algorithms to the dataset to generate avatars with "child-like" motions. Because of the actions included in Kinder-Gator, it can be applied to gait recognition, hand pose recognition, and 3D stroke gesture recognition. Furthermore, Kinder-Gator can be used in character animation to generate avatars tailored to children. This dataset is publicly available at https://jainlab.cise.ufl.edu/pose-perception.html.

2. Related Work

There have been a number of mocap datasets that target whole-body motions. Gross and Shi [GS01] created the Motion of Body (MOBO) dataset which contains RGB video images of 25 adult actors walking on a treadmill with variations (slow, fast, inclined, and with a ball). However, this dataset includes only a small number of motions and uses image sequencing to track motions, requiring a separate processing step to extract the positions of the joints. This separate processing step makes motion analysis more complex than using mocap devices in which the positions of the joints have already been identified by the device. Kolykhalova et al. [KCV*15] created the MADS dataset, which contains mocap data for five adult actors performing martial-arts, dance, and sports actions captured using a custom motion tracking device. This motion capture device provides high motion tracking accuracy, but the device is difficult to obtain due to high cost and requires people with experience in how to operate it. With the availability of low-cost technology such as the Microsoft Kinect, however, mocap data containing accurate prediction of joint motions has become easier to collect.

Bloom et al. [BMA12] created the G3D dataset which includes video, depth, and skeleton data of ten actors performing twenty gaming actions (e.g., walk, run, jump, and climb), captured using the Kinect. Leightley et al. [LYJM15] also created the K3Da dataset, which contains Kinect data for 54 adults performing clinically-relevant motions (e.g., balancing).

A limitation of these mocap datasets is that they only consider adult motions (a comprehensive list of adult mocap datasets can be found in [vB]), which could be because recruiting children for research studies or mocap sessions is difficult [FDG13]. A workaround adopted by researchers has been to recruit adult actors to perform "child-like motions." Volkova et al. [VdlRBM14] created the MPI database, which contains mocap data of eight adult actors performing motions to express the emotions of a child (e.g., anger, disgust) while listening to fairy tale stories. However, Jain et al. [JAA*16] have shown that children's movements are different from those of adults, so it is likely that even well-trained actors are still not the same as actual child actors. Kinder-Gator addresses the limitations of existing datasets by including motions from both children and adults performing a large, diverse set of 58 motions captured using the Kinect.

It is important to note that Guerra-Filho and Biswas [GFB12] created the Human Motion Database (HMD), which contains mocap data of 50 child and adult actors performing 70 different actions. However, in HMD, the actions were demonstrated to participants to maintain consistency in performance among the range of actions. In Kinder-Gator, we aimed to ensure that the participants performed the motions as naturally as possible. This approach will allow analysis of natural realistic motions as children (and adults) actually perform them.

3. Dataset Collection

We collected a total of 58 motions in our dataset. The motions were chosen by reviewing studies involving whole-body motions (e.g., [HHTR05, HKL06, NWL10]). Based on the review, we selected motions that were used in prior work, that people would be familiar with, and that we hypothesized would show differences between children and adults. We also had a set of simple "warm up" motions like waving to help get participants into the study. The motions collected from this review were classified into four categories: (a) **warm-up motions**: these motions are easy to perform and are used in day-to-day activities (9 motions); (b) **exercise motions**: these motions induce exertion when performed and are commonly used in exercise and fitness activities (14 motions); (c) **mime motions**: these motions involve the conceptualization of imaginary objects (16 motions); and (d) **communication motions**: these motions are used to convey information (19 motions) (Table 1).

