This project has received funding from the European Union's Horizon Europe research and innovation programmed under Grant Agreement No 101092889, Topic HORIZON-CL4-2022-HUMAN-01-14

SHARESPACE

Embodied Social Experiences in Hybrid Shared Spaces

Project Reference No	101092889
Deliverable	D4.5. Style Preserving Animation of Avatars
Work package	WP4: Interaction-Aware Avatar Animation and Rendering
Nature	D (Deliverable)
Dissemination Level	PU - Public
Date	May 2024
Status	V1.0
Editor(s)	Panayiotis Charalambous (CYENS) Amir Azizi (CYENS) Andreas Panayiotou (CYENS)
Involved Institutions	CYENS, UM, UJI, GOLAEM
Document Description	This deliverable presents the way gestures are expressed through movement in different cultures, genders and age groups and is a crucial factor that must be preserved in a social XR setting.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Term / Abbreviation	Definition
AI	Artificial Intelligence
Mocap	Motion Capture
VR	Virtual Reality
XR	Extended Reality
DL	Deep Learning
NSM	Neural State Machine
MOE	Mixed of Expert
ERD	Encoder–Recurrent–Decoder
PFNN	Phase-Functioned Neural Network
DDPM	Denoising Diffusion Probabilistic Model
SAAE	Sequential Adversarial Autoencoder
SHS	ShareSpace
RBFs	Radial Basis Functions

1. INTRODUCTION

1.1. Purpose of the document

The way gestures are expressed through movement in different cultures, genders and age groups is a crucial factor that must be preserved in a social XR setting. Thus, a big challenge is to animate L1-L2 avatars (L3 avatars are fully autonomous and therefore their rendering is less challenging) preserving style and intentions of participants while matching the chosen avatar's characteristics. we will develop new methods to transfer motion styles that do not violate the sensorimotor primitives (WP2).

1.2. Structure of the document

The document is organized as follows:

- The first section of this document presents the definition of style.
- The second section focuses on describing the state-of-the-art technologies for motion style transfer.
- Finally, the third section presents the chosen approach before concluding this document.

2. STATE OF THE ART

2.1. Introduction

The ShareSpace project aims to capture and replicate gestures across cultures, genders, and age groups in social Extended Reality settings. It's challenging to animate avatars while maintaining participants' styles. **STYLE** has various meanings and refers to different fields: someone fashionably wears clothes with a specific style, a painter has a specific drawing style, there are different styles of music, etc. Style is important in animation because enhances the motion diversity and adds realism and expression to character movement.
Figure 1 depicts the common trends of styles in the context of human body motion and classification of style, when seen as individual-related features[1]. Figure 2 presents the Individual related features and their occurrences, per style, based on recent reviewing of research works.

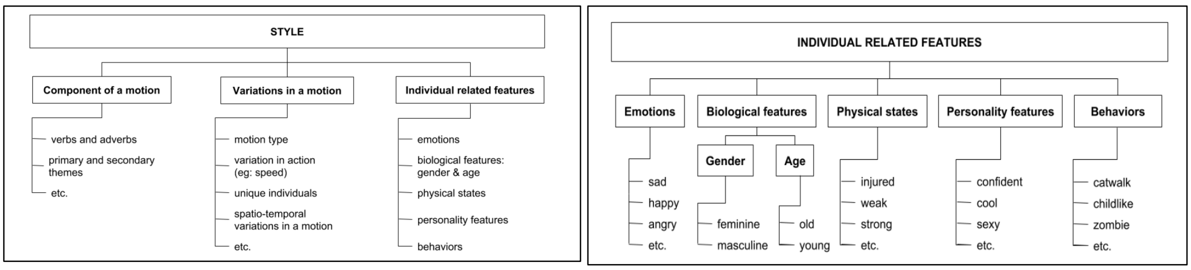

Figure 1: Common trends of the definition of style.

