Designing Grid-enabled Image Registration Services for MIAKT

Yalin Zheng, Christine Tanner, Derek L. G. Hill and David J. Hawkes

Division of Imaging Sciences, GKT Medical School, King's College London, London, SE1 9RT, UK. {**yalin.zheng**, christine.tanner, derek.hill, david.hawkes}@kcl.ac.uk

Abstract

The MIAKT project aims to develop collaborative problem solving environments in medical informatics with particular application to the management of breast cancer. This paper presents a framework of image registration Grid services that forms the key components of this project. In this work, the concept of web services was used to encapsulate the registration grid applications for better interoperability. Globus and Condor were jointly used to exploit their complementary features. A workflow was designed for management of the image registration jobs. The initial experiments in 3D breast magnetic resonance (MR) images have shown encouraging results and demonstrated the suitability of Grid technology to this type of computationally intensive applications. The image registration Grid makes it much more straightforward for different institutes to use identical registration program and protocols to register images consistently, quickly and efficiently. This can greatly improve data sharing and comparative studies in multi-center trials. The Grid presented here could be an important step for clinical applications of image registration.

Introduction

The MIAKT project (Medical Imaging and Advanced Knowledge Technologies) aims to develop collaborative problem solving environments in medical informatics [1]. The initial effort of the project is to support the screening and diagnosis of breast cancer. In addition to the fact that breast cancer is the second most common cause of death in women, the domain has been chosen because it contains a number of characteristics (i.e., large amount of data, computationally intensive image analysis and the collaborative and distributed nature of the task) that makes it especially valuable for the application of knowledge technologies and image analysis. The project has developed a number of services for the purpose of supporting multi-disciplinary meeting, i.e., ontologies of X-ray mammogram, MR mammography and breast pathology, X-ray mammography analysis services, automatic patient reporting services and grid enabled breast MR image analysis services. As part of MIAKT project, we have developed Grid services for registering breast MR images.

Breast cancer screening programmes involve taking regular X-ray mammograms and have been shown to be an effective way for early detection of breast cancer. However, x-ray mammography involves the use of ionizing radiation and has poor sensitivity for lesions in dense breasts. MR imaging was introduced about 25 ago and has been regarded as a promising supplement tool to mammography in the diagnosis of breast cancer, especially useful for monitoring young women at high risk. 3-D dynamic contrast enhanced MR is used in the breast management, that is, MR images are acquired before and after injection of contrast agent. Radiologists will classify lesions by studying the enhancement patterns of the image sequence. However, motion artifacts are often introduced during the image acquisition and these pose extra difficulties for radiologist to make decisions as the enhancement could be caused either by motion artifacts or by lesions. Therefore, minimizing these artifacts will be highly desired in clinics and this could be achieved by image registration technique.

Image registration is a procedure to establish spatial correspondence between two images by aligning one image (source image) to another (target or reference image). A linear or non-linear transformation (corresponding to rigid and non-rigid registration, respectively) between them can be determined as a result of registration [2]. Proper registration enables us to have a better understanding of the features of interest and allows integration of useful information. Applications are wide: ranging from image-guided surgery to atlas construction, segmentation propagation, monitoring changes over time and dynamic sequence analysis. The general procedure of fully automated image registration typically requires optimization of a function of the similarity between the target and source images. The intensive computations involved in image registration have to some extent made it impractical for interactive use as well as limiting its general availability [3]. Moreover, a significant feature of clinical applications is that the data to be analyzed are often from a large number of individual subjects in order to achieve statistical significant results. The requirement to carry out non-rigid registration of large numbers of dataset inevitably raised interest in finding an effective way to deliver a registration service in an efficient manner.

This work is an effort to harness the Grid technology in order to facilitate the processing of a huge number of image registration tasks and greatly shorten waiting time. This application will also make the collaboration from different institutes much easier. The paper is organized as follows. In section 2 we briefly review the grid technology, web services and the image registration algorithms. In section 3 we discuss the system design. Section 4 shows some initial results obtained from experiments with this framework and discuss some implementation issues and in section 5 we summarize the paper.

