



Interactive exploration of neuroanatomical meta-spaces

Shantanu H. Joshi*, John Darrell Van Horn and Arthur W. Toga

Laboratory of Neuro Imaging, Department of Neurology, University of California, Los Angeles, CA, USA

Edited by:

Jan G. Bjaalie, International
Neuroinformatics Coordination Facility,
Sweden; University of Oslo, Norway

Reviewed by:

Shiro Usui, RIKEN Brain Science
Institute, Japan
Jeffrey S. Grethe,
University of California, USA
Jan G. Bjaalie, International
Neuroinformatics Coordination Facility,
Sweden; University of Oslo, Norway

*Correspondence:

Shantanu H. Joshi, Laboratory of
Neuro Imaging, Department of
Neurology, University of California,
Los Angeles, CA 90095, USA.
e-mail: sjoshi@loni.ucla.edu

Large-archives of neuroimaging data present many opportunities for re-analysis and mining that can lead to new findings of use in basic research or in the characterization of clinical syndromes. However, interaction with such archives tends to be driven textually, based on subject or image volume meta-data, not the actual neuroanatomical morphology itself, for which the imaging was performed to measure. What is needed is a content-driven approach for examining not only the image content itself but to explore brains that are anatomically similar, and identifying patterns embedded within entire sets of neuroimaging data. With the aim of visual navigation of large-scale neurodatabases, we introduce the concept of brain meta-spaces. The meta-space encodes pair-wise dissimilarities between all individuals in a population and shows the relationships between brains as a navigable framework for exploration. We employ multidimensional scaling (MDS) to implement meta-space processing for a new coordinate system that distributes all data points (brain surfaces) in a common frame-of-reference, with anatomically similar brain data located near each other. To navigate within this derived meta-space, we have developed a fully interactive 3D visualization environment that allows users to examine hundreds of brains simultaneously, visualize clusters of brains with similar characteristics, zoom in on particular instances, and examine the surface topology of an individual brain's surface in detail. The visualization environment not only displays the dissimilarities between brains, but also renders complete surface representations of individual brain structures, allowing an instant 3D view of the anatomies, as well as their differences. The data processing is implemented in a grid-based setting using the LONI Pipeline workflow environment. Additionally users can specify a range of baseline brain atlas spaces as the underlying scale for comparative analyses. The novelty in our approach lies in the user ability to simultaneously view and interact with many brains at once but doing so in a vast meta-space that encodes (dis) similarity in morphometry. We believe that the concept of brain meta-spaces has important implications for the future of how users interact with large-scale archives of primary neuroimaging data.

Keywords: meta-analysis, neuroanatomical data mining, visual data mining, 3D visualization

INTRODUCTION

The past decade has seen an explosive rise in the volume of brain image scans for clinical, diagnostic as well as research purposes. Fortunately, the neuroimaging research community recognized early on that facilitating data sharing among collaborative research centers is the key to boosting neuroscientific knowledge and discovery. Drawing a parallel with genomics research which has immensely benefitted with such data sharing strategies, a position paper (Eckersley et al., 2003) even goes far to suggest the use of public domain licensing policies, not unlike the GNU public license, for neuroscience data. The consensus on the archiving and sharing of primary neuroimaging data has fostered several large-scale initiatives: The Biomedical Informatics Resource Network (BIRN), the Morphometry and Function BIRN testbed projects (Grethe et al., 2005); The NIH MRI Study of Normal Brain Development (Pediatric MRI Study) and resulting Pediatric MRI Data Repository (Evans, 2006); and The fMRI Data Center (fMRIDC) (Van Horn et al., 2001; Van Horn and Gazzaniga, 2002). Much recently, the Neuroscience Information Framework (NIF) (Hurd, 2005) has initiated the development of a comprehensive experimental, clinical and translational databases, knowledge bases, atlases etc

for processing, analysis, or simulation of brain data. Additionally the Clinical Data Interchange Standards Consortium (CDISC) (Souza et al., 2007) strives to improve data exchange across multiple domains and platforms for medical research as well as health care initiatives.

Notably, the LONI Image Data Archive (IDA) contains neuro-anatomical data from nearly 30 research projects and serves as the primary repository for large studies such as the Alzheimer's Disease Neuroimaging Initiative (ADNI). Data sharing has also indirectly benefitted and affected computational neuroscientific tool development. As algorithms get tested on more and more diverse datasets, they evolve to become more general and robust. Data sharing has been the first step in neuroinformatics research efforts and has largely been the focus of the past decade, and will continue to be so. The neuroscientific community is now getting ready to prepare for the next logical step – database integration (Forsberg and Roland, 2008). Most data storage facilities like the ones above, have implemented centralized repositories in proprietary formats. The challenge that the informatics community faces in the near future is the unification of existing large, heterogeneous neurodatabases in a user-transparent manner. This goes above and beyond

data sharing, where the user can not only access a single database, but can sift through multiple repositories at once without having the database to be localized in a central place. This is very much like the WWW, where there is an interconnection of data processing and storage nodes in a decentralized network. An important step towards this goal will be designing and standardizing robust database exchange protocols, while maintaining compatibility with privacy regulations and laws.

However, an alternate parallel goal complementing these efforts is the ability to *graphically* navigate, browse and query such aggregations of repositories. With the increasing progress in computational processing, and visualization, textual queries and interactions continue to be a severe drawback in future database access, especially with the enormity of the data involved. Recently, Herskovits et al. (Herskovits and Chen, 2008) have developed an open source implementation for a database system with data mining capabilities for managing, querying, analyzing and visualizing brain-MR images. We anticipate a compelling need for similar tools in the neuroscience community that facilitate informatics-driven approaches for users to better examine databases and explore the inter-relatedness of subjects in the population. Our goal then is to facilitate the large-scale informatics, mining, and visualization of the contents of existing neuroimaging data repositories by developing streamlined data processing workflows to decompose the contents of an archive, compare each image volume against all others in the archive, and visually display the results in a user friendly client application. We claim that the neuroimaging data itself can form the basis for such mining, that visualization of how brains relate to one another carries essential information, and that well-designed tools can permit data from outside the archive to be used as the basis for similarity-based searching.

This paper is organized as follows: the Section “Introduction” makes an argument for visual explorative interfaces for large-scale neuroimaging databases. The Section “Materials and Methods” outlines the main idea of this paper. It proposes neuroimaging workflows (see Introduction) focusing towards discriminative analysis for visualization. The Section “Materials and Methods” introduces the concept of a neuroanatomical meta-space built on top of the dissimilarity measures generated by the workflows. A meta-space is constructed in a case study (see Introduction) on a sample dataset of 400 subjects from the ADNI dataset. Finally Section “Discussion” proposes a 3D visualization environment for interactively navigating through this meta-space followed by a discussion.

