A Characteristic-Oriented User Evaluation of Immersive Virtual Reality Comparison Visualizations

Jadrian Miles
`jadrian@cs.brown.edu`
Department of Computer Science, Brown University, Providence, RI, USA

March 7, 2008

Abstract

We present a parameterized description of a subspace of the space of designs for immersive virtual reality comparison visualizations, as well as results of a user evaluation of the utility of these different parameters. Our work makes three contributions: a specific definition of comparison visualizations and their place in the scientific data cycle, advice derived from the user study for the design of such visualizations, and a first step for future formal analysis of this topic.

1 Introduction

This paper describes the design of visualizations for the comparison of scientific datasets in virtual reality. Such visualizations are important because comparison plays a central role in the scientific method. A scientist may investigate the effect of one independent variable by changing it while holding all others fixed and then comparing the results [16]. If differences in the dependent variables can be meaningfully quantified, numerical methods may suffice to detect the important effects. This is often not the case with modern high-dimensional data spaces, however, especially under an exploratory regime in which the scientist does not even know what differences might be important. Fundamental work in many newly data-rich fields involves defining and justifying new difference measures [2, 19, 14]. While this work is ongoing, practitioners may not have any reliable difference measure to use. In the absence of numerical guidance, then, one might instead combine visualizations of two related datasets together, and then visually inspect for meaningful differences. The human visual system's efficient and flexible feature-detection system makes this an appealing alternative [10].

1.1 Comparison Visualizations

A *comparison visualization* (henceforth `CV`) is a hybrid visualization of dependent variables intended to support comparison between them. We define `CV`s in the case in which values of some dependent variables have been derived in a single experiment for two different values of a set of independent variables. Such visualizations are commonplace: plotting two data series from an experiment on the same pair of axes is an example. The choices available in arranging two two-dimensional plots so that they can be compared are relatively limited—they may be superimposed or placed side-by-side, or a few more exotic treatments may be applied.

Modern scientific datasets, however, may be larger, higher-dimensional, and multi-variate. These datasets often benefit from visualization in immersive virtual reality (IVR) [24]. When dealing with virtual three-dimensional space, the design choices to be made when combining two visualizations into one are more numerous and may involve tradeoffs and context dependence. There has been significant work in the visualization community devoted to developing three-dimensional comparison visualizations for specific applications, in particular fluid flow [25, 18]. Our work instead focuses on generic operations to generate comparison visualizations for any context.

To be more specific, we use the term *visualization* to refer to a procedure for converting data into a visual rendering. A `CV` is an abstract operation that combines renderings generated by one client visualization of two related datasets into a single composite rendering. Similar operations have been studied in the past in the form of multi-view visualiza-

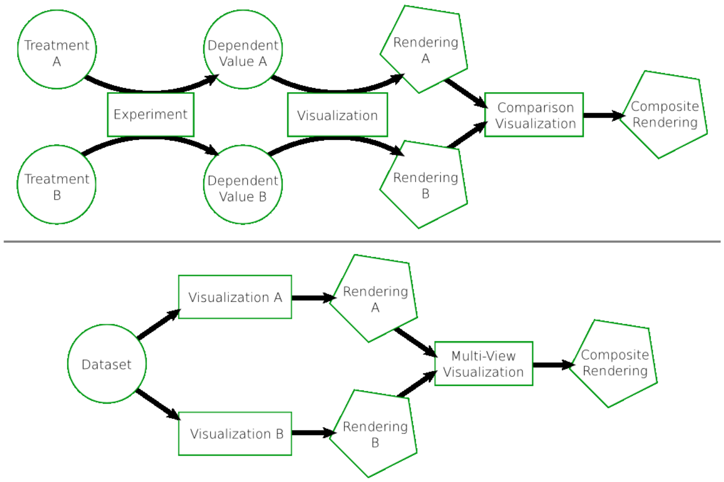


Figure 1: Data flows in comparison visualizations (top) and multi-view visualizations (bottom). A comparison visualization combines two renderings made by the same visualization technique into one composite rendering specifically for the purpose of comparing them.

tions, in which renderings of the same dataset by two different client visualizations are combined into one [20]. Figure 1 illustrates data flows for CVs and multi-view visualizations. Pagendarm and Post investigated a wider variety of mechanisms to compare two-dimensional visualizations derived from a wider range of data pipelines [17]; CVs fit into this framework. While underlying principles and implementation details may be shared among multiview visualizations, CVs as formulated here, and Pagendarm's comparative visualization framework, we believe that the focused data pipeline of the comparison visualization formulation is novel.