[1] Sarah Ribet, Hazem Wannous, Jean-Philippe Vandeborre. Survey on Style in 3D Human Body Motion: Taxonomy, Data, Recognition, and its applications. IEEE Transactions on Affective Computing, 2021,12 (4), pp.928-948. 10.1109/TAFFC.2019.2906167. hal-02420912

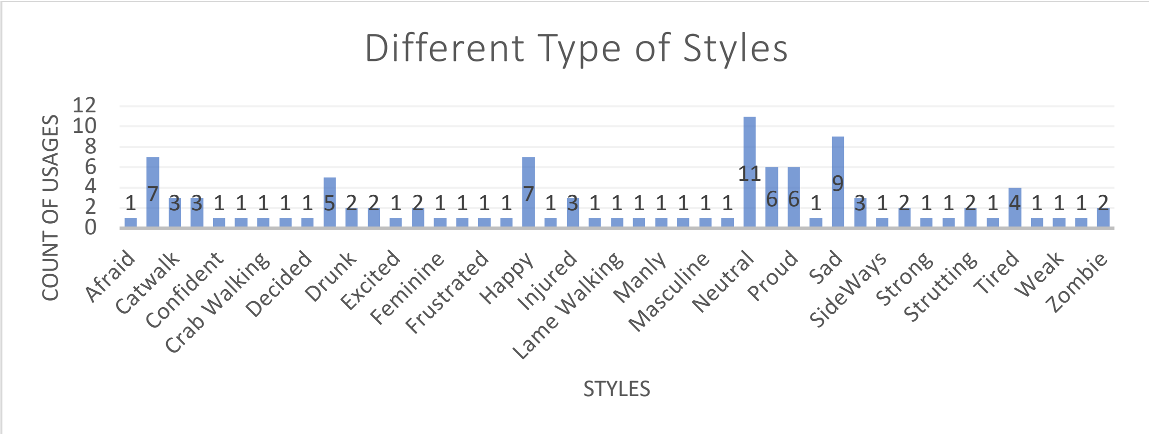

Figure 2: Styles and their occurrences.

Animation and video games require a large amount of data containing pairs of actions and styles. Actors find it difficult and time-consuming to capture all possible combinations. One solution is to generate motions instead of capturing all possible combinations of actions and styles. This process can be described with the term Motion Style Generation and involves the following three methods:

- **Motion Style Synthesis:** creates unique motion styles from scratch using generative models.
- **Motion Style Editing:** fine-tunes existing motion styles to suit artistic or functional constraints.
- **Motion Style Transfer:** transfers motion characteristics from a reference to a target animation.

Each approach serves distinct purposes within the realm of motion style manipulation.

In motion style transfer, "neutral style”,Figure 3 , refers to a basic form of motion that lacks any specific stylistic characteristics, such as exaggeration, emotion, or cultural traits. This plain, unembellished motion serves as a generic baseline for applying different styles, acting as a reference point against which other styles can be compared and measured. A neutral style provides a foundation for style transfer, offering a clean slate onto which various motion styles can be applied. For example, a neutral walking sequence can be transformed into different emotional walks, like happy, sad, or tired. This versatility is crucial in animation and games, where diverse and realistic motion behaviors are needed, and in virtual reality and robotics, where adaptable movements for avatars or robots must suit different contexts or personalities.

The process involves capturing or generating a neutral motion sequence free from stylistic influences. Then, a style transfer algorithm is applied to modify this neutral motion, imbuing it with the desired stylistic features while retaining the fundamental movement pattern. By using a neutral style as the foundation, motion style transfer can more effectively and accurately create a wide range of stylistic motions, enhancing the versatility and realism of animated characters and robots.

