Predicting User Presence in Virtual Reality Environments through Behavioral Data

Alex Dolwick
Social and Behavioral Research Branch
National Human Genome Research Institute
Bethesda, Maryland, USA
dolwick.a@gmail.com

***Abstract*—Presence, the psychological concept describing the sense of existing within a place, is critical for the external validity of behavioral research conducted using virtual reality (VR) tools. Participants using VR in behavioral studies often self-report their presence in the virtual environment after the fact. The present research aims to establish a new method for measuring presence in VR by validating it against self-report presence data. Specifically, this project aims to predict presence from nonverbal behavioral data collected automatically by virtual reality headsets. This paper describes a tool to conduct this analysis, which consists of three steps: The tool (1) imputes values for missing data using k-nearest neighbor (KNN) imputation, (2) reduces the number of features from the original high-dimensional dataset using Uniform Manifold Approximation and Projection (UMAP) and/or Principal Component Analysis (PCA), and (3) predicts presence from the reduced data using support vector regression (SVR) and/or random forest regression (RFR). Additionally, this paper describes the results of the tool on sample data. Implementation of this tool among behavioral researchers may decrease reliance of future studies on self-report data.**

***Index Terms*—virtual reality, supervised learning, behavioral research**

I. Introduction

A. Virtual Reality in Behavioral Science

Virtual reality (VR) tools are becoming increasingly common in social and behavioral science settings. Behavioral researchers have capitalized on the multiple advantages that studies in virtual environments have over studies conducted via more traditional media. Most notable among these advantages is the way in which VR tools solve the dilemma of experimental control vs. mundane realism [1]. Experimental control involves the ability of researchers to manipulate experimental stimuli so that each participant experiences the research stimuli in the same way as the other participants in her condition; studies with high experimental control tend to have high internal validity because they reduce potential noise and confounds. Mundane realism describes the extent to which stimuli in a study approximate stimuli that participants experience in their daily lives; studies with high mundane realism tend to have high external validity because they allow participants to behave similarly in response to experimental stimuli as they would to corresponding real-life stimuli [1].

The sample data used to test this tool and described in this paper were collected with the help of funding from the National Human Genome Research Institute.

Two types of environments are traditionally most common as settings for behavioral science studies: the laboratory and the field [1]. Each of these settings tends to excel at achieving one form of validity while sacrificing the other. Field experiments are almost always high in mundane realism, since they take place in real situations as opposed to artificially constructed ones. Field experiments thus achieve high external validity — participants in these experiments act naturally and results from these studies can often be applied to similar real-world situations. However, field experiments often contain variables outside of the control of the researchers and thus internal validity in these studies is often low [1]. On the other hand, lab studies are almost always high in experimental control, since the researchers can create rigid, structured procedures to reduce potential noise and confounds as much as possible. But these lab studies often involve stimuli that are not similar to those in participants' normal lives, and participants are well-aware that they are taking place in research and are in a lab. Lab studies often fail to maintain mundane realism or maximize external validity because of this [1].

VR environments can be seen as the best of both worlds in the experimental control vs. mundane realism dilemma [1]. Virtual environments are highly controlled; in fact, they can be exactly the same from participant to participant, achieving an even greater measure of consistency than lab-based studies. Furthermore, virtual environments allow participants to experience stimuli that approximate those seen in the real world to a much greater extent than anything that can be achieved in a lab setting [1]. VR thus holds much promise for behavioral researchers as it gives them greater control over realistic environments than is possible using traditional methods. The concept that allows for this happy medium is known as presence.

B. Presence in Virtual Environments

Presence, defined in the context of VR, is "the illusion of being in the virtual world" [2]. Imagine you put on a VR headset and find yourself in an elevator. The door opens and you realize you are 50 stories above the ground. In front of you is a small wooden plank suspended above the abyss. You are directed to walk out onto the plank. Then you are told to jump. How do you expect you would feel? Although many people expect that they would be unfazed because they know they are only experiencing a simulation, they quickly find that they respond to the precipice as though it were real. Many refuse to even take one step forward onto the plank. Few are able to take the leap. The immersive nature of the virtual environment elicits realistic responses to manufactured stimuli because these stimuli are perceived instinctively to be real despite the cognitive awareness that they are not [2]. This fact enables behavioral scientists to conduct highly-controlled studies and record realistic responses to stimuli.

