Establishing an International Reference Image Database for Research and Development in Medical Image Processing

Alexander Horsch[1], Micheal Prinz[2], Siegried Schneider[3], Outi Sipilä[4]
Klaus Spinnler[8], Jean-Paul Vallée[5], Irma Verdonck-de Leeuw[6]
Raimund Vogl[7], Thomas Wittenberg[8], Gudrun Zahlmann[3]

[1]Dept. of Medical Statistics and Epidemiology, Technical University of Munich, Germany; [2]Dept. of Medical Computer Sciences, University of Vienna, Austria; [3]Siemens AG, Erlangen, Germany; [4]Dept. of Radiology, Helsinki University Central Hospital, Finland; [5]Digital Imaging Unit, Informatics Center, University Hospital of Geneva, Switzerland; [6]Dept. of Otorhinolaryngology, VU Medical Center, Amsterdam, The Netherlands; [7]Tiroler Landeskrankenanstalten GmbH (TILAK), IT Dept., Innsbruck, Austria; [8]Fraunhofer-Institut für Integrierte Schaltungen, 91058 Erlangen. Email: alexander.horsch@imse.med.tu-muenchen.de

Abstract. The lack of comparability of evaluation results is one of the major obstacles of Research and Development (R&D) in Medical Image Processing (MIP). The main reason for that is the usage of different image datasets with different quality, size and Gold standard. Currently, there exist only poor and insufficient attempts to cope with this problem. Therefore, one of the goals of the Working Group on Medical Image Processing of the European Federation for Medical Informatics (EFMI WG MIP) is to develop first parts of a Reference Image Database (RID) for Medical Image Processing R&D groups until 2005. Kernel of the concept is to identify highly relevant medical problems with significant potential for improvement by MIP, and then to provide respective reference datasets. The EFMI WG MIP has primarily the role of a specifying group and an information broker, while the provider user relationships are defined by bilateral co-operation or license agreements. An explorative RID prototype has been implemented in MySQL, templates for provider user agreements have been worked out and already applied for own 'pre-RID-MIP' co-operations of the authors. First RID datasets are made available in 2003 by WG members.

1 Introduction

Medical image processing (MIP) is a steadily growing field in modern medicine. Every year, new methods (sometimes not really new, but re-invented or slightly modified old ones) and systems are presented. However, it is usually impossible to compare the performance of different algorithms, methods and applications in a sound way, because almost every R&D group uses its own image datasets.

Currently, there exist only poor and insufficient attempts to cope with this problem. Using the well-known Lena image ("'Lady with a hat"') [1], the image

processing and data communication communities have evaluated a huge number of basic methods and thereby established a de facto image reference. The task was comparably easy: one image with a variety of different features was chosen as a test image. For medical image processing, there have been only few attempts to establish appropriate reference datasets, such as Voxelman and Visible Human (these with emphasis on anatomy). 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 (from a big number of different patients) are required for to cover at least the most important medical imaging domains. And the problem is further extended, since reference images are needed for various combinations of modalities (e.g. CT, MR, endoscopy), locations (e.g. heart, brain, lung) and diseases. However, there are comparable initiatives for other objectives, e.g. clinical applications [2] or biosignal analysis [3], which give valuable ideas for how to tackle the problem.

The Working Group on Medical Image Processing of the European Federation for Medical Informatics (EFMI WG MIP), initiated in September 2001 by the main author, has set as one of its goals to develop first parts of a Reference Image Database for Medical Image Processing R&D groups (RID-MIP) until 2005. The RID-MIP concept has been described in detail in [4]. This article puts the emphasis on intermediate results and obstacles.

2 Material and Methods

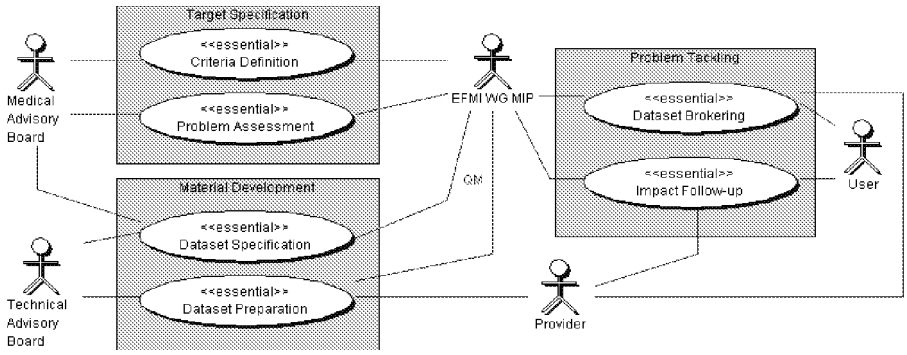
In the first phase, datasets for the following tasks are provided by WG members: 1) highly accurate classification of melanocytic lesions by skin surface microscopy, 2) highly accurate segmentation and volumetry of normal liver in CT, 3) fully automatic segmentation of stroke lesions in MRI.

The Unified Modelling Language UML [4] serves as language for graphical modelling, and the tool Together 6.0 (Together Soft Corp., Raleigh, NC, USA) as computer-based environment. The explorative RID prototype is implemented in MySQL.

