UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR ROBOTIK
UND KOGNITIVE SYSTEME

Tissue Thickness Estimation from Backscattered Light

A Novel Concept for Optical Head Tracking in Radiotherapy

Tobias Wissel

Dissertation

Universität zu Lübeck
Institut für Robotik und Kognitive Systeme

From the Institute of Robotics and Cognitive Systems
of the University of Lübeck
Director: Prof. Dr.-Ing. Achim Schweikard

Tissue Thickness Estimation from Backscattered Light

A Novel Concept for Optical Head Tracking in Radiotherapy

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

from the Department of Computer Sciences

Submitted by

Tobias Wissel
from Lutherstadt Wittenberg (Germany)

Lübeck 2015

Tobias Wissel – PhD student
Institute for Robotics and Cognitive Systems
Universität zu Lübeck
Ratzeburger Allee 160
23562 Lübeck, Germany

Board of examiners

First referee:	Prof. Dr.-Ing. Achim Schweikard
Second referee:	Prof. Dr.-Ing. Erhardt Barth
Chairman:	PD Dr. rer. nat. Gereon Hüttmann

Date of oral examination: 22.03.2016

Approved for printing. Lübeck, 31.10.2016

Zusammenfassung

In den meisten Industrienationen stellt Krebs heute nach Herz-Kreislauf-Erkrankungen die zweithäufigste Todesursache dar. Trotz erheblichen Fortschritten in Diagnose und Therapie liegt die Überlebenswahrscheinlichkeit bei kritischen Tumorarten, wie solchen im zentralen Nervensystem, immer noch unter 30 %. Für ältere Patienten fällt diese Wahrscheinlichkeit sogar unter 20 %.
In vielen Fällen ist die Strahlentherapie aufgrund von hohen Proliferationsraten und dem Risiko für Metastasen die einzige oder zumindest eine unterstützende Behandlungsmöglichkeit. Hierbei hält die sogenannte externe Strahlentherapie das Tumorwachstum mit Hilfe von mehreren sich überlagernden Strahlen hochenergetischer Röntgenstrahlung unter Kontrolle. Innerhalb eines Fraktionierungsschemas wird der Patient zunächst einer Behandlungsplanung unterzogen. Danach wird der Tumor gemäß dem Plan an mehreren Tagen pro Woche, über typischerweise 6-7 Wochen bestrahlt. Dabei muss die Patientenposition relativ zum Iso-Zentrum des Therapiegerätes zwischen Planung und Therapie genau übereinstimmen. Aufgrund von Patientenbewegung während der Bestrahlung und Repositionierungsungenauigkeiten über Sitzungen hinweg, fügen Strahlentherapeuten Sicherheitsbereiche um das Planungsvolumen herum ein. Dies stellt sicher, dass das Ziel ausreichend Dosis erhält. Um diese Bereiche jedoch so klein wie möglich zu halten und gesundes Gewebe zu schonen, ist eine sehr präzise Lokalisierung des Tumors unerlässlich für die Strahlentherapie.
Höchste Genauigkeit ist insbesondere in der kraniellen Strahlentherapie vonnöten. Diese wird entweder mit einer Immobilisierung des Patienten, bildgestützter Strahlentherapie oder einer Kombination aus beidem erreicht. Thermoplastische Masken, welche typischerweise zur Immobilisierung benutzt werden, unterliegen im Durchschnitt intra-fraktionellen Ungenauigkeiten von weniger als 1 mm und inter-fraktionellen von weniger als 3 mm. Die Masken sind unbequem und werden durchaus nicht von allen Patienten toleriert. Darüber hinaus können die Ungenauigkeiten der Immobilisierung erheblich größer sein, wenn der Patient im Laufe der Behandlung unter Gewichtsverlust leidet.
Stereoskopische Röntgenbildgebung sowie Cone Beam Computertomografie (CBCT)

gelten als Gold-Standards für die bildgestützte Strahlentherapie. Durch sie wird mittels hochgenauer Volumenregistrierung eine Kontrolle der Patienten-Vorausrichtung zu Beginn jeder Sitzung ermöglicht. Die Nutzung zur Online-Überwachung ist jedoch stark eingeschränkt. Diese Bildgebung bringt zusätzliche Strahlenbelastung mit sich und erfordert lange Aufnahmezeiten. Im Gegensatz zur Röntgenbildgebung umfasst die optische Bildgebung nichtinvasive und eher günstige Verfahren, welche Licht nutzen und schnelle Aufnahmegeschwindigkeiten erlauben. Dabei stellt die markierungslose Bewegungsverfolgung am Kopf die aussichtsreichste Variante dessen dar. Anders als Verfahren, die auf Markierungen beruhen, zielt diese Variante direkt darauf ab, die knöcherne Anatomie in ihrer Bewegung zu verfolgen. Das eigentliche Tumorziel wird dabei als starr mit dem Schädelknochen verbunden angenommen. Zwei kommerzielle Geräte sind hierzu zurzeit auf dem Markt verfügbar: die C-Rad Systeme (C-Rad, Uppsala, Schweden) und das AlignRT® System (VisionRT, London, GB). Durch Oberflächentriangulierung erreichen sie Lokalisierungsgenauigkeiten von unter 3-4 mm und 2°. Oberflächendeformationen und die Nutzung von sogenannten „open face“ Masken können allerdings zu deutlich höheren Fehlern führen.

Ein zentrales Problem der Oberflächenverfolgung bleibt deren Anfälligkeit gegenüber Oberflächenveränderungen. Das gilt insbesondere, wenn stabile und hervorstechende Landmarken fehlen. Idealerweise strebt die kranielle Strahlentherapie nach Sub-Millimeter-genauer Lokalisierung und stellt daher höchste Ansprüche an die Qualität dieser Landmarken. Aktuelle Studien deuten an, dass unter normalen Bedingungen bestimmte Gesichtsregionen, wie die Stirn, stabilere Oberflächengeometrien aufweisen und weniger anfällig gegenüber Gesichtsbewegungen sind. Dennoch kann es innerhalb des Registrierungsproblems zu Mehrdeutigkeiten, d.h. lokalen Minima der Optimierungsfunktion, kommen, wenn ein abgetasteter Oberflächenbereich zu einer Referenz registriert werden soll. Abhängig von Kopfbewegung, anatomischen Oberflächenformen oder der Deformation der Weichgewebeoberfläche kann das erwähnte Risiko sogar noch größer sein. Dies bringt zum Teil niedrige Robustheit und Ausreißer mit sich, welche insbesondere für Ziele kritisch sind, die sich weit von der Oberfläche entfernt befinden.
Das Problem der markierungslosen Bewegungsverfolgung soll in dieser Arbeit behandelt werden. Dafür soll zusätzliche Information ausgenutzt werden, mit der die Oberfläche charakterisiert werden kann. Innerhalb eines interdisziplinären Teams wurde dazu ein funktioneller Prototyp zur Oberflächenabtastung entwickelt. Er enthält einen 850 nm faser-gekoppelten Laser, welcher über eine galvanometrische Spiegeleinheit auf das Ziel gelenkt wird. Während eine Triangulationskamera zur Rekonstruktion

der 3D-Oberfläche genutzt wird, wird zusätzlich eine weitere Kamera mit hohem Dynamikbereich in den Pfad des Strahls eingekoppelt. Diese detektiert Änderungen im optischen Rückstreuverhalten von Reflexion zu Reflexion. Die Hypothese dieser Arbeit ist nun, dass die Gewebedicke an der Stirn aus der optischen Rückstreuung des Gewebes vorhergesagt werden kann. Darüber hinaus – so die zweite Hypothese – ergeben sich so Hautdickenmuster auf der Oberfläche, welche zu einer verbesserten und stabileren Oberflächenregistrierung führen.

In einem ersten Schritt wurden Monte-Carlo-Simulationen genutzt, um Licht-Gewebe-Interaktionen in einem sieben-schichtigen Hautmodell zu untersuchen. Die Simulationen dienten zur Identifikation optimaler Bedingungen, unter denen ein Maximum an Information über die Gewebedicke gewonnen werden kann. Dies wiederum lässt Rückschlüsse auf die entsprechend nötige Parametrisierung des funktionalen Prototyps zu. Im Einklang mit einschlägiger Literatur wies bei Untersuchungen für Licht im Bereich von 400 nm bis 1000 nm vor allem Licht im nahen Infrarot (NIR) Bereich eine hohe Eindringtiefe ins Gewebe auf. Dabei nimmt die relative Anzahl der tief eindringenden Photonen mit der Distanz zum Zentrum des Laserpunktes zu. Für Licht, das bereits in den oberen Hautschichten reflektiert wird, ist es wahrscheinlich, dass es das Gewebe nahe seinem ursprünglichen Eintrittspunkt wieder verlässt. Als am vielversprechendsten sind daher Reflexionen in mittlerer Distanz zum Laserpunkt anzusehen. Aus späteren Experimenten konnte ein günstiges Rückstreuinterval von 2,5 mm bis etwa 7,6 mm Abstand zum Laserpunktmaximum identifiziert werden. Dieses enthält zum Einen weniger Rückstreuung aus den oberen Hautschichten und dortige Lichtintensitäten können auf der anderen Seite noch mit einem ausreichend hohen Signal-Rausch-Abstand durch die 14 bit Kamera aufgelöst werden. Das Kamerabild wurde folglich in sieben konzentrische Ringe um das Laserpunktzentrum unterteilt. Diese dienen als sogenannte „regions-of-interest" (ROIs), in denen die Pixelintensitäten zu dann sieben Merkmalswerten akkumuliert werden. Die Anzahl der Regionen wurde zugunsten höherer Merkmalsstabilität und besserem Signal-Rausch Abstand in späteren Experimenten auf fünf reduziert.
Simulationen für Hautdicken zwischen 2,1 mm und 7,1 mm ergaben, dass die Merkmale leicht verschiedene und nichtlineare funktionale Zusammenhänge zur Änderung der Hautdicke ausweisen. Damit stellt die Hautdickenschätzung ein multivariates, überwachtes Problem des statistischen Lernens dar. Stützvektorregression (SVR) wurde genutzt, um dieses Problem zu modellieren, und war schließlich in der Lage, die Hautdicke der Simulationsmodelle mit einem mittleren quadratischen Fehler (RMSE)

von 16 µm vorherzusagen. Hierbei wurden Winkelabweichungen des Laserstrahls von orthogonaler Einstrahlung als einflussreiche Störgröße identifiziert. Da beide, der Winkel und die Hautdicke, negativ mit den Änderungen in der Rückstreuung korrelieren, verschlechtert sich der Vorhersagefehler, wenn veränderliche Einfallswinkel zugelassen werden. In einem solchen Szenario werden die besten Ergebnisse erreicht, wenn der Merkmalsraum um ein Maß für den Einfallswinkel erweitert wird (Raumkennzeichnung *A*).
Die Simulationsergebnisse wurden in einer Studie mit 30 Freiwilligen validiert und erweitert (14 männlich, 16 weiblich, Hauttypen II bis V, Alter 24-65). Von jedem Probanden wurden drei NIR-Abtastungen und eine hochauflösende Magnetresonanzaufnahme (MRT) akquiriert. Letztere ergab nach erfolgter semi-automatischer Segmentierung der Haut eine Grundwahrheit für die Hautdicke, welche nur 0,2 mm von einer manuellen Segmentierung durch geschulte Experten abwich. Um ein statistisches Modell zu lernen, wurden die Oberflächen von NIR- und MRT-Aufnahme mit Hilfe eines Beißschienenmarkers aufeinander registriert. Die zuvor erwähnten Korrelations- und Störgrößeneinflüsse konnten auch praktisch mit den resultierenden Merkmalsräumen bestätigt werden. Ein Lernverfahren analog zu den Simulationen verzeichnete Vorhersagefehler von im Mittel 0,2 mm. Dabei wurde die SVR durch Gaußprozesse (GPs) mit isotropem Matérn Kernel ersetzt, da diese als probabilistische Methode, neben einem besseren Vorhersagefehler, weitere günstige Eigenschaften aufweisen. Vorkenntnisse über das Rückstreuverhalten in der unmittelbaren Nachbarschaft eines Laserpunktes senkten den Vorhersagefehler im Mittel bis auf 0,12 mm (Raumkennzeichnung *NBH*).

Schließlich wurden die Auswirkungen auf das eigentliche Registrierproblem untersucht. Dazu wurden nach 5.000 zufällig generierten Kopfbewegungen zwischen NIR-Abtastungsoberfläche und MRT-Referenz beide Oberflächen erneut mit Hilfe einer iterativen Bestimmung nächster Nachbarn (iterative closest points, ICP) registriert. Eine, der dazu verwendeten ICP-Varianten, bestimmte dabei räumliche Korrespondenzen mit Unterstützung von Hautdickenmustern auf der Oberfläche. Sogar bei kleineren Bewegungen (< 10 mm Translation, $< 10°$ Rotation) wiesen beide Alternativen lokale Minima auf, in die der iterative Algorithmus konvergierte. Bei der Nutzung von Hautdickenmustern waren diese Minima auf ein schmaleres Intervall verteilt. Auch die absoluten Registrierfehler waren kleiner. Dabei wurden die Hautdickenmuster aus verschieden Quellen erzeugt: Die MRT-Grundwahrheit verbesserte den mittleren Fehler gegenüber einfacher Oberflächenregistrierung um einen Faktor von 29. Im Vergleich dazu erreichten die Rekonstruktion aus dem Merkmalsraum *A* einen Faktor von 5,6 und

die aus dem Merkmalsraum *NBH* einen Faktor von 7. Durchschnittliche Registrierfehler pro Proband waren stets kleiner als 1 mm. Dies lässt den Schluss zu, dass die bisherigen Vorhersagegenauigkeiten für die Hautdicke den Anforderungen genügen.
Zusammenfassend ergeben die Ergebnisse dieser Arbeit eine aussichtsreiche Bestätigung des vorgeschlagenen Konzeptes für ein verbessertes Verfahren zur Bewegungsverfolgung am Kopf. Es ist möglich, mit ausreichender Genauigkeit Hautdickeninformationen aus dem veränderlichen NIR-Rückstreuverhalten zu extrahieren. Mit dem vorgeschlagenen Ansatz wird eine robustere Oberflächenregistrierung erreicht. Zudem wird Ausreißern durch die Reduktion von räumlichen Mehrdeutigkeiten vorgebeugt. Dies wurde sogar dann beobachtet, wenn das abgetastete Gebiet klein und die Anzahl der verfügbaren Punkte eingeschränkt war. Insgesamt etabliert diese Arbeit damit eine vielversprechende Grundlage für weitergehende, größer angelegte klinische Validierungen.

Abstract

Today, cancer is ranked the second most common cause of death after cardiovascular diseases in many industrialized nations. Despite substantial advances in diagnosis and therapy, survival rates for critical tumors such as those in the central nervous system still fall below 30 %. For elderly patients this rate is even less than 20 %.
In many cases, high proliferation rates and the risk for multiple metastases make radiotherapy the definitive or at least a supportive option for treatment. Here, the external beam therapy controls or shrinks the tumor by applying superimposed beams of high energy X-rays. Within the scheme of fractionation a patient then undergoes treatment planning, before he or she is treated on several days a week for typically 6-7 weeks in a row. During each treatment session the patient position with respect to the machine isocenter has to align with the treatment plan. Due to intrafractional motion and interfractional repositioning errors, clinicians introduce extra safety margins into the treatment plan. This is to ensure that sufficient dose is still delivered to the target. To keep these margins as small as possible and to spare healthy tissue, precise localization of the tumor target is vital for radiotherapy.
High precision is particularly required in cranial radiotherapy. This is achieved with patient immobilization, image guidance or a combination thereof. Thermoplastic masks used to immobilize the patient's head typically entail intrafractional errors of on average less than 1 mm and interfractional errors of less than 3 mm. The masks are inconvenient and are also not tolerated by all patients. Moreover, immobilization errors may be significantly higher when patients experience weight loss in the course of the treatment.
The gold standard for image guidance is given by stereoscopic X-ray or cone beam computed tomography (CBCT). Thus, highly accurate volume registration can be used for pre-alignment checks at the beginning of each session. Its usage for online motion monitoring, however, is very limited. It entails additional exposure to radiation and typically suffers from slow imaging speed. In contrast to image guidance using X-rays, optical imaging is a non-invasive and rather inexpensive modality that uses light and provides fast imaging speeds. Here, marker-less optical head tracking constitutes the most promising variant. In contrast to marker-based alternatives, marker-less tracking

aims at tracking the bony anatomy of the patient, because tumor targets are very often rigidly linked to the skull bone. Two commercial options, namely the C-Rad systems (C-Rad, Uppsala, Sweden) and AlignRT® (VisionRT, London, UK), are currently available. By triangulating the skin surface, they achieve average localization accuracies of less than 3-4 mm and 2°. Surface deformation and the usage of open-face mask systems, however, may lead to substantially higher errors.

A core issue for surface tracking remains its sensitivity to changes in the surface, particularly when lacking stable and prominent landmarks. Ideally, radiotherapy aims at sub-millimeter localization accuracy and therefore places high demands on such landmarks. Recent studies gave evidence that under typical conditions certain regions of the face such as the forehead provide more stable surface geometries and are less prone to facial motion than other regions. In any of these cases, the registration problem suffers from ambiguities, i.e. local optimization minima, when matching a scanned surface patch to a reference. This risk can be even higher depending on the extent of head motion, anatomical surface shapes or deformation of the elastic soft tissue surface. This can lead to low robustness and outliers, which are particularly delicate for targets far away from the registration site.
To tackle this problem in marker-less optical tracking, this work proposes to exploit additional information with which the surface can be labeled. In synergy with interdisciplinary coworkers a functional optical scanning prototype has been developed. It consists of an 850 nm fiber-coupled laser beam which is deflected by a galvanometric mirror unit onto the target. While a triangulation camera is used to triangulate the surface geometry, another, high dynamic range camera is coupled into the beam path to detect variations in optical backscatter from spot to spot. Now, the hypothesis of this work is that, first of all, tissue thickness at the forehead can be predicted from the characteristics of optical backscatter on the skin. It is further proposed that this tissue thickness compiles patterns across the forehead surface which have a supporting effect on the registration performance.

In a first step Monte-Carlo simulations were employed to simulate the light transport in a seven layer tissue model. These were used to identify optimal conditions under which most information can be retrieved from optical backscatter and to investigate how parameters of the prototype should be defined. In the investigated spectral range between 400 nm and 1000 nm, particularly near-infrared (NIR) light was found to deeply penetrate the tissue. This is in agreement with the literature. It was found that the

relative proportion of deeply penetrating photons increases with the distance from the spot center. Light reflected from upper tissue layers is more likely to leave the tissue close to the location of incidence. Therefore, reflections at medium range from the spot center are considered most promising. In later experiments a favorable backscatter interval from approximately 2.5 mm to 7.6 mm distance from the spot center was identified. Light reflected from this interval contains less backscatter from upper layers, but can be still resolved with sufficiently high signal-to-noise ratio (SNR) by the 14 bit camera. As a consequence, the camera image was divided into seven concentric rings around the spot center. They serve as regions-of-interest (ROIs) in which the pixel intensity is accumulated to finally form seven feature values characterizing each spot. The number of ROIs was reduced to five in later experiments due to a higher feature stability and better SNR ratio.

For simulated tissue thicknesses between 2.1 mm and 7.1 mm, each feature was shown to have a slightly different and nonlinear functional relationship to the thickness. As a conclusion, the setting resembles a multivariate supervised statistical learning problem that maps features to the thickness target label. Support Vector regression (SVR) was used to model this problem and was capable of predicting the thickness of the simulation model with a root mean square error (RMSE) of 16 µm. The angular deviation of the laser beam from orthogonality on the skin surface was identified as a strong confounding factor. Since both, the angle and the tissue thickness, correlate negatively with the backscatter changes, the prediction accuracy worsens when admitting varying incident angles. Best results in this scenario were achieved by extending the feature space by the incident angle (space label *A*).

The results were experimentally validated in a study with 30 volunteers (14 male, 16 female, skin types II to V, aged 24-65). From each subject, three NIR and one high resolution magnetic resonance (MR) scan were obtained. After semi-automatic segmentation, the latter provided a tissue thickness ground truth which deviated less than 0.2 mm from manual expert delineation. For learning a statistical model, NIR and MR data were matched based on a bite marker. The aforementioned correlation and confounding effects were confirmed with the resulting feature spaces. A learning procedure equivalent to the simulations then yielded tissue thickness prediction errors of on average 0.2 mm. Here, SVR was replaced by Gaussian processes (GPs) with an isotropic Matérn kernel due to its better performance and further beneficial properties of the probabilistic framework. Prior knowledge about backscatter behavior from the spatial neighborhood decreased the prediction error down to 0.12 mm on average (space label *NBH*). In terms of prediction accuracy, the study did not indicate any sig-

nificant differences for volunteers of different skin type according to the Fitzpatrick scale.

Finally, the benefits for the registration problem were investigated. After 5,000 random movements starting from an initial, marker-based matching, the iterative closest point (ICP) algorithm was used to re-register the NIR scans to the MR reference. Unlike standard ICP, the proposed registration approach supported the identification of spatial correspondences between surfaces by the thickness patterns. Even for little motion ($< 10\,\text{mm}$ translation, $< 10^\circ$ rotation) both alternatives exhibited local minima into which the iterative algorithm converged. When using tissue thickness, these different minima were spread across a smaller interval. The absolute registration errors were smaller. Tissue thickness was used from different sources: The MR ground truth improved the mean error of pure surface registration by a factor of 29. Compared to that, the tissue reconstruction from space *A* achieved a factor of 5.6 and that of space *NBH* a factor of 7. Average registration errors were below 1 mm for each subject. This suggests that the achieved prediction accuracies meet the requirements.
All in all, the results of this work provide a promising proof of concept for the enhanced tracking approach. Additional information about the tissue thickness can be obtained from NIR optical backscatter with sufficient accuracy. The proposed approach yields more robust registration results and reduces outliers by avoiding spatial ambiguities. This was even observed when the scanned area and the number of points was limited. Overall, this work established a promising basis for larger scale clinical validation.