Technology Review

Computing Technology

Computational Grid

The concept of the Grid was introduced in the late 1990s. In the past few years there is strong interests in the development of Grid technology. The Grid technology has many desirable features such as security, resource discovery, data management, fault detection, data and processing provenance and portability. It has been shown that the Grid can tackle complex problems in reasonable time scales and with convincing performance and this has further boosted the development of Grid technologies. With the effort from IT industry, the Grid technology is evolving to a standard technology for industry rather than only for research.

Globus Toolkit is an open-architecture, open-source set of services and software libraries that support Grids and Grid applications [4]. Grid application developers can build higher-level applications on top of the low level functionality provided by Globus. It consists of three essential services to address the key issues involved in the Grid, namely, the Grid Resource Allocation and Management (GRAM), the Meta Directory Service (MDS), and the Grid Security Infrastructure (GSI). The last two years witnessed a dramatic change in Globus. In the mid of 2003, GT3 alpha, also know as Open Grid Services Architecture (OGSA) was released aiming at development of Grid Services. The latest version now is GT 3.2 and this will be further migrated into Web Services Resource Frameworks (WSRF) at the end of 2004 [5]. WSRF is the merging point of Grid and Web Services and is regarded as the future industry standards.

Condor is a software system that creates a High-Throughput Computing (HTC) environment. It effectively utilizes the computing power of workstations that communicate over a network [6]. Condor is a batch system for managing computing-intensive jobs but with the capability to manage non-dedicated machines. A number of machines can be configured as a Condor pool where there is a dedicated Condor Master machine, several Condor Submitter machines and Condor Scheduler machines. Condor provides queuing mechanisms, scheduling policies, priority schemes, and resource classifications. Users submit jobs to the Condor pool and Condor will look after the jobs for users. Condor has many attractive features such as checkpoint mechanism, dependency management and queuing.

Web Services

Web service is a collection of protocols and standards enabling applications to make calls to remote applications over the Internet [7]. The key benefit is that applications written in various languages running on various operating systems can use web services to exchange data easily between them. This interoperability is achieved by using the standard technologies know as SOAP (Single Object Access Protocol), WSDL (Web Services Description Language) and UDDI (Universal Description, Discovery and Integration).

SOAP is a lightweight protocol for exchanging messages between different services. WSDL describes the interface of particular web services by defining the syntax, the semantics, and all the various administrative aspects of the service. UDDI defines a number of lookup services aimed at allowing clients to look up and retrieve the required information to access a web service. SOAP, WSDL and UDDI are all based on XML (Extensible Markup

Language), a standard technology for data representation and integration. This is the main reason why web services can seamlessly communicate through different applications. In practice, the service provider publishes a service to an external repository (UDDI). Then the potential client can look it up from the repository and retrieve the information about the service, i.e. the format of the procedure calls and the address of a service provider, respectively. Based on the information retrieved, the client can access that service to use whatever functionality the service provides.

Image Registration

The designed Grid services currently provide both rigid and non-rigid registration services, in particular, the free-form registration algorithm extended with volume preserving constraints was used for the nonrigid registration of contrast-enhanced breast MR images in dynamic contrast enhanced volume acquisitions [8]. The optimal parameter settings and the accuracy of the algorithm at registering breast MR images have been validated on the basis of simulated gold standard deformations by employing Finite Element biomechanical models [9]. These optimal settings were provided as default settings for the image registration Grid. The Grid will however also allow the potential users to define and use their favorite settings.

System Design

Combination of Globus and Condor

As discussed earlier, there is a rapid transition over the globus toolkit. This reflects the strong driving forces from both research institutes and IT industry. Some technically unusable versions were however released and the cost for Grid developers of tracking the latest version was in our experience high. Therefore, in this project GT2.4 was selected for the basic Grid functionalities such as job submission, management and security. Web Services technology was used to encapsulate the registration applications. This could be a pragmatic solution for building Grid Services currently as Globus is still undergoing large changes. This work could be updated to the WSRF when WSRF is technically acceptable for use. Another gain of using GT2.4 is that Globus provides an easy interface for accessing Condor. The combination of Condor and Globus together can take advantage of good features from both of them. The version we have used is Condor 6.4.5 for the Linux platforms. Figure 1 illustrates this combination. The Globus component provides user functionalities such as mutual authentication, secure job submission and data transfer while Condor is managing those submitted jobs by maintaining a HTC environment.

