Tailoring Motion Recognition Systems to Children's Motions

Aishat Aloba
Department of CISE, University of Florida
Gainesville, FL, USA
aoaloba@ufl.edu

ABSTRACT

Motion-based applications are becoming increasingly popular among children and require accurate motion recognition to ensure meaningful interactive experiences. However, motion recognizers are usually trained on adults' motions. Children and adults differ in terms of their body proportions and development of their neuromuscular systems, so children and adults will likely perform motions differently. Hence, motion recognizers tailored to adults will likely perform poorly for children. My PhD thesis will focus on identifying features that characterize children's and adults' motions. This set of features will provide a model that can be used to understand children's natural motion qualities and will serve as the first step in tailoring recognizers to children's motions. This paper describes my past and ongoing work toward this end and outlines the next steps in my PhD work.

CCS CONCEPTS

• **Human-centered computing → Gestural input**; • **Social and professional topics → Children**.

KEYWORDS

whole-body motions; children; motion articulation; recognition

ACM Reference Format:
Aishat Aloba. 2019. Tailoring Motion Recognition Systems to Children's Motions. In *2019 International Conference on Multimodal Interaction (ICMI '19), October 14–18, 2019, Suzhou, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3340555.3356092

ICMI '19, October 14–18, 2019, Suzhou, China

ACM ISBN 978-1-4503-6860-5/19/10.
https://doi.org/10.1145/3340555.3356092

1 INTRODUCTION

The performance of motion recognition algorithms plays an important role in users' interactive experiences in motion-based applications. For example, in games that combine physical activity with play (exertion games), the precision of motion recognition systems has been positively associated with higher levels of immersion during game play [18]. These applications are becoming increasingly popular among children, so accurate recognition of child motion has become more important to ensure they have meaningful interactive experiences. However, motion-based recognizers usually have been trained on adults' motions [9, 19, 28]. Children and adults differ in terms of their body proportions [12] and stages of development of their neuromuscular systems [24], so children and adults will likely perform motions differently. In fact, my past work has found that naïve viewers can perceive differences between child and adult motions at levels significantly above chance, even when the motion is abstracted from all appearance cues [14]. These findings establish that child motion differs from adult motion, so motion recognition systems tailored to adults' motions will likely perform poorly for children's motions. Furthermore, these findings suggest that there are motion qualities present in children's motions that distinguish them from adults' motions. However, these motion qualities are not currently known, which makes tailoring motion recognizers to children difficult.

My PhD thesis will focus on investigating the features that characterize children's and adults' motion qualities, in order to establish a set of features that quantifies the differences between child and adult motions. Specifically, my thesis aims to answer the following research questions:

(1) Can we identify features that characterize the similarities and differences between children's and adults' motions?
(2) What inferences can we make from these features to help tailor motion recognition systems to child motion?
(3) What information about children's motion qualities can improve immersion within motion-based applications, specifically, exertion games?

The set of features identified in my work will provide a model that can be used to characterize children's natural motion qualities and will serve as the first step in tailoring recognizers to children's motions. Ultimately, this set of features can help to improve children's experiences when interacting with motion-based applications (e.g., exertion games).

2 RELATED WORK

I focus my review of prior work on human motion recognition and stroke gesture recognition. Prior work in stroke gesture recognition has investigated tailoring stroke gesture recognizers to children's gestures [23]. Therefore, I utilize prior work in stroke gesture recognition to inform my approach in tailoring recognizers to children's motions.