NEED FOR VISUAL MINING OF NEURODATABASES

There is a growing interest in content-based searches for neuroimaging because of the limitations inherent in meta-data-based systems (Nielsen et al., 2006), as well as the large range of possible uses for efficient image retrieval. Without the ability to examine image content, searches currently rely on meta-data such as captions or keywords, which may be laborious or expensive to produce manually. While textual information about images can be easily searched using existing technology, it requires humans to personally tag and annotate every image in the database. This can be impractical for very large databases. Similarly, there are added benefits for manipulating the search criteria and results visually. A visual interface will

present an opportunity to cluster, classify, and graphically represent data in ways not possible based on textual meta-data alone. We identify the following scenarios where such a graphical navigation system can be applied.

Visualizing anatomical differences and relatedness simultaneously

A single brain image scan may give rise to a variety of anatomies. Pertaining to a specific neuroscientific study, researchers may choose to directly work with MRI images, or work with suitable anatomical representations deconstructed from an MRI image. For e.g. the boundary of the volume, the cerebral cortex can be represented by a topographic two-dimensional geometrical structure (Thompson et al., 2001; Hinds et al., 2008). This structure can be further differentiated by the anatomical folds also known as the sulci and the gyri. One can further descend beneath the cortex to delineate various other structures such as the limbic system, thalamus, hypothalamus, corpus callosum (Narr et al., 2005) etc. Existing neuroimaging analysis and visualization tools restrict users to a single, individual brain image or surface for anatomical studies. While this is useful for structural analysis or evaluation of pertinent anatomies, neuroimaging studies often consist of large population of subjects and resulting brain images. Especially for large-scale statistical or discriminative analyses focusing on disease, genetic, or heritable effects and changes according to neuro-morphology, it would be useful to simultaneously visualize the morphology in an appropriate metric space resulting from the analysis. Currently most neurodatabases are accessible solely by textual queries. Furthermore there is no existing application or workflow that enables the neuroscientist to manipulate neuroinformatics search criteria, and the resulting queries and outputs in a visual manner.

Educational resource or a training environment for neuroscientists

Developments in the area of content representation, interaction, and search has been employed for graphical data with the notable example of Microsoft’s Photosynth that been used to mine the Flickr¹ photo sharing site to then graphically depict a collection of images from a spatial reconstruction of their taken vantage point. Likewise, Google Earth² displays satellite imagery, mapping, and geographic data, permitting interactive search, annotation, and other functions. In the astronomy community, the recent launch of the World wide Telescope[®] [for historical context, see (Szalay and Gray, 2001)] has revolutionized the exploration and search capabilities for astral, galactic, and planetary data obtained from multiple imaging sources. These applications continue to enhance educational instruction, both for the general public and the specialists alike. Similar tools do not yet exist in the neuroimaging community where there is a tremendous potential for computer-simulated training for neuroscientists.

Visual cataloging of neurodatabases

Visual data mining (VDM) is useful in exploratory analysis, where one has limited views and information of the data. With the recent advances in computing and storage, VDM has been

¹<http://www.flickr.com/>

²<http://earth.google.com/>

used for diverse applications such as exploring geospatial data (Keim, 2002; Keim et al., 2004), internet web resource databases (Chen et al., 2007), and analyzing business intelligence patterns (Hao et al., 2000). Such an effort currently does not exist in the field of neuroimaging. The development of visual catalogs of neuroimaging data would enable and enhance large-scale scientific interaction among users. Though some basic image viewing tools exist, we believe a different approach is needed altogether. A content-based solution is beneficial for researchers to more easily examine (dis)similarity between brains and to dynamically visualize patterns that may be indicative of the demographic and clinical attributes of the data themselves. By navigating through a virtual environment via an easy-to-use, web driven application, users will be able to examine large collections of brain data using only their computer mouse.

Visualization of atlas spaces

Individual brain anatomies, have their own local coordinate systems that measure local distortions of features such as curvatures, intensities, and surface areas. For large populations of such anatomies, most approaches construct an atlas template (Mazziotta et al., 2001) and transform all individuals to the atlas. This yields a single anatomical object that is then analyzed or visualized as a representative of the population. This approach also transforms the individual local variation to the atlas thus providing the researcher with an at-a-glance view of the variation across population. The drawback of atlas visualization is that the atlas depicts a single view of the population, and it is difficult to get an overview of the underlying dissimilarity patterns between individual subjects in the study. Often, these atlases are probabilistic in nature and thus only provide a statistical interpretation of the relationship between the template and the individual. Thus one has to continually go back and forth between the template and the individual to relate to, and observe the changes in the native brain space. Instead, a visualization scheme that simultaneously displays the atlas and the data used for its construction, in a meta-space is highly desirable. Moreover, one could technically extend this idea to multiple atlases grouped together with their respective populations.

MATERIALS AND METHODS

This section outlines the concept of a meta-space that follows from large-scale discriminative analyses on a brain population. Essential to the construction of the meta-space is the data processing framework that enables complete workflows leading from the original data in the form of images to the various metrics that attempt to classify, cluster and separate individuals in the population. Due to the enormity of the data, as well as the types of processing involved, we employ a grid-based execution environment. While large-scale distributed processing is an essential component in scientific computing, it has only recently (Rex et al., 2003; Callahan et al., 2009) evolved to adapt itself to biomedical or neuroimaging workflows. The main hindrance for adapting such technologies is the specialized knowledge required to maintain, develop, and execute applications for common neurocomputing tasks. However with latest advances in interfaces and visualization, much of these tasks have become oblivious to the end-user.

For e.g. the LONI Pipeline (Rex et al., 2003)³ is a highly flexible, distributed computing environment that enables parallelized execution of application software especially dealing with brain mapping protocols. It offers an efficient GUI interface to the user, where one can quickly build complete applications using pluggable components called pipeline modules. The pipeline user can further extend the functionality of the pipeline by developing modules in addition to the ones existing in the library. The pipeline communicates back and forth with the grid scheduler to queue up user tasks in an efficient manner. It also provides a feedback mechanism to the user where he can monitor program execution real time from the pipeline interface. Moreover the pipeline also allows the data as well as programs required for analysis, to reside on the user's local machine that launches the pipeline thereby integrating both local and remote resources in a seamless manner. Throughout this discussion, the LONI Pipeline will serve as a convenient execution environment for our architecture. We would also like to stress that the user is not restricted to the LONI computing infrastructure to take advantage of the LONI pipeline. The LONI pipeline is independent of the underlying grid computing environment and can be tailored and adapted to other suitable execution infrastructures⁴.

