

# Das MediGRID Projekt

## Gridcomputing in der medizinischen Bildverarbeitung

Michal Vossberg, Dagmar Krefting und Thomas Tolxdorff

Institut für Medizinische Informatik,  
Charité – Universitätsmedizin Berlin, 12200 Berlin  
Email: [michal.vossberg@charite.de](mailto:michal.vossberg@charite.de)

**Zusammenfassung.** Aufgrund zunehmend höherer Anforderungen an Rechenleistung und Speicherplatz in der medizinischen Bildverarbeitung ist Gridcomputing auch für dieses Arbeitsgebiet eine vielversprechende Perspektive. Im Rahmen der D-Grid Initiative des BMBF wird durch das MediGRID-Projekt eine bundesweite Grid-Infrastruktur für die Lebenswissenschaften aufgebaut, die insbesondere die hohen Anforderungen an Sicherheit und Zuverlässigkeit berücksichtigt. Das Modul Bildverarbeitung im MediGRID implementiert prototypische Anwendungen aus der medizinischen Bildverarbeitung und analysiert Möglichkeiten und Beschränkungen der Grid-Nutzung. Erste Algorithmen sind bereits im MediGRID implementiert und über ein Portal von jedem internetfähigen Webbrowser aus nutzbar.

## 1 Einleitung

Viele Anwendungen in der medizinischen Bildverarbeitung benötigen eine hohe Rechenleistung oder haben einen hohen Speicherplatzbedarf. Zusätzlich werden in der medizinischen Informatik Algorithmen entwickelt, die auch weiteren Nutzern zur Verfügung gestellt werden sollen. Diese Anforderungen kann ein Computergrid erfüllen, das den Zusammenschluss von räumlich verteilten Rechnern zu einem virtuellen Gesamtcomputer ermöglicht. Ein solches Grid soll bundesweit durch die D-Grid-Initiative aufgebaut werden, die derzeit mehr als 100 Einrichtungen gezielt in dem Bereich Grid-Forschung fördert.

MediGRID als Communityprojekt für die Medizin und biomedizinische Informatik hat das Ziel, in den kommenden Jahren schrittweise eine Grid-Infrastruktur für die biomedizinische Verbundforschung aufzubauen und nachhaltig zu betreiben. Aufgrund der besonderen Anforderungen an Sicherheit und Zuverlässigkeit sind Grid-Netzwerke im medizinischen Umfeld noch recht neu und werden vor allem in der Praxis noch nicht umfassend eingesetzt [1].

Das Modul Bildverarbeitung soll die wesentlichen Methoden medizinischer Bildverarbeitung sowie Datenstrukturen und -organisation auf das MediGRID bringen. Ziel ist es, wissenschaftliche Aufgabenstellungen in der medizinischen Bildverarbeitung effizienter zu lösen, neue Anwendungen zu erschließen und somit eScience im Bereich der Bildverarbeitung zu stärken. Das Methodenspektrum wird in drei praktisch relevanten Anwendungsszenarien exemplarisch in das MediGRID implementiert:

- Statistische Analyse funktioneller Hirnbilddaten,
- Virtuelle Gefäßchirurgie,
- 3D Ultraschallaufnahmen bei Prostatabiopsien.

Besondere Berücksichtigung finden dabei die sichere Bilddatenübertragung und -speicherung, effizienter (inhaltsbasierter) Zugriff auf Bilddaten und die Interoperabilität verschiedener Applikationen über den DICOM-Standard.

## 2 Stand der Forschung

Derzeit wird Gridcomputing vor allem von Forschungscommunities betrieben, die keine hohen Anforderungen an Benutzerfreundlichkeit und Datensicherheit stellen, beispielsweise die Hochenergiephysik. In der letzten Zeit sind jedoch verstärkt auch Benutzerschnittstellen und Konzepte zur sicheren Nutzung der Middleware entwickelt worden.

Auf der anderen Seite werden medizinische Bildverarbeitungsprobleme zurzeit hauptsächlich durch für eine Modalität oder Organ hochspezialisierte und optimierte Algorithmen gelöst. Diese sind von einer Gruppe oder Organisation lokal implementiert. Oft sind aber viele Teilschritte universeller einsetzbar und könnten auch von anderen Forschergruppen vorteilhaft eingebracht oder genutzt werden. Zusätzlich gibt es viele Dienste die völlig unabhängig vom Bildmaterial sind oder parallel zum Hauptworkflow laufen.

Einzelne Bildverarbeitungsschritte aus der medizinischen Forschung sind bereits in verschiedenen Grid-Projekten implementiert worden [2, 3]. Jedoch sind die Lösungsansätze aufgrund mangelnder nachhaltiger Infrastruktur, nichtstandardisierter Middleware und der oben genannten besonderen Anforderungen nicht weiterverfolgt worden und haben keine weitere Verbreitung in der Bildverarbeitungscommunity gefunden.

