

Fachhochschule Aachen
Campus Jülich

Fachbereich: Medizintechnik und Technomathematik
Studiengang: Scientific Programming

Berechnung und Anwendung von Kalibrierungsdaten für 3D-Polarized-Light-Imaging Messungen

Bachelorarbeit von

Kim Sontheimer

Jülich, 22. August 2017

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die Bachelorarbeit mit dem Thema **Berechnung und Anwendung von Kalibrierungsdaten für 3D-Polarized-Light-Imaging Messungen** selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, alle Ausführungen, die anderen Schriften wörtlich oder sinngemäß entnommen wurden, kenntlich gemacht sind und die Arbeit in gleicher oder ähnlicher Fassung noch nicht Bestandteil einer Studien- oder Prüfungsleistung war. Ich verpflichte mich, ein Exemplar der Bachelorarbeit fünf Jahre aufzubewahren und auf Verlangen dem Prüfungsamt des Fachbereiches Medizintechnik und Technomathematik auszuhändigen.

Ort und Datum

Unterschrift

Die Arbeit wurde betreut von:

Erstprüfer: Prof. Dr. Johannes Grotendorst

Zweitprüfer: Dr. Markus Axer

Diese Arbeit wurde erstellt am
Jülich Supercomputing Centre



in Kooperation mit dem
Institut für Neurowissenschaften und Medizin (INM-1)

Zusammenfassung

Das Verfahren des *Three-dimensional Polarized Light Imaging* (3D-PLI) wurde im Institut für Neurowissenschaften und Medizin am Forschungszentrum Jülich entwickelt. Damit ist es möglich Nervenfasern im Gehirn im Submillimeter-Bereich aufzunehmen und deren Verlauf zu rekonstruieren. Dazu werden Gehirne post mortem in 60µm dünne Scheiben geschnitten und mit einem speziell hierfür entwickelten rotierenden Polarimeter aufgenommen. Aus der transmittierten Lichtintensität kann schließlich der räumliche Verlauf der Nervenfasern rekonstruiert werden, um so einen 3D-Hirnatlas in mikroskopischer Auflösung zu erstellen. Die Qualität des Signals bei der Messung wird jedoch durch Faktoren wie Bildrauschen oder Unregelmäßigkeiten in der Aufnahmetechnik beeinflusst. Diese Einflüsse haben eine direkte Auswirkung auf die Bestimmung der Faserorientierung. Eine exakte Kalibrierung der Rohdaten nach der Aufnahme ist deshalb wichtig für die weitere fehlerfreie Verarbeitung im 3D-PLI Rekonstruktionsworkflow.

Im Rahmen dieser Arbeit wird deshalb ein Verfahren zur Kalibrierung entwickelt, das die bisher verwendete Berechnung und Anwendung von Kalibrierungsdaten optimiert. Die Berechnung der Kalibrierungsdaten wird hinsichtlich Stabilität und Möglichkeiten zur regelmäßigen Aktualisierung untersucht. Es werden Methoden entwickelt, die auftretenden Unregelmäßigkeiten in der Bildgebung zu erkennen und auszugleichen und die Kalibrierungsdaten über die Zeit aktuell zu halten. Mit dem neu implementierten Verfahren ist eine stabile und aktuelle Berechnung von Kalibrierungsdaten, bei gleichzeitig möglichst geringem Aufwand in der Neuberechnung, möglich. Eine effiziente Anwendung der Kalibrierungsdaten auf die Rohdaten des 3D-PLI wird über eine Parallelisierung der entwickelten Software für den GPU-Cluster JURECA und mit der Speicherung der kalibrierten Daten im HDF5-Dateiformat erreicht.

Inhaltsverzeichnis

1. Einleitung	1
2. 3D-Polarized Light Imaging (3D-PLI)	3
2.1. Bildgebung	4
2.2. Rekonstruktionsworkflow	5
2.3. Anforderungen an die Qualität der Aufnahmen	7
3. Berechnung der Kalibrierungsdaten	9
3.1. Gainfaktorberechnung	9
3.2. Stabilität von Gainfaktoren	10
3.2.1. Allgemeine Herleitung	10
3.2.2. Auswertung für 3D-PLI Datensätze	13
3.3. Aktualisierung von Gainfaktoren	19
3.3.1. Einzelmessung	20
3.3.2. Sliding Window Prinzip	21
3.3.3. Ausreißer Analyse	22
3.4. Auswertung und Zusammenfassung	25
3.5. Optimierungen	27
4. Anwendung der Kalibrierungsdaten	29
4.1. Speicherung der Daten	29
4.1.1. HDF5 Datenformat	29
4.1.2. Anwendung auf 3D-PLI Daten	32
4.2. Parallelisierung	35
4.2.1. Parallelisierungsstrategie	35
4.2.2. Laufzeitverhalten	37
5. Diskussion und Ausblick	43
Literatur	45
A. Konfiguration JURECA	I

Abbildungsverzeichnis

2.1. Aufbau einer Nervenzelle	3
2.2. Mikroskopie-Schema	4
2.3. Intensitätsprofil einer 3D-PLI Messung	5
2.4. Stitching	6
2.5. Orientierung der Nervenfasern im Raum	6
2.6. Fiber Orientation Map (FOM)	7
3.1. Struktogramm: Berechnung der Gainfaktoren	10
3.2. Leermessung und Gainfaktormatrix	11
3.3. Heatmap für die Anzahl benötigter Leermessungen	13
3.4. Differenz mittlere Intensitäten	15
3.5. Histogramme verschiedener Leermessungen	16
3.6. Gainfaktorvergleich bei gleichem Referenzwert	18
3.7. Gainfaktorvergleich bei unterschiedlichem Referenzwert	19
3.8. Sliding Window Prinzip	21
3.9. Vergleich verschiedener Leerbilder	23
3.10. Differenz verschiedener Leerbilder	23
3.11. Staubpartikel in einer Leermessung	24
3.12. Kalibrierungsworkflow	25
3.13. Vergleich der Ergebnisse nach Kalibrierung	26
4.1. HDF5-Gruppe	30
4.2. HDF5-Datensatz	30
4.3. HDF5: Subsets	31
4.4. HDF5: Chunking	31
4.5. HDF5: interner Link	32
4.6. HDF5: externer Link	32
4.7. PLI-HDF5-Struktur	33
4.8. SIMD Architektur	36
4.9. Gebietszerlegung Kalibrierung	37
4.10. Diagramm: Laufzeitvergleich	38
4.11. Laufzeit und Speedup gesamt	39
4.12. Speedup im Detail	40
4.13. Speedup auf mehreren Knoten	41

Tabellenverzeichnis

3.1. Berechnung Stichprobenumfang	14
3.2. Modalwerte verschiedener Leermessungen	17
3.3. Vergleich mehrerer Gainfaktoren	20
4.1. HDF5 Chunks bei 3D-PLI Daten	34
4.2. Parallelisierungsparadigmen	35

1. Einleitung

Die Funktionsweise des menschlichen Gehirns vollständig zu verstehen, ist sowohl für die neurowissenschaftliche als auch für die medizinische Forschung von großem Interesse. Weltweit gibt es viele Projekte, die sich mit dem Aufbau und der Struktur des Gehirns sowie den funktionellen Zusammenhängen im Gehirn beschäftigen. Dazu gehören beispielsweise das von der Europäischen Union geförderte *Human Brain Project* [1] oder die von den *National Institutes of Health* geförderte *BRAIN Initiative* in den USA [2].

Im Forschungszentrum Jülich beschäftigen sich unter anderem das Institut für Neurowissenschaften und Medizin (INM) und das Jülich Supercomputing Centre (JSC) gemeinsam mit dieser Thematik. Die Forschungsgruppe „Faserbahnarchitektur“ im INM entwickelt eine Technik, die es ermöglicht den dreidimensionalen Verlauf von Nervenfasern in Gehirnen in mikroskopischer Auflösung zu rekonstruieren. Als bildgebendes Verfahren wird das dort entwickelte „Three-dimensional Polarized Light Imaging“ (3D-PLI) mit dem Ziel eingesetzt, den Verlauf von Faserbündeln bis hin zu einzelnen Nervenfasern im Gehirn zu erfassen und darzustellen. Damit können wichtige Erkenntnisse über den Aufbau des Gehirns und über Verbindungen zwischen den Hirnregionen gewonnen werden. In Kapitel 2 werden die Grundlagen des 3D-PLI, der verwendeten Mikroskopie und des Rekonstruktionsworkflows genauer beschrieben.

Bei der Aufnahme von Gehirnen mit hoher Auflösung können bereits kleine Unregelmäßigkeiten in der Bildgebung das Ergebnis stark beeinflussen. Eine exakte Kalibrierung der 3D-PLI Rohbilder nach der Mikroskopie ist deshalb wichtig für eine korrekte Weiterverarbeitung im Rekonstruktionsworkflow. Daher wurde im Rahmen dieser Arbeit eine Software entwickelt um die Kalibrierung von 3D-PLI Messungen zu optimieren. Das Ziel ist sowohl die Berechnung von exakten und aktuellen Kalibrierungsdaten mit möglichst geringem Aufwand als auch eine effiziente Anwendung dieser.

In Kapitel 3 wird auf die Berechnung von Kalibrierungsdaten für das 3D-PLI eingegangen. Dabei werden die Stabilität und die Aktualisierung dieser Daten diskutiert. Dies beinhaltet die Bestimmung der Anzahl an Messungen, welche zur Kalibrierung benötigt werden, und das rechtzeitige Erkennen von Veränderungen in diesen. Mit den beschriebenen Methoden wird schließlich ein Ablauf zur Berechnung der Kalibrierungsdaten vorgestellt.

Durch die sehr große Datenmenge beim 3D-PLI ist die Verarbeitung im Rekonstruktionsworkflow sehr speicher- und laufzeitintensiv. Neben der Berechnung von Kalibrierungsdaten ist deshalb auch deren effiziente Anwendung von Bedeutung. Die Anwendung auf die Rohbilder wird in Kapitel 4 beschrieben. Es wird genauer auf die effiziente Speicherung der Rohdaten und Kalibrierungsdaten mit Hilfe des HDF5-Datenformats und auf die Laufzeitverbesserung mittels Parallelisierung eingegangen.

In Kapitel 5 werden die Ergebnisse dieser Arbeit zusammengefasst und ein Ausblick auf mögliche Folgeprojekte gegeben.

2. 3D-Polarized Light Imaging (3D-PLI)

Für eine vollständige Rekonstruktion der Anatomie des menschlichen Gehirns im Submillimeter-Bereich, auf der Ebene einzelner Nervenfasern (Axone) und ihrer Verbindungen, sind Informationen über die räumliche Ausrichtung der Nervenfasern notwendig. Die für die Rekonstruktion verwendeten Bilder müssen daher in ausreichend hoher Auflösung vorliegen, die jedoch stark durch die zur Verfügung stehenden bildgebenden Verfahren (*engl.: Neuroimaging*) limitiert ist. Ein solches Verfahren ist das 3D-PLI, bei dem die doppelbrechende Eigenschaft des Nervengewebes im Gehirn ausgenutzt wird.

Die meisten Axone im Gehirn von Wirbeltieren sind von einer Schicht aus Myelin, der sogenannten Myelinscheide, umgeben (Abb. 2.1). Diese dient unter anderem der elektrischen Isolation und spielt bei der Reizweiterleitung eine wichtige Rolle, ist jedoch auch der Grund für die doppelbrechende Eigenschaft des Nervengewebes. Dies bedeutet, dass die Myelinscheide den Polarisationszustand des Lichts verändert.

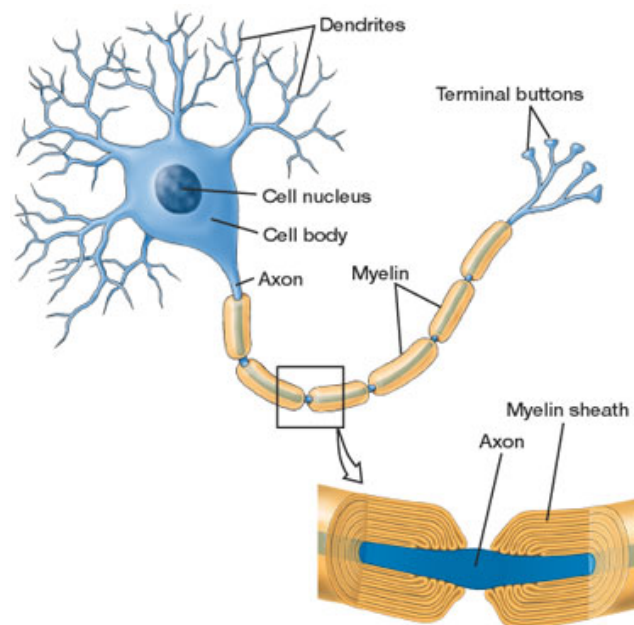


Abbildung 2.1.: Aufbau einer Nervenzelle [3]. Für das 3D-PLI von Bedeutung sind das Axon und die Myelinscheide (*engl.: Myelin sheath*), welche das Axon umgibt.

Diese Änderung korreliert mit der räumlichen Ausrichtung der Myelinscheide beziehungsweise mit der durch die Myelinscheide erzeugten optischen Achse [4]. Damit lässt sich aus der Richtung der optischen Achse die Richtung der Nervenfasern bestimmen. Dieser Zusammenhang ermöglicht es, mit Hilfe des 3D-PLI die Orientierung der Axone im Submillimeter-Bereich zu rekonstruieren [5].

Im Folgenden wird genauer auf die Bildgebung und die Verarbeitungsschritte (Rekonstruktionsworkflow) für das 3D-PLI eingegangen.

2.1. Bildgebung

Beim 3D-PLI werden Gehirne post mortem in $60\mu\text{m}$ dünne Scheiben geschnitten und danach mit einem speziell hierfür entwickelten Polarimeter aufgenommen. Abbildung 2.2 zeigt den schematischen Aufbau des Polarimeters.

Als Lichtquelle dient eine LED-Lampe, welche unpolarisiertes Licht ausstrahlt, das noch in alle Richtungen schwingt. Bei der Messung passiert das Licht zunächst einen linearen Polarisationsfilter, sodass der elektrische Feldvektor des Lichts nur noch in einer bestimmten Richtung schwingt. Über die Verzögerungsplatte wird dieses in zirkulär polarisiertes Licht transformiert. Durch die doppelbrechende Eigenschaft des Myelins in der Gewebeprobe wird das Licht beim Passieren dieser elliptisch polarisiert, mit unterschiedlicher Phasenverschiebung je nach Faserverlauf im Raum. Mit einem zweiten Polarisationsfilter, dessen Polarisationsrichtung senkrecht auf der des ersten steht, kann man schließlich die Richtung der optischen Achse an jedem Aufnahmepunkt des Gewebes ermitteln und damit die Richtung der Nervenfasern in diesem Bereich.

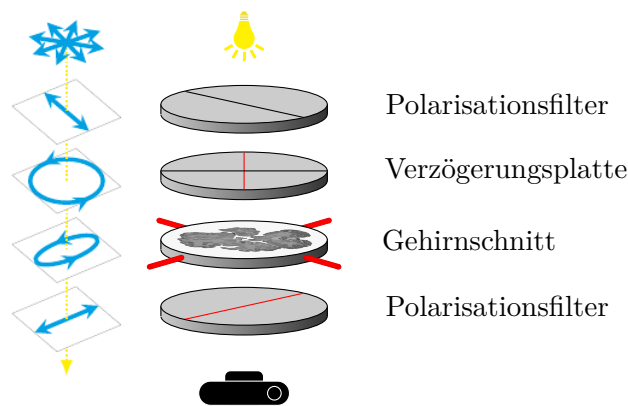


Abbildung 2.2.: Schematischer Aufbau des Polarimeters [6]. Die Filter und die Verzögerungsplatte rotieren gleichzeitig um 180° in 10° Schritten, der Gehirnschnitt ist fixiert. Bei jeder Rotation wird jeweils ein Bild aufgenommen.

Bei einer Messung rotieren Polarisationsfilter und Verzögerungsplatte gleichzeitig um 180° in 10° Schritten. Das Gehirngewebe bleibt dabei fixiert. Für jeden Rotationswinkel wird ein Bild des Gehirngewebes aufgenommen, also eine Serie von 18 Bildern. Dabei ändert sich, je nach Rotationswinkel, die aufgenommene Lichtintensität. Für jedes Bildpixel ergibt sich dann ein sinusförmiger Intensitätsverlauf (siehe Abb. 2.3), abhängig von den Eigenschaften des Gewebes an der jeweiligen Stelle. Die Aufnahme der Bilder erfolgt schließlich mit einer CCD-Kamera (*engl.: charge-coupled device*).

2.2. Rekonstruktionsworkflow

Um aus den Aufnahmen den Verlauf der Faserbahnen zu rekonstruieren, werden verschiedene Methoden der Signal- und Bildverarbeitung in mehreren Arbeitsschritten eingesetzt.

Die Bestimmung der Faserorientierung hängt stark von der Qualität der Rohbilder einer Messung ab. Durch eine **Kalibrierung** werden in einem ersten Schritt Verfälschungen der Rohdaten durch Effekte wie beispielsweise digitales Rauschen oder ungleichmäßige Beleuchtung korrigiert. Andere störende Objekte in den Rohdaten, wie beispielsweise Staubpartikel, können mit der **Independent Component Analysis** (ICA) entfernt werden.

Aus dem Intensitätsprofil wird durch die **Efficient PLI Analysis (EPA)** je Pixel ein Sinusverlauf der Lichtintensität aus den 18 Aufnahmen approximiert und drei Werte daraus gewonnen (Abb. 2.3): Die **Direktion** (Orientierung in der Schnittebene), die **Transmittanz** (Lichtdurchlässigkeit) und die **Retardierung** (Stärke der Myelinisierung).

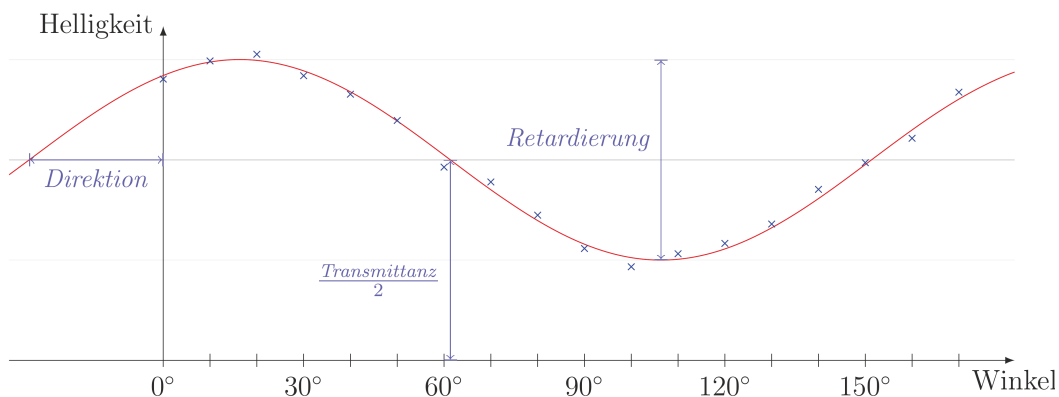


Abbildung 2.3.: Intensitätsprofil einer 3D-PLI Messung. Daraus werden die Werte Direktion, Transmittanz und Retardierung gewonnen.