DATA MINING WORKFLOW FOR DISCRIMINATIVE ANALYSIS

The data is typically stored as MRI images either in the Analyze or the NIFTI format. For the purpose of this discussion, we focus on neuroanatomical volumes, though in the future functional imaging could be incorporated. Depending upon the experimental setup, the data is usually corrected to minimize geometric distortions or non-linearities, and any non-uniform intensities resulting due to magnetic properties of the RF coils. The data can further be sharpened using histogram techniques that can further lead to a reduction of intensity non-homogeneities. The corrected MR images are then stripped of skulls, unwanted tissue, and other extra unneeded anatomical features such as the cerebellum or the brain stem. We used the Brain Extraction Tool (Smith, 2002) tool for skull stripping MRI images in our workflow, although any such similar tool can be used. All image volumes in the database are registered (Woods et al., 1998) to a standard Montreal Neurological Institute (MNI) atlas image. The resulting gray/white matter image is then processed in parallel to i) extract the cortical (gray/CSF boundary) surface (Shattuck and Leahy, 2002), and ii) extract about 56 sub-cortical features (Tu et al., 2008) such as the major gyri, hippocampus, the putamen, etc. This process exclusively gives rise to a geometrical representation that is stored in the form of a triangular mesh using a suitable file format. Henceforth in this paper, brain anatomies will be taken to mean the cortical surface as well as surface parameterizations of the individual sub structures beneath the cortex. The top portion of **Figure 1** shows the pre-processing and feature extraction steps. These steps are implemented as completely automated LONI pipeline modules. **Figure 2** shows an example of a parcellated volume colored according to different anatomies. The original input data

³<http://pipeline.loni.ucla.edu>

⁴For more information about the pipeline, the reader is referred to the article "Efficient, Distributed and Interactive Neuroimaging Data Analysis using the LONI Pipeline" by Dinov et al., 2009.

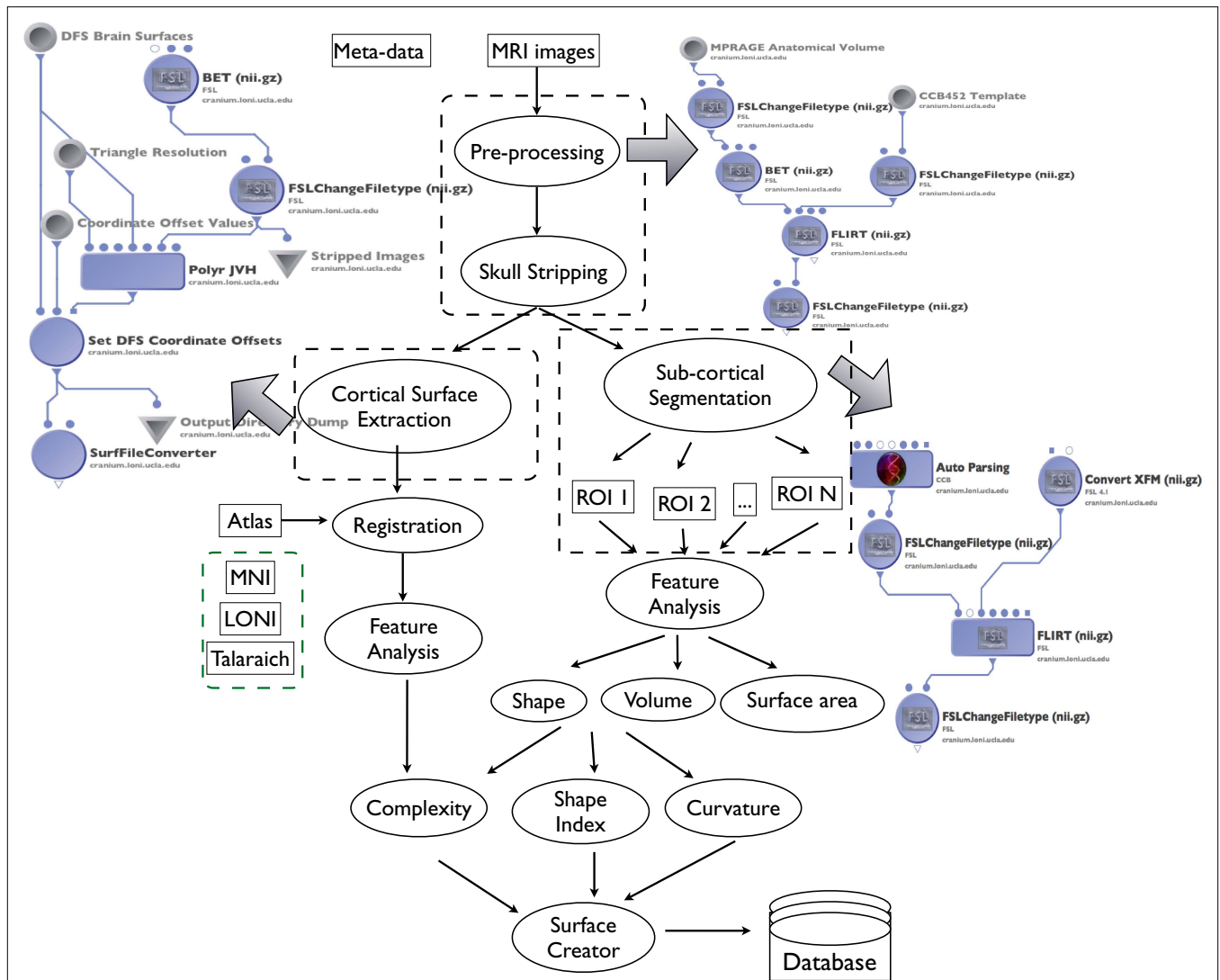


FIGURE 1 | A schematic of the data mining workflows exposed through the LONI pipeline (Rex et al., 2003). The workflow is divided into three parts, i) Processing, ii) Feature Extraction – extracting anatomical features such as cortical surfaces, sub-cortical structures etc. and having 3D mesh

representations for each feature, and iii) Feature Analysis – calculating the local curvature, shape index, cortical complexity, and encoding each surface mesh with these attributes. Each stage is implemented via pipeline without user intervention.

