# A Protocol for Conducting VR Motion Capture Studies

Arjun Talpallikar*
University of Texas at Austin

Andrew Robb†
School of Computing
Clemson University

Jerome McClendon‡
Automotive Engineering
Clemson University

## Abstract

We discuss a simple motion based protocol for conducting virtual reality (VR) user research. Previous methods made data collection difficult, which has reduced the quality and quantity of related research.
In this paper, we propose the use of an inertial tracking system and simple dimensionality reduction algorithms. Our protocol allows for rapidly testing different hypotheses comparatively cheaply. The comparatively low barriers to entry for our protocol makes open and collaborative VR user research more attainable.

**Index Terms:** 1.37 [Computer Graphics]: Virtual Reality—

## 1 Introduction

Virtual and Augmented Reality (VR/AR) technologies are rapidly increasing in popularity. VR platforms, developed by technology giants including Facebook and HTC, are beginning to reach consumers. In 2016, the augmented reality mobile app, Pokemon Go, became the most downloaded app within a week of its launch [6]

Despite this increase in popularity, little work has been carried out that examines VR users and their interaction with those systems. The rate of user studies in VR systems research is low [10], and those studies that are carried out may not follow uniform standards and protocols. General user knowledge in VR research is also lacking. We identify two underlying issues:

- The visual fidelity of VR systems has increased rapidly, which has and will continue to cause significant changes in user experience
- The rate of prior user exposure to VR systems will rapidly change over time.

These changes necessitate continual updates to the body of knowledge concerning VR users. As VR technology continues to develop and approach widespread adoption, the need for such work will only increase.
In particular, users of VR systems are no longer immobile. Head-Mounted Displays and positional tracking allow VR users to move in the virtual environment by moving in the same way in the real environment, instead of merely moving by triggering teleportation. This change necessitates significant research on motion in the virtual environment: researchers continually work to make motion feel more realistic [17].

Despite its importance, research on VR locomotion faces key issues that limit the collection and analysis of new motion data:

- Researchers collect data using complex and expensive systems
- Researchers rely on biomechanical models that can be inaccessible
- Motion data is complex, and as a result, is difficult to understand and use

Thus, in order to further research in VR user-system interaction, we propose a simple, standard protocol for easily and inexpensively conducting novel research on users motion and gait within a virtual environment. Our protocol records subjects walking with an inertial tracking system. The resulting high-dimensional data is visualized using easily implementable machine learning algorithms that are commonly used for managing data in high dimensional space.

In this paper, we first discuss and analyze previous, related work, in VR research, motion capture research, and machine learning research. Next, we describe our proposed method for VR research, and provide comparisons to previous methods used. We then describe a prototype experiment using the new protocol. Finally, we evaluate the results of our work, and discuss potential future work.

## 2 Related Works

Because we propose a new research protocol, our discussion of related work is centered around similar research protocols used by researchers interested in motion in virtual environments. However, for additional context, we have also included a brief discussion of those works' findings as well.

### 2.1 Virtual Reality User Research

Several significant bodies of VR research exist currently. Current areas of interest include interaction with other users within a virtual environment [5], digital user avatars [18], and user emotions [12], among others. Because our protocol is one for understanding user motion in the virtual space, we will focus on locomotion.

#### 2.1.1 Locomotion

Motion capture data has been analyzed using gait parameter analysis to determine the effect of a Virtual Environment (VE) on user movement [11]. This work provides a foundation for our own analysis, which uses machine learning instead of gait parameter analysis. [11] used a Vicon camera-based motion capture system, with participants walking freely between targets, in the real environment and while wearing a Head-Mounted Display (HMD), and showed that the measured gait parameters differed significantly between the real world and the virtual environment.

Domain specific research has also been conducted using variations of these methods. In particular, the connection between walking and presence was analyzed in one of the field's seminal papers, [17], which assessed a novel model for walking in a virtual environment. Presence refers to "the sensation of being in the virtual world," [15] and can often be used as a measure of the virtual environment's effectiveness. [17] found that realistic-feeling motion and transportation in a virtual environment were important for maintaining presence, which indicates that understanding motion in the virtual environment is crucial for developing effective VR systems.