Human Motion Recognition

Human motion ranges from simple limb movements (e.g., wave) to whole-body movements involving multiple limbs of the human body (e.g., walking). Recognition of human motion has been studied extensively through traditional computer vision approaches that rely on using pixels to recognize motions from images and videos (i.e., vision-based human action recognition) [9, 20]. Motion tracking devices, such as the Microsoft Kinect [17] provide accurate information about poses (i.e., positioning of the body at a specific time instance) and joint movements as users perform motions. With the provision of this information, researchers have shifted focus to approaches that rely on geometry of poses for recognition [10, 19]. Lun and Zhao termed one approach to recognize motions as template-based, which they defined as comparing an unknown whole-body gesture to a set of pre-recorded whole-body gesture templates using pattern recognition [16]. By their definition, this approach consists of recognizers that utilize direct matching approaches (e.g., dynamic time warping (DTW) [5]) and model-based approaches that utilize machine learning (e.g., support vector machines (SVM) [7]). Template-based motion recognizers, which are dependent on features selected for pattern recognition, are prone to recognition errors. For example, Riofrío et al. [19] had a recognition error rate of 15.93% when using DTW and the positions of 10 upper limb joints from the Kinect as features to recognize upper limb motions (e.g., right hand sweep). Therefore, features that reflect the distinctive characteristics of motion are important for good recognition accuracy [16]. I plan to identify features that can be used to characterize how children and adults make motions.

Tailoring Stroke Gesture Recognizers to Children's Gestures

Stroke gesture recognition researchers have found differences in how children and adults produce stroke gestures [23, 27] and have thus investigated tailoring stroke gesture recognizers to children's gestures [23]. Prior work in stroke gesture recognition has found that stroke gesture recognizers have lower recognition accuracies for children's gestures compared to adults' gestures [3]. Anthony et al. compared the recognition accuracy of stroke gestures produced by children ages 7 to 16 and adults using the $N-protractor stroke gesture recognizer [4] and found that, on average, children's gestures were recognized more poorly (81%) compared to adults' gestures (90%) [3]. Based on this finding, researchers have suggested the need to tailor stroke gesture recognizers to children's motions. For example, Shaw et al. [23] investigated the differences between children's and adults' stroke gestures using features such as path length and production time and found differences in several features. For example, children exhibit longer path lengths and higher shape errors compared to adults. The researchers also found that children are more inconsistent in how they produce gestures as characterized by higher variations in some of the features, and this inconsistency causes higher recognition errors in children's gestures compared to adults' gestures. Like stroke gesture recognizers, motion recognizers will likely perform poorly for children's motions due to differences in how children and adults perform motions [14]. I plan to analyze the motion qualities of children and adults to better understand how to tailor motion recognizers to children's motions.

3 COMPLETED WORK

In this section, I discuss some of the preliminary work I have done to understand children's motion qualities.

Creating a Dataset of Children's and Adults' Natural Motions

To characterize how users perform motions, it is important that my work focuses on natural motions to ensure that motion behaviors that are unique to children can be captured, as opposed to scripted motions, which may not capture variations in how children and adults perform motions. However, due to a lack of publicly available datasets of children's natural motions, my first step was to acquire a dataset of natural motions to analyze children's motions and to allow for comparison against adults' motions. I helped collect the Kinder-Gator dataset [1]; a dataset of 10 children and 10 adults performing 58 natural motions forward-facing the Microsoft Kinect [17]. The Kinect v1 tracks the movement of 20 joints in the body along 3-dimensions and at 30 frames per second (fps). A critical aspect for the creation of the dataset was to ensure that it encompassed a diverse range of human motions. Hence, the 58 motions I helped collect include: nine warm-up motions that are easy to perform and are used in day to day activities (e.g., raise your hand), fourteen exercise motions that induce exertion (e.g., run in place), sixteen mime motions that are used to conceptualize