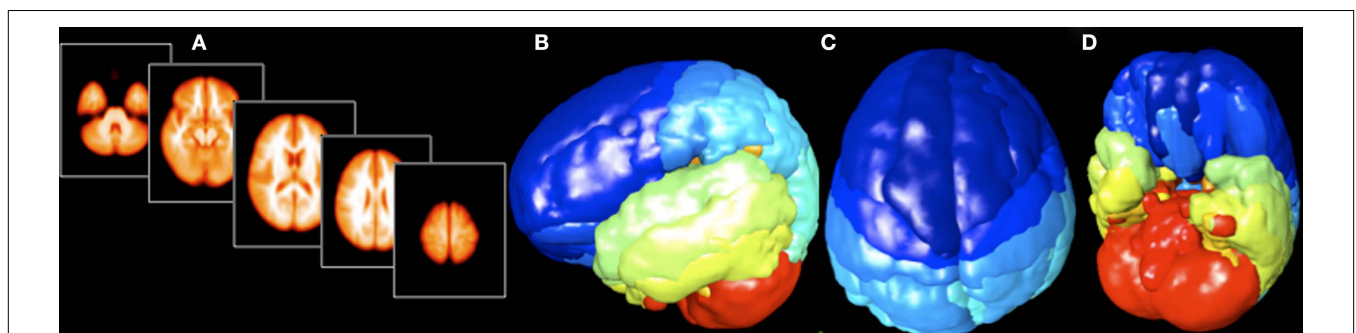


FIGURE 2 | Surface rendering of segmented sub-cortical structures labeled according to regions. (A) Examples of image slices along the axial view. **(B–D)** Parcellated cortical and sub-cortical regions along three views.

is usually accompanied by appropriate meta-tags using predefined XML schemas. The above pre-processed data is then stored hierarchically for a streamlined access in a database.

A NEUROANATOMICAL META-SPACE

The central idea behind atlas meta-spaces is the modeling of the dissimilarities between individuals in a population. The population can be analyzed all at once, or the individual subjects can be grouped according to some well-defined categories. Before performing any type of discriminatory analysis, one needs to consider an appropriate metric space of objects and define a notion of distance associated with it. Metric spaces are mathematically easier to define in case of simple tractable objects such as two-dimensional, or three-dimensional points, multidimensional vector measurements, or objects that lend themselves to functional representations in a well-defined space. In case of neuroanatomical structural data, such as cortical surfaces and sub-cortical structures, it is difficult to have a rigorous definition of a metric space of such entities directly. There are ongoing research efforts to model the geometry of the cortex, or define the shape of three dimensional closed surfaces corresponding to the sub-cortical structures. The goal is to have a mathematical representation of the shape geometry, independent of indeterminacies such as the scale, position, orientation etc. Various researchers have used harmonic functions on a sphere to represent closed genus zero surfaces. In this case the shape distance is measured by a \mathbb{L}^2 distance between the coefficients in the space of harmonic functions. Others have used level set representations for shapes of surfaces, again using the \mathbb{L}^2 metric between two signed distance representations for surfaces. Yet another approach by researchers uses global measurements such as the volume, average curvature, or the surface area of cortex, or the sub-cortical structures. Although, this may be interpreted as a gross simplification of brain geometry, numerous studies have shown the effectiveness of such simple metrics in capturing a global underlying pattern of the data. A study based on simple volume analysis of the Hippocampus (Chupin et al., 2008) was able to correctly classify 82% of Alzheimer's disease (AD) patients with respect to the elderly controls. Another study (Gosche et al., 2002) has also shown that hippocampal volume can be used as an indicator for Alzheimer neuropathology. A recent study (Dubois et al., 2007) also supports that quantitative volumetric analysis on the hippocampus was able to distinguish AD across young and old ages. We will follow a similar approach and utilize metrics that are simple to compute, and lend themselves to an easy interpretation. A few of the metrics considered in this paper are cortical complexity, shape indices, volume and surface area of the segmented structures. We then represent these quantities in the space of real numbers and adopt the standard Euclidean metric. As a specific example, for a database with N subjects and L delineated sub-volumes, given by $\{V_k^i, i = 1, \dots, N, k = 1, \dots, L\}$ we first normalize all the volumes to have a unit scale. We then calculate a N -by- N distance matrix given by,

$$D(S_i, S_j) = \frac{\sum_{k=1}^L (V_k^i - V_k^j)^2}{L} \quad (i, j = 1, \dots, N, i \neq j) \quad (1)$$

Likewise, one can in practice utilize different metrics to generate different distance matrices. In order to establish a frame-of-reference, we construct a template atlas for the population and also include it

in the distance calculation. This process yields a distance space of neuroanatomical structures with the atlas conveniently treated as the "origin". One can even define distance units in this space and define a centralized coordinate system with the atlas at the origin. This resulting distance space is extremely high dimensional and not straightforward to visualize. In the analysis stage, we will pre-store multiple such distance matrices based on different metrics for the above features in the database. As new data enters the LONI IDA, the workflow engine will automatically detect new entries and subject them to regional extraction, surface modeling, and regional measurement, etc. Random spot checking to ensure accuracy will help to reduce improper data from entering into and possibly biasing the comparison of image data-sets. Each new data set's relative distance from each of the brain volumes already in the overall distance matrix will be performed and this new information will take its place in the matrix. Upon updating of the distance matrices, the multidimensional scaling (MDS) will then be recomputed and the positions of each brain surface in the space adjusted accordingly. We expect that once fully deployed the continuous processing of new entries into the IDA and the updating of the geometric similarities will not require extensive computational loads or interfere with other jobs being processed on the LONI grid. Lastly, we will post the automated processing meta-algorithm via the community web forum so that others may download the workflow and use the LONI Pipeline on their own systems to validate results. Using the online web forum as well as client-side user interface tool, users will be able to post their reviews of the validity of meta-algorithm, note outlier subjects, or annotate interesting cases. These publicly given annotations will form additional meta-data information to be made available to other users of these tools.

