The NLM-Mayo Image Collection: Common Access to Uncommon Data

David R Holmes III*, Ellis P Workman, and Richard A Robb

Biomedical Imaging Resource, Mayo Clinic College of Medicine

Abstract. For over two decades, the National Library of Medicine (NLM) has provided support for the collection of biomedical image data for use throughout the biomedical image, visualization and analysis community. Data collected during the Visible Human Project has been utilized to advance development of medical image research, education, and other ventures. One goal of our research has been to further diversify such image data and make it openly available to the biomedical imaging community. The approach is to provide open access to a diverse collection of biomedical image data that can be used for the development and validation of new image processing and analysis techniques. With support from NLM, over 100 datasets were incorporated into the NLM-Mayo data collection. There is variation in species, anatomy, pathology, scale, and modality. In addition to providing "clinical quality" medical image data, the collection also includes newly acquired datasets of several animals, including a whole mouse with both μCT and μMR data volumes. This unique collection of data was categorized and organized into an intuitive web-based browser which allows a user to rapidly access descriptive information as well as the actual data volumes. The data collection will be made available by the NLM for distribution, vis-a-vis its Visible Human Project (VHP). Because the landscape of biomedical imaging continues to change with new, advanced image acquisition systems and techniques, continuously updating the VHP data collection seems prudent. Additional new and varied image data will be incorporated into our collection and disseminated to researchers in the medical image analysis community. This work was supported through a contract from the National Library of Medicine (NLM-HHSN276200443502C).

1 Introduction

Almost 20 years ago, the National Library of Medicine recognized the potential need for a large digital archive of medical image data. Approximately 10 years ago, Dr Michael Ackerman published a report on "Accessing the Visible Human Project" – a high resolution digital collection of medical image data of a single human (male) specimen [1]. Subsequently, the data collected under the Visible Human Project(VHP) was expanded to a second human (female) and these datasets have been utilized by numerous investigators throughout the

---

* `holmes.david3@mayo.edu`

world for many different applications. The VHP project also provided impetus for the development of the Insight Toolkit [2] which is an open-source API for biomedical image processing algorithms. In the same way that open-source software provides a common platform for algorithm development, open access image collections provide a common cohort of data to validate and compare image analysis methods. For example, one study, commonly referred to as the "Vanderbuilt Registration Study," provided the same datasets to 12 different investigators/institutions for comparison of registration methods[3]. The goal of our project was/is to collect and organize a diverse set of biomedical image types into an open access data collection for use by the biomedical image analysis community.

There are several other noteworthy data collection projects that have been completed or are currently underway. The purpose of each such project varies, however. In most cases, these other image collections have been helpful for developing and validating new algorithms, techniques, and processes. Access to these data collections range from completely open to very restrictive. Similar to the VHP project, Korean and Chinese researchers have developed the Visible Korean Human[4] and Chinese Visible Human[5], respectively. These datasets enhance the collection of whole body datasets by increasing the diversity in the data cohort. Other collection projects have targeted specific anatomy, such as the Biomedical Informatics Research Network project which focuses on the brain [6]. The Montreal Neurological Institute Brainweb project also focuses on brain anatomy [7], however access to the data is through derivatives of the raw medical image data. The UC Davis Brain Atlas Project goes beyond human anatomy and provides animal brain data as well[8].

Our goal has been to further diversify available image data collections and make them fully open to the biomedical image analysis community. In contrast to most other image data collection projects, our project focuses on variety, scope and multi-modal image datasets. Our thesis is that the data collection must be provided in a consistent and completely open manner that can be used by any investigator in the field of biomedical imaging. We believe that the NLM-Mayo image data collection will facilitate many different investigations and applications, including development and validation of new image processing algorithms (e.g., for segmentation, registration, visualization, etc.), but also for comparison and optimization of different algorithms in the context of specific image processing tasks and goals.

2 Methods

In order to provide a useful image collection with broad appeal to many different investigators, we attempted to determine appropriate inclusion criteria. Comprehensive data should span from the microscopic to macroscopic. Many different anatomic sites should be represented in the data; both normal and pathological data should be included. In the case of human data, it is important that the data be of "clinical quality" whenever possible. With the continued interest in animal imaging for the study of patho-physiologic conditions in biological systems, animal image datasets are important additions to a comprhensive collection. Finally, multi-modal data is desirable due to the many multi-spectral classification, registration and fusion algorithms designed for use with multi-modal image data (such as the brain epilepsy analysis and treatment method described in [9]).

