Grid-enabling medical image analysis

Cécile Germain-Renaud, Vincent Breton, Patrick Clarysse, Yann Gaudeau, Tristan Glatard, Emmanuel Jeannot, Yannick Legre, Charles Loomis, Isabelle Magnin, Johan Montagnat, et al.

Medical Image Analysis on Grids

C. Germain , V. Breton, P. Clarysse, Y. Gaudeau, T. Glatard , E. Jeannot, Y. Legré,
C. Loomis, , I. Magin, J. Montagnat, J-M Moureaux, A. Osorio, X. Pennec, R. Texier

***Abstract*—**
Grids have emerged as a promising technology to handle the data and compute intensive requirements of many application areas. Digital medical image processing is a promising application area for grids. Given the volume of data, the sensitivity of medical information, and the joint complexity of medical datasets and computations expected in clinical practice, the challenge is to fill the gap between the grid middleware and the requirements of clinical applications. The research project AGIR (Grid Analysis of Radiological Data) presented in this paper addresses this challenge through a combined approach: on one hand, leveraging the grid middleware through core grid medical services which target the requirements of medical data processing applications; on the other hand, grid-enabling a panel of applications ranging from algorithmic research to clinical applications.

I. Introduction

DIGITAL medical images represent a tremendous amount of data, from tens of MB for a MRI (Magnetic Resonance Imaging) image to hundreds for a multiresolution spiral CT-scan; the annual production of a single radiology department is estimated to be tens of TB per year. Moreover, data semantics is important when considering medical data, and data are often manipulated as correlated data sets. Medical data storage and retrieval thus require the manipulation of large volumes of data and meaningful associated metadata. Medical data manipulation is even more complex given the privacy constraints associated with personal data.

Algorithms for medical image analysis and diagnostic assistance tools have been developed these last fifteen years or so. Some of these algorithms have reached a high level of usability and proved to have a real impact in the clinical domain. However, their widespread adoption by clinicians is not realized yet.

Data and computing grids are an opportunity to enlarge the impact of these image processing tools and to transfer experimental research to clinical practice.

- Grids provide an infrastructure allowing multiple user communities to access and manipulate medical data.
- Grids offer the computing power needed to validate algorithms on large datasets and to process complete databases for applications requiring statistical information such as epidemiology.
- Transparent access to medium to high-end computing systems through grid middleware broadens the applicability of augmented reality as a medical tool.

Grid technologies have intensively been developed this last decade and grid middleware and standards are now emerging [11]. Filling the gap between the clinical applications and the grid middleware raises many specific issues, ranging from computer science basic research to legal concerns. Addressing these issues is an active research and technology area, and a new scientific community is emerging [4].

The authors are with CNRS, INRIA and INSERM
This work is partially funded by ACI Masses de Donnes

This paper presents the AGIR project [1]. The objectives of this multi-disciplinary project are twofold.

Algorithm research and deployment. In the short term, the deployment and availability of algorithms on a data and computing grid will ease the development, prototyping, and the validation of these algorithms.

New grid services Software development which address some of the requirements of complex medical image processing and data manipulation applications.

The method is:

- to combine the expertise from medical and computer science teams (specialized in clinical applications, medical image analysis algorithms and IT systems, grid and distributed systems) on a few paradigmatic medical applications, in order to get a cross-section of the middleware, algorithmic and medical issues;
- to leverage the research and software already developed by the participant teams in the field of medical image processing and grids.

The expected results are the integration of the solutions into a demonstrable testbed, an improved knowledge of the requirements of medical applications, algorithms and information systems towards grid middleware and designs, more powerful medical analysis algorithms, impact on national and international grid projects, and clinical deployment experiments.

AGIR started in September 2004; thus this paper reports plans and work in progress. Section II presents the overall project architecture; sections III and IV present the issues related to each architecture component; sections V,VI, VII and VIII provide descriptions of grid medical services coupled with applications.

II. Overview

The deployment of medical applications on grids can be structured following the four layers illustrated in fig. 1. The lowest layer corresponds to basic grid services. The second layer represent core services dedicated to medical applications that are not available in general purpose middleware. The third layer is the medical image processing algorithms taking advantage of the underlying grid services to process large amounts of data. The fourth and upmost layer represents clinical applications developed to address medical challenges. AGIR mainly addresses core medical services and image processing algorithms and, to some extent, clinical applications.

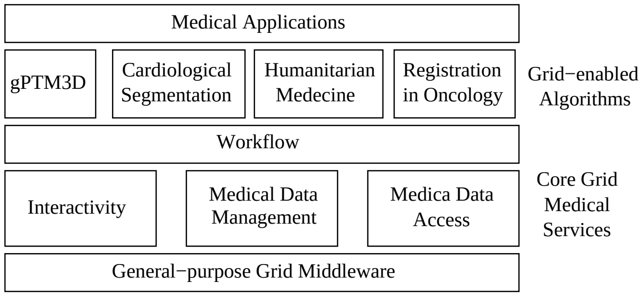

Fig. 1. A layered architecture for medical applications deployment on grids

Grids middleware is already available: truly large-scale grids exist and provide hardware and software infrastructure; real medical grids will be deployed on proven and supported grids. Thus, the project focus is definitely not on developing basic grid middleware. At this level, the goal is to assess the compatibility (or lack of) of existing production or research middleware with medical imaging requirements. AGIR is engaged in a strong collaboration with the European EGEE project, which is the basis of the initial and future testbeds. Nevertheless, the results of AGIR must not be over-determined by the underlying middleware; as far as possible, the core medical services must be WSRF-compliant; comprehensive testing of emerging middleware, such as Grid Data Farm [27] will be conducted.