1.2 Spaces of Designs

We set out to describe the space of designs of comparison visualizations for immersive virtual reality; defining a space of designs is one established way of handling the complexity of design choices. The basis of such a space is a collection of parameters or characteristics that a particular design might express. This provides a unified model for existing designs and anticipates the description of future designs. Since the space of CV designs is potentially infinite-dimensional, our work covers only a subspace. We designate this parameterized subspace of the space of designs as a *description space*.

Description spaces have been defined in other areas of visualization research. Schulz, et al. described parameters for the design of treemaps [11], while Card, et al. and Chi focused on design for information visualization in general [4, 6]. Marks, et al. developed a tool for exploring spaces of designs in visualization [15]. The rich prior work applying spaces of designs to visualization problems motivates a similar approach in this work.

1.3 User Studies

We wish to qualify our CV description space with some reckoning of the utility of different points in the space—the usefulness of a particular CV when actually using it to compare datasets. We consider the basic *comparison task* to be identifying a pair of semantically equivalent locations in the two datasets and then determining whether they are qualitatively different. Our user study does not directly measure performance on this task but instead has domain experts evaluate their utility in various hypothetical scientific research situations. Similar subjective evaluations that derive statistically meaningful results have been undertaken by Demiralp, et al. [5] and notably Giesen, et al. [9].

There are several distinct pools from which subjects may be drawn for visualization user studies, including non-experts, visualization experts, scientific domain experts, and visual design experts [1]. Each of these groups has advantages and disadvantages relative to the goals of a given study; domain experts are

Parameter	**Description**
All Visible (AV)	All points in both datasets are visible at all times.
Reduced Self-Occlusion (RSO)	The CV provides a mechanism for reducing visual occlusion of features within a dataset by other features within the same dataset
Feature Non-Occlusion (FNO)	Registered features do not visually occlude each other.
Feature Alignment (FA)	Every feature's *position* is nearby its registered feature.
Partial Feature Adjacency (PFA)	Some, but not all, of the visible features at any given time are nearby their registered features.
Total Feature Adjacency (TFA)	All the visible features at any given time are nearby their registered features.
Features in Pairs (FP)	For every *visible* feature, its registered feature is also visible, and every non-visible feature's registered feature is also non-visible.
Coded Differences (CD)	Primary visual coding is given to computed featurewise differences between the datasets.
Full Coding Space (FCS)	The CV does not use any primary visual codes, but instead leaves them open for use by the client visualization.

Figure 2: The parameters of the comparison visualization description space.

preferable for our purpose of task-dependent evaluation of the utility of visualizations. Furthermore, our user study involves tasks in an application context, as in [22], [21], and [27], rather than the traditional approach of examining simple, abstract tasks. The results of such studies are somewhat more difficult to analyze but nonetheless produce unique and meaningful insights.

Note that while the utility of the CV may depend upon properties of the client visualization, and a good CV design process ought to adapt to that, the design of the client visualization itself is considered a separate problem. Other nontrivial tasks that are assumed to be separate from the CV design process include spatial registration and temporospatial normalization. All of these are active and specialized fields of inquiry that are outside the scope of the current work [10, 8, 23].

2 Experimental Design

The design of our experiment involved two steps: first, defining the parameters for the CV description space, and second, designing a user study to evaluate these parameters.

2.1 Description Space Parameters

In order to develop a description space, one first must consider sample points within it. The literature contains a few multi-view visualization concepts that are adaptable to comparison visualization, such as 3D Magic Lenses [26] and projective Magic Mirrors [13]. By cataloguing these and a few apparent and novel ways to combine visualization outputs for comparison, we assembled a list of sample CVs.

Parameter	Comparison Visualization					
	1	2	3	4	5	6
AV	√	√	√			
RSO				√		
FNO			√	√	√	√
FA	√	√			√	
PFA					√	√
TFA	√	√				
FP	√	√	√	√		√
CD		√				
FCS	√		√	√	√	√

Figure 3: Parametric breakdown of sample immersive virtual reality comparison visualizations.

We then developed a space of parameters that fully distinguished among the sample CVs while remaining as simple as practical. Each sample comparison visualization is uniquely described by a binary distribution over these parameters; see Figure 3. Other binary distributions over the parameters describe hypothetical CVs outside our sample set We restrict our model to CVs that act on client visualization renderings in immersive virtual reality that can be interactively time-controlled, rigidly transformed, and rescaled. We further assume coordination of views [3] and equal visual importance given to both client renderings.