3.1. Study Setup

Motions in the Kinder-Gator dataset were collected using Kinect v1.0 hardware and its accompanying Kinect for Windows SDK v1.8 software. Two researchers were responsible for prompting for the next motion to be performed and controlling the Kinect software. In each study session, a participant stood within an area denoted by a square (47 x 47 inches), forward facing the Kinect and movements were only allowed within that specific area. The denotation of a specific area was to ensure that people did not move outside the tracking range of the Kinect. Participants performed all motions from a standing position. Before the start of each motion, participants stood with their arms outstretched in the form of a T-pose and then counted down from 3 to 1 while lowering their arms to their sides. The T-pose was required to get an accurate demarcation of the intended natural start of a motion.

The duration of each motion was dependent on the motion being performed. For wave your hand, walk in place, walk in place as fast as you can, run in place, run in place as fast as you can, fly like a bird, swim, climb an imaginary ladder, and do five jumping jacks, the duration was typically about five cycles (10 steps or repetitions). For motions involving making poses with the body, the duration was 3 seconds, since the experiment staff required participants to hold the pose for 3 seconds. For all other motions, the duration of the motion varied depending on the participant. Participants always returned their hands down to their sides to demarcate the end of the motion. In each session, a participant performed 58 motions. In order to ensure that participants were performing the motions as naturally as possible, participants were allowed to perform the motion free-form. That is, we did not require that the motions be performed in any predefined manner. When a participant did not understand how a motion was to be performed, one of the researchers showed an example. (This occurred for four different children on two to six actions, and two different adults on one to two actions; details are provided in the dataset). To ensure the motions performed were natural, participants performed the motion only after the researcher had stopped demonstrating an example, so as to reduce the likelihood of imitation.

3.2. Participants

We recruited ten children and ten adults via flyers, emails, and advertisements on social platforms. Recruitment and study protocol procedures were approved by our Institutional Review Board. Child participant ages ranged from 5 to 9 (mean = 6.70, SD = 1.42).

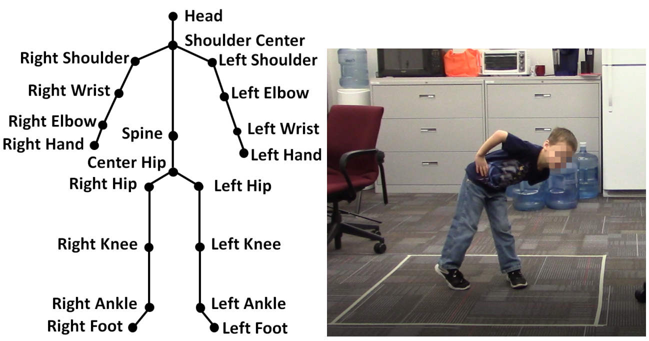

Figure 1: *(a) Joints tracked by the Kinect. (b) A child performing "put your hands on your hips and lean to the side" motion.*

Warm-up	Exercise	Mime	Communication
Raise your hand	Put your hands on your hip and lean to the side	Push an imaginary button in front of you	Point at the camera
Raise your other hand	Put your hands on your hips and lean to the other side	Swipe across an imaginary screen in front of you	Motion someone to stop
Wave your hand	Put your hands on your hips and twist back and forth	Swipe across an imaginary screen in front of you with your other hand	Motion someone to come here
Wave your other hand	Touch your toes	Fly like a bird	Draw a [circle, square, triangle] in the air
Bow	Do a forward lunge	Fly like an airplane	Draw the letter [A, C, K, M, X] in the air
Raise your arm to one side	Lift your leg to one side	Swim	Make the letter [Y, M, C, A, K, P, T, X] with your body
Raise your other arm to the other side	Lift your other leg to the other side	Kick a ball	–
Bend your knee	Walk in place	Kick a ball as hard as you can	
Bend your other knee	Walk in place as fast as you can	Kick a ball with the other leg	–
–	Run in place	Kick a ball as hard as you can with that leg	–
–	Run in place as fast as you can	Throw a ball	–
–	Jump	Throw a ball as far as you can	–
–	Jump as high as you can	Throw a ball with your other arm	–
–	Do five jumping jacks	Throw a ball as far as you can with that arm	–
–	–	Punch	–
–	–	Climb an imaginary ladder	–

Table 1: *A list of the 58 motions in the Kinder-Gator dataset.*

Five children were female. Two children were ambidextrous and none were left-handed. We focus on ages 5 to 9, since children in this age group are still growing in terms of their motor development [MKC89]. The adult participant ages ranged from 19 to 32 (mean = 23.40, SD = 4.33). Five adults were female and only one adult was left-handed (Table 2). All participants were familiar with motion interaction systems such as the Microsoft Kinect. Participants each received a $10 gift card to a local grocery store as compensation.