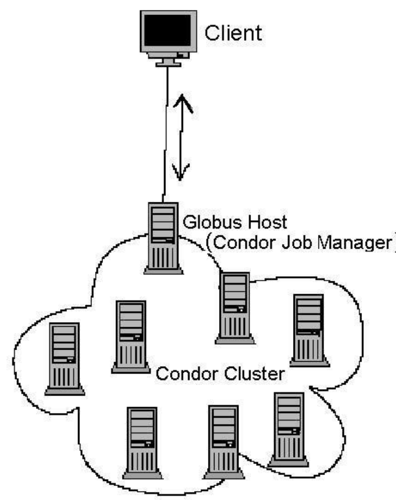


Figure 1. The combination of Globus and Condor.

Web Services

The Java code is based on AXIS 1.1 [10] and Java SDK1.4.2 to develop the Web Services. Apache Tomcat 5.1.16 was used as the Web server and servlet container [11]. Figure 2 presents part of the WSDL file describing the registration services. From Figure 2 we can see there are rigid and nonrigid registration services available. Information about these services such as access point, the parameters required to call the rigid or nonrigid registration services, etc. are also defined in this WSDL file. Based on this information, clients can dynamically invoke the services.

```
- <wsdl:portType name="Registration">
  + <wsdl:operation name="rigid" parameterOrder="targetImage sourceImage
      parameterFile divide interpolation remoteHost remotePort remoteDir
      user">
  + <wsdl:operation name="nonrigid" parameterOrder="targetImage
      sourceImage parameterFile divide controlPointSpacing mu
      interpolation remoteHost remotePort remoteDir user">
  </wsdl:portType>
+ <wsdl:binding name="registrationSoapBinding" type="impl:Registration">
- <wsdl:service name="RegistrationService">
  - <wsdl:port binding="impl:registrationSoapBinding" name="registration">
      <wsdlsoap:address location="http://is-
        210.umds.ac.uk:8080/axis/services/registration" />
    </wsdl:port>
  </wsdl:service>
```

Figure 2. The WSDL file of the Registration Services

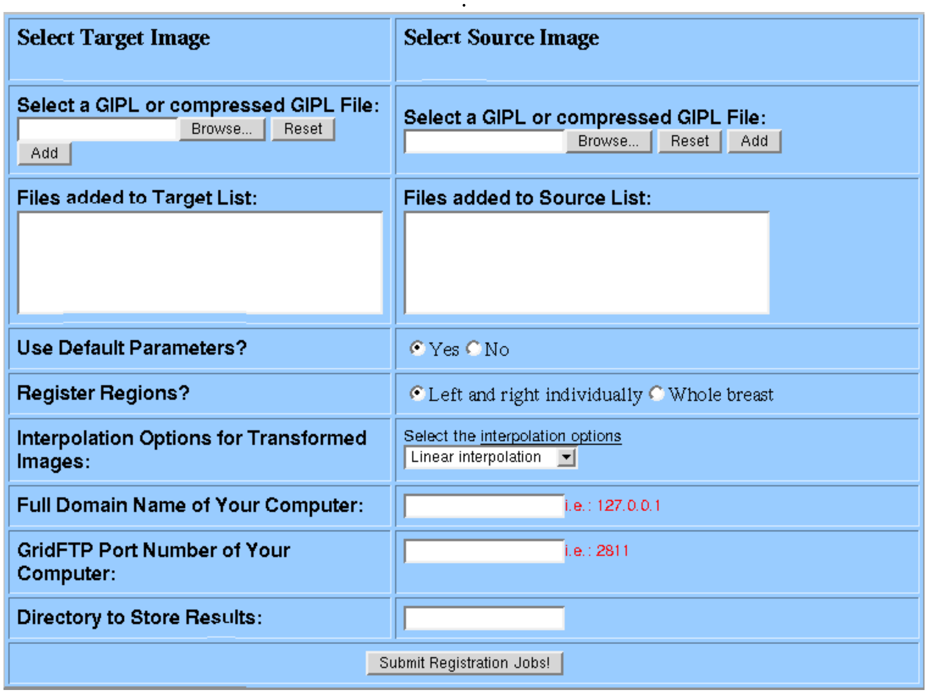


Figure 3. The Client Interface.