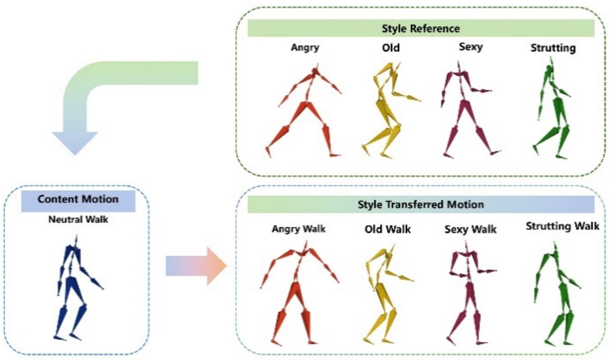

Figure 3:Neutral Style transfer to different styles[2]

[2] Qian, Z., Xiao, Z., Wu, Z., Yang, D., Li, M., Wang, S., ... & Zhang, L. (2024). SMCD: High Realism Motion Style Transfer via Mamba-based Diffusion. *arXiv preprint arXiv:2405.02844*.

2.2. Review of the approaches

Over time, numerous methods have been proposed for creating stylized human movement. Motion style transfer methods, which involve transferring the style of one animation clip to another, while retaining the content of the latter, have become increasingly popular in the past decade. Various data-driven and AI-powered techniques have successfully achieved motion style transfer for character animation. Nevertheless, the absence of paired and labeled motion data limits the widespread adoption of such approaches[3]. Pan et al. developed a motion transfer scheme using a meta-network, transferring input motion styles while maintaining primitive motion behavior, and trained a convolutional feed-forward network for the transformation function.[4] Holden et al. proposed a method, utilizing a neural network for motion style transfer, consisting of a convolutional auto-encoder and a feed-forward convolutional network for loss calculations.[5] Similarly, Smith et al. Presented another method for motion style transfer, generating stylized motion from diverse motion sequences without retargeting or action labels.[6] Aberman et al. developed a motion style transfer framework that can extract styles from video clips, 3D-animated characters, and 2D projections of 3D motions. The framework uses a deep convolutional neural network to train a universal style extractor.[7] Xia et al. introduced an online learning method for motion style transfer, which involves transferring input motion data into output-style frames. The method uses local collections of auto-regressive models to match input poses from training data, extracting input poses from output styles through linear transformations, and building new local models for each successive pose.[8] Mason et al. focused on a transfer learning method, utilizing neural networks to transform existing models into new ones. The model uses few-shot learning from limited data, trained on a few styles using a phase-functioned neural network (PFNN), and extracts style-agnostic attributes to produce output.[9] Dong et al. introduced the adult-to-child (Audlt2child) motion style transfer algorithm, to taggle the challenges of transferring style to children, due to the differences in skeleton size, limb dimensions, and movement speeds, and the difficulty in motion-capturing for children due to their differing understanding levels.[10] Tao et al. proposed a style encoder–recurrent–decoder (ERD) framework for real-time online motion style transfer. The framework uses styles as input streams and embeds prior frame knowledge in the style transfer module's memory, generating high-quality style transfer.[11] Chang et al. presented the denoising diffusion probabilistic model (DDPM) for styled motion synthesis, a diffusion-based solution that models content and style in a shared representation, employing adversarial training to harmonize predictions and enable global movement perception.[12] Jiang et al. introduced a motion puzzle framework that transfers motion styles by body parts, extracting styles from various motions and translating them to desired body parts. The framework preserves specific motion content by controlling individual body parts' styles.[13] Wang et al. developed a neural network-based architecture for motion style transfer, consisting of a sequential adversarial autoencoder (SAAE) and a style discriminator. The SAAE

[3] Neverova, N.; Wolf, C.; Lacey, G.; Fridman, L.; Chandra, D.; Barbello, B.; Taylor, G.W. Learning Human Identity from Motion Patterns. IEEE Access 2016, 4, 1810–1820.

[4] Jian, P.; Huaijiang, S.; Yue, K. Fast human motion transfer based on a meta network. Inf. Sci. 2021, 547, 367–383

[5] Holden, D.; Habibie, I.; Kusajima, I.; Komura, T. Fast Neural Style Transfer for Motion Data. IEEE Comput. Graph. Appl. 2017, 37, 42–49.

[6] Smith, H.J.; Cao, C.; Neff, M.; Wang, Y. Efficient Neural Networks for Real-time Motion Style Transfer. Proc. ACM Comput. Graph. Interact. Tech. 2019, 2, 1–17.

