Concept and Roadmap for Establishing an International Reference Image Database for Medical Image Processing R&D Groups

A. Horsch[1], T. Wittenberg[2], K. Spinnler[2]

[1]Institute of Medical Statistics and Epidemiology (IMSE)
Technical University of Munich, Germany

[2]Fraunhofer Institute for Integrated Circuits (IIS-A)
Erlangen, Germany

ABSTRACT

One of the ‘chronic’ obstacles within the field of medical image processing is the lack of reference image datasets for research and development (R&D) groups, in order to evaluate new methods and applications so that the results are comparable in a reliable, i.e. quantitative way. Whereas the development of basic image processing methods over many years frequently have been evaluated on the famous Lena image [1][2], there have been only few attempts to establish appropriate reference data for the medical field [6][7]. One important reason for this is the much more difficult task: for medical image processing purposes reference datasets from only some representative images to those with up to thousands of images are required for to cover at least the most important medical imaging domains. The recently initiated Working Group on Medical Image Processing of the European Federation for Medical Informatics (EFMI WG MIP) has as one of its goals to develop first parts of a Reference Image Database for Medical Image Processing R&D groups (RID-MIP) during the next three years. The concept of RID-MIP is based on three kinds of datasets: clinical trial datasets, elk test datasets, modality simulator datasets. Due to a draft roadmap the following milestones have to be rolled up: identification and specification of important datasets to be provided, selection and preparation of three highly relevant elk test datasets, dissemination, evaluation of impact on R&D.

1. INTRODUCTION

Worldwide many groups in the large and steadily growing interdisciplinary field of medical image processing are working on different aspects (such as storage, archiving, compression and communication, data standards, image enhancement, segmentation, classification, visualization), with different modalities and in various medical domains. There is an immense variety of research activities, from basic research over clinical evaluation and algorithm design to clinical studies. It can be observed, however, that most of the groups deal with their own images acquired from one or more modalities (e.g. CT, MR, X-ray, endoscopy, microscopy, video). The acquisition, documentation and usage of own image datasets by almost each group leads to the following problems:

- Algorithms developed by a doctoral student, research group or a R&D company in this field are usually based on images taken from one sole (locally available) image acquisition unit (e.g. a Siemens Somatom CT). All these algorithms tend to be optimized solely for *this* machine and can seldom be used on other machines without more or less essential modifications.
- Therefore, the direct comparison of similar research activities (e.g. segmentation of stroke lesions in MRI scans of the brain) is very rarely possible, since each research group has its own, usually incompatible, dataset.
- Image data sets obtained within only one research center do never represent the medical variety needed for large clinical studies (variance in sex, age, compliance, anamnestic data, etc), and are therefore not apt for statistically well-funded clinical statements.
- On the other hand, if the problem of ‘sparse local data acquisition’ is solved by the organization of multi-center studies, these studies usually lack of comparability of the image datasets used, since there are seldom the image acquisition units of exactly the same type available in more than two clinics or hospitals.

In order to improve this situation it seems reasonable to start building an international Reference Image Database for Medical Image Processing R&D groups (RID-MIP) comprising appropriate datasets for the scientific community.

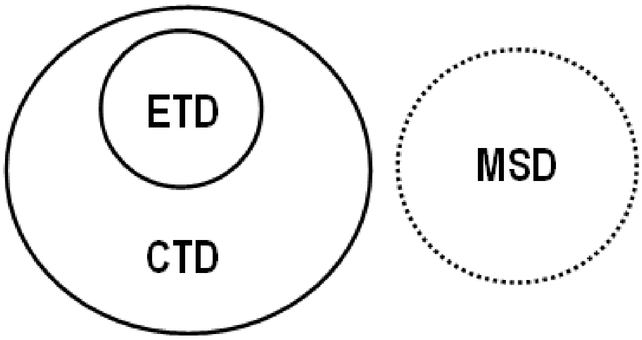

Figure 1: Types of datasets. Size of circles symbolizes the relative size of datasets. For MSDs the size can in principle easily be varied.