Mit Hilfe der Transmittanzbilder wird bei der **Segmentierung** die räumliche Abgrenzung von Hintergrund und Gehirngewebe ermittelt. Der verwendete Region-Growing-Algorithmus erstellt eine Maske, welche die Bildpixel entweder dem Hintergrund oder dem Gehirn zuordnet [7]. Dies verringert die Menge der weiter zu verarbeitenden Daten, da der Hintergrund für die Analyse keine Rolle spielt.

Um die benötigte Genauigkeit im Submillimeter-Bereich zu erreichen, werden die Schnitte zusätzlich mit einem Mikroskop aufgenommen [8]. Mit diesem **Polarizing Microscope (PM)** wird eine Auflösung von $1,33 \times 1,33 \mu\text{m}$ pro Pixel erreicht. Die Gehirnschnitte können dadurch jedoch nicht im Gesamten aufgenommen werden und werden deshalb in kleine Bereiche (Kacheln) der Größe 2048×2048 Pixel unterteilt, die nacheinander mit dem PM aufgenommen werden.

Bei der Aufnahme in Kacheln wird außerdem mit etwa 30% Überlappung auf-

genommen. Damit wird sichergestellt, dass keine Informationen an den Rändern dieser Kacheln verloren gehen. Beim *Stitching* werden die einzelnen Kacheln eines Schnittes dann zu einem Gesamtbild zusammengesetzt (Abb. 2.4). Dazu werden in den Randbereichen benachbarter Kacheln markante Stellen markiert und die Bilder so aufeinander gelegt, dass diese Stellen übereinander liegen und ein konsistentes Bild entsteht.

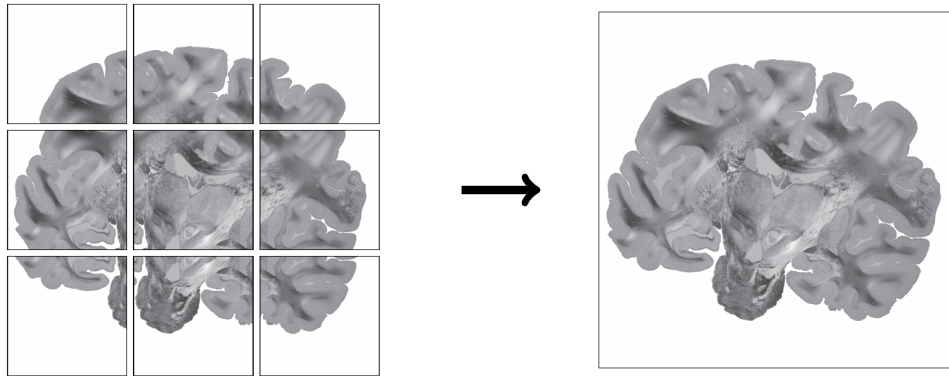


Abbildung 2.4.: Beispiel für das Stitching von neun Kacheln [9]. Die überlappenden Randbereiche jeder Kachel werden aufeinander geschoben und ein konsistentes Bild erzeugt.

Nach dem Stitching wird schließlich die Richtung der Fasern im Raum je Schnitt berechnet. Die Richtung setzt sich aus der Direktion und *Inklination* zusammen. Die Direktion φ beschreibt dabei den Winkel der Projektion der Faser in der Schnittebene, die Inklination α den Winkel zwischen Schnittebene und Faser (Abb. 2.5). Der Winkel α wird mit Hilfe der Retardierungswerte berechnet.

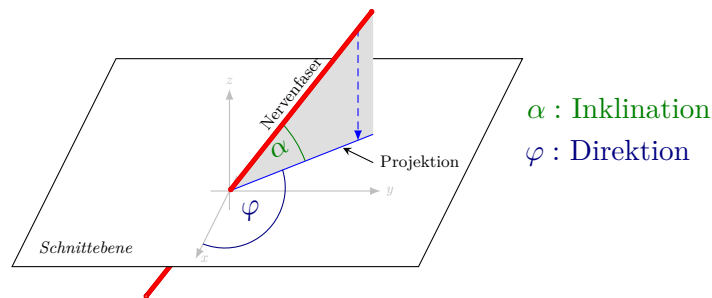


Abbildung 2.5.: Die Orientierung der Nervenfasers im Raum kann durch die Inklination (Winkel zwischen Schnittebene und Faser) und Direktion (Winkel in der Schnittebene) beschrieben werden [6].

Die Darstellung eines Punktes im dreidimensionalen Raum kann durch seine sphärischen Koordinaten erfolgen, also durch den Abstand vom Ursprung und zwei Winkel. Da die Länge der Faser hier keine Rolle spielt, kann für jedes Pixel mit α und φ also ein Vektor für die Richtung der jeweiligen Faser im Raum bestimmt werden. Mit diesen Richtungsinformationen können dann die Pixel anhand der Richtung,

beispielsweise im HSV-Farbraum, eingefärbt werden. Dadurch ergibt sich die **Fiber Orientation Map (FOM)** (Abb. 2.6).

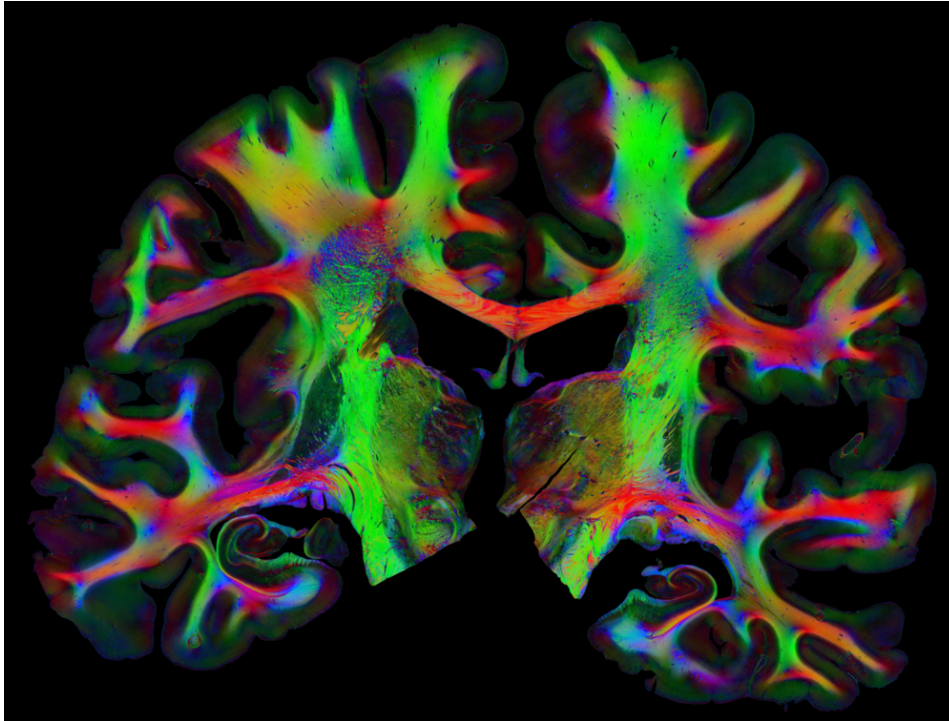


Abbildung 2.6.: Die FOM eines Schnittes des menschlichen Gehirns im HSV-Farbraum [6]. Die Farben stehen repräsentativ für die Faserorientierung.

Die einzelnen Schnitte mit den zugehörigen Richtungsinformationen müssen abschließend wieder zu einem Gesamtvolumen zusammengesetzt werden, um den Verlauf von Nervenfasern über die Schnitte hinaus extrahieren zu können. Dieser Vorgang wird **Registrierung** genannt. Dabei müssen mehrere Faktoren berücksichtigt werden: Beispielsweise sind Lage und Orientierung der Schnitte auf dem Objektträger nicht immer gleich, es muss also eine entsprechende Ausrichtung vorgenommen werden. Auch wird durch den Schneideprozess das Gewebe des Gehirns nicht linear verformt. Eine Rekonstruktion durch einfaches Aufeinanderlegen der Schnitte ist somit nicht möglich. Mit Hilfe von sogenannten Blockface-Bildern, einer Aufnahme des Gehirns *vor* dem jeweiligen Schnitt, kann über verschiedene Transformationen die ursprüngliche Form des Schnittes wiederhergestellt werden.

2.3. Anforderungen an die Qualität der Aufnahmen

Das sinusförmige Intensitätsprofil eines jeden Bildpixels ist entscheidend für die korrekte Bestimmung der Faserorientierung. Aus diesem Grund ist es wichtig, eine gleichmäßige Ausleuchtung des Objektträgers zu haben, die Lichtintensität ungestört aufnehmen zu können und die verbleibenden Effekte durch eine möglichst ex-

akte Kalibrierung auszugleichen. Die Qualität des Signals im optischen System der 3D-PLI-Bildgebung wird im Allgemeinen, wie bei vergleichbaren Verfahren, durch folgende Faktoren beeinflusst:

Statische Effekte

Zu den statischen Effekten gehören immanente Unregelmäßigkeiten in der Aufnahmetechnik. Helligkeitsunterschiede durch ungleichmäßige Beleuchtung (Vignettierung), Eigenschaften der CCD-Kamera (Thermik, Empfindlichkeit der Pixel) aber auch eine (geringe) Inhomogenität in den Polarisationsfiltern sind einige Beispiele, die zu Unterschieden in der Lichtintensität der Pixel eines Bildes führen.

Spontane Effekte

Zu den spontanen Effekten gehören beispielsweise Fremdkörper in der Aufnahmetechnik, welche nicht dauerhaft vorliegen, sondern spontan auftreten und wieder verschwinden können. Verunreinigungen wie Staubpartikel auf dem Sensor können zum Beispiel zwischen zwei Messungen hinzukommen, verschwinden oder „wandern“. Daher sollten diese Effekte rechtzeitig erkannt und ausgeglichen werden. Damit ist neben der Genauigkeit auch die regelmäßige Aktualisierung von der für die Kalibrierung verwendeten Daten wichtig.

Kontinuierliche Effekte

Längerfristige Veränderungen in der Aufnahmetechnik müssen ebenfalls durch eine regelmäßige Aktualisierung der Kalibrierungsdaten ausgeglichen werden. Beispielsweise muss, da beim 3D-PLI LED-Lampen als Lichtquelle eingesetzt werden, deren Lebensdauer berücksichtigt werden. Aufgrund des langen Zeitraums bis zum Ausfallzeitpunkt wird bei LED-Lampen im Gegensatz zu anderen Leuchtmitteln die Lebensdauer als Zeitraum in Stunden definiert, bis die ursprüngliche Lichtleistung auf einen festgelegten Schwellenwert abgesunken ist [10]. In Bezug auf das 3D-PLI bedeutet dies, dass über die Zeit hinweg die gemessene Lichtintensität abnimmt und größere Zeiträume zwischen zwei Kalibrierungen zu Ungenauigkeiten führen.

Als Kalibrierung wird im Folgenden nach der Definition des JCGM [11] ein zweistufiger Prozess bezeichnet: Im ersten Schritt werden die Korrekturfaktoren (**Gainfaktoren**) für das optische System ermittelt. Der zweite Schritt beinhaltet die *Anwendung* dieser Kalibrierungsdaten auf die Rohdaten einer Messung. Die Berechnung der Kalibrierungsdaten wird in Kapitel 3 behandelt, deren Anwendung in Kapitel 4.

Das Ziel dieser Arbeit ist es, neue Methoden zur Berechnung von exakten und aktuellen Gainfaktoren zu entwickeln und deren Anwendung auf die Rohdaten zu optimieren.

3. Berechnung der Kalibrierungsdaten

Für die Berechnung der Kalibrierungsdaten, den Gainfaktoren, werden Leermessungen (*Flatfields*, auch bekannt als Weißbilder) aufgenommen. Eine Leermessung ist eine Aufnahme ohne Objekt im Versuchsaufbau. Dadurch sollen auftretende Unregelmäßigkeiten in der Aufnahmetechnik erfasst werden, um die in Kapitel 2.3 beschriebenen statischen und spontanen Effekte auszugleichen.

Die Leermessung sollte unter gleichen Bedingungen wie die späteren Aufnahmen stattfinden, da beispielsweise Faktoren wie die räumliche Ausrichtung und die Belichtungszeit des Kamerasensors Einfluss auf die Abbildungsfehler haben. Für die Aufnahmen des 3D-PLI bedeutet dies, dass für unterschiedliche Belichtungszeiten und alle 18 Rotationswinkel ρ jeweils eigene Gainfaktoren berechnet werden müssen. Um die erwähnten kontinuierlichen Effekte auszugleichen, ist zusätzlich eine regelmäßige Wiederholung der Gainfaktorberechnung nötig.

In diesem Kapitel werden deshalb die Stabilität sowie die Aktualisierung der Gainfaktoren untersucht und in Bezug auf das 3D-PLI diskutiert.

3.1. Gainfaktorberechnung

Die Gainfaktoren für die Bildsensorkalibrierung werden über das Verhältnis der gemessenen Intensitätswerte der Pixel in einer Leermessung zu einem Referenzwert berechnet (3.1). Da die Aufnahme einer einzelnen Leermessung Bildrauschen enthält, sollten mehrere Leermessungen gemacht werden, um den Erwartungswert der Intensität eines jeden Bildpixels korrekt ermitteln zu können. Als Referenzwert wird der Modalwert über alle Leermessungen gewählt, also die häufigste Intensität über alle Pixel. Dies minimiert die Anzahl zu verändernder Pixel bei der Anwendung der Gainfaktoren.

Mit 18 Rotationswinkeln ρ beim 3D-PLI entspricht ein Set aus n Leermessungen somit $n \cdot 18$ Leerbildern und die Gainfaktoren ergeben sich damit über

$$g_n(\rho_i) = \frac{I_{ref}}{\bar{I}_n(\rho_i)}, \quad i \in [1; 18] \quad (3.1)$$

mit der Bildmatrix $\bar{I}_n(\rho_i)$ als Mittelwerte (arithmetisches Mittel) über n Leermessungen bei Rotationswinkel ρ_i und dem Wert I_{ref} als Modalwert aller Leermessungen.

Bei kleinen Abweichungen des Mittelwertes eines Pixels vom Modalwert ist der Gainfaktor somit nahe an 1.0 und der Wert dieses Pixels wird bei Anwendung der Gainfaktoren nur gering verändert.

Für die Bestimmung der Gainfaktoren aus n Leermessungen der Bild-Dimension $x \times y$, wird im Algorithmus (Abb. 3.1) zunächst die Berechnung der Mittelwerte pixelweise je Rotationswinkel durchgeführt. Unter Verwendung des Medians der Modalwerte aller 18 Mittelwert-Matrizen werden schließlich die Gainfaktoren berechnet und in einer $18 \times x \times y$ Matrix abgespeichert.

$modes \leftarrow$ Feld zum Speichern der Modalwerte	
$means \leftarrow$ Feld zum Speichern der Mittelwert Matrizen	
Für alle 18 Winkel ρ	
	Alle n Leermessungen für Winkel ρ einlesen
	Bilde Histogramm über die Leermessungen
$max \leftarrow$ Modalwert des Histogramms	
$I_range \leftarrow$ alle Peaks $\geq 99\%$ von max	
$mode \leftarrow \frac{Minimum(I_range) + Maximum(I_range)}{2}$	
$I_flat \leftarrow$ pixelweiser Mittelwert der Leermessungen	
$mode$ in $modes$ abspeichern	
I_flat in $means$ abspeichern	
$I_ref \leftarrow$ Median von $modes$	
$gain \leftarrow \frac{I_ref}{means}$	

Abbildung 3.1.: Struktogramm: Berechnung der Gainfaktoren. Für jeden Winkel wird zunächst der Modalwert der n Bilder dieses Winkels bestimmt. Der Median aller 18 Modalwerte wird als I_{ref} gewählt. Die Gainfaktor Matrix ergibt sich schließlich nach 3.1.

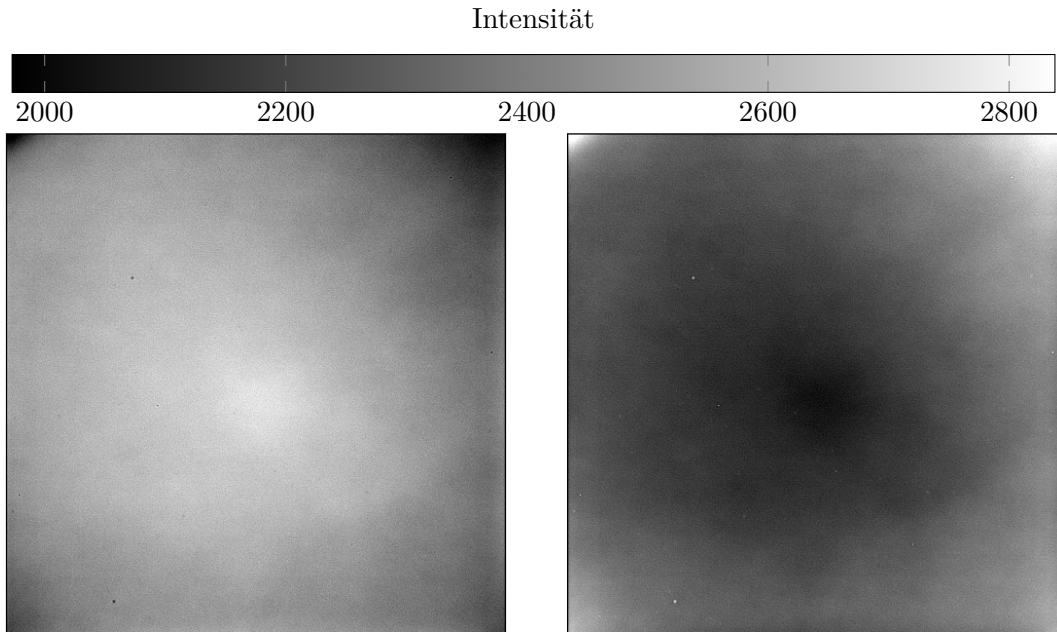
Das Ergebnis einer Auswertung für $n = 100$ Leermessungen (1800 Leerbilder) ist in Abbildung 3.2 exemplarisch für den Winkel $\rho_i = 0^\circ$ gezeigt. Sowohl im Bild der mittleren Intensitätswerte aus den 100 Leermessungen als auch in den daraus ermittelten Gainfaktoren sind die auftretenden Unregelmäßigkeiten bei erhöhtem Bildkontrast zu erkennen.