III. Core Medical Services

Reaching the high level of data and computing resource integration needed for complex medical applications requires addressing considerable challenges. Opening the system to multiple user communities raises the problem of medical data privacy. Sharing medical data sets for processing raises the problem of transparent access to heterogeneous and distributed data sources. Seamless integration of the grid services into the usual workflow or tools of the physician or medical researcher requires integrating heterogeneous delay-oriented multitasking into the grid exploitation models.

A. Interactivity

A critical system requirement is the need to move Grids from exclusive batch-oriented processing to general purpose processing, including interactive tasks. This is essential for any activity that involves humans, such as on-demand data product generation, exploration of datasets and visualization. For example, in high-degree-of-freedom optimisation problems such as non-linear image registration or radiation treatment planning, computational steering can be used to help algorithms avoid local minima. These interactions may trigger asynchronous computation or data retrieval operations.

B. Medical data management

Grid access to medical images implies concurrent access and tranfers of very large data sets: 3D images or even sequences of 3D images. Moreover, medical data encompasses not only medical images but a rich amount of semantic metadata as well, e.g. the patient medical record. The medical data manager needs to easily identify data sets according to clinically relevant research criteria that involve both metadata and image contents. On the other hand, core middleware data managers for grids (such as the Internet Backplane Protocol or the EGEE SRM) provide basic functionalities for storing, replicating or caching individual files. The initial step to bridge this gap is to grid-enable the access to DICOM servers. An EGEE-enabled DICOM interface based on DCMTK is being implemented, built from previous experiments in the MEDIGRID project [16]. The next step will be to take into account the performance requirements of medical applications, particularly concerning the semantics of queries and prefetching of data sets for grid computations (latency) and the concurrent accesses to DICOM servers (bandwidth). Finally, grid security must be levelled in order to actually reap the health care benefits of the grid. Current Virtual Organization based systems can provide a binary patient privacy protection (full or denied access), but might be more difficult to adapt to differentiated access, from the attending physician to pedagogical or epidemiological use.

C. Medical data access Protocols

Transmission of the tremendous amount of medical data involved in medical applications can easily saturate existing systems. The general goal of this subtask is to define an analogue of the streaming protocols used for remote access to multimedia data, which can be termed as a medical streaming protocols. As in multimedia protocol, the key idea is to focus the resource usage on the most useful data, while the Grid system can exploit the delays related to human interaction or remote processing in the background to refine the information finally delivered.

The three components of these protocols are compression algorithms, coupling compression and transfer, and finally optimization targeting medical image exploration. Compression can be optimized depending on the application use of data: fast encoding and decoding should be privileged for data transmission while longer but more efficient and in some cases lossy compression algorithms are suited for long term storage. It is thus important to allow for a dynamic adaptation of the compression method and ratio depending on the resource performance speed and the application usage. Specializing compression algorithms to medical images or datasets offers opportunities to go further than general-purpose compression algorithms. For instance, only a cursory inspection can result in eliminating some datasets, when a medical record is interactively or automatically browsed for selecting the most relevant images. Moreover, some applications need an optimal quality only within a region of interest (ROI) of a selected image or set of images. Typical examples of this situation are: navigation through a registered volume, with continuous visualization of the corresponding planes; medical windowing; situations where not all the information provided by

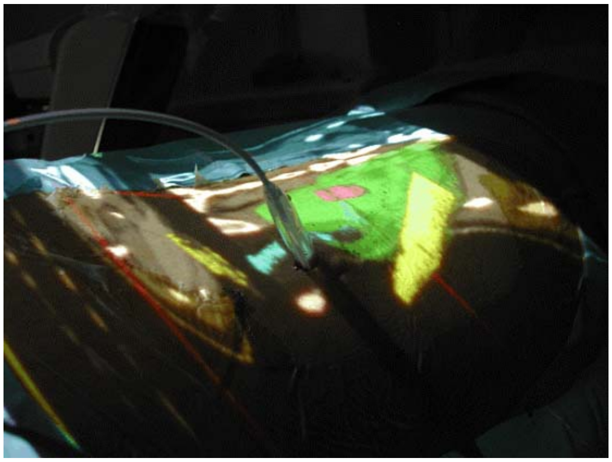

Fig. 2. Augmented reality applied to percutaneous nephrolithotomy, Tenon hospital. October 2003.

complex segmentation or registration algorithms can be simultaneously displayed. For these situations, a selection of the relevant data could lead to very efficient data transmission schemes. In the specific context of interactivity, automatic selection through intelligent prefetch mechanisms could exploit the way data are explored and extrapolate the exploration characteristics.

IV. Gridifying Medical Applications

Four medical applications are the primary targets of AGIR: PTM3D, an interactive medical analysis software focusing on volume reconstruction and measurement; cardiac image segmentation; remote medicine, exemplified by humanitarian medical development using grid technologies; and image registration applied to oncology.