Currently we have developed a simple web interface for clients to invoke the services. Figure 3 shows one of these forms. This interactive interface enables the user to assemble the registration tasks by specifying target images, source images, image registration protocols and parameters (the tested default parameter settings are also provided), etc and to submit their registration tasks.

Workflow

Workflow is an automation procedure to manage computational jobs according to a set of predefined rules. There are some basic requirements for it such as the need to support sequential and parallel flows, looping and conditionals, recursion and data managements. Given the substantial volume of data and computational jobs involved in the Grid application, workflow becomes a key component in the Grid technology. In our work, we mainly focused on the dependency problems among a cohort of registration jobs and this will be discussed in detail next.

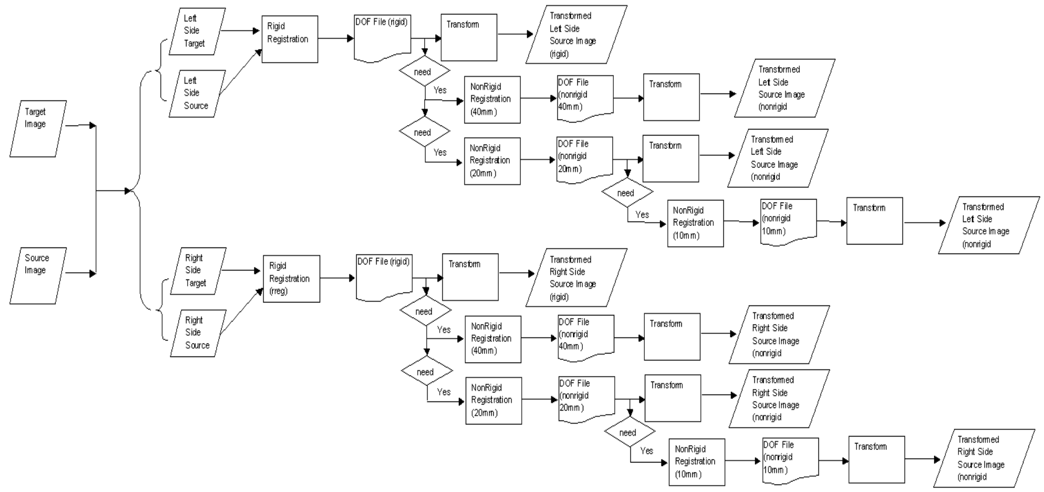


Figure 4. Workflow for non-rigid registration of breast MR image.

In breast MR image registration regions including the left or the right sides of the breast are often considered individually in order to minimize any possible motion interdependency between them and to reduce the computational demands. There are different job dependencies between rigid registration and nonrigid registration jobs. Any rigid registration job will not start until the target image, source image, parameter file or any required initial transformation estimates are ready. This becomes more complicate for the case of nonrigid registration. As we discussed earlier, rigid registration is often used for providing initial transformation estimation for the nonrigid registration. Generally nonrigid registration jobs will therefore only launch after the rigid registration that provides the initial estimate has completed. Furthermore, different control point spacings can be used in nonrigid registration. Registration with finer control point spacing has been shown to improve the alignment of the whole breast, while the coarser spacing improves registration accuracy of the enhancing lesions [8]. In our application, default parameter settings for nonrigid registrations with control point spacings of 40 mm, 20 mm and 10 mm were provided. Nonrigid registrations with 40mm and 20mm control point spacings will use the estimated transformation from rigid registration while nonrigid registration with 10mm control point will take the estimate from the nonrigid registration with 20mm control point spacing. If the user chooses to have the registered images as well as the transformation results, the program to generate these images should be called after the complete of corresponding registration program. Figure 3 presents the workflow for nonrigid registration.