Contents

Mathematical Notation and Indices

Symbol	Description
A_{ph}	absorption array of photon energy in tissue (see subsubsection 3.2.2.2)
b_j	the j-th feature dimension in a D-dimensional feature space
b_i	the i-th feature dimension in a D-dimensional feature space
$\boldsymbol{B}$	design matrix within the feature space (see chapter 2)
C	regularization constant balancing between complexity and goodness of fit (see subsubsection 2.2.1.1)
d_{si}	tissue thickness and target label at spot i(see chapter 2)
D	dimension of input features
D'	dimension of the features transformed into high-dimensional kernel space
$\mathcal{D}$	labeled data with $\mathcal{D} = \{\boldsymbol{B}, \boldsymbol{d}_s\}$
$f(\cdot)$	functional mapping from feature or kernel space to the target labels (see chapter 2)
F	polynomial degree in polynomial functions (see subsubsection 2.2.2.4)
g	anisotropy factor for scattering media (see subsection 3.2.1)
$k(\cdot,\cdot)$	kernel or covariance function (see GP or SVR)
$\mathbf{K}$	covariance or Gram matrix (see GP models)
$L_*(\cdot)$	loss function or empirical risk (see Table 2.1)
$\mathcal{L}$	Lagrange function of an optimization problem
$\mathbf{I}$	identity matrix
$m_{GP}(\cdot)$	mean function (see GP)
n	iid noise superposing measured data and being drawn from a specific distribution (see chapter 2)
n_{fld}	number of folds in a CV testing scheme
n_{rep}	number of repetitions in a CV testing scheme
N,N_p,N_q,N_{fld}	number of data samples, subscripts p, q and fld may denote an

	affiliation to a specific set
$p(x)$	probability of x
$p(x\|y)$	probability of x given y
P_{cld}, Q_{cld}	point clouds being finite sets of vectors in $\mathbb{R}^{3\times1}$, i.e. points in 3D space
$q(\cdot)$	approximated probability distribution
r	Euclidean difference between two feature vectors $\|\mathbf{b}-\mathbf{b}'\|$ (see subsubsection 2.2.2.4)
$\boldsymbol{r}$	triplet of angles representing rotations around the three coordinate axes $[r_x, r_y, r_z]^T$ (Tait Bryan angles in yaw-pitch-roll convention)
$\boldsymbol{R}$	orthogonal rotation matrix in 3D space with determinant 1
R_D	diffuse reflection (see subsection 3.2.2)
R_D^k	diffuse reflection restricted to photons which reached at most tissue layer k (see subsubsection 3.2.2.2)
R_{sp}	specular reflection (see subsubsection 3.2.2.2)
s_c	multiplicative scaling factor for kernel functions (see subsubsection 2.2.2.4)
s_{ph}	current photon step size (see subsubsection 3.2.2.2)
$\boldsymbol{t}$	a translational offset with scalar elements $[t_x, t_y, t_z]^T$
T_t	total transmittance (see subsection 3.2.2)
${}^Q\mathcal{T}_P$	homogenous transformation matrix from coordinate space Q into P
$\boldsymbol{w}$	slope or weight vector of the general regression model (see Equation 2.4)
$\boldsymbol{u}$	inducing variables (see subsubsection 2.2.2.7)
w_0	explicit linear offset in the SVR model
α	incident angle of the laser beam (deviation from orthogonality
α_i^*	Lagrange multipliers for the loss constraints of the SVR model (see subsubsection 2.2.1.2)
β_0	constant offset for linear kernel functions (see subsubsection 2.2.2.4)
$\delta(k,j)$	Kronecker-Delta ($\delta(k,j)=1$ for $k=j$, $\delta(k,j)=0$ for $k\neq j$)
η_i^*	Lagrange multipliers for the $\xi_i^{(*)} \geq 0$ of the SVR model (see subsubsection 2.2.1.2)
γ	length scale kernel parameter (see subsubsection 2.2.2.4)
$\boldsymbol{\varphi}$	functional mapping from feature to kernel space (see chapter 2)
$\boldsymbol{\Phi}$	design matrix in kernel space (see chapter 2)
ε	parameter of the ε-insensitive loss function: width of ε-tube used in SVR
μ_a	absorption coefficient (see subsection 3.2.1)

μ_s	scattering coefficient (see subsection 3.2.1)
μ_t	total attenuation coefficient being the sum of μ_a and μ_s (see subsubsection 3.2.2.2)
ψ_{ph}	azimuth angle after a scattering event (see subsubsection 3.2.2.2)
$\mathfrak{P}$	penalty term for regression analysis (see Table 2.1)
$\mathfrak{R}$	structural risk (see subsubsection 2.2.1.1)
σ	standard deviation
σ^2	variance
Σ_w	covariance matrix of the weights $\boldsymbol{w}$ in Bayesian regression
θ_{ph}	altitude or deflection angle after a scattering event (see subsubsection 3.2.2.2)
$\xi_i^{(*)}$	collective term for slack variable in SVR
ζ	uniformly distributed random variable
Γ	Gamma distribution
$\mathcal{N}$	normal distribution
$\mathcal{U}$	uniform distribution

Superscript indices

$\bar{x}$	mean of variable x
$\boldsymbol{A}^{-1}$	the inverse of matrix $\boldsymbol{A}$
$\boldsymbol{A}^{T}$	the transpose of matrix $\boldsymbol{A}$
${}^{Q}\boldsymbol{p}$	a 3D vector or point that resides in coordinate space Q

Subscript indices

$d_{s\star}$	target label prediction (see chapter 2)
$\boldsymbol{b}_\star$	input features from unseen data (see chapter 2)
$k_{SoR}(\cdot,\cdot)$	SoR kernel approximation (see subsubsection 2.2.2.7)
$\mathbf{Q}_{SoR}$	Gram matrix of the SoR approximation (see subsubsection 2.2.2.7)
$\boldsymbol{m}_{ph}$	mean position of the beam profile distribution (see subsubsection 3.2.2.2)
$r_{x,y}$	Pearson's correlation coefficient between x and y
w_{ph}	energy weight of a photon packet in MCML (see subsubsection 3.2.2.2)
$ux_{ph}, uy_{ph}, uz_{ph}$	directional cosines of photon moving direction

	(see subsubsection 3.2.2.2)
x_{ph}, y_{ph}, z_{ph}	current photon location in the tissue (see subsubsection 3.2.2.2)
$\sigma_\star^2$	target label variance predicted by a probabilistic algorithm (see GP models)
σ_n^2	variance deviation of Gaussian noise
$\mathbf{\Sigma}_{ph}$	variance around the mean position of the beam profile distribution (see subsubsection 3.2.2.2)

Operators

$\langle \cdot, \cdot \rangle$	inner product of two vectors
$\{\cdot\}$	set of elements
$\lvert\cdot\rvert$	absolute value of elements
$tr(\cdot)$	the trace of a matrix
$diag(\cdot)$	the diagonal of a matrix
$STD(\cdot)$	standard deviation of a variable

1 Introduction

This first chapter provides a general introduction to the problem of marker-less optical head tracking in radiation therapy (RT). It highlights the relevance and special need for highly precise treatment of cancer in the head region. Therefore, stereotactic radiation therapy (SRT) of the head region is first put into the broader context of other treatment options and particularly RT (cf. section 1.1). Second, a typical workflow for SRT is described in detail within section 1.2. This subsection also elaborates on immobilization and motion compensation during treatment and motivates the role of marker-less optical head tracking.
Based on weaknesses of state-of-the-art systems, section 1.3 discusses the purpose of this work. After pointing out the main problems of the existing approaches, a new concept for marker-less optical head tracking will be proposed which aims at tackling these weaknesses. The feasibility of this concept will be investigated in terms of the main research questions presented in this subsection. Finally, section 1.4 outlines the organization of the following chapters.

1.1 Cancer in the Head Region - Relevance and Treatment Options

Worldwide, 8,201,575 people died as a results of cancer in 2012 [86]. According to the International Agency for Research on Cancer of the World Health Organisation (WHO), the same statistics reveal 14,067,894 cases of newly diagnosed cancer in the same year. The highest rates per 100,000 people occur in more developed regions of the world such as Australia, North America or Europe. Statistics published by the corresponding national institutions confirm these numbers. According to the Surveillance, Epidemiology and End Results (SEER) database of the National Cancer Institute, 443 per 100,000 people (men and women) were newly diagnosed with cancer in the United States alone in 2011 [205]. In the same year, 351 females and 440 males per 100,000 people in each case developed cancer in Germany according to the Robert Koch Institut [343]. With an absolute number of 220,914 people dying, cancer was ranked as the second

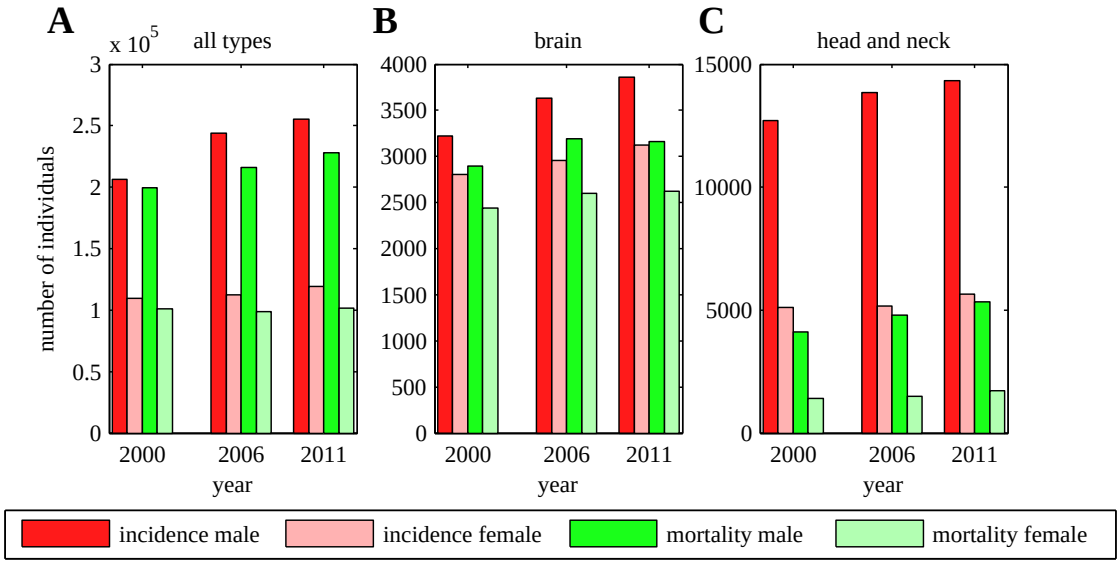


Figure 1.1: Gender specific number of cases for cancer incidence and mortality (years 2000, 2006, and 2011) in Germany. Data are presented for **A:** all types of cancer, **B:** CNS cancer, and **C:** HN cancer (data by courtesy of the Robert Koch-Institut, Berlin, Germany [343]).

most common cause of death after cardiovascular diseases in Germany [274]. Despite enormous research spendings, this is also the case for the United States (584,881 cases [48]) and other industrialized nations [205].

Figure 1.1 shows the gender specific development of incidence and mortality in Germany for the years 2000, 2006, and 2011. Notwithstanding ongoing research in the field, the absolute numbers for all types of cancer increase particularly for the male population. Similar trends are observed for brain and head and neck (HN) cancer, which are particularly in need of highly precise and careful treatment, since they affect the head region. Brain tumors account for 85-90 % of all central nervous system (CNS) tumors, while HN cancer includes tumors of the mouth, lips, nasal cavity, sinuses, salivary glands, throat, larynx, and lymph nodes in the neck [205]. Although HN cancer reaches higher incidence rates especially for males, fig. 1.1 indicates rather low mortality. In contrast, mortality comes quite close to incidence for brain tumors, which indicates a lower survival rate. Table 1.1 outlines a similar picture for the five year survival rate in 2010. CNS cancer has only about half the survival rate of cancer on average and only exceeds that of tumors in lung and liver. In addition table 1.1 illustrates the five year prevalence rate, which describes how many per 100,000 people were newly diagnosed with a specific type of cancer or were already under treatment in the five preceding years.

Table 1.1: Gender specific prevalence (per 100 000 people) and five year survival rate for 2010 of different cancer types in Germany. Data by courtesy of the Robert Koch-Institute [343].

Type of cancer	Prevalence rate		Survival rate [%]	
	Male	Female	Male	Female
All cancer types	1992.4	1867.2	60	66
Colon	293.3	237.2	64	66
Liver	18.6	7.2	16	13
Larynx	29.4	4.2	65	63
Lung	123.4	64.6	16	21
Uterus	-	111.7	-	81
Prostate	699.7	-	92	-
CNS	16.9	13.6	27	29

The survival rate for CNS cancer is lower than for cancer on average within all age groups (cf. fig. 1.2). This discrepancy strongly grows for patients of 45 years or older - falling even below 12 % for elderly people of 65 years and older. Incidence is higher for males and more frequent for whites than for blacks [205]. Secondary tumors in the brain, i.e. metastases, outnumber the number of primary brain tumors 10:1 and are the most common type of intracranial tumors in adults. Among others, metastases in the brain are predominantly developed by patients suffering from lung cancer which has got one of the highest incidence rates [217]. These key facts make intracranial tumors a critical and partly delicate issue of high relevance.

A decision for a certain treatment option typically depends on histology, localization and spread of the disease. Here, the WHO classification of tumors of the nervous system remains a gold standard and guideline for neuro-oncology [177]. Tumors are classified with respect to morphological as well as immunohistochemical features and are graded according to malignancy into four groups I-IV. The grading describes the potential for proliferation and risk of secondary tumors, and therefore provides a guideline for treatment [293]. Further information about crucial tumor properties such as volume, proliferation, metabolism, oxygenation or vascularization can be obtained from advanced imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT), or positron emission tomography (PET).

As one treatment option, surgical resection is generally recommended and particularly applicable for benign tumors of group I [23, 205]. However, especially for malignant

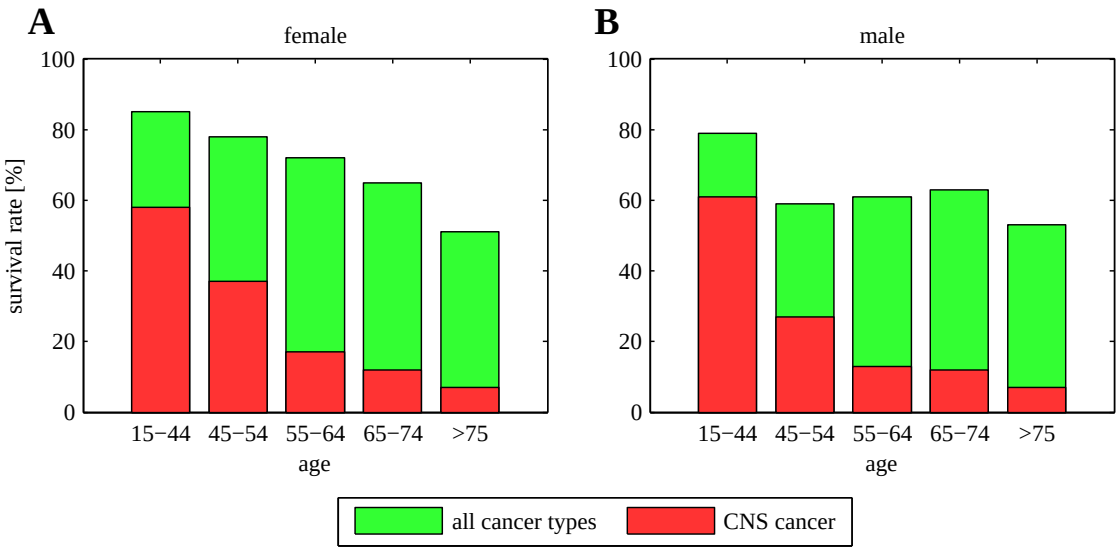


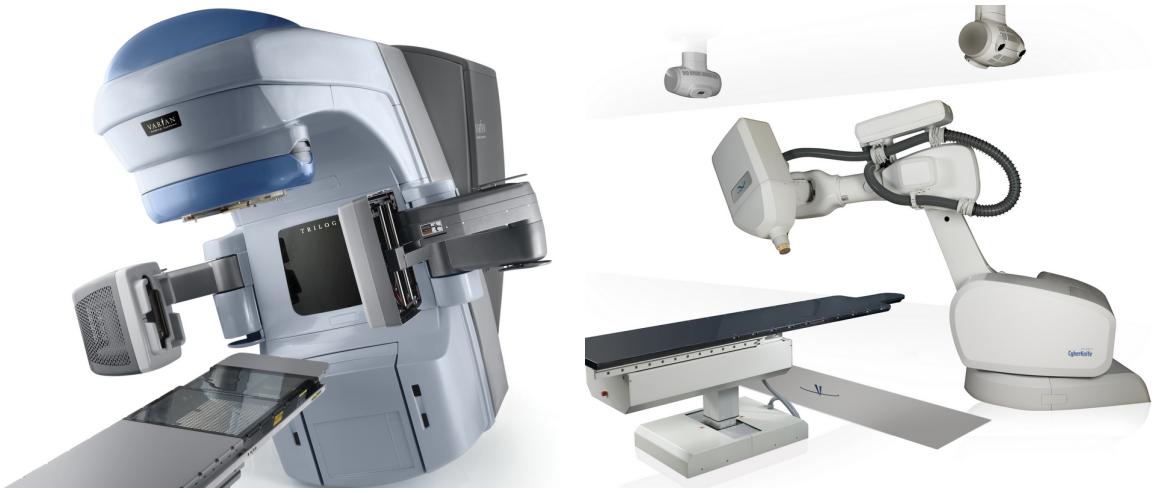
Figure 1.2: Gender specific five-year survival rates for all cancer types in relation to CNS cancer (2009-2010, **A:** female, **B:** male).

tumors of high proliferation rate and the risk of multiple metastases (grades III and IV), irradiation constitutes the definitive or at least a supportive option for treatment. The latter can be done in a pre- or post-operative setting, i.e. either kill residual cancer cells or make an unresectable tumor amendable to surgery [23]. Accounting for 60 % of all neoplasms, Astrocytic tumors, such as Glioblastomas (grade IV), are the most common primary tumor.

For irradiation, high-energy X-rays or γ-rays are directed onto the target to evoke ionizing effects on molecules of the desoxyribonucleic acid (DNA), enzymes or membranes [319]. This disturbs cell metabolism, leads to DNA mutations, and therefore loss of cell division capabilities. Since malignant cells have higher proliferation rates than normal tissue, repair mechanisms are not left enough time before replication. Ultimately, this leads to cell death, i.e. tumor growth control or shrinkage [23].

Radiation is either delivered by Brachytherapy using small implanted sources of moderate radiation, or by external beam radiation therapy (EBRT). The latter applies γ-rays percutaneously from a certain distance by means of an external treatment beam. In order to prevent normal tissue from being damaged, precise tumor irradiation is vital, particularly for intracranial tumors such as CNS or HN cancer.

As a third treatment option, chemotherapy very often goes along with surgery and EBRT. Here, cytotoxic agents increase tumor cell killing and moreover result in syner-



A: Trilogy® System (Image by courtesy of Varian Medical Systems, Inc., All rights reserved. [304])

B: Cyberknife® System (Image by courtesy of Accuray, Inc., All rights reserved. [3])

Figure 1.3: Typical examples of gantry-based, isocentric (**A**, from Varian Medical, Inc.) and robotic, non-isocentric (**B**, from Accuray, Inc.) linear accelerators (LINACs).

gistic effects of chemo- and radiosensitization, mainly through increased inhibition of DNA-repair mechanisms [23]. Particularly for HN tumors, positive effects have been reported [225].
In this context and with respect to the fact that nearly two-thirds of all cancer patients nowadays receive RT during their illness [6], Bernier and colleagues [23] state that radiation oncology will surely remain a key modality in the treatment and management of cancer during the next century.

1.2 Stereotactic Radiotherapy

1.2.1 Planning and Treatment

In the course of this work, the term conventional RT will be used for the clinically most common type of EBRT, consisting of a planning and treatment phase. The presented concept for tracking in this work will relate to this specific kind of clinical workflow. In EBRT radiation is applied percutaneously through high energy photon beams. These are generated by LINACs, which accelerate electrons using a standing microwave within the so-called waveguide. The electron energy increases during acceleration and typically ends up at several MeV. This energy, in the medical context often just called Mega Voltage (MV), describes the energy an electron would absorb after going through

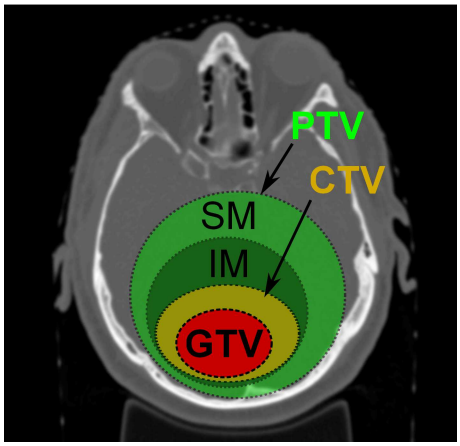


Figure 1.4: Simplified sketch of volumes and margins defined during treatment planning.

the voltage of several 10^6 V. After passing several deflection units, the electrons hit a target, where the impact generates electromagnetic waves (i.e. photons) of high energy. This radiation is often named MV X-rays, but actually corresponds to γ-rays in a strict physical sense.
Typical treatment devices are isocentric, gantry-based systems such as the Trilogy® System (Varian Medical Systems, Inc., [304]) shown in fig. 1.3A. The isocenter is the center of rotation for the gantry. The therapeutic beam exits the machine at the end part of the gantry. In contrast to that, robotic LINACs such as the one used in the Cyberknife® System (cf. fig. 1.3, Accuray, Inc. [3]) are capable of non-isocentric irradiation. The project work presented here is part of a research partnership with Varian Medical, Inc [304]. Consequently, all subsequent conceptual considerations and discussions of the marker-less tracking approach will first and foremost refer to gantry-based systems as shown in fig. 1.3A. Although these systems shall form the context, it should be emphasized that the general concept is not restricted to them, but easily transferable to others.

Planning

In the planning phase a specific dose is prescribed to the tumor target. Based on that, a plan is designed describing how this dose will be delivered. For this purpose a simulator device is used. It recreates the actual treatment machine with its isocenter and geometry [101, 277].
For planning purposes, the simulator is typically equipped with a CT device which

images the tumor target with respect to the machine isocenter. Nowadays, state-of-the-art planning may also use other imaging modalities such as MRI or PET [223]. This is particularly the case for treatments of the head region. Tumors in the head region are treated by so-called cranial RT which will be the main focus here. In contrast, the treatment of body tumors is named stereotactic body radiation therapy (SBRT).
Within the acquired planning CT, volumes and margins illustrated in fig. 1.4 are defined [55]. In this context the Gross Tumor Volume (GTV) labels the gross visible extent and location of malignant growth. Around this main part of the tumor, the Clinical Target Volume (CTV) additionally covers subclinical microscopic malignant parts and metastases, which also have to be eliminated. It is important to note that the CTV is a purely anatomical concept [133]. This means that it is independent of the treatment choice and defines a volume that definitely needs to be treated to achieve the aim of therapy: cure or palliation.
In supplements to earlier reports [55, 133], the International Commission on Radiation Units and Measurements, also defines additional safety margins resulting in the Planning Target Volume (PTV) [134]. As the union set of all safety margins, this volume takes the net effect of all possible geometric variations and inaccuracies into account. It is defined to select beam size and beam arrangements and ensures that the prescribed dose is actually delivered to the CTV. This means that the volume size depends on the chosen treatment technique, in order to compensate for effects of organ and patient motion, as well as patient setup inaccuracies.
Two different safety margins allow for two different sources of uncertainty: First, the internal margin (IM) covers effects that originate from physiological processes, e.g. bladder/stomach filling or respiratory deformations. Variations in size, shape and position of the CTV in relation to anatomical reference points are taken into account. Second, the set-up margin (SM) accounts for technical uncertainties. This includes variations in patient positioning and beam alignment during planning and all treatment sessions, mechanical inaccuracies of the equipment (e.g. gantry, immobilization etc.), human errors or beam geometry selection [134].

In the aforementioned context, accurate patient positioning with respect to the treatment plan as well as the quality of motion detection and compensation, directly links to the size of these margins (the SM to be precise) [307]. Margin reduction was shown to potentially lead to a better sparing of healthy tissue and dose escalation. As a result, patient care directly benefits from it [259]. Thus, a central motivation for intracranial tumor tracking in general is, (1) to reduce these margins, (2) to deliver more

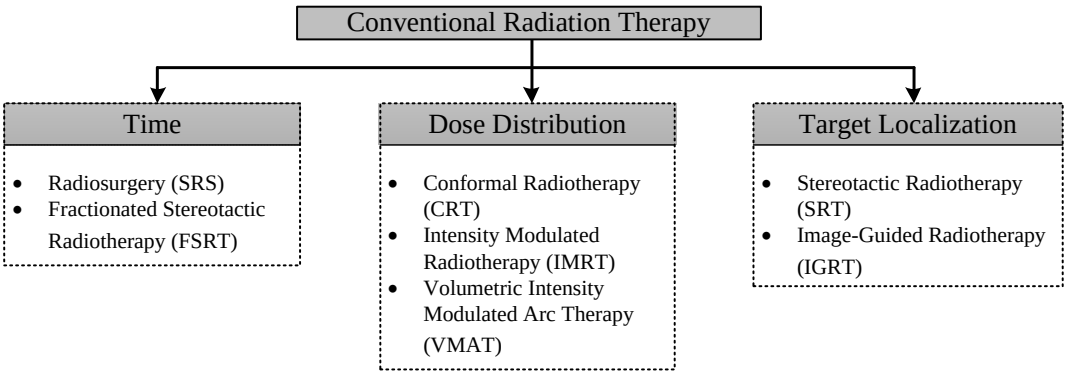
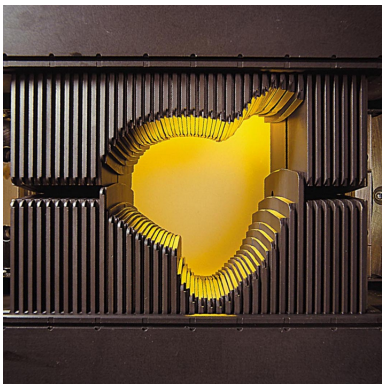


Figure 1.5: Overview chart showing different extensions of conventional radiation therapy. The extensions are sorted into three categories: (1) dose delivery schemes over time, (2) spatially optimized dose distribution, and (3) target localization improvements. The extensions are not mutually exclusive and overlap. The list does not claim completeness.

dose to the actual target, and (3) to spare healthy tissue from absorbing unnecessary dose.

Finally, after contouring the target along these guidelines, the treatment planning system will model the dose distribution within these regions. Here, different extensions of conventional RT follow different approaches as summarized in fig. 1.5. Conformal radiation therapy (CRT) exploits the superposition of multiple beams or beam shaping by multileaf collimators (MLCs) to achieve a dose distribution conforming to the actual shape and size of the tumor. In a similar way, intensity-modulated radiation therapy (IMRT) modulates the dose intensity across subregions instead of having a constant dose level within the entire volume. This 2D or 3D dose painting allows for a better sparing of normal tissue and higher conformity to the tumor shape.
Mainly automated treatment planning software [211] then optimizes beam positioning according to the chosen treatment scheme. This positioning also depends on specific organs at risk (OAR) which may be closely located to the tumor target. A treatment scheme that particularly links to gantry-based LINACs is volumetric intensity modulated arc therapy (VMAT). This fast scheme sculpts the dose distribution by going through a minimal number of rotations (arcs) around the tumor isocenter. In each arc the gantry stops at optimized positions and activates the beam (stop and shoot approach).

A: HexaPOD™ robotic couch system (Image by courtesy of Elekta, AB, All rights reserved. [76])



B: Millennium™ MLC (Image by courtesy of Varian Medical, Inc., All rights reserved. [304])

Figure 1.6: Typical examples of a robotic couch system and an MLC which can be used to compensate motion or alignment errors.

Treatment

At the beginning of a treatment session (called fraction), the patient is typically immobilized and then positioned with respect to the LINAC isocenter (see sec. 1.2.2 for more detail). Repositioning and alignment corrections that were found to be necessary, can be carried out manually, but most often using e.g. a robotic couch or MLCs (cf. fig. 1.6). In an ideal case, the tumor isocenter identified in the simulator during planning then precisely coincides with the treatment device isocenter. Since this can only be achieved with limited accuracy, the PTV adds the SM which covers this and other uncertainties. Tryggestad and colleagues [296] found that not only positioning uncertainties, but also small differences between the treatment and simulation device contribute to the overall inaccuracies.