*e-mail: arjuntalpallikar@utexas.edu
†e-mail:arobb@clemson.edu
‡e-mail:jmcclen@g.clemson.edu

Similar work was done in [13], where gait parameters and other variables were assessed to determine the effect on presence. [13] extends on previous work by the authors, who previously examined the effect of stressful environments on gait parameters, again using a Vicon camera system for motion capture and an HMD. [8] compared the effects of an isometric mapping and a non nonisometric mapping of motion from the real world to the virtual environment on gait parameters, using a GAITrite walkway system and an HMD.

## 2.2 Motion Capture and Gait Analysis

While locomotion analysis in the sphere of VR research has largely been limited to gait parameter analysis, machine learning techniques have been applied to gait analysis in other fields. Such projects are generally focused on gesture recognition [22] [3], a classification problem, which seeks to break data into pre-defined groups. This has been done both with a simpler feed-forward neural network [22] and with a deep-learning methodology that uses a temporal encoding layer to better generalize results [3].

Because these methodologies require large datasets, other projects have worked on data augmentation, supplementing real motion capture data with realistic computer-generated synthetic data [4].

Just as in [11], [8], and [14], these projects typically use expensive custom infrared multi-camera systems, or rely on data from the CMU Motion Capture Database, which was collected using one such system ( [3], [4], [11]). Those that instead use intertial systems refrain from large translatory motion (such as walking) and focus on small hand movement [22].

### 2.2.1 Motion Capture Visualization

Just as our improvements to data collection build on previous contributions, so too do our improvements in motion capture visualization. Principal Component Analysis, a linear dimensionality technique, has been shown to successfully retain keyframes of motion capture data while reducing unwanted noise [23]. Visualization methods using colored shapes to represent motion have also been shown to be successful for visualizing individual recordings of human motion [19]. Likewise, [20] also showed that abstract visualizations, in this case inspired by slit camera photography, can be successful for visualizing human motion data.

## 2.3 Limitations of Existing Work

### 2.3.1 Data Collection

As we have mentioned earlier, motion capture data collection has generally been done through multi-camera systems. These systems are expensive and immobile - the North American authorized retailer for some commonly used systems indicates that prices vary from $8,000 to $250,000 [1]. The high-end Vicon systems used in several of the cited papers [13] [11] needs to be custom ordered, and generally costs $100,000 or more [2]. These high costs create a significant barrier to entry for motion capture recording.

The lack of a standard form of walking trial is also a limiting factor on collaboration and openness in motion capture studies in VR research. When each experimental protocol differs greatly, comparing and replicating results becomes more difficult.

### 2.3.2 Data Visualization

Previous work in motion capture visualization was successful in visualizing a single movement, but did not intuitively extend to visualizing many different sets of data, limiting its viability in the area of VR research.

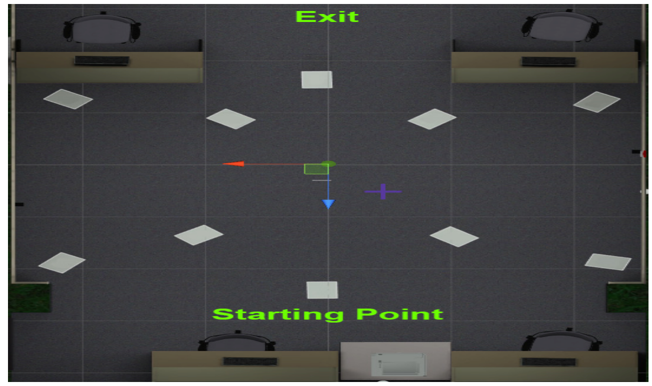


Figure 1: VR Environment

# 3 Protocol Concept

The goal our proposed protocol is to overcome these aforementioned issues. Our work focuses on two parts: improvements on data collection and improvements on data visualization.

## 3.1 Data Collection

Data is collected by recruiting subjects to complete walking trials. Subjects complete an initial survey, if necessary for the study being performed. Subjects then wear an inertial tracking suit, as well as a hip-mounted HTC VIVE Tracker. Participants are tracked using Valve's Lighthouse basestation tracking system. Researchers should create a high-fidelity virtual environment to mirror the physical environment. Subjects complete two sets of walking trials, one in the physical environment and one in the virtual environment while wearing a Head-Mounted Display. Each trial involves walking between targets several times. Participants are instructed to start and stop via recorded auditory alert. See figure 1 for our virtual environment and layout of targets.

