CS6244 Project Final Report
Neural Human and Object Interaction

Yan Zhiwen A0164790X

Chen Hanlin A0262444H

1 Project Description

This project aims to simulate the physical interaction between human neural radiance fields (NeRF) with geometric objects. Given the dynamic neural human representation and a well-defined geometric object, we aim to simulate the motion of the geometric object under the physical interaction with the neural human. The motion of the object includes the angular and linear velocity of the object, under the linear and angular acceleration caused by contact force with the neural human. The force feedback on the neural human is not reflected since it is a reconstruction of the real-world.

2 Motivations

Neural radiance fields have become a new approach to reconstructing dynamic 3D human sequences with a volumetric representation. Novel view synthesis, playback, editing, and novel pose generation have been explored to further extend the application of this method. However, the problem of physical interaction between dynamic human NeRF and other objects remains largely unexplored.

The physical interaction between dynamic neural representation and geometric objects, or even another neural representation, is more sophisticated form of understanding and manipulating this new 3D representation. If this innovation is successful, it can enable a wide range of applications in areas including AR/VR, remote technical support, gaming, etc. A simple example would be the user reconstructed with dynamic human NeRF moving and playing with virtual objects floating in the virtual 3D space. The user can stream this process as a performance, or play a game in this form.

A key difference between this neural interaction with other existing skeleton-based human-object interactions is that it is fundamentally not limited to human-object interaction. The neural radiance field can be used to reconstruct any object in the real-world and interact with objects from the virtual environment. A user can play basketball at home with his own basketball bat without the system even detecting the bat explicitly. The user can even use his/her book or legs to hit the ball if he/she wants. This enables a brand new level of immersive interaction and marks an interesting advancement in the use of the neural radiance field.

3 Scope of Project

The project focuses on the physical interaction simulation between a dynamic human NeRF with a geometric object. If time allows, it will be extended to the interaction with another static NeRF object. The project will not focus on improving the reconstruction quality of the existing dynamic human NeRF models. As this is the first work in this direction, the physical simulation will not be very polished.

Formally, we define the problem as the following: given a dynamic human NeRF model $f_{NeRF}(x, d, t) \rightarrow \sigma, c$ and a geometric object with an initial surface or volume representation $f_{obj}(\theta)$ at t_0, predict the position x_{obj}, rotation R_{obj}, and motion v_{obj}, ω_{obj} of the object under the physical contact with the human NeRF at any time t, where the $x, d, t, \sigma, c, \theta$ represents position, viewing direction, time, density, color, and geometric parameters respectively.

4 Novelty

There has not been any work focusing on the physical interaction between dynamic human/object NeRF with other objects. The closest previous work(2) only applies to the physical interaction with a static NeRF. Our proposed algorithm, however, captures the dynamic motion of human in real-life and transfer the motion to a physical simulation space.

This naturally leads to the challenges of calculating and handling the dynamics of the moving human representation. To simulate physical interaction in a volumetric representation, we need to find the velocity of each point in the space occupied by the human at any time and transfer it to a force acting on the geometric object if they are in contact. This is more difficult than the existing physical interaction work where all the point velocities in the 3D space are known as it is a fully parametric representation.

5 Methodology

We tackle the challenge of physical interaction in two steps:

1. Calculate the velocity of any point in the dynamic neural human representation and the geometric object.
2. Calculate the contact force caused by relative point motion and apply it to the geometric object in the form of force and torque.

5.1 Point motion estimation

For the point motion estimation, we aim to predict the 3D velocity of any point in the dynamic 3D human NeRF at any given time. The typical dynamic human NeRFs warp any point x in the observation space at any time t to a common canonical space point x' for rendering as shown in Fig. 1(4). This process usually exploits the human parameterized models like SMPL as a pose prior. Our algorithm can use this to establish point correspondence from any time t to a few frames before and after the target frame, to calculate a temporary 3D velocity of any point.

More specifically, we use a dynamic neural human representation with forward and backward transformation functions for each point. We use a forward deformation field $d(x_{obs}, t) \rightarrow x_{can}$ to warp any point from the observation space to canonical space and a backward deformation $d'(x_{can}, t) \rightarrow x_{obs}$ to warp any point from the canonical space to observation space, which is shown in 2. We choose the forward-backward model from SNARF (1) as our backbone, which enforces a cycle consistency between a SMPL blending-based forward model and a root-finding-based backward model for a more accurate mapping even with extreme poses.