2. CONCEPT

2.1. Overall approach

The overall approach for building a RID-MIP consists of the following basic ideas: Experienced providers (single institution or a group of institutions) prepare well-defined and validated image datasets as common reference for R&D groups. The datasets meet certain quality criteria that are established by interdisciplinary groups of experts on the respective medical imaging domain. Provider and user of a dataset make a bilateral co-operation agreement or contract. Depending on the type of the dataset (see 2.2) and the aim of the research or development work (scientific, commercial) such bilateral agreements or contracts will vary significantly, especially with respect to licensing.

As to scientific methodology-oriented work, medical image processing tasks (solution expectable) or challenges (big gap between state of the art and a solution) are formulated in connection with appropriate free of license fee datasets. R&D groups all over the world then have the chance to apply their already available methods or to develop new methods in order to solve such a task or challenge as good as possible. The results of different groups can then be presented e.g. on periodically held sessions in the framework of established conferences. Due to the defined starting conditions, the methods or systems presented can objectively be compared.

As to development of applications for the use in clinical practice, R&D groups and companies shall use appropriate datasets for sound clinical validation of prototypes, pre-products or products. This can require license fees.

The role of EFMI WG MIP is a facilitating one. The working group shall propose tasks and challenges, initiate and support the specification, preparation and dissemination of appropriate datasets, and finally evaluate the impact of the initiative.

2.2. Different types of datasets

Within the RID-MIP three different kinds of data sets shall be available (figure 1).

2.2.1. Clinical Trial Datasets

A *Clinical Trial Datasets* (CTD) is big image dataset suitable as input of clinical trials for to validate the performance of MIP applications before transfer into clinical usage. CTDs serve as a basis for making good validation studies in the sense of clinical trials on big sets of data in order to measure the performance and outcome of MIP applications in prototype, pre-product or product stage. Clinical trial datasets can serve as an impartial reference for the comparison of different MIP systems.

It is essential for a CTD that the number and variety of images it contains is sufficient to make statistically sound statements about the quality of a solution to a task or challenge. Usually a MIP system should be able to cope with a certain variety of image qualities, e.g. with images from different acquisition units, in order to be robust enough for clinical usage. However, there can be exceptions, where a system is optimized for certain models of acquisition units. In such a case CTD subsets can be used.

2.2.2. Elk Test Datasets

An *Elk Test Dataset* (ETD) is a minimum dataset in the sense that it comprises not more than necessary images for to meet the requirement of reflecting the domain- and modality-specific image variety. The purpose of ETDs is to foster the development of methods for medical image analysis by serving as a reference for method comparison. The emphasis lies on basic methods such as preprocessing or segmentation methods. The statistical properties and the variety of elk test datasets should be similar to the corresponding large clinical trial dataset to ensure consistent evaluation results. It might even be reasonable to define an ETD as a suitable subset of a CTD, if such a CTD is already available.

2.2.3. Modality Simulator Datasets

A *Modality Simulator Dataset* (MSD) is a dataset artificially generated by a modality simulator. Based on an algorithmic model of the modality's physical behavior, the gold standard for the evaluation of MIP systems is mathematically derived and does not need to rely on a human expert. Therefore such simulators allow the accurate quantitative evaluation of MIP systems, at least in the sense of more or less realistic phantoms, depending on the quality of the underlying model. In general, modality simulators can be supposed to be suitable for generating ETD-like datasets. Nevertheless, each model means a certain simplification of reality. The clinical evaluation of MIP systems can most probably not be performed on the basis of modality simulator datasets.

An example for a modality simulator is the MRI simulator for normal brains and brains with MS lesions in [3]. It has been used to generate a Simulated Brain Database (SBD) which contains a set of realistic MRI data volumes for normal anatomy and for multiple sclerosis. For both, full 3-dimensional data volumes have been simulated using T1, T2 and proton-density-weighted sequences and a variety of slice thicknesses, noise levels, and levels of non-uniformity.