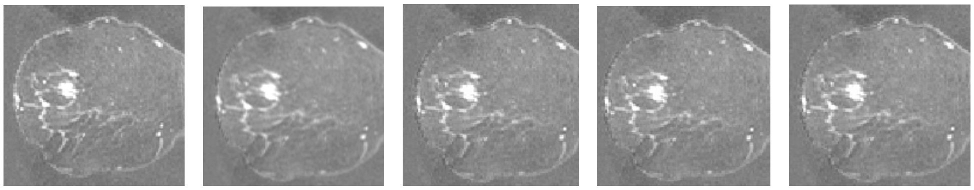
Results and Discussions

The image registration Grid has been validated, before processing a large number of registration jobs, by ensuring that the same numerical results were obtained as those computed directly from the Condor application for the same dataset.

Before we move on, it is necessary to address firewall issues. Firewall is a common means to protect the security of the network by appropriate packet filtering. Different institutions have their own firewalls to protect themselves from malicious attacks from outside. When the service provider and the client are on different sides of a firewall, if the firewall is not configured for the communication between them, their communication will be blocked. By their nature, Grid applications require communication between client and service provider. Therefore, firewall configurations become a key issue for the Grid technology.

Test One

To avoid the complexity with configuring the firewall, this test was conducted on the system within our secure site. In this test, a Condor cluster with 40 Linux machines was involved (see [3] for technical details). Figure 5 shows an example slice of the difference images.



(a) (b) (c) (d) (e)

Figure 5. Illustrative coronal slices of the right breast MR images showing the differences before and after registrations. From left to right are (a) before registration; (b) after rigid registration; after nonrigid registration with (c) 40 mm, (d) 20 mm and (e) 10 mm control point spacing.

Job No.	Target Image	Source Image	Side	Registrion Type	Status
1	b024A-04 giplZ	b024A-09 gipl.Z	left	rigid	DONE
2	b024A-04 giplZ	b024A-09 gipl.Z	right	rigid	DONE
3	b024A-04 giplZ	b024A-08 gipl.Z	left	rigid	DONE
4	b024A-04 giplZ	b024A-08 gipl.Z	right	rigid	DONE
5	b024A-04 giplZ	b024A-07 gipl.Z	left	rigid	ACTIVE
6	b024A-04 giplZ	b024A-07 gipl.Z	right	rigid	DONE
7	b024A-04 giplZ	b024A-06 gipl.Z	left	rigid	DONE
8	b024A-04 giplZ	b024A-06 gipl.Z	right	rigid	DONE
9	b024A-04 giplZ	b024A-05 gipl.Z	left	rigid	ACTIVE
10	b024A-04 giplZ	b024A-05 gipl.Z	right	rigid	ACTIVE

(a) status of rigid registration jobs

Job No.	Target Image	Source Image	Side	Registrion Type	Status
1	b024A-04.giplZ	b024A-09.giplZ	left	nonrigid: 40mm	ACTIVE
2	b024A-04.giplZ	b024A-09.giplZ	right	nonrigid: 40mm	ACTIVE
3	b024A-04.giplZ	b024A-08.giplZ	left	nonrigid: 40mm	ACTIVE
4	b024A-04.giplZ	b024A-08.giplZ	right	nonrigid: 40mm	PENDING
5	b024A-04.giplZ	b024A-07.giplZ	left	nonrigid: 40mm	WAITING
6	b024A-04.giplZ	b024A-07.giplZ	right	nonrigid: 40mm	ACTIVE
7	b024A-04.giplZ	b024A-06.giplZ	left	nonrigid: 40mm	PENDING
8	b024A-04.giplZ	b024A-06.giplZ	right	nonrigid: 40mm	PENDING
9	b024A-04.giplZ	b024A-05.giplZ	left	nonrigid: 40mm	WAITING
10	b024A-04.giplZ	b024A-05.giplZ	right	nonrigid: 40mm	WAITING

(b) status of non-rigid registration jobs

Figure 6. Job running status for 5 MR images to the same reference image (WAITING denotes the task is waiting for necessary data, ACTIVE denotes the task is running, DONE means the task has been completed and PENDING denotes a running job is suspended because of current shortage of available computers in Condor pool).