With the backward deformation field $d(x_{obs}, t) \rightarrow x_{can}$ and the forward deformation field $d'(x_{can}, t) \rightarrow x_{obs}$, we can then calculate the point velocity on the neural human body v_{hx} by subtracting the position of the same point at the current frame and the last frame:

$$v_{hx} = \Delta x_h = x - d'(d(x_{obs}, t), t-1). \tag{1}$$

To calculate the relative velocity, we also need the point velocity on the geometric object. Luckily, the geometric object is a virtual object where its object linear velocity v_g and angular velocity ω_g are known. We calculate the velocity of a point that is at radius r_x from the center of mass of the object as:

$$v_{gx} = v_g + r_x \times \omega_g. \tag{2}$$

With both the point velocity on the neural human body v_{hx} and the geometric object v_{gx}, we can easily calculate the relative velocity v_x as:

$$v_x = v_{hx} - v_{gx}. \tag{3}$$

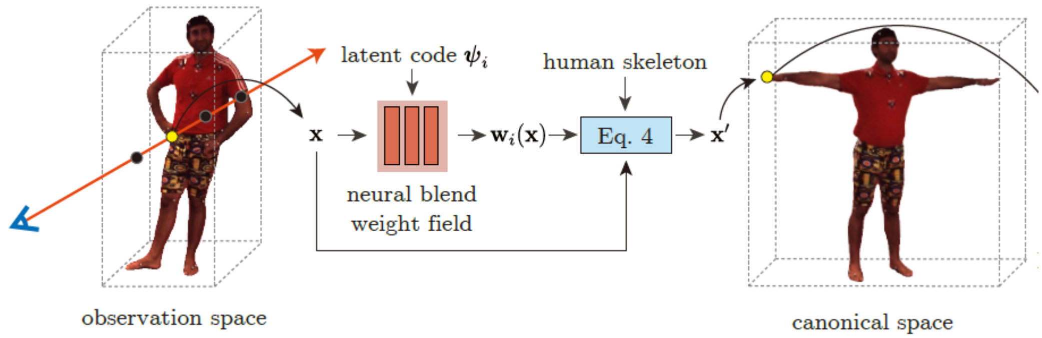

Figure 1: Dynamic human NeRF warps the point in the observation space to the canonical space.

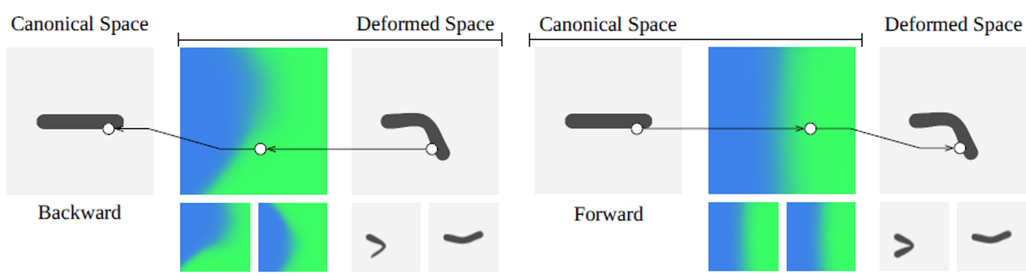

Figure 2: An illustration of a forward backward wrapping between the observation space and the canonical space.

5.2 Physical Simulation

For the physical interaction, we need to first check for collision and then calculate the force on all sampled points to form a resultant force on the geometric object.

For collision checking, we sample N points uniformly on the geometric object since it is in a much more regular shape than the neural human. A point is a colliding point between the neural human $f_{NeRF}(x,t)$ and the geometric object $f_{obj}(x,t)$ if both of them occupy this point. This means the occupancy boolean status $o_t(x)$ at point x and time t is calculated as:

$$o_t(x) = \{f_{NeRF}(x,t) > \sigma_{thresh} \ and \ f_{obj}(x,t) > 0\}. \tag{4}$$

With the collision detected, we can then calculate the force and torque on each of the point and aggregate them to become the resultant force and torque acting on the entire geometric object. More specifically, the force is calculated as:

$$F = \lambda_1 Mean_{collide}(v_{hx} - v_{gx}). \tag{5}$$

The torque is calculated as:

$$\tau = \lambda_2 Mean_{collide}(r_x \times (v_{hx} - v_{gx})). \tag{6}$$

The hyper-parameter λ_1 and λ_2 represent the scaling of the force and the inertia of the geometric object. With the force and torque available, the linear and angular acceleration of the geometric object can be easily calculated by dividing the mass.