Robotic Couch The couch system is capable of automatically maneuvering in 4 or 6 degrees of freedom. Positioning ranges typically fall within ± 4 cm and $\pm 3°$ with accuracies of $<$0.1 mm, and $< 0.1°$, respectively [102, 253]. While many devices still ignore pitch and roll rotations, recent studies confirm that state-of-the-art motion compensation extensively makes use of all 6 degrees of freedom to achieve the best possible positioning [132, 287]. Examples are given by the 6D HexaPOD™ couch from Elekta AB [76, 197] (cf. fig. 1.6A) or the Varian PerfectPitch™ 6D couch [253, 304]. First attempts towards automatic head adjusters were also made [323]. The adjustment unit makes small corrections

to the head position without moving the entire couch.

Multi-Leave Collimators MLCs were originally employed by CRT or IMRT and can be used to adjust the beam shape using multiple movable lead leaves (cf. fig. 1.6B). In this way alignment errors and even intrafractional motion [152] in the low millimeter range can be compensated. Positioning and re-positioning typically account for most of the time during a treatment session, while actual irradiation of the target takes seconds or a couple of minutes only [101].

Treatment Schemes Sources of localization uncertainties also arise from the treatment scheme, i.e. the way in which dose is delivered over time. For head and spine – the targets relevant for this work – cancer was mainly treated by stereotactic radiosurgery (SRS) in the past. Following the definition of the Swedish neurosurgeon Lars Leksell, high dose was delivered in a single session with steep fall-off dose gradients [168]. A more modern definition states that target tumors are inactivated or eradicated in up to five sessions [15, 47].
This and particularly the evolution of fractionated stereotactic radiation therapy (FSRT) (cf. fig. 1.5) in recent decades, made reproducible re-positioning and high precision localization from fraction to fraction a necessity for successful modern RT. FSRT divides the treatment process into a sequence of fractions with dose delivery usually once a day and five times a week. Special schemes referred to as hyper- or hypofractioning may differ from that. With 1.8-2.0 Gy, each fraction applies only a low dose compared to SRS, which also allows to treat larger target volumes [72]. The reasoning behind that is provided by the four R's: repair, redistribution, reoxygenation and repopulation [23]: Patient survival rates are increased when normal tissue is exposed to a lower dose per fraction. Low proliferation rates provide enough time for repair before replication. This is not the case for fast proliferating tumor cells, where unrepaired damage is lethal. Further on, tumor cells in the S phase of the cell cycle are more resistant. More dose in only that phase would not be effective. Finally, radiation effectiveness is higher after reoxygenation of the cells. Practically, Buatti et al. [41] found that fractioning avoids skin soreness, neurocognitive decline or alopecia. Therefore, fractionation improves the treatment results, but also sets new challenges for localization and positioning techniques.

Stereotaxy Improvements were achieved by the introduction of stereotaxy. The resulting schemes of SRS and SRT exploit surrogates that are ideally rigidly linked to the target in order to define a coordinate frame. Earlier on, these surrogates were markers at

a metal frame, mask or localizer box rigidly attached to the cranium. The requirement of a rigid attachment is never completely fulfilled. The error depends on the chosen immobilization scheme as the next subsection will point out. The tumor location within the stereotactic coordinate frame is typically acquired by X-ray based imaging during planning. The stereotactic coordinate frame in turn can then externally be localized with imaging modalities such as stereo X-ray, cone beam computed tomography (CBCT) or optical imaging.
Modern SRT aims at marker-less approaches to bypass the marker and partially the immobilization problem. The patient's anatomy, e.g. the skull bone, is directly localized as a tumor surrogate [296]. All this contributed to the advent of image guided radiation therapy (IGRT). Marker-less optical head tracking being one of these modern localization options will be the main focus of this thesis.

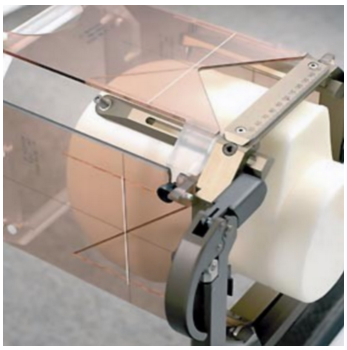
1.2.2 Tracking and Compensating Head Motion

State-of-the-art RT ensures precise target localization by two means: immobilization, image guidance, or a combination thereof. Ever since, both are closely linked and are important factors in stereotaxy and IGRT.

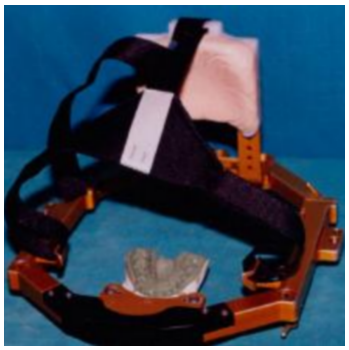
1.2.2.1 Immobilization

Immobilization restricts target motion and provides reproducible target positioning between planning and actual treatment. Moreover, it is very often also part of the stereotactic concept. Immobilization devices may carry marks or tattoos that can be used by image guidance or the clinician as an external reference for the stereotactic coordinate system. This aids patient alignment with respect to the machine isocenter. The simplest case is given by the room lasers. Initially aligning the isocenter marks on the immobilization device to the room laser crosshair, is part of almost every clinical workflow. Afterward, more sophisticated image guidance may be used.
The following paragraphs will describe four main immobilization devices in the chronological order of their introduction: (1) stereotactic surgical frames, (2) relocatable stereotactic frames, (3) thermoplastic masks, and (4) open-face masks. Except for relocatable stereotactic frames, variants of these categories are still used in everyday clinical practice. Examples are shown in fig. 1.7 and a summary is given in table 1.2.

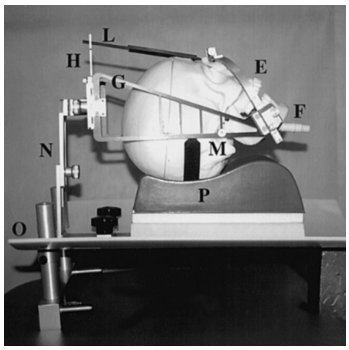
Stereotactic surgical frames With the advent of cranial radiosurgery, highly precise target localization was required. Early immobilization was therefore adopted from exist-



A: Leksell® Coordinate Frame G (Image by courtesy of Elekta, AB, All rights reserved. [76])

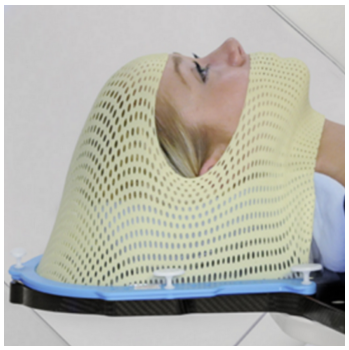


B: Gill-Thomas-Cosman relocatable frame (Image by courtesy of [233], Creative Commons Attribution 3.0 License)

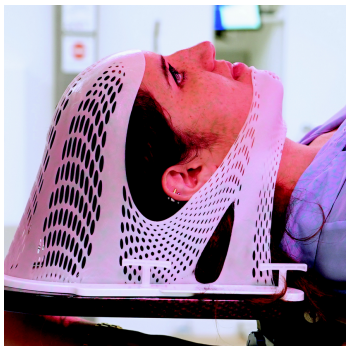


C: Laitinen stereoadapter (Reprinted from [148], Copyright (2001), with permission from RSNA®)

D: Orfit HP thermoplastic 3-point mask system (Image by courtesy of Orift Industries n.v., All rights reserved. [212])



E: Typical open-face mask for SRT (Image by courtesy of Varian Medical, Inc., All rights reserved. [304])



F: Typical open-face mask used for marker-less optical tracking (Image by courtesy of Vision RT, Ltd., All rights reserved. [310])

Figure 1.7: Evolution of immobilization devices from frame-based to frame-less in cranial SRT.

ing devices in neurosurgery. These were invasively attached to the skull bone by screws or fiber glass pins and were highly robust against target dislocation [169]. Patient set-up was only possible under local or even general anesthesia and mild sedation [339]. Common examples are the Brown-Robert-Wells [39, 123, 159, 179] or Riechert-Mundiger frame [121, 203]. As most of them, the former consisted of a metal head ring which could invasively be fixed to the patients head. A CT-localizer frame equipped with intersection rods could be rigidly placed on top. With the frame center and the position of these rods visible in each CT slice, precise stereotactic target localization was possible [159]. Apart from these two frames and modifications thereof [179], several other frames of similar design were adopted [27, 129, 257]. One particularly popular frame was invented by

Lars Leksell and evolved from these early frames [169, 183, 339]. It consists of an aluminium frame equipped with four pins which were placed in previously drilled holes to attach the frame to the skull. Figure 1.7A shows the metal frame with a localizer box on top. The frame with an immobilization accuracy of less than 0.1 mm in all three dimension is still used in modern SRS. Table 1.2 provides an overview of common devices as well as their strengths and weaknesses.

Relocatable stereotactic frames A major drawback was given by the fact that these frames could only be worn over short time intervals [104] and that they were hardly relocatable. Attempts were made to use frames across several sessions by reusing previously drilled holes and tattoos as orientation marks [257], or by simply leaving the metal frame in place for days [119]. These were, however, accompanied by skin infections and high patient discomfort.
An increasing number of FSRT treatments entailed the development of noninvasive and relocatable stereotactic frames [104, 118]. Two relocatable frames compatible with the Brown-Robert-Wells mount were introduced: the Gill-Thomas-Cosman frame [72, 104, 108, 171] and the Laitinen stereoadapter [62, 117–119, 163]. The former consists of a U-shaped base plate to which a dental tray is attached. While the dental impression fits to the patient's upper jaw, occipital fixation is achieved by an adjustable head rest (cf. fig. 1.7B). Finally three quick-release nylon straps fully immobilize the head [104]. The frame can be combined with a localizer box, which is visible in CT images due to metallic markers.
The Laitinen stereoadapter achieves immobilization by ear plugs and nasion support (cf. fig. 1.7C). Ear plugs are pressed against the auditory meati and can be adjusted with millimeter scales. Aluminium triangles with four transverse arms are fixated on both sides of the head. The arms as well as localizer pins are made to show up in CT and hence to define the stereotactic coordinate frame in which the target is defined.
Studies for both relocatable frames have shown typical re-positioning accuracies below 1 mm [62, 104, 163]. However, occasionally high positioning errors of several millimeters render these frames inapplicable for high dose, single fraction applications such as SRS [160]. These errors have been found to predominantly occur for patients with poor upper dentition [119, 160] or due to scalp motion within the frame [119]. This emphasizes the need for a highly skilled operator and moreover comprehensive patient cooperation. The latter makes this immobilization scheme not applicable to patients who are not able to carefully follow instructions such as very young children [72]. A possibly required anesthesia is inapplicable due to the mouth bite, which is also not tolerated by all

adults. Modifications have been proposed by Kooy et al. [160] and Dunbar et al. [72]. Both frames have been found applicable for claustrophobic patients since the head is not covered during treatment [108].
Although the high effort entails setup times of up to 20 min [104], the general clinical workflow has become more flexible. Imaging and planning can now be separated from the actual treatment. With these frames the treatment could also be distributed across several fractions.

Thermoplastic masks In modern SRT, thermoplastic masks adopt the advantages of relocatable frames, but tackle some of the most severe drawbacks (cf. fig. 1.7D). They nowadays constitute the most commonly used immobilization scheme. They consist of thermoplastic material which is made moldable in a warm water bath. The mask can then be molded directly on the patient's face who is typically in supine position. The material hardens within minutes and can be re-used across several fractions.
First attempts with plastic material have already been made in modifications of relocatable frames for treating children [72, 160] and were soon extended to thermoplastic full masks in combination with bite bars [33, 41, 260]. In early immobilization systems, the actual mask was fixed to a metallic horseshoe frame [198, 240, 260] for additional stability. Recent studies, however, argue that the metallic frame increases geometric distortions of the mask when attaching it to the treatment device [296]. Due to slight differences between treatment and simulation device, they hence directly entail systematic interfractional errors (re-positioning errors from fraction to fraction compared to the planning) [103, 241]. To separate these from random errors, El-Gayed and colleagues [75] proposed to measure interfractional errors by the signed mean and its standard deviation. Apart from these geometric distortions, Li et al. [173] also found random translational shifts of up to 9 mm induced by the process of locking the mask to the treatment table. While the way of locking varies, there are typically either three, four or five fixation points. Five-point masks include a shoulder immobilization, which has, however, not been found to significantly improve immobilization accuracy for brain tumors [103]. Modern masks from manufacturers such as Orfit Inductries n.v. [212], BrainLAB AG [35], Civco Medical Solutions Inc. [60], or others usually do not rely on metallic frames and are hence often referred to as frame-less immobilization. Together with modern IGRT, frame-less masks can separate target localization from immobilization without any stereotactic localizers [41]. With short setup times and low manufacturing efforts, these masks easily integrate into the FSRT workflow. Rotondo et al. found on average 5 min setup time to be necessary [241]. A questionnaire also

indicated only slight discomfort felt by most patients.
Interfractional repositioning errors mainly occur due to patient motion in the mask and because the mask cannot identically be fitted to the head over several fractions. Using the mask or a frame attached to the mask as a surrogate for positioning, then results in incorrect target alignment. Common initial positioning, however, still uses the room laser crosshair to align tattoo marks on the mask to the treatment room isocenter. These are placed on the mask during planning. For the room lasers alone, Stephenson reported 2 mm variance over a six months period. Even though suffering from the similar problems, marker-based optical tracking represents another approach for initial alignment [176, 273, 315]. Here, reflective markers are attached to the mask.
Thus, quantification of interfractional and also intrafractional errors is an essential, yet challenging part for clinical workflow design. Nevertheless, it has not been done consistently throughout the literature [202]. Approaches vary by patient population, error measure, clinical workflow protocol and imaging modality used as a ground truth. Interfractional alignment was measured using either optical tracking [41, 155, 173, 273], portal imaging [103, 260, 263, 320], or CT/CBCT imaging [31, 97, 111, 150, 176, 187, 198, 240, 296, 305]. Ideally – to only investigate errors originating from the mask itself – anatomical landmarks or internal fiducials [110, 154] are compared to external landmarks such as mask field edges, markers on the mask or the mask fixation frame. This is true for some studies that used portal imaging or CT/CBCT [111, 198, 240, 260], but may not be the case for other studies [31, 41, 187, 273, 296]. Thus, the errors presented there may also include inaccuracies from the room lasers or manual/automatic re-alignment errors using e.g. the couch system. Peng et al. [220] measured considerably higher interfractional errors when initially aligning with the room lasers instead of markers attached to a mouth bite. Another error source is given by the imaging resolution. Further, portal imaging has only a limited view through the beam's eye and rather poor soft tissue contrast.
All this makes most studies very difficult to compare. Most errors presented in these clinical studies are probably higher than the actual interfractional accuracy of the mask. They necessarily include other error sources, but nevertheless reveal that positioning according to surrogates on the mask leads to incorrect target alignment. Additional verification is imperatively needed. Table 1.2 gives a rough overview of intervals for 3D translational and rotational errors which were found on average in most studies. The values in brackets include an error spread of one standard deviation. Occasionally higher errors have been reported for some cases [103, 296, 320]. Most studies also distinguish between superior-inferior (SI), lateral (LAT), and anterior-posterior (AP) directions.

While some studies emphasize particularly high errors in SI direction [240], the specific distribution of the errors across the three spatial dimensions strongly depends on the type of mask, the material and operator skills or protocols. Errors considerably increase when the patient experiences swellings or weight loss in the course of the treatment [97, 107, 240, 279]. This is even more likely when RT is combined with chemotherapy [23]. Furthermore, particularly these cases render mask shrinkage or sharp edges on the mask problematic.
Intrafractional errors, i.e. motion within the mask during dose delivery, are lower than interfractional errors. They have been measured by either comparing pre- and post treatment CBCT (motion range only) or continuous optical tracking during treatment [150, 273, 296] (cf. table 1.2).

All these errors can be used to compute safety margins and to ensure that 95 % of the dose will be delivered to the CTV [302]. Common margins added to the CTV are about 2-3 mm [198].

Open-face masks One major drawback of full-head masks is the claustrophobic distress a patient is exposed to. This full coverage of the face also prevents marker-less optical approaches from directly tracking the skin surface. By relaxing this restriction, studies have shown that excluding the shoulders from immobilization or cutting eyes and mouth free does not influence the overall immobilization accuracies, but reduces the claustrophobic distress [263, 320]. Velec et al. made similar experiences with so-called skin-sparing masks which left the jaw and lower neck free [305]. Removing parts of the mask was further found to reduce skin toxicity. Lee et al. discovered that skin absorbs 18 % more dose when covered by the mask due to its bolus effect [167].
As a consequence, some departments [173, 215] started migrating to so-called open-face masks. As shown in fig. 1.7E, they leave the facial area free and are highly suitable for claustrophobic patients. Only a minor worsening of the immobilization errors was found in recent investigations by Li et al. [173]. Rather than tracking the mask surface, they also provide access to the face for marker-less optical tracking. Manufacturers of marker-less tracking devices such as VisionRT, Ltd therefore recommend using these masks [215, 310] (cf. fig. 1.7F). This is reasonable since studies indicated large differences when comparing the motion of the mask surrogate with the actual motion happening inside the mask [306].
Finally, open-face masks have raised mild concerns about a higher risk of mask twists and geometric distortion, particularly when weight loss is experienced. Furthermore,

Table 1.2: Types of immobilization devices in SRT. The table lists type, strengths, weaknesses as well as typical accuracy intervals.

type	strengths	weaknesses	accuracy[1]	references
stereotactic surgical frames	- highly accurate - applicable in other fields	- invasive - anesthesia required - hardly relocatable - short time wearable	<0.1 mm	- Leksell [169, 183, 339] - Brown-Robert-Wells [39, 123, 159, 179] - Riechert-Mundinger [121, 203] - other [27, 129, 257]
relocatable stereotactic frames	- relocatable - noninvasive - accurate on average - head exposed - fits to existing mounts	- requires patient cooperation - occasionally high dislocation - needs skilled operator - setup effort - only partly tolerated	<1 mm (<1.6 mm)	- Gill-Thomas-Cosman [72, 104, 108, 171] - Laitinen [62, 117–119, 163]
thermoplastic full-masks	- relocatable - noninvasive - low setup time - moderate comfort	- claustrophobic distress - face covered - rare skin toxicity - mask shrinkage - weight loss/swelling	<3 mm/0.7° (<4-5 mm/1.5°) **<1 mm** **(<1.5 mm/1°)**	- Orfit [97, 150, 176, 320] - BrainLAB [87, 198, 240, 273] - Other [31, 111, 187, 234, 298] [33, 41, 78, 103, 241, 260, 305]
open-face masks	- same as full-mask - low claustrophobic stress - face unexposed and accessible for tracking	- slightly more motion - mask may twist - restricted laser mark placement	<2.5 mm/1.8° (<5-6 mm) **<1.5 mm/0.5°** **(<2 mm/1°)**	- manufacturers [60, 212, 310] - studies [155, 173, 215, 220, 221, 305]

[b]Normal font refers to interfractional and **bold font to intrafractional** 3D errors. Values given in brackets correspond to intervals of average plus standard deviation in addition to only average intervals without brackets.

alignment tattoos for the room lasers are more difficult and less reliably to set on the restricted mask area.
Thus, alternative approaches propose to establish a rigid link to the skull bone by customized mouth bites. This additional fixation allows to substantially reduce the head proportion covered by the mask. Once the mouth bite is equipped with markers, it can also be optically tracked for positioning. Although Buatti et al. [41] revealed a reproducibility for mouth bite insertion of less than 0.5 mm on average, there is a high variance across patients. More recently, Wang and colleagues obtained errors of less than 2 mm for 87.5 % of all patients when positioning patients with optical markers attached to a bite block [316]. Reproducible insertion was found particularly problematic for patients with poor dentition [215, 316]. Further disadvantages involve the bite effort and the resulting fatigue as well as occasional oral toxicity, increased swallowing motion and incompatibility with anesthesia [173].

1.2.2.2 Image Guidance

Target localization distinguishes two different cases: (1) re-positioning the patient with respect to the LINAC isocenter at the beginning of each fraction (interfractional positioning), and (2) tracking target motion during the treatment (intrafractional tracking).
As mentioned earlier in this section, re-positioning is initially done by aligning marks on the immobilization device to the treatment room lasers which intersect at the machine isocenter. This constitutes the worst option due to interfractional inaccuracies of the immobilization device, but roughly positions the tumor target within the field of view (FoV) of more sophisticated imaging modalities used in IGRT. The imaging modalities include: stereoscopic X-ray, CBCT, portal imaging, marker-based optical tracking and marker-less optical tracking. These are described in the following. Note that other modalities such as ultrasound (US) or electromagnetic imaging [250] are not described due to their minor relevance for for cranial SRT.

X-ray-based Imaging In the context of this thesis, X-ray imaging will always refer to kilo Voltage (kV) imaging, since it agrees with the definition of X-rays in a strict physical sense. While classical X-ray systems still required films to obtain the images, modern systems almost exclusively rely on flat panel detectors and fluoroscopy for fast, digital acquisition [21]. Possible approaches include stereoscopic X-ray [145, 176], kVCBCT [143, 210] or in rare cases CT-on-rails [180]. Most treatment systems are equipped with the former two approaches. Widely used examples are the Cyberknife® system (cf. fig. 1.3B, Accuray, Inc. [3]), the Varian On-Board Imager® (OBI) (cf. fig. 1.3A, Varian

Medical, Inc. [304]), or the Elekta Synergy® system [76]. Some systems such as the latter two are capable of 2D and 3D imaging.

2D imaging is accomplished by stereoscopic X-ray imaging. Here, X-ray images from orthogonal directions (e.g. at defined gantry positions) result in two projections that describe a 3D volume in the FoV. These images can then be registered to the planning CT by generating 2D digitally reconstructed radiographs (DRRs) from the 3D planning volume [222]. This is termed 2D to 3D fusion. A typical clinical workflow first aligns the patient with respect to the room lasers. Second, X-ray images are acquired and, third, interfractional alignment errors obtained from image registration to the planning CT are compensated with the robotic couch system, for instance.
Since stereoscopic X-ray imaging is faster than CBCT and entails lower imaging doses, it can also be used to compensate intrafractional motion during the treatment. Typical examples suitable for frame-less treatment are the BrainLAB Novalis® system [35, 145, 176] or the Cyberknife® 6D skull tracking [153]. Due to the extra dose delivered and image acquisition time, continuous tracking is typically limited to frequencies below 1 Hz. Treatment without mask immobilization is possible, but bears a risk for sudden motion that cannot be captured by the imaging speed. Another impairing restriction for gantry-based solutions is that they are only capable of acquiring images at limited gantry positions, i.e. views. Mask-less treatment is therefore hardly used.

The workflow for CBCT positioning is equivalent to the one described for stereoscopic imaging [296]. In contrast to ceiling mounted imaging equipment, these gantry-based approaches are capable of also acquiring 3D volumetric CBCT images. Due to very long acquisition and image reconstruction times (2-5 min according to [107, 278, 279]), intrafractional monitoring is not feasible. Most time is spent for reconstruction [278].
The registration accuracy with CBCT images depends on three major technical factors. First, most treatment systems provide either 3 or 6-degrees of freedom image fusion with the planning CT. 6D fusion was found more accurate for the type of motion usually occurring within the clinical application [176]. Second, the imaging resolution is high (voxel size of 1 mm or better [52]), but limited. In addition, the planning CT typically has slice thicknesses of 1-3 mm [97]. Third, the calibration to the machine isocenter. For state-of-the-art calibration procedures, manufacturers state deviations between imaging and machine isocenter of less than 1 mm, whereas quality assurance was shown to achieve accuracies down to 0.3 mm [221]. For the Varian On-Board Imager®, the calibration typically yields errors below 0.5 mm [98]. Finally, the gantry motion also

induces isocenter shifts of up to 0.7 mm while moving [145].

Therefore, and because registration relies on volumetric, intensity-labeled fusion, CBCT or X-ray in general is seen as a robust gold standard for localization and tracking in FSRT and SRS [172, 215, 220]. The required imaging dose, however, is seen as one of the central drawbacks. Studies report a common dose exposition of 3-10 mGy per acquisition [204, 264, 272, 278]. Particularly accumulated over more than 30 fractions, this is seen problematic [107]. Alternative guidance approaches are possibly needed to reduce the frequency of X-ray based acquisitions required up to date [279]. Although stereoscopic imaging requires less dose as such, intrafractional tracking with this method can be judged along similar lines.

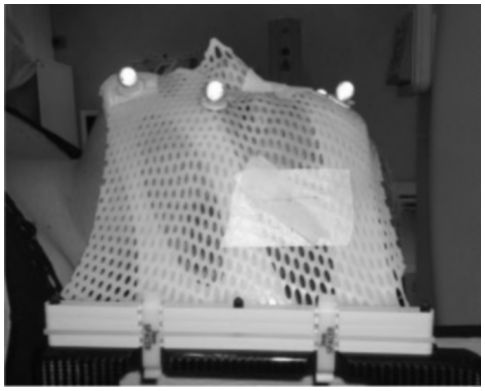
Portal Imaging In the context of this thesis, portal imaging refers to MV γ-ray imaging. This type of imaging is done through the eye (named "port") of the therapeutic beam. In theory this constitutes the most accurate localization technique, because the image precisely corresponds to what is actually treated. The imaging isocenter is identical with the machine isocenter and is unaffected by gantry motion. No extra dose is delivered to the patient, unless additional verification scans, not being part of the original treatment plan, are acquired. This on the other hand involves high extra dose. In this context, there is also the option for MV CBCT [64].
However, image resolution is lower when compared to CBCT [54] and the FoV is very limited. Consequently, image registration to a kV planning CT only relies on a very small sector. Moreover, portal imaging also suffers from poor soft tissue contrast and bad detector efficiency [278]. Ideally, bony landmarks or implanted fiducials can be used for better image registration. Internal fiducials have been proposed for cranial SRT [110, 154], but are reluctantly used due to their invasiveness.[1]
While classical MV imaging was done using port films, the evolution of electronic portal imaging using amorphous silicon flat panels now enables LINAC-integrated, fully digitized image acquisition [11].

