Using Virtual Reality to Improve Performance and User Experience in Manual Correction of MRI Segmentation Errors by Non-experts

Dominique Duncan¹ · Rachael Garner¹ · Ivan Zrantchev¹ · Tyler Ard¹ · Bradley Newman² · Adam Saslow² · Emily Wanserski² · Arthur W. Toga¹

Published online: 20 July 2018
© Society for Imaging Informatics in Medicine 2018

Abstract

Segmentation of MRI scans is a critical part of the workflow process before we can further analyze neuroimaging data. Although there are several automatic tools for segmentation, no segmentation software is perfectly accurate, and manual correction by visually inspecting the segmentation errors is required. The process of correcting these errors is tedious and time-consuming, so we present a novel method of performing this task in a head-mounted virtual reality interactive system with a new software, Virtual Brain Segmenter (VBS). We provide the results of user testing on 30 volunteers to show the benefits of our tool as a more efficient, intuitive, and engaging alternative compared with the current method of correcting segmentation errors.

Keywords Virtual reality · Segmentation · MRI · Imaging · Quality control

Introduction

Segmentation of magnetic resonance imaging (MRI) scans into labeled anatomical regions is a crucial aspect of the neuroimaging workflow. The goal of segmentation is to algorithmically divide MRI voxels into brain or non-brain tissues, specific tissue types, including gray and white matter, or specific neural structures, e.g., hippocampus [17]. This allows neuroimaging researchers to examine individual structures within the cerebral cortex and compute volumetric analysis. Manual segmentation is widely considered the “gold standard” for segmentation [20], but is time-consuming to learn and complete, requires considerable expertise in neuroanatomy and MRI, and can lead to inter-variability and intra-variability if segmenters do not follow

consistent protocols [14]. Thus, automatic segmentation software has been developed to process images more quickly with greater consistency.

One such software is FreeSurfer [9], an open source software package developed by the Athinoula A. Martinos Center for Biomedical Imaging that is used for the creation of computerized 3D models of the brain derived from MRI scans (<https://surfer.nmr.mgh.harvard.edu/>). FreeSurfer’s segmentation algorithm first distinguishes cortical white and gray matter and then the resulting image is segmented using a Markov random field model and surface-based atlas [9]. We chose to focus on FreeSurfer because it is the most widely used automatic software for segmentation, as it shows a stronger correlation to hand-tracing than other automated segmentation programs, including FSL-FIRST [19]. However, it is still imprecise, requiring users to correct errors manually in individual MRI slices.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10278-018-0108-5>) contains supplementary material, which is available to authorized users.

✉ Dominique Duncan
Dominique.Duncan@loni.usc.edu

¹ Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, 2025 Zonal Ave., Los Angeles, CA 90033, USA

² RareFaction Interactive, 1725 Camino Palmero Street, 410, Los Angeles, CA 90046, USA

Related Work

Previous work has analyzed FreeSurfer’s segmentation errors with regard to the hippocampus, a region of interest in the study of epilepsy [1] and Alzheimer’s Disease [10], indicating that overestimation amounted to 23–29% in healthy adults [3] and 51.53–60.38% in children [24].

Numerous studies have sought out computational methods that attempt to correct this overestimation. One model

improved segmentation of the hippocampus by 2–6% by devising a hybrid algorithm that incorporated probabilistic atlases and anatomical landmarks [5]. Another method improved automatic segmentation errors by incorporating a wrapper method, which learned patterns of systematic errors that could be applied to new images, producing 72% fewer errors than with FreeSurfer alone [28]. These methods, among others [29, 33, 34], have introduced computational analysis to better classify different tissues as well as detect and remove mislabeled partial volume voxels.

Motivations

Although these methods of intervention have improved FreeSurfer's segmentation, manual intervention is still required to ensure precision and accuracy. Though many researchers are familiar with volumetric segmentation, the repetitive and unengaging task is often assigned to lab newcomers, such as undergraduate researchers. Our lab has invested significant resources in training student researchers to complete error correction via manual segmentation in FreeSurfer, but we have experienced high turnover due to the unengaging and time-consuming nature of the task. Training has become a constant concern and strain on experts in the field. Developing a more efficient and engaging segmentation correction process may decrease the time required to correct each brain and simultaneously improve research assistant retention; it may even increase the number of students, researchers, and potentially non-experts willing and able to complete the task.