To form a more plausible physical effect, we also add in the constant gravity force $F_{gravity}$ to the object all the time pointing in the negative y direction so that the object keeps falling down without contacting the human body.

5.3 Repulsion Force

Another effect that we have observed is that the object tends to fall through the human body if the relative velocity between the object and the human body is small. This is caused by us only modeling the contact force derived from relative motion, but not the static contact force.

To solve this issue, we propose another repulsion force F_{rep} to penalize any inter-penetration between the geometric object and the human body. The repulsion force is in the reverse direction from the center of the mass of the object to any point in the inter-penetrated volume between the object and the human body. More specifically, this force is calculated as:

$$F_{rep} = -\lambda_3 Mean_{collide}(r_x). \tag{7}$$

6 Experiments

We demonstrate the interaction visualization (Figure 3,4,5,6) between dynamic neural humans reconstructed in the AMASS(3) dataset. The color of the mesh visualized represents the estimated velocity of the point on the human body. Since the frames are sampled in fixed intervals of the motion sequence, it might not seem very continuous. Please refer to the project website for the videos for a better experience.

Unfortunately, it is difficult for us to do any quantitative evaluations within the limited time. There is no existing dataset with multiview dynamic human and virtual object interaction ground truth at the moment.

7 Limitations and Extensions

There are a few main limitations to the current project. The first limitation is the inaccurate physical simulation. The current project only produces a physically plausible interaction simulation without any accurate physics engine. It might be difficult for the current system to simulate friction and articulated object interaction.

Another limitation is the object interacted with is limited to a simple geometric object. This can be extended to other representations including another object in neural representation. The other object can also be articulated neural object interacting with the human body, but the kinematics of the object needs to be understood by the dynamic neural representation first. This is yet to be solved for neural reconstruction as well.

The third limitation is that the dynamic neural representation is limited to the human body. In the future, this project can be extended to any generic dynamic neural representations which can provide a lot more flexibility to its applications.

This is a very interesting direction of work, and I intend to continue doing this for my own research. I trust that the problem and idea introduced in the project would be well respected although it is a course project.

8 Conclusion

For this project, we proposed a novel problem statement of the volumetric physical interaction between a dynamic neural human representation with a geometric object. The motion of the object under the interaction is simulated by calculating the relative motion and force between the object and the neural human. This project is a crude prototype to a wide range of applications that are yet to be explored.

References

[1] Xu Chen, Yufeng Zheng, Michael J Black, Otmar Hilliges, and Andreas Geiger. Snarf: Differentiable forward skinning for animating non-rigid neural implicit shapes. In *International Conference on Computer Vision (ICCV)*, 2021.

Figure 3: The "aist" motion sequence interaction with the cube.

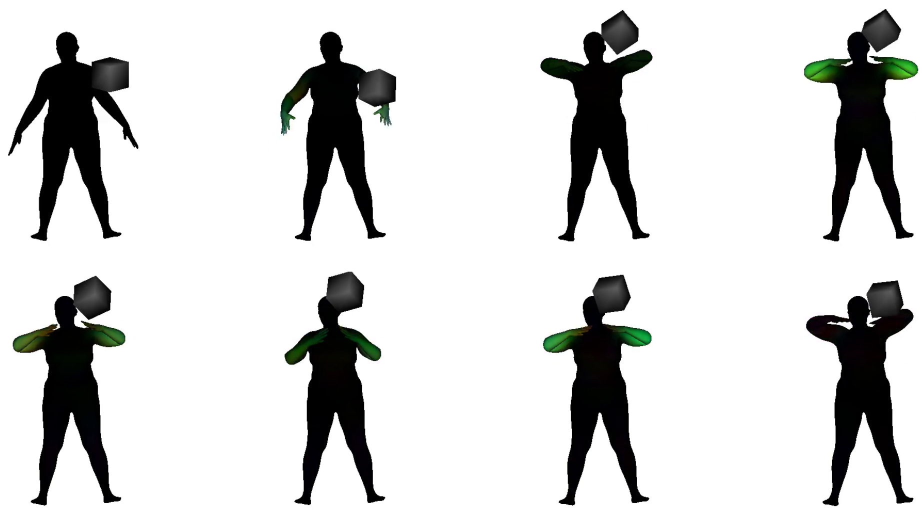

Figure 4: The "chicken" motion sequence interaction with the cube.

