Toward Validation Databases for Medical Imaging: Engineering a Scientific Rendezvous

T.S. Yoo

National Library of Medicine, National Institutes of Health, Bethesda, MD, 20894, USA
yoo@nlm.nih.gov

Extended Abstract: The broad questions of content-based retrieval on image archives, segmentation and registration of medical data and computer-assisted diagnosis (CAD) tacitly introduce complex problems of validation. How do we, as a community, provide the necessary technical framework to support the diverse initiatives in image-based medicine? Setting aside the statistical and software engineering aspects of such a foundation, what are the requirements for a medical image database that might support validation of segmentation, registration, CAD, and content-based data retrieval? I will briefly introduce several elements that should be addressed in the evolution of such a resource.

Building Community through the creation of Scientific Rendezvous

There have been several examples of accelerating progress in computer science research through the creation of a concept that I am introducing as a “scientific rendezvous.” The most notable cases have been open-source software projects that have stimulated education, research, and training. The creation of the Unix operating system along with the long-term support of the University of California at Berkeley did more to disseminate advanced operating system design elements among academic and research institutions than any particular text or curriculum [4]. The open architecture of the IBM PC also introduced a common basis over which a community dialog could be conducted. Perhaps the most successful sponsored project that led to the formation of a community dialog and later to an entire industry was the development of internet tools; HTML was the core of the most successful of these tools, and the open nature of HTML opened the way for derivative products such as Netscape Communicator and Internet Explorer. The National Library of Medicine (NLM), at the U.S. National Institutes of Health, has participated in multiple efforts of this nature, including GenBank, the public gene sequence repository supported by he NLM’s National Center for Biotechnology Information. Finally, the Visible Human Project has produced a pair of studies of human anatomy that have provided common ground from which to extend multidisciplinary research in the fields of anatomy, medical imaging, visualization, and other areas of computer science. [7]

The strength of a scientific rendezvous is its ability to promote and sustain dialog for whatever field it supports. The demand for common ground usually leads to the development of standards and conventions; what I am advocating is that before a field is mature enough to require standards, it can benefit from some nucleus around which scientific and engineering conversations can revolve. Principal attributes of the most successful scientific rendezvous are open architecture (software or hardware), extensible infrastructure and control languages, and explicit support for the substance of the core material including distribution mechanisms, open licensing, tutorials and documentation, tools, and bug fixes.

Most of the systems, curricula, and datasets which have become the common basis for information exchange have evolved spontaneously as need and opportunity have converged. They seldom represent good business models. Indeed, the idea of making the components accessible and open places limits on individual control on any element, inhibits the exclusive generation of profit, and invites competition. Anyone planning and building a potential scientific rendezvous should attempt to include as many of the elements of previous successful projects and recognize that neither control nor profit are useful motives or goals in such an endeavor.

Basic Requirements: A Growing Need for Public Data

With the growth of medical computer vision applications, there is an emerging need to supply the research community with a variety of data, organized in multiple ways, to foster scientific growth and sustain academic momentum [6,10]. Clarke has published work in the *accuracy* (as defined by Udupa and Hermann [8]) of MRI segmentation by comparing automatic vs. visual processes [2,3]. Falcao and Udupa have evaluated *accuracy* and *efficiency* of their semi-automatic 2D segmentation methods [5]. In related work, West and Fitzpatrick have created a test suite for validating rigid registration algorithms on multimodal head-and-neck data. Their project provides a very useful public research service [9]. However, these and other validation studies have been conducted using sources of proprietary data.

As medical image analysis continues to mature, the pressure for shared image databases that support research will increase. What is required is a public resource, or several public resources, capable of encouraging discussion and an exchange of ideas between institutions. Although no one can assure success of this type of initiative, several elements can be built into the design to help promote breadth and utility to a broad spectrum of users. Here are some of the basic requirements.

Public data: As with any scientific endeavor, results must be corroborated and compared with the existing body of research. Common image databases will help to foster cross-institution research and evaluation of methods and techniques. It is essential that future test and training datasets be shared among institutions. Heretofore, the trend has always been toward the protection of proprietary data. Public datasets have been available, but never in sufficient quantity nor with sufficient variety to sustain research in medical image processing.

The internet and its many distributed resources enables the sharing of data. Mechanisms for promoting inter-institution collaboration are already in place; it is the existing research mindset and administrative culture that continues to impede collaborative research. Shared datasets may provide a structured, unambiguous interface or locus that will serve as a scientific rendezvous, a meeting place over which many minds may meet. In order for this to happen, however, the database must be public, limit the liability of its contributors and users, and be freely licensed permitting incorporation in publications and research programs.

Longitudinal studies: A common mistake in medical image analysis is to consider the image a single study. From the clinical standpoint, the patient is the single case, and all of the film that contributes to the evolution and resolution of the pathology are the record of the treatment and healing of the subject. While many conditions are transient, there are a greater variety of chronic illnesses for which a single dataset only shows a brief cross-section of the manifestation of the disease in a patient at a single point in time. Computer aided diagnosis, treatment planning, and other image-based research must be evaluated in the context of the disease. In the cases of metastatic disease, chronic conditions, and the general issues surrounding growth and aging, longitudinal studies will have a deep impact on clinical understanding, providing that imaging tools are available and that a broad enough sampling of the populace can be provided. Therefore, the perception of the image database as a series of images should be discarded in favor of the view of the database as a series of patients and their conditions.