In this study, we present Virtual Brain Segmenter (VBS), a novel tool that utilizes virtual reality (VR) as an innovative, user-friendly interface for correcting images. If VR better represents the 3D structure of MRI data than traditional 2D computer interfaces, VBS may be more intuitive to users and lessen the burden of training novices in a field that experiences high turnover.

Benefits of Virtual Reality

Virtual reality has proven to be an effective research tool in many medical domains, including cognitive behavioral therapy to treat social anxiety disorder [13] and claustrophobia reduction for patients in MRI scans [12], and as an instructional tool to teach human anatomy [8].

We have utilized VR over other methods because it allows users to make detailed observations in naturalistic navigation. Users can walk through multiple slices of MRI at a time, reorient the image by turning their heads or walking around, and grab the image to bring it closer or farther away, providing an immersive experience that extends beyond the capabilities offered by typical cursor interaction. Rather than scrolling and zooming in a traditional 2D environment, VR offers users a

360-degree view of each voxel, simplifying the process to determine which voxels comprise specific neural structures. VR provides depth information, which can help users detect subtle errors by providing contextual 3D information that is not present in 2D interfaces [4]. In addition, VR's immersive environment removes peripheral visual stimuli, which helps people to remain engaged in the task by reducing distractions [30]. This 3D analysis may be especially helpful in correcting complex errors that are more challenging to discern in a 2D environment.

Manual correction is required in a variety of neural regions, most commonly along the borders of neural and non-neural tissues. In order to address the effectiveness of VBS as a correctional tool while minimizing the duration of the explanatory and experimental trials, a small region of interest was chosen. One specific error that FreeSurfer consistently produces is labeling the optic nerve as brain tissue instead of nerve tissue. This error tends to be more easily visible to a non-expert compared with more nuanced regions, such as the basal ganglia, so novices commonly learn to make segmentation corrections on the optic nerve before mastering more complex regions. Since it is approachable for novice users, the optic nerve was chosen as the region of interest for this study. If this method proves to be efficient and effective for correcting segmentation errors on the optic nerve, as seen in Figs. 1 and 2, further studies may be conducted in other regions where segmentation is flawed.

Immersive virtual reality is an important element of VR applications that we have implemented in VBS. Immersive VR, a perception of being physically present in a non-physical world, has shown to be more powerful and more tuned to the human sensorium in comparison to the traditional interaction with desktop computers [26]. Immersive VR not only outperforms desktops but also provides the user with a better understanding of the visualization [2].

Recent advancements in technology have led to the HTC Vive, an advanced immersive VR system that has a head-mounted display with 2160×1200 pixel resolution, 110-degree field of view, a 15 ft. $\times$ 15 ft. tracking area, and a 90 Hz refresh rate, making it ideal for this experiment.

The overarching goal of this investigation is to test whether FreeSurfer or VBS is a more time-efficient and accurate correctional process, effectively exploring if VR could be a viable tool in neural segmentation. In this study, we have two main objectives. First, we test the efficiency and accuracy of segmentation error correction on the optic nerve that arise from FreeSurfer's automated processing. Accuracy of the intervention is calculated against a corrected baseline that is verified by each member of the research team. The second objective of this experiment is to gauge the user's preference between the standard correction procedure in FreeSurfer and VBS. User preference is assessed through a brief survey, which participants complete after correcting segmentation errors on the optic nerve using both tools.

Fig. 1 The blue region in the 3D rendering of the brain on the left is the portion of the optic nerve that is incorrectly included in FreeSurfer's segmentation, and the brain on the right shows that portion of the optic nerve removed

System

The Unity game engine [31] was chosen for development with the HTC Vive as the target release platform. An iterative development strategy was chosen where research, design, implementation, and feedback loops informed the content of subsequent iterations.