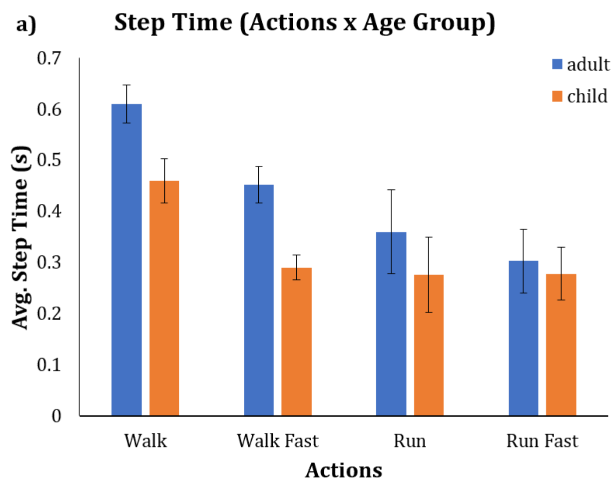

Figure 1: Figure 1: Effect of action type and age group on step time. Children generally move faster than adults.

imaginary objects (e.g., climbing an imaginary ladder), and nineteen communication motions that are used for conveying information (e.g., motion someone to stop). All future research directions for my work will utilize this dataset.

Characterizing the Differences between Child and Adult Motion using Gait Features

In my past work, we conducted a perception study to investigate if naïve viewers can perceive the differences between child and adult motion when the motion is abstracted from appearance cues [14]. Our results showed that naïve viewers can perceive the difference between child and adult motion at levels significantly above chance. Additionally, we found that for dynamic motions, such as walking and running, naïve viewers can perceive the differences between children and adults with about 70% accuracy. These findings suggest that there are perceivable cues that can be used to differentiate child motion from adult motion. Hence, we hypothesized that there are features that quantify these differences between children's and adults' motions. To test our hypothesis, I initially concentrated on analyzing walking and running motions, since naïve viewers in our perception study had high accuracy for these motions [14]. Furthermore, the gait literature has identified temporal (time) features and spatial (distance) features that characterize walking and running motions [13]. In our study, we selected four temporal features (step width, step height, relative step height, and walk ratio) and five spatial features (step time, cycle time, cycle frequency, step speed, and cadence). We used these features to analyze walking and running in the Kinder-Gator dataset namely: walk in place, run in place, walk in place as fast as you can, and run in place as fast as you can. For each participant-motion pair, we computed the nine gait features and used a repeated-measures Analysis of Variance (ANOVA) to analyze the main effect of age group (child vs. adult) and motion type. We found no significant differences between children and adults for spatial features [2]. On the other hand, we found a significant difference between child and adult motions for all temporal features except step speed: children complete walking and running motions in less time (Figure 1) and with higher energy compared to adults [2].

4 RESEARCH PLAN

Although the results above verify my hypothesis that there are features that quantify the differences between children's and adults' motions, it is difficult to generalize these features to a broader set of motions because they are optimized for analyzing gait and rely on the periodicity of the motion. Therefore, gait features may not be suitable for motions that are not periodic, such as exercise motions (e.g., "a jump" or "kick"). Hence, I plan to identify features that are generalizable to a broader set of motions. Vatavu et al. [25] proposed spatial features that quantify how much users are moving in space, kinematic features that quantify the time associated with performing the motion, and appearance features that quantify the expressiveness of the motion. These features focus on describing the motions on a global level by focusing on the poses (i.e., positioning of the body at a specific time instance) that make up the motion. In addition to these features, I plan to identify a set of features that will describe motions on a joint level (geometric features). Geometric features will quantify properties of the motion paths necessary for performing the motions (e.g., length, shape, curvature). I posit that exploring these four types of features (spatial, kinematic, appearance, and geometric) will provide knowledge regarding motion characteristics (e.g., the degree of movement of each joint), which can then be used to quantify the differences between children's and adults' motion qualities. In this section, I will describe my continued work and future research directions to answer my research questions.

Prior research in stroke gesture recognition has found differences between children's and adults' stroke gestures. I posit that stroke gestures and motions share a representational similarity: stroke gestures consist of lines, curves, and corners [6] with a point cloud representation in 2D space while motions consist of poses with a point cloud representation in 3D space. Therefore, both stroke gestures and motions include a set of data points moving in space over time, so methods from stroke gesture recognition to identify geometric features can inform our research into identifying geometric features that characterize whole-body motions.