This set of applications is the basis for studying and developing the grid services needed for storage, retrieval, and processing of medical data on grids. Our postulate is that different medical applications have common basic requirements that are not yet covered by existing grid middleware. The goal is to grid-enable these applications up to the point where they can be used in their usual context, either clinical research or medical practice. Moreover, these applications provide a cross-section of various medical scenarios which can benefit from the grid:

Algorithm research and deployment. First, algorithms can be developed and tested on larger datasets. Second, the spreading and sharing of algorithms is eased, allowing users to experiment and compare existing techniques on common datasets. In a longer term, grid-enabled algorithms will be more easily deployed in the medical world for clinical use. This will help dissemination of image processing techniques and real scale experimentation. Finally, new applications are expected to emerge from such a grid infrastructure. Epidemiology or other data intensive applications will benefit from the availability of large datasets and medical knowledge databases.

Image guided diagnosis and surgical planning. Interactive image analysis is potentially a major beneficiary of the high-performance computation available on a grid. Combining the medical user expertise and the resource of the grid in compute and data intensive tasks is a promising way to transfer experimental research first to clinical practice, and then to routine clinical practice.

Augmented reality Pre-operative data, computed geometries, intra-operative images or the patient body itself can be combined to optimize a surgical intervention or therapy planning (fig. 2) [21]. Registration, intra-operative MRI for brain surgery, and multimode fusion are typical examples of this approach.

V. Interactivity and Volume Reconstruction

A. A case for interactive grids

Many clinical or medical imaging applications require both human supervision and high-end computing and data storage resources. Seamless integration of the grid power to the everyday desktop use, when the desktop is insufficient, requires to revisit, not the grid concepts, but many trends in grid software. Departing from the computing centres model means to put emphasis not on work throughput, but on response time at all levels of the grid protocols, while retaining the key advances of Grid technology allowing to build virtual organizations: security, unique login, cross-domain access etc. The general goal is thus to integrate the response-time constraints into the grid scheduling protocol stack.

Two classes of services can be considered to tackle this issue

Application-level schedulers Many interactive applications amenable to grid involve pleasant parallelism. These applications often require fine-grained dynamic scheduling, at a much finer time scale than what grid scheduling can provide. Application level [3] is here analogous to the user-level thread schedulers versus kernel-level ones in scheduling for a processor. Just as a user-level thread scheduler avoids the penalties of the kernel calls, a grid application-level scheduler will avoid the traversal of the grid information systems for executing each and every small parallel task.

Soft real-time scheduling(SRTS) The grid resources must be shared across interactive jobs on one hand, and across interactive and batch jobs on the other hand. Other contexts also are presented with the need to combine best-effort and QoS, namely network traffic management [5], [8], [19], and multimedia job scheduling on general purposes system [6], [26]. However, the constraints on the scheduling policy targeting interactive jobs might be more stringent. Typically, the delays incurred by the batch jobs should remain bounded and small; interactive jobs should do not degrade resource use nor need extensive reorganization of existing policies governing the batch jobs.

QoS for grids has been explored in [15], [10] and is currently addressed by the WS-agreement GGF protocol [9]. While

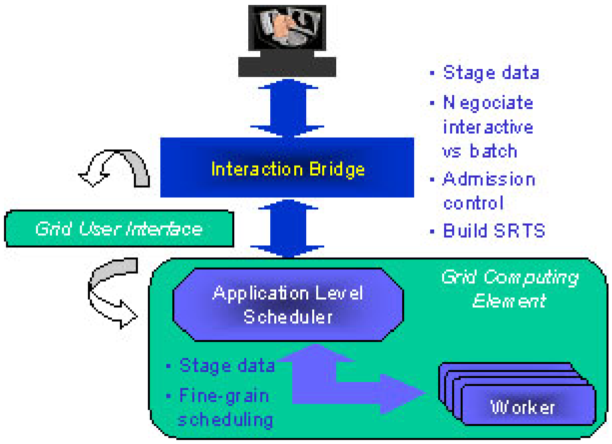

Fig. 3. gPTM3D architecture

some components of grid SRTS could fit in this framework, e.g. admission control for interactive applications (if sufficient resources are available, an interactive session is admitted), fine-grained time-sharing requires the more reactive behavior of user-level autonomous agents.

B. *gPTM3D*

PTM3D [25] is a fully featured DICOM images analyzer developed at LIMSI. PTM3D transfers, archives and visualizes DICOM-encoded data; besides moving independently along the usual three axes, the user is able to view the cross-section of the DICOM image along an arbitrary plane and to move it. PTM3D provides computer-aided generation of three-dimensional representations from CT, MRI, PET-scan, or echography 3D data. A reconstructed volume (organ, tumor) is displayed inside the 3D view. The reconstruction also provides the volume measurement, required for therapeutic decisions. The system currently runs on standard PC computers and it is used on-line in radiology centres. A cluster-enabled volume recontruction is described in [14].

The first step in grid-enabling PTM3D (gPTM3D) is to speedup compute-intensive tasks, such as the volume reconstruction of the whole body used in percutaneous nephrolithotomy planning (fig. 2). The volume reconstruction module has been coupled with EGEE with the following results:

- the overall response time is compatible with user requirements (less than 2 minutes), while the sequential time on a 3GHz, 2MB memory PC is typically 20 minutes.
- the local interaction scheme (stop, restart, improve the segmentation) remains strictly unmodified.