As MRI data are recorded in a voxel-based manner, our team's first step was to integrate an I/O library for the NIfTI-1 (Neuroimaging Informatics Technology Initiative) file format. The data layer is represented internally as a 3D array of scalar values between 0 and 255. Look up tables (LUTs) are used to associate scalar values recorded in an MRI's voxels with RGB color values. Various LUTs are often employed to provide optimal perceptual differences relative to the task at hand. For example, a LUT that colors most of the optic nerve purple and gray matter green enables improved boundary detection as compared with typical grayscale.

In order to remain data and LUT agnostic, our team implemented a system through which MRI files and the LUTs on which they rely are described in a JSON (JavaScript Object Notation) file; these resources are loaded dynamically at runtime. Using this approach, we can load and display any MRI file that uses the NIfTI-1 format with any LUT the user chooses to apply. This provides potential utility for the tool outside of its intended scope.

In developing the VR-based volume editing software, several computational challenges required special attention. Our

MRI data, which is a typical acquisition size, is 256 voxels per axis, totaling 16,777,216 data points. Despite our development machines' excellent hardware for rendering high-end real-time graphics (i7-6800k CPU, GTX 1080 GPU), our dataset's size and the eventual need for computationally expensive rendering (i.e., transparency) required that we take into account our design decisions' impact on frame rate. Our goal was to develop a visualization and interaction model that enabled rapid identification of voxels associated with the optic nerve, and a precise, ergonomic way to eliminate them using the HTC Vive's input device.

To increase the efficiency of the segmentation task, we aimed to provide an interactive 3D view into the interior of the brain, which enables accurate editing of the data faster than FreeSurfer's single-layer approach. In order to view multiple layers (i.e., scan slices), we first created a single layer of 256×256 cubes to represent the data for that layer. Each cube was scaled down by approximately 60%, creating gaps between cubes and allowing the user to see behind the current layer. We then created six additional layers behind this layer, gradually increasing the scale of the cubes so that the viewer can no longer see between or behind the seventh layer. This enables the user to effectively view several 3D layers simultaneously, which is a vast improvement when editing 3D data. The inclusion of a pial surface mesh (the boundary between the brain and cerebrospinal fluid) illustrated in yellow in Fig. 3 also acts as a visual aid for identifying non-cortical structures.

Fig. 2 The figure on the left shows the cerebral tissue after FreeSurfer's initial segmentation, and the figure on the right shows the manually corrected cerebral tissue. In this view, the coloring scheme represents different intensities to code for various tissue types, i.e., gray vs. white matter. This color scheme better represents differences in tissue types, especially to non-experts, than a grayscale color scheme

Fig. 3 VBS interface used for treatment B

Participants have several options to orient themselves in the immersive environment. Head movements shift the visual frame of reference to face the head's changing orientation. Users can also physically walk through the 15 ft. × 15 ft. tracking area to walk through the interior of the brain. Zooming features allow for the traditional expand and contract visuals mimicked in on-screen computer softwares.

Interactivity is achieved via two Vive controllers, each held in one hand. One controller allows users to scroll through the layers dynamically and move the entire volume's position; the second controller or "wand" is topped with a red sphere that turns voxel cubes white to mark them for removal. As the wand is moved forward by the user in space, the foreground layers fade away, allowing the user to view and edit the layers clearly in the background. This editing process can be completed with a single gross motor motion of the controller.

User Study

To test the effectiveness of VBS, a within-subject design experiment was conducted. Participants were exposed to two treatment options, segmentation correction via FreeSurfer and VBS, to evaluate the effectiveness of our novel interface compared with the widely used and well-regarded visualization software.

Thirty participants over the age of 18 participated in the study. Participants either voluntarily responded to flyers posted around campus or were recruited via word of mouth. They received no compensation for their participation. Participants had no prior experience with FreeSurfer or VBS, although some had prior exposure to VR as an entertainment device. We chose to test our system with participants inexperienced with segmentation both to eliminate training bias and because our goal is to develop a protocol that requires a less intensive training regimen that is needed to master segmentation error correction in FreeSurfer.