3.2. Stabilität von Gainfaktoren

Um die spontanen und kontinuierlichen Effekte (siehe Kapitel 2.3) in den Aufnahmen zu erfassen und zu korrigieren, sollte die Berechnung der Gainfaktoren regelmäßig wiederholt werden. Änderungen wie die Lage und Anzahl von Staubpartikeln oder eine über die Zeit schwächer werdende Lichtquelle können damit einbezogen und die Gainfaktoren aktuell gehalten werden. Jedoch ist eine regelmäßige Messung je nach Anzahl benötigter Leermessungen zeitaufwendig und speicherintensiv. Aus diesem Grund bietet sich eine sinnvolle Beschränkung der Anzahl verwendeter Leermessungen an.

3.2.1. Allgemeine Herleitung

Mit Hilfe von statistischen Methoden kann berechnet werden, welche Anzahl an Leermessungen mindestens benötigt wird, um hinreichend stabile Gainfaktoren zu erhalten. Für die Berechnung der Gainfaktoren wird für jedes Bildpixel je Rotationswinkel das arithmetische Mittel aus den Leermessungen als Abschätzung für den Erwartungswert genommen (siehe Abb. 3.1). Dafür wird für n Leermessungen



(a) Die mittleren Intensitätswerte aus 100 Leermessungen, $\bar{I}_{100}(0^\circ)$ (b) Die daraus berechneten Gainfaktoren $g_{100}(0^\circ)$

Abbildung 3.2.: Die Leermessungen wurden mit dem PM des 3D-PLI aufgenommen. Abb. (a) zeigt die Unregelmäßigkeiten in den Leermessungen. Die Gainfaktoren in Abb. (b) gleichen diese auf den Referenzwert I_{ref} aus. Zur Verdeutlichung der Unregelmäßigkeiten (Vignettierung und Staubpartikel) ist der Kontrast in beiden Bildern erhöht worden.

jedes Bildpixel als unabhängige und identisch verteilte Zufallsvariable X_i , $i = 1, \dots, n$ mit unbekanntem Erwartungswert μ und unbekannter Standardabweichung σ angesehen:

$$X_i \sim (\mu, \sigma), \quad \mu \in \mathbb{R}, \quad \sigma \in \mathbb{R}^+, \quad i = 1, \dots, n.$$

Das Stichprobenmittel eines Bildpixels $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ über n Messungen ist somit ein erwartungstreu und konsistenter Punktschätzer für den Erwartungswert μ der Zufallsvariable. Da bei unbekannter Varianz σ^2 die Stichprobenvarianz σ_n^2 kein erwartungstreu Schätzer der Verteilung ist, muss die korrigierte Stichprobenvarianz $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ verwendet werden. Die Abweichung des Mittelwerts vom tatsächlichen Erwartungswert μ folgt somit der *Studentschen t-Verteilung* mit $n - 1$ Freiheitsgraden

$$T = \frac{\bar{X}_n - \mu}{s_n / \sqrt{n}}$$

Für große n nähert sich die t-Verteilung der Normalverteilung an und unter Verwendung des Zentralen Grenzwertsatzes nach Lindeberg–Lévy ergibt sich schließlich:

$$\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} = Z_n \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1) \quad (3.2)$$

Unter Vorgabe eines Fehlers, also der gewünschten Genauigkeit der Abschätzung, und einer Sicherheitswahrscheinlichkeit, lässt sich nun die benötigte Anzahl an Messungen bestimmen. Für einen absoluten Fehler e und eine Fehlerwahrscheinlichkeit α bzw. Sicherheitswahrscheinlichkeit $1 - \alpha$, lässt sich μ durch das arithmetische Mittel mit

$$P(\mu - e \leq \bar{X}_n \leq \mu + e) = 1 - \alpha \quad (3.3)$$

abschätzen und nach Gleichung 3.2 gilt:

$$\begin{aligned} & P\left(\frac{(\mu - e) - \mu}{s_n/\sqrt{n}} \leq Z_n \leq \frac{(\mu + e) - \mu}{s_n/\sqrt{n}}\right) = 1 - \alpha \\ \Leftrightarrow & P\left(Z_n \geq \frac{-e \cdot \sqrt{n}}{s_n}\right) - P\left(Z_n \geq \frac{e \cdot \sqrt{n}}{s_n}\right) = 1 - \alpha \\ \Leftrightarrow & P\left(Z_n \leq \frac{e \cdot \sqrt{n}}{s_n}\right) - (1 - P\left(Z_n \leq \frac{e \cdot \sqrt{n}}{s_n}\right)) = 1 - \alpha \\ \Leftrightarrow & 2P\left(Z_n \leq \frac{e \cdot \sqrt{n}}{s_n}\right) - 1 = 1 - \alpha \\ \Leftrightarrow & P\left(Z_n \leq \frac{e \cdot \sqrt{n}}{s_n}\right) = \frac{2 - \alpha}{2} \\ \Leftrightarrow & P\left(Z_n \leq \frac{e \cdot \sqrt{n}}{s_n}\right) = 1 - \frac{\alpha}{2} \\ \Leftrightarrow & \frac{e \cdot \sqrt{n}}{s_n} = z_{1-\frac{\alpha}{2}} \\ \Leftrightarrow & n = \left(z_{1-\frac{\alpha}{2}} \cdot \frac{s_n}{e}\right)^2 \end{aligned}$$

wobei z_q das q -Quantil der Standardnormalverteilung ist.

Zu beachten ist dabei die Abhängigkeit vom Schätzwert s_n , also die Kenntnis über den zu bestimmenden Stichprobenumfang n . Daher muss das Ergebnis \hat{s}_N aus einer vorherigen Messung mit möglichst großem Stichprobenumfang N genutzt werden, um die wahre Standardabweichung σ möglichst genau abzuschätzen.

Damit ist

$$n \geq \left(z_{1-\frac{\alpha}{2}} \cdot \frac{\hat{s}_N}{e}\right)^2 \quad (3.4)$$

eine Abschätzung des Stichprobenumfangs n , sodass \bar{X}_n eine Realisation der Zufallsvariable X_i im zentralen Schwankungsintervall $[\mu - e; \mu + e]$ mit Sicherheitswahrscheinlichkeit $1 - \alpha$ annimmt.

Für die spätere Berechnung der Gainfaktoren kann das über 3.4 bestimmte n als Anzahl benötigter Leermessungen verwendet werden.

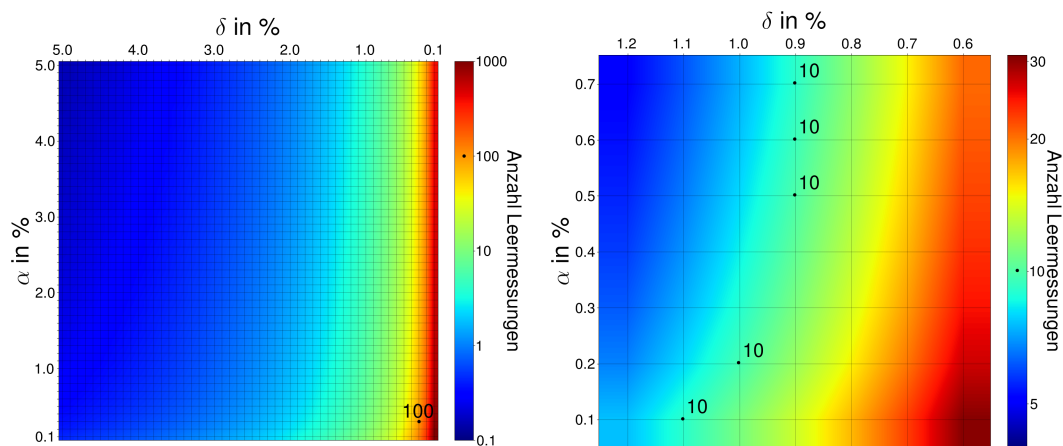
3.2.2. Auswertung für 3D-PLI Datensätze

Bestimmung des Stichprobenumfangs

Für die Bestimmung der benötigten Anzahl Leermessungen für das 3D-PLI, wurde eine Stichprobe von 100 Leermessungen verwendet und für jedes Pixel der Mittelwert $\hat{\mu}_{100}$ und die Stichprobenvarianz \hat{s}_{100}^2 berechnet. Damit lässt sich für jedes Bildpixel die benötigte Anzahl n an Messungen bestimmen, um den wahren Erwartungswert μ durch das arithmetische Mittel \bar{X}_n abschätzen zu können. Der Fehler e wurde auf einen prozentualen Anteil δ des jeweiligen $\hat{\mu}_{100}$ aus den 100 Leermessungen festgelegt. Mit der nach 3.4 bestimmten Anzahl n liegt das arithmetische Mittel \bar{X}_n also im Intervall

$$[\mu - \hat{\mu}_{100} \cdot \delta; \mu + \hat{\mu}_{100} \cdot \delta]$$

In Abbildung 3.3 ist der Verlauf der benötigten Anzahl an Messungen bei gegebenem α und δ aufgetragen. Es wurde zunächst für alle Bildpixel unabhängig voneinander die Anzahl n in Abhängigkeit des jeweiligen α und δ bestimmt. Damit für kein Pixel die Schätzung außerhalb des gewollten Schwankungsintervalls liegt, wurde das maximale n als benötigte Anzahl gewählt und eingetragen.



(a) Die Anzahl benötigter Leermessungen für $\alpha, \delta \in [5\%; 0.1\%]$.

(b) Ein vergrößerter Ausschnitt mit $\alpha, \delta \in [1.2\%; 0.6\%]$.

Abbildung 3.3.: Die maximal benötigte Anzahl n in Abhängigkeit von α und δ . In (a) ist exemplarisch die erreichbare Genauigkeit mit $n = 100$ Leermessungen markiert; in (b) verschiedene mögliche α - δ -Kombinationen, die zu $n = 10$ benötigten Leermessungen führen.

Die bisherige Berechnung der Gainfaktoren aus $n = 100$ Leermessungen entspricht somit beispielsweise einer Genauigkeit von 99.7% ($\delta = 0.3\%$) mit 99.7% ($\alpha = 0.3\%$) Sicherheitswahrscheinlichkeit. Erhöht oder verringert man die Anzahl der verwendeten Leermessungen, erhöht oder verringert sich entsprechend die erzielte Genauigkeit und Sicherheitswahrscheinlichkeit der gemessenen Mittelwerte.

Der logarithmische Verlauf in Abbildung 3.3 veranschaulicht den starken Einfluss von kleinen Änderungen in α und δ auf die Anzahl benötigter Leermessungen. Für eine Berechnung der Gainfaktoren mit $n = 10$ Leermessungen wird beispielsweise eine Genauigkeit von 99% ($\delta = 1\%$) mit 99.8% ($\alpha = 0.2\%$) Sicherheitswahrscheinlichkeit erreicht. Das Ergebnis der Auswertung für dieses α und δ ist in Tabelle 3.1 gezeigt. Es wurden die berechneten Mittelwerte $\hat{\mu}_{100}$ für die Intervallbreite und Standardabweichung \hat{s}_{100} der Stichprobe aus 100 Leermessungen verwendet und schließlich nach 3.4 das n für alle Pixel bestimmt.

	Minimum	Mittelwert	Maximum
$\hat{\mu}_{100}$	1765	2514	2838
\hat{s}_{100}	10.81	17.14	25.51
n	1.72	4.46	9.67

Tabelle 3.1.: Die Berechnung des Stichprobenumfangs mit Sicherheitswahrscheinlichkeit 99.8% ($\alpha = 0.2\%$) und Genauigkeit 99% ($\delta = 1\%$). Aus der vorliegenden Stichprobe aus 100 Leermessungen wurde für alle $18 \cdot 2048 \cdot 2048$ Pixel der Mittelwert $\hat{\mu}_{100}$ und die Standardabweichung \hat{s}_{100} bestimmt und die Anzahl n der benötigten Messungen nach 3.4 berechnet.

Dies bedeutet, im Vergleich zu $n = 100$ Leermessungen, ein Genauigkeitsverlust von nur 0.7 Prozentpunkten und etwa gleiche Sicherheit, bei einem Zehntel benötigter Messungen.

Auswertung verschiedener Sets aus Leermessungen

Für einen detaillierten Vergleich von Gainfaktoren, berechnet aus einer unterschiedlichen Anzahl verwendeter Leermessungen wurden aus der vorliegenden Stichprobe (100 Leermessungen) das Set $flat_{100}$ mit allen 100 Leermessungen und mehrere verschiedene Sets $flat_{10}$ aus je 10 Leermessungen der selben Stichprobe gebildet. Das Set $flat_{100}$ dient dabei als zur Verfügung stehende genaueste Abschätzung für die wahren Erwartungswerte der Intensitäten aller Pixel. Nach dem in Abbildung 3.1 beschriebenen Verfahren werden aus den Leermessungen zunächst die Histogramme über die mittleren Intensitäten und die Referenzwerte für jeden Winkel berechnet.

Um die Abweichung der Sets $flat_{10}$ zum Set $flat_{100}$ zu quantifizieren, wurden die absoluten Differenzen aus den mittleren Intensitäten $|\bar{I}_{100} - \bar{I}_{10}|$ gebildet. Mit der gewählten Genauigkeit $\delta = 1\%$ bei 10 Leermessungen, werden also Differenzen kleiner 1% um die mittleren Intensitäten des Set $flat_{100}$ (\bar{I}_{100}) erwartet. Diese liegen zwischen 1765 und 2838 (siehe Tab. 3.1). Die Differenzen sollten also maximal im Bereich 17 bis 28 (1% von \bar{I}_{100}) liegen. In Abbildung 3.4 ist exemplarisch für eines der Sets $flat_{10}$ das Histogramm über die absoluten Differenzen gezeigt. 99.8% der Differenzen sind kleiner als 15.91 und damit innerhalb des akzeptierten Bereichs um die approximierten wahren Erwartungswerte. Die Ergebnisse dieser Auswertung für weiterer Sets $flat_{10}$ sind am Ende des Kapitels in Tabelle 3.3 gelistet. Sie bestätigen die Stabilität und Genauigkeit für diese Sets aus 10 Leermessungen.

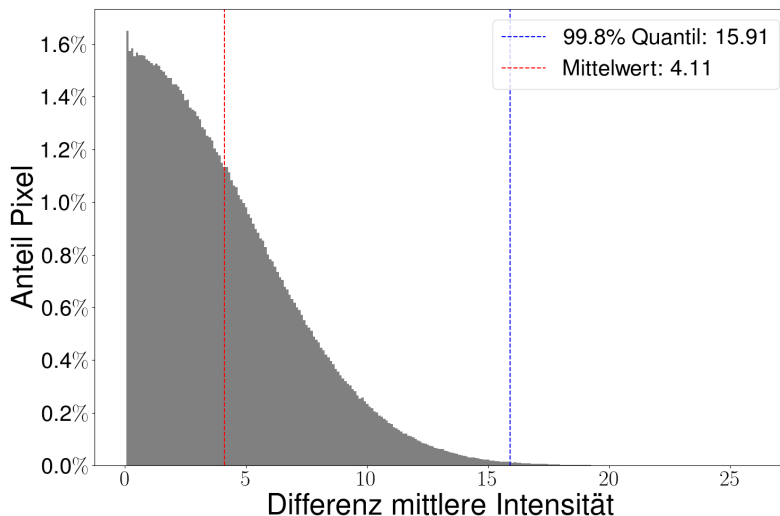


Abbildung 3.4.: Histogramm über die absoluten Differenzen der mittleren Intensitäten $|\bar{I}_{100} - \bar{I}_{10}|$. Mit $\delta = 1\%$ und erwarteten Mittelwerten im Bereich 1765-2838 (vgl. Tab. 3.1) werden Differenzen < 17 erwartet. 99.8% der Differenzen liegen im 1% Intervall um die Mittelwerte.

Aus den Histogrammen der mittleren Intensitäten wird im verwendeten Algorithmus (Abb. 3.1) nun der Referenzwert berechnet. Für alle Rotationswinkel werden dazu der maximale Peak und die Peaks $\geq 99\%$ dessen bestimmt. Finden sich mehrere Peaks in den Histogrammen eines Winkels, wird die Mitte aus dem kleinsten und

dem größten Intensitätswert der Peaks als Modalwert für diesen Winkel gewählt. Unterschiede in der Bestimmung des Referenzwertes zwischen verschiedenen 10er Stacks traten dabei auf, wenn mehrere Peaks gefunden wurden. In Abbildung 3.5 ist dieser Fall exemplarisch anhand der Histogramme der Sets $flat_{100}$ (100 Leermessungen), $flat_{10A}$ und $flat_{10B}$ (jeweils 10 Leermessungen) für die Rotationswinkel 10° und 50° gezeigt. Das Histogramm von $flat_{10A}$ hat bei Winkel 50° den maximalen Peak bei 2526 und einen Peak $> 99\%$ dessen bei 2489. Der Modalwert für diesen Winkel ist somit 2507. Für die Sets $flat_{100}$ und $flat_{10B}$ ist der entsprechende Peak minimal kleiner 99% des maximalen Peaks.

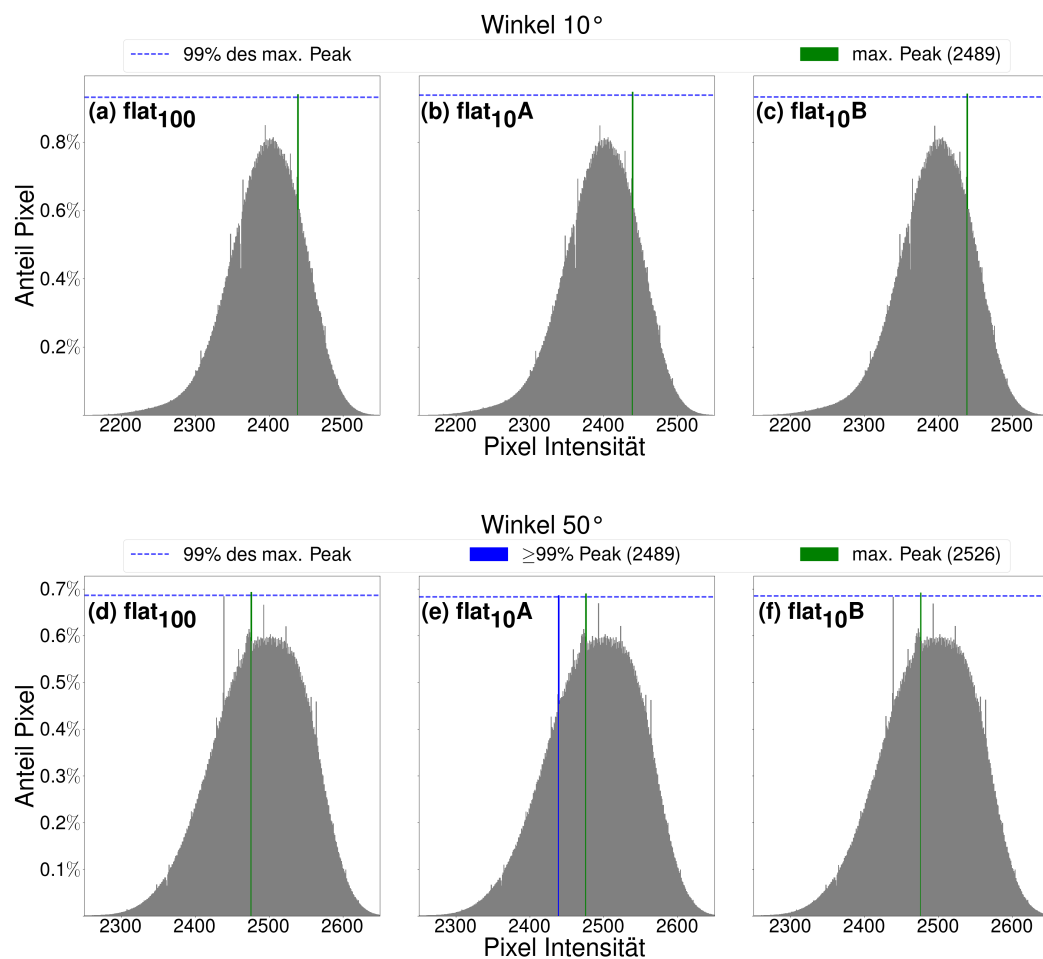


Abbildung 3.5.: Die Histogramme verschiedener Sets im Vergleich. (a) und (d) aus 100 Leermessungen; (b), (c), (e) und (f) aus jeweils 10 Leermessungen. Auswirkungen auf die Berechnung der Gainfaktoren haben die Unterschiede bei den häufigsten Intensitätswerten. Mit $\lfloor (2489 + 2526) / 2 \rfloor = 2507$ ergibt sich der Modalwert 2507 für $flat_{10A}$ bei 50° (e), während der Modalwert 2526 von (d) und (f) dem maximalen Peak des jeweiligen Histogramms entspricht.