The parameters are defined in Figure 2 and illustrated by examples from the training phase of the user study in Figure 5. Some terminology is necessary to understand the definitions. A "feature" is any individual element of the rendering. The visual relationship between each pair of semantically equivalent features in the datasets is an important point of distinction among CVs, and so we refer to these pairs with the simpler term "registered features". We distinguish between a feature's *position* and its *visibility*, which may be controlled independently. Even a feature that is not rendered to the display at a given time is considered to have a position; when the feature is made visible, this position stays consistent.

Note from the definitions in Figure 2 that the parameters are not all independent; for example, AV and RSO are mutually exclusive, as are PFA/TFA and CD/FCS. Because of these dependences, the description space has the undesirable property of including combinations of parameters that describe impossible visualizations. Note also that primary coding for differences (CD) is not possible in some usage contexts, since, as mentioned above, differences are not always well defined.

2.2 User Study

Five pairs of expert users, each one seniority-matched, were recruited from various scientific disciplines. Each pair underwent training to gain familiarity with navigating in virtual reality and to understand the definitions of the parameters. This training took place in the CAVE virtual reality environment [7]. The experimenter guided the users through various VR illustrations of visualizations prepared in CavePainting [12], a freehand modeling program that runs in the CAVE. Two-dimensional views of sample illustrations are in Figure 5 at the end of the paper.

Upon completing the training, the rest of each session was video recorded. We asked the subjects to develop an imaginary but detailed scientific visualization scenario by discussing it aloud with each other. This allowed us to observe the thought process of the expert users without disrupting it [28]. The scenario was to include a phenomenon to study, an independent variable under variation, the dependent variables to be visualized, and how they would be visualized.

Finally, the subjects were asked to evaluate, on a 7-point scale, the usefulness they would expect of each CV parameter when comparing their datasets for their particular scientific purpose.

3 Results

The subjects chose a variety of scientific usage scenarios. In the results, we refer to each pair by the primary subject of its usage scenario.

Pig — Study pig jaw movement while chewing: soft vs. hard food. Visualize a skull model with position of the jaw relative to spatially registered craniums and with color-coded tooth occlusion on tooth surfaces.

Bat — Study bat movement during flight: slow vs. fast flight. Visualize an anatomical model with muscle activation, bone flexing, joint angles, wing distortion, orientation, and potentially other variables.

Knee — Study cartilage recovery after meniscus surgery: little vs. much tissue removed. Visualize 3D reconstruction of the knee based on MRI, with cartilage thickness color-coded.

Wing — Study steady flow over an airfoil: low vs. high Reynolds numbers. Visualize lift and drag with a single glyph each, pressure on wing surface with color, pressure in flow with color-coded translucent isosurfaces, streamlines locally colored by speed, reattachment points on the wings labeled, and vorticity with glyphs.

Brain — Study brain signal source estimation: EEG vs. fMRI. Visualize signal intensity throughout the brain with color.

To account for per-subject scoring bias, we normalized the scores from each pair of subjects by subtracting the mean and dividing by the standard deviation, resulting in z-scores with mean zero and standard deviation 1. The full z-score summary is in Figure 4.

4 Discussion

We note that for the Coded Differences and Full Coding Space parameters, the scores under Bat are sign-reversed by a fairly large margin from the other subjects' scores (see entries marked ^{a} in Figure 4). A T-test is not possible in this case because we have

Param	**Pig**	**Bat**	**Knee**	**Wing**	**Brain**	μ	σ	μ_{maj}	σ_{maj}
AV	-1.28	-1.92	-0.50	-1.27	-1.38	-1.2	0.5		
RSO	0.85	1.08	0.63	0.88	0.76	0.84	0.17$^{\mathrm{c}}$		
FNO	0	-0.04	0.63	0.88	1.19	0.53	0.54		
FA	0.85	-0.04	1.20	0.16	-0.10	0.41	0.58		
PFA	-1.28	-1.17	-1.64	0.88	-0.95	-0.83	0.99$^{\mathrm{b}}$		
TFA	0.85	0.71	0.06	-1.27	-0.95	-0.12	0.96$^{\mathrm{b}}$		
FP	0.43	1.08	-1.07	0.16	0.76	0.27	0.83$^{\mathrm{b}}$		
CD	0.85	-0.04$^{\mathrm{a}}$	1.20	0.88	1.19	0.82	0.51	1.03$^{\mathrm{a}}$	0.19$^{\mathrm{a,c}}$
FCS	-1.28	0.33$^{\mathrm{a}}$	-0.50	-1.27	-0.52	-0.65	0.67	-0.9$^{\mathrm{a}}$	0.44$^{\mathrm{a}}$