Experimental Design

As both error correction methods require background information and instructions for operating the hardware and software, participants were given 5 min to read an introductory handout that described the purpose of this study and their objective: to manually remove the optic nerve from segmented data that mischaracterized the optic nerve as brain tissue rather than nerve tissue. The handout also included specific guidelines on how to operate each treatment protocol, including screen captures of each treatment visualization and descriptions of relevant buttons and their functions. Treatment A utilized the software FreeView, a widely used FreeSurfer visualization tool, through which users manually correct neural segmented volumes via a typical 2D screen and mouse interface. All participants completed treatment A using a 13-in. MacBook Air and trackpad interface, the standard model used by research assistants for this task. Treatment B utilized VBS, our VR tool designed to provide the same capabilities of manual volume correction in a 3D immersive environment via the HTC Vive VR hardware.

After reading the introductory handout, 15 participants were assigned to begin with treatment A and 15 with treatment B. In both treatments, subjects were notified that their objective was to complete segmentation error corrections on the optic nerve in the image presented in the fastest and most precise way possible. They were also notified that they could stop the experiment at any time; participation was voluntary and participants' consent was obtained before each phase of the experiment began.

For each treatment, participants were given a 5-min demonstration of FreeSurfer and VBS to familiarize themselves with the software's interface. They were also given a 5-min period to practice identifying the optic nerve and removing relevant voxels under the guidance of an experimenter. Once the instruction and practice periods were complete, a timed trial began. All participants began both trials seeing the data

from a fully zoomed out viewpoint, positioned at the point in the volume corresponding with the first layer where the optic nerve is visible. When timing began, participants used the cursor or remote to select voxels they thought should be removed from segmentation. Timing concluded when the participant verbally signaled completion of the task.

Results in both treatments were judged by comparing the voxels participants selected for removal to a key of correct voxels completed by an expert and verified by each investigator. To evaluate participant performance in VBS, we defined criteria to determine true positives (TP), false positives (FP), and false negatives (FN) against the perfectly corrected baseline. TPs exist where the participant correctly deleted voxels that were also deleted in the baseline. If a participant failed to delete a voxel that was deleted in the baseline, it is considered a FP. If a participant deleted a voxel that was not erased in the baseline, it is considered a FN. These values are calculated through arithmetic analysis in the National Institute for Mental Health's Analysis of Functional NeuroImaging (AFNI) programs. The 3dcalc program in AFNI allows for voxel-by-voxel comparison of datasets, allowing us to compare the participants' voxel transformations with those in the baseline. Once these values are obtained using AFNI, we calculated sensitivity ($TP/(TP + FN)$) and precision ($TP/(TP + FP)$) for each participant.

After completing both treatments, participants were asked to fill out an anonymous four-question survey. The survey included a question to indicate which of the two treatments the participant preferred as well as three free-response questions inviting participants to offer suggestions to improve both methods of correcting segmentation errors and the experiment as a whole.

Results

The average time for participants to complete treatment A was 231.86 (SD = 73.46) seconds, and the average time for participants to complete treatment B was 163.81 (SD = 75.04) seconds (Fig. 4). On average, participants completed treatment B 68.05 s faster. A paired *t* test was conducted on the results, producing a value of $t = 4.1544$ ($n = 30$, $df = 29$). From the two-tailed distribution, we calculated $p = 0.0002627$, indicating that completing the task with VBS was significantly faster than using FreeSurfer.

Participants were able to remove the optic nerve 68 s faster using VBS than using FreeSurfer ($p < 0.001$, highly statistically significant). Individual results can be seen in Table 1. Participants completed the VBS task with a mean precision of 0.85 (SD = 0.107) and sensitivity of 0.86 (SD = 0.118). Ganz et al. demonstrated that non-experts manually correct errors in FreeSurfer with a mean sensitivity of 82% and precision of 42% [11], so our results suggest that non-experts may be able to complete similar corrections with 4% higher sensitivity and 43% higher precision in VBS. Because we chose a region of

interest that comprises only 69 voxels out of the 16,777,216 voxel grid, we did not calculate accuracy and specificity values; the large number of true negatives would have skewed the values to be very high. While our task was to edit a small portion of the brain, users had the ability to edit the entire volume, resulting in the true-negative being largely non-meaningful.