Marker-based Optical Tracking Optical tracking overcomes some of the major drawbacks of X-ray-based tracking. The target is localized using light and does therefore not add any additional ionizing radiation to the treatment. Marker-based tracking relies on a small set of markers that are tracked by a ceiling-mounted camera. Markers can be

[1] This also makes electromagnetic, beacon-based tracking as performed by Calypso® a possible [61, 250], but questionable choice for cranial RT.

Figure 1.8: Integration of the Novalis® ExacTrac® system into a gantry-based treatment system. The system consists of two ceiling mounted aSi flatpanel X-ray detectors and an optical tracking camera to track passive markers. (Image by courtesy of BrainLAB, AG, All rights reserved. [35])

reflective marker spheres (”passive markers”) or light emitting diodes (LEDs) (”active markers”), whereas the former are more widely spread. Active markers usually emit non-visible light, e.g. in the infrared (IR) or near-infrared (NIR) range. Robust 6D tracking becomes feasible with the use of more than three markers [315]. Using this concept, optical tracking was also introduced to reduce cost, personnel, complexity of SRT as well as patient discomfort [61].
There are two systems commercially available: the BrainLAB ExacTrac® system and the Varian/Zmed RadioCameras. Both are passive marker-based tracking systems.
The ExacTrac® system can be combined with stereoscopic X-ray imaging resulting in the Novalis® localization system. This system has been investigated by several research groups [16, 145, 155, 176, 182, 273, 315, 340]. It consists of two IR cameras (depending on the released version a Polaris™ [206] or other model) which are mounted onto the ceiling (cf. fig. 1.8). The system is capable of tracking 5-7 marker spheres with a frequency of 20-30 Hz. The imaging speed is therefore faster than for X-ray-based approaches. It can be used for both, positioning and intrafractional motion monitoring. The latter enables gating, i.e. switching the therapeutic beam off for large deviations from the isocenter.



A: Passive marker spheres tracked by the ExacTrac® camera are the attached to a thermoplastic mask [273]. (Creative Commons Attribution License)

B: Passive marker spheres tracked by the RadioCameras (Zmed/Varian Inc., Ashland, MA) are attach to a mouth bite. (Image by courtesy of Varian Medical, Inc., All rights reserved. [304])

Figure 1.9: Typical examples of marker-based tracking approaches.

The marker identification error is specified as 0.3 mm [145, 273]. The markers are not required to be placed in a rigid regime and can therefore be attached to the chest or for SRT to the thermoplastic mask (cf. fig. 1.9A). This compromises the overall tracking accuracy with respect to the target. Tracking surrogate motion of the mask does not provide reliable information about the internal target motion. This was supported by Linthout et al. [176] and Spadea et al. [273], who both confirmed that the optical tracking system reported less motion than actually found by 6D target fusion between planning CT and verification X-ray images. Thus, the overall tracking accuracy is similar to the intrafractional errors reported for thermoplastic masks and ranges between 1 mm and 3 mm [145, 273, 315].

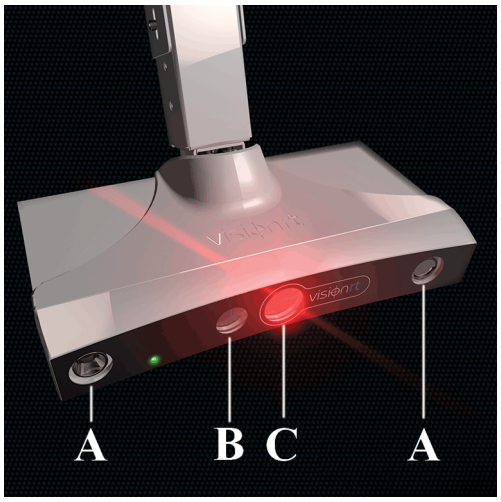
The second system, the RadioCameras (Varian/Zmed) [304], also uses a Polaris™ [206] IR tracking system and an array of more than three passive marker spheres [61, 149, 189, 190, 220, 244, 292, 315]. The markers are placed in a rigid regime onto a customized maxillary bite block. The mouth bite is then fixed to the upper jaw of the patient allowing only partial mask coverage for the rest of the head (cf. fig. 1.9B). A slightly modified version called SonArray® can be used for US or customized applications.
The unambiguous marker constellation of the rigid regime can be stored in files for reference. The stored geometry can then be tracked during the treatment and finally be registered to the marker locations extracted from the planning CT. Inaccuracies of the stored reference geometry correspond to the marker identification errors. The extraction

accuracy from CT was found to be <0.5 mm [315]. Due to the bite block and rigid fixation to the skull, the accuracy of the RadioCameras is higher than for the ExacTrac® system. The reproducibility of the bite block placement was reported to be on average below 0.5 mm [41]. Further advantages and disadvantages for using a bite block have been discussed for the immobilization devices already. Most important, D'Ambrosio and colleagues [61] emphasized that bite effort and fatigue made treatment times above 15 min hardly possible.

Marker-less Optical Tracking The discrepancy between surrogate and actual target motion constitutes one of the major drawbacks in marker-based optical tracking [220]. Therefore, it is desirable to directly track the motion of the patient's anatomy. In case of cranial RT, tumor motion typically has a high correlation with the corresponding motion of the skull. Therefore, X-ray-based image registration mainly relies on registering the bony orbit from different recordings [221]. In contrast to this volumetric approach, marker-less tracking registers surface information. The surface is usually given by the skin surface and is assumed to be a representative for the target motion. Kim et al. [156] showed that registering surfaces can be problematic. Limited information as compared to volumetric fusion makes the registration more sensitive to changes in surface shape or to large motion. The latter is even more important when the surface is smooth without distinct landmarks. By comparing with CBCT fusion, he showed for real data from HN patients and simulated motion that theoretical inaccuracies were on average 2.7 mm, and can be as large as 5.2 mm. Although the situation may be better for intracranial tumors, this finding stresses the major weakness of marker-less approaches.
On the other hand, it does not rely on additional radiation or artificial surrogates, is a fast way to monitor even intrafractional motion and is inexpensive as far as the device is concerned. By providing the advantages of marker-based optical tracking and tackling its main drawback, the marker-less approach has got a high potential.

The majority of devices relies on the principle of triangulation. Alternative approaches such as time-of-flight cameras have been proposed [227], but still remain a minority. Triangulation can be either based on natural features of the patient (vision-based stereo-camera photogrammetry [341]), or on projected patterns which are observed by a camera [65, 174, 254]. An excellent overview is given by Chen et al. [56]. In FSRT, projection methods – so-called laser triangulation – have prevailed due to their superior robustness. Possible projections involve speckle patterns [24], scanning laser lines [213] or single laser spots [79], and are observed by one or multiple cameras.

A: Typical treatment scenario: two ceiling mounted cameras image the surface geometry of the patient's face which is exposed for tracking by the open-face mask.



B: The AlignRT® [310] device. A - stereo-camera system for triangulation, B - gray-scale texture camera, C - speckle projector.

Figure 1.10: The AlignRT® [310] marker-less optical tracking system. (Images by courtesy of VisionRT, Ltd., All rights reserved. [310])

Both, low-cost consumer products as well as dedicated tracking devices have been proposed for usage in SRT. The former include the Microsoft Kinect®, for instance, which, however, suffers from a very limited accuracy (1 cm depth resolution at 2 m) [247, 338]. Apart from that, mainly two commercially available systems have been established in FSRT so far: (1) the Sentinel™ or Catalyst™ systems from C-Rad (Uppsala, Sweden) [44], and (2) the AlignRT® system from VisionRT (London, UK) [310]. These shall be discussed in the following.

AlignRT® The AlignRT® system consists of one to three ceiling-mounted cameras in typical distances of 1.8 m-2.7 m to the patient [24, 254]. To avoid target occlusion during gantry motion, and to generally increase the surface coverage, one camera is mounted on each side, and one in front of the treatment couch. Figure 1.10A illustrates an SRT treatment scenario with two cameras.
Each of these cameras consists of two charge-coupled device (CCD) cameras for stereoscopic imaging, a gray-value texture camera and a speckle generator (cf. fig. 1.10B). A speckle pattern is either statically or dynamically projected onto the patient. The pattern (as indicated by the light red illumination in fig. 1.7F and fig. 1.10A) consists of about 10,000 points of 1-3 mm spacing [220]. These patterns generate unambiguous landmarks

on the surface which are observed by the stereo-camera system. The identification of the same points in both camera images forms the basis for laser triangulation and hence 3D surface reconstruction. The gray-value camera is only used for visualization of the treatment scene and not for alignment [24].
The system first generates a reference surface offline and then registers surfaces acquired online (the so-called "real-time mode") to it. The reference can either be a surface reference extracted from the planning CT, or a high resolution optical scan recorded by the camera system itself. Studies have shown, that the optical reference tends to provide better registration accuracies with respect to target alignment [220]. Beforehand, the chosen surface reference can be restricted to only include specific regions-of-interest (ROIs) or be manipulated to exclude certain areas such as the eyes, the nostrils or the mouth [50]. The acquisition frequency depends on the size of the chosen ROI and may vary in real-time mode between 0.1 and 7.5 frames per second [24, 50, 220]. This can be exploited for respiratory motion compensation or gating in SBRT. It is called GateCT® or GateRT® in this context [151, 252, 254].
Before the system can be used, a calibration with respect to the machine isocenter is performed. This is carried out by acquiring images from a special calibration plate which is aligned to the isocenter using the room lasers. The calibration errors with respect to the isocenter were reported to be less than 0.5 mm [172] and the calibration errors among different cameras less than 1 mm [24]. The latter merges the surface point clouds between all cameras and ensures small alignment errors among them. Stability tests for 57 h of operation revealed shifts of the camera coordinate system that were below 0.5 mm [24].
The accuracy of the registration and the capability to precisely report target motion was tested against X-ray-based imaging such as CBCT by several groups. In these, initial patient positioning was done with AlignRT® and afterward verified using CBCT [50, 172, 220]. Initial phantom tests yielded errors of less then 1 mm [49] or even down to 0.1 mm and 0.1° on average [172] in all spatial dimensions.
Average alignment errors in real patient scenarios were mainly found to be less than 2.5 mm [50, 172, 220], but may, with up to several millimeters, be occasionally higher [50, 107, 156]. Increasing errors have particularly been reported for increasing distances to the isocenter [49, 220].
The system has also been successfully integrated into the clinical workflow [25, 173, 175, 254]. An average time for patient setup using surface scanning of 14 min was achieved [50, 215]. Here, the total setup time of 26 min also included an additional patient alignment using CBCT (11-12 min) for cross-checking. Setup times with both modalities were therefore comparable (median treatment time was 40 min for comparison [50]).

A: C-Rad [44] Sentinel™ system consisting of one laser line projector and a triangulation camera.

B: C-Rad [44] Catalyst™ system consisting of a laser line projector and a triangulation camera.

Figure 1.11: Two optical marker-less tracking devices by C-Rad [44]. Images by courtesy of C-Rad, Ltd., All rights reserved. [44]).

In terms of robustness, eye blinks have not been found to constitute a significant problem [50]. General concerns were, however, discussed with respect to deformations of the surface geometry. Rigid registration to the reference surface might result in substantial misalignments. Surface changes can be caused by facial expressions [50, 174], weight loss/gain, or medication [220]. Lee et al. reported a median weight loss of 3.3 % across the treatment time with daily weight loss between 0.15 % and 0.22 % for HN patients [166]. In fact, studies, which stated average errors below 2.5 mm, also reported cases with alignment errors between 3 mm and 10 mm [50, 107, 220]. These outliers may originate from the aforementioned error sources. Indeed, Gopan and colleagues confirmed that some ROIs such as the cheek bones or the forehead give more stable registration performance than others [107]. Further issues compromising the robustness of the system are suboptimal camera calibration, poor definition of the body contour as extracted from the CT scan, insufficient structure information within the ROI or the availability of only one camera due to occlusion [50].

C-Rad Systems The Swedish company C-Rad [44] developed two marker-less tracking systems: the Sentinel™ and the newer Catalyst™ system. Until recently [43], the

Sentinel™ system was also distributed by LAP GmbH Laser Applikationen [164] under the brand name Galaxy system [34, 201, 202]. Due to a dispute about malfunctions, the cooperation was canceled. Being competitors for AlignRT®, the devices have recently been ordered by a couple of clinics [77, 137–140]. Both systems work in a similar manner as compared to the AlignRT® system, but in contrast rely on the projection of a scanning laser line instead of a speckle pattern [213].

The Sentinel™ system shown in fig. 1.11A consists of a complementary metal-oxide-semiconductor (CMOS) BCi4 LS camera (C-Cam Technologies [42]) and a 690 nm laser line sweeping across the target. The galvanometric deflection unit is calibrated relative to the camera which enables laser triangulation of a 3D surface. The laser scans 40 cm, i.e. 30 contours in 2 s time with a nominal resolution of 0.2 mm (for earlier versions Moser et al. mention a deviating lateral resolution of 0.5 mm [202]). The scanner weighs 8 kg [213]. Similar to AlignRT® the motion tracking can either rely on a CT surface or high resolution optical reference. The latter has been found to be of higher accuracy [213, 214]. For long operation times a baseline drift for the system of up to 3 mm was measured by Moser et al. [201].
Phantom tests revealed accuracies below 1.5 mm in all spatial dimensions [201, 213] and a calibration accuracy with respect to the machine isocenter of less than 0.5 mm [201]. In scenarios with real patients suffering from HN or brain cancer, the system yielded positioning errors below 4 mm and 2.1° in all spatial dimensions [278]. In another study, Moser et al. found 75 % of all 3D errors below 3.2 mm and deviations in the roll angle of below 2° [202]. Both studies, however, may be compromised by the fact that they scanned the mask surface and not the face directly. This, on the one hand, does not reflect the real internal target motion as corrected for by the CBCT gold standard, and also only provides an uneven and suboptimal surface for the laser scan.

The Catalyst™ system shown in fig. 1.11B represents the successor to the Sentinel™ system. In contrast to the latter, it consists of three cameras positioned similarly to the AlignRT® system with respect to the treatment couch. For each camera, integration time and camera gain can be adjusted by the user to cope with different lighting conditions, surface shapes and colors. This was successfully tested by Stieler et al. [279]. He also confirmed reproducibility and tracking accuracy of less than 1 mm with a phantom as well as less than 3.5 mm and 2° average deviation from a CBCT ground truth for positioning three HN cancer patients [279]. As for earlier studies on the other system, these results may also be compromised by scanning the mask surface and not the

patient's face.
The Catalyst^{TM} system provides laser projection with blue (minimal skin penetration - used for laser triangulation), red and green lasers. With the latter two, certain parts of the body can be illuminated to indicate correct and incorrect alignment or to guide the operator.
Finally, large errors of up to 13.4 mm were recorded with patients experiencing notable weight loss. Scanning the mask surface has surely contributed to these high values. However, similar problems with surface reproducibility were also reported by Pallotta et al. [214] for thoracic regions when treating overweight patients.

Summary

Table 1.3 presents the localization approaches discussed in the last paragraphs in an overview chart. The approaches are compared according to eight main categories, whereas the color-coding (from red, orange, yellow to green) indicates a rating (with red being most negative and green most positive). The categories are as follows.

Imaging Modality This category is weighted neutral and lists the general concept which is used for imaging.

Imaging Dose The imaging dose indicates the extra amount of dose which is used for imaging only. The optimal case corresponds to no extra dose for the patient.

Detection Capability It is described what kind of motion can sufficiently be monitored: interfractional, intrafractional or both.

Registration The planning reference can be registered to data recorded online in various ways. The most reliable approach is given by volumetric fusion, which exploits intensity changes across a large number of points. Surrogate and surface motion only use limited information and the target may be quite distant to the registered region. It is more favorable to track patient anatomy (e.g. the skin surface) directly instead of external surrogates.

Accuracy Related to the previous category this indicates the uncertainty with which the registration can be performed and the target be located.

Table 1.3: Advantages and disadvantages of different target localization approaches for IGRT.

	portal imaging	CBCT & stereoscopic X-ray	marker-based optical tracking	marker-less optical tracking
Imaging Modality	MV γ-rays	kV X-rays	optical	optical
Imaging Dose	high	low-medium	none	none
Detection Capability	interfractional	inter- and limited intra-fractional	inter- and intra-fractional	inter- and intra-fractional
Registration	volumetric	volumetric	surrogate	surface
Patient Comfort	mask immobilization	mask immobilization	mask or bite bar needed	potentially without immobilization
Accuracy	medium	high	low-medium	medium
Imaging speed	medium	slow-medium	real-time	real-time
Operating Expense	moderate cost, but low impact on workflow	costly, but minimal impact on workflow	inexpensive, but high impact on workflow	inexpensive and low impact on workflow

Imaging Speed Ideally, new information about target motion can be acquired continuously and in real-time. The latter term describes a case in which the imaging speed is faster than typically expected target motion.

Operating Expense On the one hand, this covers the cost for the device and all necessary components. Some components may be already available at the treatment site. Further on, the impact on the clinical workflow is rated: necessary personnel, needed manual interventions or additional steps, necessary patient or marker preparation, manufacturing of customized components etc.

All in all, a general tendency can be seen. While X-ray-based imaging is the most accurate and is used as a gold standard, optical tracking approaches on the other hand are capable of flexible and fast monitoring. They require no extra dose and are typically inexpensive compared to X-ray devices. In fact, marker-less tracking potentially reduces the workflow overhead, since preparation or even the need of immobilization devices can be reduced. However, one of the major drawbacks is given by their limited reliability, robustness and hence accuracy with respect to target localization.

1.3 Purpose of this Work

The previous sections set this work into its broader context. In a nutshell, all subsequent investigations will relate to fractionated external beam therapy of intracranial targets only. This mainly refers to FSRT. In order to ensure precise irradiation in compliance with the treatment plan, localization of the target and tracking of its motion is vital. Section 1.2.2.2 pointed out strengths and weaknesses of different ways for tracking head motion. This thesis will exclusively focus on marker-less optical surface registration, which overcomes some severe drawbacks of other methods such as slow tracking speed or the exposure to additional radiation. Concretely, the work being presented will be dedicated to the main limitation of optical surface tracking: moderate accuracy with respect to target alignment and – along with that – the risk of occasional misregistrations.

Problem Statement Registering surfaces bears the risk of higher tracking errors when in fact targets underneath the surface need to be aligned. Unlike typical volume registration, surface registration separates target and registration site. Therefore, high accuracy demands and a low risk for outliers is required. The two core problems for head tracking and an intuition for the proposed solution are illustrated in fig. 1.12.

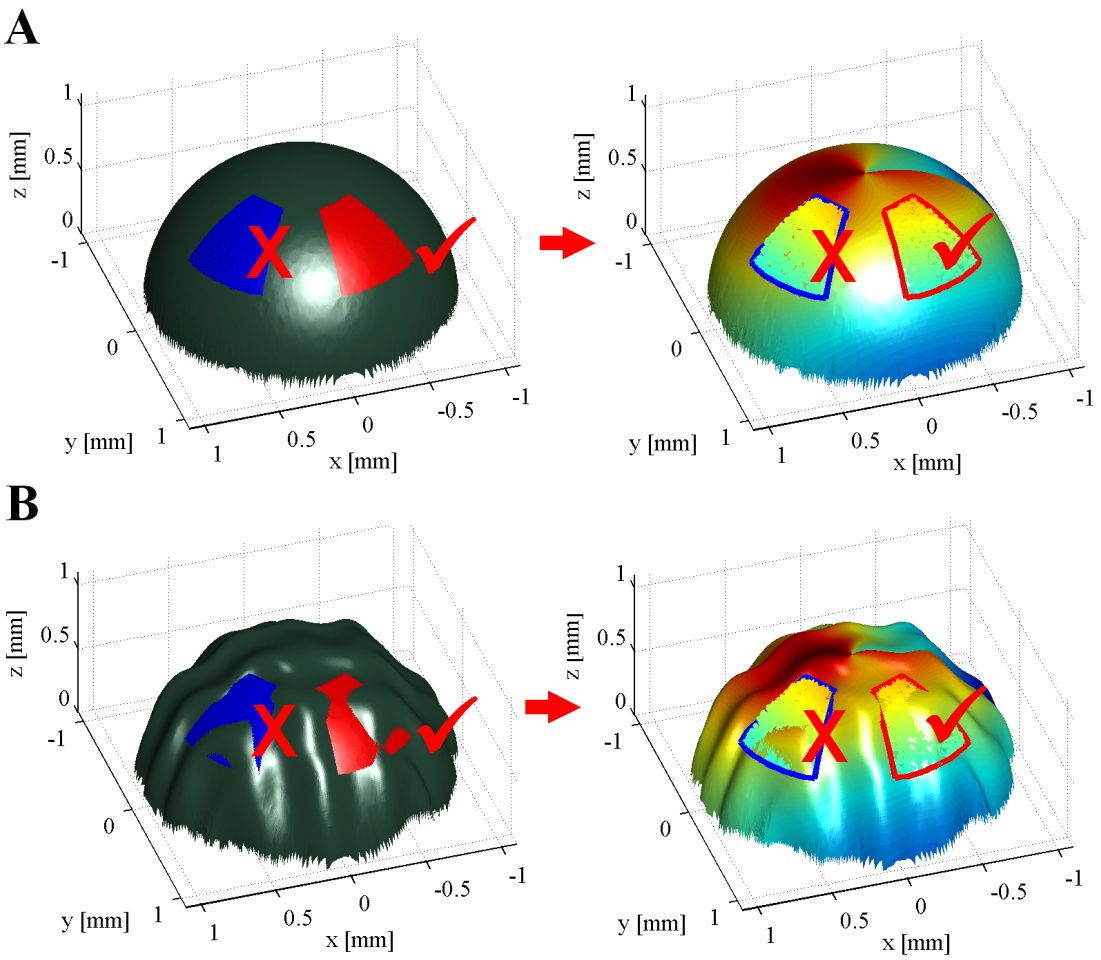


Figure 1.12: Core problems of pure surface registration and the intuition for a possible solution. **A:** A patch scanned from a sphere needs to be registered to the reference (gray object). In this extreme case, any possible alignment on the sphere is a local minimum (e.g. blue location) with the same registration error as the global one, i.e. the correct alignment (red). Additional information denoted by the color patterns is the only way to tell the minima apart. **B:** A patch scanned from a non-rigid surface needs to be registered to the reference (gray object). Deformation of the surface prevents a perfect match. Moreover, the correct alignment (red) has a slightly higher spatial registration error than the (misleading) global optimum (blue). Again the color helps to identify the correct choice. (Reprinted from [331], Copyright (2015), with permission from Springer).

First of all, the registration process is iterative and, depending on the surface characteristics, a highly non-convex optimization. Globally, there may be more than one good fit of a scanned patch to the reference surface. Depending on the initial pose, i.e. the head

motion, the iterative registration may converge into local minima and end up trapped in them. In an extreme case, the problem is ambiguous, in that there are several solutions which are equally good (cf. matching patches to a sphere in fig. 1.12A).
Second, the registration is not rigid for a human head. The soft tissue surface may deform, e.g. after weight loss or even slight changes in the facial expression. The two surfaces to be registered may also come from different modalities. This entails different systematic and random acquisition errors. Therefore, it is very likely that a perfect match does not exist. These influences could even transform the correct alignment into a local minimum and generate a global one somewhere else. This means that the alignment having the lowest surface registration error may not coincide with the correct target alignment anymore. This is illustrated in fig. 1.12B.
This scenario worsens, if the surfaces include only very little landmarks or if the landmarks can only insufficiently be resolved by the scanning process. The forehead region has been recommended for surface tracking since motion artifacts are less severe than in other facial regions [107]. On the other hand, surface scans from that region may still suffer from the aforementioned ambiguity problems due to the lack of landmarks. In FSRT, the thermoplastic mask may also only allow limited access to the patient's face.
This work proposes to use additional surface information for fixing degrees of freedom poorly defined in the registration process. This is denoted by the color overlay in fig. 1.12. This additional dimension rules out ambiguities and introduces additional landmarks. This supports the registration process and increases the likelihood for identifying the correct alignment.

