Blob Bodies: Application of Style Transfer to Generate Body Tracking Art

Lucia Rhode
The Cooper Union
lucia.rhode@cooper.edu

Nicole Joseph
The Cooper Union
nicole.joseph@cooper.edu

Abstract: Body tracking has become a popular medium for interactive art, and this project aims to investigate the potential for using pose segmentation and style transfer to imitate simulation art. In interactive body tracking art, it is common to use physics-based interactions and artistic rendering tools in addition to body tracking to achieve a certain artistic look that responds to the observer's motions. This project aims to examine whether these tools can be circumvented by using style transfer, and whether the resulting artwork can be convincing enough to trick the viewer into believing that there are artistic renderings or simulations involved. Additionally, this project aims to determine a balance between interactivity and realism, and explore the feasibility of implementing pose estimation and style transfer in real time with limited computing power.

1. Introduction

Interactive art is an artistic medium that engages the observer and allows them to become a part of the artwork. In recent times, interactivity is often achieved through the use of technology such as computers, sensors, and interfaces that take in inputs from the observer or the environment to respond to motion, heat, or touch. With the emergence of motion capture and body tracking technologies, it has become possible to map the movement of humans onto digital elements, creating a more immersive and interactive experience.

A popular form of interactive body tracking art seen on YouTube often involves mapping human joints onto particles, having particles react to human movements, or having beautiful renderings react to human motion (see references [1]-[3]). However, achieving this effect often requires complex coding of physics-based simulations or the use of artistic renderings tools like Unity and Blender. In this research paper, we aim to investigate whether similar results can be achieved using a combination of pose estimation and style transfer deep learning algorithms, which have the potential to simplify the process of creating interactive

body tracking art, allowing for experimentation to become more accessible. Through this exploration, we aim to identify the potential benefits and limitations of using these techniques.

2. Related Work

In the entertainment industry, marker-based 3D tracking is a widely used method for motion capture in video games and animated films. This method involves the use of multiple optical cameras and reflective markers placed on the human body to triangulate the 3D position of the subject. While highly accurate, marker-based motion capture requires expensive proprietary equipment and is time consuming to set up and calibrate, making it a less accessible option for artists outside of the big-budget entertainment industry. As an alternative, pose estimation has gained popularity as a low-cost and open source method for body tracking. Pose estimation is marker-less and can be achieved using a single camera. While pose estimation is less accurate than marker-based tracking and is more susceptible to occlusions, it is a useful tool for artistic purposes, especially when the focus is on conveying a general sense of motion rather than precise measurements.

3. Methodology

a. Pose Estimation - BlazePose

For this project, human pose estimation was implemented using BlazePose to track the movement of the human body in real-time [4]. BlazePose is a lightweight convolutional neural network architecture for human pose estimation. Blazepose is capable of estimating the relative 3D position of 33 body key points and predicting a full body segmentation mask to distinguish the human from the background.

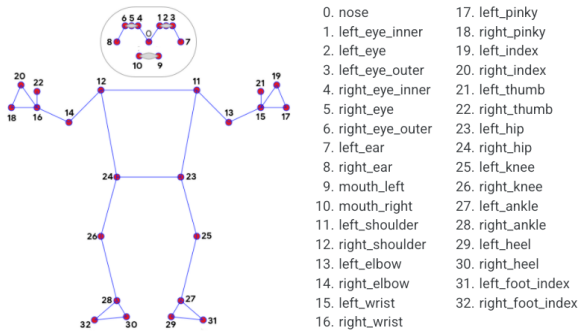

Figure 1

This pose estimation model has been implemented in Google's MediaPipe platform, which provides a cross-platform pipeline framework for machine learning solutions for live and streaming videos. MediaPipe is designed for use on Android and IOS, and offers a python API, making it suitable for our purpose. The combination of BlazePose's real-time performance and MediaPipe's compatibility with mobile devices make it ideal for this project, as pose estimation can be used in real-time without requiring significant computing power.

b. Selfie Segmentation

MediaPipe's Selfie Segmentation was used to identify and segment the prominent humans in the scene of an image or video. The technology is specifically designed for use cases such as selfie effects and video conferencing, where the person is close (< 2 meters) to the camera. It is also capable of running in real time, making it well-suited for the purposes of this project. The API provides two modes: general and landscape.

Both modes are similar in functionality, but the landscape model has a faster inference speed, while the general model is more accurate. The output options for Selfie Segmentation include a binary mask and a full body segmentation mask, which can be used to distinguish the person from the background. MediaPipe's Selfie Segmentation was chosen for its real-time performance and accurate human segmentation capabilities of the full body when they are close to the camera.

