Style Translation to Create Child-like Motion

Yuzhu Dong, Aishat Aloba, Lisa Anthony and Eakta Jain

Department of Computer and Information Science and Engineering
University of Florida

Abstract

Animated child characters are increasingly important for educational and entertainment content geared towards younger users. While motion capture technology creates realistic and believable motion for adult characters, the same type of data is hard to collect for young children. We aim to algorithmically transform adult motion capture data to look child-like. We implemented a warping based style translation algorithm, and show the results when this algorithm is applied to adult to child transformation.

1. Introduction and Background

Video games and electronic entertainment have become a routine part of life for a large majority of children today. Games can serve as an avenue for relaxation, education, and fitness [Olson 2010]. As a result, animated child characters are becoming increasingly important for the associated industries. Motion capture data (mocap) is widely used in animating avatars. While adult mocap databases are abundant and cover different types of motions, the lack of child-like motion capture data poses a challenge. Motion capturing children is difficult and expensive compared to motion capturing adults.

Previous work has shown that people are able to identify if a motion was performed by a child or an adult [Jain et al. 2016]. We aim to create believable motion for child avatars by stylizing mocap data collected on adult actors in the same way that previous works have stylized normal walks to crouch walks, or a sideways shuffle [Hsu et al. 2005], for example. This is different from retargeting motion to a differently proportioned skeleton [Gleicher 1998]. [Dong et al. 2017] showed that linearly scaling adult motion in space and time dimensions makes it more likely for viewers to think that such motion came from child actors than adult actors. However, the accuracy levels indicated that there are still stylistic differences between adult and child motion that are not captured by linear scaling. Our work explores a non-linear warping based style translation algorithm. In particular, we implement [Hsu et al. 2005]'s method which transforms input motion into a new style through a linear invariant model. Our main contribution is to apply this method towards the new problem of transforming adult motion to appear child-like.

2. Method

We use the database of [Aloba et al. 2018] for matched pairs of child and adult motion. We first translate the joint position in world coordinates into local coordinates as in [Dong et al. 2017]. We denote the frame number as i, the joint number as j, the input adult motion as u, the output child motion as v. Therefore, the joint position along the x-axis in local coordinates is written as ${}^{x}u_i^j$ for adult motion, and the output child motion is written as ${}^{x}v_i^j$. The matrix u is used to describe the entire motion of all the joints in the skeleton, where each frame is a row of the matrix. Each joint is considered separately. The vector $\mathbf{u}$ is concatenated by the position of the jth joint across all the frames, $\mathbf{u} = [{}^{x}u_1^j, {}^{x}u_2^j, {}^{x}u_i^j \ldots, {}^{x}u_T^j]^T$. The data is normalized to zero mean and unit standard deviation.

Computing correspondences Because frame correspondences are the key to successful style translation, iterative motion warping is performed to find correspondences between the two stylistically different motions, child and adult motion. Each iteration has two steps: dynamic time warping followed by space warping. The two steps were iterated until the cost function comes to convergence.

The iterative motion warping procedure minimizes the following cost function:

$$E(\mathbf{a}, \mathbf{b}, \mathbf{W}) \equiv \|\mathbf{W}(\mathbf{U}\mathbf{a} + \mathbf{b}) - \mathbf{v}\|^2 + \|\mathbf{F}\mathbf{a}\|^2 + \|\mathbf{G}\mathbf{b}\|^2 \quad (1)$$

where $\mathbf{U}$ is $diag(\mathbf{u})$, $\mathbf{W}$ is the time warping matrix, $\mathbf{a}$ is the scale vector, and $\mathbf{b}$ is the offset vector. The matrices $\mathbf{F}$ and $\mathbf{G}$ provide weighted finite differences of $\mathbf{a}$ and $\mathbf{b}$ respectively.

Dynamic time warping. Dynamic time warping (DTW) minimizes the Euclidean distance between the joint positions of the input and output motions. For each frame i of the input motion and frame i' of the output motion, the Euclidean distance between the input motion and the output motion is $d(i,i') = \sqrt{(u_i - v_{i'})^2}$, and the overall distance from $d(1,1)$ to $d(T,T')$ is $d = \sum d(i,i')$.

While [Hsu et al. 2005] only allow repetitions, we modified the time warping step to include deletions as well. This modification was necessary for the type of alignment we needed to see: for example, we noticed that a child's jump had a longer anticipatory phase and a shorter actual jump compared to an adult's jump.

Space warping. In the space warping stage, we equate the first

DOI: 10.2312/egp.20181023

derivative of cost function E with respect to $\mathbf{a}$ and $\mathbf{b}$ to zero. The parameter $\mathbf{W}$ from the time warping stage is fixed. This leads to the following normal equation:

$$\begin{bmatrix} \mathbf{U^T W^T W U} + \mathbf{F^T F} & \mathbf{U^T W^T W U} \\ \mathbf{W^T W U} & \mathbf{W^T W} + \mathbf{G^T G} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{U^T W^T v} \\ \mathbf{W^T v} \end{bmatrix} \quad (2)$$

We solve this equation via LU (lower-upper) factorization using MATLAB's built in function *linsolve*.