[2] Simon Le Cleac'h, Hong-Xing Yu, Michelle Guo, Taylor A. Howell, Ruohan Gao, Jiajun Wu, Zachary Manchester, and Mac Schwager. Differentiable physics simulation of dynamics-augmented neural objects, 2022.

[3] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, pages 5442–5451, Oct. 2019.

[4] Sida Peng, Junting Dong, Qianqian Wang, Shangzhan Zhang, Qing Shuai, Xiaowei Zhou, and Hujun Bao. Animatable neural radiance fields for modeling dynamic human bodies. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14314–14323, 2021.

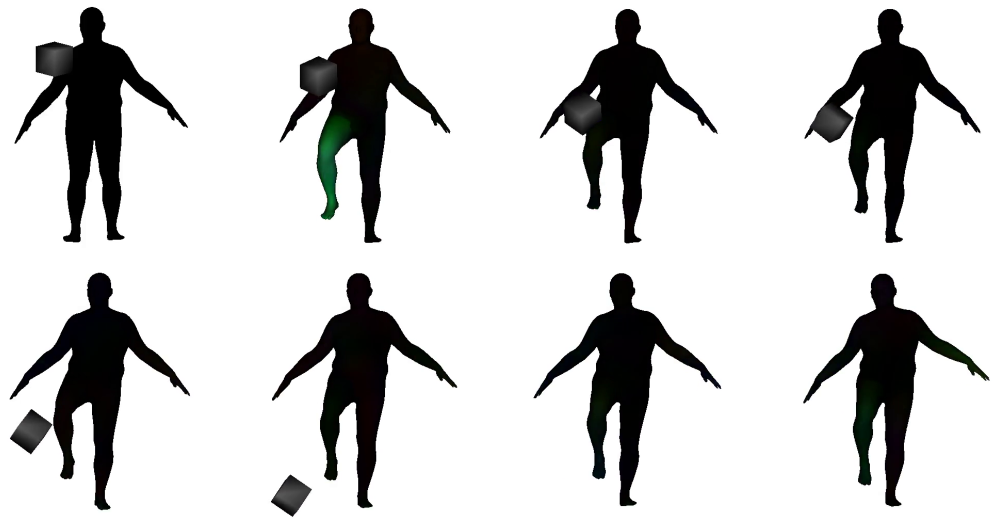

Figure 5: The "stand" motion sequence interaction with the cube.

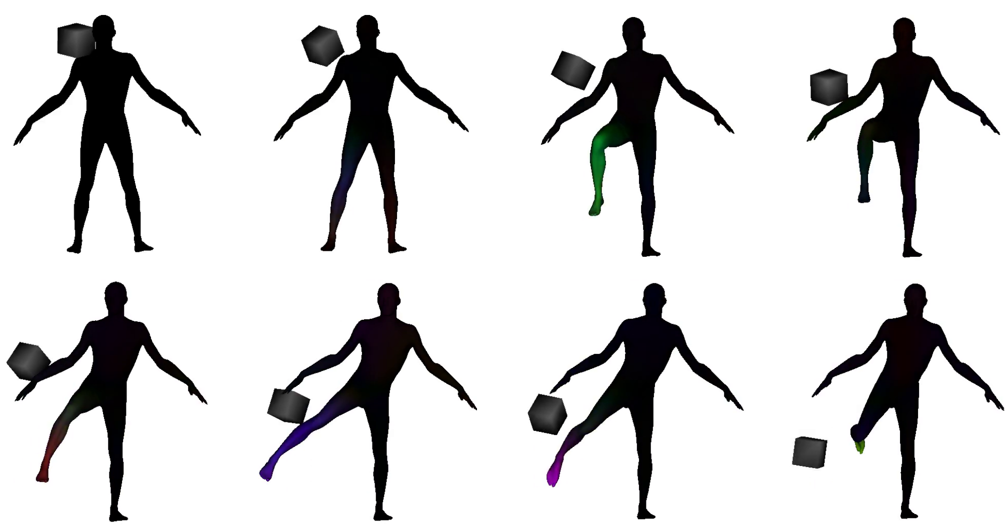

Figure 6: The "tilt" motion sequence interaction with the cube.