c. Style Transfer

Style Transfer is a technique in computer vision that allows for the recomposition of the content of an image in the style of another image. Given a content image and a style reference image as input, the output image preserves the core elements of the content image, but appears to be painted or filtered in the artistic style of the reference image. Style transfer can be applied in a variety of contexts, including photo and video editing, commercial art, gaming, and virtual reality. Neural Style Transfer introduced the use of deep convolutional neural networks in this artistic stylisation task in "A Neural Algorithm of Artistic Style" and utilized the VGG-19 architecture [5]. Earlier versions of Neural Style Transfer required thousands of iterations just for the artistic stylisation of a single image. Fast Style Transfer addressed this issue by training a single model to stylize an image in a single feed forward pass and combining other deep learning techniques for style transfer [6]-[7]. Fast Style Transfer can stylize an image in as little as a fraction of a second; this is especially useful for video stylizations where consecutive frames are stylized and restitched together to produce a final output video. Fast Stable Style Transfer takes this a step further by preserving style invariance, temporal consistency, and the style and content features of inputs in videos. By using the AdaIN operator, the image-transformation network becomes both content and style invariant, meaning that the encoder and decoder parameters are preserved across styles [8].

d. Tensorflow Hub

For these experiments, Tensorflow 2 and the pre-trained Fast Style Transfer Deep Neural Network model [9] from the Tensorflow Hub were used. The Tensorflow Hub is a repository of pre-trained models that one can fine-tune and deploy immediately. The pre-trained style transfer model can be run directly on the computer if a Nvidia GPU is available, or on Google Colab, which gives free access to a GPU.

4. Discussion

a. Experiments

For this experiment, binary segmentation masks were generated for several videos of individual dancers. To improve segmentation around boundaries, a joint bilateral filter was applied using OpenCV. The filter replaced the intensity of each pixel with a weighted average of intensity values from nearby pixels, resulting in a noise-reducing, smoothing effect. To create an interesting visual effect, three copies of the binary segmentation masks were generated and placed in the same frame using OpenCV. When using the landscape mode and not applying the bilateral filter,

this produced an interesting “flame-like” (C. Curro, personal communication, December 15, 2022) particle effect around the silhouettes, as shown below in Figure 2.

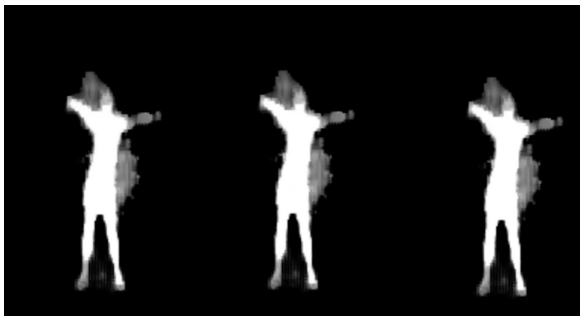

Figure 2
[https://www.youtube.com/watch?v=wrWl-1tBsHM]

However, for the input to the style transfer model, we opted to use the individual black and white segmentation masks. These videos of the segmentation masks produced were used as the content to the style transfer, and well-known art pieces such as Pillars of Creation, a photograph taken by the Hubble Space Telescope, and The Scream by Edvard Munch were applied as the style. After passing a binary segmented video, in which the dancer’s silhouette is white and the background is black, through the style transfer model, it was found that the style did not attach itself to the dancer’s figure in a clear manner, as shown in Figure 3.

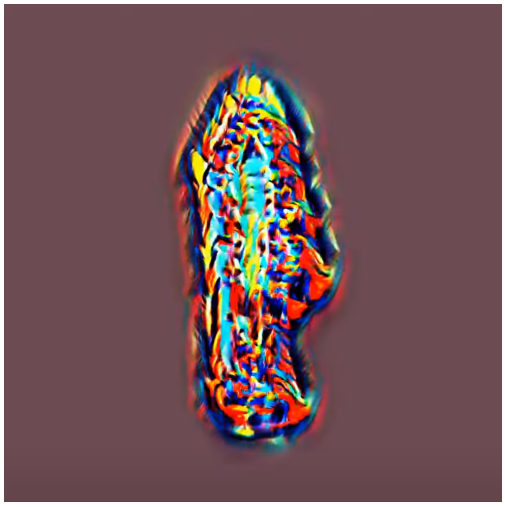

Figure 3
[https://www.youtube.com/shorts/E6WIIDl4OQI]

As an improvement, the binary segmentation mask was changed to output the dancer’s silhouette in black and the background in white. This allowed the dancer’s figure and limbs to be more apparent, as shown below in Figure 4.

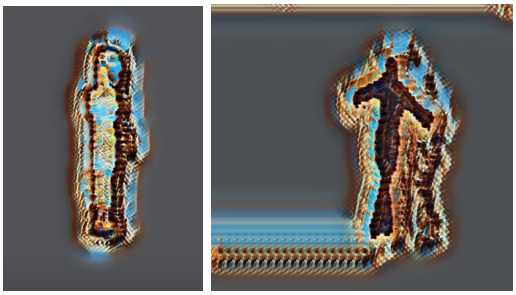

Figure 4
[https://www.youtube.com/shorts/bAfRELEbF5c],
[https://www.youtube.com/shorts/Wf6_QgmRiQI]

We also experimented with customizing the background by setting a background color, loading an image or video as the background with the same width and height of the input image, and blurring the input image through

image filtering. The output in Figure 5, shows the addition of a static background image. While interesting to look at, the physics-based interactions effect was not achieved using a static image and appears more prominent in Figure 4. Further experimentation with dynamic backgrounds is recommended.