2.1 Data Collection and Categorization

We took three approaches to image data collection – incorporation of existing relevant and high quality data, necessary conditioning of existing data, and acquisition of new data. Because of the extensive history of the BIR and associated collaborators, much of the data was retrieved from the BIR archives as well as the archives of our collaborators. Some of the collected datasets have been investigated within a given research group but have not before been openly available to the medical image analysis community. In some cases, the type of data (or acquisition methods) have not been available to other researchers in the field due to unique imaging capabilities of a given institution. One such example is dynamic heart, lung, and circulatory image data collected with the Dynamic Spatial Reconstructor [10]. This data, collected in the 1980s and 1990s, remains unique due to the custom high performance scanner used to collect it.

In addition to existing normal and especially obtained medical image data with pathology, sometimes requiring special conditioning (e.g., reconstruction, formatting and organizing), we determined that the data collection project would benefit greatly from new image data which can be acquired from high resolution small-animal imaging systems. A whole mouse was scanned with both μCT and μMR at voxel resolutions of 0.02mm^3 and 0.125mm^3, respectively. Because of the scale at which this data was acquired, it was necessary to collect and organize the whole dataset into several individual volumes of data. Data volumes were acquired in overlapping sections to allow for realignment into a single volume dataset.

The images collected for our project were categorized in several ways. After dividing the datasets into human and animal types, both categories were further indexed by anatomical region. Within each anatomical region, datasets were differentiated by modality and by pathology. In several cases, multi-modal data was available for individual subjects.

2.2 Data Management and Presentation

The data obtained in the NLM-Mayo image collection project was acquired using a variety of techniques and imaging hardware. The original file formats of the data varied greatly. In order to provide the data in a common and easily accessible manner, three data formats were chosen for data storage within the collection – Analyze 7.5, Analyze Volumefile, and MetaIO. These formats were selected because they are widely recognized and utilized throughout the biomedical image research community. In addition, each of these methods can use a two file approach in which one of the files contains header information while the other file contains the raw image data. Accordingly, each dataset contains three different header files and one raw image file. A description of the format of the three header files can be found in [11–13].

All data for this project were converted using the AVW medical image processing library developed by the BIR[14]. The library currently can recognize and convert among 30 different common file formats, as well as several less known, but useful, formats. During the conversion process, all patient data was anonomized according to pre-defined guidelines and HIPPA standards. With the exception of volumetric reconstruction, the data was maintained in its raw data form whenever possible. Unprocessed (e.g., not interpolated) data was unavailable in some cases. Image parameters (such as voxel resolution, etc.) were stored within the header record. Additional information about the imaged anatomy and pathology was also recorded.

Data was organized hierarchically according to the classifications defined during the collection. Representative images were collected and combined with the image parameters and pathological conditions in order to provide an anectdotal text description of each dataset. The data was organized into a website and a special intuitive browser developed which allows rapid access for review and downloading of individual datasets.

3 Results

An overview of the image data in the NLM-Mayo collection is provided in Table 1. Over 100 different datasets were collected and categorized into 59 groups for the project. In several cases, there is multi-modal data for the same subject, including one patient study containing 5 different modality datasets.

Table 1

Overview of image data in the NLM-Mayo data collection. The collection includes both normal and pathological datasets.

Species	Human Anatomy	Modalities
Human	Abdomen	MR (MRA, T1, T2)
Canine	Brain	CT (3D, 4D, contrast-enhanced)
Dolphin	Cells	Confocal Microscopy
Mouse	Chest	PET
Rabbit	Ear	SPECT
	Hand	Ultrasound
	Heart	μCT
	Knee	μMR
	Larynx	
	Liver	
	Prostate	

Snapshots of the interface of the web-based browser are shown in Figure 1. Each dataset is indexed and can be rapidly accessed by clicking on links in the left hand column. Details of the datasets are shown on the right, including examples of 2D sections and 3D renderings.

Torn Anterior Cruciate Ligament
MR knee scan showing a torn ACL.

