

Neuroimaging: SPM als verteilte Komponente in Grid- und Cluster-Architekturen

Michael Luchtman, Sebastian Baecke, Johannes Bernarding, Lama Naji

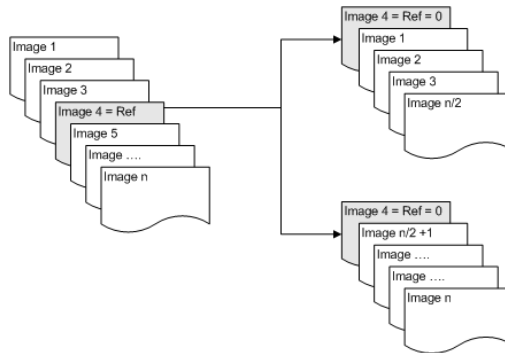
Institut für Biometrie und Medizinische Informatik
Medizinische Fakultät der Otto-von-Guericke-Universität Magdeburg
Email: luchtman@googlemail.com

Zusammenfassung. Statistical parametric mapping (SPM) ist ein umfangreiches, auf Matlab basierendes Softwarepaket zur bildgestützten Analyse in der funktionellen Hirnbildgebung. Es dient zum Nachweis von Aktivitätsänderungen in Hirnarealen bei Durchführung definierter Aufgaben und Wahrnehmung sensorischer Stimuli mittels fMRT, SPECT oder PET. Je nach Datenmodalität sind unterschiedliche, zeitintensive Vorverarbeitungsschritte erforderlich. Im vorgestellten Projekt wurde eine Infrastruktur zur Parallelverarbeitung der Daten mittels SPM entwickelt, die außer Matlab keine weitere Software benötigt. Es konnte eine signifikante Reduktion der Auswertzeit entwickelt werden. Das Konzept ist skalierbar und erlaubt somit in Rechnerclustern eine weitere Reduktion der Auswertzeit.

1 Einleitung

SPM baut auf den Modulen der Vorverarbeitung, der statistische Analyse und der visuellen Darstellung auf. Die einzelnen Bausteine stellen unterschiedliche Anforderungen: während die visuelle Darstellung technisch sehr rudimentär ist und wenig Rechenzeit benötigt, ist die Vorverarbeitung, abhängig von der Anzahl der auszuwertenden Bilddaten, der verwendeten Interpolationsverfahren und der definierten Iterationsschritte, sehr zeitintensiv. Allerdings können die Vorverarbeitungsschritte Registrierung, Normalisierung auf ein Standard-Template und Glättung auf Untermengen der Bilddaten ausgeführt und somit parallelisiert werden. Das Modul Realign dient der Elimination von Bewegungen während der Aufnahme durch Registrierung auf eine Referenzbild der Zeitserie[1]. Mit dem Modul Normalize werden die Strukturen der gemessenen Hirnbilddaten auf ein ideales „Durchschnitts“-Gehirn (MNI-Template) [2] transformiert. Dies dient zum Vergleich der Ergebnisse verschiedener Probanden und ermöglicht eine Gruppenanalyse [1]. Das Modul Smooth dient der Glättung der Original-Bilddaten, wodurch das Rauschen minimiert und die statistische Analyse unabhängiger von der genauen Lage und Ausdehnung der aktivierten Hirnareale verschiedener Probanden ist [1].

Abb. 1. Auswahl und Transfer eines Referenzdatensatzes (hier beispielhaft Bildserie 4) auf einen Cluster. Zur Vereinfachung ist die minimale Clustergröße mit zwei Rechnern gewählt. Die Referenz wird allen aufgeteilten Untermengen als erste fMRI-Aufnahme vorangestellt. Image i entspricht dem Volumendatensatz zum Zeitpunkt i



2 Stand der Forschung und Fortschritt durch den Beitrag

Nur wenige Arbeiten haben sich der genannten Thematik gewidmet und interessante Lösungen erarbeitet [3]. Aufgrund der Verwendung sehr komplexer Technologien, wie MPI und CORBA, sind diese Verfahren nicht uneingeschränkt geeignet, um sie in bestehende Cluster- oder Grid-Architekturen zu integrieren. Der vorliegende Beitrag stellt einen Ansatz vor, welcher unabhängig von der existierenden Netzwerk-Architektur eine parallele Berechnung von funktionellen Hirnbilddaten realisiert. Das Konzept verzichtet dabei bewusst auf die Verwendung einer zusätzlichen proprietären Middleware, um das gezeigte Design für alle Umgebungen offen zu gestalten.