Proposal Figure 1.13 presents a general proposal of how this can be achieved in a clinical scenario. The key idea is to exploit variations in the optical backscatter returned by the optical scan of the skin surface. These changes should contain information about changing optical properties of the skin across the forehead region. One of the sources for backscatter variations is the variation of tissue thickness. One may imagine that thicker skin would reflect less light due to absorption and scattering than very thin skin. The thickness measure also entails the advantage of being available from cranial magnetic resonance (MR) or CT after segmentation, too. Knowledge about this measure would therefore allow to support registration to the planning reference (absolute tracking) and between online optical scans (relative tracking).
This work proposes to generate a model during treatment planning, which can infer tissue thickness from optical backscatter. Using this model, tissue thickness can also be obtained for scans during treatment later on. The tissue labeled scans or the tissue

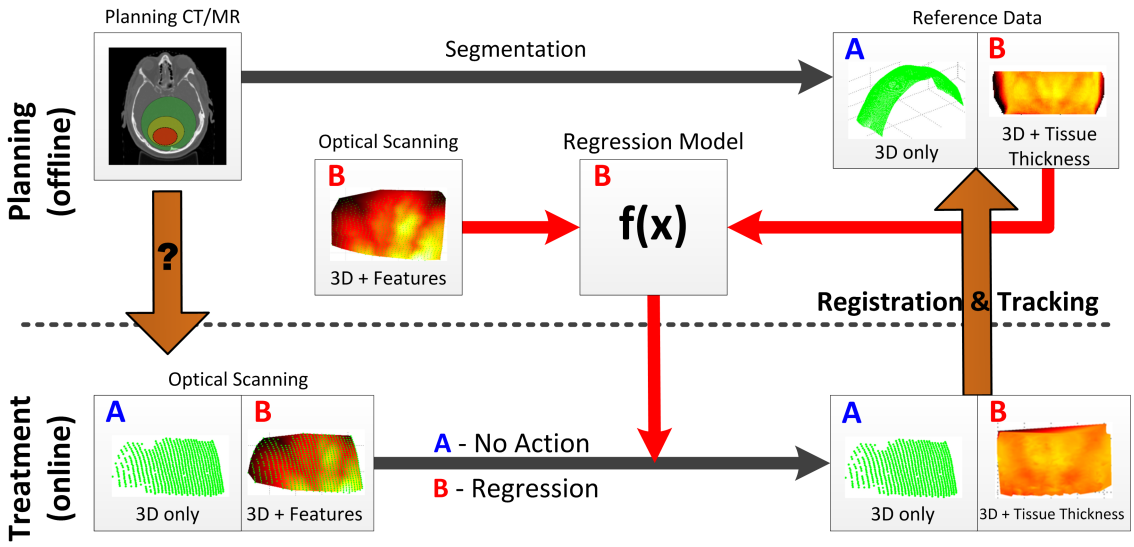


Figure 1.13: Proposal of the enhanced marker-less tracking concept. The additional information in fig. 1.12 will be extracted from recorded optical features. These contain information about the tissue thickness, which can also be segmented from CT or MRI scans. The left orange arrow shows the desired transformation between online scans and planning reference, while the right arrow illustrates how this transformation is computed. The "A" path denotes classical, pure surface registration, and "B" the enhanced proposal. A model which predicts tissue thickness from optical information is generated during planning and applied during treatment. This gives rise to corresponding structural information on reference and online scan, which is then exploited for tracking. (Reprinted from [331], Copyright (2015), with permission from Springer).

labeled planning reference can then be used for enhanced registration and therefore motion tracking.

This work is exclusively concerned with the evaluation of this concept: How can information be retrieved from optical backscatter? What are optimal conditions for this retrieval and what are challenges? How does this information affect the registration process?
The development of optical or related hardware is beyond the scope of this work. Details about this distinct topic are the main concern of the work of Patrick Stüber [281–283, 285] and can be found elsewhere. Insights will only be given and referenced where necessary for the understanding. The same applies for the triangulation software framework. The concept of how to obtain the 3D surface geometry is the main concern of the thesis of

Benjamin Wagner [312] and equivalently beyond the scope of this work.
The main concern here is information retrieval. For enhanced tracking, only a proof of concept will be given. The goal is not to develop dedicated tracking algorithms, but to demonstrate the concept with state-of-the-art surface registration algorithms.

Research Questions This work will approach the proposal in terms of three main research questions. These are summarized with their corresponding sub-questions below.
First, RQ 1 will deal with the theoretical basis for the concept. It will investigate optimal specifications of an optical assembly for acquiring optical features with high information content. Based on simulations, it will be investigated how information is encoded in the backscatter, which disturbance factors have to be considered, and finally how the information can be converted into a tissue thickness scalar.
Second, RQ 2 will evaluate the transfer of these findings into practice. Challenges in a real world scenario will be discussed and the agreement with theoretical findings evaluated. Central to this point is the question of how accurate tissue thickness patterns can be reconstructed and which factors have an influence on this accuracy.
Third, RQ 3 looks into the link between accurate prediction of tissue thickness patterns and the actual benefit for the registration accuracy.

RQ 1	**Is there a valid basis for extracting information about tissue thickness from optical backscatter?**	chapters 3, 6
RQ 1.1	What are the most suitable hardware parameters for an optical setup?	
RQ 1.2	How is information encoded in backscattered light and how can it be optimally translated into informative features?	
RQ 1.3	What are possible disturbance quantities?	
RQ 1.4	How can informative features be used to retrieve a reproducible pattern to support surface tracking?	
RQ 2	**To which extent does the theoretical basis hold for real data recorded from the forehead?**	chapters 4, 6
RQ 2.1	How are the simulation results reflected in real data?	
RQ 2.2	Which statistical learning approach is most suitable for retrieving information about tissue thickness from optical features?	
RQ 2.3	Which disturbances have an influence on the estimation accuracy and how can they be handled?	
RQ 2.4	Are there indications of gender, age or skin type affecting the learning outcome?	
RQ 3	**Does head-tracking gain from incorporating tissue thickness information?**	chapters 5, 6
RQ 3.1	Where are concrete benefits for standard matching algorithms?	
RQ 3.2	Which impact has imperfect data on these benefits?	

1.4 Organization

Following this introductory chapter, there will be five further chapters. At the beginning of each chapter the main content will be summarized and the structure will be outlined. In turn, the end of each chapter will conclude the main ideas discussed in more detail within the actual chapter. Conclusions will be given as answers to the aforementioned research questions. This directly interprets the main findings with respect to the objectives set above.

Chapter 2: Generally, methods and materials will be presented in their corresponding chapters. Apart from that, this second chapter will outline the key methodology of this work. These methods are central for the understanding and appear across multiple sections. A general data processing pipeline will be introduced first, to illustrate the main modules involved in any proposed concept. Subsequently, algorithms for statistical learning namely Support Vector regression and Gaussian processes as well as surface registration algorithms will be introduced. These are options for the modules of the processing pipeline.

Chapter 3: This part will be concerned with answering RQ 1. After reviewing the anatomy and physiology of the forehead and the human skin, approaches for simulating light-tissue interactions will be described. Special focus is directed to Monte-Carlo simulations, which will be used to analyze how information about the tissue thickness is encoded in the optical backscatter. The chapter will define optimal conditions for information retrieval and propose how statistical learning can be used to convert it into a thickness measure. Main disturbance factors will be taken into consideration. Parts of this chapter have been published in [324] and [325].

Chapter 4: The theoretical findings will be investigated on real world data. This will answer RQ 2. Therefore, this chapter starts with describing the optical assembly and framework for data acquisition. Further on, the design of a subject study involving 30 volunteers will be pointed out. This includes an elaboration on the tissue thickness ground truth and how it is obtained from MRI scans.
After investigating general aspects of light-tissue interaction in practice, they will be compared to the simulations. Based on that, the accuracy for predicting tissue thickness from optical features will be studied. Unlike the simulated data, which were exclusively processed with Support Vector regression, convenient properties of Gaussian processes will

also be discussed here. This discussion is extended to remaining challenges, e.g. how to handle larger amounts of data. First light is shed on possible solutions.
Overall, this chapter will cover one of the two major requirements for the enhanced tracking concept: The optimal reconstruction of tissue labeled surfaces from optical backscatter information. Variations and influencing factors within the subject group such as gender, age and skin type will be taken into account as far as possible. Parts of this work have been published in [327–332].

Chapter 5: The second major point of the proposed concept is treated in chapter 5. It answers RQ 3 by investigating the benefits tissue thickness has on the registration performance. A test concept will be introduced which will use a standard algorithm for surface registration and simulated motion on real data. This way, an exhaustive analysis of possible subject motion can be made. A reliable comparison between pure surface tracking and tissue supported tracking can be obtained. Special concern is directed to the question of how prediction errors of the tissue thickness affect the registration performance. In conclusion, it will be discussed in which scenarios tissue support is particularly useful and how it influences the registration process. Parts of this work have been published in [335].

Chapter 6: The final chapter summarizes the main findings and links back to the motivation of this first introductory chapter. It concludes the results in terms of the research questions risen above and assesses the outcome of this work. Remaining challenges will be taken into consideration and directions for future work will be suggested. Alternative fields of application will be indicated.

2 Key Methodology

The following chapters will be kept as self-contained as possible for the sake of readability. That means that materials and methods are only discussed within the chapter they are actually used in. This chapter, however, will make an exception. It will describe and discuss key methodology in detail. Key methods are those being used throughout the entire manuscript and hence being of highest relevance. Almost all subsequent chapters will require background knowledge about the techniques described here. Such an understanding will also facilitate understanding implications of later findings and experimental outcomes.

First, the general processing chain from raw data to informative features and finally tissue thickness predictions is discussed. This will review general conventions and terms of statistical learning and artificial intelligence. The data processing chain is not necessarily unique for the optical data processed here, but will discuss general steps such as pre-processing, feature extraction, feature transformation, regression and interactions thereof.

The second section will detail two specific techniques for statistical learning: Support Vector regression and Gaussian process regression. The subsections will present both, the weight-space as well as the Bayesian view on the regression problem[1]. Finally, the last part will discuss testing methods, error measures and how to estimate the generalization error of a specific technique on given data.

The third and last section will elaborate on point cloud registration. Its relevance is given by the central motivation of this work: how to improve tracking approaches which merely rely on 3D spatial information by incorporating additional information such as tissue thickness distributions. In particular, the Iterative-Closest-Point algorithm will be discussed as the most widely used algorithm for point cloud registration and tracking.

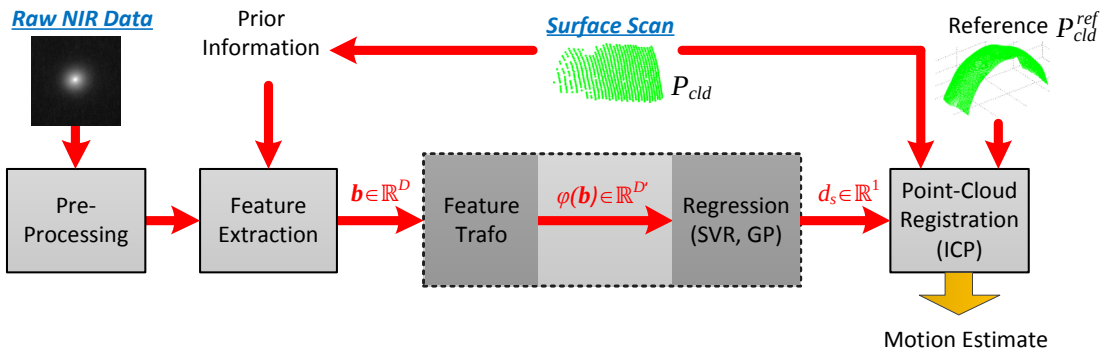


Figure 2.1: General data processing chain. Raw data is first pre-processed, before informative features $\boldsymbol{b}$ can be extracted. Raw data input provided by measurements are underlined and highlighted in blue italics. Regression Analysis relates these to the tissue thickness d_s - possibly after implicitly transforming the features $\boldsymbol{b}$ into a high dimensional space using $\varphi(\cdot)$ (dotted box). The regression output is finally used to support point cloud registration.

2.1 Data Processing Chain and Notation

Machine Learning, as it will be used to approach the main goal of this work, typically entails a general processing chain. The chain will take raw measurement data as an input, extract useful information from it, and learn statistical dependencies within the data. In the context of this work the raw data will be twofold: (1) a 3D point cloud $P_{cld} := \{\boldsymbol{p}_i \in \mathbb{R}^{3\times 1} | \boldsymbol{p}_i = [p_{xi}, p_{yi}, p_{zi}]^T\}_{i=1\ldots N}$, and (2) N 2D images containing optical backscatter data for each point in P_{cld}. Then, the aim of the machine learning is to take the information extracted from backscatter data, combine it with prior information possibly obtained from the experimental conditions, the hardware setup or the point cloud P_{cld}, and finally to relate it to a physiological measure such as tissue thickness. This physiological measure can then be overlaid onto P_{cld}. This overlay is then used for improved and more robust point cloud registration with a reference point cloud P_{cld}^{ref}. Figure 2.1 illustrates these general interrelations. The raw data inputs are highlighted in underlined blue italics. The main blocks of this chain will be briefly discussed in the following.

Pre-Processing This step ensures data consistency and validity before it is used as an input for the next step of the chain. Missing or distorted data are detected and artifacts removed. Data which do not comply with the expected standards are hence discarded.

[1]The distinction between these general categories has been chosen in agreement with [237].

Apart from rejecting invalid data, this step may also provide first data conditioning. In case of the given NIR data, the image is centered around the centroid of the spot and image dimensions are adjusted by appropriately cutting the image size. Details will be discussed in subsequent chapters.

Feature Extraction The pre-processed raw data contains relevant information as well as overhead which is irrelevant for the target quantity. The target quantity in this work is the tissue thickness $d_s \in \mathbb{R}$. It will be referred to as *target label*. Extracting information from pre-processed raw data will break it down to D quantities per target label i. These quantities carry information relevant with respect to the target label and are called *features* $\boldsymbol{b} \in \mathbb{R}^{D\times 1}$. Depending on their information content, features may differ in their relevance.
With features from N data samples the entire input can be summarized with design matrix $\boldsymbol{B} \in \mathbb{R}^{D\times N}$ and target label vector $\boldsymbol{d}_s \in \mathbb{R}^{1\times N}$. This gives rise to the input data set $\mathcal{D}$

$$\mathcal{D} = \{(\boldsymbol{b}_i, d_{si})\}_{i=1\ldots N} = (\boldsymbol{B}, \boldsymbol{d}_s) \quad \text{with} \quad \mathcal{D}\{\in \mathbb{R}^{D\times N} \times \mathbb{R}^{1\times N}\}. \tag{2.1}$$

The N samples are distributed in an D dimensional feature space in this case. Parts of the variance in this feature space are due to covariation with the target label.

Feature Transformation Very often it is possible to transform these samples into another – in most cases higher dimensional – space. This may facilitate modeling a functional relationship between features and labels, particularly when this relationship is complex and highly nonlinear. For reasons explained later on, this new space will be called kernel space of new dimension D' with typically $D < D'$ [255]. The transformation is as follows:

$$\boldsymbol{\varphi} : \mathbb{R}^D \longmapsto \mathbb{R}^{D'} \tag{2.2}$$

Earlier definitions can be extended to an adapted design matrix $\boldsymbol{\Phi} \in \mathbb{R}^{D'\times N}$. The transformation is hardly ever done explicitly, but implicitly by the statistical learning approach using the so-called kernel trick [255]. This is indicated by the dotted box in fig. 2.1.

Statistical Learning Statistical learning has two categories: classification and regression. While classification maps the input features to a finite set of classes, regression tries to learn a continuous function f mapping features to labels.

$$f : \mathbb{R}^{D'} \longmapsto \mathbb{R} \tag{2.3}$$

Thus, regression generalizes classification as it maps to an infinite number of classes. Function f is learned by minimizing the deviation between given target labels $\boldsymbol{d}_s$ and the target label predictions $\boldsymbol{d}_s^\star$ obtained from the current hypothesis for f. This is called supervised learning. In the context of supervised learning, target labels are also referred to as ground truth. Predicted target labels $\boldsymbol{d}_s^\star$ and input features $\boldsymbol{b}_\star$ for this prediction will be labeled with a star $\star$.
Various regression techniques have been published up to date [122]. Most of them can be traced back to a data model, where each measurement of a target label d_s is corrupted by noise $n \in \mathbb{R}$ and where the underlying function f can be expressed in terms of a weighted superposition of basis functions $\boldsymbol{\varphi}$. The weight vector is given by $\boldsymbol{w}$ and assumptions about the nature of n may vary by learning technique.

$$d_s = f(\boldsymbol{b}) + n = \boldsymbol{w}^T \boldsymbol{\varphi}(\boldsymbol{b}) + n = \langle \boldsymbol{w}, \varphi(\boldsymbol{b}) \rangle + n \tag{2.4}$$

Here $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors. This work will employ two regression techniques namely Support Vector regression (SVR) and Gaussian processes (GPs). This choice is not exclusive and there are other approaches which may work equally well. Nevertheless, both regression techniques are quite explicit in modeling the data and their optimization aims at a complexity-accuracy tradeoff. Further on, it will be shown that GPs can be understood as a non-parametric generalization of many other approaches.

General Notation As done in this first part of the chapter, notation of mathematical terms will be kept as follows. Scalar values or functions with a scalar output, i.e. values in $\mathbb{R}^1$, are denoted by small letters in normal font. Bold font and small letters are used for vectors or functions with a vector output in $\mathbb{R}^{n\times 1}, n > 1$. Finally, matrices e.g. in $\mathbb{R}^{n\times m}, n > 1, m > 1$ are labeled by capital letters and bold font. Let f be a function with scalar output, then $f(\boldsymbol{b}_i)$ is the functional value of f at $\boldsymbol{b}_i$ and $\boldsymbol{f}(\boldsymbol{B})$ with $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N]$ are the functional values of f at all $\boldsymbol{b}_i$ concatenated into a vector. Furthermore, let $\boldsymbol{\varphi}$ be a function with vector output, then $\boldsymbol{\varphi}(\boldsymbol{b}_i)$ is the vector output of $\boldsymbol{\varphi}$ at $\boldsymbol{b}_i$ and $\boldsymbol{\Phi}(\boldsymbol{B})$ with $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N]$ are the vector outputs of $\boldsymbol{\varphi}$ at all $\boldsymbol{b}_i$ concatenated into a matrix.

Table 2.1: Overview of common regression techniques. SVR together with other regression approaches is categorized in the context of loss and penalty functions.

Approach	Loss Function L_*	Penalty Term $\mathfrak{P}$
Least-Squares Fit	$L_2 = \sum_i (d_{si} - d_{si\star})^2$	none
Least Absolute Deviations [30]	$L_1 = \sum_i \lvert d_{si} - d_{si\star} \rvert$	none
Ridge Regression [126]	$L_2 = \sum_i (d_{si} - d_{si\star})^2$	$\lVert \boldsymbol{w} \rVert_2^2$
LASSO [290]	$L_2 = \sum_i (d_{si} - d_{si\star})^2$	$\lVert \boldsymbol{w} \rVert_1$
Elastic Net [345]	$L_2 = \sum_i (d_{si} - d_{si\star})^2$	$\lVert \boldsymbol{w} \rVert_1$ and $\lVert \boldsymbol{w} \rVert_2^2$
SVR	L_ε	$\lVert \boldsymbol{w} \rVert_2^2$

2.2 Statistical Learning Techniques

2.2.1 Support Vector Regression

2.2.1.1 The Weight-space View.

The weight-space view on regression looks for a function f which fits the labeled data $\mathcal{D}$ best under certain assumptions. This is achieved by constructing an optimization problem or, more precisely, by minimizing a structural risk functional $\mathfrak{R}$. This functional consists of two parts: (1) the empirical risk expressed by a loss function L_*, and (2) a penalty term $\mathfrak{P}$ for regularizing the complexity of the solution:

$$\min_f(\mathfrak{R}) = \min_f(C \cdot L_*(f) + \mathfrak{P}(f)) \tag{2.5}$$

where C is the regularization constant which balances the optimization problem between loss and complexity penalty, i.e. the larger C, the more weight is put on minimizing the distance between the function f and the given data set $\mathcal{D}$. A small C, on the other hand, leads to very smooth functions. In the simplest case $\mathfrak{P}$ is a function of the weights $\boldsymbol{w}$ and penalizes high values within $\boldsymbol{w}$.
The loss function is a function of the differences between predicted labels $d_{s\star}$ and the corresponding ground truth d_s. Its choice varies by regression technique and may depend on the expected noise distribution or on whether a sparse set of basis functions is desired or not. Typical examples are shown in fig. 2.2A such as *L2*-loss (used by least-squares or ridge regression), *L1*-loss (used by least absolute deviations), or ε-insensitive loss L_ε

(used by SVR). Table 2.1 compares common regression techniques according to their structural risk approach and sets SVR into their context.

Just like ridge regression, SVR employs the data model given in eq. 2.4 combined with ε-insensitive loss and an *L2* penalty on the weights $\boldsymbol{w}$. As illustrated in fig. 2.2A, the loss function L_ε would assign zero loss for samples within an ε-tube around the fitted function f. This implies that the deviation of a predicted value from the ground truth label is at most ε. These deviations are handled in terms of slack variables $\xi_i^{(*)}$. While the slack for deviations smaller than ε is zero, it grows linearly with this deviation, starting with zero at deviation ε (cf. eq. 2.6 and fig. 2.2B). Here $\xi_i^{(*)}$ is a collective term for both ξ_i and ξ_i^*, which refer to cases where either the prediction was larger or smaller than the ground truth, respectively.

$$\xi_i^{(*)} := \begin{cases} 0 & \text{if } |d_{si} - \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle - w_0| \leq \varepsilon \\ |d_{si} - \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle - w_0| - \varepsilon & \text{otherwise} \end{cases} \tag{2.6}$$

Here w_0 is the explicit constant offset originating from $f(\boldsymbol{b}_i) = d_{si\star} = \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle + w_0$ (eq. 2.4 absorbed the offset into $\boldsymbol{w}$). The overall loss L_ε for a function f fitted to N data samples then corresponds to:

$$L_\varepsilon(f) = \sum_{i=1}^{N} (\xi_i + \xi_i^*) \tag{2.7}$$

2.2.1.2 The SVR-Model

The L_ε loss is not a smooth function. Therefore, the resulting structural risk

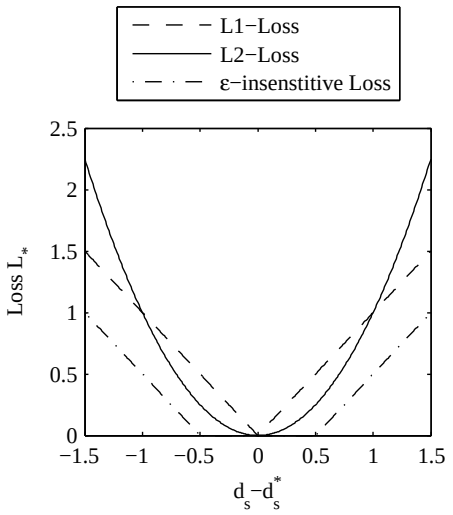
$$\mathfrak{R} = C \cdot L_\varepsilon(f) + \mathfrak{P}(f)$$

needs to be expressed in terms of a constrained optimization problem.

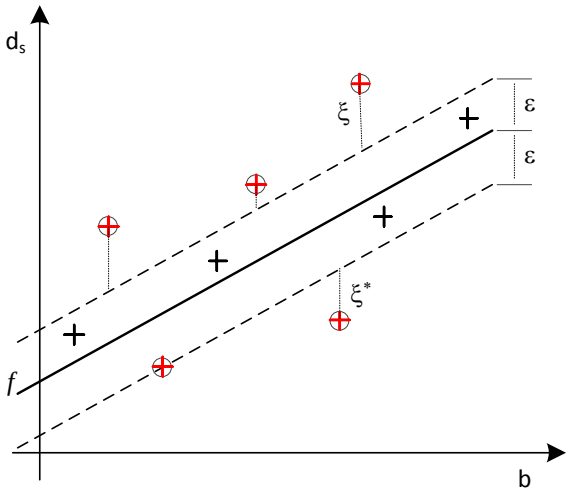
$$\min_{\boldsymbol{w}} \left[C \sum_{i=1}^{N} (\xi_i + \xi_i^*) + \frac{1}{2} \|\boldsymbol{w}\|_2^2 \right] \tag{2.8}$$

$$\text{subject to} \begin{cases} d_{si} - \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle - w_0 \leq \varepsilon + \xi_i \\ \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle + w_0 - d_{si} \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

This problem is convex and equivalent to a quadratic programming (QP) problem. There are no local minima and the same parameter set will always result in the same solution



A: The ε-insensitive loss function L_ε and others in comparison.



B: The ε-tube around a function f fitted to data samples. Support Vectors are circled in red.