In order to explore the dissimilarities between brain volumes, we need to project the dissimilarity matrix into an appropriate 2D or 3D space. There are numerous techniques to project high dimensional data into lower dimensional spaces for analysis or visualization. As discussed above, one could calculate principal coefficients, or principal factors explaining the maximum observed population variability in terms of a few determining factors. For visualization purposes, only the first 3 eigen projections can be used to display objects in a 3D space. Sophisticated visualization tools (Swayne et al., 2003) exist for performing such high dimensional data visualizations, as well as plotting multivariate statistics of the data. However, these tools usually represent objects by points in 3D space and thus limit the interaction with the original objects themselves. Moreover, since the construction of our meta-space relies on dissimilarities among neuro-structures, we will use the multidimensional scaling (Kruskal and Wish, 1978) approach for projecting the dissimilarity matrix into a 3D space. Multidimensional scaling is an optimization technique that projects a high dimensional dissimilarity matrix into a low dimensional space that most accurately represents the pair-wise distances between the objects. This is achieved by minimizing a cost function that minimizes a Euclidean cost between the original dissimilarity matrix and a set of low dimensional (3D in our case) vectors. Additionally, since most studies come equipped with meta-data tags along with the images, one can easily perform comparative analyses of individual brain locations with the mean brain locations for each categorical meta-data type. For example, in case of an Alzheimer's study, this implies that a brain whose standardized distance from the mean

AD patient location is smaller than the normal subject will increase the likelihood that the brain belongs to an AD patient. **Figure 3** shows an illustrative visualization of the meta-space after the MDS projection of pair-wise distances between a group of brains with respect to an atlas.

Case study for a subset of ADNI dataset

As a case study for our framework, we sampled the LONI IDA and identified three groups of subjects obtained from the ADNI dataset. Subjects included $N = 244$ mild cognitively impaired (MCI) subjects, $N = 56$ Alzheimer's Disease (AD) patients, and $N = 100$ normal control subjects, for an overall group of $N = 400$ neuro-anatomical Magnetization-Prepared Rapid Acquisition Gradient-Echo (MPRAGE) image volumes. All images in this example were

scanned using a 3T MR scanning platform, although, a mixture of data from across scanner manufacturers and field strengths would also be possible. The distance matrix computed using Eq. 1 is visualized in **Figure 4A**. This distance matrix is projected to a three dimensional space using MDS and the results is displayed in **Figure 4B**. The spheres are drawn with a radius equal to 5% of the standard deviation from the population mean. From the MDS analysis, we extracted the first three latent dimensions which accounted for more than 66% of the distance variation between subjects (50%, 10%, and 6%, respectively). No inferential statistical test thresholding (e.g. T -tests, F -tests, etc) was performed and no significance-levels were determined concerning the differences between groups as "group" variables were not specified *a priori*. Rather, all data were considered equally in terms of processing

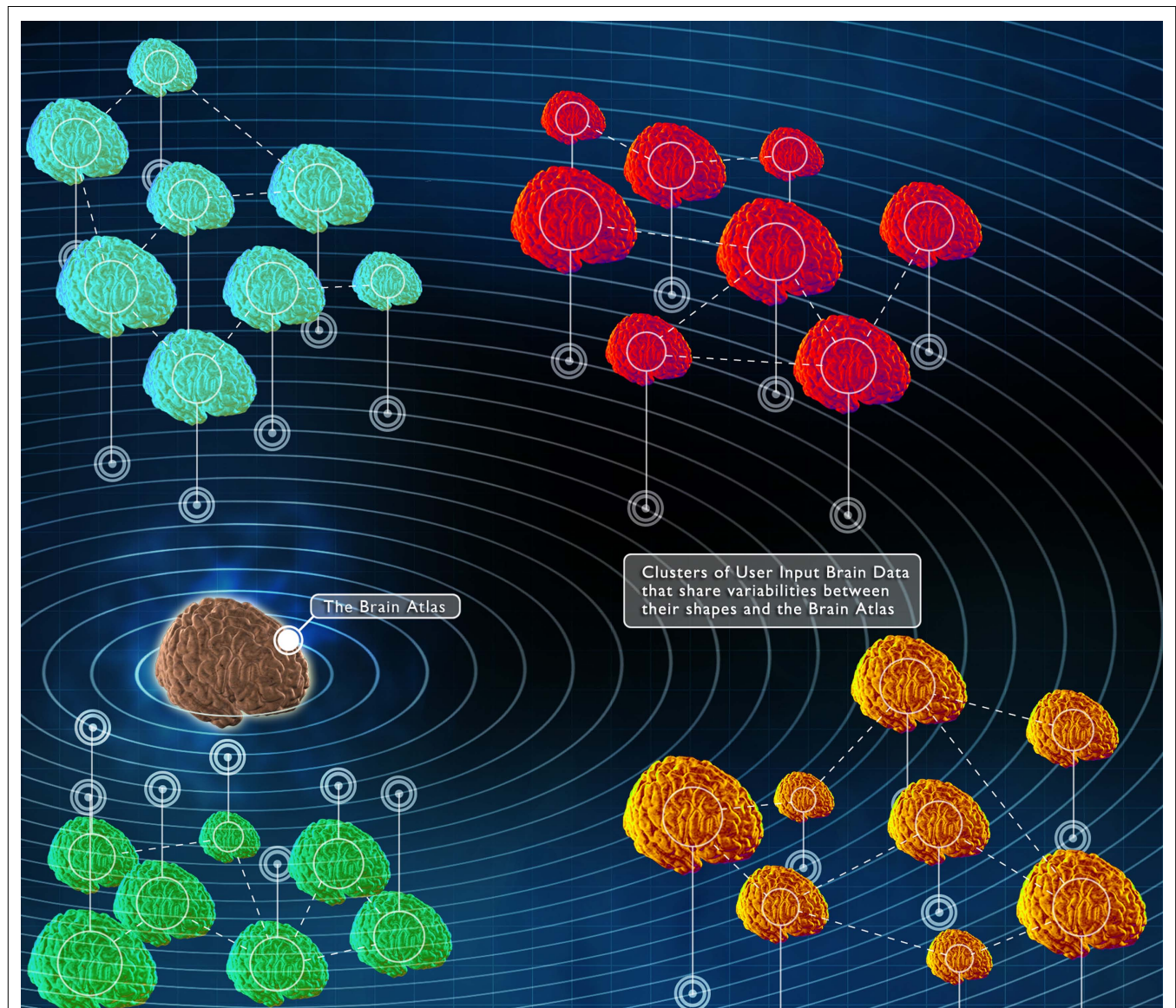
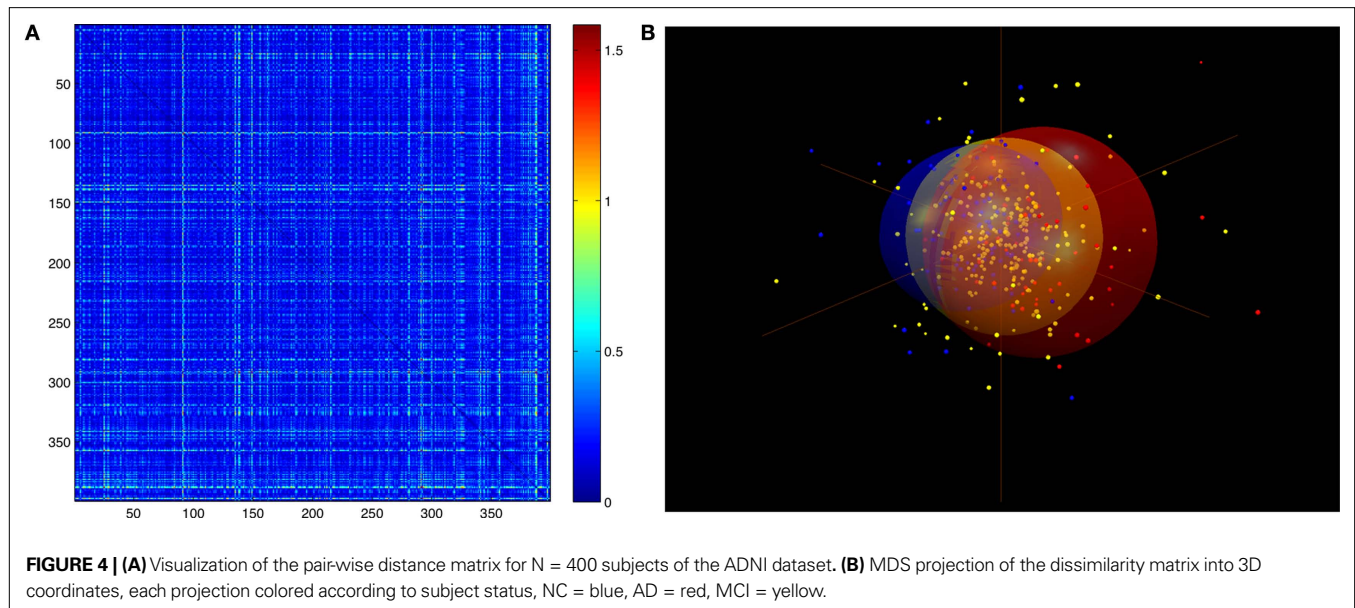