Simplifying Motion Representations

However, features from stroke gestures are not directly applicable to motions, since they differ in terms of number of paths. Stroke gestures are defined by a single gesture path, which is the path the finger takes on the surface over time. In contrast, motions involve many joints with each joint having a motion path; however, not all joints are essential to the performance of the motion. For example, to raise one's hands, the hand and wrist joint are essential to the motion, but the foot joint is not necessary for the performance of the motion. Therefore, any motion path in the foot joint is either due to tracking errors from the sensing device or unintentional movements from the user, so utilizing the full set of joints could lead to incorrect inferences regarding how users perform motions.

For this reason, I identified the joints that users are actively moving intentionally. I posited that these joints will have higher movement variability compared to the movements in joints that are moving unintentionally. To identify the joints with higher variability, I computed the standard deviation of the joint positions over time and partitioned the result into two clusters using a K-means clustering algorithm [11]. I selected the joints in the cluster with the higher mean as the joints that the user is moving intentionally and simplified the motion representation to include only these joints (filterJoint representation). I applied the filterJoint and original representations on a subset of 11 motions from the Kinder-Gator dataset that [1] are non-overlapping in terms of motion type and limb direction. Inspired by elements of stroke gesture recognizers, such as $1 [26]), I developed a template-based recognizer for motions and used this recognizer to compare the accuracies of both representations. The recognition accuracy is determined by matching a user's motions to other users' motions, so it is an indication of between-user consistency in motion. Therefore, a higher recognition accuracy implies that the representation includes features that are more discriminative among motion types. I found that our filterJoint representation achieved a significantly higher recognition accuracy compared to the original representation ($t(8) = 8.03$; $p < 0.001$). This finding suggests that the filterJoint representation is more suitable for understanding how users perform motions. Furthermore, an investigation of the degree of agreement in the actively moving joints, calculated as the number of unique joint combinations used to perform a motion, showed that adults exhibit significantly higher levels of agreement compared to children ($t(10) = 5.88$; $p < 0.001$). This finding suggests that children are more inconsistent in how they perform motions compared to adults.

Identifying Features

My future research will involve identifying a set of features that quantifies the geometric properties (e.g., length, shape, curvature) of the paths in the filterJoint representation. This set of features will enable an understanding of the variations in how adults and children move joints during the motion. For example, a feature can measure the degree of difference between the shape of users' motion paths and the shapes of a representative set of motion paths, created by averaging across the motion paths of participants. If children have a higher degree of difference compared to adults, then this would mean that children are more inconsistent in how they move their joints. Once I have identified all geometric features, I will use an Analysis of Variance (ANOVA) to check the effect of category (child vs. adult) using motions from the Kinder-Gator dataset [1]. The features with a significant difference are the features that effectively quantify the differences between child and adult motion.

Validating Features

For validation, I will use all the features (geometric, spatial, kinematic, appearance) to train a binary classifier (classes: child, adult) to explore the relationship between the most discriminative features from the classifier and the features with significant differences between children and adults based on the results from our statistical test. My expectation is that the most discriminative features should correspond to the features that show significant differences and the least discriminative features should correspond to features with no significant differences. After this validation, we will be able to fully understand the similarities between children's and adults' motions (RQ1). For the validation, I plan to consider different machine learning classifiers, such as Support Vector Machines (SVM [7]), since this classifier has been used in the motion recognition literature to classify motions [22], k-nearest neighbor [8], and decision trees [21].