Figure 2.2: Data model and concept of SVR.

irrespective from the initial seed. The solution, i.e. minimum, to this primal convex problem is equivalent to the maximum of the Lagrange function $\mathcal{L}(\boldsymbol{w}, w_0, \xi_i^{(*)}, \alpha_i^{(*)}, \eta_i^{(*)})$ which joins the objective function and constraints of the primal into one objective [268].

$$\begin{aligned}\mathcal{L} = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\alpha_i(\varepsilon + \xi_i - d_{si} + \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle + w_0) \\ -\sum_{i=1}^{N}\alpha_i^*(\varepsilon + \xi_i^* - \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle - w_0 + d_{si}) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*)\end{aligned} \tag{2.9}$$

The Lagrange multipliers[2] or dual variables need to fulfill a positivity constraint $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$. Furthermore, the derivatives of the Lagrangian with respect to the

[2]The Lagrange multipliers can be interpreted as positive scaling factors needed to achieve equality between the gradients of the objective and the corresponding constraint at optimality.

primal variables vanish for optimality:

$$\frac{\partial \mathcal{L}}{\partial w_0} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) \stackrel{!}{=} 0 \tag{2.10a}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \boldsymbol{w} + \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\boldsymbol{\varphi}(\boldsymbol{b}_i) \stackrel{!}{=} 0 \tag{2.10b}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)} \stackrel{!}{=} 0 \tag{2.10c}$$

Inserting eq. 2.10b and eq. 2.10c into eq. 2.9 yields the dual description of the the QP problem.

$$\max_{\alpha_i^{(*)}} \left[\frac{1}{2}\sum_{i,j}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\left\langle \boldsymbol{\varphi}(\boldsymbol{b}_i), \boldsymbol{\varphi}(\boldsymbol{b}_j)\right\rangle - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} d_{si}(\alpha_i - \alpha_i^*) \right] \tag{2.11}$$

$$\text{subject to } \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0$$

Since the primal optimization problem has no equality, but inequality constraints, the Karush-Kuhn-Tucker (KKT) conditions have to be met when applying the Lagrangian[3]. These require that the products of dual variables $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ and the primal constraints have to be zero. The reason is that the multipliers need to vanish for points lying not at the constraint boundary, i.e. for which the problem is locally unconstrained [208].

$$\alpha_i(\varepsilon + \xi_i + \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle + w_0 - d_{si}) = 0 \tag{2.12a}$$

$$\alpha_i^*(\varepsilon + \xi_i^* - \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i)\rangle - w_0 + d_{si}) = 0 \tag{2.12b}$$

$$\underbrace{(C - \alpha_i)}_{\eta_i}\xi_i = 0 \tag{2.12c}$$

$$\underbrace{(C - \alpha_i^*)}_{\eta_i^*}\xi_i^* = 0 \tag{2.12d}$$

The dual problem and these conditions finally directly give rise to the following important observations.

1. Due to $\eta_i^{(*)} = C - \alpha_i^{(*)}$ and $\eta_i^{(*)} > 0$ we obtain the box constraints $\alpha_i^{(*)} \in [0, C]$.

[3] Strictly speaking the KKT conditions generalize the concept of Lagrange multipliers to inequality constraints.

2. The function $\boldsymbol{\varphi}(\boldsymbol{b}_i)$ only occurs within the inner product $\langle \boldsymbol{\varphi}(\boldsymbol{b}_i), \boldsymbol{\varphi}(\boldsymbol{b}_j) \rangle$.
3. For samples outside the ε-tube we have $\xi_i^{(*)} \neq 0$ and hence $\alpha_i^{(*)} = C$ from eq. 2.12c and eq. 2.12d.
4. From eq. 2.12a, eq. 2.12b and $\xi_i \xi_i^* = 0$ follows $\alpha_i \alpha_i^* = 0$.
5. $\alpha_i \in (0, C)$ and $\xi_i = 0$ yield $w_0 = d_{si} - \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i) \rangle - \varepsilon$.
6. $\alpha_i^* \in (0, C)$ and $\xi_i^* = 0$ yield $w_0 = d_{si} - \langle \boldsymbol{w}, \boldsymbol{\varphi}(\boldsymbol{b}_i) \rangle + \varepsilon$.

2.2.1.3 Kernel Trick

Since $\boldsymbol{\varphi}(\boldsymbol{b}_i)$ appears only in form of an inner product with itself, kernel functions $k(\boldsymbol{b}_i, \boldsymbol{b}_j)$ can be introduced [255].

$$k(\boldsymbol{b}_i, \boldsymbol{b}_j) = \langle \boldsymbol{\varphi}(\boldsymbol{b}_i), \boldsymbol{\varphi}(\boldsymbol{b}_j) \rangle \qquad \text{with} \qquad k : \mathbb{R}^D \times \mathbb{R}^D \longmapsto \mathbb{R} \tag{2.13}$$

This means it is not necessary to know $\boldsymbol{\varphi}(\boldsymbol{b}_i)$ explicitly. It suffices to have an analytic expression of how to compute the inner product in its output space. This implicit feature transformation, i.e. without explicitly computing it, is efficient and saves storage space. Valid kernel functions have to fulfill Mercer's theorem [196] which requires symmetric and positive definite functions k. While for some k the corresponding $\boldsymbol{\varphi}$ can be obtained, this is not necessarily possible for all k. One example is the radial basis function (RBF) kernel function which will be described in the next section. Here $\boldsymbol{\varphi}$ can be expanded into an infinite series, which would correspond to a mapping into an infinite dimensional space. This property is called non-degenerate.

For a finite set of data samples $\boldsymbol{b}_i$ the Gram matrix $\boldsymbol{K}$ can be defined as

$$\boldsymbol{K}(\boldsymbol{B}, \boldsymbol{B}) = [k(\boldsymbol{b}_i, \boldsymbol{b}_j)]_{i=1\ldots N, j=1\ldots N} = \begin{bmatrix} k(\boldsymbol{b}_1, \boldsymbol{b}_1) \ldots k(\boldsymbol{b}_1, \boldsymbol{b}_N) \\ \vdots \qquad \ddots \qquad \vdots \\ k(\boldsymbol{b}_N, \boldsymbol{b}_1) \ldots k(\boldsymbol{b}_N, \boldsymbol{b}_N) \end{bmatrix}. \tag{2.14}$$

This matrix is also called kernel or covariance matrix in the context of GPs.

2.2.1.4 Support Vectors

Either α_i or α_i^* is always zero. Both are zero only for samples which fall within the ε-tube. Further follows that samples exactly located on the tube have corresponding multipliers in the open interval $\alpha_i^{(*)} \in (0, C)$. Samples outside the tube always have $\alpha_i^{(*)} = C$ (see

fig. 2.2).
All samples $\boldsymbol{b}_i$ for which $\alpha_i^{(*)} \neq 0$ are called Support Vectors (SVs). With eq. 2.10b the weight vectors equals:

$$\boldsymbol{w} = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \boldsymbol{\varphi}(\boldsymbol{b}_i) \tag{2.15}$$

The final prediction formula can be obtained based on the relationship between the coefficients $\alpha_i^{(*)} \in [0, C]$ and the SVs, eq. 2.4 and eq. 2.15.

$$\boxed{f_\star = f(\boldsymbol{b}_\star) = \sum_{i \in SV} (\alpha_i - \alpha_i^*) \langle \boldsymbol{\varphi}(\boldsymbol{b}_i), \boldsymbol{\varphi}(\boldsymbol{b}_\star) \rangle + w_0 = \sum_{i \in SV} (\alpha_i - \alpha_i^*) k(\boldsymbol{b}_i, \boldsymbol{b}_\star) + w_0} \tag{2.16}$$

Equation 2.16 is called Support Vector expansion, since the output of function value $f_\star$ for an unknown sample $\boldsymbol{b}_\star$ is expressed by a weighted sum of its similarities to the samples $\boldsymbol{b}_i$ from the set of known SVs. The inner product described by the kernel function can thus be interpreted as a (possibly nonlinear) similarity measure in a high dimensional space. As mentioned before, input features $\boldsymbol{b}_\star$ with unknown target label are labeled by a star $\star$, since they are subject to prediction.

2.2.1.5 Optimization

The optimization problem can be solved with any QP solver. A stopping criterion for these iterative algorithms is given by the duality gap (difference between $\mathcal{L}$ and $\mathfrak{R}$ which vanishes for optimality) or the fulfillment of the KKT conditions.
In this work the sequential minimal optimization (SMO) algorithm introduced by Platt [228] has been used. This sequential solver is dedicated to the problem at hand and iterates through the coefficients $\alpha_i^{(*)}$ and optimizes $\mathcal{L}$ for a pair of two coefficients in a step-wise manner. The implementation was adopted from [53]. There also exist solvers which can adapt an existing model by adding new data to it [181].

2.2.2 Gaussian Process Regression

2.2.2.1 The Bayesian View

Section 2.2.1 introduced the weight-space view on regression. This subsection will show how a probabilistic – so-called Bayesian – point of view generalizes the concept of structural risk minimization with loss and penalty term. This will finally lead to the definition of GPs.

The Bayesian view relies on not more than three fundamental equations for the probabilities of some events a and b: $p(a, b)$ ("a and b"), $p(a \cup b)$ ("a or b"), and $p(a|b)$ ("a given b"):

$$p(a,b) = p(a) \cdot p(b) \quad \text{(product rule)} \tag{2.17}$$

$$p(a \cup b) = p(a) + p(b) \quad \text{(sum rule)} \tag{2.18}$$

$$p(a|b) = \frac{p(b|a) \cdot p(a)}{p(b)} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}} \quad \text{(Bayes' rule)} \tag{2.19}$$

The product rule is valid for statistical independence between a and b. In Bayes' rule, the *prior* expresses the prior belief into how a may be distributed. The likelihood describes how a behaves given b[4] and the marginal likelihood gives the likelihood where b has been integrated ("marginalized") out. The term "integrating out" refers to the fact, that a variable can be eliminated from a distribution by integrating this distribution over all possible values for this variable.

Bayesian regression adopts the data model from eq. 2.4 with the special case of noise n drawn from a Gaussian distribution $\mathcal{N}$ with zero mean and standard deviation σ_n^2:

$$d_s = f(\boldsymbol{b}) + n = \boldsymbol{w}^T \boldsymbol{\varphi}(\boldsymbol{b}) + n \tag{2.20}$$

$$n \sim \mathcal{N}(0, \sigma_n^2) \tag{2.21}$$

Since the distribution of $n_i = d_{si} - \boldsymbol{w}^T \boldsymbol{\varphi}(\boldsymbol{b}_i)$ is known, the probability for a value d_{si} given input data $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N]$ and some weight vector $\boldsymbol{w}$ can be obtained as follows (being the likelihood for $\boldsymbol{w}$):

$$p(n) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(d_{si} - \boldsymbol{w}^T \boldsymbol{\varphi}(\boldsymbol{b}_i))^2}{2\sigma_n^2}\right) := p(d_{si}|\boldsymbol{b}_i, \boldsymbol{w}) \tag{2.22}$$

and given all (statistically independent) training data $\boldsymbol{B}$:

$$p(\boldsymbol{d}_{si}|\boldsymbol{B}, \boldsymbol{w}) = \prod_{i=1}^{N} p(d_{si}|\boldsymbol{b}_i, \boldsymbol{w}) = \frac{1}{(2\pi\sigma_n^2)^{\frac{N}{2}}} \exp\left(-\frac{|\boldsymbol{d}_{si} - \boldsymbol{w}^T \boldsymbol{\Phi}(\boldsymbol{B})|^2}{2\sigma_n^2}\right) \tag{2.23}$$

[4] Although the probability expresses the event b given a, the term likelihood always denotes a function $a = f(b)$ for such a distribution.

In addition to this likelihood, the belief about the prior distribution of the weights is also assumed to be a Gaussian distribution:

$$\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_w) \tag{2.24}$$

This assumes that the weights are normally distributed around zero and that large weights are very unlikely. The prior is equivalent to regularization or a penalty term. The prior belief can vary for different regression techniques, different data or requirements.
Applying Bayes' rule results in the posterior distribution $p(\boldsymbol{w}|\boldsymbol{d}_{si}, \boldsymbol{B})$, i.e. the belief about which values for $\boldsymbol{w}$ are how likely after having seen the data.

$$p(\boldsymbol{w}|\boldsymbol{d}_{si}, \boldsymbol{B}) = \frac{p(\boldsymbol{d}_{si}|\boldsymbol{B}, \boldsymbol{w}) \cdot p(\boldsymbol{w})}{p(\boldsymbol{d}_{si}|\boldsymbol{B})} \tag{2.25}$$

The marginal likelihood in the denominator does not depend on the weights and constitutes a constant normalization factor. Since the posterior is subject to optimization in order to find an optimal $\boldsymbol{w}$, this constant factor can be dropped. It does not change the optimization result. Using eq. 2.23 and eq. 2.24, the product in the numerator therefore yields (after rearranging by "completing the square" and dropping further constant factors) [237]:

$$p(\boldsymbol{w}|\boldsymbol{d}_s, \boldsymbol{B}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{w} - \bar{\boldsymbol{w}})^T(\sigma_n^{-2}\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \boldsymbol{\Sigma}_w^{-1})(\boldsymbol{w} - \bar{\boldsymbol{w}})\right) \tag{2.26}$$

where $\bar{\boldsymbol{w}} = \sigma_n^{-2}(\sigma_n^{-2}\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \boldsymbol{\Sigma}_w^{-1})^{-1}\boldsymbol{\Phi}\boldsymbol{d}_s^T$. Equation 2.26 corresponds to a standard normal distribution with mean $\bar{\boldsymbol{w}}$ and covariance $\boldsymbol{A}^{-1}$:

$$\boxed{p(\boldsymbol{w}|\boldsymbol{d}_{si}, \boldsymbol{B}) \propto \mathcal{N}(\bar{\boldsymbol{w}}, \boldsymbol{A}^{-1}) = \mathcal{N}(\sigma_n^{-2}\boldsymbol{A}^{-1}\boldsymbol{d}_s^T\boldsymbol{\Phi}, \boldsymbol{A}^{-1})} \tag{2.27}$$

where $\boldsymbol{A} = \sigma_n^{-2}\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \boldsymbol{\Sigma}_w^{-1}$. The distribution above is fundamental, since many common regression techniques can be derived from it. To obtain a regression model such as eq. 2.4 most approaches approximate the weight vector $\boldsymbol{w}$ by its maximum a posteriori (MAP) estimate $\bar{\boldsymbol{w}}$. This estimate is computed by maximizing the negative logarithm of eq. 2.27 with respect to $\boldsymbol{w}$.
Setting $\boldsymbol{\Sigma}_w$ to the identity matrix $\boldsymbol{I}$ yields the analytic solution for ridge regression. Note that the regularization constant C then corresponds to the inverse of the noise variance σ_n^{-2}. Further, setting it to a diagonal matrix the elements of which approach infinity yields

Table 2.2: Bayesian view of common regression techniques. By re-interpreting the columns of table 2.1 the table shows how common regression techniques derive from the Bayesian framework.

Approach	Likelihood	Prior
Least-Squares Fit	Gaussian noise	flat
Least Absolute Deviations [30]	Laplacian noise	flat
Ridge Regression [126]	Gaussian noise	Gaussian
LASSO [290]	Gaussian noise	Laplacian
Elastic Net [345]	Gaussian noise	sum of Laplacian and Gaussian
SVR	$\propto \exp(-L_\varepsilon)$	Gaussian

the least squares solution. The latter case is equivalent to assuming a flat, i.e. uninformative prior distribution over the weights in eq. 2.24: $\boldsymbol{w} \sim \mathcal{U}(-\infty, +\infty)$.
Table 2.2 demonstrates that all regression techniques from table 2.1 can be derived as special cases from the Bayesian framework by assuming distributions for the noise/likelihood in eq. 2.23 or prior distribution over the weights in eq. 2.24.
Parameters such as the variance of the measurement noise σ_n^2 as well as parameters of the kernel function $k(\cdot,\cdot)$ or Gram matrix $\boldsymbol{K}(\cdot,\cdot)$ are called hyperparameters. The belief about their values could be expressed by further priors, i.e. priors on the priors. Along these lines further techniques can be derived as by Tipping et al. [291]. He set a Gamma distribution prior on the variance of weights $\boldsymbol{w}$ and obtained the so-called Relevance Vector machine.
With the MAP estimate, techniques such as ridge regression select only **one** possible solution for the weight vector: the most likely one given the data. In contrast, GPs average across **all** possible solutions for $\boldsymbol{w}$ according to their probability. The equations above yield the predictive distribution, i.e. the probability distribution $p(f_\star|\boldsymbol{b}_\star, \boldsymbol{B}, \boldsymbol{d}_s)$ for an unknown functional value $f_\star$ associated to features $\boldsymbol{b}_\star$. The following consideration uses eq. 2.20 to define $f_\star = \boldsymbol{w}^T\boldsymbol{\varphi}(\boldsymbol{b}_\star)$. Given $\boldsymbol{b}_\star$ and $\boldsymbol{w}$, $f_\star$ can only take one possible value the probability of which is one. The predictive distribution is hence obtained by scaling the argument $\boldsymbol{w}$ in eq. 2.27 by $\boldsymbol{\varphi}(\boldsymbol{b}_\star)$.

$$\begin{aligned} p(f_\star|\boldsymbol{b}_\star, \boldsymbol{B}, \boldsymbol{d}_s) &= \int p(f_\star|\boldsymbol{b}_\star, \boldsymbol{w}) p(\boldsymbol{w}|\boldsymbol{B}, \boldsymbol{d}_s) d\boldsymbol{w} \\ &= \mathcal{N}\left(\sigma_n^{-2}\boldsymbol{\varphi}(\boldsymbol{b}_\star)^T\boldsymbol{A}^{-1}\boldsymbol{\Phi}\boldsymbol{d}_s, \boldsymbol{\varphi}(\boldsymbol{b}_\star)^T\boldsymbol{A}^{-1}\boldsymbol{\varphi}(\boldsymbol{b}_\star)\right) \\ &= \mathcal{N}\left(\bar{f}_\star, var[f_\star]\right) \end{aligned} \tag{2.28}$$

The Gram matrix $\boldsymbol{K}(\cdot,\cdot) = \boldsymbol{\Phi}^T\boldsymbol{\Sigma}_w\boldsymbol{\Phi}$ is introduced analogously to sec. 2.2.1. Thus, the kernel trick can be used here as well to avoid computing the feature transformation $\boldsymbol{\varphi}(\boldsymbol{b})$ explicitly.
This is done by noting that $\boldsymbol{A} = \sigma_n^{-2}\boldsymbol{\Phi}(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}_w^{-1}$ and replacing $\boldsymbol{A}$ in eq. 2.29. As a shorthand $\boldsymbol{\varphi}(\boldsymbol{b}_\star) = \boldsymbol{\varphi}_\star$ is used.

$$\begin{aligned} p(f_\star|\boldsymbol{b}_\star, \boldsymbol{B}, \boldsymbol{d}_s) &= \mathcal{N}\big(\boldsymbol{\varphi}_\star^T\boldsymbol{\Sigma}_w\boldsymbol{\Phi}(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{d}_s, \\ &\quad \boldsymbol{\varphi}_\star^T\boldsymbol{\Sigma}_w\boldsymbol{\varphi}_\star - \boldsymbol{\varphi}_\star^T\boldsymbol{\Sigma}_w\boldsymbol{\Phi}(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Sigma}_w\boldsymbol{\varphi}_\star\big) \qquad (2.29) \\ &= \mathcal{N}\big(\boldsymbol{k}_\star(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{d}_s, \\ &\quad k(\boldsymbol{b}_\star, \boldsymbol{b}_\star) - \boldsymbol{k}_\star(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{k}_\star\big) \qquad (2.30) \end{aligned}$$

The last expression provides the full predictive distribution, i.e. indicates how likely a specific prediction output $f_\star$ is. In this work the mean $\bar{f}_\star$ of this distribution will be used as the prediction output. The predictive variance $var[f]_\star$ then provides an uncertainty measure for this prediction. The mean $\bar{f}_\star$ can finally be written analogously to SVR as a weighted superposition of similarities between the training data and the desired prediction output. Note that a kernel takes the features as an input, but outputs the covariance between the corresponding target labels.

$$\boxed{\begin{aligned} \bar{f}_\star &= \boldsymbol{k}_\star \underbrace{(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{d}_s}_{\boldsymbol{\alpha}=[\alpha_1,\ldots\alpha_N]} \\ \bar{f}_\star &= \sum_{i=1}^{N}\alpha_i \cdot k(\boldsymbol{b}_i, \boldsymbol{b}_\star) \end{aligned}} \qquad (2.31)$$

For the computation of the prediction above, the inverse of $\boldsymbol{K} \in \mathbb{R}^{N\times N}$ needs to be computed. The computational effort is of order $\mathcal{O}(N^3)$, which becomes prohibitive for large data sets, i.e. large N. The framework above has been implemented in the `gpml` toolbox [238], which has also been adopted for this work. Note that for computing the error measures, it will be assumed that $\bar{d}_{s\star} = \bar{f}_\star$. This is true for any noise distribution with zero mean and enables to compare predictions $f_\star$ with a possible noisy ground truth d_s. On average there is no difference between them.

2.2.2.2 Function Space View and Gaussian Processes

A different view on the previous considerations is the function-space view. For completeness it will be briefly described in the following. According to the data model, each d_{si} can be interpreted as a sample from a Gaussian distribution with some mean and variance.

Having more than one sample $\boldsymbol{d}_s \in \mathbb{R}^N$ results in a multivariate Gaussian distribution with a mean vector and the Gram matrix $\boldsymbol{K}$ as the covariance matrix, i.e. the behavior of one sample relates to the behavior of others. Increasing the number of samples N to infinity lets the vector change into a function f. The multivariate Gaussian distribution then turns into a Gaussian process – a distribution over functions with mean function $m_{GP}(\cdot)$ and covariance function $k(\cdot, \cdot)$.

$$f(\boldsymbol{b}) \sim \mathcal{GP}(m_{GP}(\boldsymbol{b}), k(\boldsymbol{b}, \boldsymbol{b}')) \tag{2.32}$$

This is often called non-parametric regression, since there exists no explicit weight vector. The process above describes the prior belief about how the desired function might look like and which properties it has. The mean function is typically set to zero and all variation is modeled by the covariance.
If data $\boldsymbol{B}$ and $\boldsymbol{d}_s$ are observed, information of f at certain points is received. Other points $f_\star$ at position $\boldsymbol{b}_\star$ are not known and need to be predicted. This scenario, with interest in only a finite set of samples from f, can be expressed as a multivariate Gaussian distribution again.

$$p(\boldsymbol{d}_s, f_\star) = p\left(\begin{bmatrix}\boldsymbol{d}_s \\ f_\star\end{bmatrix}\right) = \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix}\boldsymbol{K}(\boldsymbol{B}, \boldsymbol{B}) + \sigma_n^2 \boldsymbol{I} & \boldsymbol{k}_\star = \boldsymbol{k}(\boldsymbol{B}, \boldsymbol{b}_\star) \\ \boldsymbol{k}_\star = \boldsymbol{k}(\boldsymbol{B}, \boldsymbol{b}_\star) & k(\boldsymbol{b}_\star, \boldsymbol{b}_\star)\end{bmatrix}\right) \tag{2.33}$$

After re-arranging the exponent of this multivariate distribution and integrating out $\boldsymbol{f}$ (see the appendix in [237] for details), it is possible to show that eq. 2.30 and eq. 2.31 follow from these assumptions.

2.2.2.3 Linear Example

A small linear example in 1D shall illustrate the aforementioned framework. Measurement units will be omitted for simplicity. Let the true underlying function f be linear with slope $w_1 = 0.5$ and offset $w_0 = 1$.

$$f(b) = w_1 \cdot b + w_0 = 0.5 \cdot b + 1 \tag{2.34}$$

Now, $N = 23$ measurements $\boldsymbol{d}_s \in \mathbb{R}^{1 \times N}$ at positions $\boldsymbol{b} \in \mathbb{R}^{1 \times N}$ are acquired. The b_i are sampled equidistantly within $[-1, 10]$ in steps of 0.5. The corresponding target labels d_{si} have been obtained by computing $f(b_i)$ and corrupting it with Gaussian noise ($\sigma_n^2 = 3$). Figure 2.3 illustrates the true function as a black line and the data samples in green.
The prior distribution on the weights in eq. 2.24 was chosen as a 2D Gaussian centered at the origin with variance 1. The iso-contours of equal probability are thus concentric

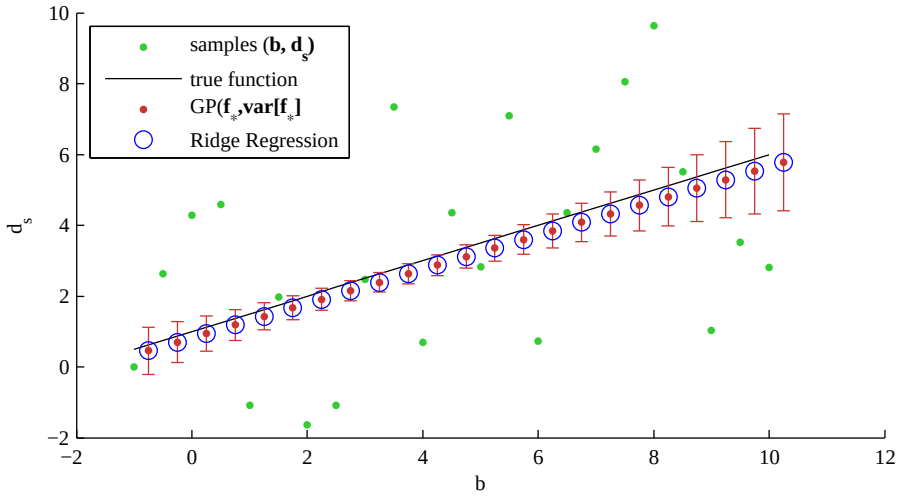


Figure 2.3: GP example with a linear function. The plot shows (1) the true linear function as a black solid line, (2) measured data samples in green, (3) the ridge regression MAP solution in blue circles, and (4) the GP predictive distribution with mean and standard deviation in red.

circles around the origin and there is no covariation between both weights.