Die Modalwerte aller Winkel für $flat_{100}$, $flat_{10}A$ und $flat_{10}B$ sind in Tabelle 3.2 gelistet. Der Median dieser 18 Werte wird als Referenzwert I_{ref} für die Gainfaktorberechnung gewählt, um die Messung für alle Rotationswinkel auf den selben Intensitätswert zu normieren. Der abweichende Modalwert bei 50° von $flat_{10}A$, resultiert in diesem Fall in einem kleineren Referenzwert I_{ref} und damit einer Normierung auf einen kleineren Wert. Eine Diskussion der möglichen Auswirkungen wird in Kapitel 3.5 geführt.

Winkel in $^\circ$	$flat_{100}$	$flat_{10}A$	$flat_{10}B$
0	2445	2467	2445
10	2489	2489	2489
20	2489	2489	2489
30	2489	2489	2489
40	2489	2489	2489
50	2526	2507	2526
60	2614	2614	2614
70	2614	2614	2614
80	2614	2614	2614
90	2614	2614	2614
100	2614	2614	2614
110	2614	2614	2614
120	2593	2614	2593
130	2543	2543	2543
140	2526	2526	2526
150	2489	2489	2489
160	2489	2489	2489
170	2489	2489	2489
I_{ref}	2526	2516	2526

Tabelle 3.2.: Die Modalwerte aller 18 Rotationswinkel für $flat_{100}$, $flat_{10}A$ und $flat_{10}B$ sowie I_{ref} als Median dieser. Für diese Stichproben ist I_{ref} der mittlere Wert der Modalwerte bei 50° und 140° (9. und 10. Wert der sortierten Referenzwerte). Unterschiedliche Modalwerte sind grau markiert.

Vergleich der Gainfaktoren

Aus den Mittelwerten der Pixel des Sets $flat_{100}$ und mehrerer Sets $flat_{10}$, sowie den jeweiligen Referenzwerten I_{ref} , wurden die entsprechenden Gainfaktoren GF_{100} und GF_{10} berechnet, diese auf ihre Verteilung hin verglichen und pixelweise die absoluten Differenzen gebildet. Bei gleichem Referenzwert I_{ref} wird auf den gleichen Wert normiert und bei ähnlichen Mittelwerten der Pixel sind eine nahezu identische Verteilung und Differenzen nahe 0 zu erwarten. Unterschiedliche I_{ref} bei der Berechnung zweier Gainfaktoren sollten sich nur auf die Lage der Verteilungskurve in x-Richtung auswirken, der Verlauf der Kurve sollte identisch bleiben.

Der Vergleich der GF_{100} mit den GF_{10} soll diese Vermutung bestätigen und die Genauigkeit von Gainfaktoren, die aus nur einem Zehntel der bisher verwendeten Leermessungen berechnet wurden, verdeutlichen. Im Vergleich der Histogramme der

GF_{100} - mit einer GF_{10} -Matrix bei gleichem Referenzwert $I_{ref} = 2526$ (Abb. 3.6 links) sind die Verteilungen sowohl in Lage als auch Verlauf nahezu identisch. Der glattere Verlauf im Histogramm der GF_{100} kommt durch die 10fache Menge an Leermessungen zur Berechnung der Gainfaktoren zustande. Auch das Histogramm über die absoluten Differenzen (Abb. 3.6 rechts) entspricht dem erwarteten Ergebnis. Die Abweichungen liegen in der Größenordnung 10^{-3} , was sich ebenfalls durch die unterschiedliche Anzahl an Leermessungen und die daraus resultierenden Mittelwerte der Pixel erklären lässt.

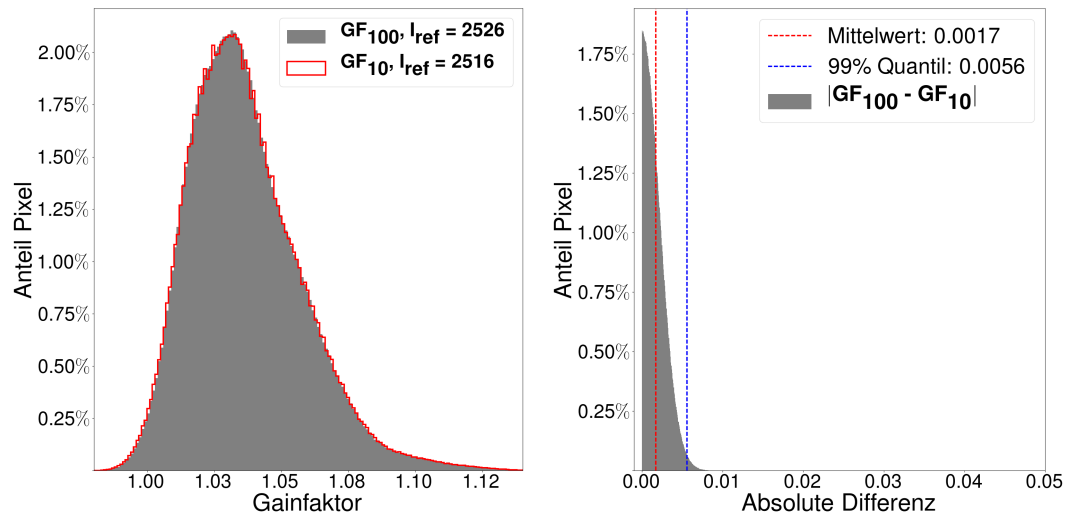


Abbildung 3.6.: Vergleich der Gainfaktoren berechnet aus 100 und aus 10 Leermessungen, exemplarisch für einen Winkel. Der Referenzwert I_{ref} ist für beide der selbe. Das Histogramm auf der linken Seite zeigt die nahezu identische Verteilung der beiden Gainfaktor-Matrizen. Rechts zu sehen ist das Histogramm der absoluten Differenzen der Gainfaktor-Matrizen pixelweise mit einer durchschnittlichen Abweichung von 0.0017 und 99% aller Abweichungen kleiner als 0.0056.

Der Vergleich der GF_{100} - mit einer GF_{10} -Matrix bei unterschiedlichen Referenzwerten ist in Abbildung 3.7 zu sehen. Die Verteilungskurve der GF_{10} Matrix im Histogramm ist nach links verschoben, da in diesem Fall auf einen kleineren Referenzwert I_{ref} normiert wird. Die Verteilungskurve ist jedoch nahezu identisch. Entsprechend ergibt sich ein größeres Maximum der absoluten Differenzen der beiden Matrizen, da die Mittelwerte aller Pixel nun etwas von den Werten der GF_{100} Matrix abweichen.

Die Auswertung mehrerer Gainfaktoren GF_{10} aus 10 Leermessungen, im Vergleich zu den bisherigen Gainfaktoren aus 100 Leermessungen, lieferte vergleichbare Ergebnisse. Die fast identischen Verteilungskurven und sehr kleinen absoluten Differenzen bestätigen die Stabilität der Gainfaktoren und der darauf basierenden Kalibrierung. Unterschiedliche Referenzwerte I_{ref} traten dabei nur selten auf.

Eine tabellarische Auflistung der durchgeführten Vergleiche zeigt Tabelle 3.3, exemplarisch für acht verschiedenen Sets aus je 10 Leermessungen. Ausgewertet wurden

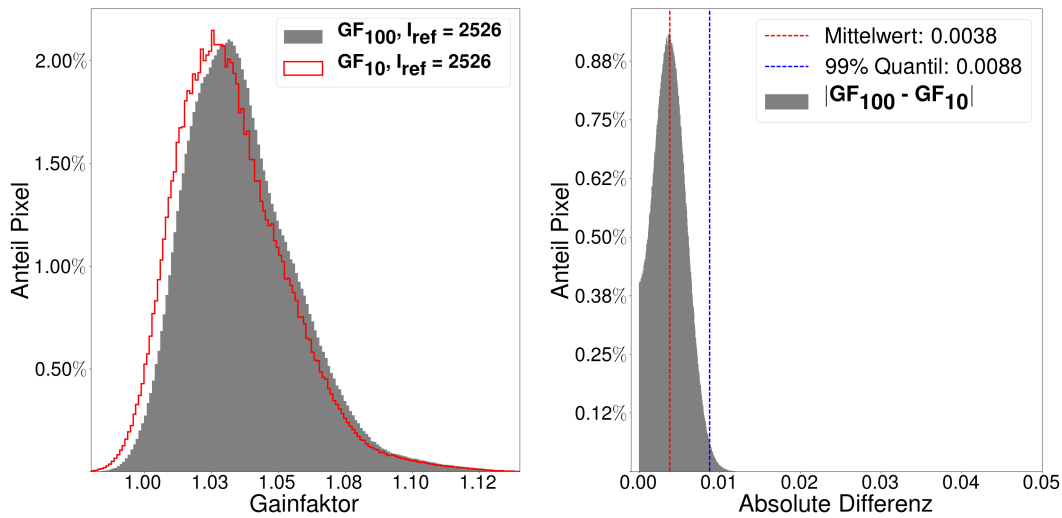


Abbildung 3.7.: Vergleich der Gainfaktoren aus 100 und aus 10 Leermessungen bei unterschiedlichem Referenzwert I_{ref} . Der kleinere I_{ref} Wert der GF_{10} sorgt für die Verschiebung in x-Richtung nach links. Das Histogramm der Differenzmatrix auf der rechten Seite zeigt die erwartete Verteilung mit nach rechts verschobenem Maximum der Differenzen.

dafür die mittleren Intensitäten \bar{I}_{10} der Leermessungen und die daraus berechneten Gainfaktoren GF_{10} , jeweils im Vergleich zum Set aus 100 Leermessungen (\bar{I}_{100} und GF_{100}). Über die absoluten Differenzen $|\bar{I}_{100} - \bar{I}_{10}|$ wurde die mittlere Differenz und das 99.8% Quantil dieser berechnet. Wie auch in Abbildung 3.4 weichen 99.8% aller Pixel um weniger als 1% vom jeweiligen Mittelwert ab. Anschließend wurde jeweils die Anzahl der abweichenden Modalwerte aller 18 Winkel bestimmt und der Referenzwert I_{ref} berechnet. Unterschiedliche Referenzwerte bzw. I_{ref} im Vergleich zu GF_{100} ergaben sich nur selten.

Mit der hier gewählten Sicherheitswahrscheinlichkeit 99.7% und Genauigkeit 99%, lassen sich also sehr stabil Gainfaktoren aus nur einem Zehntel der Anzahl an Leermessungen berechnen.

Die Ergebnisse zeigen, dass die GF_{10} sowohl zueinander als auch zu den bisherigen Gainfaktoren GF_{100} sehr ähnlich sind. Minimale Schwankungen in den Leermessungen (vgl. Abb. 3.5) können eine Normierung auf einen anderen Intensitätswert bedeuten. Das benötigte Ergebnis, ein gleichmäßig belichteter Hintergrund, wird jedoch ebenfalls erzielt.

3.3. Aktualisierung von Gainfaktoren

Während die Stabilität der Gainfaktoren primär für den Ausgleich der beschriebenen statischen Effekte wichtig ist, müssen bei der Berechnung auch die spontanen Effekte durch Änderungen im optischen System und die kontinuierlichen Effekte berücksichtigt werden. Die Aktualisierung von Gainfaktoren beinhaltet, dass solche Veränderungen erkannt und eine entsprechende regelmäßige Neuberechnung stattfindet.

	1	2	3	4	5	6	7	8
mittlere Differenz	4.12	4.11	4.11	4.11	4.11	4.11	4.11	4.11
99.8% Quantil	15.95	15.95	15.93	15.93	15.94	15.94	15.94	15.94
Abw. Modalwerte	3	0	0	0	0	1	0	0
I_{ref}	2517	2526	2526	2526	2526	2526	2526	2526

Tabelle 3.3.: Vergleich acht verschiedener Sets aus jeweils 10 Leermessungen mit dem Set aus 100 Leermessungen. Es sind die durchschnittlichen Differenzen der mittleren Intensitäten und das 99.8% Quantil dieser gelistet sowie die Anzahl abweichender Modalwerte zum Set aus 100 Leermessungen und der Referenzwert auf welchen normiert wird (I_{ref}).

Aktuell werden zur Berechnung der Gainfaktoren mit dem in Kapitel 3.1 beschriebenen Verfahren 100 Leermessungen verwendet. Eine ständige Neumessung ist jedoch, abhängig von der Häufigkeit, sehr zeitaufwändig und speicherintensiv. Deshalb wird nur bei Änderungen im Aufbau, wie beispielsweise eine andere Kamera oder Lichtquelle, eine Neukalibrierung vorgenommen. Für den zweiten Schritt der Kalibrierung, der Anwendung der Kalibrierungsdaten, werden daher die Gainfaktoren aus den Leermessungen verwendet, welche zum Zeitpunkt der Aufnahme die neuesten schon verfügbaren waren.

Diese können jedoch älteren Datums sein. Spontane und kontinuierliche Effekte werden in einem solchen Fall aktuell nicht berücksichtigt und die Aufnahmen werden damit potentiell fehlerhaft kalibriert. Größere Veränderungen in den Aufnahmen werden zwar unter Umständen erkannt, jedoch erst durch Artefakte in den Bildern bei der Weiterverarbeitung der Rohdaten, also nach der Anwendung der Gainfaktoren oder später im Rekonstruktionsworkflow.

Im Folgenden wird erläutert wie spontane und kontinuierliche Effekte besser in der Kalibrierung berücksichtigt werden können.

3.3.1. Einzelmessung

Neben den Sets aus mehreren Leermessungen zur Kalibrierung wird aktuell vor jeder Aufnahme eines Gehirnschnittes eine einzelne Leermessung durchgeführt, also eine Aufnahme ohne Objekt für jeden der 18 Rotationswinkel. Eine solche Einzelmessung wäre für die Berechnung von Gainfaktoren zwar nicht ausreichend, kann aber für eine kontinuierliche Aktualisierung der Gainfaktoren genutzt werden. Sollte beispielsweise ein Staubpartikel auf dem Sensor der Kamera hinzukommen, wegfallen oder 'gewandert' sein, würde die Anwendung von Gainfaktoren, die vor diesem Zeitpunkt berechnet wurden, auf Aufnahmen nach diesem Zeitpunkt an der entsprechenden Stelle falsch kalibrieren. Auch offensichtliche Abweichungen wie sehr dunkle Aufnahmen durch Fehler in der Kamera oder der Lichtquelle wären in der Einzelmessung ersichtlich. In einem solchen Fall kann eine Neukalibrierung oder Wiederholung der Messung vorgenommen werden. Durch die in Kapitel 3.2 beschriebene

Beschleunigung aufgrund einer deutlich verringerten Anzahl tatsächlich benötigter Leermessungen, verringert sich außerdem der Aufwand einer Neukalibrierung.

Die Verwendung der Einzelmessung ist somit ein Kompromiss zwischen einer vollständigen Neukalibrierung vor *jeder* Aufnahme und einem möglichst geringen zeitlichen Aufwand. Der Einsatz der Einzelmessung zur Aktualisierung wird im folgenden Kapitel 3.3.2 genauer diskutiert, die Bestimmung von Veränderungen durch die Verwendung der Einzelmessung in Kapitel 3.3.3.

3.3.2. Sliding Window Prinzip

Eine Möglichkeit die regelmäßige Aktualisierung der Gainfaktoren vorzunehmen ist, die zum Zeitpunkt der Aufnahme eines Gehirnschnittes neuesten Einzelmessungen zu verwenden. Das Set aus n Leermessungen für die Berechnung der Gainfaktoren würde sich damit aus den n neuesten Einzelmessungen zusammensetzen anstelle von n Leermessungen, die alle direkt nacheinander und unabhängig von tatsächlichen Aufnahmen aufgenommen wurden. Mit jeder neuen Aufnahme eines Schnittes und damit einer neuen Einzelmessung könnten neue, aktuellere Gainfaktoren für diesen Schnitt berechnet werden. Der Begriff *Sliding Window* bezeichnet somit ein 'Fenster' über ein Set aus den aktuellsten Leermessungen, welches bei einer Neuberechnung von Gainfaktoren die neueste Einzelmessung hinzunimmt und die älteste Einzelmessung aus der Berechnung ausschließt (Abb. 3.8). Die beschriebenen kontinuierlichen Effekte würden so ausgeglichen werden.

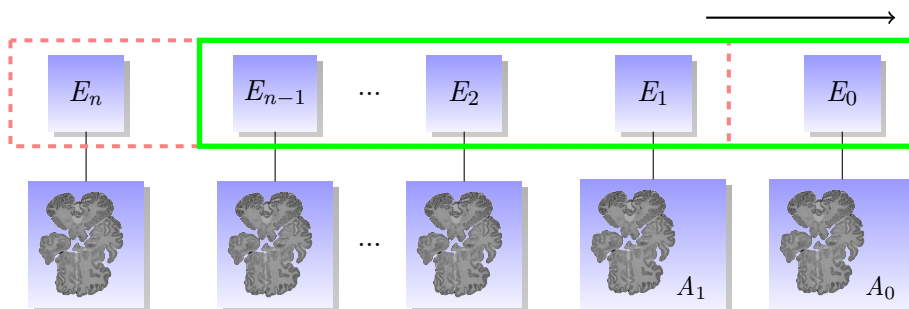


Abbildung 3.8.: Das Sliding Window Prinzip. Die Gainfaktoren zur Aufnahme A_1 setzen sich aus den n aktuellsten Einzelmessungen (E_1 bis E_n) zusammen. Kommt eine neue Aufnahme (A_0) und Einzelmessung (E_0) hinzu, verschiebt sich das Fenster und die Gainfaktoren zur neuen Aufnahme A_0 setzen sich aus E_0 bis E_{n-1} zusammen.