The challenge that arises with presence in VR is how to measure it accurately. While VR environments do elicit high levels of presence in users, the sense of being there can differ from environment to environment and from user to user. Thus is it important to measure presence in behavioral studies using VR to ensure that the behaviors of all participants represent their response to stimuli they perceive as realistic. The most common measurement tools of presence in VR users are self-report questionnaires. After a participant exits the virtual environment, they are asked questions about the extent to which they felt present in the VR world. Self-report questionnaires, while exceedingly common in behavioral research, are vulnerable to participant biases [3]. Additionally, asking participants questions about their sense of presence in the VR world after they exit the VR environment may not lead to accurate responses since the participants are basing their answers off memory instead of current experience. Several methods have been proposed to resolve the problems inherent in self-report. Some studies have developed methods to ask presence questionnaires within the VR environment instead of after the participant's time in VR [4], but this does not resolve the problem of bias. Other researchers have developed physiological measures of presence in VR [5], but these may not work with non-stressful VR stimuli. The most promising research to date has utilized fMRI to understand brain activation correlating with presence in VR [6], but future research still needs to be done to understand ways in which presence can be directly measured without the need for self-report data.

C. The Current Research

The present study aims to test the use of nonverbal behavioral data automatically collected by the VR headset as a potential measure of VR user presence. VR systems must collect continuous data on the position and rotation of the headset in order to update the portion of the virtual environment in view for the user. These data are collected along six degrees of freedom (6DOF): the user's position is measured along the x-, y-, and z-axes, and the rotation of the user's head is measured along the axes of yaw, pitch, and roll (Fig. 1). These 6DOF data are collected continuously and make up a detailed log of participant movement behavior within the virtual environment.

The first step in understanding whether 6DOF data can be applied as a measure of presence for VR users is to validate it against previous self-report measures of presence. While these measures have their weaknesses, their data represents the scientific community's best understanding of the abstract psychological concept of presence. Past research has validated non-self-report measures of presence by establishing correlations between data from novel measures and data from self-report measures [6]. If patterns in 6DOF behavioral data can be indicators of variance in cognition in virtual environments, it is important to establish that these patterns represent the concept that we know of as presence.

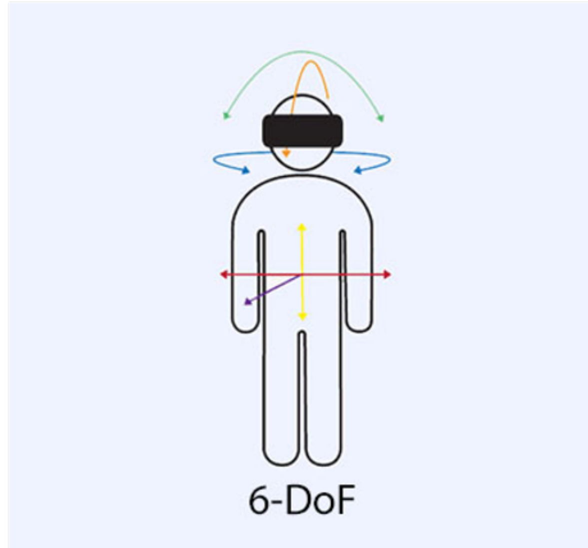

Fig. 1. The six degrees of freedom along which user nonverbal behavior is automatically measured by virtual reality (VR) systems.

This paper describes testing of a tool designed to use 6DOF behavioral data from participants in a VR environment to predict their self-reported presence following VR use. Features (the 6DOF data) will be preprocessed and then input into supervised learning algorithms to predict the presence outcome.

II. Methods

A. Data Collection

Data used as a sample for this tool were originally collected for use in the Health Communication and Behavior Unit (led by Dr. Susan Persky) in the National Human Genome Research Institute. Participants (N=232) entered a VR Buffet environment (Fig. 2) and were instructed to make a plate of food by clicking their controller to select items to put on their plate. 6DOF data were collected at 0.5-second intervals.

