

# Simulating Anthropomorphic Upper Body Actions in Virtual Reality using Head and Hand Motion Data

Dustin T. Han\*

Shyam Prathish Sargunam †

Eric D. Ragan‡

Texas A&M University, USA

## ABSTRACT

The use of self avatars in virtual reality (VR) can bring users a stronger sense of presence and produce a more compelling experience by providing additional visual feedback during interactions. Avatars also become increasingly more relevant in VR as they provide a user with an identity for social interactions in multi-user settings. However, with current consumer VR setups that include only a head mounted display and hand controllers, implementation of self avatars are generally limited in the ability to mimic actions performed in the real world. Our work explores the idea of simulating a wide range of upper body motions using motion and positional data from only the head and hand motion data. We present a method to differentiate head and hip motions using information from captured motion data and applying corresponding changes to a virtual avatar. We discuss our approach and initial results.

**Index Terms:** H.5.1 [Information interfaces and presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

## 1 INTRODUCTION

Representational self avatars in virtual environments can provide users with a stronger sense of presence and body ownership allowing for a more immersive experience [4, 1]. Studies also found that it is not visual realism of the avatar that contributes most to body ownership, but rather visuomotor synchrony and sense of agency [5]. With the recent developments in the use of virtual reality as a basis for social training and networking, there is also an increasing need to represent ourselves digitally. We are therefore interested in understanding how we can better track human body motions to create avatars that more closely represent the actions of our real world counterparts. Motion tracking is a widely used technique across various industries to capture body movements of human actors. Unfortunately, several motion capture systems today require the extensive use of markers placed on the body and often require actors to put on full body suits. On the other end of the spectrum, devices like the Kinect that utilize only depth cameras to track motion can be inaccurate without investment in larger tracked spaces and complicated setups using multiple cameras [2]. Moreover, access to large open spaces and full body suits are not readily available to a majority of consumers. To provide a balance between accessibility and accurate motion tracking, we have systems with limited body tracking where only tracking of the head from the head-mounted display and the hands from hand-held controllers are present.

One major limitation of only having head and hand tracking is the inability to differentiate between head and body rotations. Reus et al. [3] proposed a solution to this constraint using a *comfort pose function*. This algorithm uses the assumption that if a user's head

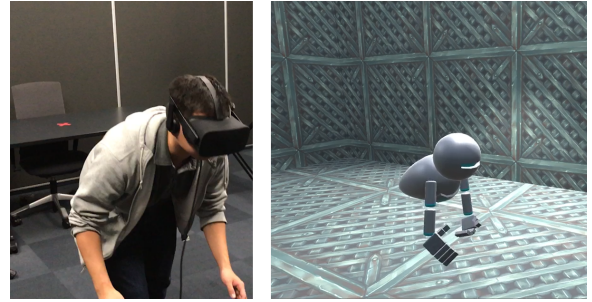


Figure 1: Images demonstrating the *leaning forward* motion simulated by our application. The left shows a user physically performing the motion, and the right shows the simulated virtual avatar.

faces a specific direction for a long duration, their body is most likely facing the same direction as well. Slater and Usoh [4] discuss a more simplified approach that works off of the idea that people often do not look down at their bodies when looking far right or left. This technique reorients the body to match the head only after the head has rotated past 60 degrees.

Although the previously mentioned techniques present solutions to the head and body alignment issue, they do not necessarily provide a synchronous tracking system and instead rely on assumptions of human behavior to give the illusion of correct alignment of the head and body. Relying on user inattention is also not a valid solution for use in multi-user settings where a self-avatar can be seen by other users at any given time. We present an approach to the head and body alignment issue that will be able to determine body and head motions in real time through action detection based off of review of prerecorded data of various upper body actions. By distinguishing different ways the head and hands move during different types of actions, we created a set of activation criteria to enable avatar actions based on different tracker motions.

## 2 METHOD

The goal for our initial tests was to differentiate between four common upper body actions: (1) *leaning forward*, (2) *looking down*, (3) *full body turning*, and (4) *looking left and right*. *Leaning forward* is the motion of bending forward from your hips, while *looking down* is the motion of rotating your head down from the neck pivot. *Full body turning* is the action of reorienting the entire body around the vertical axis, whereas *looking left and right* only includes pivoting the head from the base of the neck.

Our first step was finding characteristics of the actions to create activation criteria for each. To collect a preliminary motion data set, we recorded ourselves performing the actions (Figure 1) using an Optitrack Flex 13 capture system for hand and body tracking.

### 2.1 Leaning Forward and Looking Down

After visually plotting the positional data of the head during the actions, we noticed a main difference between a *lean forward* and *look down* motion was the steepness or difference in tangents along

\*e-mail: dthan@tamu.edu

†e-mail: shyam.prathish@tamu.edu

‡e-mail: eragan@tamu.edu

the arcs of the two different actions. Therefore, the activation criteria for these actions compares the realtime motion to the average tangent for each motion's arc from the test data. To calculate these tangents, we compute a quadratic Bézier curve formed by every set of three consecutive positional data points from the HMD. To apply this approach, activation of avatar motions must happen in real time; for example, the distinction between a *lean forward* or *look down* motion, must be determined within the very first few frames of motion. For this reason, we calculate the average tangents based on the first 45 frames of motion data. In addition to motion tangents, we consider the change in horizontal head distance to differentiate between *leaning forward* and *looking up and down*. This horizontal head distance threshold was also calculated based off of the first 45 frames of recorded test data.