Location: /NLMDATA/Knee/TornACL/MR_TornACL.img (2.3 MB)

Volume Geometry
Volume Dimension:
- 256 x 256 x 18

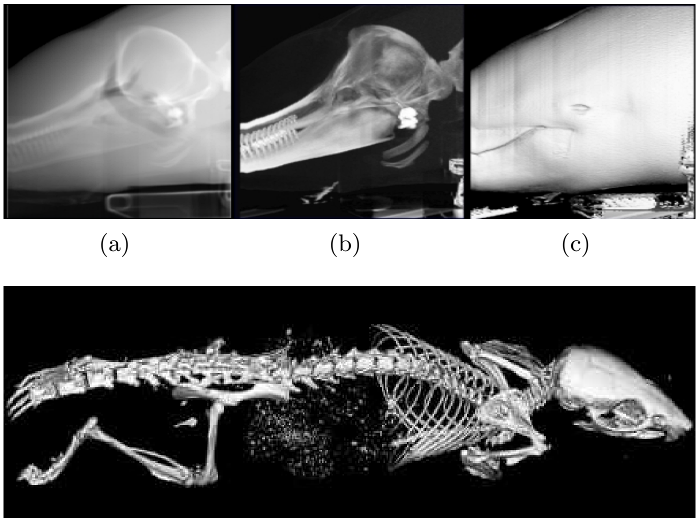
Voxel Dimension:
- 0.625 mm x 0.625 mm x 5 mm

Data Maximum Value: 2364
Data Minimum Value: 0
Pixel Data Type: 16-bit signed
Plane of Acquisition: Transverse

(a) (b)

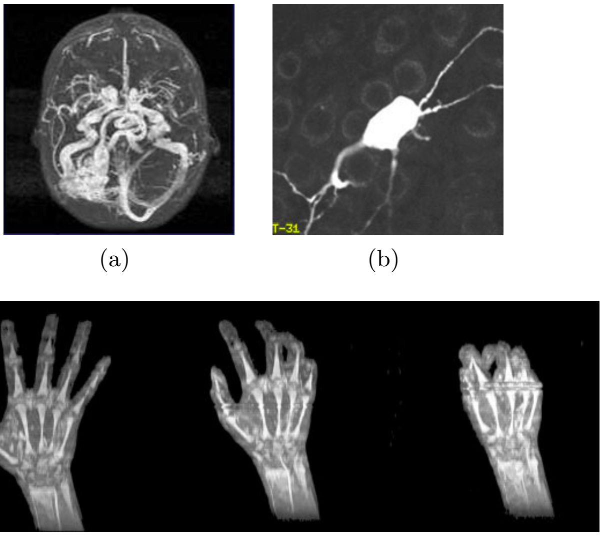
Fig. 1. Web-based interface to the NLM-Mayo data collection. In the top center of the page, links point to overview information about the project. In (a), the left frame contains a column of indexed links to detailed descriptions of all of the different datasets. The main frame is used to display images from each of the datasets. In (b), a representative example is shown of the content within the selected dataset. In addition to providing text details describing the dataset, several images are included for visual reference, both 2D and 3D.

Several unique image volumes are included within the collection. Figure 2 shows some of the animal data. In Figure 2(a)-(c), renderings of a dolphin head are shown. In addition to high-resolution CT data , SPECT data of the dolphin was acquired. Two such dolphin heads are in the collection. In Figure 2(d) a rendering of the whole mouse from μCT is shown. Rendering this data is challenging due to the large number of data samples in the volume. Within the human data, several different types of images are represented. Figure 3 shows more example images from the collection. Figure 3(a) was generated from a MR dataset of a patient with a large Arteriovenous Malformation (AVM). Figure 3(b) is confocol microscopy data of a single mesenteric neuron. Figure 3(c) shows three renderings of an MRI 4D dataset of a grasping hand.


(a) (b) (c)

(d)

Fig. 2. Renderings illustrating animal data in the NLM-Mayo collection. In (a-c), different renderings are shown of a dolphin head from a CT scan. In (d), the skeleton of a whole body mouse is rendered (after thresholding) from μCT data.



(a) (b)

(c)

Fig. 3. Image examples from the NLM-Mayo collection of human anatomy with pathology, cells, and motion. A large AVM from CT is shown in (a). Confocal microscopy data of a single mesenteric neuron is shown in (b). Renderings of an MRI 4D dataset of a moving hand are illustrated (c).