This first step has implemented fine-grained parallelism and data-flow execution on top on a large scale and file-oriented grid system. The architecture based on Application Level Scheduler/Worker agents shown in fig 3 is fully functional on EGEE. The Interaction Bridge (IB) acts as a proxy between the PTM3D workstation, which is not EGEE-enabled, and the EGEE world. When opening an interactive session, the PTM3D workstation connects to the IB; in turn, the IB launches a scheduler and a set of workers on an EGEE node, through fully standard requests via an EGEE User Interface; a stream is established between the scheduler and the PTM3D front-end through the IB. When the actual volume reconstruction is required, the scheduler receives contours; the Scheduler/Worker agents follow a pull model, each worker computing one slice of the reconstructed volume at a time, and sending it back to the scheduler, which forwards them to IB from where they finally reach the front-end. The next step will be to implement a scheme where the IB and the scheduler cooperate to respectively define and enforce a soft real-time schedule.

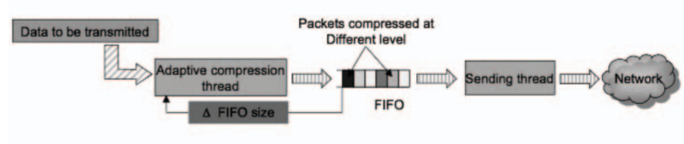

Fig. 4. AdOC Algorithm: emission process (reception process is symmetric but does not monitor the queue size)

VI. Adaptive Online Compression and Cardiac Image Segmentation

This service provides pipelining of data transfer and compression. It is an *adaptive* service as the compression level dynamically changes according to the environment and the data.

A. *The AdOC Algorithm*

The AdOC algorithm (fig. 4) has been proposed by Jeannot, Knutsson and Bjorkman in [17]. It is a general-purpose, user-level, portable algorithm suited not only for grid computing but also for any data transfer application. It is mainly based on two ideas:

- Compression and communication overlap. Overlapping compression with communication allows the compression time to become mostly invisible to the user.
- Dynamic adaptation of the compression level. The environment (CPU/network speed, data, etc.) is subject to change with time. Therefore, the available time to compress/decompress data changes during the data transfer.

Communication and compression overlap is implemented by multithreading. The sending process is made of two threads. One thread compresses the data; the other thread sends the data to the network. A queue is used to store data shared by each thread. On the sending side, the compression thread stores data in the queue, the emission thread reads this data and sends it to the network. On the receiving side, a symmetric mechanism is provided. The adaptation is performed by monitoring the variation of the FIFO queue on the sender side. The idea is the following:

- If the queue size increases, this means that the network and the receiver consume data slower than it is produced by the compression thread. Some more time is therefore available for compression: the compression level is then increased.

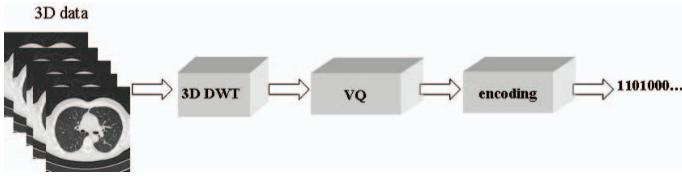

Fig. 5. The lossy Compression Scheme

- If the size of the queue decreases, it means that the network and the receiver consume data faster than it is produced by the compression thread. It is required to decrease the compression level in order to generate packets at a greater rate.

B. *The AdOC Library*

The AdOC library is a set of user level functions that enables on the fly compression. It provides functions for sending or receiving data through a socket using the AdOC algorithm. The function are similar to the `read`, `write` and `close` POSIX system calls and respect their semantics. This library can be used both for transmitting (using AdOC at both ends of the socket) or for storage on a server (using AdOC only on the client side). Compression can be optimized depending on the application use of data: fast encoding and decoding should be privileged for data transmission while longer but more efficient compression algorithms are suited for long term storage.

C. *Lossy Compression for Image Transmission*

So far, the AdOC library provides only lossless compression based on the Ziv-Lempel algorithm [32]. However when dealing with images, lossy compression is a interesting alternative. Concerning medical images, lossless algorithms are preferred to lossy ones for obvious diagnosis purposes. However, in some cases, a non-optimal quality of information is sufficient, and lossy algorithms can be thus an efficient response to storage or transmission problems. Here we propose the design of a lossy compression scheme based on three well-known steps: discrete wavelet transform (DWT), quantization, and binary encoding, as represented in fig. 5. We have shown [30] that this scheme could outperform the JPEG2000 standard as well as the famous SPIHT algorithm [28]. In order to exploit time correlations of volumetric data, we use 3D DWT as well as vector quantization (VQ). One of the key points of this scheme is the multiresolution analysis associated to DWT which yields several sub-volumes (or sub-bands in 2D) to be compressed. Thus, a bit allocation procedure is necessary to assign a compression bit rate to each sub-volume under the constraint to reach the total compression rate, while minimizing the total distortion. As a consequence, progressivity can be easily performed, by transmitting initially low resolution sub-volumes. This property can be used advantageously as a function of the network speed to adapt the response time.