Alle drei ausgewählten Szenarien der medizinischen Bildverarbeitung beinhalten komplexe, rechenintensiven Algorithmen und umfassen große Datenmengen. Das hier ausgewählte Projekt zur Bildverarbeitung von 3D-Ultraschallbildern von Prostatabiopsien umfasst in seinem vollen Umfang neben Segmentierungs-, Registrierungs-, Klassifikations- und Visualisierungsalgorithmen auch die Anbindung an ein PACS, den Zugriff auf eine Datenbank innerhalb des Grids sowie den Anschluss an ein Image Retrieval System [4].

Ziel des Projektes ist die Bestimmung der räumlichen Lage der Gewebeproben im Prostatavolumen zur besseren Diagnose und Therapieplanung von Prostatakarzinomen, eine der häufigsten Krebserkrankungen von Männern. Besonders rechenintensiv ist die 2D-3D-Registrierung der Ultraschallaufnahmen. Desweiteren wird die Segmentierung der Biopsienadel auf zahlreiche Bildsequenzen angewendet, die ohne Parallelisierungsaufwand auf mehrere Computer aufgeteilt werden können. Die Vorteile einer Benutzung von MediGRID liegen deshalb auf der Hand:

1. Schnellere Ausführung der rechenaufwändigen Algorithmen durch Aufteilung des Problems und Verteilung auf Hochleistungsrechner im Grid,

2. Nutzung zusätzlicher Methoden, wie Klassifizierung oder Segmentierung, ohne Eigenentwicklung oder umständliche Hard- und Softwareinstallationen,
3. Nutzung verteilten Speicherplatzes zur Speicherung der voluminösen 3D Bilddatensätze und -filme.

### 3 Methoden

Die Algorithmen des Ultraschallprojektes, die bisher ins Grid implementiert sind, sind Teilschritte der Segmentierung und der Registrierung. Die Segmentierungsalgorithmen sind in Matlab 7.1. geschrieben, die Registrierung verwendet ITK 2.9. Für die Verwendung im Grid sind die Matlab-Skripte kompiliert und können so lizenzfrei als ausführbare Programme auf die Gridrechner verteilt werden. Die Algorithmen werden so angepasst, dass sich die Aufgaben durch Angabe entsprechender Parameter in kleinere Teilprobleme zerlegen lassen. Diese Teilaufgaben sind unabhängig voneinander und können beliebig auf die zur Verfügung stehenden Gridrechner verteilt werden.

Die Gridversionen der Algorithmen werden in einem lokalen Testbed, bestehend aus vier Arbeitsplatzrechnern, entwickelt und anschließend auf das MediGRID gebracht. Beteiligte Gridrechner sind in der derzeitigen Anfangsphase bei MediGRID zwei Hochleistungsrechencluster am Konrad-Zuse-Zentrum, Berlin und an der Gesellschaft für wissenschaftlichen Datenverarbeitung mbH in Göttingen. Prinzipiell wird freie Software eingesetzt, die unter GNU- oder ähnlichen Lizenzmodellen laufen. Als Middleware wird das weitverbreitete Toolkit Globus [5] in der Version 4.0 eingesetzt. Vorteile des Globus Toolkits ist einerseits seine starke Verbreitung und andererseits die konsequente Umsetzung moderner, diensteorientierter Technologien, wie beispielsweise WeBServices. Der Datentransfer wird über das gridftp-Protokoll gewährleistet. Diese FTP-Variante ist optimiert für Zugriffe in Gridnetzwerken und besonders sicher und effizient für höchsten Datendurchsatz. Beim Datenmanagement kommt der SRB zum Einsatz [6], der es erlaubt, transparent Dateien auf entsprechenden Datenservern zu speichern und wieder zu laden. Der Datenbankzugriff erfolgt über OGSA-DAI, das einen effizienten und sicheren Zugriff auf SQL-Datenbanken im Grid ermöglicht. Die Authentifizierung und Autorisierung der Benutzer und beteiligten Rechner wird über Globus-GSI mithilfe von DFN-Zertifikaten realisiert. Dieser Sicherheitsstandard findet unter anderem auch im Signaturgesetz Verwendung und bietet nach derzeitigen Erkenntnissen ausreichend Sicherheit zur Authentifikation der beteiligten Benutzer und zur Autorisierung ihrer gewünschten Handlung. Der Zugang zum Grid und die Steuerung der Anwendungen erfolgt über ein Webportal, das als webbasierte Anwendung von jedem internetfähigem Webbrowser aus verfügbar ist. Das in MediGRID verwendete Webportal basiert auf einem Gridsphere-Server [7]. Neben dem Zugang zum Grid und seinen Basisdiensten bietet das Webportal auch den Zugang zu den einzelnen Anwendungen. Die Entwicklung der anwendungsspezifischen Oberflächenkomponenten und ihre Integration ins Gesamtportal erfolgt in Java nach dem JSR168-Portlet Standard. Die Algorithmen, die ebenso wie die entsprechenden Bibliotheken auf den Gridrechnern installiert sein müssen, werden über einen Webservice angesprochen.