2.3. Requirements

To assure compatibility and adequately usage of the RID-MIP, for all its datasets – depending on their corresponding tasks or challenges –descriptions of content, modality and reference are required. In this context these requirements have the following meaning:

- *Contents:* this data type consists of two types of information:
  1) the *public* information about the medical domain (e.g. cardiology), the location (e.g. heart) and the diagnosis (e.g. ischemic stroke) and others, seen in the image;
  2) the *confidential* information about the patient's basic records (age, sex, race, etc), as well as the corresponding anamnestic data. The confidential part has to be obtained in an anonymous way.
- *Modality:* information about the image acquisition unit and the recording parameters used. The image acquisition unit can either be a complete unit (e.g. Siemens SOMATOM Volume Zoom CT) or a chain of optical devices, where each individual part can have an influence on the final image (e.g. Wolf 15mm C-mount objective, Wolf 90^o rigid Laryngosope, Wolf 250 W light-source, FAST screen-machine video-adapter). The recording parameters can either be unit-depended (amount of illumination, recording time) or object-dependent (contrast dose).
- *Reference ('gold standard'):* information obtained from at least one expert or from the best available method to make correct decisions. This information serves as a reference to measure the quality of an algorithm or a method. Examples can be a manual segmentation of a cancerous area/volume in the brain from a MRI scan, the classification of individual tumor cells from a microscopic view, or a histology for the decision benign or malignant tumor.

Not all information will be needed for all dataset scenarios. Table 1 shows the necessity of each information type regarding its use (modality simulation, development of algorithms, clinical trials)

Table 1: Information Requirements in dependence of dataset scenario

	Modality simulation	Algorithm development	Clinical trials
Public data	Yes	Yes	Yes
Confidential data	No	No	Yes
Modality	Yes	Yes	Yes
Reference	Depending	Yes	Depending

Much, but not all of the information requirements are already fulfilled for modalities using the internationally established DICOM standard, such as CT, angiography and MRI. Other imaging modalities, such as endoscopy, microscopy, or infra-red camera videos (e.g. used in gait analysis) make use of completely different documentation standards, proprietary data formats, no standards at all, or are not supported by existing DICOM standard. Therefore a unique image data acquisition documentation standard will have to be established, e.g. using XML.

2.4. Licensing policy

The datasets for to evaluate and compare basic MIP methods, i.e. especially ETDs, shall be provided free of charge. According to the objective of primarily evaluating pre-products and products, and due to the fact that it is a huge amount of work to prepare these big datasets, CTDs will usually be licensed by the dataset producer.

In order to support sound agreements between providers and users of the datasets, EFMI WG MIP will provide templates for co-operation agreements and contracts. Such agreements typically contain issues like subject (i.e. exact identification of the dataset and the kind of rights that are granted to the user); title of the study that's intended to be performed on basis of the dataset; binding of dataset usage to the named study; term of the agreement; restrictions on how to handle the dataset in the study to assure its reference character; good scientific practice (especially with respect to archiving of primary data for studies [5]); the appropriate co-operation with the dataset provider (e.g. scientific discussion, correct citation of the dataset, co-authorship of publication); conditions for irregular termination of the agreement. Contracts about CTDs will also contain a license fee agreement, at least if there is a commercial interest behind the use of the dataset, e.g. in order to validate a pre-product or a product.

2.5. Standardized naming of datasets

It seems to be reasonable to have a standardized way of naming the datasets that are available from different providers. There should be a full name and an abbreviation, as well as a version number. As to give a first suggestion of components included in such a dataset standard name: Medical domain (e.g. dermatology), modality (e.g. skin surface microscopy), subject of imaging (e.g. melanocytic lesions), number of images or image series (e.g. 300), version number (e.g. V1.0), date of issue (e.g. 020224 in the format yymmdd). Then, for example, *Dermatology* - *Skin Surface Microscopy* - *Melanocytic Lesions* - *300* could be the full name of a dataset and *D-SSM-ML-300-V1.0.020224* its abbreviation.