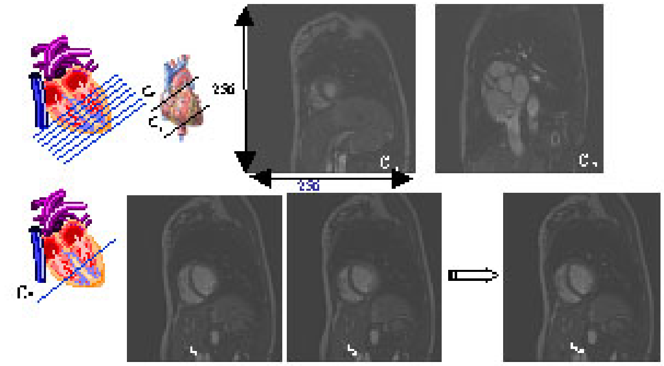

Fig. 6. MRI acquisition composed of 7 short axis slice levels and 30 cardiac phases at each level.

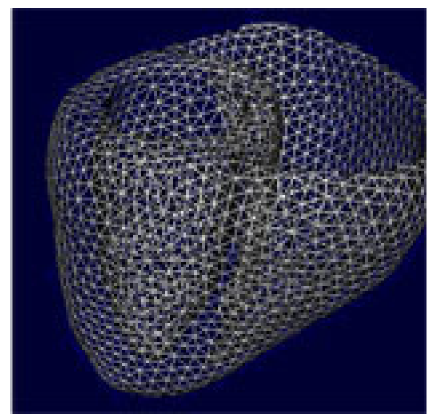

Fig. 7. The bi-cavity heart template

D. *3D+time Segmentation of Magnetic Resonance Cardiac images*

Accurate analysis of the cardiac function relies on tomographic image acquisitions. Current MRI cardiac examination comprises dynamic short axis acquisitions at several slices that cover the whole heart. Figure 6 shows a typical MRI acquisition of 210 images, with dimension 256x256, short coded. The analysis of the cardiac function relies on quantitative global parameters such as the volume evolution of the left ventricular cavity, the ejection fraction, as well as local parameters such as the wall thickening. The estimation of these parameters requires the extraction of the heart contours which requires computer-assisted image processing methods. As in the previous example, the goal of a grid-enabled method is quasi real time processing, in order to make the tool usable in a clinical context. We present first the segmentation alorithm, and next the relation with adaptive on-line compression.

The contour extraction process from the images is based on the so-called deformable model approach, which uses an a priori model of the object to be extracted. This model is deformed iteratively to fit the image content. The proposed model presents the advantage of allowing the simultaneous extraction of both the endocardial and epicardial surfaces [18], [23]. The concept, named elastic deformable

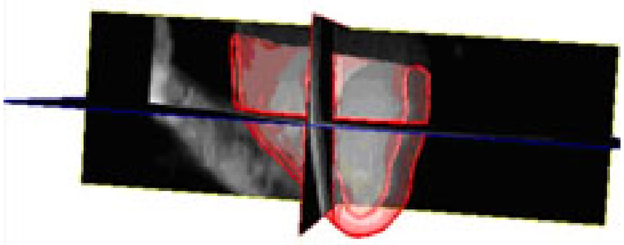

Fig. 8. Illustration of a segmentation result: heart model in red + 3D image

template, combines a topological and geometric model of the object to be segmented, and a constitutive equation (linear elasticity) defining its behavior under applied external image forces that pushes the model interfaces towards the image edges. In this context, the a priori model is a bi-cavity geometrical mesh that results from the manual segmentation of a reference data set (fig. 7).

The equilibrium of the model is obtained through the minimization of a global energy functional $E = E_{elastic} + E_{data}$ where $E_{elastic}$ represents the deformation energy of the model and E_{data} is the energy due to the external image forces.

Internal Energy term. The object is considered as a linear elastic body. Its elastic energy can be expressed as:

$$E_{elastic} = \frac{1}{2} \int_{\Omega} \sigma^t \epsilon \; d\Omega, \quad (1)$$

where σ, ϵ are the 3D strain and deformation vectors respectively and Ω is the model domain. Moreover, the material is considered as isotropic and is completely defined by the Young modulus Y and the Poisson coefficient ν.

External Image Energy. The object is submitted to a 3D boundary force field $\mathbf{t}$. The expression of the external energy E_{data} is :

$$E_{data}(u) = \int_{\Gamma} (\mathbf{t}.u) \; d\Gamma, \quad (2)$$

with Γ the border of the object domain Ω. The force field is either derived from the gradient of a potential function P (such that $\mathbf{t} = -grad(P)$) computed from an edge map, or a specific force field called gradient vector flow (GVF) [31] which is sometimes more efficient regarding the initialization and the convergence to edges.

Global energy minimization. The model is decomposed into tetrahedral elements. The segmentation is obtained through the minimization of the global energy functional E using the Finite Element Method (FEM).

Template initialization The segmentation procedure requires a rough positioning of the initial template into the data. This is performed by an affine registration of the model with a Volume of Interest (VOI), centered on the heart region and extracted from an isotropic interpolated volume data.

Figure 8 shows an example of the 3D segmentation.