3.3. Data Collection

The 58 motions in our dataset were collected using Kinect v1.0. The Kinect tracks 3D positions of twenty joints along three dimensions (x:horizontal, y:vertical, z:depth) with the corresponding timestamps, at 30 frames per second. The joints tracked by the Kinect are shown in Figure 1a. The joint positions are recorded in meters and the timestamp is recorded in milliseconds. A total of 19 RGB videos and 1159 motion trials (58 motions x 20 participants) are included in our dataset; RGB video for all actions for one adult (ID: 934) and skeleton data for jump high for one adult (ID: 565) is missing due to a software error. The total time it took to perform the motions ranged from 247s to 363s (M = 301s, SD = 37.2) for children and from 302s to 424s (M = 344s, SD = 35.3) for adults.

CID	Sex	Age	Hand	Grade	AID	Sex	Age	Hand	Edu. Compl.
337	F	5	B	Pre-K	565	F	19	R	High school
595	M	5	B	Pre-K	577	F	19	R	Some college
106	M	6	R	K	604	F	20	R	Some college
290	M	6	R	K	976	M	20	R	Some college
342	F	6	R	K	734	M	22	R	Undergrad
474	F	6	R	1	876	F	23	R	Undergrad
169	M	8	R	2	921	M	26	L	Undergrad
103	F	8	R	3	888	F	25	R	Grad
723	M	8	R	3	970	M	28	R	Grad
644	F	9	R	4	934	M	32	R	Grad

Table 2: *Demographics for the children and adults in our dataset.*

To ensure that the data collected is in a format that facilitates easy retrieval and analysis, we performed some data post-processing. In the post-processing stage, the data was re-organized such that each row corresponds to the positions of all the joints at one frame. The first column has the timestamp such that the difference between the last row and the second row gives the duration of the motion in milliseconds; the first row is the header. Subsequent columns have the x, y, and z positions of each joint as recorded by the Kinect. The last two columns have the ID of the participant and the motion label; the values are always the same in each row for one motion. Then, each motion's data has been saved as a .csv file.

4. Discussion

Kinder-Gator is a dataset of child and adult motions tracked using the Kinect. The accuracy of the Kinect has been validated for adults, and the invariance of its tracking algorithm to body shapes and poses [SFC*11], coupled with its prevalence in gaming applications for children make the Kinect suitable for children. Since Kinder-Gator contains many different types of motions, we expect that researchers will extract different motion subsets for analysis. As an example, a subset of this dataset was used by Jain et al. [JAA*16] to conduct a perception study to identify whether naïve viewers can perceive the difference between child and adult motion when presented with point-light displays of the motions. The study was conducted using four children (IDs: 290, 337, 644, 723) and four adults (IDs: 734, 921, 934, 970) and six motions from our dataset: walk in place, run in place as fast as you can, jump as high as you can, wave your hand, fly like a bird, and do five jumping jacks. Survey responses showed that naïve viewers could perceive the differences at levels significantly above chance. Jumping jacks and jump high showed the highest variability between children and adults. This finding suggests that there are indeed quantifiable variations between child and adult motion. These types of results show that Kinder-Gator can be an effective dataset for understanding the differences between child and adult motion.