$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_0 \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{2.35}$$

This corresponds to the linear kernel as a covariance function $k(b, b') = b \cdot b'$. Given the data, eq. 2.27 can be used to compute the posterior distribution on the weights. The 2D posterior distribution and the marginal distributions for each single weight are shown in fig. 2.4. It can be seen, that the prior with circular contours turned into ellipses with covariation between the weights. The MAP estimates for each weight, i.e. mean and standard deviation, can be read from the marginal distributions in the figure. The most likely estimates given the data were $w_1 = 0.483 \pm 0.14$ and $w_0 = 0.821 \pm 0.69$. The mean estimates are also the weights obtained as the optimal solution by ridge regression. For the GPs, the computation of the weight posterior is not necessary and also not recommended for more complicated covariance functions.
Instead, the predictive distribution from eq. 2.30 is computed directly. The predictive distribution has been obtained at unknown positions $b_i + 0.25$. Figure 2.3 illustrates the predictive distribution with mean and standard deviation in red error bars. It also shows the solution for ridge regression in blue circles. Both solutions have been computed assuming that the noise variance $\sigma_n^2 = 3$ was known. This is one reason why the solution

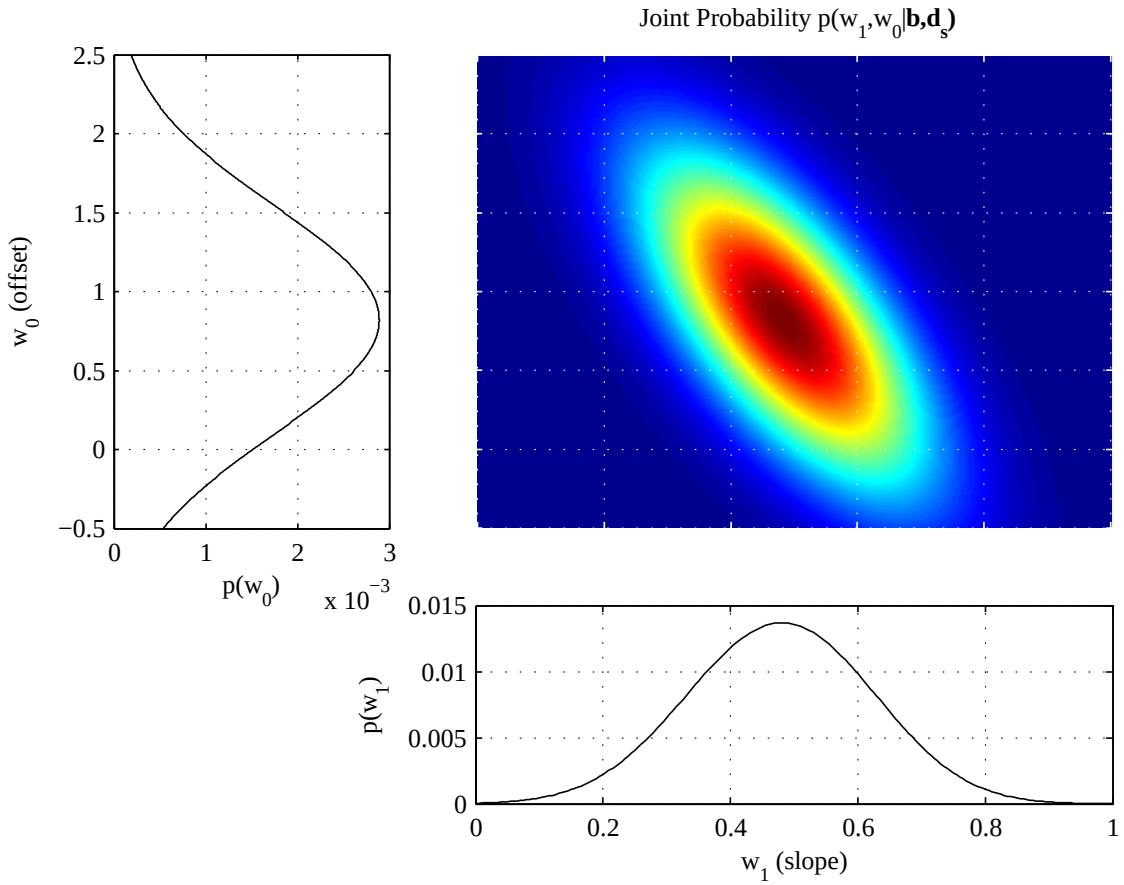


Figure 2.4: GP example - posterior distribution on the weights. The posterior of the weights is a 2D Gaussian. After integration out the other weight, 1D marginal distributions are shown for each weight separately.

for ridge regression and the GP solution coincide. In a real scenario this quantity needs to be estimated (for the GPs as discussed later) or guessed (for ridge regression). Further on, ridge regression for more complex functions would require a precise model φ and a good guess for its parametrization.
In addition, the predictive distribution – particularly the standard deviation bars – show that the predictions in the central part are more reliable than the predictions at the margins of the data set. This is reasonable, since these cases are lacking informative covariates (training data) to their left or to their right. Overall, the solutions only slightly deviate from the true underlying function despite the noise. The deviations originate from the fact that only a finite and very limited number of data samples was available to carry out the inference. With more or differently sampled training data, the outcome may vary.

2.2.2.4 Kernel Functions

Kernel functions are analytic expressions for inner products in potentially high dimensional spaces. A valid kernel has to fulfill Mercer's theorem [196]. The simplest case is given by the polynomial kernel given in eq. 2.36,

$$k_P(\boldsymbol{b}, \boldsymbol{b}') = s_c \cdot \left(\beta_0 + \boldsymbol{b}^T \boldsymbol{\Sigma}_w \boldsymbol{b}'\right)^F \tag{2.36}$$

where F is the polynomial degree, s_c a general scaling factor, $\boldsymbol{\Sigma}_w$ the covariance matrix parameter for scaling the variation in certain input dimensions, and β_0 an offset parameter. The case $\beta_0 = 0$ is called homogeneous polynomial kernel. The linear example earlier made use of this kernel with:

$$\beta_0 = 0,\ F = 1,\ s_c = 1,\ \text{and } \boldsymbol{\Sigma}_w = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

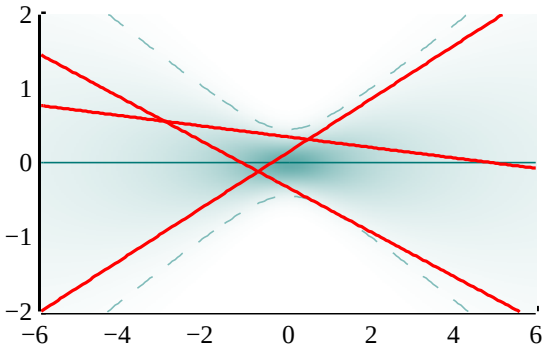
This kernel is not translationally invariant, since its output changes for increasing arguments. Stationary kernels, as they will be used in this work, only depend on the difference r between input vectors $r = \|\mathbf{b} - \mathbf{b}'\|$. Functions having this property are also called RBFs. However, in a less strict sense the term RBF is also used synonymously for the squared exponential (SE) kernel, especially in the SVR community. Equation 2.37 lists the SE kernel function.

$$k_{SE}(\boldsymbol{b}, \boldsymbol{b}') = s_c \cdot \exp\left(-\frac{r^2}{2\gamma}\right) \tag{2.37}$$

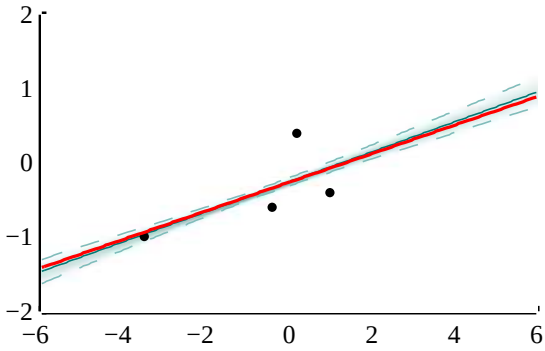
The SE kernel is a infinitely differentiable function and is hence very smooth. It is defined by a scaling parameter s_c and a length scale parameter γ. The latter indicates the average distance between samples in the input space beyond which the correlation between these samples drops dramatically. The SE or RBF kernel is the most widely used kernel in the field of machine learning.
However, Stein and colleagues [276] argue that the strong smoothness assumption is not realistic for real world scenarios. He therefore introduced the so-called Matérn kernel class which has a higher local flexibility. This kernel takes a simpler form when parametrized with half integers $\nu = p + 0.5$ [237]:

$$k_{Mat_\nu}(\boldsymbol{b}, \boldsymbol{b}') = s_c \cdot \exp\left(-\frac{\sqrt{2\nu}r}{\gamma}\right) \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{i=0}^{p} \frac{(p+i)!}{i!(p-i)!} \left(\frac{\sqrt{8\nu}r}{\gamma}\right)^{p-i} \tag{2.38}$$

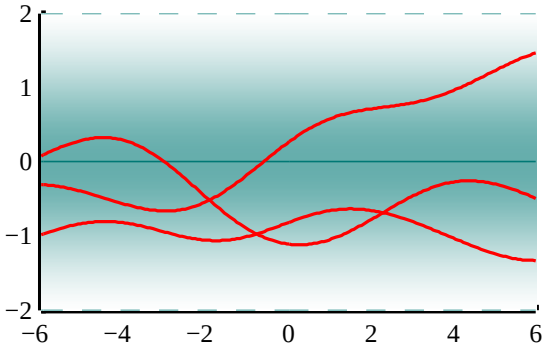
Here Γ denotes the Gamma function and p an arbitrary integer. The Matérn kernel is at most ν times differentiable and thus rougher than the SE kernel. The roughness can



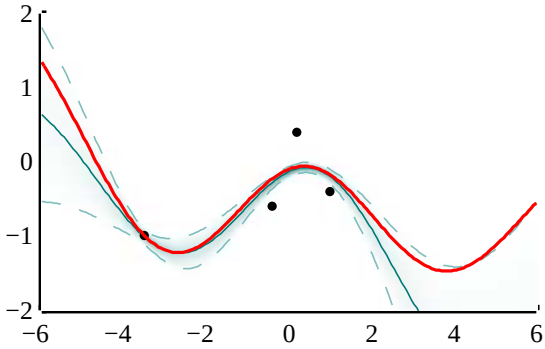
A: Functions for the linear kernel sampled from a prior ($w_i \sim \mathcal{N}(0, 0.05)$).



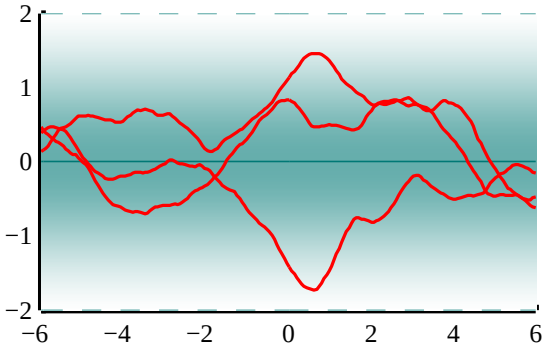
B: Functions for the linear kernel sampled from the posterior given the data.



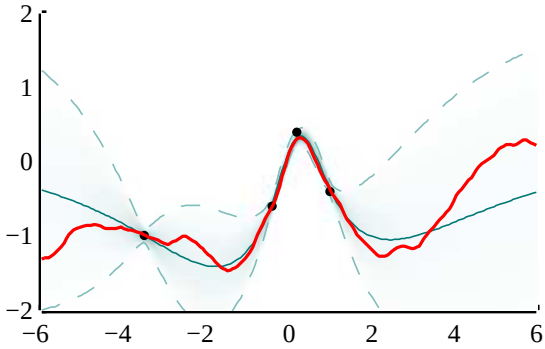
C: Functions for the SE kernel sampled from a prior ($s_c = 1, \gamma = 3$).



D: Functions for the SE kernel sampled from the posterior given the data.



E: Functions for the Matérn kernel sampled from a prior ($s_c = 1, \gamma = 3, \nu = 1.5$).



F: Functions for the Matérn kernel sampled from the posterior given the data.

Figure 2.5: Prior and posterior samples from a Gaussian process with different covariance functions. The solid blue line corresponds to the mean function, the dashed to the upper and lower standard deviation bounds, the red lines to exemplary samples from the GP and the shading denotes the probability for a data residing in that area. Observed experimental data is indicated by black dots.

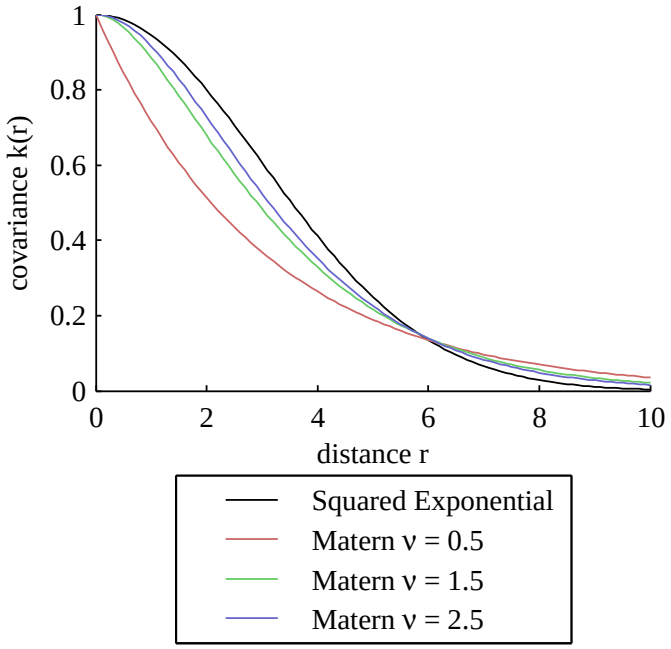


Figure 2.6: Kernel output for different kernel functions and parametrization. The similarity output of the SE and three Matérn kernels is shown in dependency of r.

be controlled by ν. While $\nu \to \infty$ gives the SE kernel in the limit, this work here will make use of the most widely-used parametrization $\nu = 1.5$. The kernel then simplifies as follows.

$$k_{Mat}(\boldsymbol{b}, \boldsymbol{b}') = s_c \cdot \left(1 + \frac{\sqrt{3}r}{\gamma}\right) \exp\left(-\frac{\sqrt{3}r}{\gamma}\right) \tag{2.39}$$

The left column of fig. 2.5 illustrates typical functions that were sampled from GPs characterized by different covariance functions. While these functions represent examples from the prior belief, the right column shows samples from the posterior distribution once data has been observed. It can be seen that the prior belief covers many different functions up to a certain complexity. A bound for the latter is defined by the covariance function. Once data is observed, the set of likely functions given the data is narrowed down to a subset. The algorithm is more certain about the true underlying function.
Both the SE as well as the Matérn kernel express similarities between samples $\boldsymbol{b}$ and $\boldsymbol{b}'$. Figure 2.6 illustrates the differences between different Matérn kernels and the SE kernel. The higher ν, the higher is the similarity output for samples being close together. On the other hand, samples far from each other are evaluated to be more similar for smaller ν. Overall, this defines the smoothness properties discussed earlier.

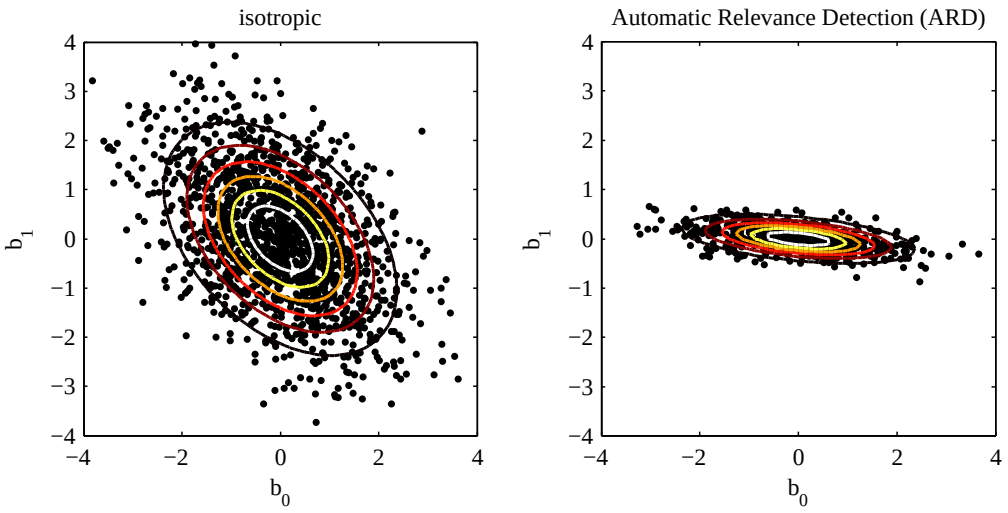


Figure 2.7: Isotropic and automatic relevance detection (ARD) weighting of feature dimensions. The data $\boldsymbol{b}_i$ (black) is sampled from a Gaussian distribution (cf. iso-probability-lines). In the right plot of the same data, the vertical axis is always scaled by $\gamma_2 = 0.25$ and the horizontal axis by $\gamma_1 = 1$. The behavior of the ARD kernel for arbitrary data would change after applying some rotation matrix $\boldsymbol{R}$ to $\boldsymbol{b}_i$ or reordering the feature dimensions. The data would look different from the kernel's perspective due to the axis-dependent scaling.

2.2.2.5 Automatic Relevance Detection (ARD)

For many applications, different features can be differently important for making predictions. One may wish to ensure that more relevant features have more impact on the prediction – particularly if some features contain very little information, but a lot of noise. The kernel functions presented in eq. 2.37 and eq. 2.39 are isotropic. This means the kernel behaves in an identical way irrespective of the feature dimension. Rotations in the input space have no impact. In order to provide so-called ARD, both kernels can be generalized in the way shown below.

$$k_{SE}(\boldsymbol{b}, \boldsymbol{b}') = s_c \cdot \exp\left(-\frac{|\boldsymbol{G}^{-1}\left(\boldsymbol{b} - \boldsymbol{b}'\right)|^2}{2}\right) \tag{2.40}$$

$$k_{Mat}(\boldsymbol{b}, \boldsymbol{b}') = s_c \cdot \left(1 + \sqrt{3}|\boldsymbol{G}^{-1}\left(\boldsymbol{b} - \boldsymbol{b}'\right)|\right) \exp\left(-\sqrt{3}|\boldsymbol{G}^{-1}\left(\boldsymbol{b} - \boldsymbol{b}'\right)|\right) \tag{2.41}$$

where

$$\boldsymbol{G} = \begin{bmatrix} \gamma_1 & 0 & \dots & 0 \\ 0 & \gamma_2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \gamma_D \end{bmatrix}.$$

The scaling for each dimension can be changed by introducing individual length scale parameters γ_i. This can be used to weight the relevance of specific dimensions by assigning a higher or lower weight to them. Figure 2.7 compares the isotropic case described earlier on with the ARD case. In the latter, one dimension is assigned a small weight. This reduces the impact of the feature's variance on the total sum in the $|\cdot|$ operator and hence also its relevance. The isotropic case can be constructed by setting $\gamma_1 = \gamma_2 = \ldots = \gamma_D = \gamma$ yielding eq. 2.37 and eq. 2.39.

All kernel parameters such as the length scale parameters are considered as hyperparameters and can be learned from the data or just be assigned fixed values.

2.2.2.6 Optimization and Hyperparameters

The noise variance σ_n^2 and the parameters of the covariance function $k(\cdot,\cdot)$ such as scaling value and length scales are unknown. They can be set to fixed values or be optimized within the Bayesian framework.
A convenient way is given by minimizing the negative log marginal likelihood (NLML) $p(\boldsymbol{d}_s|\boldsymbol{B})$. The expression for this probability already occurred in the upper left element of the matrix shown in eq. 2.33. As a function of the hyperparameters, this probability indicates the likelihood that the measured target labels $\boldsymbol{d}_s$ were generated from the input features $\boldsymbol{B}$ under the current model. The model is defined by the hyperparameters. The NLML can be derived as:

$$\log p(\boldsymbol{d}_s|\boldsymbol{B}) = -\frac{1}{2}\boldsymbol{d}_s^T\left(\boldsymbol{K} + \sigma_n^2\boldsymbol{I}\right)^{-1}\boldsymbol{d}_s - \frac{1}{2}\log\left|\boldsymbol{K} + \sigma_n^2\boldsymbol{I}\right| - \frac{N}{2}\log 2\pi \tag{2.42}$$

For minimization, conjugate gradient descent was used [208]. The gradient descent has been repeated 20 times with randomly sampled starting points for the set of hyperparameters. The optimal set of hyperparameters was selected as the one having the smallest NLML of all 20 gradient descent solutions. Note that this optimization is unbiased and only based on the training data. Unlike for SVR, the hyperparameter optimization is non-convex and can suffer from local minima. This issue is eased by the repetitive procedure. However, *all* parameters can be optimized in this manner, whereas SVR employs

convex optimization for the Lagrange multipliers, but requires grid search for others like the kernel parameters.

2.2.2.7 Sparse Approximations

The evaluation of eq. 2.30, i.e. the predictive distribution, requires to compute the inverse of the Gram matrix $\boldsymbol{K} \in \mathbb{R}^{N \times N}$. The computational cost for this scales with $\mathcal{O}(N^3)$. Therefore, the computational time grows cubically with the number of data samples. For a few thousand samples this becomes prohibitive. Two approaches evolved to tackle this shortcoming [209]:

1. The application of sparse approximation methods approximating the full predictive distribution and involving matrices of lower rank.

2. Matrix-vector multiplication methods for optimizing the computation of the inversion problem itself.

The following considerations will focus on the first approach only. Further on, the approximations will only use inductive methods [258], which exploit information from the training data only.

Subset of Data (SoD) The simplest way to reduce computational effort is achieved by discarding some of the training data. Selecting only a subset of size M from all N samples decreases the size of the Gram matrix to $M \times M$ with $M < N$. This very crude way is the baseline against which more sophisticated approximation methods are compared. Although the assumption that some of the training data contains redundant information may be true for many cases, the actual selection of the subset is still of importance.

The Concept of Inducing variables More sophisticated sparse approximation methods generally rely on the following assumption: Unknown data of the underlying function f can be fully described by a finite subset of so-called *inducing variables* $\boldsymbol{u} \in \mathbb{R}^{1 \times M}$ at input feature vectors $\boldsymbol{B_u} \in \mathbb{R}^{D \times M}$. These variables can be part of the training set of size $N > M$ or can be completely unknown data. In the following it will be assumed that the training data has sufficiently captured these variables.
The general idea is that the number of inducing variables M is smaller than the size of the training set N. Computations such as in eq. 2.30 can be expressed in terms of the inducing variables, which saves computational time and data storage demands.
Since the inducing variables fully describe the unknown data from f, it can be assumed

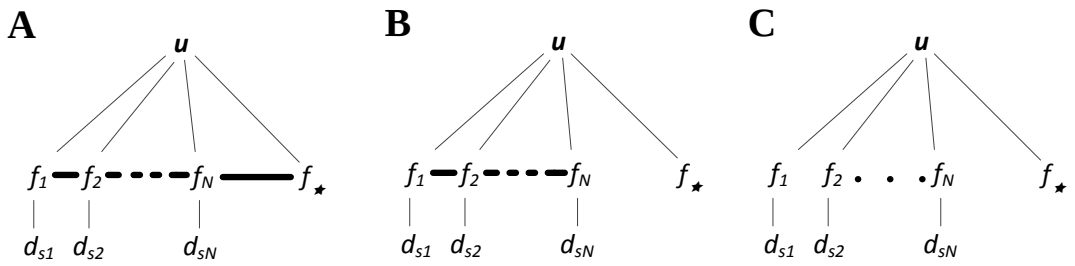


Figure 2.8: The concept of inducing variables. **A:** Fully connected model without approximation. Functional values $f_i = f(\boldsymbol{b}_i)$ of the training data set statistically relate to each other and also to the desired test target label $f_\star$. The observations are obtained from the corresponding functional values after adding iid noise. **B:** The general sparse approximation scenario. Training and test labels can only communicate via the inducing variables. Given $\boldsymbol{u}$ they are independent [209]. **C:** This case extends case **B**. It shows the dependencies for fully independent training conditional (FITC) approximation.

that the known training data with $\boldsymbol{B}$ and $\boldsymbol{d}_s$ is statistically independent from the test data $f_\star$ given $\boldsymbol{u}$. Figure 2.8 illustrates this schematically. Edges in the graph denote statistical dependencies among the data. Case **A** on the left shows the general scenario and case **B** the approximation using the independence assumption. Mathematically, the transition can be expressed by introducing latent variables $\boldsymbol{u}$ in the joint distribution of eq. 2.33:

$$p(\boldsymbol{d}_s, f_\star) = \int p(\boldsymbol{d}_s, f_\star|\boldsymbol{u})d\boldsymbol{u} \approx \int p(f_\star|\boldsymbol{u})p(\boldsymbol{d}_s|\boldsymbol{u})p(\boldsymbol{u})d\boldsymbol{u} \stackrel{!}{=} q(\boldsymbol{d}_s, f_\star) \qquad (2.43)$$
$$\text{with } p(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}(\boldsymbol{B_u}, \boldsymbol{B_u}))$$

where $q(\boldsymbol{d}_s, f_\star)$ is the approximated, i.e more efficient, joint distribution, from which the new predictive distribution is derived. The distributions $p(f_\star|\boldsymbol{u})$ and $p(\boldsymbol{d}_s|\boldsymbol{u})$ are the test and training conditional, respectively:

$$\begin{aligned} p(\boldsymbol{d}_s|\boldsymbol{u}) &= \mathcal{N}(\boldsymbol{K}(\boldsymbol{B}, \boldsymbol{B_u})\boldsymbol{K}(\boldsymbol{B_u}, \boldsymbol{B_u})^{-1}\boldsymbol{u}, \boldsymbol{K}(\boldsymbol{B}, \boldsymbol{B}) - \boldsymbol{Q}_{SoR}(\boldsymbol{B}, \boldsymbol{B}) + \sigma_n^2\boldsymbol{I}) \\ p(f_\star|\boldsymbol{u}) &= \mathcal{N}(\boldsymbol{k}(\boldsymbol{b}_\star, \boldsymbol{B_u})\boldsymbol{K}(\boldsymbol{B_u}, \boldsymbol{B_u})^{-1}\boldsymbol{u}, k(\boldsymbol{b}_\star, \boldsymbol{b}_\star) - k_{SoR}(\boldsymbol{b}_\star, \boldsymbol{b}_\star)) \end{aligned} \qquad (2.44)$$

where $\boldsymbol{Q}_{SoR}(\boldsymbol{b}_i, \boldsymbol{b}_j) = \boldsymbol{k}(\boldsymbol{b}_i, \boldsymbol{B_u})\boldsymbol{K}(\boldsymbol{B_u}, \boldsymbol{B_u})^{-1}\boldsymbol{k}(\boldsymbol{B_u}, \boldsymbol{b}_j)$ can be shown to be the approximation of the full Gram matrix $\boldsymbol{K}$ resulting from the subset of regressors (SoR) approximation method (see below). An intuitive way of understanding eq. 2.44 is as follows. As illustrated by case **B** in fig. 2.8 there is no mutual dependency of $\boldsymbol{d}_s$ and $f_\star$ in eq. 2.44. The general structure of mean and variance is similar to eq. 2.30. The variance is the full Gram

matrix $\boldsymbol{K}$ minus the variance proportion which is explained by $\boldsymbol{u}$, since the distribution is conditioned on $\boldsymbol{u}$. Assuming that $\boldsymbol{u}$ would comprise the entire training set $\boldsymbol{B_u} = \boldsymbol{B}$, matrices in the expression would cancel out. In the training conditional the mean would turn into the functional values $\boldsymbol{f}$ and the variance into the noise term $\sigma_n^2 \boldsymbol{I}$ only. For $\boldsymbol{u}$ being the empty set the prior for the target labels in eq. 2.33 is obtained.
Sparse techniques now aim at finding approximations $q(\boldsymbol{d}_s|\boldsymbol{u})$ and $q(f_\star|\boldsymbol{u})$ for $p(\boldsymbol{d}_s|\boldsymbol{u})$ and $p(f_\star|\boldsymbol{u})$, respectively.