Style translation. Given a sequence of input frames, we estimate a sequence of output frames as follows:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}'_t + \mathbf{e}_t \\ \mathbf{v}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}'_t + \mathbf{f}_t \end{cases} \quad (3)$$

where $\mathbf{u}' = \mathbf{W}\mathbf{u}$, $\mathbf{x}_t$ is the state of the system at frame t, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ are constant system matrices. $\mathbf{e}_t^j$, $\mathbf{f}_t^j$ are white noise terms. We treat each joint separately and trained different models for different joints. The training data are fed into MATLAB's system identification method, *ssest* [Ljung 1998], to estimate parameters in matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$. The model order is set to 40. The state vector $\mathbf{x}_t$ is initialized as zero vector. Once the model parameters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ have been learned, any input adult motion capture sequence can be style translated using this model.

The result of our algorithm was translated into Euler angles using the method described in [Dong et al. 2017]. The result of our algorithm was then converted into a .bvh file where it can be directly imported into Maya.

3. Results

We show the results of the style translation algorithm on the action "Jump High" from the dataset of [Dong et al. 2017]. Child actor 723 and adult actor 734 are used as training data, while adult actor 934 is the new input data. Because these motion sequences are very short (40-50 seconds), we repeated each sequence 28 times after iterative motion warping, before estimating the system matrices.

The result of the motion "Jump High" is shown in Figure 1. The skeleton in green is the translated motion. The figure captures the moment when both adult and child are bending to the lowest point in the landing phase of the jump. Compared to adult 934 which has the green skeleton, the result of the algorithm in the red skeleton and the child 723 in the blue skeleton both bend lower. Also, adult 934 does not bend his knees to the same extent. Finally, the motion of child 723 and the result of our algorithm both spend longer in the landing phase because it takes longer to bend to a lower position.

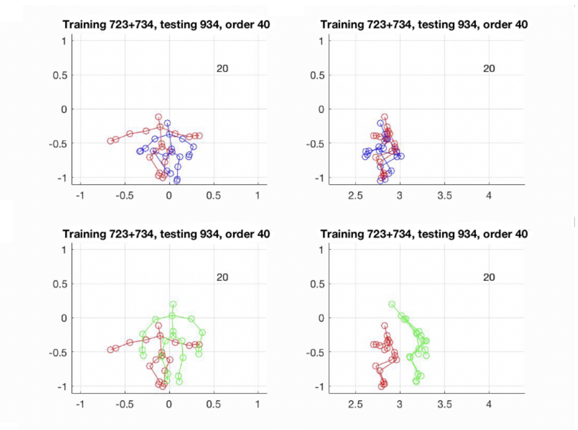

Figure 1: *Child motion (red), adult motion (green), our result (blue).*

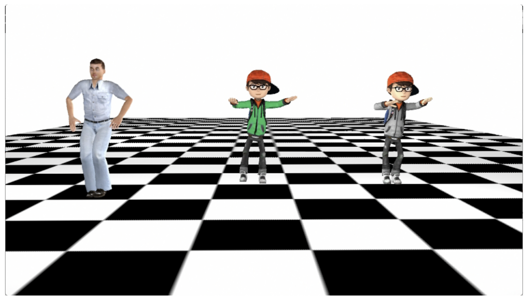

Figure 2: *Training data for adult motion is from actor 934, and for child motion is from actor 723. Our result motion is in green.*

In the rendered result in Figure 2, the arm of child 723 (grey jacket) and our result (green jacket) both swing to a larger extent compared to adult 934. Thus, we see that the gesture and the timing of our translated motion is very similar to child motion.

4. Conclusion

In this paper, we generated motion for animated child characters by algorithmically transforming adult motion capture data to child-like motion capture data. We showed that a warping based style translation algorithm could create child-like motion that mimics the original child motion. This method can be used for a variety of applications, such as games, movies, educational content, training and simulation environments. We aim to apply this algorithm to more actions and actors, and to evaluate how close the result is to actual child motion by conducting a perception study.

References

[Aloba et al. 2018] Aloba, A., Flores, G., Woodward, J., Shaw, A., Castonguay, A., Cuba, I., Dong, Y., Jain, E., and Anthony, L. 2018. Kindergator: The UF kinect database of child and adult motion. In *Eurographics*, [To Appear]. 1

[Dong et al. 2017] Dong, Y., Aloba, A., Paryani, S., Rana, N., Anthony, L., and Jain, E. 2017. Adult2child: dynamic scaling laws to create child-like motion. In *Proceedings of the Tenth International Conference on Motion in Games*, ACM, No. 13. 1, 2

[Gleicher 1998] Gleicher, M. 1998. Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ACM, 33–42. 1

[Hsu et al. 2005] Hsu, E., Pulli, K., and Popović, J. 2005. Style translation for human motion. In *ACM Transactions on Graphics (TOG)*, vol. 24, ACM, 1082–1089. 1

[Jain et al. 2016] Jain, E., Anthony, L., Aloba, A., Castonguay, A., Cuba, I., Shaw, A., and Woodward, J. 2016. Is the motion of a child perceivably different from the motion of an adult? *ACM Transactions on Applied Perception (TAP) 13*, 4, Article No. 22. 1

[Ljung 1998] Ljung, L. 1998. System identification. In *Signal analysis and prediction*. Springer, 163–173. 2

[Olson 2010] Olson, C. K. 2010. Children's motivations for video game play in the context of normal development. *Review of General Psychology 14*, 2, 180. 1