Figure 6 shows a table captured from the web interface updating users with the progress of their registration jobs. From Figure 6 we can see that 5 source images were registered to the same target image. There are total 10 rigid and non-rigid registration jobs respectively for registering both sides of the breast. At this time point, Figure 5 (a) shows that all rigid registrations, except for job 5, 9 and 10, have been completed. Figure 5 (b) shows the corresponding non-rigid registration. Due to the dependencies on rigid registration, job 5, 9 and 10 are still waiting for the initial estimation from the rigid registration. Of the other 7 jobs, 4 are active while the other 3 are pending which means that there are no available resources for them. It is worth mentioning that there are jobs from other colleagues running in the Condor cluster during these tests.

As introduced earlier, 40 execute Condor machines ideally enable 40 registration jobs to run simultaneously so that it will significantly reduce the waiting time for large batch jobs (1/40 waiting time). In practice it will take slightly longer as there are always more than one user using the system. For example, in a test the total CPU time required by 12 jobs was 127.8 hours while the total waiting time was 30.9 hours. This shows that a significant time saving can be achieved despite the use by others. The maximum waiting time for all submitted registration jobs largely depends on the job that takes the longest time as the average time spent setting up becomes negligible if jobs take a long time to execute. However, how to evaluate the overall system performance remains to be investigated in the future.

Test Two

In this test, we have the Institute of Cancer Research (ICR), UK involved as client for the remote access of the services. We have built a miniature system in our demilitarized zone (DMZ) and this system comprises a Globus Server and a Condor cluster with 5 execute machines. Globus 2.4 was installed on a desktop machine at ICR and this machine was configured as Globus client. Following the requirements provided by Globus team [12], we have configured the firewalls at both sites for the communication between server and client. We have also opened HTTP port for the computer at ICR to access our web interface.

In this study, 3-D contrast-enhanced MR breast image sequences collected for the MARIBS project [13] (Magnetic Resonance Imaging in Breast Screening) were used. Led by ICR, the MARIBS project recruited about 500 women below the age of 50 at high genetic risk of breast cancer per year for three years and the annual MRI and x-ray mammography will be collected for up to five years. 17 screening centers from all regions of UK participated in this study.

We have conducted some tests on the system. Client at ICR submitted their registration jobs to the KCL Computer resources. The system will transfer the images from the client to the server side and the results are transferred back to the client after the registrations were completed. The identical registration results were obtained compared to the local tests on the same dataset. An import issue for the remote access of computer resources is the transfer rate of data. We have tested this locally and across the network between KCL and ICR. In the local test, we transferred one file between Globus client and server machines and repeated for 10 times. The average transfer rate is 9.88 Mbps. Then we transferred the same file between KCL and ICR. The average transfer rate is 0.75Mbps. The major reason for this large difference might come from the network speed. We have a 100Mbps network connection locally while the client at ICR just has a 10Mbps network. The above result was obtained on transferring a file of large size. We also tested transferring seven files between KCL and ICR. The total size of these seven files is as same as the file used in the previous tests. The average speed over 10 tests is 0.685 Mbps. The loss of transfer speed could be the overhead involving more initializations of GridFTP servers.

The success of the test is encouraging, as this a starting point for the access from other screening centers and could eventually lead to multiple center trials. This will be valuable as all data will be processed by identical registration program and thus the results are comparable; also this can be a sensible way to process large data set through the Grid.

Discussions

In this work we have used GT2.4 rather than GT3. The reasons for doing this are multifold. Although GT3 was developed to provide Grid services for solving the stateless and non-transient problems inherent with web services, this solution however is not welcomed by the IT industry because it is incompatible with Web Services. Another reason is that it did not solve some fundamental problems such as resource management and job prioritization. Finally GT3 is evolving into WSRF and it will be unwise to use the unstable versions during this transition. In fact, it should not be very difficult for us to adapt to WSRF when it is finally ready.

At present, potential users still have to configure Globus. This is unlikely to be practical as clients will not be computer experts. Future work will investigate the feasibility of using the HTTPS protocol for authentication and data transfer which would allow the clients to access the services just with a valid UK e-Science certificate. Furthermore, it would reduce the firewall configuration work for clients.

Conclusion

In this paper we presented a framework of image registration services. Globus and Condor were jointly used to exploit their complementary features. A workflow was designed for management of the image registration jobs. The use of web services concept provides better interoperability with other services. The initial experiments in 3D breast MR images have shown encouraging results and demonstrated the suitability of Grid technology to this type of computationally intensive applications and the potential to be used for registering large amount of image dataset and multiple screening center trials.