The geometrical mesh is typically composed of about 6000 nodes and 25000 cells. The system handled for the problem solving is therefore of size 324MFloats ([$3 \times$ NumberOfNodes^2). The program uses the PETSC library for linear algebra operations. The typical computing time for one frame is about 3 minutes on a PENTIUM M, 1.7GHz, 1GB RAM.

Currently, once the segmentation has been performed on the first frame of the sequence, the resulting model is used as the initial template for the next frame, thus the computation is sequential. Adaptive lossless compression will be used to adapt the transmission speed of the consecutive frames to the performance of the network and computing resources.

In order to reduce the computing time for the segmentation of individual cases, a multi-resolution strategy can be applied to each 3D image of the sequence. Classical multiresolution processes construct a hierarchy of reduced resolution images by low-pass filtering and subsampling the original image a number of times; integrating such a crude sampling inside the datapath (transfer plus computation) from the DICOM server to the Computing Elements is a first simple compression strategy. However, the compression scheme described in VI.*C* can also be used to provide a flow of frames of increasing resolutions (streaming), with better correlation to the image intrinsic features.

Grid interaction is also applicable for this algorithm. User interventions may help bootstrapping the segmentation through user assisted definition of a VOI and checking the initial positioning of the template at a client station. Finally, a dynamic follow-up of the segmentation process might be of interest, and is easy, because only the template geometry has to be updated for visualization purpose.

VII. Workflow Management and the Evaluation of Registration Algorithms

Medical protocols define a medical data acquisition procedure which aims at providing suitable data for a given medical examination in a reproducible manner. Data acquired within a given protocol framework can be compared and reproduced. Beyond medical protocols for data acquisition, medical data analysis tasks now often involve processing of the acquired images for enhancing, interpreting and extracting quantitative information from the data. The processing of medical data involves several elementary tasks with inter-dependencies such as data denoising, acquisition bias correction, data modelling and measurements on models

Depending on the analysis objectives, the chain of elementary tasks to apply is not necessarily linear and there is often a graph of interconnected tasks to deal with. High-level tools for expressing and handling the computation flow are therefore expected to ease medical experiment development. Workflow processing is a thoroughly researched area. When dealing with medical experiments, the user often needs to process datasets made of e.g. hundreds of individual images. The workflow management is therefore data driven and the scheduler responsible for distributing the computational load should take into account the input dataset as well as the workflow graph topology. In the context of the AGIR project, we are studying workflow processing of medical datasets on grids. We are mostly interested in workflow management systems that are compliant to recent grid standards (and especially Web Services which will largely be involved in the forthcoming WSRF specification), interoperable, and able to deal with large datasets [2], [7], [12], [13], [29].

Grid-enabled workflows provide a tool for easily setting up medical experiments on large datasets. This is tremendously important for areas with large datasets processing needs such as epidemiology or model computation and statistic extraction from populations of data. Other applications areas which involve repeated computation (e.g. Monte Carlo simulation or algorithm parameter optimization) are also concerned. We similarly see workflows as an important milestone to design algorithm evaluation procedures. In the absence of ground truth (which is usually the case in medical image processing), evaluation of the accuracy and robustness of image processing algorithms is very difficult. A solution that has been proposed recently is to establish a bronze standard [20] by considering the "exact result" as an unknown variable that has to be estimated as well (along with its accuracy).

The target application is medical image registration. Registration aims at transforming one image to make it fit another one according to some similarity criterion and a suitable transformation set. Rigid registration is a particular case where the input image is transformed by a single translation and rotation to the target image, thus bringing two different data in a single frame. Nonrigid registration aims at introducing more subtle transformations capable of compensating for differences in shapes between the two images. Medical image registration is a very important step for analyzing and comparing multiple data. The accuracy and robustness of medical image registration algorithms is also difficult to assess.

The main idea to apply the bronze standard technique to registration is to use many different registration methods (different from the one to be evaluated) to register all possible pairs of images in order to better exploit the redundancy of the information. The deployment of such a method on the grid is a perfect example of a dynamic workflow involving the core medical grid services provided by the middleware. One needs to find the available image data of a given type, registration services that are able to deal with them, and lots of resources to perform all the computations.

The final goal of this task is to design and develop a grid service for the registration of medical images that will be part of a complete processing chain going from the image acquisition at the hospital, to processing and analysis on distant computing resources, to storage in a dedicated information system, and finally to use by a practitioner. The medical context is brain cancer therapy, and especially radiotherapy. In this application, the grid registration service will be used to realign images of a same patient at different time-points for follow-up (rigid and slightly non-rigid registration), and to warp a generic atlas into a specific image patient (non-rigid registration with interaction) for segmenting the brain structures in order to precisely define the areas to irradiate and those to avoid.

Fig. 9. French and chinese clinicians ausculting a chinese child during the last Chain of Hope mission in Chuxiong, PRC

VIII. Medical Data management and Humanitarian Medicine

On-line access to medical data is of real interest in clinical practice. On-line display of medical records enables remote diagnosis by medical experts who may not be available in any health center and the ability to browse medical databases for research purposes. Grids are an opportunity to record and make globally available tremendous amounts of medical data. However, the sensitivity of these data enforce the use of strict access control policies and privacy protection procedures to deny unauthorized users access to the data. Medical data encompass medical images (usually stored in DICOM format) and associated records (usually stored in associative databases). Both images and records are confidential and must not be accessible to any grid user without proper access rights verification. Records are even more sensitive than images as they may contain personal and identifying information. A grid medical data manager therefore has to protect patient information from any unauthorized user. This can efficiently be done by well established cryptographic techniques ensuring that data are not available in a readable form neither on disk, nor during network transmission. However, encryption of data require cryptographic keys management. Several techniques have been proposed in the literature to ensure protection of these keys [24] but none is implemented today in any full scale grid infrastructure to our knowledge (only prototypes have been deployed).