[7] Aberman, K.; Weng, Y.; Lischinski, D.; Cohen-Or, D.; Chen, B. Unpaired motion style transfer from video to animation. ACM Trans. Graph. 2020, 39, 64.

[8] Xia, S.; Wang, C.; Chai, J.; Hodgins, J.K. Realtime style transfer for unlabeled heterogeneous human motion. *ACM Trans. Graph.* 2015, *34*, 119:1–119:10.

[9] Mason, I.; Starke, S.; Zhang, H.; Bilen, H.; Komura, T. Few-shot Learning of Homogeneous Human Locomotion Styles. Comput. Graph. Forum 2018, 37, 143–153.

[10] Dong, Y.; Aristidou, A.; Shamir, A.; Mahler, M.; Jain, E. Adult2child: Motion Style Transfer using CycleGANs. In Proceedings of the Motion, Interaction and Games (MIG), North Charleston, SC, USA, 16–18 October 2020.

[11] Tao, T.; Zhan, X.; Chen, Z.; van de Panne, M. Style-ERD: Responsive and coherent online motion style transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.

[12] Chang, Z.; Findlay, E.J.; Zhang, H.; Shum, H.P. Unifying Human Motion Synthesis and Style Transfer with Denoising Diffusion Probabilistic Models. arXiv 2022, arXiv:2212.08526.

[13] Jang, D.K.; Park, S.; Lee, S.H. Motion puzzle: Arbitrary motion style transfer by body part. ACM Trans. Graph. (TOG) 2022, 41, 1–16.

encodes the input sequence's style/emotion label, while the style discriminator extracts it from its encoding representation.[14]

3. Proposed Method for Motion Style Transfer

At this stage of SHS, for style motion transfer we're using the method outlined in this article[15]. This work presents a method for modeling style and effect in human motion by utilizing user-defined parameters from experienced animators, to edit motion sequences and synthesize variations in walking styles such as happy, sad, feminine, masculine, energetic, and tired. This approach involves converting neutral walks into different variants through Gaussian radial basis functions (RBFs) provided by animators, which are then compiled into an expert-driven set of features. These features are then applied to motion sequences and evaluated through a perception study to validate their effectiveness, efficiency, scalability, and inversion properties in perceiving style and affect from human motion. The steps of the proposed method are as follows:

1. Collect inputs from experienced animators to edit motion sequences using Gaussian radial RBFs for synthesizing variations in walking styles.
2. Compile the RBF edits provided by animators into an expert-driven set of features that can transform neutral styles into happy, sad, feminine, masculine, energetic, and tiring variants.
3. Apply the expert-driven feature set to motion sequences and conduct a perception study to validate the effectiveness, efficiency, scalability, and inversion properties of the proposed models in perceiving style and affect in human motion.

Table 1 shows the specific features that need modification to achieve each style based on the input file.