Participants were able to spend as much time as they wanted in the VR Buffet. Data are organized so that a row of the data file represents one participant and a column of the data file represents one degree of freedom at one 0.5-second time point. The data file extends rightward until the last 0.5-second time point of VR use for the participant who spent the longest in the VR Buffet. (This time point happens to be 462.5 seconds.) Participants who did not spend as much time in the VR Buffet have missing values for each time point after the end of their VR use. For example, if participant A spent 232 seconds in the VR Buffet, each column recording a degree of freedom for

time points from 232.5 seconds onward has a missing value for that participant.

The outcome variable for self-reported user presence was obtained via a 5-item Likert-type questionnaire with available responses ranging from ”1–not at all” to ”5–extremely.” Items on this questionnaire were as follows:

- To what extent did you feel involved in the virtual world?
- To what extent did you feel like you were inside the virtual world?
- To what extent did you feel surrounded by the virtual world?
- To what extent did it feel like you visited another place?
- How much did the virtual world seem like the real world?

Responses to all five items were averaged so that the final response variable of presence ranged from 1 to 5. Sixteen participants who did not take the presence questionnaire were removed from analysis, resulting in a final analytic sample of 216.

B. Tool Design

The goal of the present tool is to perform supervised learning in order to predict an outcome variable (i.e., presence, in the sample data) from a number of feature columns (i.e., 6DOF output, in the sample data). A requirement for this tool to work is that raw data be organized in ”tidy” format— that is, each observation (e.g., participant) has its own row, and each feature has its own column. This tool consists of three steps:

- Imputation of values for missing data.
- Dimensionality reduction of features.
- Supervised learning to predict outcome variable.

All code for this tool and the sample data used to run it can be found on GitHub (https://github.com/adolwick/BIOF509-Project).

The tool is set up to run imputation of missing values using any imputer, but for the sample data, KNN Imputation was implemented using scikit-learn. Before imputation, data were scaled using scikit-learn's MinMaxScaler. Data were separated by degree of freedom before scaling since the 6DOF have different ranges of possible values (due to the fact that the positional 3DOF are measured in meters and the rotational 3DOF are measured in degrees). After scaling and imputation, the number of unique values in each column was measured. Since the time points at which participants exited the VR environment followed a roughly normal distribution, columns describing time points past 300 seconds or so had very few non-missing values (Fig. 3). Thus, the imputed values for these columns contained low variance. Such columns are not useful for analysis, so a minimum number of unique values (for the sample data, 30) was established. All columns failing to meet that criterium were removed from the dataset.

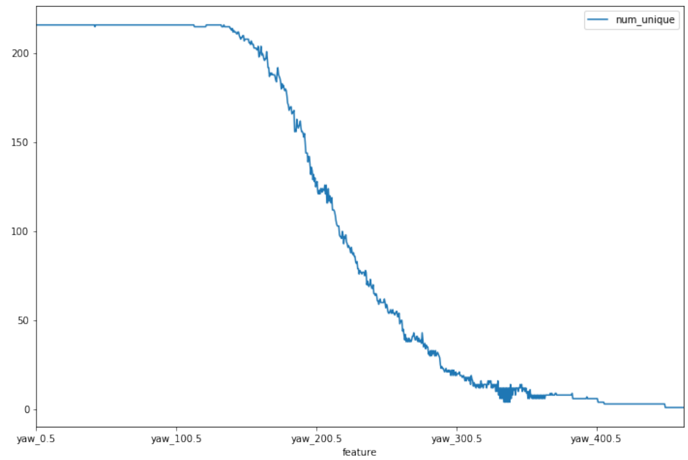

Fig. 3. Graph of unique values per feature in the data file for yaw values after scaling and imputation.