Figure 4: Z-scores for CV design parameters within subject groups. μ_{maj} and σ_{maj} are the mean and standard deviation for the majority-population scores where one subject group's scores differed from the others. Superscripts refer to discussion in the text.

only one sample for the hypothesized second group, but (CD, Bat) varies from the majority mean μ_{maj} by $5.63\sigma_{\mathrm{maj}}$ and (FCS, Bat) varies by $2.79\sigma_{\mathrm{maj}}$. It is notable that the Bat scenario involved a visualization of many variables, for most of which differences are not well defined. Observation of the Bat subjects during the user study indicates that they anticipated the difficulty of automatically coding for differences and instead preferred a full coding space.

The scores for Partial Feature Adjacency, Total Feature Adjacency, and Features in Pairs exhibit especially high variances (see entries marked $^{\mathrm{b}}$ in Figure 4), which we interpret as an indication that the meaning of these parameters was unclear to the subjects. We also observed this difficulty during the training and testing portions of the user study.

The variance for Reduced Self-Occlusion and for the majority scores of CD are particularly low and the means are positive (see entries marked $^{\mathrm{c}}$ in Figure 4). This indicates that the subjects were generally in agreement that RSO and CD (where feasible) would be especially useful parameters for a comparison visualization to express.

As noted in [1], domain experts have the disadvantage as user study subjects of being difficult to schedule. Our ten-subject user study was conducted over the course of several months to get just five data points per parameter. Statistical significance is difficult to achieve in such a regime. Additionally, an acknowledged shortcoming of a user study based on subjective evaluation of tools by expert users is that they are biased toward established tools. The description space has the power to describe CVs that have never been implemented, so it may be unwise to rely on evaluations based only on existing tools. Feedback from one pair of subjects did suggest an improvement to the list of parameters, however: the preservation of orientation between the two client renderings is not necessarily guaranteed by all CVs, but may be a desirable characteristic.

The results of the user study highlight some lessons for future experimental designs of this type. Models of the space of CV designs should be carefully constructed to have pairwise independent parameterizations; this simplifies analysis and also prevents the description space from including impossible configurations. User studies must also be designed carefully; two options present themselves. The effects of individual parameters could be isolated by comparing utilities in a controlled fashion between contrived CVs that differ by only one parameter. Utility could be measured more accurately but in a less general way by defining specific atomic tasks; this design also benefits from the use of a larger, less specialized subject population.

5 Conclusion

This work makes three main contributions. It formulates the specific concept of comparison visualizations and describes their place and significance in the scientific data pipeline. It highlights the utility of tools for interactively reducing self-occlusion in interactive virtual reality comparison visualizations and the context dependence of other design decisions. It also serves as a guide for developing future systematic investigations of the design of comparison visualization, including the construction of description spaces, which must be independent, and the design of user studies, which must be carefully controlled.

References

[1] Acevedo, D., Jackson, C., Laidlaw, D. H., and Drury, F. Using visual design experts in critique-based evaluation of 2d vector visualization methods. *IEEE Transactions on Visualization and Computer Graphics* (2008). In Press.

[2] Bookstein, F. L. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis 1* (April 1997), 225–243.

[3] Boukhelifa, N., Roberts, J., and Rodgers, P. A coordination model for exploratory multiview visualization. In *Proc. International Conference on Coordinated and Multiple Views in Exploratory Visualization* (July 2003), pp. 76–85.

[4] Card, S., and Mackinlay, J. The structure of the information visualization design space. In *Proc. IEEE Symposium on Information Visualization* (October 1997), pp. 92–99, 125.

[5] Çağatay Demiralp, Jackson, C., Karelitz, D., Zhang, S., and Laidlaw, D. H. CAVE and fish-tank virtual-reality displays: A qualitative and quantitative comparison. *IEEE Transactions on Visualization and Computer Graphics 12*, 3 (May 2006), 323–330.

[6] Chi, E. H. A taxonomy of visualization techniques using the data state reference model. In *Proc. IEEE Symposium on Information Visualization* (2000), pp. 69–75.