Data is streamed from the inertial tracking suit and the positional sensor to Unity. From Unity, data is recorded in a csv, and then imported to Python 3 using the Jupyter Notebook interface.

## 3.2 Data Visualization

Motion capture data is difficult to explore because of a combination of a high number of observations (data is likely recorded at 60 hz or more) and a high number of features (rotations are recorded for several body parts).

Thus, the key to visualizing this data is dimensionality reduction, as was demonstrated by [23]. Data can be visualized using Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE), two dimensionality reduction and visualization techniques that are described below.

### 3.2.1 PCA

Briefly, PCA seeks to display high-dimensional data in a lower-dimensional space. The transformation selects the linear combinations of features that best explain the variation in the dataset. A brief conceptual explanation is included below. A more thorough and mathematically rigorous explanation is provided in [16].

- Standardize data
- Calculate the Eigenvectors and Eigenvalues from the covariance matrix
- Select the eigenvectors of the largest eigenvalues
- Project the original data to the new subspace using a projection matrix constructed from the selected eigenvectors

PCA is easy to implement, and can be used to explore data without destroying its structure.

#### 3.2.2 t-SNE

Like PCA, t-SNE seeks to display high-dimensional data in a lower-dimensional space. However, t-SNE does not retain the structure of the data, but instead produces useful exploratory visualizations that attempt to keep neighboring points together. As before, a brief conceptual description is included, while a more thorough explanation is provided in [21].

- Determine the distance between points, and then model that distance as a probability model
- Map this probability model in a lower dimensional space
- Minimize the difference between the two probability models by minimizing the sum of the Kullback-Leibler divergences

t-SNE is easy accessible, and can be used to produce high-quality visualizations for exploratory analysis.

## 4 Prototype Experiment

In order to test our proposed experimental protocol, we conducted a simple prototype experiment that is described below. Because of the time limitations of an 8 Week Summer REU program, the scale of the prototype experiment is limited.

### 4.1 Overview

Our experiment was designed to replicate the foundational study described in [11], which used a traditional camera based system and gait parameter analysis to show that walking in VR differed from walking in the real world. Our hypothesis, based on the previously done work, was that walking in VR would differ significantly from walking in the real world.

### 4.2 Methods

For this experiment, we implemented the protocol described in this paper. Specifically, the experiment took place in a fully-tracked 5 meter x 4 meter room. As is described above, participants' motion was tracked using the Perception Neuron inertial tracking suit, along with the HTC VIVE Hip tracker and the Valve Lighthouse tracking system. Participants first completed a 3 part demographic survey. They then completed two sets of 40 trials, one while using the HTC VIVE HMD and one in the real world, walking between targets laid out to match figure 1. Targets were labeled with two sets of letters that were selected to reduce confusion: [JYXEFUOKMR] and [CALWSIGHWP]. After each trial, participants labeled from memory a blank map with the positions of the letters, and completed brief post-trial surveys. Participants were incentivized with $15 gift cards from a major online retail store.

Body part position, body part rotation, participant position, and trial information data was streamed from the Perception Neuron tracking system to Unity and recorded as CSV data. This data was refined in Python 3 Jupyter notebooks. Because the distribution of each feature differed significantly, a normalization procedure was essential to process the data. Without such a procedure, some features would significantly dominate others in procedures involving distance metrics. To select a procedure, we tested three statistical procedures. Max normalization, standardization, and standardization to the range of [0,1] were each applied to individual real world trials, individual virtual world trials, individual combined trials, and trials across multiple participants. Max normalization was selected, because its visualizations included the least distortion. The data was then analyzed using two clustering methods briefly described below. If clustering analysis of an individual's VR and real world data successfully separated the VR data from the real world data, that would be evidence to support our hypothesis.

#### 4.2.1 k-Means

As before, a brief description of the algorithm is provided below. A more thorough explanation is available in [9]. k-Means is a commonly used clustering technique that seeks to categorize unlabeled data into groups. After specifying that you are looking for k clusters, randomly place k centroids. Then, repeat the following steps until the algorithm converges or a specified stopping point is reached:

- Assign each data point to its closest centroid
- Move the centroid to the average of its cluster

Eventually, the centroids will stop shifting and each point will be labeled.