Future work will proceed in two directions. So far, the tests were based on rather small systems and only one remote client were involved to test the system. In the future we will undertake more extensive tests. For example, we are investigating the feasibility of accessing computer resources with over 2000 processors provided by the National Grid Service (NGS) [14]. This means that we could further accelerate the computational ability of the system and to make registration of images from other screening centers feasible. These tests will be essential for us to evaluate the system performance such as scalability, reliability and time-saving factors. It is believed that ultimately the image registration Grid will provide a resource that can provide a powerful, yet userfriendly tool for clinicians registering a large number of images.

ACKNOWLEDGMENT

This work was supported by U.K. EPSRC e-Science MIAKT project under Grant GR/R85167.

REFERENCES

[1] Grid enabled knowledge services: collaborative problem solving environments in medical informatics, http://www.gridoutreach.org.uk/docs/pilots/griden.htm

[2] J. V. Hajnal, D. L. G. Hill and D. J. Hawkes, *Medical Image Registration*, CRC Press, 2001.

[3] Y. Zheng, C. Tanner, D. L. G. Hill, D. J. Hawkes, M. J. White, M. Khazen and M. O. Leach, *Alignment of dynamic contrast enhanced MR volumes of the breast for a multicenter trial: an exemplar grid application*, Medical Imaging, SPIE 2004, 14-19 February 2004, San Diego, CA USA.

[4] Globus: http://www.globus.org/

[5] K. Czajkowski, D. Ferguson, I. Foster, J. Frey, S. Graham, T. Maguire, D. Snelling and S. Tuecke, *From Open Grid Services Infrastructure to WS-Resource Framework: Refactoring & Evolution*, http://www-fp.globus.org /wsrf/OGSI to WSRF 1.0.pdf

[6] Condor: http://www.cs.wisc.edu/condor/

[7] S. Cable, B. Galbraith, R. Irani, M. Hendricks, J. Milbury, T. Modi, A. Tost, A. Toussaint and J. Basha, *Professional Java Web Services*, Wrox Press Inc., 2002.

[8] C. Tanner, J. A. Schnabel, A. Degenhard, A. Castellano-Smith, C. Hayes, M. O. Leach, D. R. Hose, D. L. G. Hill and D. J. Hawkes, *Validation of Volume-Preserving Non-Rigid Registration: Application to Contrast-Enhanced MR-Mammography*, MICCAI'02, Tokyo, Japan, vol. 2489 of Lecture Notes in Computer Science, Springer Verlag, pp. 307-314, 2002

[9] J. A. Schnabel, C. Tanner, A. D. Castellano-Smith, A. Degenhard, M. O. Leach, D. R. Hose, D. L. G. Hill and D. J. Hawkes, Validation of Non-Rigid Image Registration using Finite Element Methods: Application to Breast MR Images, IEEE Transactions on Medical Imaging, vol. 22(2), pp. 238-247, 2003.

[10] Axis: http://ws.apache.org/axis/

[11] Tomcat: http://jarkata.apache.org/tomcat/

[12] Von Welch, Globus Toolkit Firewall Requirements, version 5, July, 2003, http://www.globus.org/security/v2.0/Globus%20Firewall%20Requirements-5.pdf

[13] J. Brown, D. Buckley, A. Coulthard, A. K. Dixon, J. M. Dixon,D. F. Easton, R. A. Eeles, D. G. R. Evans, F. G. Gilbert, M. Graves, C. Hayes, J. P. R. Jenkins, A. P. Jones, S. F. Keevil, M. O. Leach, G. P. Liney, S. M. Moss, A. R. Padhani, G. J. M. Parker, L. J. Pointon, B. A. J. Ponder, T. W. Redpath, J. P. Sloane, L. W. Turnbull, L. G. Walker and R. M. L. Warren, Magnetic Resonance Imaging Screening in Women at Genetic Risk of Breast Cancer: Imaging and Analysis Protocol for the UK Multicentre Study, Magnetic Resonance Imaging, vol. 18, pp.765-776, 2000.

[14] National Grid Service, http://www.ngs.ac.uk/