3 Methoden

Das Gesamtkonzept beruht darauf, die Bilddaten in Untermengen aufzuteilen, die auf den einzelnen Rechnern unabhängig vorverarbeitet werden können. Da SPM unter Matlab läuft, ist hierzu nur die Installation von Matlab auf den einzelnen Rechnern erforderlich. Nach der Vorverarbeitung werden die Bilddaten zusammengeführt und statistisch ausgewertet. Da hierzu alle Daten vorliegen müssen, kann dieser Schritt nicht parallelisiert werden. Zur Aufteilung, zum bidirektionalen Datentransfer und zur Remote-Steuerung der Auswertung wurde ein Prototyp entwickelt, welcher das vorliegende Konzept realisiert.

Für die Bewegungskorrektur wird ein Volumendatensatz der Zeitserie ausgewählt, welcher für alle Untermengen als Referenz dient. Anschließend wird im Cluster über externe Matlabschnittstellen eine Batchroutine verteilt, welche die Prozesse Registrierung, Normalisierung und Glättung auf jedem der Remote-Rechner initialisiert und ausführt.

Zur Auswertung wurden verschiedene funktionelle Bilddatensätze mit folgenden Parametern verwendet: 64x64x32 Voxel bei 16 Bit Farbtiefe und bis zu 400 Bilder in einer Zeitserie. Die Daten wurden zur Vorverarbeitung von DICOM in das Analyze-Format überführt.

4 Ergebnisse

Im Rahmen des Projektes MediGRID (<http://www.medigrid.de>) wurde ein erster Prototyp für die Evaluierung des vorgestellten Konzeptes realisiert. Ziel war die Prüfung der Integrierbarkeit der parallelisierten Vorverarbeitung in bestehende Grid-Architekturen [4].

4.1 Realisierung Prototyp

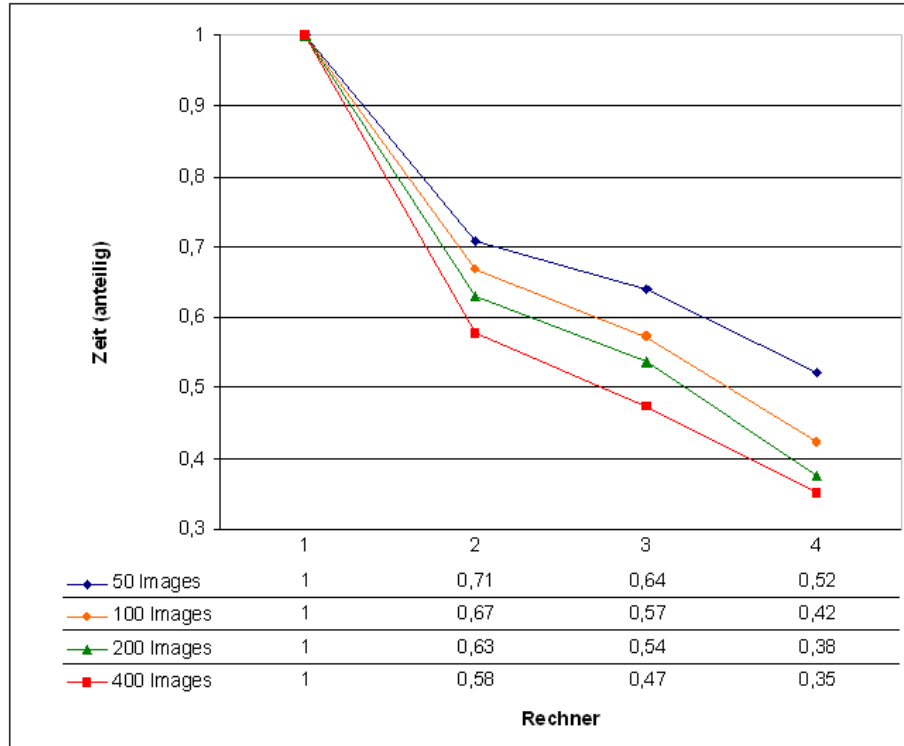
Als Grundlage für die Umsetzung eines Prototypen wurde ein einfacher Cluster mit vier Computern auf Basis von Windows XP verwendet (Intel Pentium D, 2GB RAM). Auf den Einzelrechnern wurde Matlab 7 und SPM2 installiert. Alle Knoten im Cluster wurden per Hub über ein 100 MBit Ethernet verbunden. Es wurde mit Microsoft Visual Studio eine Applikation erstellt, welche die Verteilung der Images auf die gegebenen Ressourcen koordiniert und gleichzeitig auch per DCOM den externen Kommandozeileninterpreter von Matlab zur Verfügung stellt. Das Konzept erlaubt es, dass die beschriebene Schnittstelle im Rahmen eines Unix/Linux-Cluster oder innerhalb einer Grid-Architektur auch durch andere Methoden, wie Web Services oder RPC, ersetzt werden kann.