		Themes					
		Happy	*Sad*	*Energetic*	*Tired*	*Feminine*	*Masculine*
Common features	*1*	Shoulders: increased swing along Z	Shoulders: tilted along $-Y$	Knees: increased swing along Y	Shoulders: tilted along $-Y$	Hip: increased swing along X	Shoulders: increased swing along Z
	2	Wrists: increased swing along Z	Head: tilted along $-Y$	Head: tilted along $+Y$	Head: tilted along $-Y$	Ankles: tilted, R along $+X$, L along $-X$	Knees: tilted, R along $-X$, L along $+X$
	3	Knees: increased swing along Y	Neck: tilted along $-Y$	Elbows: increased swing along X	Ankles: decreased swing along Z	Knees: tilted, R along $+X$, L along $-X$	Ankles: tilted, R along $-X$, L along $+X$
	4	Head: increased swing along X	Shoulders: decreased swing along Z	Elbows: increased swing along Z	Head: tilted along $+Z$	Torso L: increased swing along X	Elbows: tilted, R along $-X$, L along $+X$
	5	Wrists: increased swing along X	Wrists: decreased swing along Y	Shoulders: increased swing along Y	Wrists: tilted along $-Y$	Thighs: tilted, R along $+X$, L along $-X$	Ankles: increased swing along Y
	6	Hip: increased swing along X	Torso U: tilted along $-Y$	Shoulders: increased swing along Z	Ankles: decreased swing along Y	Torso U: increased swing along X	Knees: increased swing along Y
	7	Knees: increased swing along X	Head: tilted along $+Z$	Wrists: increased swing along Z	Neck: tilted along $-Y$	Thighs: increased swing along Y	Head: increased swing along Z
	8	Neck: tilted along $+Y$	Neck: tilted along $+Z$	Wrists: increased swing along X	Neck: tilted along $+Z$	Shoulders: tilted along $-Y$	Neck: increased swing along Z
	9	Head: tilted along $+Y$	Elbows: tilted along $-Y$	Thighs: increased swing along Z	Torso L: tilted along $+Z$	Elbows: increased swing along X	Shoulders: increased swing along Z
	10	Head: tilted along $-Z$	Hip: tilted along $-Y$	Thighs: increased swing along Y	Torso U: tilted along $+Z$	Wrists: increased swing along X	Elbows: increased swing along Z

Table 1: Specific features that need modification to achieve a specific style.

4. Results

We select Feminine and Old styles and apply the method presented in Section 3, following these steps:

[14] Wang, Q.; Chen, M.; Artières, T.; Denoyer, L. Transferring style in motion capture sequences with adversarial learning. In Proceedings of the European Symposium on Artificial Neural Networks, ESANN, Bruges, Belgium, 25–27 April 2018.

[15] S. A. Etemad and A. Arya, "Expert-Driven Perceptual Features for Modeling Style and Affect in Human Motion," in *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 4, pp. 534-545, Aug. 2016, doi: 10.1109/THMS.2016.2537760.

1. Use the **Pymotion**[16] library to load and edit the BVH animation files. (PyMotion is a Python library built for easy handling and processing of motion data in either NumPy or PyTorch.)
2. Modify the specified joints based on the provided features to achieve the desired style. This involves adjusting rotation matrix values.
3. Save the modified BVH animation file containing the desired styled motion.

Figure 4 illustrates the pipeline.

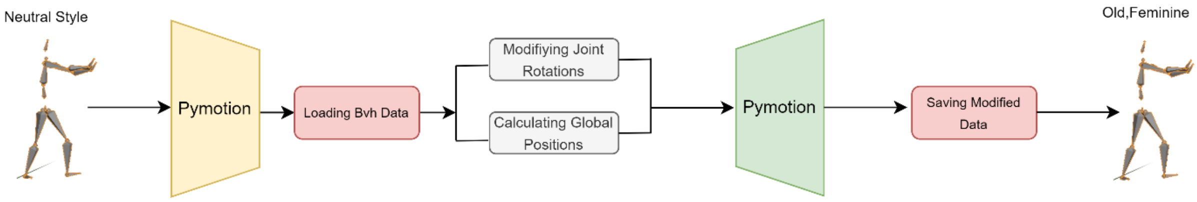

Figure 4: Steps to create a new style from the input neutral style with Pymotion.

5. CONCLUSION

Numerous studies have explored the modeling of motion styles using recorded sequences combined with machine learning techniques. However, our approach diverges by drawing on the expertise of experienced animators. This ensures that our models are not only efficient but also highly effective. Moving forward, we'll assess the proposed features across various skeleton models and actions to gauge their practicality and significance. Furthermore, we aim to develop a tool using these features to automatically recognize and locate motion sequences. By comparing it with existing systems, we'll gain insights into the adaptability of our features across diverse types of motion.

6. VIDEO

At this stage, using our proposed method, we applied motion style transfer to create old-style and feminine-style animations. You can view the results in the video linked below.

https://www.youtube.com/watch?v=Pc4dosbmKuI&t=4s

[16] https://github.com/UPC-ViRVIG/pymotion