FIGURE 3 | An illustration of distributions of brain surfaces in an atlas meta-space. The atlas can be treated as the origin. The locations of the brain surfaces are derived using MDS applied to the distance matrix of discriminative features. A radial coordinate system is shown for convenience, in practice any other informative reference frame can be used.



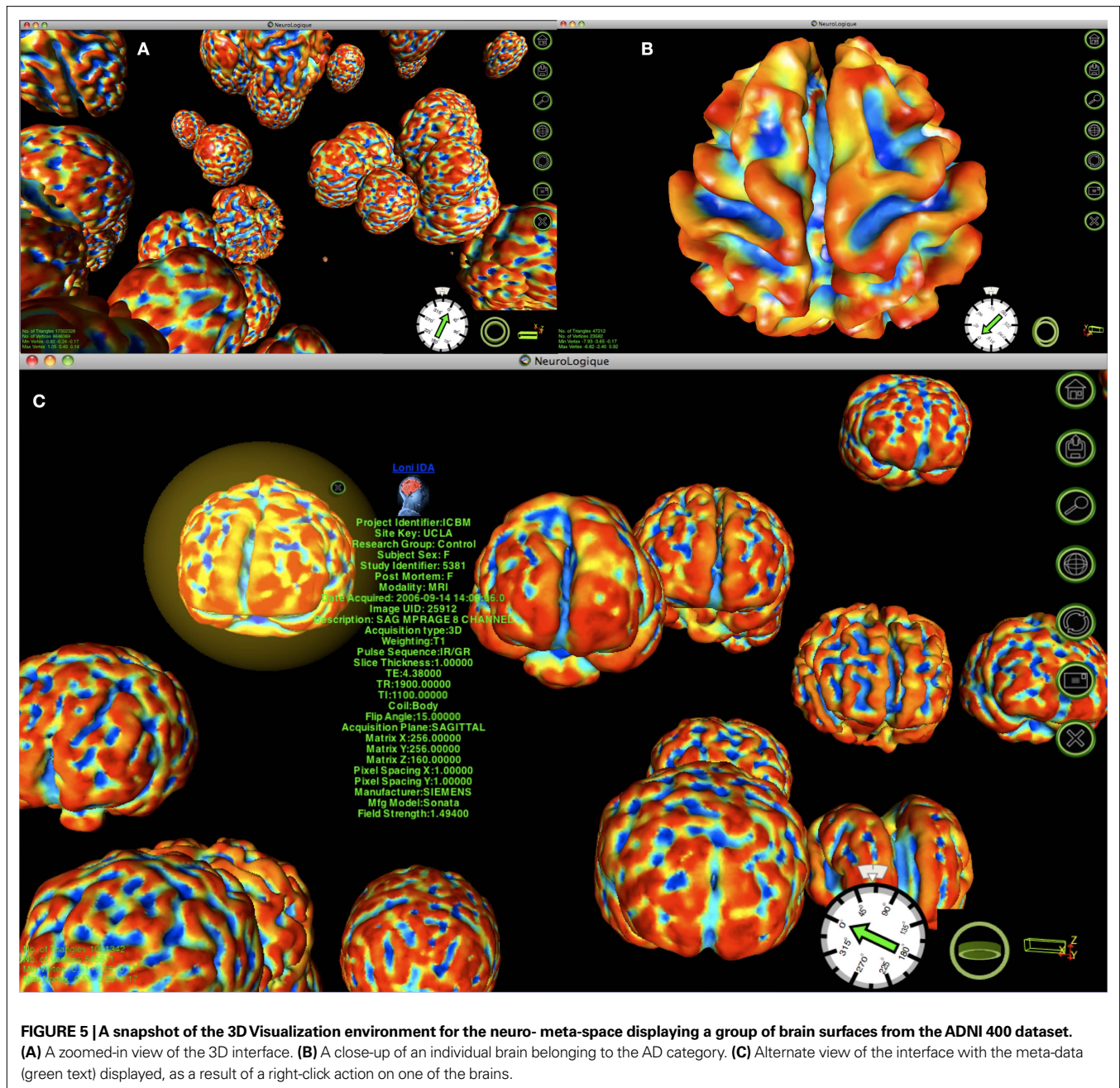
and MDS analysis. Subjects segregated maximally along the first principle axis. Along this dimension, normal subjects were clearly distinguishable from AD patients, whereas MCI patients were observed to overlap both normal and AD distributions. Each extracted brain surface was positioned in space by its MDS coordinate triad. This served to offset each brain from the origin, to position brains that are similar near one another, and those that are dissimilar far from each other. In this way a user can graphically examine similar brains to identify similar meta-data characteristics from those brains that might have bearing on etiology of disease, demographic factors, etc. The intention in processing these data in this manner was to determine and demonstrate whether the imaging data, based upon the characteristics of content-based regional brain geometry, would separate themselves in a manner that would be obvious to an end-user. Other metrics, however, besides the profile of regional volumes, might also segregate subjects equally well or even better. Different metrics of distances can lead to differing patterns of results. This scheme sets the stage for systematic evaluation of which metric discriminate between subjects best, best express individual variability, or classify subjects to heretofore unappreciated classifications based on anatomical similarity, meta-data factors, etc.