Abb. 1. Snapshot des MediGRID-Webportals



## 4 Ergebnisse

Seit einigen Monaten werden sukzessive Teilschritte der Ultraschall-Anwendung für die Arbeit mit MediGRID umgesetzt und von uns für die Experimente benutzt. Die Umsetzung beinhaltet sowohl die Anpassung der Algorithmen, die Entwicklung der Webservices als auch die Portalentwicklung (Abb. 1). Konkret lassen sich derzeit über das Webportal Bilddaten im Grid ablegen, Jobs sowohl im Batchbetrieb als auch direkt auf manuell ausgewählten Rechnern starten, ihre Ausführung überwachen und nach Beendigung auf die Ergebnisse zugreifen. Weiterhin ist es möglich, die Ergebnisse in der Dokumentationsdatenbank im Grid eintragen zu lassen oder abzufragen. Zusätzlich besteht nun die Möglichkeit, auf die Anwendung von jedem internetfähigem Webbrowser aus zuzugreifen. So können die Experimente nun einfach und ohne größere Installationen von zuhause, unterwegs oder anderen Büros aus gestartet und kontrolliert werden.

## 5 Diskussion

Zusammenfassend können wir im Fall des Ultraschallprojektes schon nach kurzer Zeit sehr positive Erfahrungen mit der Benutzung von MediGRID ziehen. Mit geringem Entwicklungsaufwand können die angebotenen Hochleistungs-Ressourcen im Grid genutzt und die Anwendung einer breiteren Benutzerschicht, wie beispielweise den an dem Projekt beteiligten Ärzten, zugänglich gemacht werden. Da allerdings das Sicherheitskonzept bisher noch nicht vollständig im MediGRID umgesetzt ist, können zur Zeit nur anonymisierte Daten verarbeitet werden. Neben zahlreichen Erweiterungen, die in MediGRID geplant sind, etwa eine Workflowengine zur automatischen Aufteilung und komfortablen Überwachung von Jobs im Grid, stehen speziell im Modul Bildverarbeitung die folgenden Entwicklungen an: Nach den ersten Erfahrungen wird die Anwendung 3D-Ultraschall nun weiter ausgebaut, und um diagnoseunterstützende Dienste erweitert. Der Schwerpunkt hierbei wird weniger im Zeitvorteil liegen als in der Tatsache dass es überhaupt möglich sein wird diese Anwendungen ohne lokale Implementationen zu nutzen. Außerdem wird im Bereich Datenmanagement an der Umsetzung des DICOM Protokolls, dem weltweiten Standard zur Bildübertragung in der Medizin, im Grid zur Anbindung von PACS Systemen gearbeitet. Neben dem vorgestellten Anwendungsszenario werden die schon beschriebenen zwei weiteren Projekte aus dem Modul Bildverarbeitung, Funktionelle Hirnbilddaten und Virtuelle Gefäßchirurgie, implementiert, um einen möglichst umfassenden Teil von typischen Methoden der Bildverarbeitung beispielhaft in MediGRID umzusetzen.

## Danksagung

Diese Arbeit ist Teil des MediGRID Projektes und wird gefördert vom BMBF, Förderkennzeichen 01AK803F.

## Literaturverzeichnis

1. MediGRID. GRID-Computing für die Medizin und Lebenswissenschaften; 2006. BMBF Förderkennzeichen 01AK803F, <http://www.medigrid.de>.
2. GEMSS. Grid-Enabled Medical Simulation Services; 2005. [Http://www.ccrlnece.de/gemss/index.html](http://www.ccrlnece.de/gemss/index.html).
3. EGEE. Enabling Grids for E-science; 2006. <Http://www.eu-egce.org>.
4. IRMA. Image Retrieval in Medical Applications; 2005. <Http://irma-project.org>.
5. Foster I, et al. Globus toolkit version 4: Software for service-oriented systems. LNCS 2005;3779:2–13.
6. Rajasekar A, et al. Storage resource broker: Managing distributed data in a grid. Computer Society of India Journal 2003;33(4):42–54.
7. Novotny J, Russell M, Wehrens O. GridSphere: An Advanced Portal Framework. <http://www.gridisphere.org>; 2005.