Of the 30 individuals who participated in the experiment, 28 completed the follow-up survey. All participants wrote that they preferred correcting the images using VBS over FreeSurfer. Many participants explained that the novelty of the VBS experience made it more exciting and fun; some wrote that it made the task feel more like a game.

In addition to increased enjoyment, 10 participants also specifically stated that VBS was more intuitive or easier to interact with than the FreeSurfer interface, even having never used a VR headset before. One participant specifically noted that the error correction was “more intuitive with 3D modules” that were “clearly provided by the VR environment,” supporting our hypothesis that the 3D environment provided by VR (specifically the HTC Vive, which allows for movement and motion) may better lend itself to analysis of 3D structures, such as the brain.

In the survey, participants were encouraged to provide specific feedback as to how each treatment could be improved. With regard to FreeSurfer, six participants noted that treatment A may have been improved by allowing participants to use another input mechanism, such as a touch screen device or a device that utilizes a pen stylus (i.e., a Wacom tablet) rather than the MacBook Air's built-in trackpad. Two participants also stated that they had no suggestions for improvements to the FreeSurfer program, because the system is “outdated,” and the only way to assuage its tediousness would be to incentivize precise and fast completion with rewards.

Participants offered fewer criticisms about treatment B but did indicate several changes that could be made to improve VBS. Five participants suggested that we gamify the voxel deletion through the incorporation of sound effects and a scoreboard to make the process more time-incentivizing and entertaining. Other suggested improvements included reversing the up/down command buttons to coincide with the direction of slice progression, adding outlines around voxels to see their partitioning more clearly, and introducing an undo voxel deletion option (rather than a global undo button). Six participants complained that the visuals were slightly blurry, although it is likely that this resulted from the headset being improperly situated. Three participants listed only one criticism: they wished they had more than 5 min to practice in the VR environment.

Discussion

This study demonstrates the feasibility of using VR as a correctional tool for errors that arise from automated segmentation and how spatial immersion enables users to correct

Fig. 4 Time to complete correction of errors in FreeSurfer and Virtual Reality. Each point represents one participant. Points below the 45-degree line indicate participants who completed segmentation faster with VBS than FreeSurfer

segmentation errors quickly and accurately. These results were obtained during participants' first attempts at removing the optic nerve; we hypothesize that results would have improved had we also recorded trials for a second, third, or fourth attempt. With treatment B averaging more than 1 min faster than treatment A in this initial trial, we see this tool as potentially quite useful for correcting automated segmentation errors more effectively than a typical 2D screen/cursor interface, especially with further improvements to the functionality.

Computationally, our cube-layers-based approach with spacing in between cubes is expensive, but interactive frame rates are still maintained. The HTC Vive also boasts higher resolution than other VR systems, so the images that it produces appear clearer to participants. This is an important aspect with regard to this study since participants are expected to erase individual voxels with precision.

From a usability perspective, the ability to walk within a defined boundary may make the HTC Vive the most suitable VR headset over competitors, such as the Oculus Rift. The ability to move around promotes an active, standing posture (although the task could be completed sitting down as well). For long periods of segmentation, a standing posture is beneficial, as it has been shown that workstations that support non-seated positioning (i.e., standing desks) reduce upper back/neck pain and improve mood states [21]. Arm fatigue was not found to be a significant issue; users employed gross motor skills followed by frequent resting rather than continuously elevating their arms.