[7] Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proc. ACM SIGGRAPH* (1993).

[8] Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., and Frackowiak, R. S. J. Spatial registration and normalization of images. *Human Brain Mapping 3*, 3 (1995), 165–189.

[9] Giesen, J., Mueller, K., Schuberth, E., Wang, L., and Zolliker, P. Conjoint analysis to measure the perceived quality in volume rendering. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (November 2007), 1664–1671.

[10] Johnson, C. R., Moorhead, R., Munzner, T., Pfister, H., Rheingans, P., and Yoo, T. S., Eds. *NIH-NSF Visualization Research Challenges Report*, 1st ed. IEEE Press, Los Alamitos, CA, USA, 2006.

[11] Jörg Schulz, H., Luboschik, M., and Schumann, H. Exploration of the 3D treemap design space. IEEE Visualization'07, Poster Compendium, 2007.

[12] Keefe, D. F., Acevedo, D., Moscovich, T., Laidlaw, D. H., and LaViola, J. CavePainting: A fully immersive 3D artistic medium and interactive experience. In *Proc. ACM Symposium on Interactive 3D Graphics* (2001).

[13] König, A., Doleisch, H., and Gröller, E. Multiple views and magic mirrors—fMRI visualization of the human brain. Tech. Rep. TR-186-2-99-08, Institute of Computer Graphics, Vienna University of Technology, February 1999.

[14] Krekel, P. R., Botha, C. P., Valstar, E. R., Bruin, P. W. d., Rozing, P., and Post, F. H. Interactive simulation and comparative visualisation of the bone-determined range of motion of the human shoulder. In *Proc. 17th Conference on Simulation and Visualization* (Erlangen, Germany, March 2006), T. Schulze, G. Horton, B. Preim, and S. Schlechtweg, Eds., SCS Publishing House, pp. 275–288.

[15] Marks, J., Andalman, B., Beardsley, P. A., Freeman, W., Gibson, S., Hodgins, J., Kang, T., Mirtich, B., Pfister, H., Ruml, W., Ryall, K., Seims, J., and Shieber, S. Design galleries: a general approach to setting parameters for computer graphics and animation. In *Proc. ACM SIGGRAPH* (1997), pp. 389–400.

[16] Myers, J. L., Ed. *Fundamentals of Experimental Design*, third ed. Allyn and Bacon, Boston, MA, USA, 1979.

[17] Pagendarm, H.-G., and Post, F. Comparative visualization - approaches and examples. In *Visualization in Scientific Computing*, M. Göbel, H. Müller, and B. Urban, Eds. Springer, 1995.

[18] Pagendarm, H.-G., and Walter, B. Competent, compact, comparative visualization of a vortical flow field. *IEEE Transactions on Visualization and Computer Graphics 1*, 2 (Jun 1995), 142–150.

[19] Privitera, C., and Stark, L. Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 9 (Sep 2000), 970–982.

[20] Roberts, J. C. Multiple view and multiform visualization. In *Proc. SPIE* (2000), vol. 3960, pp. 176–185.

[21] Saraiya, P., North, C., and Duca, K. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics 11*, 4 (July 2005), 443–456.

[22] Swan, J.E., I., Gabbard, J., Hix, D., Schulman, R., and Kim, K. P. A comparative study of user

performance in a map-based virtual environment. *Proceedings of IEEE Virtual Reality 2003* (March 2003), 259–266.

[23] Tanawongsuwan, R., and Bobick, A. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2001), vol. 2, pp. 726–731.