Dimensionality reduction in this tool can be conducted using either Uniform Manifold Approximation and Projection (UMAP) or Principal Component Analysis (PCA). For either algorithm, the tool allows for customization of important paramters (e.g., number of components). The tool also enables visualization of PCA or UMAP results to give researchers feedback on which parameters work well given their data (Fig. 4 and 5). After UMAP or PCA, the output data is scaled so it can be used for future analyses. In the case of the sample data, a 20-component PCA was decided on for use in supervised learning algorithms to predict presence. These 20 components explained 64 percent of the variance in the original dataset (Fig. 4).

Supervised learning can be conducted via this tool using support vector regression (SVR) and/or random forest regression (RFR) in scikit-learn. Parameters of both models are customizable: different kernels can be used for SVR, and a grid search outputs the ideal parameter values for RFR out of a list of candidate values input by the researcher. Any scoring methods available for regression models in scikit-learn can be used to evaluate SVR or RFR models. Feature importances in the RFR model can be visualized. For the sample data, the scaled PCA data was imported, and both models were tested on their ability to predict VR user presence from the features.

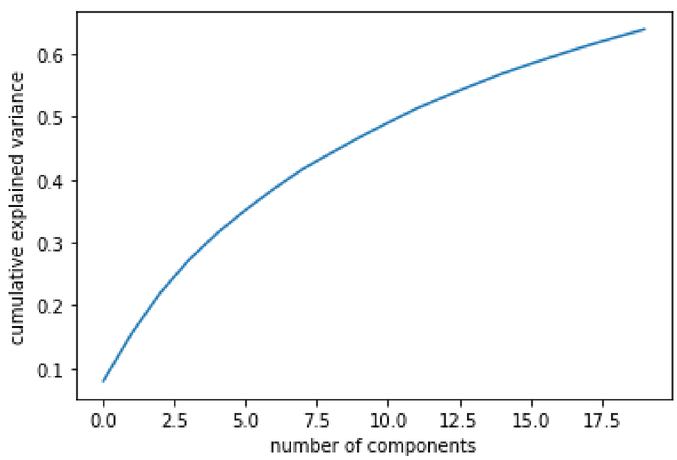

Fig. 4. Proportion of variance explained with each successive component in a 20-component PCA.

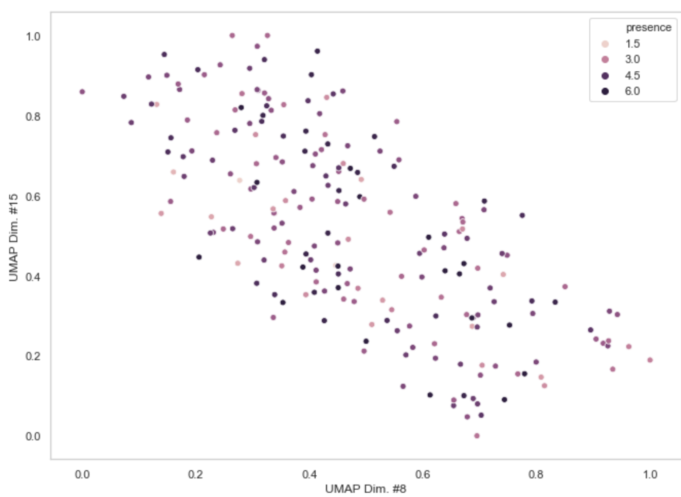

Fig. 5. Clustering of data on two dimensions of a 20-component UMAP.

One RFR model was used, with max depth = 2 and number of estimators = 80. These values were settled upon after trial and error with inputting different ranges of values into the grid search. Two SVR models were used, one with an RBF kernel and one with a linear kernel. 5-fold cross-validation was used in all instances. Models were compared using root mean squared error (RMSE) and R-squared.

III. Results

Model comparison results are visualized in Table 1.

TABLE I
MODEL SCORES

Model	**RMSE**	**R-squared**[a]
Random Forest	0.75	-0.11
SVR (kernel=rbf)	0.76	-0.18
SVR (kernel=linear)	0.79	-0.27

[a]Negative R-squared values indicate poor model fit.