Tailoring Motion Recognizers to Children's Motion Qualities

The next step will be to make inferences about how children perform motions. Specifically, I want to explore the relationship between age and type of motion being performed and consistency in how children perform motions. To accomplish this, I will analyze child and adult motions based on the features that I find to be different between children and adults. This analysis will involve conducting a repeated measures ANOVA on each feature with a within-subjects factor of motion type and between-subjects factor of age; children in the Kinder-Gator dataset [1] have ages in the range of 5 to 9. The results from this test will provide information regarding: (a) if any children within this range (5 to 9) perform motions consistently like adults, (b) the motions which children tend to perform consistently regardless of their ages, and (c) the age groupings in which children share the most similarity in how they perform motions. I will use this information to extend motion recognition systems to children's motion qualities (RQ2). For example, if the results show that younger children perform motions similarly to each other but differently from older children, then this may imply that motion recognizers may need to consider age groupings during the recognition process. For example, template-based recognizers can include weights during the matching process so that children's motions are more likely to be matched to motions from children with similar age groupings.

Improving Immersion in Exergames

Prior work has associated precision of motion recognition systems to increased immersion within exergames. To investigate the effect of motion recognition systems on children's immersion when interacting with exergames, I will extend an existing exertion game to either include a standard motion recognizer trained on adults' motions or the motion recognizer tailored to children's motions (RQ3). Next, I will conduct a within-subjects experiment wherein children's immersion will be measured after interacting with both exergames. Prior work has asserted that immersion in games can be measured subjectively using questionnaires and objectively using task completion times [15]. I plan to use both approaches to measure children's immersion in exergames.

Conclusion

My work thus far has established that children's motions differ from adults' motions. My continued work will identify the features that characterize the differences between child and adult motions and extend motion recognition algorithms to children's motion qualities. Features identified from my research have the potential to improve recognition of children's motions and improve children's immersion within exertion games.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Lisa Anthony, and my committee members, Dr. Jaime Ruiz and Dr. Eakta Jain for their support on this work. I would also like to thank members of the INIT lab and RUIZ lab at the University of Florida for their help.

REFERENCES

[1] Aishat Aloba, Gianne Flores, Julia Woodward, Alex Shaw, Amanda Castonguay, Isabella Cuba, Yuzhu Dong, Eakta Jain, and Lisa Anthony. 2018. Kinder-Gator: The UF kinect database of child and adult motion.. In *EG 2018 - Short Papers.*, Olga Diamanti and Amir Vaxman (Eds.). 13–16. https://doi.org/10.2312/egs.20181033

[2] Aishat Aloba, Annie Luc, Yuzhu Dong, Rong Zhang, Eakta Jain, and Lisa Anthony. 2019. Quantifying differences between child and adult motion based on gait features. In *International Conference on Human-Computer Interaction (HCII '19)*. 385–402. https://doi.org/10.1007/978-3-030-23563-5_31

[3] Lisa Anthony, Quincy Brown, Jaye Nias, Berthel Tate, and Shreya Mohan. 2012. Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces (ITS '12)*. ACM Press, 225. https://doi.org/10.1145/2396636.2396671

[4] Lisa Anthony and Jacob O Wobbrock. 2012. $ N-protractor: a fast and accurate multistroke recognizer. In *Proceedings of Graphics Interface (GI '12)*. Canadian Information Processing Society, 117–120.

[5] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. In *KDD Workshop on Knowledge Discovery in Databases*. 359–370.

[6] Xiang Cao and Shumin Zhai. 2007. Modeling human performance of pen stroke gestures. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07)*. ACM, 1495–1504.

[7] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.

[8] Thomas Cover and Peter Hart. 1967. Nearest neighbor pattern classification. *IEEE transactions on information theory* 13, 1 (1967), 21–27.

[9] Janto F. Dreijer and Ben M. Herbst. 2008. Action classification using the average of pose changes. In *Symposium of the Pattern Recognition Association of South Africa*. 85–85.

[10] Tomasz Hachaj and Marek R. Ogiela. 2014. Rule-based approach to recognizing human body poses and gestures in real time. *Multimedia Systems* (2014), 81–99. https://doi.org/10.1007/s00530-013-0332-2

[11] John A Hartigan and Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 1 (1979), 100–108.