4.2 Validierung

Abbildung 2 veranschaulicht den Geschwindigkeitsvorteil bei der Nutzung eines Clusters im Gegensatz zur Einzelplatzlösung. So benötigt die Vorverarbeitung eines Datensatzes von 400 Bilder in einem Cluster mit vier Knoten inklusive aller Kopiervorgänge nur noch knapp ein Drittel der ursprünglich auf einem Rechner verwendeten Zeit. Die Messungen umfassten dabei den vollständigen Prozess der Vorverarbeitung inklusive aller notwendigen Verteilungen der Daten. Dies beinhaltete sowohl das divergente Kopieren der Ausgangsdaten auf die verschiedenen Knoten des Clusters, als auch die konvergente Rückführung der berechneten Images.

Zu jeder Messung wurde der gleiche Satz an Images verwendet. Die Verteilung auf die einzelnen Knoten im Cluster erfolgte bei einem Messdurchgang von einem vorher definierten Knoten. Die Ergebnisse wurden durch Wiederholen bestätigt. Dabei dauerte die vollständige Vorverarbeitung für eine Messung von 400 Bildserien unter Verwendung von vier Knoten im Durchschnitt 280 Se-

Abb. 2. Prozentuale Entwicklung der Dauer der Vorverarbeitung abhängig von der Zahl der Knoten und Bildserien im Cluster



kunden¹. Innerhalb der Prozedur fielen dabei für das Transferieren² der Images 28 Sekunden an. Das reine Preprocessing benötigte dementsprechend 252 Sekunden. Der gleiche Vorverarbeitungsschritt auf einem einzelnen Rechner, ohne das Kopieren von Bildinformationen, dauerte für 400 Images durchschnittlich 799 Sekunden. Es wird deutlich, dass zeitlich fixe Prozeduren, wie z.B. die Berechnung der Transformationsmatrix für die Überführung des Referenzbildes in den Talairach-Raum, mit steigender Zahl zu berechnender Images an Bedeutung verlieren, d.h. je höher die Zahl der MRT-Daten desto größer ist auch der Geschwindigkeitsvorteil bei deren Vorverarbeitung.

¹ Die Messung wurde durch den langsamsten Rechner im Cluster(Bottleneck) terminiert. Nach 280 Sekunden, war dementsprechend der letzte Knoten im Cluster mit der Berechnung und Transferierung fertig.

² Im Rahmen der Transferierung wurden die originären Bilddaten auf den Remote-Host und die vorverarbeiteten Images wieder auf den Quellknoten kopiert.

5 Diskussion

Die Berechnungen zur Auswertung von funktionellen Hirnbilddaten nehmen trotz aller technologischer Fortschritte, abhängig von der jeweiligen Stichprobe, immer noch zu viel Zeit ein. Aufgrund der Verwendung von immer höher aufgelösten Hirnbilddaten, wird sich an diesem Zustand bei der Verwendung einer Einzelplatzlösung nicht viel ändern. Mit pSPM existiert ein auf MPI basierendes Modul, welches bereits parallele Vorverarbeitung ermöglicht. Allerdings bieten nicht alle Grid-Architekturen und -Konzepte eine Unterstützung für Applikationen auf Basis von MPI. Im Gegensatz dazu ist unser Konzept aufgrund seiner offenen Struktur in herkömmliche Grid-Architekturen integrierbar. Darüberhinaus publizierte keine der genannten Arbeiten verwertbare Benchmarkergebnisse [3]. Unsere Resultate zeigen, dass das Verfahren realisierbar ist und es bei gleichbleibender Qualität einen signifikanten Geschwindigkeitsvorteil mitbringt. Der Ansatz ist einfach skalierbar und plattformunabhängig. Bei Verwendung größerer Cluster ist eine weitere signifikante Reduktion der Rechenzeit zu erwarten.

Literaturverzeichnis

1. Huettel SA, Song AW, McCarthy G. Functional Magnetic Resonance Imaging. 1st ed. Sinauer Associates, Inc; 2004.
2. Talairach J, Tournoux P. Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging. 1st ed. Thieme Medical Publishers; 1988.
3. May M, Munz F, Ludwig T. CORBA-basierte verteilte Berechnung medizinischer Bilddaten mit 1118-9. Procs BVM 2000; 213-217.
4. Foster I. The grid: Computing without bounds. Scientific American 2003;288:78-85.