Unter Umständen liegen größere Zeiträume zwischen den Aufnahmen von Gehirnschnitten. Gainfaktoren berechnet aus Einzelmessungen sind in einem solchen Fall nicht so aktuell wie solche basierend auf einem kompletten Set aus Leermessungen, die direkt vor der Aufnahme gemacht werden. Eine ständige Neukalibrierung eines vollständigen Sets aus n Leermessungen ist jedoch, abhängig von Häufigkeit und Anzahl benötigter Leermessungen, sehr zeitaufwändig und speicherintensiv. Die Verwendung des *Sliding Window Prinzips* ist somit der erwähnte Mittelweg aus voll-

ständiger Neukalibrierung vor jeder Aufnahme und einem geringen zeitlichen Aufwand bzw. Speicherbedarf.

3.3.3. Ausreißer Analyse

Bei der Verwendung der Einzelmessungen für die Berechnung der Gainfaktoren muss gewährleistet werden, dass es sich bei diesen nicht um Ausreißer handelt, die zum Beispiel durch falsche Einstellungen oder Beleuchtung entstanden sind. Um Ausreißer feststellen zu können, wird daher eine Einzelmessung mit einem Set aus Leermessungen verglichen. Die neu hinzukommende Einzelmessung kann dafür mit den vorherigen Einzelmessungen verglichen werden. Damit ist eine Art Selbstkontrolle gegeben; jede korrekte Einzelmessung wird in das Set zur Kontrolle der nächsten aufgenommen.

Für die Ausreißer-Analyse spielen zwei Kriterien eine wichtige Rolle: Einerseits sollten für alle 18 Rotationswinkel die Einzelmessungen auf globaler Ebene verglichen werden, andererseits sollten auf den Bildern lokale Änderungen auf Pixelebene erkannt werden, zum Beispiel Staubpartikel mit wenigen Pixeln Durchmesser. Dafür ist in beiden Fällen ein Schwellenwert für signifikante beziehungsweise nicht signifikante Veränderungen festzulegen.

Globale Änderungen

Als globale Änderungen werden bildweite Unterschiede zu vorherigen Leermessungen definiert. Dabei gilt es zu unterscheiden, ob eine potentiell fehlerhafte Einzelmessung vorliegt oder eine nicht signifikante Abweichung, wie der Trend zu dunkler werdenden Bildern durch den beschriebenen Leuchtkraftverlust der LED-Lampe. Ein sehr dunkles oder sehr helles Bild deutet beispielsweise auf einen Fehler in der Belichtung hin und die Einzelmessung sollte nicht zur Kalibrierung verwendet werden. Zusätzlich lassen sich solche Fälle gut mit einer Qualitätssicherung verbinden: Die untersuchte Einzelmessung wird direkt vor der Aufnahme des Gehirnschnittes aufgenommen, diese weist folglich mit hoher Wahrscheinlichkeit den gleichen Fehler auf und sollte somit wiederholt werden.

Optisch unauffällige Leermessungen, welche global jedoch von bisherigen abweichen, können mit dem *Sliding Window Prinzip* erkannt werden. Dazu gehören Veränderungen im Helligkeitsprofil des Bildes, beispielsweise Unterschiede in den Unregelmäßigkeiten. Abbildung 3.9 zeigt ein Beispiel für global verschiedene und global sehr ähnliche Leerbilder, jeweils für den gleichen Winkel. Der Kontrast in den Bildern ist gleichermaßen erhöht worden, um die Unterschiede zu verdeutlichen.

Die mittlere Differenz der Leerbilder von August und Dezember (Abb. 3.10a) ist deutlich größer als die mittlere Differenz der Leerbilder von Dezember und März (Abb. 3.10b). Bildet man vor der Berechnung der Gainfaktoren mit dem *Sliding Window Prinzip* die Differenzen der neuen Einzelmessung zu den vorherigen Einzelmessungen, können damit globale Ausreißer dieser Art erkannt werden. Dazu kann, für eine automatische Überprüfung, auch ein Schwellenwert festgelegt werden, ab welcher Abweichung ein Ausreißer vorliegt. In einem solchen Fall muss eine genauere Untersuchung auf Pixelebene nicht durchgeführt werden und entsprechende

Maßnahmen können vorgenommen werden. Zunächst bietet sich dafür eine Wiederholung der Messung an. Falls dies nicht zum gewünschten Ergebnis führt, etwa weil eine dauerhafte Veränderung vorliegt, sollte eine Neukalibrierung vorgenommen werden.

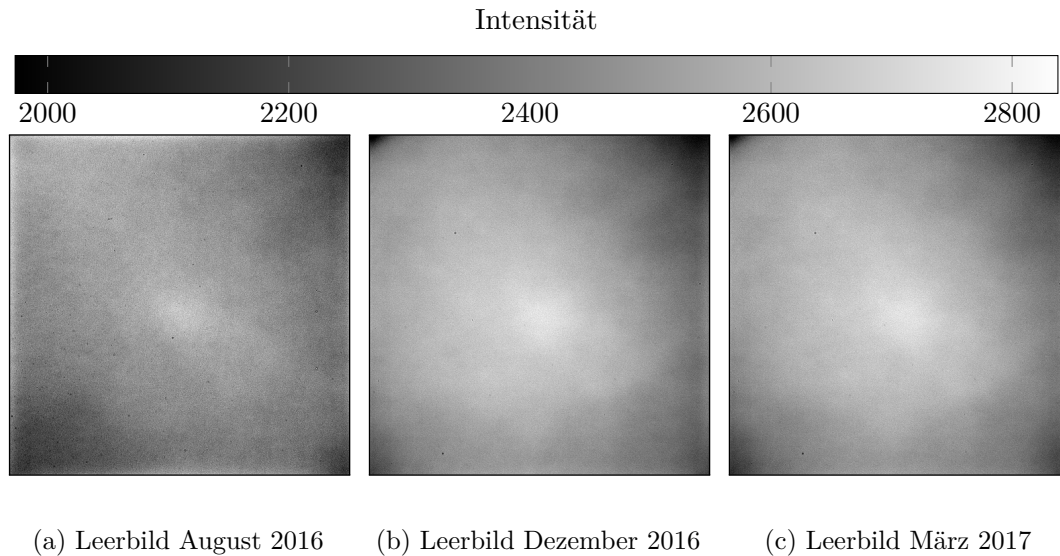


Abbildung 3.9.: Vergleich von Leerbildern aus verschiedenen Leermessungen. Es sind eine unterschiedliche Helligkeitsverteilung und Vignettierung zwischen (a) und (b) zu erkennen, jedoch optisch keine Unterschiede zwischen (b) und (c). Zur Verdeutlichung der Unterschiede ist der Kontrast gleichermaßen erhöht worden.

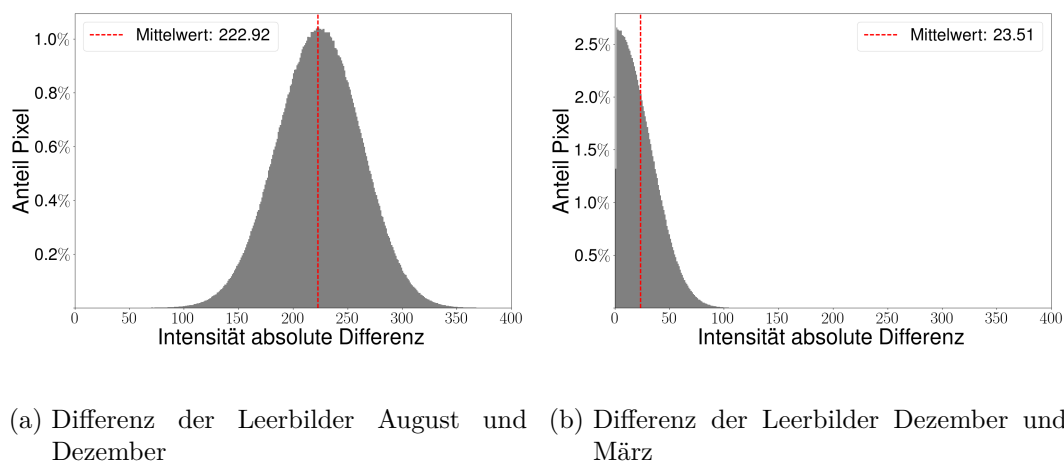


Abbildung 3.10.: Die Histogramme über die pixelweise absoluten Differenzen der Leerbilder aus Abbildung 3.9. Der deutlich größere Mittelwert der Differenzen der Leerbilder von August und Dezember weist auf die globale Abweichung hin.

Lokale Änderungen

Sind zwei Leermessungen nahezu identisch, wurden also keine globalen Abweichungen gefunden, kann zusätzlich auf Pixelebene verglichen werden, um lokale Änderungen zu erkennen. Dabei gilt es zu beachten, dass ein pixelweiser Vergleich zu genau wäre, da Bildrauschen als potentielle lokale Änderung detektiert würde, jedoch bei zu großflächigem Vergleich kleine lokale Abweichungen unter Umständen nicht erkannt werden. Es muss also ein Mittelweg für die Genauigkeit der Überprüfung gefunden werden.

Weiterhin muss ein Schwellenwert für die noch akzeptierte Abweichung zwischen den Leerbildern festgelegt werden, also die Grenze zwischen minimalen lokalen Änderungen und natürlichem Bildrauschen.

Über Erfahrungswerte kann ein $x \times y$ Pixel großer Bereich definiert werden, welcher die einzelnen Leerbilder abschnittsweise miteinander vergleicht. Ist beispielsweise ein Staubpartikel zwischen zwei Leermessungen hinzugekommen, verschwunden oder gewandert, unterscheiden sich die Messungen an dieser Stelle. In Abbildung 3.11 ist ein Ausschnitt aus einem Leerbild mit Staubpartikel gezeigt. Ein Bereich von etwa 10×10 Pixeln um einen Staubpartikel ist dunkler und weist eine geringere mittlere Intensität auf als ein Bereich gleicher Größe ohne Staubpartikel. Bewegt

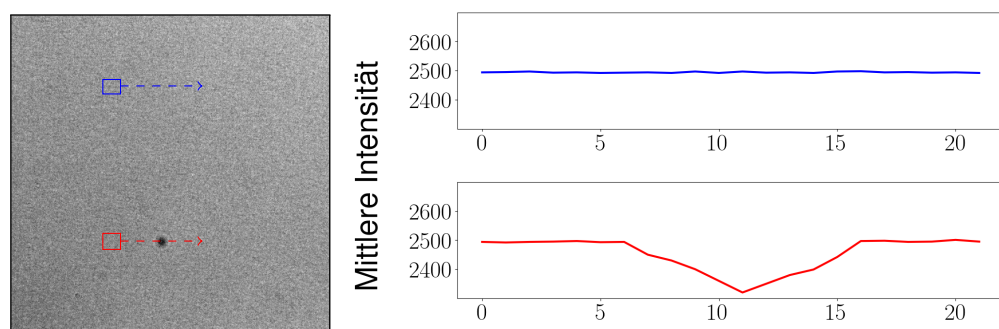


Abbildung 3.11.: Links ein Ausschnitt aus einem Leerbild mit einem Staubpartikel von etwa 10×10 Pixeln. Der Verlauf der mittleren Intensitäten eines Bereichs mit und ohne Staubpartikel ist auf der rechten Seite gezeigt. Der Staubpartikel kann durch ein lokales Minimum identifiziert werden.

man nun solche Fenster pixelweise über die Einzelmessungen und bildet jeweils die mittleren Intensitätswerte, können Unterschiede festgestellt werden. Beispielsweise würde ein Fenster, welches über einen Staubpartikel bewegt wird (rot), im Verlauf der mittleren Intensitäten als lokales Minimum erkennbar sein (Abb. 3.11 rechts unten). Das gleiche Fenster in einem anderen Leerbild ohne Staubpartikel an dieser Stelle (blau), würde einen Verlauf ohne Minimum zeigen (Abbildung 3.11 rechts oben).

3.4. Auswertung und Zusammenfassung

Mit den in Kapitel 3.2 und 3.3 beschriebenen Methoden und Verfahren sind eine sinnvolle Begrenzung der Anzahl zu verwendender Leermessungen und eine steti-ge Aktualisierung der Gainfaktoren möglich. So werden bei der Berechnung der Gainfaktoren für die Kalibrierung sowohl statische als auch kontinuierliche Effekte ausgeglichen. Eine zusätzliche Kontrolle der Gainfaktoren bietet die Erkennung von spontanen Effekten. Dies lässt sich in einem *Kalibrierungsworkflow* zusammenfassen (Abb. 3.12).

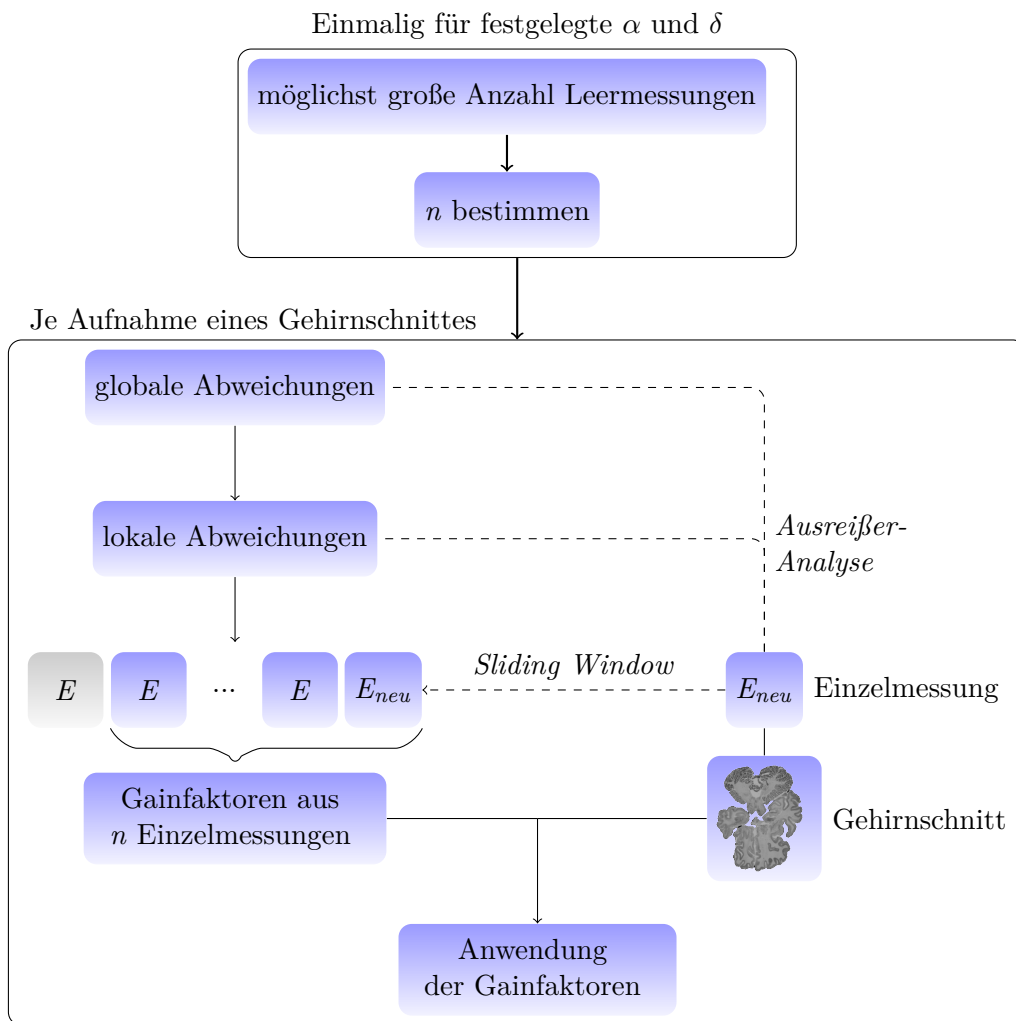


Abbildung 3.12.: Der entwickelte Kalibrierungsworkflow. Zunächst muss einmalig mit festgelegtem α und δ die Anzahl benötigter Leermessungen bestimmt werden. Bei jeder Aufnahme eines Gehirnschnittes mit Einzelmessung kann eine Ausreißer-Analyse durchgeführt werden und schließlich die Berechnung der Gainfaktoren mit dem *Sliding Window Prinzip*.

Zur Validierung dieser Methoden und Verfahren wurden exemplarisch Gainfaktoren aus zehn aufeinanderfolgenden Einzelmessungen, aufgenommen in einem Zeitraum von 20 Tagen, gebildet und mit den bisher verwendeten Gainfaktoren verglichen, die auf 100 an einem Tag aufgenommenen Leermessungen basieren. Dazu wurden beide Gainfaktoren auf die aktuellste Leermessung angewandt, um so gleichmäßig belichtete Bilder ohne Unregelmäßigkeiten zu erhalten. Die Auswertung dieser Bilder ermöglicht eine Quantifizierung der Reduzierung der Inhomogenität und damit einen Vergleich verschiedener Gainfaktoren. Ausgewertet werden dabei die Verteilungen der Intensitäten vor und nach der Kalibrierung (siehe Abb. 3.13).

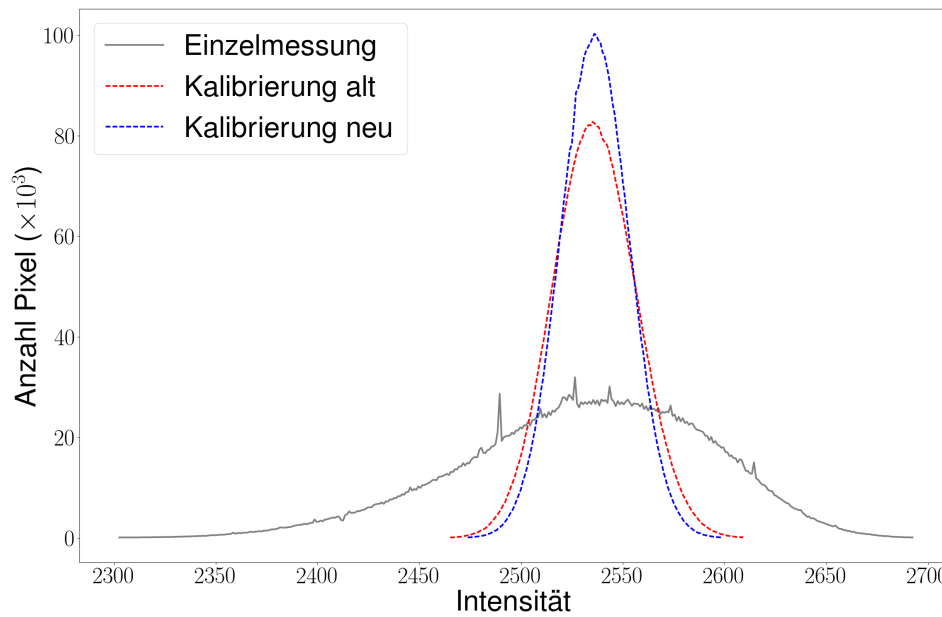


Abbildung 3.13.: Die Verteilungen der Intensitäten einer Einzelmessung *vor* (grau) und *nach* (rot/blau) der Kalibrierung, exemplarisch für einen Winkel. Die Gainfaktoren mit den vorgestellten neuen Methoden aus 10 Einzelmessungen (blau), führen mit nur einem zehntel an Leermessungen der bisherigen Gainfaktoren (rot) zu einem besseren Ergebnis.