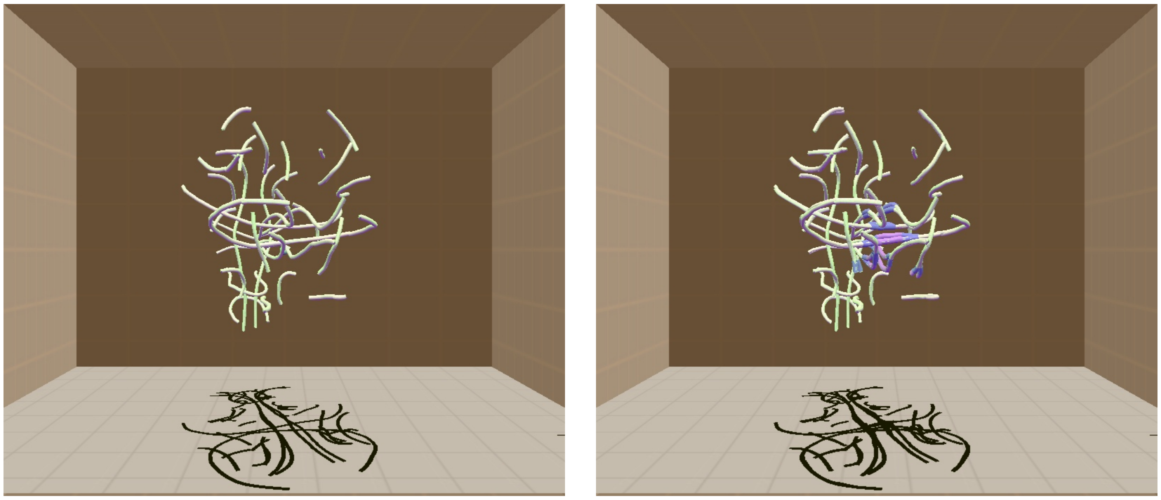
[24] van Dam, A., Forsberg, A., Laidlaw, D. H., LaViola, J., and Simpson, R. M. Immersive VR for scientific visualization: A progress report. *IEEE Computer Graphics and Applications 20*, 6 (November/December 2000), 26–52.

[25] Verma, V., and Pang, A. Comparative flow visualization. *IEEE Transactions on Visualization and Computer Graphics 10*, 6 (November 2004), 609–624.

[26] Viega, J., Conway, M. J., Williams, G., and Pausch, R. 3d magic lenses. In *Proc. ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 1996), ACM, pp. 51–58.

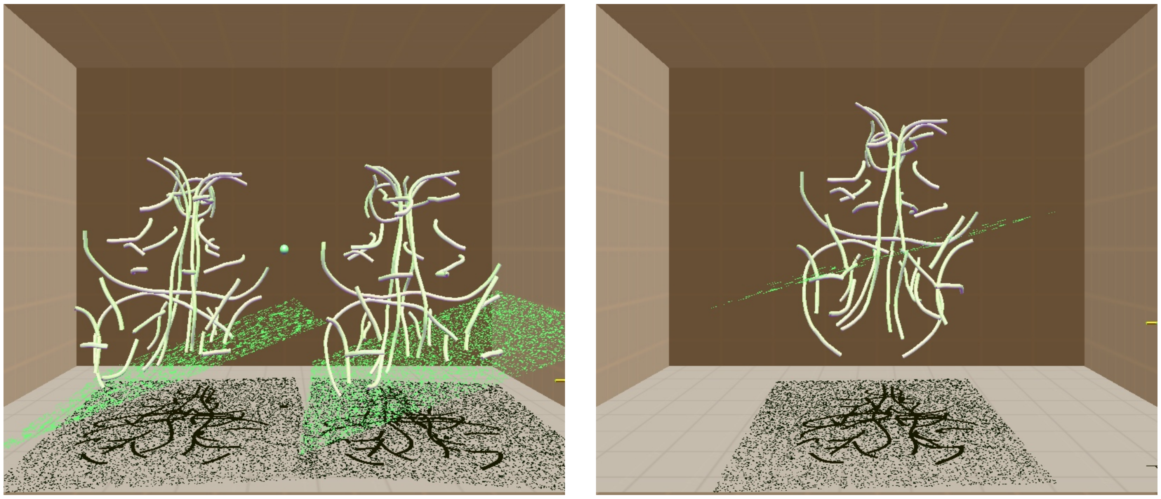
[27] Vote, E., Acevedo, D., Laidlaw, D. H., and Joukowsky, M. Discovering Petra: Archaeological analysis in VR. *IEEE Computer Graphics and Applications 22*, 5 (September/October 2002), 38–50.

[28] Wildman, D. Getting the most from paired-user testing. *Interactions 2*, 3 (1995), 21–27.



(a) A single brain visualization, showing major white matter tracts.

(b) A CV with AV, FA, TFA, FP, and CD.



(c) A CV (superior view) with RSO, FNO, FP, and FCS.

(d) A CV (superior view) with FNO, FA, PFA, and FC. Note that one dataset is rendered on one side of the plane approximately parallel to the ground, and the other is rendered on the other side. This effect is easier to see in VR.

Figure 5: Illustrations of the various CV design parameters. Note that the green planes are included for clarity in the 2-D rendering only and represent interactive interface elements that would not normally be visible in virtual reality.