Generalized Motion Synthesis for Multiple Avatars

Siddhant Arora
Department of Computer Science
University of Texas at Austin
Austin, Texas
siddhant.arora@utexas.edu

Rohit Neppalli
Department of Computer Science
University of Texas at Austin
Austin, Texas
rohit.neppalli@utexas.edu

Amal Babu
Department of Computer Science
University of Texas at Austin
Austin, Texas
amalbabu@utexas.edu

Kunal Mody
Department of Computer Science
University of Texas at Austin
Austin, Texas
kunalmody@utexas.edu

***Abstract*—Human motion is difficult to create in a realistic manner. One of the problems is that humans don't walk perfectly uniformly; instead each human has their own unique motion. In this proposal, we build upon the idea of motion synthesis, or creating realistic human motion. There are existing models that can create realistic walking motions from large data sets of multiple people, of from physically valid models. What these approaches lack is the representation of individuality in humans. In general, humans walk differently. Most current models aggregate the walking motions of multiple people into one general purpose walking motion. We propose using a style transfer method that would reflect the differences in individual's walking mannerisms, in such a way that if the model is trained on one person there will be visible differences when trained on another person who walks in a significantly different manner. We believe that the style transfer method in particular will allow the style of each person's walk to be recognizable.**

***Index Terms*—motion synthesis, generative models, style transfer, virtual reality**

I. Introduction

Generating realistic human motion in a virtual environment is a challenging task. While historical approaches exist (e.g. using statistical models to predict joint movement [2] or using models that employ the laws of physics [3], there are fewer that implement concepts from a modernized research area: artificial intelligence. In addition, none of the prior approaches are geared towards generating motion in Virtual Reality, much less navigating from a given start to a given destination. Thus, we are proposing a style transfer model that will allow a person to provide base motion as an input and synthesize motion in VR that can navigate environments while mimicking the input user's style. This work can significantly help in the development and testing of VR-based interfaces like SEAN, and can also assist in synthesizing realistic motion for HRI experiments (e.g. the BWI project at UT Austin).

II. Background

In [2] and [5], Wang discusses several historical and modern approaches useful for motion synthesis and prediction.

In [3], Wei et. al augment upon prior statistical approaches using the concept of a "force field" that is applied onto a user, modifying the movement based on external and/or environmental factors. By modeling the joints in a human body as a 48-dimensional vector and applying dynamic system modeling with statistical priors, they are able to produce a realistic and scientifically-correct motion.

In [4], an approach to synthesize both 3D and 2D animations using a content (base) motion and a style motion is detailed with the use of a generative-styled network and discriminator. This is largely what we seek to model our approach off of.

In [6], Holden et al. are able to generate motion by training a motion manifold (which consists of a convolutional autoencoder) independently of another control network. Though this approach is fruitful, modernized approaches seem to discourage the use of CNNs due to their limitations in what they can generate.

In [1], a VR interface that could be useful for testing out synthesized motion against real human agents is described.

III. Approach

Our group is heavily modeling our model off of the existing paper Unpaired Motion Style Transfer from

Video to Animation.

Currently, style transfer approaches exist for 2D animations. The program is provided a base image with a series of motions of a stick figure character, and a desired style, and will transfer that style onto the base. We are attempting to apply this same methodology to a 3D environment. One benefit of this style transfer approach is that you end up with novel animations and not simply replications of whatever training data you provide.

i **Generating Inputs**
The style transfer approach requires 2 inputs: a basic, general walking animation and walking motion for each human that we want to represent(the style data. For the general data, we will record one person walking in a fairly uniform, calm manner and copy that as the basic motion. For the style data, we will record exaggerated motions. Both of these inputs will feed into our model. The inputs will consist of pose information about all of the joints that are being tracked, either relative or absolute. The general motion only has to be recorded once and can be used by any style for training, while each style data set must be recorded uniquely for training.

ii **Style Transfer Pipeline**
Our model is largely based on the existing style transfer approach [4]. Our model will require only 3 dimensional inputs while the Aberman model supported both 2-D and 3-D. The model will take in both the default walking motion data and the style walking data. The style walking data will merge with the default data to create an output that demonstrates a stylized version of the default walk. The output will return data in a similar format to that of the inputs; aka a vector of joint positions per time step.

iii **Creating Animation from Outputs**
After receiving the joint poses for each time step, we will use Unity, Blender, or Maya to animate a human-like figure from time step to time step. We will then test to make sure there are noticeable differences for each model that is trained on a different walking style, most likely from participant surveys. We will also attempt to extend this walking animation to HRI studies; for example, seeing if a real human can interpret the VR avatar motion correctly when it is making a turn.

Future Work: Eventually we could apply this technique to different base motions, such as turning left, turning right, running, etc. and generate motion from the style motion. This way, you can generate a wide range of motions without having to provide various versions of a given walking style.

IV. Timeline

TABLE I Project Timeline

9/28	Have the existing model fully running
10/12	Generate our own data using VR trackers in lab
10/21	Render model output through Unity/VR interface
10/27	Existing model can successfully run on our own training data
10/28	Initial Paper due
11/02	Program Committee Meeting
11/04	Initial Presentation
11/18	Incorporate multiple base motions; allows for turning motion in HRI experiments
12/02	(tentative) Finish editing model using improvements listed in paper; may also incorporate model into other experiments at this time
12/11	Final Paper and Presentation

V. Conclusion

We predict that the style transfer model will create virtual reality agents that don't copy the style of the human they were trained on, but will imitate the *style* of the human they were trained on. We believe that this is key in creating realistic human motion; instead of one general walking motion, agents should learn motion from individual humans. We predict that humans who interact with these agents in virtual reality will feel more comfortable interpreting their motion than standard VR avatar motion. In general, our goal is to create inconsistent motion that varies depending on the human it was trained on; and expect that this inconsistency will produce naturalistic walking motions.

References

[1] N. Tsoi, M. Hussein, J. Espinoza, X. Ruiz, and M. Vázquez, "SEAN: Social Environment for Autonomous Navigation," Arxiv https://arxiv.org/pdf/2009.04300.pdf (accessed September 10, 2021).

[2] Qi Wang. Statistical Models for Human Motion Synthesis. Modeling and Simulation. Ecole Centrale Marseille, 2018. English. ffNNT : 2018ECDM0005ff. fftel-02071347f

[3] X. Wei, J. Min, and J. Chai, "Physically valid statistical models for human motion generation," in ACM Transactions on Graphics, vol. 30, Issue 3, Article No.: 19, pp. 1–10.

[4] K. Aberman, Y. Weng, D. Lischinski, D. Cohen-Or, and B. Chen, "Unpaired Motion Style Transfer from Video to Animation," in ACM Transactions on Graphics, vol. 39, Issue 4, Article No.: 1, pp. 1–10.

[5] Qi Wang, Thierry Artières. Motion Capture Synthesis with Adversarial Learning. Intelligent Virtual Agents, Aug 2017, Stockholm, Sweden. ffhal-01691463f

[6] Holden, D, Saito, J & Komura, T 2016, 'A Deep Learning Framework for Character Motion Synthesis and Editing', ACM Transactions on Graphics, vol. 35, no. 4, 138. https://doi.org/10.1145/2897824.2925975