#### 4.2.2 Growing Neural Gas

Growing Neural Gas is a variant of the Self Organizing Map that learns the topology of a dataset. Again, a brief description is provided here - please refer to [7] for a more thorough explanation. We first initialize two random nodes and select a value N. We then repeat the following steps until the algorithm converges:

- For each point in the dataset, select the closest node
- Bring that node closer to the point
- Bring nodes connected to the selected node closer to the point
- Every N iterations, add a new node.
- If a certain condition is met (such as a maximum distance between nodes), connect the nodes.
- If that condition no longer holds, destroy the edge between those nodes.

Eventually, nodes will stop being added. The result is a set of graphs that model the original dataset. The number of graphs returned is the number of clusters the algorithm has found.

### 4.3 Participants

10 participants were selected. 3 participants had 20 or more hours of VR locomotion, and 7 participants had less than 1 hour of VR locomotion. Participants included other summer REU students, undergraduate students, and graduate students recruited largely from the School of Computing.

## 5 Evaluation of Results

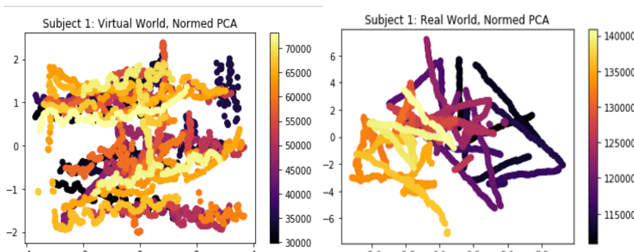


Figure 2: PCA, Subject 1 Trials

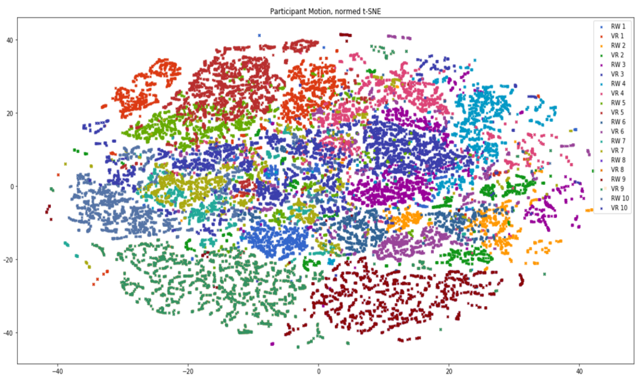


Figure 3: t-SNE, All Data

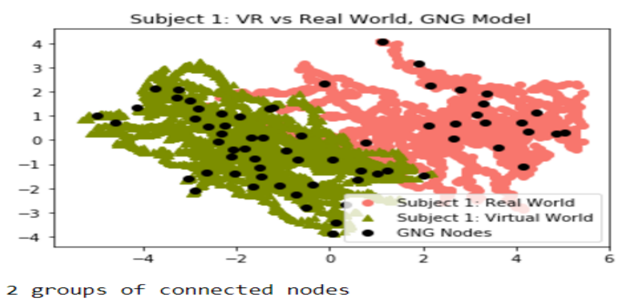


Figure 4: GNG, Subject 1

### 5.1 Prototype Experiment Results

Exploratory data analysis was conducted with PCA on individuals' data. VR data appeared to have significantly more variation than the PCA data (figure 2). Additional visualization was conducted using t-SNE, which showed distinct separation between each dataset. k-Means analysis failed to produce useful clusters and was abandoned. GNG analysis was conducted on each individuals' data. 50% of participants had data converge with two clusters (figure 4). 30% of participants had data that would likely converge with two clusters with correct model parameters, while 20% of participants had data that would not successfully be labeled. This successful clustering is consistent with our hypothesis, and with the previous work done.

### 5.2 Evaluation of Protocol

The inertial tracking suit was significantly less expensive than the camera-based counterparts, costing only $1500. Despite this, useful data was collected, and a successful experiment was run.
Visualization techniques used also produced useful results, as seen in figure 2 and figure 3. These results indicate that the proposed protocol is an inexpensive and simple way to conduct motion capture trials in VR research.

### Acknowledgments

The authors wish to thank Dr. Oyewole Oyekoya for his help in facilitating the REU experience. The authors also wish to thank the SAWIAGOS group for their help and support. This work was supported in part by NSF ACI Award 1359223.