4 Discussion

The purpose of this project was to collect a number of diverse biological and medical image datasets for public use in the evaluation of biomedical image processing and analysis methods. The selection and inclusion process included consideration of acceptance criteria. As with any type of data, images should be representative of the relevant population (whenever possible). They should be varied across several parameters including type, location, scale, resolution, anatomy, physiology, pathology and modality. They should be of sufficient quality to be a surrogate for typical clinical data. We have been able to achieve this goal by relying on the extensive archives and image acquisition facilities of the BIR and its collaborators around the world who have been collecting for many years both structural and functional multi-dimensional multi-modality image data. In addition, the acquisition of new data, such as high-resolution multi-modal whole mouse images, dolphin brain images, human physiology and pathology images, and microscopic cell volume images, provide unique and unusual data for testing of image processing algorithms being developed for a variety of scientific and/or clinical purposes.

The collection and consolidation of "open" biomedical image data is an ongoing process. New imaging methods will be developed and new advanced hardware will be built. It is important that projects like the VHP continue in order to provide an extensive up-to-date and relevant database of images that researchers in the field can draw upon for the design, validation and refinement of new algorithms. As this collection grows, it will be critically important to provide a formal schema for the classification of data which will clearly cross-reference type, anatomy, physiology, biology, pathology, scale, and modality. Adequate resources should be made available for common storage and rapid distribution of the image data collection.

5 Acknowledgments

This project involved many internal and external collaborators of the Mayo Biomedical Imaging Resource. Their contributions enhanced the variety and quality of the data collection. The generosity of these collaborators reflects the interest and support of the biomedical imaging community for the open source/data approach to research. We recognize and thank the National Library of Medicine for their financial support of this effort (Contract NLM-HHSN276200443502C).

References

1. Ackerman, M.: Accessing the visible human project. D-lib Magazine **Oct** (1995)
2. Yoo, T., Ackerman, M., Lorensen, W., Schroeder, W., Chalana, V., Aylward, S., Metaxes, D., Whitaker, R.: Engineering and algorithm design for an image processing api: A technical report on itk - the insight toolkit. In Proc. of Medicine Meets Virtual Reality (2002) 586–592

3. West, J., Fitzpatrick, J., Wang, M., et al: Comparison and evaluation of retrospective intermodality brain image registration techniques. Journal of Computer Assisted Tomography **21(4)** (1997) 554–566
4. Park, J., Chung, M., et al.: Visible korean human: Improved serially sectioned images of the entire body. IEEE Trans Med Imaging **24(3)** (2005) 352–360
5. Zhang, S., Heng, P., et al.: Creation of the chinese visible human data set. The Anatomical Record **275B** (2003) 190–195
6. Grethe, J., Baru, C., Gupta, A., James, M., et al: Biomedical informatics research network: building a national collaboratory to hasten the derivation of new understanding and treatment of disease. Stud Health Technol Inform **112** (2005) 100–9
7. Montreal Neuologic Institute: Brainweb. (http://www.bic.mni.mcgill.ca/brainweb/)
8. UC Davis: The brain atlas project. (http://nir.cs.ucdavis.edu/)
9. Brinkmann, B., OConnor, M., OBrien, T., Mullan, B., So, E., Robb, R.: Dual-isotope spect using simultaneous acquisition of 99mtc and 123i radioisotopes: A double-injection technique for peri-ictal functional neuroimaging. Journal of Nuclear Medicine **40(4)** (1999) 677–684
10. Robb, R., Ritman, E.: High-speed synchronous volume computed tomography of the heart. Radiology **133** (1979) 655–661
11. BIR, Mayo Clinic: Analyze 7.5 file format. (http://www.mayo.edu/bir/PDF/ANALYZE75.pdf)
12. BIR, Mayo Clinic: Avw_volumefile format. (AVW Programmers Guide)
13. CADDLab, University of North Carolina: Metaio: Medical image i/o made simple. (http://caddlab.rad.unc.edu/software/MetaIO/MetaIO-Introduction.htm)
14. Hanson, D., Robb, R., et al: New software toolkits for comprehensive visualization and analysis of three-dimensional multimodal biomedical images. Journal of Digital Imaging **10(3 Suppl 1)** (1997) 229–30