3. ROADMAP

The roadmap suggested for to establish the first version of the RID-MIP consists of the following steps:

3.1. Identification and specification of important datasets

In this first step, important datasets required in actual MIP R&D are to be identified, named and specified. This includes the identification of most relevant domains and problems, the definition of general as well as domain- and modality-specific assessment criteria for single images and whole datasets, and a commented list of presently available modality simulators.

3.2. Selection of first three highly relevant ETDs

On basis of the specification from step 1, three relevant and not too difficult to fulfill ETD specifications should be chosen for development. In order to increase the chances for good acceptance and dissemination, the development should be done preferably by international subgroups of EFMI WG MIP.

3.3. Dissemination of these first ETDs

The EFMI WG MIP will try to promote dissemination and application of the ETDs in order to achieve the intended impact on MIP R&D activities (Internet, conferences, fairs).

3.4. Evaluation of impact on R&D

During a three-year observation study, starting some months after dissemination of the datasets, the outcome shall be measured in terms of number and quality of publications based on the datasets.

In further development steps, the collection of ETDs shall be expanded and development, dissemination and evaluation of CTDs shall be started. At latest in this stage, attempts to get funding from the European Commission seem to be reasonable.

4. TIME SCHEDULE

It is intended to specify the first ETDs from March to August 2002 and to present the results at the MIE 2002. The preparation of these ETDs shall follow from October 2002 to February 2003, so that they can be presented at the Special Topic Spring Conference of EFMI in March 2003. Dissemination is planned for April to August 2003.

5. CONCLUDING REMARKS

Although the MIP-RID is an activity of the European EFMI WG MIP, a co-operation with Korea would be highly welcome! Actually, it is hoped, intended and highly welcome that the RID-MIP should be accessible, maintained and used by R&D groups all over the world.

The Visible Human Project [6] has stimulated a lot of excellent work in the field of medical image processing. The ongoing Visible Korean Human Project [7] will probably have further stimulating effect. Whereas these projects with their few individual study objects mainly help in accelerating basic methodological progress and new insights in anatomy, the RID-MIP initiative will hopefully help in fostering the comparability of MIP applications' performance for clinical use.

ACKNOWLEDGEMENTS

The authors want to thank the colleagues from Finland, The Netherlands, Switzerland and Germany who have already joint the WG or have shown deep interest.

REFERENCES

[1] Munson DC jr. (editor-in-chief): A Note on Lena. IEEE Transactions on Image Processing. Vol. 5. No. 1, Jan. 1996

[2] The Lena Story: http://ndevilla.free.fr/lena/ ; http://www-2.cs.cmu.edu/~chuck/lennapg/lenna.shtml

[3] Kwan RKS, Evans AC, Pike GB: An Extensible MRI Simulator for Post-Processing Evaluation. Visualization in Biomedical Computing (VBC'96). Lecture Notes in Computer Science, Vol. 1131, Springer, 1996, 135-140

[4] McConnel Brain Imaging Centre: BrainWeb interface to Simulated Brain Database (SBD) and custom MRI simulations: www.bic.mni.mcgill.ca/brainweb/

[5] DFG Ehrenkodex für gutes wissenschaftliches Verhalten. Sicherung und Aufbewahrung von Primärdaten (Empfehlung 7), DFG 16.12.1997, http://www.dfg.de/aktuell/pressemitteilungen/Archiv/presse_1997_31.html

[6] Visible Human Project http://nlm.nih.gov/research/visible/visible_human.html

[7] Min Suk Chung, Woo Sup Hwang, Jin Seo Park, Jin Yong Kim, Young Hwa Ahn, Jae Keun Kim: Visible Korean Human. The Fifth Koera-Germany Joint Workshop on Advanced Medical Image Processing, May 17-18, 2001