Die Reduzierung der Inhomogenität (R_{cal}) kann dabei über die Halbwertsbreiten (FWHM, engl.: *Full Width at Half Maximum*) der Verteilungen in Abbildung 3.13 quantifiziert werden [12]. Mit $FWHM_{cal}$ und $FWHM_{raw}$ als Halbwertsbreiten *nach* und *vor* der Kalibrierung, ergibt sich

$$R_{cal} = 1 - \frac{FWHM_{cal}}{FWHM_{raw}} \quad (3.5)$$

als prozentuale Reduzierung der Inhomogenität durch die jeweiligen Gainfaktoren. Die Auswertung mit den bisherigen Gainfaktoren (rot) ergab eine Reduzierung von etwa 68%. Mit den neuen Gainfaktoren (blau) wurden etwa 74% erreicht.

3.5. Optimierungen

Zu den vorgestellten Verbesserungen in der Berechnung der Kalibrierungsdaten und dem entwickelten Rekonstruktionsworkflow können verschiedene Optimierungen vorgenommen werden. Im Folgenden werden zwei einfach zu implementierende Optimierungen vorgestellt. Weitere mögliche Erweiterungen werden in Kapitel 5 diskutiert.

Konstanter Normierungswert

Wie schon in Tabelle 3.2 gezeigt wurde, können kleine Unterschiede in den Leermessungen die Normierung beeinflussen. Auf alle Kacheln eines Gehirnschnittes werden die gleichen Gainfaktoren angewandt. Der Schnitt wird damit auf den gleichen Intensitätswert normiert. Mit dem implementierten *Sliding Window Prinzip* müssen für die Aufnahme eines Schnittes mit neuer Einzelmessung jedoch neue Gainfaktoren berechnet werden. Dieser Schnitt wird damit unter Umständen auf einen anderen Wert als der vorherige normiert. Zwei aufeinanderfolgende Gehirnschnitte wären zwar in sich konsistent, unterschiedliche absolute Intensitätswerte können aber zu Problemen bei der Registrierung führen. Um die Normierungen auf verschiedene Referenzwerte aufeinanderfolgender Schnitte zu vermeiden, ist eine mögliche Optimierung, einen festen Intensitätswert zu bestimmen. Als Orientierung kann beispielsweise der Mittelwert bisheriger I_{ref} verwendet werden (≈ 2500). Jedoch hätte diese Methode auch einen Nachteil: Bei der Anwendung der Gainfaktoren wurde bisher auf eine Minimierung der Veränderungen geachtet. Mit der Wahl des Modalwertes zur Normierung wurde dies erreicht. Mit einer Normierung auf einen festen Wert werden im Vergleich dazu mehr Pixel um einen größeren Betrag angepasst (aufgehellt oder verdunkelt).

Gewichtung der Einzelmessung

Bei n verwendeten Leermessungen für die Kalibrierung und unter Verwendung des *Sliding Window Prinzips*, hat die neu hinzukommende Einzelmessung ein Gewicht von $1 : n$ bei der Berechnung der Mittelwerte jedes Pixels. Zeigt die neue Einzelmessung nun messbare, jedoch nicht kritische Abweichungen zu den anderen $n - 1$ Leermessungen, würde dies die daraus resultierenden Gainfaktoren nur gering beeinflussen. Die Kalibrierung mit diesen Gainfaktoren würde somit die aktuellen Verhältnisse nicht vollständig korrigieren. Eine stärkere Gewichtung der aktuellsten Einzelmessung bei der Berechnung der Gainfaktoren ist möglich, jedoch würden damit auch Ausreißer in dieser stärker gewichtet.

Je nach festgelegter Anzahl Leermessungen zur Berechnung der Gainfaktoren, wäre eine einfache Optimierung eine regelmäßige vollständige Neukalibrierung. Gerade bei fehlender Aktualisierung mit Einzelmessungen auf Grund größerer Zeiträume zwischen den Aufnahmen, wäre dies eine wichtige Maßnahme. Werden in einem zu definierenden Zeitraum keine Aufnahmen und damit Leermessungen gemacht, sollte vor der nächsten Aufnahme statt einer Einzelmessung neu kalibriert werden. Die aktuellsten und genauesten Gainfaktoren würden natürlich durch eine vollständige Kalibrierung mit n Messungen vor *jeder* Aufnahme erreicht werden, was jedoch sehr zeit- und speicherintensiv wäre.

4. Anwendung der Kalibrierungsdaten

Nachdem die Gainfaktoren im ersten Schritt der Kalibrierung berechnet wurden, beinhaltet der zweite Schritt die Anwendung dieser Kalibrierungsdaten, also die Berechnung von kalibrierten Rohdaten für die Weiterverarbeitung im Rekonstruktionsworkflow (siehe Kapitel 2.2). Dabei sind zwei Aspekte von Bedeutung: Eine effiziente *Speicherung* der Daten und eine effiziente *Anwendung* der Gainfaktoren auf die Rohbilder. Im Folgenden wird deshalb auf die Wahl des Datenformats und die Parallelisierung des zweiten Kalibrierungsschrittes eingegangen.

4.1. Speicherung der Daten

Für die Speicherung der 3D-PLI Daten gilt es zu beachten, dass die in Kapitel 2.1 beschriebene Aufnahmetechnik des Polarisationsmikroskops sehr große Datenmengen erzeugt. Die Aufnahme eines Gehirns besteht aus etwa 1500 Schnitten mit jeweils etwa 1500 Kacheln. Pro Kachel werden 18 Bilder aufgenommen. Dies ergibt über 40 Millionen Rohbilder zur Weiterverarbeitung für ein einziges menschliches Gehirn. Da auch die Zwischenergebnisse des Rekonstruktionsworkflows verfügbar sein sollen, kommen zu den Rohdaten noch die gleiche Menge kalibrierter Daten hinzu sowie die daraus berechneten Daten wie Retardierung, Direktion, Transmittanz und Inklination. Unabhängig vom benötigten Speicherplatz ist bereits eine so große Anzahl an Dateien ein Problem für viele Dateisysteme. Aus diesem Grund wurde das *Hierarchical Data Format 5* (HDF5) als Datenformat für das 3D-PLI ausgewählt.

4.1.1. HDF5 Datenformat

Das *Hierarchical Data Format* (HDF) wurde vom National Center for Supercomputing Applications (NCSA) entwickelt. Das Ziel war es eine architektur-unabhängige Software zu entwickeln, um Daten wissenschaftlicher Anwendungen auf verschiedenen Computerplattformen zu speichern und verwenden zu können [13].

Das aktuelle Datenformat HDF5 ist ein vollständig hierarchisches Datenmodell, welches die Speicherung komplexer, mehrdimensionaler Daten zusammen mit zugehörigen Metadaten ermöglicht. Diese werden als Objekte in einer HDF5-Datei ähnlich wie in einem Dateisystem abgespeichert. Jedes Objekt hat einen Pfad innerhalb der Datei, unter dem es gespeichert ist. Dabei gibt es keine Begrenzung in Anzahl und Größe der Objekte. Weiterhin unterstützt die HDF5-API den parallelen Zugriff über das *Message Passing Interface* (MPI)¹.

¹<https://support.hdfgroup.org/HDF5/PHDF5/>

Gruppen

Gruppen dienen der Organisation von Objekten innerhalb einer Datei. Jede HDF5 Datei besitzt ein Wurzelverzeichnis „/“, unter welchem alle weiteren Objekte abgespeichert werden. Gruppen sind vergleichbar mit Verzeichnissen in UNIX-Dateisystemen und können weitere Objekte und Gruppen enthalten. Dabei gibt es keine Beschränkung in der Verschachtelungstiefe. Gruppen können über ihren Namen und den Pfad adressiert werden.

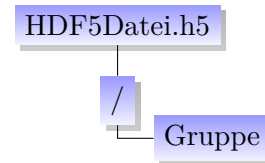


Abbildung 4.1.: HDF5-Gruppe.

Datensätze

Ein HDF5-Datensatz (*engl.: dataset*) ist zusammengesetzt aus den eigentlichen Daten und Metadaten (Abb. 4.2). Wie auch Gruppen haben die Datensätze einen Namen, mit welchem auf sie zugegriffen werden kann.

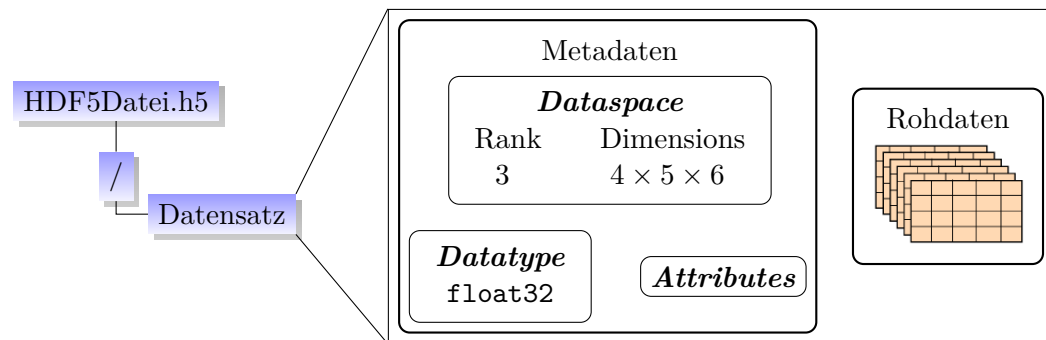


Abbildung 4.2.: Ein HDF5-Datensatz besteht aus Rohdaten und beschreibenden Metadaten [13]. Dieses Beispiel zeigt einen dreidimensionalen Datensatz („rank=3“) der Größe $4 \times 5 \times 6$ aus `float32` Werten.

Der *Dataspace* eines Datensatzes muss beim Anlegen des Objektes festgelegt werden und beschreibt die logische Struktur der Daten. Damit trotzdem ein effizienter Zugriff auf sehr große Datensätze, welche nicht auf einmal in den Arbeitsspeicher passen würden, möglich ist, kann über den *Dataspace* auf sogenannte *Subsets* (Unterbereiche) zugegriffen werden (Abbildung 4.3). Die dafür benötigten Informationen sind:

- Offset: Abstand des Unterbereichs vom Ursprung in jeder Dimension
- Count: Die Anzahl der Unterbereiche in jeder Dimension
- Stride: Der Abstand zwischen den Unterbereichen
- Block: Die Größe der Unterbereiche

Das Auswählen von Unterbereichen ist außerdem für zwei weitere Eigenschaften von HDF5 von Bedeutung: Zum einen kann ein paralleler Zugriff auf einen Daten-

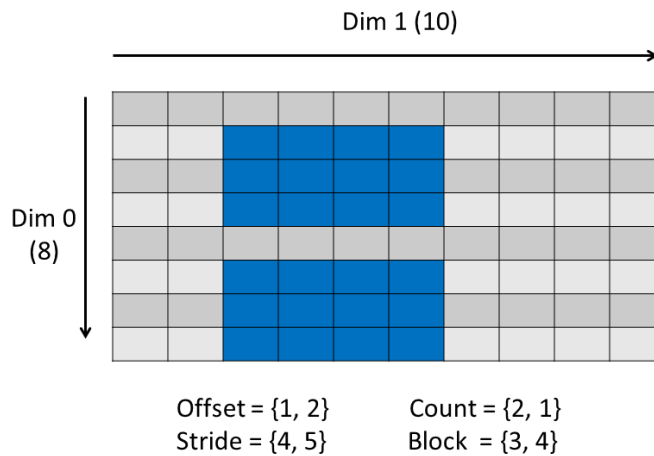


Abbildung 4.3.: Beispiel für den Zugriff auf Unterbereiche eines Datensatzes der Dimension 8×10 . Mit den angegebenen Parametern würde auf den markierten Bereich zugegriffen werden.

satz durch das Auswählen von nicht überlappenden Unterbereichen erfolgen, zum anderen kann der Zugriff auf Unterbereiche durch *Chunking* beschleunigt werden.

Chunking bezeichnet das Einteilen eines Datensatzes in Bereiche, welche linear im Speicher liegen (Abb. 4.4). Beim Zugriff auf Unterbereiche des Datensatzes müssen nur die daran beteiligten Chunks in den Cache geladen werden. Gerade bei sehr großen Datensätzen, die abschnittsweise gelesen und geschrieben werden, kann die Wahl von *Chunks* passender Größe den Zugriff beschleunigen.

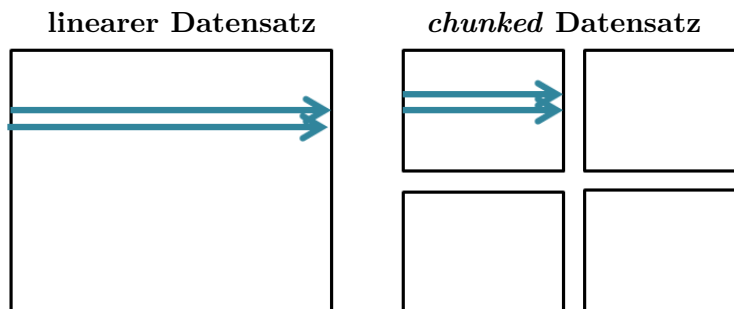


Abbildung 4.4.: Die Daten eines HDF5-Datensatzes werden standardmäßig linear abgespeichert (links). Die Einteilung eines Datensatzes in *Chunks* (rechts) speichert diese jeweils linear ab. Der Zugriff auf diese Bereiche ist damit schneller.

Attribute

Optional können zu den Objekten (Gruppen und Datensätze) Attribute abgespeichert werden. Sie bestehen aus einem Namen und einem Wert und dienen der Be-

schreibung des Objektes. Sie haben zwar wie Datensätze einen *Dataspace* und Datentyp, unterstützen aber keine weiteren Eigenschaften, um den Speicherbedarf klein zu halten.

Links

Links können zwischen Objekten innerhalb einer HDF5-Datei (interne Links) und Objekten verschiedener HDF5-Dateien (externe Links) gesetzt werden. Bei Verwendung eines Links verhält sich das verlinkte Objekt genauso wie beim direkten Zugriff. Mit internen Links können Daten unter verschiedenen Pfaden innerhalb einer Datei erreicht werden, ohne diese mehrfach abspeichern zu müssen (Abb. 4.5).

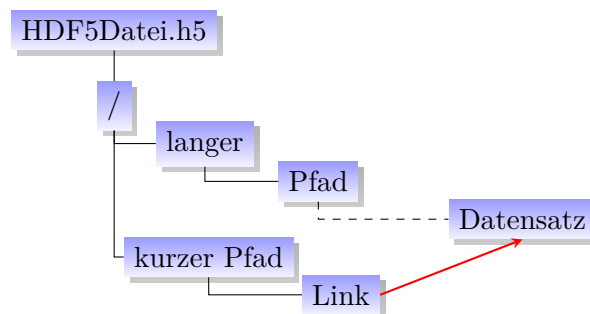


Abbildung 4.5.: Ein Interner Link auf einen Datensatz in einer HDF5-Datei. Der Zugriff kann über `"/langer/Pfad/.../Datensatz"` und `"/kurzer Pfad/Datensatz"` erfolgen.

Externe Links ermöglichen es, Daten auf verschiedene Dateien zu verteilen und über Links zu verknüpfen (Abb. 4.6). Für den Benutzer der HDF5-API verhält sich der Zugriff so, als wären alle Daten in einer Datei.

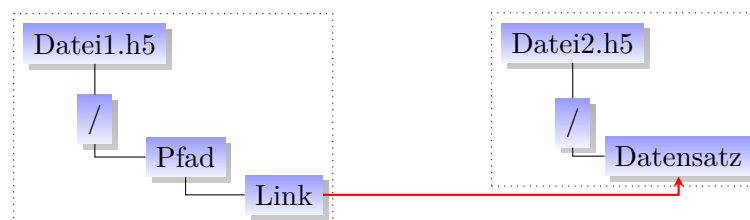


Abbildung 4.6.: Ein Link zwischen zwei HDF5-Dateien. Der Datensatz ist sowohl direkt über `"/Datensatz"` in Datei2.h5 erreichbar als auch über den Link `"/Pfad/Datensatz"` in Datei1.h5.

4.1.2. Anwendung auf 3D-PLI Daten

Für die Daten des 3D-PLI werden die Rohbilder eines Schnittes und die Zwischenergebnisse des Rekonstruktionsworkflows jeweils in eigenen Datensätzen abgespeichert. Zusätzlich werden diese Datensätze auf mehrere Dateien verteilt. Damit ist eine ausreichende Trennung der Daten vorhanden und für den Zugriff muss nicht

jedes Mal eine sehr große Datei geöffnet werden. Eine zentrale HDF5-Datei mit dem Namen des Datensatzes, z.B. `Brain_001.h5`, enthält externe Links zu Dateien mit den jeweiligen Schnittnummern und diese jeweils weitere Verlinkungen zu Dateien mit den Rohdaten und Zwischenergebnissen (Abb. 4.7). Damit sind die Daten gut organisiert und es ist ein einfacher Zugriff seitens des Benutzers über die verlinkten Dateien möglich.

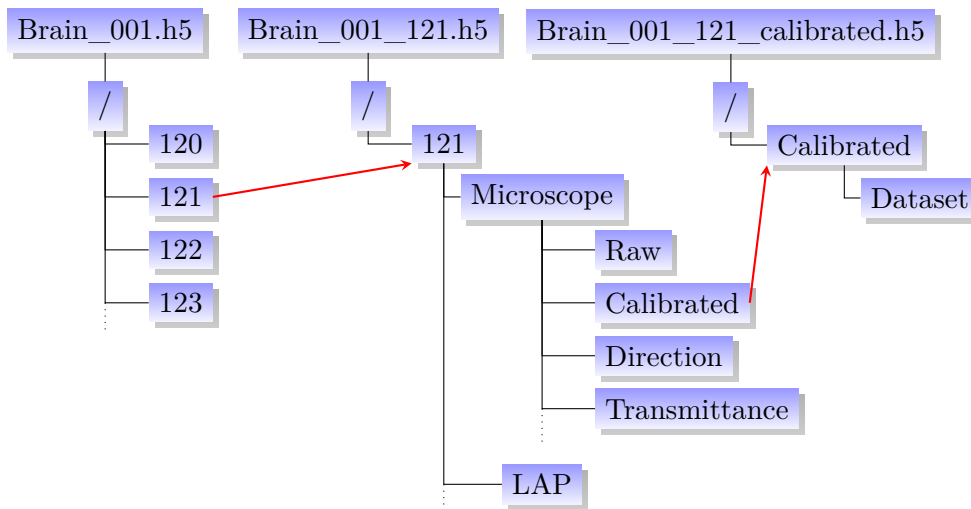


Abbildung 4.7.: Die gewählte HDF5 Struktur in vereinfachter Darstellung. Der Zugriff erfolgt je Gehirn über eine Datei (z.B. `Brain_001.h5`), in der alle Schnitte verlinkt sind (z.B. `Brain_001_121.h5`), mit weiteren Verlinkungen zu den jeweiligen Daten (z.B. `Brain_001_121_calibrated.h5`).