### References

[1] Metamotion.com.
[2] vicon.com.
[3] J. Bütepage, M. J. Black, D. Kragic, and H. Kjellström. Deep representation learning for human motion prediction and classification. *CoRR*, abs/1702.07486, 2017.
[4] C. C. Charalambous and A. A. Bharath. A data augmentation methodology for training machine/deep learning gait recognition algorithms. *CoRR*, abs/1610.07570, 2016.
[5] W. Chen, N. Ladeveze, C. Clavel, D. Mestre, and P. Bourdot. User cohabitation in multi-stereoscopic immersive virtual environment for individual navigation tasks. In *2015 IEEE Virtual Reality (VR)*, pp. 47–54, March 2015. doi: 10.1109/VR.2015.7223323
[6] A. M. Clark and M. T. G. Clark. Pokmon go and research. *International Journal of Qualitative Methods*, 15(1), 2016. doi: 10.1177/1609406916667765
[7] J. Holmström. Growing neural gas–experiments with gng–gng with utility and supervised gng. Master's thesis, Uppsala University, 2002.
[8] O. Janeh, E. Langbehn, F. Steinicke, G. Bruder, A. Gulberti, and M. Poetter-Nerger. Walking in virtual reality: Effects of manipulated visual self-motion on walking biomechanics. *ACM Trans. Appl. Percept.*, 14(2):12:1–12:15, Jan. 2017. doi: 10.1145/3022731
[9] M. Jeevan. Possible the simplest way to explain k-means algorithm, May 2015.
[10] S. J. J. Kim. A user study trends in augmented reality and virtual reality research: A qualitative study with the past three years of the ismar and ieee vr conference papers. In *2012 International Symposium on Ubiquitous Virtual Reality*, pp. 1–5, Aug 2012. doi: 10.1109/ISUVR.2012.17
[11] B. Mohler, J. Campos, M. Weyel, and H. Blthoff. Gait parameters while walking in a head-mounted display virtual environment and the real world. In *13th Eurographics Symposium on Virtual Environments and 10th Immersive Projection Technology Workshop*, 2007. doi: 10.2312/PE/VE2007Short/085-088
[12] A. Naz, R. Kopper, R. P. McMahan, and M. Nadin. Emotional qualities of vr space. In *2017 IEEE Virtual Reality (VR)*, pp. 3–11, March 2017. doi: 10.1109/VR.2017.7892225
[13] L. Phillips, V. Interrante, M. Kaeding, B. Ries, and L. Anderson. Correlations between physiological response, gait, personality, and presence in immersive virtual environments. *Presence: Teleoper. Virtual Environ.*, 21(2):119–141, apr 2012.
[14] L. Phillips, B. Ries, M. Kaeding, and V. Interrante. Gait parameters in stressful virtual environments. 2010.
[15] M. J. Schuemie, P. V. D. Straaten, M. Krijn, and C. A. V. D. Mast. Research on presence in virtual reality: A survey. *CyberPschology & Behavior*, 4(2), 2001.
[16] J. Shlens. A tutorial on principal component analysis. *CoRR*, abs/1404.1100, 2014.
[17] M. Slater, M. Usoh, and A. Steed. Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 2(3):201–219, Sept. 1995. doi: 10.1145/210079.210084
[18] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 67–76, March 2016. doi: 10.1109/VR.2016.7504689
[19] H. Sunada, K. Yokoyama, T. Hirata, T. Matsukawa, and H. Takada. Visualization system for the information measured by motion capture system. In *2016 11th International Conference on Computer Science Education (ICCSE)*, pp. 21–26, Aug 2016. doi: 10.1109/ICCSE.2016.7581548
[20] Y. Tashiro and T. Saitoh. A study on motion visualization system using motion capture data. In *17th International Conference on Artificial Reality and Telexistence (ICAT 2007)*, pp. 314–315, Nov 2007. doi: 10.1109/ICAT.2007.53
[21] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9, 2008.
[22] D. Xu. A neural network approach for hand gesture recognition in virtual reality driving training system of spg. In *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, pp. 519–522, 2006. doi: 10.1109/ICPR.2006.109
[23] Y. Yang, L. Zeng, and H. Leung. Keyframe extraction from motion capture data for visualization. In *2016 International Conference on Virtual Reality and Visualization (ICVRV)*, pp. 154–157, Sept 2016. doi: 10.1109/ICVRV.2016.33