## 2.2 Head Turning and Full Body Turning

To distinguish between *head turning* and full *body turning*, we consider hand movement for the activation criteria for *full body turning*. During a full body turn, the left and right hands of a user generally move in opposite directions. But the hands could move in such opposite directions even when a user is not turning, such as when a user swings both arms while physically walking. To address such cases, our activation criteria for *full body turning* also considers head orientation in addition to opposite hand movements. The virtual gaze direction must differ from the virtual torso direction by at least 60 degrees to activate. Once *full body turning* is confirmed, the virtual torso is rotated around the vertical axis by the angle offset, which in most cases is approximately 60 degrees.

## 2.3 Applying Motion to Avatar

In our approach, the *looking down* and *looking left and right* actions only involve head movements and not torso movement. The virtual head's position and rotation are updated one-to-one based on the tracking data from headset. On the other hand, the torso is rotated around the horizontal axis when *leaning forward* is detected and around the vertical axis when *full body turning* is detected. When *leaning forward* is detected, rotations are continuously applied to the hip joint based on the angle created between the up vector and the vector formed between the current hip and head positions (see Figure 1). Whereas, when *full body turning* is detected, the hip joint is rotated around the vertical axis by the angular difference between the forward directions of the torso and head. These rotations continue to be applied to the hip joint until the system has detected a new action or the user returns to the default standing pose.

Because our method requires checking a set of initial frames of motion to detect an action, the motion is not entirely one-to-one with real world motions for actions that involve rotations from the hip which in this case are the *lean forward* and *full body turning*. This causes a slight delay before the action is detected, which creates a pop in the torso as it attempts to match the current orientation of the real world torso at the time of detection. We addressed this issue in our approach by interpolating the hip from the upright position to the target angle and target position that would orient the torso appropriately from the head's current position. The interpolation creates a drifting effect of the body as it reaches its position. This drifting improves the popping problem, but additional testing is needed to assess noticability.

## 3 PRELIMINARY RESULTS

For preliminary testing, we built a simple avatar and placed it in a virtual environment with mirrors to provide visual feedback of the actions performed (see Figure 2). Testing used an Oculus Rift CV1 display, and the environment and inverse kinematics for the avatar arms were implemented in the Unity engine (5.4.1).

We conducted initial tests to distinguish between the *lean forward* and the *look down* actions. Our system was able to detect the



Figure 2: Avatar motions are tested within a mirrored environment to provide visual feedback of body actions

*lean forward* action with about an 80% consistency rate when testing the system ourselves. We will conduct a more thorough test for success rate in the future as we further develop the system. There were cases where certain orientations of the head while leaning affected the *lean forward* detection. If *lean forward* motion was performed too slow, it may not be detected, so physical speed variation is another factor to consider in future tests.

The activation criteria to distinguish between *fully body turning* and *looking left and right* still needs refinement. The current criteria assumes that the user's arms remain at the side of the torso (as in a neutral standing pose), but our goal is to distinguish between these actions while allowing the head and hands to freely rotate during any action.

## 4 CONCLUSION AND FUTURE WORK

We present a method to detect and simulate upper body actions using minimal tracking data from only the head and hands. Further work is needed to distinguish between a wider range of head and body motions. Collection and analysis of motion data from additional participants will be on going to generate more robust test data and to determine if there are any significant differences from person to person that may pose any issues in action detection. We will continue to study other actions and apply the technique to distinguish them from each other. Furthermore, we are interested in applying our technique to other VR systems and avatars of varying shapes and sizes within a multi-user setting.

## REFERENCES

- [1] M. Gonzalez-Franco, D. Perez-Marcos, B. Spanlang, and M. Slater. The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment. In *2010 IEEE virtual reality conference (VR)*, pages 111–114. IEEE, 2010. 1
- [2] A. Pfister, A. M. West, S. Bronner, and J. A. Noah. Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis. *Journal of medical engineering & technology*, 38(5):274–280, 2014. 1
- [3] V. Reus, M. Mello, L. Nedel, and A. Maciel. Correcting drift, head and body misalignments between virtual and real humans. *SBC Journal on Interactive Systems*, 4(2):55–65, 2013. 1
- [4] M. Slater and M. Usoh. Body centred interaction in immersive virtual environments. *Artificial life and virtual reality*, 1:125–148, 1994. 1
- [5] W. Steptoe, A. Steed, and M. Slater. Human tails: ownership and control of extended humanoid avatars. *IEEE transactions on visualization and computer graphics*, 19(4):583–590, 2013. 1