[12] Donald F. Huelke. 1998. An overview of anatomical considerations of infants and children in the adult world of automobile safety design. *Association for the Advancement of Automotive Medicine* 42 (1998), 93–113. https://doi.org/10.1145/982452.982461 arXiv:arXiv:1011.1669v3

[13] Wan Noorshahida Mohd Isa. 2005. Analysis on spatial and temporal features of gait kinematics. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*. IEEE, 130–133.

[14] Eakta Jain, Lisa Anthony, Aishat Aloba, Amanda Castonguay, Isabella Cuba, Alex Shaw, and Julia Woodward. 2016. Is the motion of a child perceivably different from the motion of an adult? *ACM Transactions on Applied Perception* 13, 4 (jul 2016), 1–17. https://doi.org/10.1145/2947616

[15] Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International journal of human-computer studies* 66, 9 (2008), 641–661.

[16] Roanna Lun and Wenbing Zhao. 2015. A survey of applications and human motion recognition with microsoft kinect. *International Journal of Pattern Recognition and Artificial Intelligence* (2015). https://doi.org/10.1142/s0218001415550083

[17] Microsoft. [n.d.]. Kinect for Windows. https://developer.microsoft.com/en-us/windows/kinect

[18] Jasmir Nijhar, Nadia Bianchi-Berthouze, and Gemma Boguslawski. 2012. Does movement recognition precision affect the player experience in exertion games?. In *International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN '12)*. 73–82. https://doi.org/10.1007/978-3-642-30214-5_9

[19] Santiago Riofrío, David Pozo, Jorge Rosero, and Juan Vásques. 2017. Gesture recognition using dynamic time warping and kinect: A practical approach. In *International Conference on Information Systems and Computer Science*. 302–308.

[20] Mikel D. Rodriguez, Javed Ahmed, and Mubarak Shah. 2008. Action MACH: A spatio-temporal maximum average correlation height filter for action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*. 6–6. https://doi.org/10.1109/CVPR.2008.4587727

[21] S Rasoul Safavian and David Landgrebe. 1991. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics* 21, 3 (1991), 660–674.

[22] Christian Schuldt, Ivan Laptev, and Barbara Caputo. 2004. Recognizing human actions: a local SVM approach. In *Proceedings of the International Conference on Pattern Recognition (ICPR '04)*, Vol. 3. IEEE, 32–36.

[23] Alex Shaw and Lisa Anthony. 2016. Analyzing the articulation features of children's touchscreen gestures. In *ACM International Conference on Multimodal Interaction (ICMI '16)*. 333–340. https://doi.org/10.1145/2993148.2993179

[24] David H. Sutherland. 1997. The development of mature gait. *Gait Posture* 6, 2 (1997), 163–170. https://doi.org/10.1016/S0966-6362(97)00029-5

[25] Radu-Daniel Vatavu. 2017. Beyond features for recognition: human-readable measures to understand users' whole-body gesture performance. *International Journal of Human–Computer Interaction* 33, 9 (2017), 713–730.

[26] Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. 2007. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proceedings of the ACM symposium on User interface software and technology (UIST '07)*. https://doi.org/10.1145/1294211.1294238

[27] Julia Woodward, Alex Shaw, Annie Luc, Brittany Craig, Juthika Das, Phillip Hall Jr., Akshay Holla, Germaine Irwin, Danielle Sikich, Quincy Brown, and Lisa Anthony. 2016. Characterizing how interface complexity affects children's touchscreen interactions. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM Press, 1921–1933.

[28] Hee Deok Yang, A. Yeon Park, and Seong Whan Lee. 2006. Human-robot interaction by whole body gesture spotting and recognition. In *International Conference on Pattern Recognition*, Vol. 4. 774–777. https://doi.org/10.1109/ICPR.2006.642