Dong et al. [DPR*17] used the same subset of motions and participants to generate "child-like" motions from adult motions using dynamic scaling laws. Observations of these characters suggested that dynamically scaled adult avatars are similar to adults in terms of coordination, but similar to children in terms of the pace of the motion. Dong et al. [DAAJ18] also used a style translation algorithm to achieve the same goal. The algorithm trains a model using child and adult motions such that it translates an adult motion to a child motion. Rendered characters of the motions showed that the translated adult motion is more similar to that of a child than an adult. Taken together, these findings show that Kinder-Gator can be used to test such cross-generational morphing methods.

5. Conclusion

Kinder-Gator is a dataset of 58 motions performed by ten children (ages 5 to 9) and ten adults (ages 19 to 32) recorded using the Kinect v1.0. Kinder-Gator can be used in research fields such as animation, whole-body interaction, and recognition. One of the main intended applications for Kinder-Gator is creating believable child characters for games and animated movies. Style translation methods [HPP05, XWCH15] can be leveraged along with this dataset to enable cross-generational morphing and create compelling avatars. In addition to animation, our dataset can also be used to create more robust recognizers that can recognize both child and adult motions. Due to the diverse set of motions in the dataset, subsets of the dataset can be employed in: (a) gait recognition to quantify gait features that are different between children and adults (e.g., walk in place), (b) human activity recognition to detect different motion types (e.g., motion someone to stop), and (c) stroke gesture recognition to create recognizers that can recognize both 2D and 3D stroke motions (e.g., draw the letter A in the air). The goal of our dataset is to encourage research investigating the differences and similarities between child and adult motions to enable future work in recognition and animation.

References

[Ani13] ANIMATIONADDICTS: The challenge of animating underaged adults. http://www.animation-addicts.com/2013/01/30/animating-kids/, 2013. Accessed: 2017-11-12. 1

[BMA12] BLOOM V., MAKRIS D., ARGYRIOU V.: G3D: A gaming action dataset and real time action recognition evaluation framework. *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (2012), 7–12. URL: http://10.1109/CVPRW.2012.6239175. 2

[DAAJ18] DONG Y., ALOBA A., ANTHONY L., JAIN E.: Style translation to create child-like motion. *Eurographics* (2018), [to appear]. 1, 4

[DPR*17] DONG Y., PARYANI S., RANA N., ALOBA A., ANTHONY L., JAIN E.: Adult2child: Dynamic scaling laws to create child-like motion. In *Proceedings of the International Conference on Motion in Games* (New York, NY, USA, 2017), MIG '17, ACM, pp. 13:1–13:10. URL: http://doi.acm.org/10.1145/3136457.3136460, doi:10.1145/3136457.3136460. 1, 3

[FDG13] FOSS E., DRUIN A., GUHA M. L.: Recruiting and retaining young participants: Strategies from five years of field research. In *Proceedings of the International Conference on Interaction Design and Children* (New York, NY, USA, 2013), IDC '13, ACM, pp. 313–316. URL: http://doi.acm.org/10.1145/2485760.2485798, doi:10.1145/2485760.2485798. 2

[GFB12] GUERRA-FILHO G., BISWAS A.: The human motion database: A cognitive and parametric sampling of human motion. *Image Vision Comput. 30*, 3 (Mar. 2012), 251–261. URL: http://dx.doi.org/10.1016/j.imavis.2011.12.002, doi:10.1016/j.imavis.2011.12.002. 1, 2

[Gle08] GLEICHER M.: More motion capture in games – can we make example-based approaches scale? In *International Workshop on Motion in Games* (2008), Springer, pp. 82–93. URL: http://dx.doi.org/10.1007/978-3-540-89220-5_9. 1

[GS01] GROSS R., SHI J.: *The CMU Motion of Body (MoBo) Database*. Tech. Rep. 1, 2001. doi:10.1109/MP.2008.4430762. 1