Die einzelnen Kacheln eines Schnittes werden außerdem in Dateien mit *Chunks* in der Größe der Kacheln gespeichert, um den Zugriff bei kachelweiser Auswertung zu beschleunigen. Dies unterstützt auch die parallele Verarbeitung eines Schnittes (siehe Kapitel 4.2).

Tabelle 4.1 zeigt die gewählten Chunk-Größen der HDF5-Dateien für die auf Kacheln arbeitenden Schritte des 3D-PLI (bis zum *Stitching*). Beim Polarisationsmikroskop haben die Kacheln die Dimension 2048×2048 . Für die Kalibrierung werden für die Verarbeitung der Kacheln alle 18 Bilder benötigt, die Chunks also in der Größe $18 \times 2048 \times 2048$ angelegt. Nach der *Efficient-PLI-Analysis (EPA)* ist nur noch ein Bild je Kachel vorhanden, also ein zweidimensionaler Datensatz und damit Chunks der Größe 2048×2048 . Da nach dem *Stitching* die Schnitte nicht mehr gleichmäßig in Kacheln unterteilt vorliegen, ist eine einheitliche Einteilung in Chunks ab diesem Verarbeitungsschritt nicht mehr möglich.

Verarbeitungsschritt	Daten	Bilder/Kachel	Größe der Chunks
Kalibrierung	Rohdaten Gainfaktoren kalibrierte Daten	18	(18, 2048, 2048)
EPA	Direktion Transmittanz Retardierung	1	(2048, 2048)
Segmentierung	Maske, Gewebe	1	(2048, 2048)
Stitching			

Tabelle 4.1.: Die Speicherung der Daten des 3D-PLI in HDF5-Dateien. Alle in Kacheln der Dimension 2048×2048 vorliegenden Daten werden in HDF5-Dateien mit Chunks entsprechender Größe gespeichert. Ab dem Stitching sind die Schnitte nicht mehr gleichmäßig in Kacheln unterteilt. Eine Einteilung in Chunks fester Größe ist damit nicht mehr möglich.

4.2. Parallelisierung

Die beim 3D-PLI erzeugten Daten haben einen Speicherbedarf im Terabyte Bereich. Bereits die Rohdaten eines Schnittes mit 1500 Kacheln benötigen 216GB (1500 Kacheln je 18 Bilder, 8MB pro Bild). Durch den begrenzten Arbeitsspeicher eines Rechners müssen die Daten deshalb nacheinander geladen und verarbeitet werden. Die Hinzunahme mehrerer Prozessoren mit jeweils separatem Arbeitsspeicher würde die gleichzeitige Verarbeitung einer größeren Menge an Daten ermöglichen. Damit kann eine schnellere Berechnung der kalibrierten Daten beziehungsweise eine geringere Laufzeit des aktuell für die Kalibrierung verwendeten Programms erreicht werden. Aus diesem Grund wurde im Rahmen dieser Arbeit die Anwendung der Gainfaktoren parallelisiert. Eine Parallelisierung der Berechnung der Gainfaktoren würde bei einer Laufzeit von nur wenigen Sekunden keine signifikante Verbesserung erreichen. Im Folgenden wird auf die Anforderungen an die parallele Verarbeitung der Daten eingegangen und eine geeignete Parallelisierungsstrategie gewählt. Abschließend wird die Laufzeit des entwickelten parallelen Programms analysiert.

4.2.1. Parallelisierungsstrategie

Nach der Klassifikation von Michael Flynn werden Rechnerarchitekturen nach der Anzahl der Befehls- (*instruction stream*) und Datenströme (*data stream*) unterteilt [14]. Unterschieden werden dabei vier verschiedene Arten:

	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD

Tabelle 4.2.: Die Klassifikation von Rechnerarchitekturen nach Michael Flynn. Die Bezeichnungen ergeben sich durch die Anzahl der zugrundeliegenden Befehls- (*instruction*) und Daten- (*data*) Ströme.

Die Anwendung der Gainfaktoren auf 3D-PLI-Daten ist eine elementweise Multiplikation der Rohdaten aller Kacheln eines Gehirnschnittes mit der immer gleichen Gainfaktor-Matrix, also eine Rechenoperation angewandt auf verschiedenen Daten. Damit eignet sich die SIMD (*Single Instruction, Multiple Data*) Architektur am besten für die Kalibrierung (Abb. 4.8). Da keine Abhängigkeiten zwischen den Daten hinsichtlich ihrer Verarbeitung bestehen, liegt zusätzlich eine, für Bildverarbeitung typische, Datenparallelität vor. Auf SIMD Architekturen können also parallel durch eine einzelne Anweisung Datenmengen beliebiger Größe und in beliebiger Reihenfolge geladen werden sowie die Verarbeitung dieser erfolgen.

Für eine geeignete Parallelisierungsstrategie muss schließlich noch die Art der Arbeitsaufteilung festgelegt werden. Zunächst muss dafür zwischen Funktions- und Gebietszerlegung unterschieden werden. Bei der **Funktionszerlegung** werden die Rechenoperationen, also der *Instruction Pool*, aufgeteilt und von verschiedenen Prozessoren berechnet. Bei der **Gebietszerlegung** werden die zu verarbeitenden Daten

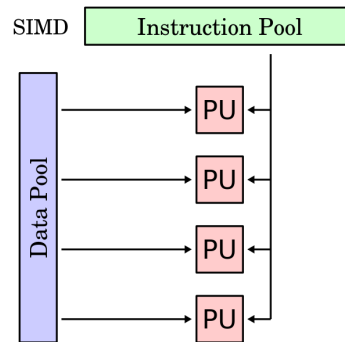


Abbildung 4.8.: SIMD Architektur. Der *Instruction Pool* beinhaltet die pixelweise Multiplikation der Rohdaten mit den Gainfaktoren. Die einzelnen Rohbilder aus dem *Data Pool* können parallel geladen und von den *Processing Units (PU)* gleichzeitig verarbeitet werden.

auf die Prozesse aufgeteilt. Da die Anwendung der Gainfaktoren nur eine elementweise Multiplikation beinhaltet, ist eine Funktionszerlegung in diesem Fall nicht anwendbar. Durch die Aufnahme eines Gehirnschnittes in Kacheln liegt auch bereits eine Gebietszerlegung in gleichgroße Teilgebiete vor. Es ist also kein zusätzlicher Rechenaufwand für die Zerlegung notwendig und dieses Verfahren bietet sich damit an.

Mit der Wahl einer Gebietszerlegung muss nun entschieden werden, ob die Teilgebiete dynamisch, also während der parallelen Verarbeitung oder statisch zu Beginn auf die Prozesse verteilt werden. Eine dynamische Verteilung der Daten wird durch das **Master-Worker** Paradigma realisiert. Dabei übernimmt ein Prozess die Rolle des Masters und verteilt die gesamte Arbeit dynamisch auf die anderen Prozesse (Worker). Der Vorteil ist eine gleichmäßige Aufteilung der Last, wenn die Gesamtarbeit nicht bekannt oder die Verarbeitung der Teilgebiete unterschiedlich aufwändig ist. Bei der Kalibrierung ist jedoch die Menge der Daten bekannt und der Aufwand für die Verarbeitung einer Kachel ist für alle Kacheln gleich. Eine ständige Kommunikation zwischen dem Master-Prozess und den Worker-Prozessen ist also nicht nötig.

Eine geeignete Parallelisierungsstrategie für die Kalibrierung ist damit eine **statische Gebietszerlegung** eines Gehirnschnittes. Um die zu verarbeitenden Kacheln gleichmäßig auf die Prozesse zu verteilen, wurde in der Implementierung eine zyklische Aufteilung gewählt (Abb. 4.9).

Da auch die Multiplikation selbst keine Abhängigkeiten zwischen den Daten erzeugt, ist keine Kommunikation zwischen den Prozessen nötig. Der Ablauf der parallelen Anwendung der Gainfaktoren lässt sich demnach in folgende Schritte zusammenfassen:

1. Einlesen der Gainfaktoren: Jeder Prozess benötigt die selben Gainfaktoren zur Berechnung.
2. Anlegen der globalen HDF5-Datei für die kalibrierten Daten.
3. Zyklische Aufteilung der Kacheln auf die Prozesse.

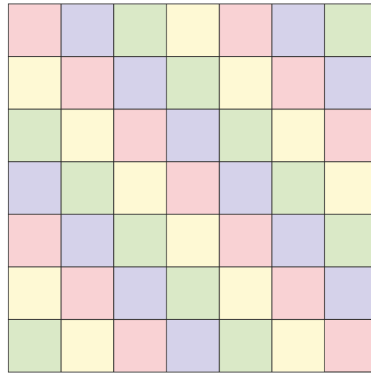


Abbildung 4.9.: Zyklische, zweidimensionale Gebietszerlegung auf vier Prozesse. In Bezug auf das 3D-PLI entspricht dies der Aufteilung eines Gehirnschnittes aus 7×7 Kacheln (ein Rotationswinkel) auf die Prozessoren.

4. Jeder Prozess verarbeitet kachelweise seinen Anteil; je Kachel:
 - 4.1. Einlesen der Rohdaten.
 - 4.2. Elementweise Multiplikation der Rohdaten mit den Gainfaktoren.
 - 4.3. Schreiben der Ergebnisse für die Kachel in eine gemeinsame HDF5-Datei.

Jeder Prozess liest je Kachel 18 Dateien ein (*task-local files*), also eine Bild-Datei für jeden Rotationswinkel. Für das Speichern der Ergebnisse wird parallel auf eine gemeinsame HDF5 Datei zugegriffen. Dies ist in der HDF5-API mit MPI-I/O realisiert. Die parallele Anwendung der Gainfaktoren wurde daher mit MPI implementiert und getestet.

4.2.2. Laufzeitverhalten

Das Laufzeitverhalten eines Programmdurchlaufs lässt sich im Wesentlichen in die folgenden drei Abschnitte aufteilen: das Einlesen der Daten, die Kalibrierung und das Speichern der Ergebnisse. Dabei werden nur einmal zu Beginn die Gainfaktoren eingelesen und die Aufteilung der Kacheln auf die Prozesse vorgenommen. Die iterative Arbeit für jeden Prozess besteht aus dem Einlesen der Rohdaten, der elementweisen Multiplikation mit den Gainfaktoren und dem Schreiben der Ergebnisse in eine HDF5-Datei.

Mit nur einer Rechenoperation, der Multiplikation, besteht der Großteil des Aufwands aus I/O-Operationen. Eine Auswertung der durchschnittlichen Laufzeiten in der Verarbeitung einer Kachel zeigt, dass unter 10% der Verarbeitungszeit einer Kachel für die Multiplikation benötigt wird (Abb. 4.10), jedoch rund 55% der Zeit für das Einlesen und 36% für das Schreiben nötig sind. Die Unterschiede für das Lesen und Schreiben lassen sich dadurch erklären, dass das Schreiben der Ergebnisse pro Kachel nur eine I/O-Operation ist, beim Lesen der Rohdaten einer Kachel jedoch 18 Bilder geöffnet und geschlossen werden müssen. Für diesen Vergleich wurden 200 Kacheln sequentiell verarbeitet, mit einer Auswertungszeit von etwa 1.1s pro Kachel.

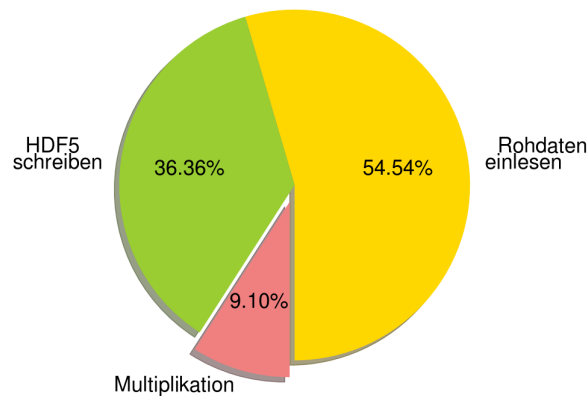


Abbildung 4.10.: Auswertung der Laufzeiten der Verarbeitungszeit einer Kachel. Die einzige Rechenoperation der Prozesse ist die elementweise Multiplikation von Rohdaten und Gainfaktoren. 90% der Laufzeit wird für I/O-Operationen benötigt.

Die Ergebnisse dieser Auswertung deuten auf ein mögliches *Bottleneck* für die Skalierung hin, da mit steigender Anzahl verwendeter Prozesse auch die Menge an Daten, welche parallel gelesen und geschrieben werden, steigt. Die Rohdaten einer Kachel und die kalibrierten Daten sind dreidimensionale Matrizen der Dimension (18, 2048, 2048) aus `int16` bzw. `float32` Werten. Konkret werden pro Kachel und damit pro Prozess 144 MB Rohdaten gelesen und 288 MB kalibrierte Daten geschrieben. Zusätzlich liegen nochmals 288MB Daten für die Gainfaktoren im Arbeitsspeicher. Insgesamt ist demnach eine volle Auslastung der I/O-Kapazität zu erwarten, bevor die maximale Auslastung der CPUs erreicht wird.

Anhand des Laufzeitvergleichs wird auch offensichtlich, dass eine weitere Parallelisierung der Berechnung einzelner Kacheln (beispielsweise pro Winkel) keine signifikante Beschleunigung in der Kalibrierung ergeben würde. Die Berechnungszeit pro Kachel liegt bereits im Bereich von 100ms. Eine potentiell schnellere Berechnung durch die weitere Aufteilung einer Kachel auf die Prozesse würde sich durch den zusätzlichen Kommunikations-Overhead (Aufteilung der Rohdaten und Sammeln der Ergebnisse) nicht lohnen.

Für eine ausführliche Analyse des Laufzeitverhaltens wurde mit einem Testdatensatz die Auswertung von 2000 Kacheln auf JURECA² simuliert.

Ergebnisse

Zunächst wurde das Skalierungsverhalten auf einem Knoten auf JURECA ausgewertet. Es wurden die Gesamtlaufzeit der Kalibrierung sowie die Laufzeit für das Einlesen der Rohdaten, die Berechnung und das Speichern der Ergebnisse gemessen. Dabei wurde die Auslastung des Knotens zwischen einem und 24 maximal verfügbaren Prozessen getestet.

² *Jülich Research on Exascale Cluster Architectures*, Details zur Konfiguration in Anhang A

Durch die große Menge an gleichzeitig eingelesenen und geschriebenen Daten wird bei steigender Anzahl verwendeter Prozesse die I/O- und Arbeitsspeicher-Kapazität stärker ausgelastet. Es ist also ein Abflachen des Speedups (sequentielle Auswertungszeit / parallele Auswertungszeit) zu erwarten. Die Messergebnisse bestätigen diese Vermutung: Bis etwa zwölf Prozesse ist eine Verbesserung in der Gesamtlaufzeit (Abb. 4.11a) und im Speedup (Abb. 4.11b) zu erkennen. Darüber hinaus bleibt die Laufzeit gleich und der Speedup flacht entsprechend ab.

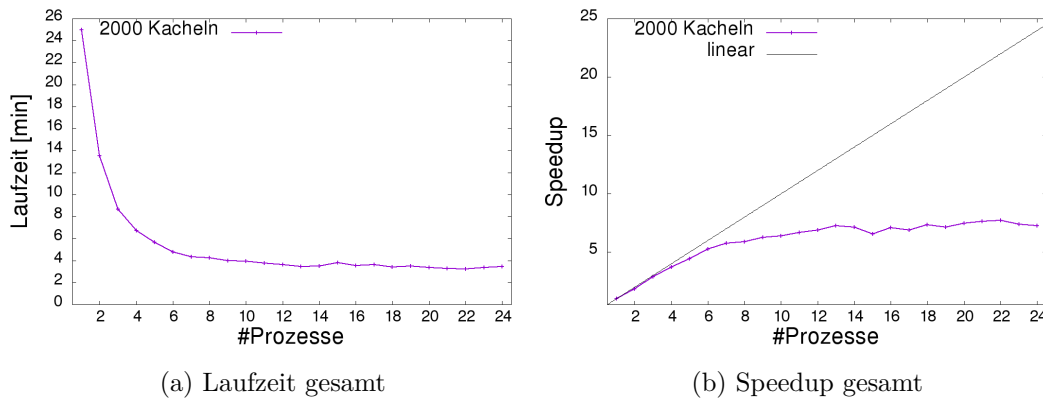
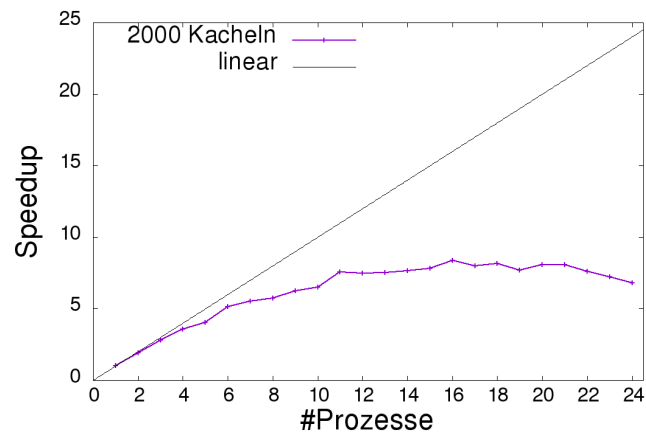


Abbildung 4.11.: Gemessene Laufzeit und Speedup der Auswertung von 2000 Kacheln auf einem Knoten. Ab etwa zwölf Prozessen ist kein bemerkbarer Speedup mehr vorhanden.