Additionally, the fact that the method was preferred by many participants for its entertainment component indicates that VBS may be a more engaging method of segmentation error correction. This could have positive implications in terms of recruiting

Table 1 Time trial results for participants in Virtual Reality and FreeSurfer

Participant	Virtual reality (minutes)	FreeSurfer (minutes)	Participant	Virtual reality (minutes)	FreeSurfer (minutes)
1	3:41.85	3:45.12	16	1:46.94	6:00.18
2	2:53.36	3:44.35	17	4:31.36	4:28.39
3	2:20.53	5:05.89	18	2:46.94	2:34.66
4	4:45.39	5:09.87	19	2:48.04	3:16.88
5	3:59.10	2:06.65	20	1:34.72	1:56.81
6	1:42.15	4:25.84	21	2:05.31	3:03.86
7	3:28.82	2:54.59	22	0:50.51	2:53.91
8	5:13.29	3:42.23	23	1:18.54	3:51.73
9	1:35.20	3:11.49	24	1:46.59	3:30.88
10	1:26.27	3:48.03	25	1:09.92	5:00.56
11	3:22.30	4:08.08	26	4:11.53	6:20.20
12	3:52.92	3:00.39	27	2:07.45	2:22.56
13	2:33.52	3:38.32	28	1:20.19	2:47.26
14	3:21.16	6:40.86	29	2:21.06	4:27.63
15	5:04.93	5:14.61	30	1:53.53	2:44.04

and retaining correction specialists, especially if we expand the project to include features that gamify VBS.

In discussions with participants after the completion of the experiment, some noted that gamifying the experience could improve motivation and make the task seem less tedious than it may be in the designed interface. However, it should be noted that performance could decrease if participants are more concerned with time than precision.

Several games have emerged over the past few years as popular tools for crowdsourcing various tasks to combat the time-intensive nature of different pre-processing or basic analysis steps on massive datasets. Some examples of these are FoldIt for protein structure prediction [6], EyeWire for mapping 3D neurons in a retina [15, 18], and EteRNA for designing RNA sequences that fold into a target shape [7, 22]. The FoldIt game resulted in the publication of a highly influential journal paper [6] with over 57,000 players as co-authors.

A recent study assessed whether utilizing non-experts was a feasible strategy to correct errors that arise from automatic processing of MR cortical surfaces [11]. Images were first processed using FreeSurfer then corrected by the crowd (found via the crowdsourcing platform Amazon Mechanical Turk); users were asked to draw boxes around all errors detected. The results showed that the task could be completed quickly by non-experts with a mean sensitivity of 82% and precision of 42%. The study emphasized that this method would be useful to minimize segmentation experts' work to focus only on error correction while error detection could be completed by the crowd [11].

These projects emphasize the significant role that carefully engineered games can play in processing useful scientific data. We hope this study will be the beginning of a multi-stage project to create a gamified virtual reality program that can be utilized by a crowd to correct segmentation errors. With this as our long-term goal, we completed this study to first explore how feasible it is to utilize VR as a correctional tool.

Correcting segmentation errors has been an important area of interest in neuroimaging for several decades, but despite significant research efforts, automated segmentation protocols are still imprecise and inaccurate. Developing efficient segmentation correction methods could aid volumetric analysis of structures implicated in neurological disorders. Hippocampal atrophy has been reported as an early biomarker for Alzheimer's disease, fronto-temporal dementia, and semantic dementia [27], suggesting that accurate volumetric analysis could support early diagnosis and intervention. Similarly, accurate localization and quantification of gliomas and glioblastomas, which can be difficult due to diffuse structure and diverse presentation, could improve diagnostics and treatment plans [16]. Accurate segmentation could have far-reaching benefits in neuroscience, spanning from research in development and cognitive processes to degenerative diseases, such as Alzheimer's disease, autism spectrum disorder, and epilepsy.

Future Work

In the next phase of this project, we aim to make specific improvements to VBS's software. Some of these improvements were minor suggestions made in the user surveys: reversing the up/down command keys to coincide with the direction of slice progression, adjusting the brush size, and modifying the undo voxel deletion tool. We also plan to include a training phase in which participants are provided with active feedback regarding the accuracy of their segmentation corrections. This could act as a periodic quality assurance test by utilizing previously corrected scans; it could also offer more training and familiarization for non-neuroscience trained users.

Once we implement these additional features, we intend to gamify the program, emulating other projects, such as FoldIt and EyeWire. We plan to begin by adding sound effects and a scorekeeping system. If this successfully engages a non-expert audience, we may then explore methods to transform the interface into a different environment altogether (i.e., users would move through a natural scene that is more familiar than the brain to non-experts). With any modification, we hope to make the software visually interesting, engaging, and accessible to a wide audience.