Many of the most powerful demonstrations of treatment include "before-and-after" images, showing the results of an intervention or medical treatment. Pre-and-post operative data would be invaluable in gauging the clinical outcome of the effects of any computer-aided diagnosis or computer-based measurement. While quantitative and qualitative metrics (tumor volume, object surface area, calcification density, etc.) can be generated for almost any situation, condition, or application, the clinical outcome is always the final and most important metric. If the analysis and measurement of image features is to be used as a medical decision aid, it is essential that we correlate the features that we measure with their clinical significance. If a treatment is chosen, the results of that treatment should be compared with the same imaging tools to evaluate trends, efficacy, and ultimately the impact on the patient.

Patient information: In keeping with the notion that we are creating databases of patients rather than databases of images, we should strive to collect and retain a broad representation of the patient including information normally acquired during a history and physical examination. Simple information such as height, weight, age, sex, and other gross demographic information should be available. Additional history regarding the use of treatments and the course of any represented pathologies should also be included. Finally, if a diagnosis has been made, a treatment prescribed, a biopsy recommended, or other action taken, the results of that action should appear with the patient data. Outcomes up to and including mortality should be reported in any data repository associated with patient images if that repository is to be used for algorithm validation. The challenge arises when the inclusion of patient information is at odds with the planned public nature of the database.

Beyond the basic requirements

Assuming an open research dialog as the goal, there are other requirements for a validation database. I can only enumerate a few of them here:

- Decentralization of access/control
- Extensible search mechanisms
- Explicit multi-institution support
- Expert Data Correlated with Images
- Training vs. Testing Data
- Renewable Data

There are special considerations when supporting algorithm validation. Truth as well as comparative metrics must be provided. Many algorithms based on statistics or connectionist approaches require copious training data, including expert information or truth associated with each data point. Frequent renewal of this data is essential to keep pace with changing technology and to suppress bias introduced by the finite size of the database. Public data also suggests decentralized access control. Finally maintenance, bug fixes, and support will remain a continuing necessary cost.

Looking toward the future

NLM is working on related initiatives in image processing tools. The Insight project is currently under development. It is a public domain, open-source API which is intended to serve as a repository for algorithms and methods for segmentation and registration. We have a particular interest in validation of these methods, and we are exploring the requirements for validation studies, their metrics, necessary test and training data, and the public dissemination of these resources [1,10]. NLM will continue to grow its procedures and its support for research in this vital area.

Acknowledgements

I would like to thank the NLM Visible Human Project Insight Team (P.I.s: B. Lorensen (GE CRD), W. Schroeder (Kitware), V. Chalana (Insightful), S. Aylward (UNC-CH) D. Metaxas, (U Penn) and R. Whitaker, (UUtah)). Many of the ideas published here reflect the collective thinking of the Insight group.

Selected References

1. Ackerman, M.J., T.S. Yoo, D. Jenkins. 2000. The visible human project: from data to knowledge. Computer Assisted Radiology and Surgery (Proceedings of CARS2000), H. U. Lemke, et al. eds., Elsevier Press, Amsterdam: 11-16.
2. Clarke, L. P., R. P. Velthuizen, S. Phuphanich, J. D. Schellenbeg, J.A. Arrington, M. Silbiger. 1993. MRI measurement of brain tumor response: comparison of visual metric and automatic segmentation. Magnetic Resonance Imaging, 11(1) 95-106.
3. Clarke, L. P., R. P. Velthuizen, M. Clark, J. Gaviria, L. Hall, D. Goldgof, R. Murtagh, S. Phuphanich, S. Brem. 1998. MRI measurement of brain tumor response: comparison of visual metric and automatic segmentation. Magnetic Resonance Imaging, 16(3) 271-279.
4. P. Collinson. 1986. UNIX: The Cult. Proceedings of Usenix 1986. 22-28.
5. Falcao, A.X., J.K. Udupa, F.K. Miyazawa. 2000. An ultra-fast user-steered Image segmentation paradigm: live wire on the fly. IEEE Trans. Med. Imaging. 19(1) 55-61.
6. Gee, J. C. 2000. Performance evaluation of medical image processing algorithms. In: Hanson, K.M., ed., Proc. SPIE Med. Im. 2000, Bellingham, WA:SPIE, 3979, 19-27, 2000.
7. Spitzer, V., M.J. Ackerman, A.L. Scherzinger, and D. Whitlock, 1996, The visible human male: a technical report, J. of the Am. Medical Informatics Assoc., 3(2) 118-130.
8. Udupa, J.K., G.T. Herman, eds. 2000. 3D Imaging in medicine. CRC, Boca Raton, FL.
9. West, J.B., J.M. Fitzpatrick, et al. 1997 Comparison and evaluation of retrospective inter-modality brain image registration techniques. J. Comp. Assist. Tomography. 21. 554-566.
10. Yoo, T.S., M.J. Ackerman, M. Vannier. 2000. Toward a common validation methodology for segmentation and registration algorithms. in: S. Delp, A. DiGioia, B. Jaramaz, eds., Medical Image Computing and Computer-Assisted Intervention (Proc. MICCAI 2000), Lecture Notes in Computer Science, Volume 1935, (Springer-Verlag: Berlin), 422-431.