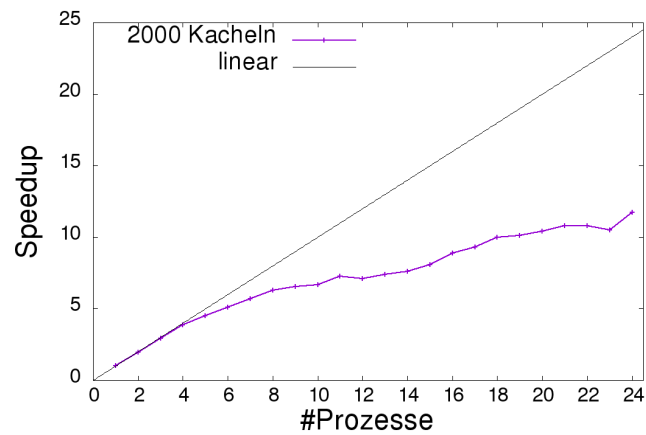
Für eine detaillierte Auswertung dieses Verhaltens wurden jeweils die Laufzeiten des Einlesens der Rohdaten, der Multiplikation mit den Gainfaktoren und des Schreibens der Ergebnisse in eine HDF5-Datei gemessen. Da die Multiplikation keine Datenabhängigkeiten erzeugt und keine Kommunikation benötigt, ist hier der Speedup nahezu linear (Abb. 4.12c). Bei den I/O-Operationen ist jedoch bei steigender Anzahl Prozesse die Auslastung des Knotens bemerkbar. Durch den Laufzeitanteil von über 90% haben sie auch den größten Einfluss auf die Gesamtlaufzeit. Der Einbruch im Speedup beim Einlesen der Rohdaten (Abb. 4.12a) ist zusätzlich stärker bemerkbar als beim Schreiben der Ergebnisse (Abb. 4.12b). Dies liegt hauptsächlich an der erwähnten Tatsache, dass pro Prozess beim Einlesen auf 18 Dateien (eine pro Rotationswinkel) zugegriffen wird, der Schreibvorgang aber nur eine I/O-Operation ist. Zusätzlich wird das Schreiben durch die optimierte HDF5 API für parallelen Zugriff (s. Kapitel 4.1.1) unterstützt.

Ein weiterer Test zur Laufzeitüberprüfung wurde auf mehreren Rechenknoten durchgeführt. Um die Auslastung des Arbeitsspeichers und der I/O-Kapazität ab etwa der Hälfte der verfügbaren Prozesse auf einem Knoten zu überprüfen, wurde bei steigender Anzahl an Knoten mit jeweils 6, 12 und 24 Prozessen (viertel, halbe und volle Auslastung) die Laufzeit gemessen.

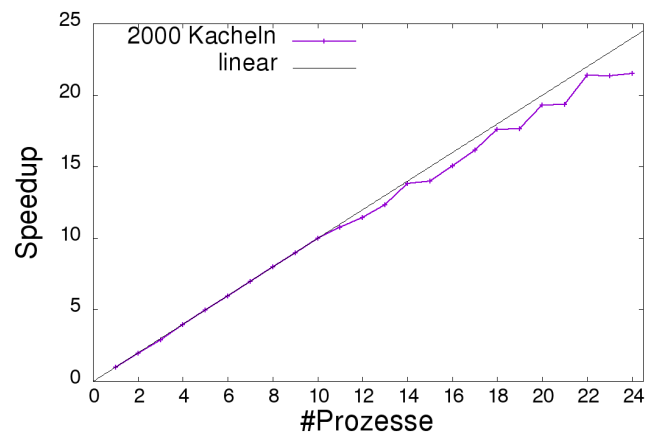
Die Ergebnisse (Abb. 4.13) zeigen, dass mit mehreren Knoten bei voller Auslastung ein nur sehr geringer Speedup im Vergleich zu halber (viertel) Auslastung vorhanden ist. Wird jedoch bei gleicher Anzahl insgesamt verwendeter Prozesse die Berechnung auf doppelt (viermal) so viele Knoten verteilt, wird damit ein besserer Speedup erzielt.



(a) Einlesen der Rohdaten



(b) Schreiben der Ergebnisse



(c) Berechnung (Multiplikation)

Abbildung 4.12.: Speedup der Auswertung von 2000 Kacheln. Das Einlesen der Rohdaten (a) und Schreiben der Ergebnisse (b) zeigt ein Abflachen ab etwa 12 Prozessen. Die Multiplikation (c) skaliert dagegen nahezu linear.

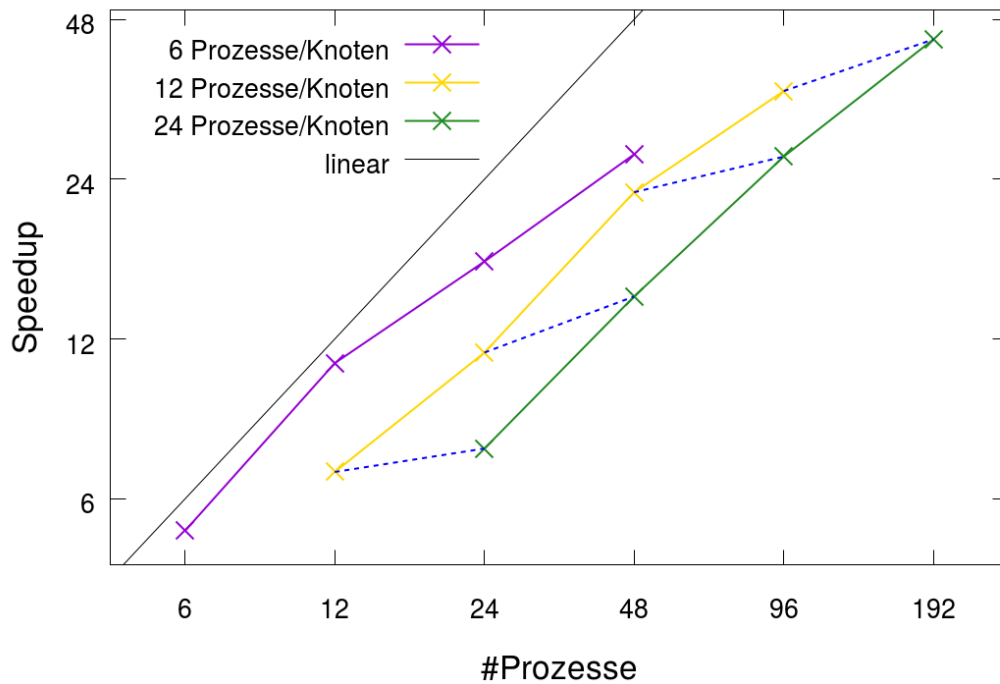


Abbildung 4.13.: Der Speedup der Auswertung von 2000 Kacheln auf mehreren Knoten. Es wurde auf 1, 2, 4 und 8 Knoten mit jeweils 6, 12 und 24 Prozessen/Knoten die Laufzeit gemessen. Dies entspricht der viertel, halben und vollen Auslastung der Knoten. Mit nur der halben (viertel) Auslastung pro Knoten ist ein deutlich besserer Speedup erreichbar, als bei voller Auslastung. Der theoretische Speedup zwischen der halben und vollen Auslastung eines Knotens (blaue Linien) bei gleicher Anzahl verwendeter Knoten ist jeweils nur minimal und zeigt damit das gleiche Verhalten wie die Messung auf einem Knoten.

5. Diskussion und Ausblick

Im Rahmen dieser Arbeit wurde ein Verfahren zur Kalibrierung von 3D-Polarized Light Imaging (3D-PLI) Messungen weiterentwickelt und implementiert. Das Ziel war es, die bisher verwendete Berechnung und Anwendung der Kalibrierungsdaten zu optimieren.

Beim 3D-PLI werden histologische Gehirnschnitte mit einem Mikroskop unter polarisiertem Licht aufgenommen. Mit diesen hochauflösenden Aufnahmen werden über verschiedene Methoden der Signal- und Bildverarbeitung Richtungsinformationen der Nervenfasern berechnet und so ein räumliches Modell des Gehirns, ein 3D-Hirnatlas, erstellt (siehe Kapitel 2). Dabei muss beachtet werden, dass die Aufnahmen durch verschiedene Faktoren beeinflusst werden. Als erster Schritt nach der Mikroskopie ist eine exakte Kalibrierung der Rohdaten damit für eine korrekte Rekonstruktion des aufgenommenen Gehirns sehr wichtig.

Der bisher verwendete Kalibrierungsalgorithmus berechnet aus Leermessungen sogenannte Gainfaktoren, welche mit den Rohdaten multipliziert und diese so auf einen Referenzwert normiert werden. Dabei sollen auftretende Unregelmäßigkeiten in den Rohbildern ausgeglichen werden (siehe Kapitel 3.1). Da die Aufnahme einer großen Menge an Leermessungen jedoch sehr zeit- und speicherintensiv ist, wurde zunächst mit Hilfe statistischer Methoden ein Verfahren zur Bestimmung der Anzahl tatsächlich benötigter Leermessungen entworfen. Unter Vorgabe einer Sicherheitswahrscheinlichkeit und eines Fehlers können aus diese Weise stabile Gainfaktoren mit der minimal nötigen Anzahl Leermessungen berechnet werden (siehe Kapitel 3.2). Bei einer Sicherheitswahrscheinlichkeit von 99.8% und einem Fehler von 1% reichen so beispielsweise 10 Leermessungen aus, anstatt der bisher verwendeten 100, was einer Sicherheitswahrscheinlichkeit von 99.7% und einem Fehler von 0.3% entsprechen würde. Damit auch sich über die Zeit verändernde Unregelmäßigkeiten berücksichtigt werden, wurden zusätzliche Verfahren entwickelt, um die Gainfaktoren regelmäßig zu aktualisieren und Veränderungen zu erkennen, ohne jedes Mal eine aufwändige Neukalibrierung vornehmen zu müssen. Dazu werden einzelne Leermessungen genutzt, welche vor jeder Messung eines Gehirnschnittes aufgenommen werden (siehe Kapitel 3.3). Mit diesen können Ausreißer zu vorherigen Messungen gefunden und eine kontinuierliche Aktualisierung der Gainfaktoren erreicht werden.

Im zweiten Schritt der Kalibrierung, der Anwendung der Gainfaktoren auf die Rohdaten, müssen diese geladen, verarbeitet und als kalibrierte Daten wieder abgespeichert werden. Die große Datenmenge erfordert dabei eine effiziente Speicherung und Berechnung. Um die Menge der erzeugten Daten zu minimieren und für einen einfachen Zugriff auf diese seitens des Benutzers, wird für die Kalibrierung, so wie auch in den darauf folgenden Schritten des Rekonstruktionsworkflows, das HDF5-Datenformat verwendet (siehe Kapitel 4.1).

Die Laufzeit der Kalibrierung wird maßgeblich durch das Einlesen der Rohdaten und das Abspeichern der Ergebnisse beeinflusst, die zusammen etwa 90% der Ge-

samtlaufzeit ausmachen. Die große Menge an Rohdaten eines Gehirnschnittes, aufgenommen in Kacheln, kann nicht auf einmal geladen und verarbeitet werden. Da bei der Berechnung jedoch keine Datenabhängigkeiten erzeugt werden und die HDF5 API einen parallelen Zugriff auf die gespeicherten Daten unterstützt, wurde die Anwendung der Gainfaktoren mit MPI parallelisiert (siehe Kapitel 4.2). Das entwickelte Programm wurde auf JURECA getestet und die Laufzeit gemessen. Es zeigte sich ein nahezu linearer Speedup bis etwa 12 Prozessen je Knoten. Danach machte sich die volle Auslastung des Arbeitsspeichers durch ein Abflachen des Speedups bemerkbar.

Zusammenfassend optimiert die entwickelte Software die Kalibrierung von 3D-PLI Rohdaten unter den Aspekten der Stabilität und regelmäßigen Aktualisierung. Es können mit dem neu implementierten Verfahren mit geringem Aufwand stabile Kalibrierungsdaten berechnet werden. Die Verwendung aktueller Leermessungen erlaubt die Erkennung von Veränderungen in den Aufnahmen und ermöglicht eine kontinuierliche Aktualisierung der Kalibrierungsdaten. Die Berechnung der kalibrierten Daten für die weitere Verarbeitung im Rekonstruktionsworkflow des 3D-PLI wurde parallelisiert und die kalibrierten Daten werden schließlich in HDF5 abgespeichert.

Der entwickelte Kalibrierungsworkflow bietet auch die Möglichkeit verschiedener Optimierungen und Erweiterungen. Neben der in Kapitel 3.5 angesprochenen einheitlichen Normierung auf einen immer gleichen Referenzwert und der unterschiedlichen Gewichtung verwendeter Einzelmessungen je nach Aktualität, können die einzelnen Schritte in der Berechnung und Anwendung der Kalibrierungsdaten unabhängig voneinander angepasst werden.

Die Analyse einer Einzelmessung auf globale und lokale Abweichungen zu vorherigen Leeraufnahmen kann erweitert werden, um weitere Überprüfungen vorzunehmen. Auch bietet sich hier die Einbindung in eine umfassendere Qualitätssicherung an, sodass auffällige Leermessungen früh erkannt werden und, falls nötig, der Gehirnschnitt neu gemessen oder eine Neukalibrierung vorgenommen werden kann.

Die Bestimmung der verwendeten Anzahl an Leermessungen für die Berechnung der Gainfaktoren setzt die Festlegung einer Sicherheitswahrscheinlichkeit voraus, mit welcher der erwartete Intensitätswert eines Bildpixels korrekt geschätzt wird. In der automatischen Aktualisierung mit Hilfe der Einzelmessungen, dem *Sliding Window Prinzip* (siehe Kapitel 3.3.2), kann eine zusätzliche Kontrolle eingeführt werden: Bei eigentlich n benötigten Leermessungen können beispielsweise die aktuellsten $n + 2$ Leermessungen verwendet werden und pixelweise zwei Ausreißer verworfen werden, also zum Beispiel der kleinste und größte Messwert für dieses Pixel. Damit wäre eine zusätzliche Sicherheit bei der Bestimmung des erwarteten Intensitätswertes eines jeden Bildpixels gegeben, bei einem, je nach gewähltem Verfahren, nur geringen Mehraufwand.

Zuletzt ist eine alternative Berechnung der Gainfaktoren möglich. Der verwendete Algorithmus stellt eine gängige Methode der Bildkalibrierung dar [15][16][17]. Eine Erweiterung des Verfahrens oder die Verwendung eines anderen Verfahrens könnten jedoch in Bezug auf das 3D-PLI möglicherweise bessere Ergebnisse liefern, was zu untersuchen wäre. Benötigt ein gewähltes alternatives Verfahren auch weiterhin mehrere Leermessungen, könnte dieses ohne weitere Änderungen im entwickelten Kalibrierungsworkflow eingesetzt werden.

Literatur

- [1] *Human Brain Project*. URL: <https://www.humanbrainproject.eu/> (besucht am 07/2017).
- [2] *BRAIN Initiative*. URL: <https://www.braininitiative.nih.gov/> (besucht am 07/2017).
- [3] premedhq. *Insulation of Axons: Myelin Sheaths and Schwann Cells*. URL: <https://www.premedhq.com/myelin-sheath-schwann-cells> (besucht am 11/2016).
- [4] A. C. Caseiro-Filho B. de Campos Vidal M. L. Silviera Mello und C. Godo. „Anisotropic properties of the myelin sheath.“ In: *Acta Histochemica* 66(1) (1980). DOI: 10.1016/S0065-1281(80)80079-1.
- [5] M. Axer; K. Amunts; D. Grässel; C. Palm; J. Dammers; H. Axer; U. Pietrzyk und K. Zilles. „A novel approach to the human connectome: Ultra-high resolution mapping of fiber tracts in the brain.“ In: *NeuroImage* 54(2) (2011). DOI: [dx.doi.org/10.1016/j.neuroimage.2010.08.075](https://doi.org/10.1016/j.neuroimage.2010.08.075).
- [6] O. Bücker. *Von der Kachel zum Volumen - (Nerven-)Fasern auf der Spur*. Jahresabschlusskolloquium am JSC, Forschungszentrum Jülich. 2015.
- [7] A. Westhoff. „GPU-accelerated Segmentation of high-resolution Human Brain Images acquired with Polarized Light Imaging“. Masterarbeit. Fachhochschule Aachen, 2013.
- [8] J. Reckfort; H. Wiese; U. Pietrzyk; K. Zilles; K. Amunts; M. Axer. „A multiscale approach for the reconstruction of the fiber architecture of the human brain based on 3D-PLI.“ In: *Frontiers in Neuroanatomy* 9 118 (2015). DOI: 10.3389/fnana.2015.00118.
- [9] C. Palm; M. Axer; D. Grässel; J. Dammers; J. Lindemeyer; K. Zilles; U. Pietrzyk; K. Amunts. „Towards ultra-high resolution fibre tract mapping of the human brain - registration of polarized light images and reorientation of fibre vectors.“ In: *Frontiers in Human Neuroscience* 4(9) (2010). DOI: 10.3389/neuro.09.009.2010.
- [10] Manfred Scholdt. *Temperaturbasierte Methoden zur Bestimmung der Lebensdauer und Stabilisierung von LEDs im System. Vol.4*. KIT Scientific, 2013.
- [11] Joint Committee for Guides in Metrology. *International vocabulary of metrology – Basic and general concepts and associated terms (VIM) - Definition: 2.39*. Publikation BIPM - <http://www.bipm.org/en/publications/guides/vim.html> (last visited June 2017). 2012.
- [12] J. Dammers; M. Axer; D. Gräßel; C. Palm; K. Zilles; K. Amunts; U. Pietrzyk. „Signal enhancement in polarized light imaging by means of independent component analysis“. In: *Neuroimage* 49.2, 1241–1248 (2010).

- [13] HDF Group. *HDF5 Support*. URL: <https://support.hdfgroup.org/HDF5/> (besucht am 07/2017).
- [14] Michael J Flynn. „Some computer organizations and their effectiveness“. In: *IEEE transactions on computers* 100.9 (1972), S. 948–960.
- [15] J. M. ZWIER u. a. „Image calibration in fluorescence microscopy“. In: *Journal of Microscopy* 216.1 (2004), S. 15–24. DOI: 10.1111/j.0022-2720.2004.01390.x.
- [16] Georg Emrich. *Flatfield Korrektur für CCD Aufnahmen*. URL: <https://www.astronomie.at/tips/flatfield.htm> (besucht am 07/2017).
- [17] Dean Jacobsen. *CCD Image Calibration*. URL: <http://www.astrophoto.net/calibration.php> (besucht am 07/2017).

A. Konfiguration JURECA



Hardware Eigenschaften JURECA¹:

- 1872 Rechenknoten
 - Pro Knoten: Zwei Intel Xeon E5-2680 v3 Haswell 12-Kern Prozessoren mit 2,50 GHz
 - 75 Knoten mit zwei NVIDIA K80 GPUs (2×4992 CUDA Kerne, 2×24 GiB GDDR5 Speicher)
 - DDR4-SDRAM (2133 MHz) Technologie:
 - 1605 Knoten mit 128 GiB Speicher
 - 128 Knoten mit 256 GiB Speicher
 - 64 Knoten mit 512 GiB Speicher
- 12 Visualisierungsknoten
 - Zwei Intel Xeon E5-2680 v3 Haswell CPUs pro Knoten
 - Zwei NVIDIA K40 GPUs pro Knoten
 - 10 Knoten mit 512 GiB Speicher
 - 2 Knoten mit 1024 GiB Speicher
- Loginknoten mit 256GiB Speicher pro Knoten
- 45216 CPU Kerne
- 100 GiB pro Sekunde Speicheranbindung

¹http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JURECA/JURECA_node.html,
Zuletzt besucht am 10. August 2017