None of the models performed well on the data, although the RFR model performed slightly better than the SVRs on both metrics. The RFR model's RMSE value of 0.75 indicates that, with the presence scale having a range of four (from 1 to 5), the model erred by about three-quarters of a point on average. R-squared measures goodness of fit of the model in this context, and negative values are possible, but demonstrate that all three models fit poorly to the data.

IV. Discussion

None of the supervised learning algorithms used in the present analysis predicted presence well. However, this does not mean that it is not possible to predict VR user presence using 6DOF data. There are several steps to take from here to explore this tool to the fullest in order to understand whether presence in VR can be predicted from nonverbal behavioral data.

The raw data is one element of this analytic process that can be improved. Currently, the sample dataset for this tool includes 6DOF data from one study. The Health Communication and Behavior Unit has completed four VR Buffet studies in which 6DOF data were collected. One future step to take involves combining these datasets and measuring presence across all four studies. This may resolve a problem present at the PCA step of the pipeline: in the current analysis, 20 principal components explained 64 percent of the data. The number of participants whose data were collected for this analysis was 216, and a benchmark for avoiding overfitting is ten observations per feature. A benchmark for retaining sufficient variance in the data is 85-95 percent. In the present analysis, retaining 85 percent of the variance required 56 components and retaining 95 percent required 113 components. With data from four studies, perhaps both benchmarks can be met for future analyses.

At the imputation level, different criteria for sufficient within-feature variance should be tested to ensure that all useful data is retained and no data with insufficient unique values is kept. This can be done in conjunction with establishing a better understanding of feature importances in PCA.

At the feature selection and dimensionality reduction step, the author plans to better understand which original features are important for which components. The tool has functionality to examine this question at present, but it has not been fully explored. Understanding patterns in the importance of original features for each principal component may help elucidate directions for future analyses. Similarly, bolstering the UMAP visualization functionality of the tool will allow for graphing 3 components at once on a 3D graph, and iterative visualization could allow for quickly examining many combinations of dimensions on a multi-dimensional UMAP. In the current analysis, PCA was utilized for analysis over UMAP since UMAP clusters did not appear to develop well. Further experimentation with UMAP needs to occur to see under which conditions it can cluster the data. It is also possible that future analyses will make use of more manual feature selection processes (such as inputting descriptive statistics of 6DOF over time as features).

Perhaps the greatest limitation to this tool for use with 6DOF data from VR environments is that the algorithms used in the tool are not capable of examining sequential patterns in time-series data. For this reason, the first priority for future directions on this tool will be to implement a recurrent neural network (RNN). Using RNN will enable analysis of time-series patterns, which are how behavior works. Understanding behavior over time should be a much more effective approach than using PCA and UMAP to create linear combinations of features as if they were independent. Future work with RNN and other algorithms will provide more insight into whether nonverbal behavioral data can be used to predict user presence in VR environments.

Acknowledgment

The author thanks Dr. Susan Persky for her mentorship and for allowing the sample data to be used for this project. Additionally, the author thanks James Anibal and Yein Christina Park for their teaching, feedback, and assistance in development of this tool.

References

[1] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, “Immersive virtual environment technology as a methodological tool for social psychology”, Psychological Inquiry, vol. 13, no. 2., pp. 103–124, 2002.

[2] M. Slater, “Immersion and the illusion of presence in virtual reality,” British Journal of Psychology, vol. 109, pp. 431–433, 2018.

[3] T. F. van de Mortel, “Faking it: Social desirability response bias in self-report research,” Australian Journal of Advanced Nursing, vol. 25, no. 4, pp. 40–48, 2008.

[4] V. Schwind, P. Knierim, N. Haas, and N. Henze, “Using presence questionnaires in virtual reality,” CHI 2019, May 2019.

[5] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, Jr., “Physiological measures of presence in stressful virtual environments,” in Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH ’02), 2002.

[6] M. Clemente, B. Rey, A. Rodriguez-Pujadas, A. Barros-Loscertales, R. M. Banos, C. Botella, ..., and C. Avila, “An fMRI study to analyze neural correlates of presence during virtual reality experiences,” Interacting with Computers, vol. 26, no. 3, pp. 269–284, 2013.