3 Results

Overall concept. Figure 1 illustrates the overall concept of the initiative. The *target specification process* comprises the definition of criteria for the assessment of medical problems and their MIP parts, as well as the assessment itself. Result of the process is a list of highly relevant medical problems and corresponding MIP tasks and challenges, for which reference datasets should be made available with high priority. The *material development process* consists of the specification and preparation of datasets. Four kinds of datasets are distinguished: system trial datasets, elk test datasets, modality simulator datasets, phantom datasets. The *problem tackling process* comprises dataset brokering and impact follow-up. The processes are co-ordinated by EFMI WG MIP, whereas the provider user relationship itself is always a matter of a bilateral agreement or contract.

Fig. 1. The overall concept of the RID-MIP initiative



Benefit for the dataset provider. Why should an image producing party become a provider of the RID-MIP? 1) An important reason can be to get a solution for a problem from the user(s), if this is agreed in the co-operation between provider and user(s). 2) Co-operation with the aim to fuse the new ideas of the dataset user with the experience of the dataset provider and common publication of the results can also be a benefit. 3) A third reason might be the interest in fostering scientifically sound evaluation of MIP methods and systems. 4) Next reason can be to initiate further development after having finished a project where image datasets have been acquired. The objectives of the new task need not necessarily be the same as in the original project. It is a matter of the provider user agreement to create an interesting co-operation. 5) Then, a reason can be given by concrete incentives like, for example, to receive a software tool from EFMI WG MIP to support the technical preparation of the datasets for submission to the RID or for direct shipping to the user, especially if this tool has also additional benefit for the provider, for example as an image viewer and editor. 6) If the user is a company and the purpose of dataset usage is to evaluate a commercial product, then license fees can be the reason for to provide a dataset. 7) Finally, providers will have access to other datasets in their domain and can use this for quality control by comparing the image acquisition quality of their own datasets with other ones.

Delivery paths. Principally, there are two different delivery paths for datasets: 1) delivery by post via a storage medium (usually CD-ROM or DVD), 2) download directly from the WG website or a website of the dataset provider. It is up to the provider to decide whether he wants a) to keep the image dataset delivery under his own management, or b) to let the whole dataset be stored on the WG server for authorized (i.e. after bi-lateral agreement or contract has been signed) download, or c) to let the images and additional information on each image of the dataset be stored in the RID for authorized complete or selective download.

The explorative prototype. A first restricted version of the reference database in MySQL for internal usage within EFMI WG MIP as working prototype on

the Internet has been developed in 2002. It comprises a) the dataset classification and management, b) the provider and user management, c) the follow-up management (impact), d) management of images and image-related additional information. Parts a-c of the database are dealing with management information on datasets as a whole, while part d deals with the images themselves. The prototype is available for WG members on the internal area of the WG website in order to discuss the further development.

The agreement templates. Templates for provider user agreements have been prepared in German and English language. They consist of the following paragraphs: §1 Subject (timely limited, non-exclusive usage license for the dataset); §2 User license (user is allowed to use the dataset in the framework of the scientific study referenced in this document); §3 Binding to purpose; §4 Validity timespan; §5 Reference character of the dataset; §6 Usage by a third party; §7 Media and copies; §8 Reference copy; §9 Fees; §10 Co-operation; §11 Publication of the results; §12 Runtime and irregular termination. The full text templates can be downloaded from the WG website. They have already been used for several co-operations.

The website. On the website *www.efmi-wg-mip.net* detailed information about the state of the work is made available to the public. Detailed conceptual documents, the roadmap, work plans and the explorative prototype are available in the internal area for WG members.

4 Discussion

There are at least two critical points beyond any problems with motivation or management: 1) The Gold standard. For many tasks, the creation of a reliable Gold standard is difficult. For example, to create the Gold standard segmentation for a system trial dataset with a hundred brain MRI examinations of MS patients is a huge amount of work, and the result is still a kind of a fuzzy standard. 2) To provide reference image datasets together with a task or challenge is probably not enough for many problems; it is also necessary to give (more or less mandatory) guidelines or recommendations for how to perform the evaluation of methods and systems on basis of the dataset, especially concerning the study design and the appropriate statistical methods and packages. The comparability of evaluation results will suffer from divergent evaluation procedures. (In medical and pharmaceutical research, these issues are considered very carefully.)

5 Conclusion

The first step of conceptual work on the way to a comprehensive reference image database for medical image processing has been done. The group of currently 11 active members from 6 European countries has to be enlarged. Applications for funding in the IST part (Information Society Technologies. Integrating and Strengthening the European Research Area) of the Sixth Framework Programme

of the European Commission will hopefully be successful in order to bring the necessary power on the next steps of the initiative.

References

1. Munson DC jr: A Note on Lena. IEEE Transactions on Image Processing. 5(1):editorial, 1996
2. Horsch A: Jetzt im Web - der Health Informatics Software Catalogue der EFMI. mdi Forum der Medizin-Dokumentation und Medizin-Informatik 4(2):64–65, 2002
3. Penzel T, McNames J, deChazal P, Raymond B, Murray A, Moody G: Systematischer Vergleich von Algorithmen der EKG-Analyse zur Erkennung schlafbezogener Atmungsstörungen. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 33(2-3):126–127, 2002
4. Prinz M, Horsch A, Schneider S, Sipild O, Spinnler K, Vallee JP, Verdonck-de Leeuw I, Vogl R, Wittenberg Th, Zahlmann G: A Reference Image Database for Medical Image Processing. Proceedings of the 2nd conference of the Österreichische Wissenschaftliche Gesellschaft für Telemedizin, a-telmed 2002, OCG-Schriftenreihe, Vienna, 45–51, 2002