Since 2000, the Chain of Hope and Shanghai No.9 hospital have collaborated to send neurosurgeons and anaesthetists to the province of Yunnan (PRC). The purpose is to improve the pediatric neurosurgical treatments and aftercare in the disadvantaged provinces of China.

To help them in this exchange of data, an application called CHINA (Collaboration between Hospitals for International Neurosurgery Applications) was developed to allow the storage of relevant medical data for distant diagnosis and post-surgery follow-up. Health records are composed by medical images (CT scan, MRI, Radiography) and volatile data stored on an Oracle database at CNRS-IN2P3 Computing Centre in Lyon and accessible through a web portal located in Clermont-Ferrand. Access to the portal requires authentication and data transfer to be encrypted.

The medical relevance of the application was successfully evaluated with French and Chinese clinicians during one medical mission in China during the fall of 2004 (fig. 9). This application is going to be used to implement and test AGIR developments related to medical data management. A server for Shanghai hospital No.9 is presently being configured so that health records can be stored securely in Clermont-Ferrand and in Shanghai. Data management services will be deployed on the two servers acting as a prototype grid.

IX. Conclusion

Parallel and distributed hardware and software have become natural tools for research in hard science and engineering. The relevant skills are widespread, and end-users are able to select the adequate tools for their needs, from clusters of workstations to ultra-high performance machines. To some extent, biological sciences have also turned to high performance computing, but more massively when web-based clients hide the ugly details of the computing machinery and data access. At the other end of the landscape, content providers are also massive users of high performance computing.

Clinical research and practice is mainly to discover that computing may be more than the limited parallelism offered by a high-end workstation. Obviously, the privacy protection constrainsts hamper evolution towards grids. However, focusing on this sole issue could mask a more fundamental obstacle. The parallel and distributed systems have co-evolved with the hard science applications, on an ascending ladder of trials and improvements. Bridging the gap between clinical research and practice on one hand, and computer science and engineering on the other, has proved more difficult. One of the reasons is probably that the opportunities to gather truly multi-disciplinary teams like e.g. the BIRN project [22] are much less frequent than in the hard science field.

The basic option of the multi-disciplinary AGIR project is that the medical habits regarding computing practice (interactivity, data protection, compliance to standards) must be considered as first-order requirements, not user-interface or practical issues; one of its goal is to demonstrate that grid-enabling advanced image analysis algorithms can be not only realized, but actually used, on a production grid. Initial results on some components - interactive volume reconstruction on EGEE, grid-enabling a DCMTK interface, compression algorithms and methods - are very promising, and we hope to integrate these and further research in the near future.

References

[1] Agir web site. http://www.aci-agir.org.

[2] I. Altintas, C. Berkley, E. Jaeger, M. Jones, B. Ludaescher, and S. Mock. Kepler : Towards a grid-enabled system for scientific workflows. In *GGF10*, March 2004.

[3] F. Berman, R. Wolski, and H. Casanova et al. Adaptive Computing on the Grid Using AppLeS. *IEEE Trans. Parallel Distrib. Syst.*, 14(4):369–382, 2003.

[4] D. Berry, C. Germain-Renaud, D. Hill, S. Pieper, and J. Saltz. Report on workshop IMAGE 03: Images, Medical Analysis and Grid Environments. TR UKeS-2004-02, UK National e-Science Centre, http://www.nesc.ac.uk/technical_papers/UKeS-2004-02.pdf, Feb. 2004.

[5] S. Blake, D. Black, , M. Carlson, E. Davies, Z. Wang, and W. Weiss. An Architecture for Differenciated Services, 1998. IETF RFC 2475.

[6] A. Chandra, M. Adler, and P. Shenoy. Deadline fair scheduling: Bridging the theory and practice of proportionate-fair scheduling in multiprocessor servers. In *Proc. of the 7th IEEE Real-Time Technology and Applications Symposium*, 2001.

[7] E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, K. Vahi, K. Blackburn, A. Lazzarini, A. Arbree, R. Cavanaugh, and S. Koranda. Mapping Abstract Complex Workflows onto Grid Environments. *Journal of Grid Computing*, 1(1):9 – 23, 2003.

[8] Z. Duan, Z. Zhang, Y. T. Hou, and L. Gao. A Core Stateless Bandwidth Broker Architecture for Scalable Support of Guaranteed Services . *IEEE Trans. Parallel and Distributed Systems*, 15(2):167–182, 2004.

[9] J. Pruyne et al. WS-Agreement specification draft, June 2004. https://forge.gridforum.org/projects/graap-wg/document/draft-ggf-graap-agreement-0.9.pdf/en/1.

[10] I. Foster, M. Fidler, A. Roy, V. Sander, and L. Winkler. End-to-End Quality of Service for High-end Applications. *Computer Communications*, 27(14):1375–1388, 2004.