Once modifications are made, we plan to share the software publicly and utilize a crowdsourcing platform for users to correct segmentation errors, which has been shown to be successful in other domains [23, 25, 32]. The software is compatible with a variety of VR platforms so that the maximum number of users would have access. Ideally, once crowdsourced, we would obtain a very large dataset encompassing manual corrections, which would enable a wide variety of improvements and techniques. For example, we could optimally tune inter-rater reliability and guide error-prone users to tutorials that address the particular mistakes they may make. Furthermore, machine learning techniques could be trained through these manual corrections to provide improvements to the automated segmentation processes.

Funding Information This work is supported by the National Institutes of Health grants P41-EB015922, U54EB0020406, and the University of Southern California Provost's Postdoctoral Scholar Research Grant.

References

1. Anand KS, Dhikav V: Hippocampus in health and disease: An overview. *Ann Indian Acad Neurol* 15:239–246, 2012
2. Bayyari A, Tudoreanu ME: The impact of immersive virtual reality displays on the understanding of data visualization. In: *Proc. ACM Sym. on Virtual Reality Software and Technology*:368–371, 2006
3. Cherbuin N, Anstey K, Reglade-Meslin C, Sachdev PS: In vivo hippocampal measurement and memory: A comparison of manual tracing and automated segmentation in a large community-based sample. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0005265>, April 16, 2009