[HHTR05] HÖYSNIEMI J., HÄMÄLÄINEN P., TURKKI L., ROUVI T.: Children's intuitive gestures in vision-based action games. *Commun. ACM 48*, 1 (Jan. 2005), 44–50. URL: http://doi.acm.org/10.1145/1039539.1039568, doi:10.1145/1039539.1039568. 2

[HKL06] HWANG B.-W., KIM S., LEE S.-W.: A full-body gesture database for automatic gesture recognition. In *International Conference on Automatic Face and Gesture Recognition* (April 2006), FGR '06, pp. 243–248. URL: http://dx.doi.org/10.1109/FGR.2006.8, doi:10.1109/FGR.2006.8. 2

[HPP05] HSU E., PULLI K., POPOVIĆ J.: Style translation for human motion. *ACM Trans. Graph. 24*, 3 (July 2005), 1082–1089. URL: http://doi.acm.org/10.1145/1073204.1073315, doi:10.1145/1073204.1073315. 4

[JAA*16] JAIN E., ANTHONY L., ALOBA A., CASTONGUAY A., CUBA I., SHAW A., WOODWARD J.: Is the motion of a child perceivably different from the motion of an adult? *ACM Trans. Appl. Percept. 13*, 4 (July 2016), 22:1–22:17. URL: http://doi.acm.org/10.1145/2947616, doi:10.1145/2947616. 1, 2, 3

[KCV*15] KOLYKHALOVA K., CAMURRI A., VOLPE G., SANGUINETI M., PUPPO E., NIEWIADOMSKI R.: A Multimodal Dataset for the Analysis of Movement Qualities in Karate Martial Art. In *Proceedings of the International Conference on Intelligent Technologies for Interactive Entertainment* (2015). URL: http://dx.doi.org/10.4108/icst.intetain.2015.260039. 1

[LYJM15] LEIGHTLEY D., YAP M. H., J. COULSON Y. B., MCPHEE J. S.: Benchmarking Human Motion Analysis Using Kinect One: an open source dataset. *IEEE Conference of Asia-Pacific Signal and Information Processing Association*, December (2015), 1–7. URL: http://dx.doi.org/10.1109/APSIPA.2015.7415438. 2

[MKC89] MAGALHAES L. C., KOOMAR J. A., CERMAK S. A.: Bilateral motor coordination in 5-to 9-year-old children: A pilot study. *American Journal of Occupational Therapy 43*, 7 (1989), 437–443. 3

[NWL10] NORTON J., WINGRAVE C. A., LAVIOLA JR. J. J.: Exploring strategies and guidelines for developing full body video game interfaces. In *Proceedings of the International Conference on the Foundations of Digital Games* (New York, NY, USA, 2010), FDG '10, ACM, pp. 155–162. URL: http://doi.acm.org/10.1145/1822348.1822369, doi:10.1145/1822348.1822369. 2

[SFC*11] SHOTTON J., FITZGIBBON A., COOK M., SHARP T., FINOCCHIO M., MOORE R., KIPMAN A., BLAKE A.: Real-time human pose recognition in parts from single depth images. In *CVPR 2011* (June 2011), pp. 1297–1304. URL: http://dx.doi.org/10.1109/CVPR.2011.5995316, doi:10.1109/CVPR.2011.5995316. 3

[vB] VAN BOXTEL J. J.: Free mocap databases. http://jeroenvanboxtel.com/MocapDatabases.html. Accessed:2017-12-28. 2

[VdlRBM14] VOLKOVA E., DE LA ROSA S., BULTHOFF H. H., MOHLER B.: The MPI emotional body expressions database for narrative scenarios. *PLoS ONE 9*, 12 (2014), e113647. 1, 2

[XWCH15] XIA S., WANG C., CHAI J., HODGINS J.: Realtime style transfer for unlabeled heterogeneous human motion. *ACM Trans. Graph. 34*, 4 (2015), 119:1–119:10. URL: http://doi.acm.org/10.1145/2766999, doi:10.1145/2766999. 4