[11] I. Foster, C. Kesselman, J. M. Nick, and S. Tuecke. Grid Services for Distributed System Integration. *Computer*, 35(6):37–46, 2002. extended version: The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, GGF June 2002.

[12] I. Foster, J. Voeckler, M. Wilde, and Y. Zhao. Chimera: A Virtual Data System for Representing, Querying and Automating Data Derivation. In *14th Conference on Scientific and Statistical Database Management*, Edinburgh, Scotland, July 2002.

[13] N. Furmento, A. Mayer, S. McGough, S. Newhouse, T. Field, and J. Darlington. ICENI : Optimisation of component applications within a Grid environment. *Journal of Parallel Computing*, 28(12):1753 – 1772, 2002.

[14] C. Germain, R. Texier, and A. Osorio. Exploration of Medical Images on the Grid. *Methods of Information in Medecine*, 2005. To appear.

[15] I.Foster, J. Insley, G. Laszewski, C. Kesselman, and M.Thiebaux. Distance Visualization: Data Exploration on the Grid. *Computer*, 32(12):36–43, 1999.

[16] I. E. Magnin J. Montagnat, V. Breton. Partitioning medical image databases for content-based queries on a grid. *Methods of Information in Medecine*, 2005. To appear.

[17] E. Jeannot, B. Knutsson, and M Björkman. Adaptive Online Data Compression. In *IEEE High Performance Distributed Computing (HPDC'11)*, pages 379 – 388, Edinburgh, Scotland, July 2002.

[18] T. Mäkelä, Q. C. Pham, P. Clarysse, J. Nenonen, J. Lötjönen, O. Sipilä, H. Hänninen, K. Lauerma, J. Knutti, T. Katila, and I.E. Magnin. A 3D model-based registration approach for the PET, MR and MCG cardiac data fusion, medical image analysis. *Medical Image Analysis*, 7(3):377–389, 2003.

[19] K. Nahrstedt, Hao hua Chu, and S. Narayan. Qos-aware resource management for distributed multimedia applications. *J. High Speed Netw.*, 7(3-4), 1999.

[20] S. Nicolau, X. Pennec, L. Soler, and N. Ayache. Evaluation of a new 3d/2d registration criterion for liver radio-frequencies guided by augmented reality. In N. Ayache and H. Delingette, editors, *International Symposium on Surgery Simulation and Soft Tissue Modeling (IS4TM'03)*, volume 2673 of *Lecture Notes*

in Computer Science, pages 270–283, Juan-les-Pins, France, 2003. INRIA Sophia Antipolis, Springer-Verlag.

[21] A. Osorio, O. Traxer, S. Merran, F. Dargent, X. Ripoche, and J. Atif. Real time fusion of 2D fluoroscopic and 3D segmented CT images integrated into an Augmented Reality system for percutaneous nephrolithotomies (PCNL). In *InfoRAD 2004, RSNA'04*, Nov. 26 - Dec. 4, 2004. Award: Cum Laude.

[22] S.T. Peltier and M.H. Ellisman. *The Biomedical Informatics Research Network*, chapter 8. Elsevier, 2004.

[23] Q. C. Pham, F. Vincent, P. Clarysse, P. Croisille, and I.E. Magnin. A FEM-based deformable model for the 3D segmentation and tracking of the heart in cardiac MRI. In *2nd International Symposium on Image and Signal Processing and Analysis (ISPA 2001)*, pages 250–254, Pula, Croatia, 2001.

[24] L. Seitz, J.-M. Pierson, and L. Brunie. Key Management for Encrypted Data Storage in Distributed Systems. In *IEEE Security in Storage Workshop*, pages 20–30, Washington DC, USA, October 2003.

[25] V. Servois, A. Osorio, and J. Atif et al. A new pc based software for prostatic 3d segmentation and volume measurement. application to permanent prostate brachytherapy (ppb) evaluation using ct and mr images fusion. InfoRAD 2002 - RSNA'02, 2002.

[26] A. Srinivasan and J. H. Anderson. Efficient scheduling of soft real-time applications on multiprocessors. *Jal. Embedded Computing*, 1(3):1–14, 04.

[27] Osamu Tatebe, Youhei Morita, Satoshi Matsuoka, Noriyuki Soda, and Satoshi Sekiguchi. Grid Datafarm Architecture for Petascale Data Intensive Computing. In *2nd IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2002.

[28] D. S. Taubman and M. W. Marcellin. *JPEG 2000, Image Compression, Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2002.

[29] I. Taylor, M. Shields, I. Wang, and R. Philp. Grid Enabling Applications Using Triana. In *Workshop on Grid Applications and Programming Tools – In conjunction with GGF8 jointly organized by : GGF Applications and Testbeds Research Group and GGF User Program Development Tools Research Group*, Seattle, USA, June 2003.

[30] T. Voison, L. Guillemot, and J.M. Moureaux. Image compression using lattice vector quantization with code book shape adapted thresholding. In *ICIP 2002, IEEE International Conference on Image Processing*, Rochester, NY, USA, September 2002.

[31] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.

[32] Jakob Ziv and Abraham Lempel. Compression of Individual Sequences via Variable-Rate Coding. *IEEE Transactions on Information Theory*, 24(5):530–536, 1978.