4. Cho I, Dou W, Wartell Z, Ribarsky W, Wang X: Evaluating depth perception of volumetric data in semi-immersive VR. In: IEEE Virtual Reality Workshops. <https://doi.org/10.1145/2254556.2254606>, 2012
5. Chupin M, Hammers A, Liu RS Colliot O, Burdett J, Bardinet E, Duncan JS, Garnero L, Lemieux L: Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *Neuroimage*, 46:749–761, 2009
6. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, Leaver-Fay A, Baker D, Popović Z, players F: Predicting protein structures with a multiplayer online game. *Nature*, 466:756–760, 2010
7. Cordero P, Lucks JB, Das R: An RNA mapping database for curating RNA structure mapping experiments. *Bioinformatics*, 28:3006–3008, 2012
8. Fairén M, Farres M, Moyes J, Insa E: Virtual reality to teach anatomy. In: Proc. Eurographics. <https://doi.org/10.2312/eged.201710262017>, 2017
9. Fischl B. FreeSurfer. *Neuroimage*, 62:774–81, 2012
10. Fox NC, Warrington EK, Freeborough PA, Hartikainen P, Kennedy AM, Stevens JM, Rossor MN: Presymptomatic hippocampal atrophy in Alzheimer's disease. *Brain*, 119:2001–2007, 1996
11. Ganz M, Kondermann D, Andrulis J, Knudsen GM, Maier-Hein L: Crowdsourcing for error detection in cortical surface delineations. *Int J Comput Assist Radiol Surg* 12:161–166, 2017
12. Garcia-Palacios A, Hoffman HG, Richards TR, Seibel EJ, Sharar SR: Use of virtual reality distraction to reduce claustrophobia symptoms during a mock magnetic resonance imaging brain scan: A case report. *Cyberpsychol Behav* 10:485–488, 2007
13. Herbelin B, Riquier F, Vexo F, Thalmann D: Virtual reality in cognitive behavioral therapy: A preliminary study on social anxiety disorder. In: Proc. Int. Conf. on Virtual Systems and Multimedia, 2002
14. Jack, Jr CR, Theodore WH, Cook M, McCarthy G: MRI-based hippocampal volumetrics: data acquisition, normal ranges, and optimal protocol. *Magn Reson Imaging* 13:1057–1064, 1995
15. Kim JS, Greene MJ, Zlateski A, Lee K, Richardson M, Turaga SC, Purcaro M, Balkam M, Robinson A, Behabadi BF, Campos M, Denk W, Seung HS, Eyewirers: Space-time wiring specificity supports direction selectivity in the retina. *Nature* 509:331–336, 2014
16. Li Z, Wang Y, Yu J, Shi Z, Guo Y, Chen L, Mao Y: Low-grade glioma segmentation based on CNN with fully connected CRF. *J Healthc Eng*, <https://doi.org/10.1155/2017/9283480>, June 13, 2017
17. Lötjönen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, Rueckert D, Alzheimer's Disease Neuroimaging Initiative: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 49:2352–2365, 2010
18. Marx V: Neuroscience waves to the crowd. *Nature methods* 10: 1069–1074, 2013
19. Morey RA, Petty CM, Xu Y, Pannu Hayes J, Wagner, II HR, Lewis DV, LaBar KS, Styner M, McCarthy G: A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* 45:855–866, 2009
20. Pardoe HR, Pell GS, Abbott DF, Jackson GD: Hippocampal volume assessment in temporal lobe epilepsy: How good is automated segmentation? *Epilepsia* 50:2586–2592, 2009
21. Pronk NP, Katz AS, Lowry M, Payfer JR: Reducing Occupational Sitting Time and Improving Worker Health: The Take-a-Stand Project, 2011. *Prev Chronic Dis*, <https://doi.org/10.5888/pcd9.110323>, October 11, 2012
22. Rowles TA: Power to the people: Does Eterna signal the arrival of a new wave of crowd-sourced projects? *BMC Biochem* 14:1, 2013
23. Shamir L, Yerby C, Simpson R, von Benda-Beckmann AM, Tyack P, Samarra F, Miller P, Wallin J: Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *J. Acoust. Soc. Am.* 135:953–962, 2014
24. Shoemaker D, Buss C, Head K, Sandman CA, Davis EP, Chakravarty MM, Gauthier S, Pruessner JC: Hippocampus and amygdala volumes from magnetic resonance images in children: Assessing accuracy of FreeSurfer and FSL against manual segmentation. *Neuroimage* 129:1–14, 2016
25. Sun C, Rampalli N, Yang F, Doan AH: Large-scale classification using machine learning, rules, and crowdsourcing. In: Proc. VLDB Endowment 7:1529–1540, 2014
26. Van Dam A, Laidlaw DH, Simpson RM: Experiments in immersive virtual reality for scientific visualization. *Computer and Graphics* 26:535–555, 2002
27. Van de Pol LA, Hensel A, van der flier WM, Visser PJ, Pijnenburg YA, Barkhof F, Gerz HJ, Scheltens P: Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease. *J Neurol Neurosurg Psychiatry* 77:439–442, 2006
28. Wang H, Das SR, Suh JW, Altinay M, Pluta J, Craige C, Avants B, Yushkevich PA, Alzheimer's Disease Neuroimaging Initiative: A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation. *Neuroimage* 55: 968–985, 2011
29. Wang JY, Ngo M, Hessel D, Hagerman RJ, Rivera SM: Robust machine learning-based correction on automatic segmentation of the cerebellum and brainstem. *PLoS ONE*, <https://doi.org/10.1371/journal.pone.0156123>, May 23, 2016
30. Wang Y, Otitoju K, Liu T, Bowman DA: Evaluating the effect of real-world distraction on user performance in virtual environments. In: Proc. ACM Symp. Virtual Reality Software and Technology;19–26, 2006
31. Williams D. Cubiquity for Unity (Voxel Engine). Retrieved September 26, 2016 from <https://www.assetstore.unity3d.com/en/#!/content/12689>
32. Wu H, Sun H, Fang Y, Hu K, Xie Y, Song Y, Liu X: Combining machine learning and crowdsourcing for better understanding commodity reviews. In: AAAI; 4220–4221, 2015
33. Xue H, Srinivasan L, Jiang S, Rutherford M, Edwards AD, Rueckert D, Hajnal JV: Automatic segmentation and reconstruction of the cortex from neonatal MRI. *Neuroimage* 38:461–477, 2007
34. Zheng W, Chee MWL, Zagorodnov V: Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. *Neuroimage* 48:73–83, 2009