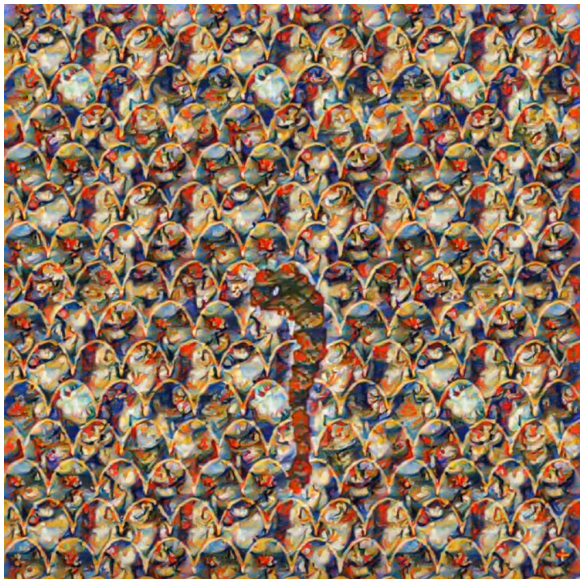

Figure 5
[https://www.youtube.com/shorts/CWp-5SaOJPQ]

b. Hardware Limitations for Real-time Video

Hardware limitations significantly impacted the real-time experiments conducted for this project. One such limitation included the laptop web-camera used for video input, which compromised the resolution and frame rate of the videos fed into the Selfie Segmentation model. The low resolution of the web-cam resulted in blurry and pixelated images, making it difficult to accurately segment the dancer's body in the scene.

Another limitation was the lack of a NVIDIA GPU, which hindered the use of Cuda and resulted in slower processing times and lower-quality masks. To mitigate this issue, higher quality videos that were previously recorded on a phone were used. However, without GPU, the average processing time for each frame run through the style transfer model was about 20 seconds per frame. This meant that a 9 second with a frame rate of 30 frames per second took 90 minutes to process.

These hardware limitations ultimately impacted the performance and accuracy of the Selfie Segmentation model, as well as the potential for real-time due to the limitations of processing speeds of style transfer. To improve the results in future studies, it will be necessary to use more advanced hardware, including a higher-resolution camera and a GPU, to ensure better image quality and faster processing times.

5. Conclusion

In conclusion, the combination of pose segmentation and style transfer can be used to create visually interesting body-tracking art that has the potential to imitate more complex physics-based simulations and artistic renderings. During the exploration of this technique, it was found that body segmentation works effectively in real-time, while style transfer was found to be very slow due to the hardware limitations. However, with access to better hardware, using pose estimation in conjunction with style transfer remains a viable, low-cost, and

open-source alternative for creating interactive body-tracking art without the need for motion capture, coding physics-based interactions, or using artistic rendering tools.

Overall, this research demonstrates the potential of using pose estimation and style transfer to create interactive body-tracking art. Future work could focus on improving the efficiency of style transfer models for video and real-time use as well as exploring more filtering and background techniques to make the art feel even more immersive and visually stimulating.

References

[1] Žiga Trontelj. Particles - Interactive new-media art installation (Body tracking) (Nov 20, 2019). Accessed: 1 December 2022. [Online Video]. Available: https://www.youtube.com/watch?v=VPYxQFViuag

[2] Ray Chen. Interactive Tracking and Particle Demo (Feb 8, 2018). Accessed: 1 December 2022. [Online Video]. Available: https://www.youtube.com/watch?v=gT76k-Fbtdk

[3] Touch world-sh Mark. 5 particles use Kinect and Processing (Mar 3, 2020). Accessed: 1 December 2022. [Online Video]. Available: https://www.youtube.com/watch?v=kJg3-5Y-emA

[4] Bazarevsky, Valentin, et al. "Blazepose: On-device real-time body pose tracking." arXiv preprint arXiv:2006.10204 (2020). URL: https://arxiv.org/abs/2006.10204

[5] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." arXiv preprint arXiv:1508.06576 (2015). URL:https://arxiv.org/abs/1508.06576

[6] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." European conference on computer vision. Springer, Cham, 2016. URL:https://arxiv.org/abs/1603.08155

[7] Ulyanov, Dmitry, Andrea Vedaldi, and Victor Lempitsky. "Instance normalization: The missing ingredient for fast stylization." arXiv preprint arXiv:1607.08022 (2016). URL: https://arxiv.org/abs/1607.08022

[8] Huang, Xun, and Serge Belongie. "Arbitrary style transfer in real-time with adaptive instance normalization." Proceedings of the IEEE international conference on computer vision. 2017. URL: https://arxiv.org/abs/1703.06868

[9] Ghiasi, Golnaz, et al. "Exploring the structure of a real-time, arbitrary neural artistic stylization network." arXiv preprint arXiv:1705.06830 (2017). URL:https://doi.org/10.48550/arXiv.1705.06830