INTERACTIVE VISUALIZATION OF NEURO- META- SPACES

Finally, we provide the user with a fully interactive 3D exploration experience by allowing visual navigation of the meta-space. As seen earlier, the meta-space comprises of 3D projections of pair-wise distances among the population along with the atlas template. Intuitively, this can also be thought of as scaling, stretching and collapsing a set of 3D points (corresponding to brain anatomies) such that their pair-wise distances in the higher dimensional space are accurately approximated. Thus we have the target locations for all brain anatomies after the MDS procedure converges. We now simply scale and translate the corresponding brain cortical surfaces extracted in the data mining workflows to the appropriate location

in the 3D space. The end result is a graphical rendering of a large volume of brain surfaces all at once. The visualization display is dynamic, thus enabling the user to rotate, zoom, and pan the view in real time. Additionally, the user can also navigate through the meta-space, thus discovering and verifying the brain surface geometry simultaneously in relation to its neighbors. Each brain surface is accompanied by an XML description of its meta-data that can be quickly displayed on the screen to get more information about the individual brain.

A growing challenge to the visualization environment is the rapidly accumulating data. Both long-term storage and memory requirements for data multiply progressively with increase in the sheer data volume. Real time visualization of large data-sets presents numerous difficulties with regards to limited processing power and computer memory. For e.g. a triangular mesh parameterization of a moderate resolution cortical surface roughly includes 250 K triangles and 100 K vertices. A floating point representation for the geometry alone requires about 1.5 MBytes of storage, while attributes such as colors and normals are represented at an additional cost. For a brain volume database in excess of 500 brains, the storage requirements start becoming prohibitive for any real time manipulation of data. For this reason, it is necessary to represent the data in a multi-resolution manner. There is an ongoing research effort in the area of triangular mesh simplification for visualization or compression for storage purposes. For our visualization interface, we have implemented the quadric error mesh simplification strategy (Garland and Heckbert, 1997) that keeps on contracting edges defined by vertex-pairs until the desired number of faces are achieved. The multi-resolution representation and rendering enables faster response times, and facilitates better user interaction. Currently the surface geometry is stored as triangular meshes with faces, vertices, and colors. We also allow surfaces to be colored according to attributes for each vertex. These can represent measures such as cortical thickness, functional activity, or other statistics. **Figure 5**



shows the functioning prototype of our visualization interface. The visualization environment is a desktop-based application designed in C++ and Open GL[®] and is available on all Windows, Mac OS X, and Unix-based platforms. The OpenGL pipeline conveniently provides a built-in framework for polygonal rendering and transformations.

DISCUSSION

We foresee the development of graphical visualization tools that enable and enhance scientific interaction with large-scale databases, as the next step in neuroimaging informatics. Though some basic image viewing tools exist, we have argued for a need for a next gen-

eration visual interaction framework. We have also demonstrated a content-based solution that can be applied to any such archive in order for researchers to more easily examine dissimilarity between brains and to dynamically visualize patterns in the degree of proximity between brains that may be indicative of the demographic and clinical attributes of the data themselves. In fact, all throughout our approach, we have made as few assumptions about the data as possible, and really let the data segregate itself based upon the characteristics of regional shape and geometry. A key component of this framework is the fully interactive, 3D visualization environment. By navigating through a virtual environment via an easy-to-use, web driven application, users will be able to examine large collections of

brain data using only their computer mouse. The underlying data distribution manifested through classification and collocated with the respective brain anatomies would be a very valuable tool for data processing, mining and interactive visualization of large-scale neuroanatomical databases. This will form a common frame-of-reference for neuroimaging informatics that is (a) familiar to most neuroimaging scientists, (b) provides a navigable space in which to position brain data, and (c) allows measurement of brain dissimilarity to be visually represented.

Our plan now is to (i) apply this meta-workflow to the thousands of MR anatomical images contained in the LONI IDA to obtain cortical surface and partition shape statistics, (ii) measure the pair-wise distances between the shapes obtained from the individual MR volumes, (iii) apply multidimensional scaling (MDS) and related decompositions of the matrix of pair-wise distances to determine which brains are most related, and (iv) broaden the concept of the standard brain atlas space to extend beyond the boundaries of the atlas to form a large space, analogous to a celestial coordinate system, wherein the atlas is centered at the origin and the individual

brain surface representations are distributed in clusters with respect to it. We also plan on enhancing the user interface and scaling its performance with the increasing data.

ACKNOWLEDGMENTS

This research was partially supported by the National Institute of Health through the National Center for Research Resources (NCRR) for Center for Computational Biology (CCB), Grant U54 RR021813. Additional support was provided by Award Number RC1MH088194 from the National Institute of Mental Health. This study utilized the LONI Pipeline environment (<http://pipeline.loni.ucla.edu>), which was developed by the Laboratory of Neuro Imaging and partially funded by NIH grants P41 RR013642, R01 MH71940 and U54 RR021813. Additionally, we thank Mr. Vaughn Greer, Laboratory of Neuro Imaging for the graphic in **Figure 3**. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

REFERENCES

- Callahan, M., Cole, M. J., Shepherd, J. F., Stinstra, J. G., and Johnson, C. R. (2009). A meshing pipeline for biomedical computing. *Eng. Comput.* 1, 115–130.
- Chen, J., Zheng, T., Thorne, W., Zaiane, O. R., and Goebel, R. (2007). Visual data mining of web navigational data. In *IV'07: Proceedings of the 11th International Conference Information Visualization*, Washington, DC, IEEE Computer Society, pp. 649–656.
- Chupin, M., Chetelat, G., Lemieux, L., Dubois, B., Garnero, L., Benali, H., Eustache, F., Lehericy, S., Desgranges, B., and Colliot, O. (2008). Fully automatic hippocampus segmentation discriminates between early Alzheimer's disease and normal aging. In *the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008*. ISBI 2008., pp. 97–100.
- Dinov, I., Van Horn, J. D., Lozev, K. M., Magsipoc, R., Petrosyan, P., Liu, Z., MacKenzie-Graham, A., Eggert, P., Parker, D. S., and Toga, A. W. (2009). Efficient, distributed and interactive neuroimaging data analysis using the LONI pipeline. *Front. Neuroinform.* 3, doi:10.3389/neuro.11.022.2009.
- Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., and Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* 8, 734–746.
- Eckersley, P., Egan, G., DeSchutter, E., Yiyuan, T., Novak, M., Sebesta, V., Matthiessen, L., Jaaskelainen, I., Ruotsalainen, U., Herz, A., Hoffmann, K.-P., Ritz, R., Ravindranath, V., Beltrame, F., Amari, S.-i., Usui, S., Lee, S.-Y., van Pelt, J., Bjaalie, J., Wrobel, A., daSilva, F., Gonzalez, C., Grillner, S., Verschure, P., Dalkara, T., Bennett, R., Willshaw, D., Koslow, S., Miller, P., Subramaniam, S., and Toga, A. (2003). Neuroscience data and tool sharing. *Neuroinformatics* 2, 149–165.
- Evans, A. C. (2006). The NIH MRI study of normal brain development. *Neuroimage* 1, 184–202.
- Forsberg, L. and Roland, P. (2008). 1st incf workshop on neuroimaging database integration. In *Nature Precedings*. available at: <http://dx.doi.org/10.1038/npre.2008.1781.1/>
- Garland, M. and Heckbert, P. (1997). Surface simplification using quadric error metrics. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, New York, NY, ACM Press/Addison-Wesley Publishing Co, pp. 209–216.
- Gosche, K. M., Mortimer, J. A., Smith, C. D., Markesbery, W. R., and Snowdon, D. A. (2002). Hippocampal volume as an index of Alzheimer neuropathology: Findings from the nun study. *Neurology* 10, 1476–1482.
- Grethe, J. S., Baru, C., Gupta, A., James, M., Ludaescher, B., Martone, M. E., Papadopoulos, P. M., Peltier, S. T., Rajasekar, A., Santini, S., Zaslavsky, I. N., and Ellisman, M. H. (2005). Biomedical informatics research network: building a national collaboratory to hasten the derivation of new understanding and treatment of disease. *Stud. Health Technol. Inform.* 112, 100–109.
- Hao, M. C., Dayal, U., and Hsu, M. (2000). Visual data mining for business intelligence applications. In *WAIM '00: Proceedings of the First International Conference on Web-Age Information Management*, London, Springer-Verlag, pp. 3–14.
- Herskovits, E. H., and Chen, R. (2008). Integrating data-mining support into a brain-image database using open-source components. *Adv. Med. Sci* 2, 172–181.
- Hinds, O., Polimeni, J. R., Rajendran, N., Balasubramanian, M., Wald, L. L., Augustinack, J. C., Wiggins, G., Rosas, H. D., Fischl, B., and Schwartz, E. L. (2008). The intrinsic shape of human and macaque primary visual cortex. *Cerebral. Cortex* 11, 2586–2595.
- Hurd, N. A. (2005). Neuroscience Information Framework.
- Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Trans. Vis. Comput. Graph.* 1, 1–8.
- Keim, D. A., Panse, C., Sips, M., and North, S. C. (2004). Visual data mining in large geospatial point sets. *IEEE Comput. Graph. Appl.* 5, 36–44.
- Kruskal, J. B., and Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA, Sage Publications.
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N.,
- LeGoualher, G., Boomsma, D., Cannon, T., Kawashima, R., and Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 1412, 1293–1322.
- Narr, K. L., Bilder, R. M., Toga, A. W., Woods, R. P., Rex, D. E., Szeszko, P. R., Robinson, D., Sevy, S., Gunduz-Bruce, H., Wang, Y., DeLuca, H., and Thompson, P. M. (2005). Mapping cortical thickness and gray matter concentration in first episode schizophrenia. *Cerebral. Cortex* 6, 708–719.
- Nielsen, F. A., Balslev, D., and Hansen, L. (2006). Data mining a functional neuroimaging database for functional segregation in brain regions. In *Danske Konference i Mønstergenkendelse og Billedanalyse*.
- Rex, D. E., Ma, J. Q., and Toga, A. W. (2003). The LONI pipeline processing environment. *Neuroimage* 3, 1033–1048.
- Shattuck, D. W., and Leahy, R. M. (2002). BrainSuite: an automated cortical surface identification tool. *Med. Image Anal.* 2, 129–142.
- Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 3, 143–155.
- Souza, T., Kush, R., and Evans, J. P. (2007). Global clinical data interchange standards are here! *Drug Discov. Today* 3–4, 174–181.
- Swayne, D. F., Buja, A., and Lang, D. T. (2003). Exploratory visual analysis of graphs in ggobi. In *CompStat: Proceedings in Computational Statistics*, 16th Symposium, pp. 477–488.

- Szalay, A., and Gray, J. (2001). The World-Wide Telescope. *Science* 5537, 2037–2040.
- Thompson, P. M., Mega, M. S., Woods, R. P., Zoumalan, C. I., Lindshield, C. J., Blanton, R. E., Moussai, J., Holmes, C. J., Cummings, J. L., and Toga, A. W. (2001). Cortical change in alzheimer's disease detected with a disease-specific population-based brain atlas. *Cerebral. Cortex* 1, 1–16.
- Tu, Z., Narr, K. L., Dollar, P., Dinov, I., Thompson, P. M., and Toga, A. W. (2008). Brain anatomical structure segmentation by hybrid discriminative/generative models. *IEEE Trans. Med. Imaging* 4, 495–508.
- Van Horn, J. D., and Gazzaniga, M. S. (2002). Opinion: Databasing fMRI studies towards a 'discovery science' of brain function. *Nat. Rev. Neurosci.* 4, 314–318.
- Van Horn, J. D., Grethe, J. S., Kostelec, P., Woodward, J. B., Aslam, J. A., Rus, D., Rockmore, D., and Gazzaniga, M. S. (2001). The functional magnetic resonance imaging data center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 1412, 1323–1339.
- Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., and Mazziotta, J. C. (1998). Automated image registration: I. general methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr.* 1, 139–152.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 02 April 2009; paper pending published: 24 June 2009; accepted: 05 October 2009; published online: 06 November 2009. Citation: Joshi SH, VanHorn JD, and Toga AW (2009) Interactive exploration of neuroanatomical meta-spaces. Front. Neuroinform. 3:38. doi: 10.3389/neuro.11.038.2009 Copyright © 2009 Joshi, Van Horn, and Toga. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.*