Subset of Regressors (SOR) The SoR assumes that the predictive sample $f_\star$ is deterministically described by a weighted sum of similarities (the regressors) between the test input $\boldsymbol{b}_\star$ for the desired $f_\star$ and the subset $\boldsymbol{B_u}$. Unlike for other approaches, there is no uncertainty involved. Similarities are defined by the kernel function. The approximations for $q_{SoR}(\boldsymbol{d}_s|\boldsymbol{u})$ and $q_{SoR}(f_\star|\boldsymbol{u})$ can hence be derived as follows.

$$\begin{aligned} q_{SoR}(\boldsymbol{d}_s|\boldsymbol{u}) &= \mathcal{N}\left(\boldsymbol{K}(\boldsymbol{B},\boldsymbol{B_u})\boldsymbol{K}(\boldsymbol{B_u},\boldsymbol{B_u})^{-1}\boldsymbol{u}, \sigma_n^2\boldsymbol{I}\right) \\ q_{SoR}(f_\star|\boldsymbol{u}) &= \mathcal{N}\left(k(\boldsymbol{b}_\star,\boldsymbol{B_u})\boldsymbol{K}(\boldsymbol{B_u},\boldsymbol{B_u})^{-1})\boldsymbol{u}, 0\right) \end{aligned} \tag{2.45}$$

With eq. 2.43, this yields a joint distribution which contains the SoR Gram matrix from above.

$$q_{SoR}(\boldsymbol{d}_s, f_\star) = \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{Q}_{SoR}(\boldsymbol{B},\boldsymbol{B}) + \sigma_n^2\boldsymbol{I} & \boldsymbol{q}_{SoR}(\boldsymbol{B},\boldsymbol{b}_\star) \\ \boldsymbol{q}_{SoR}(\boldsymbol{b}_\star,\boldsymbol{B}) & q_{SoR}(\boldsymbol{b}_\star,\boldsymbol{b}_\star) \end{bmatrix}\right) \tag{2.46}$$

Detailed descriptions are given in [51, 209, 232].

Fully Independent Training Conditional (FITC) Among other techniques such as SoR, FITC was recommended as the best choice by Quiñonero-Candela et al. [209]. It was hence adopted in this work.
As the name suggests, the method assumes a second approximation for the training conditional, while the test conditional is $q_{FITC}(f_\star|\boldsymbol{u}) = p(f_\star|\boldsymbol{u})$. The approximation was first proposed by Snelson [271] and is given in eq. 2.47 [209].

$$\begin{aligned} &q_{FITC}(\boldsymbol{d}_s|\boldsymbol{u}) = \\ &\mathcal{N}(\boldsymbol{K}(\boldsymbol{B},\boldsymbol{B_u})\boldsymbol{K}(\boldsymbol{B_u},\boldsymbol{B_u})^{-1}\boldsymbol{u}, \operatorname{diag}\left(\boldsymbol{K}(\boldsymbol{B},\boldsymbol{B}) - \boldsymbol{Q}_{SoR}(\boldsymbol{B},\boldsymbol{B})\right) + \sigma_n^2\boldsymbol{I}) \end{aligned} \tag{2.47}$$

Given $\boldsymbol{u}$, the target labels of the training set $\boldsymbol{d}_s$ are fully independent, i.e. the covariance matrix of the training conditional is a diagonal matrix with zeros off the diagonal.

Table 2.3: Storage demands, computational time for training the GP model as well as computing the predictive mean and variance are compared for the full model, SoD and FITC (cf. [51]).

Method	Storage	Training	Predictive Mean	Predictive Variance
Full Model	$\mathcal{O}(N^2)$	$\mathcal{O}(N^3)$	$\mathcal{O}(N)$	$\mathcal{O}(N^2)$
SoD	$\mathcal{O}(M^2)$	$\mathcal{O}(M^3)$	$\mathcal{O}(M)$	$\mathcal{O}(M^2)$
FITC	$\mathcal{O}(MN)$	$\mathcal{O}(M^2N)$	$\mathcal{O}(M)$	$\mathcal{O}(M^2)$

Schematically this is illustrated in case **C** in fig. 2.8. This gives rise to the approximate joint distribution, which is the FITC replacement for eq. 2.33:

$$q_{FITC}(\boldsymbol{d}_s, f_\star) = \tag{2.48}$$
$$\mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{Q}_{SoR}(\boldsymbol{B},\boldsymbol{B}) - \operatorname{diag}\left(\boldsymbol{K}(\boldsymbol{B},\boldsymbol{B}) - \boldsymbol{Q}_{SoR}(\boldsymbol{B},\boldsymbol{B})\right) + \sigma_n^2 \boldsymbol{I} & \boldsymbol{q}_{SoR}(\boldsymbol{B}, \boldsymbol{b}_\star) \\ \boldsymbol{q}_{SoR}(\boldsymbol{b}_\star, \boldsymbol{B}) & k(\boldsymbol{b}_\star, \boldsymbol{b}_\star) \end{bmatrix}\right)$$

Equation 2.48 shows that the SoR approximation is taken for the training conditional, where the diagonal is replaced with the true diagonal of $\boldsymbol{K}(\boldsymbol{B},\boldsymbol{B})$. The matrix part related to the desired test label in the lower right is not subject to approximation. Thus, strictly speaking, due to different treatment of training and test data, FITC does not count as a proper GP anymore.

After generalizing the SoR Gram matrix $\boldsymbol{Q}_{SoR}$ to a kernel function $k_{SoR}(\boldsymbol{b}_i, \boldsymbol{b}_j) := \boldsymbol{k}(\boldsymbol{b_i}, \boldsymbol{B_u})\boldsymbol{K}(\boldsymbol{B_u}, \boldsymbol{B_u})\boldsymbol{k}(\boldsymbol{B_u}, \boldsymbol{b}_j)$, FITC can be implemented by just replacing the original kernel function $k(\cdot,\cdot)$ in eq. 2.31 by the adapted kernel function $k_{FITC}(\cdot,\cdot)$.

$$k_{FITC}(\boldsymbol{b}_i, \boldsymbol{b}_j) = k_{SoR}(\boldsymbol{b}_i, \boldsymbol{b}_j) + \delta_{ij}\left[k(\boldsymbol{b}_i, \boldsymbol{b}_j) - k_{SoR}(\boldsymbol{b}_i, \boldsymbol{b}_j)\right] \tag{2.49}$$

where $\delta(k,j)$ is the Kronecker-Delta ($\delta(k,j) = 1$ for $k = j$, $\delta(k,j) = 0$ for $k \neq j$). FITC reduces the computational complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(M^2N)$, Table 2.3 finally lists the savings in storage as well as training and prediction time for the full model, subset of data (SoD) and FITC. For the complexity of the predictive quantities, it is assumed that the inverse of the Gram matrix has been computed during training and is stored for prediction.

Determining the Inducing Variables The quality of the approximation is related to the choice of the inducing variables. The problem of selecting the best subset from the training data can be approached in different ways. The easiest way is to define

Algorithm 1 k-Means

Initialize
    Set mean vectors $\boldsymbol{m}_1^{(1)}, \dots \boldsymbol{m}_M^{(1)}$ randomly to M training vectors $\boldsymbol{b}_j,\ j \in [1, N]$
    Set $i = 1$
for $i = 1 : N_{Iter}$
    $S_k^{(i)} = \left\{ \boldsymbol{b}_j : \left\| \boldsymbol{b}_j - \boldsymbol{m}_k^{(i)} \right\| \leq \left\| \boldsymbol{b}_j - \boldsymbol{m}_{k^*}^{(i)} \right\| \forall k^* = 1, \dots, M; k^* \neq k \right\}$
    $\boldsymbol{m}_k^{(i+1)} = \frac{1}{S_k^{(i)}} \sum_{\boldsymbol{b}_j \in S_k^{(i)}} \boldsymbol{b}_j$
endfor
$S_{\boldsymbol{u}} = \{\boldsymbol{b}_j : \|\boldsymbol{b}_j - \boldsymbol{m}_k\| \leq \|\boldsymbol{b}_l - \boldsymbol{m}_k\| \, \forall l = 1, 2, \dots N; \forall k = 1, 2, \dots M\}$

M and then to randomly sample data from the training data. This approach does not incorporate knowledge about how information is distributed across training samples.
A second way is given by unsupervised learning e.g. via k-means. This method clusters the training data and selects a set $S_{\boldsymbol{u}}$ consisting of the M training samples closest to the M cluster centers. From this set of vectors the matrix $\boldsymbol{B}_{\boldsymbol{u}}$ can be composed. Overall, this aims at an evenly sampled feature space maintaining an optimum of information.
Here, SoD and FITC are tested with both approaches. Methods with random initialization will be denoted SoD$_R$ and FITC$_R$. K-means initialization is labeled SoD$_K$ and FITC$_K$.

The k-means procedure is summarized in algorithm 1. Essentially, k-means iterates between two steps. First, it assigns all N samples to M sets S_k. Each sample $\boldsymbol{b}_j$ is assigned to the set corresponding to the closest centroid $\boldsymbol{m}_k$. The initial centroids are randomly chosen samples from the data. Second, new centroids are computed with this new set assignment.
The variables $\boldsymbol{u}$ can also be treated as hyperparameters and estimated during training. This, however, significantly increases the search space dimension for the conjugate gradient descent. The optimization results would be less stable, unless the number of repeated optimization runs with varying starting points is increased. This will again affect the computation time required for training the GP model. More sophisticated approaches have been proposed by Smola et al.[269], who approximated the posterior over the weights $\boldsymbol{\alpha}$ from eq. 2.31 and then selected the $\boldsymbol{u}$ by a greedy algorithm maximizing the NLML.

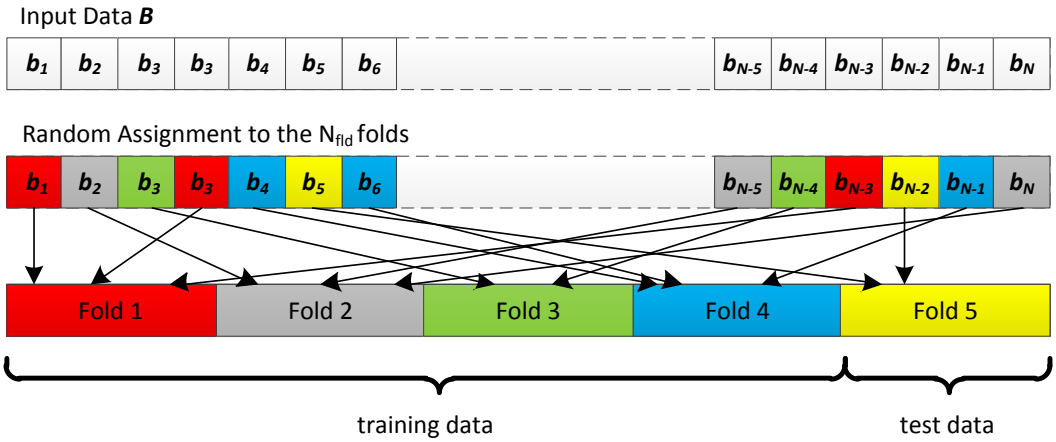


Figure 2.9: Visualization of the cross-validation (CV) testing scheme. This scheme will be used for testing based on a single frame. The data samples are randomly assigned to one of n_{fld} folds (illustrated above for $n_{fld} = 5$). One fold is used for testing and the remaining for training a regression model. The scheme then iterates through all folds for assigning the test set, i.e. each sample will exactly once be treated as an unknown test sample.

2.2.3 Considerations on Testing

Testing Schemes The supervised learning techniques described earlier handle labeled data. Each data set consists of N feature vectors $\boldsymbol{B}$ (e.g. optical features, input) and target labels $\boldsymbol{d}_s$ (tissue thickness, output). Experimentally speaking, these data may comprise different cases:

1. One single frame from one measurement with up to 1024 scanned laser spots
2. Up to three different measurements with 1-3 frames each à 1024 laser spots

The term frame will relate to one forehead scan with a 32×32 laser spot grid. A measurement refers to a collection of one or more frames for the same head pose without voluntary head motion. Different measurements relate to different head poses with voluntary head motion in between. The testing scheme will depend on this case distinction.

To get an estimate for the generalization error, the data will be separated into training data and test data. The training set is used by the statistical learning algorithm to learn a model. In order to evaluate how well the model generalizes on unseen data, the test set

will be used to compute the generalization error. Overfitting and biased error measures are avoided since the test set contributes no information during model training.

Cross-Validation (CV) In the first case there is only one frame available. Repeated cross-validation (CV) will be used to estimate the generalization error. As illustrated in fig. 2.9, the N data samples in $\boldsymbol{B}$ are randomly assigned to one of n_{fld} so-called folds. Each fold is a data subset comprising $\frac{N}{n_{fld}}$ data samples. The fraction will be rounded to the next smaller integer N_{fld} and remaining samples will be distributed across the existing folds. Therefore, the number of samples per fold is either N_{fld} or $N_{fld}+1$.
For testing, one fold is chosen as the test set and all other $n_{fld}-1$ folds are used for training the model. The CV scheme will then iterate the test set through all folds to cover the entire data set. Thus, each sample turns into an unknown test sample exactly once. This results in a fair and reproducible estimate for the generalization error.
The error estimate will depend on the random assignment of the data samples to a specific fold. Therefore, the CV procedure described before will be repeated n_{rep} times with different random assignments. By taking the mean over all repetitions, the random variance will be averaged out. The final scheme is called n_{rep}-times-n_{fld}-fold-CV. For all tests discussed hereafter a 5-times-10-fold-CV scheme was employed.

Evaluation across Measurements (AM) If the data set comprises more than one measurement, no CV scheme will be applied. Instead, the training can contain several frames from more than one measurement. One measurement not included in this set will be used for testing. Two different scenarios will be distinguished:

1. AM1 set - test set and training set contain data from at most one measurement each
2. AM2 set - the test set is a single measurement, the training set contains all remaining measurements (typically two)

Grid Search and Hyperparameters Section 2.2.2.6 described how hyperparameters of a GP are estimated by minimizing the NLML. This results in unbiased estimates for the parameters, since the NLML only depends on the training data. Initial parameter values for the gradient descent were set at random.
In SVR the testing schemes above were used to optimize the SVR parameters C, ε and γ. This is done by computing the generalization error for a "grid" of parameters and then selecting the parameters corresponding to the lowest generalization error. This is called *grid search*.

Strictly speaking, this will result in too optimistic estimates of the generalization error, since the parameters are optimized with the test set. This data is not available in a real scenario. An unbiased optimization would make use of additional CV runs on the training data. This is called nested CV and introduces an additional test set within the training data set – called validation set [170]. This was however found to be computationally prohibitive. Instead, all SVR evaluations will optimize the generalization error in two steps. First, a 2D grid for γ and ε is tested. Meanwhile, C is set to 200 being an initial guess in the medium range. The optimal parameter set for the two parameters is then used to optimize C in a 1D grid.
To account for the bias, the difference between nested CV and the CV scheme just described will be shown and discussed for a single data set. This provides insight into the extent of the bias and allows to judge about later results accordingly [326].

Error Measures The error will be computed as the difference between a predicted target label (the tissue thickness) $d_{s_\star}$ or, more precisely, $\bar{d}_{s_\star}$ and the corresponding ground truth for this data sample d_s, i.e. $d_s - d_{s_\star}$. Note that SVR and GPs estimate the underlying functional value $\bar{f}_\star$ and not the noise corrupted $\bar{d}_{s_\star}$. The error, however, computes the difference with respect to the noise corrupted ground truth measurement d_s. This is still valid, since the noise distributions assumed in both regression models are of zero mean. On average, there is $\bar{f}_\star = \bar{d}_{s_\star}$ as explained in eq. 2.31. By further averaging across multiple data samples, the noise superposing the ground truth is averaged out.
All subsequent error measures will make use of the absolute error (AE), because there is no special penalty on exceeding or falling below the ground truth.

$$AE = |d_s - d_{s_\star}| \tag{2.50}$$

The 90 % error bound (I90) then indicates the value which is only exceeded by 10 % of all absolute errors in the data set. 90 % of the data have an absolute error below the I90 bound.

$$p(AE \leq I90) = 0.9 \tag{2.51}$$

For a data set containing N data samples the absolute errors $\{AE_i\}_{i=1\ldots N}$ can be computed. The mean absolute error (MAE) is then obtained as follows.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} AE_i = \frac{1}{N}\sum_{i=1}^{N} |d_{si} - d_{si_\star}| \tag{2.52}$$

Equivalently the root mean square error (RMSE) can be defined.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_{si} - d_{si_\star})^2} \tag{2.53}$$

Both the MAE and the RMSE will be computed for error evaluation. They may range from zero to infinity and are measured in $[mm]$. As a linear score, the MAE indicates where most of the error mass is located. The RMSE constitutes a more conservative measure. Due to its quadratic dependency on individual errors, it penalizes outliers with a higher weight. Generally, MAE $\leq$ RMSE is true. Both are only equal if all error values have the same magnitude.

2.3 Point Cloud Registration

2.3.1 Point Clouds and Homogenous Transformations

A set of vectors in 3D space $\mathbb{R}^3$ can be interpreted as a set of points in this space. This set of points will be named point cloud or surface. Without an explicit calibration of any kind, the relationship between two different point clouds $P_{cld} := \{\boldsymbol{p}_i \in \mathbb{R}^{3\times1} | \boldsymbol{p}_i = [p_{xi}, p_{yi}, p_{zi}]^T\}_{i=1\ldots N_p}$ and $Q_{cld} := \{\boldsymbol{q}_i \in \mathbb{R}^{3\times1} | \boldsymbol{q}_i = [q_{xi}, q_{yi}, q_{zi}]^T\}_{i=1\ldots N_q}$ is unknown. These point clouds may correspond to different views from the same object or can be acquired using different sensor modalities and coordinate spaces. For this work, we assume P_{cld} to be a point cloud triangulated from a laser scanning device, and Q_{cld} to be a high resolution surface from an MRI scan. We typically have $N_p << N_q$.
Finding a spatial relationship between the two clouds is called registration. Two point clouds are said to be registered, if they can be brought into coincidence by *one* rigid[5] transformation [59]. This registration is hardly ever perfect, since the two surfaces represent noisy measurements from different modalities and were recorded from deformable soft tissue. This makes registration a (typically non-convex) optimization problem.
The optimization aims at finding a transformation matrix that transforms P_{cld} into the coordinate space of Q_{cld}, where both coincide:

$${}^{Q}\boldsymbol{p} = {}^{P}\mathcal{T}_{Q} \cdot {}^{P}\boldsymbol{p} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix} \cdot \begin{bmatrix} {}^{P}\boldsymbol{p} \\ 1 \end{bmatrix} = \boldsymbol{R} \cdot {}^{P}\boldsymbol{p} + \boldsymbol{t} \tag{2.54}$$

where the left upper case letter denotes the coordinate space in which the 3D point $\boldsymbol{p}$ resides. Further on, ${}^{P}\mathcal{T}_{Q} \in \mathbb{R}^{4\times4}$ is the transformation matrix from space P to Q, $\boldsymbol{t} =$

[5] A rigid transformation includes only three translational and three rotational degrees of freedom. There is no shearing or zooming.

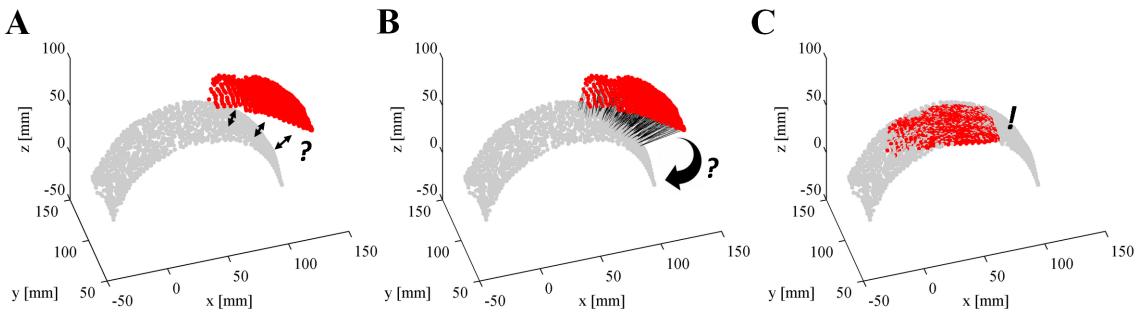


Figure 2.10: Iterative Closest Point Algorithm. The plots illustrate one ICP iteration: **A:** Two point clouds need to be registered. In a first step point-to-point correspondences are estimated. **B:** A transformation matrix can be computed using the correspondences from the last step, **C:** After several iterations, convergence, i.e. a registration of the point clouds, is achieved.

$[t_x, t_y, t_z]^T \in \mathbb{R}^3$ is the translational offset and $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ an orthogonal rotation matrix with determinant 1:

$$\boldsymbol{R} = \begin{bmatrix} \cos r_y \cos r_z & -\cos r_y \sin r_z & \sin r_y \\ \cos r_x \sin r_z + \sin r_x \sin r_y \cos r_z & \cos r_x \cos r_z + \sin r_x \sin r_y \sin r_z & -\sin r_x \cos r_y \\ \sin r_x \sin r_z - \cos r_x \sin r_y \cos r_z & \sin r_x \cos r_z + \cos r_x \sin r_y \sin r_z & \cos r_x \cos r_y \end{bmatrix}$$

With the triplet of angles $\boldsymbol{r} = [r_x, r_y, r_z]^T \in \mathbb{R}^3$, it describes the rotation that can be decomposed into three subsequent rotations around the coordinate axes of space Q (yaw-pitch-roll or ZY'X" convention according to the Tait Bryan group for order of rotations).

2.3.2 Iterative-Closest-Point (ICP) Algorithm

For finding the transformation matrix as described above, two optimization problems have to be solved. Figure 2.10 illustrates both of them. First, the two point clouds contain different numbers of points and are arbitrarily orientated with respect to each other. To compute a transformation matrix which registers both surfaces, point-to-point correspondences have to be known. That means, knowledge is required about which point $\boldsymbol{p}_i$ in P_{cld} corresponds to which point $\boldsymbol{q}_j$ in Q_{cld}. After the correspondences have been determined, a transformation matrix can be found, which minimizes some error measure between the point pairs.
Since it is hard to solve both problems at once, the iterative closest point (ICP) algorithm solves the overall problem iteratively in two steps. Algorithm 2 outlines the general idea of ICP in pseudo code, where $\boldsymbol{P} \in \mathbb{R}^{3\times N}$ is a matrix that contains all points $\boldsymbol{p}_i$ as column vectors and k is the iteration index.

Algorithm 2 Iterative Closest Point Algorithm

Initialize
    Find initial correspondences (e.g. at random)
    Set $\mathcal{T}^0$ to be the identity matrix
    Set $\boldsymbol{P}^0 = [\boldsymbol{p}_i]_{i=1...N}$
    Set $k = 1$
loop until: registration error $< \epsilon_{threshold}$
    (1) Matching step: compute $\mathcal{T}^k$ by minimizing the error between corresponding points
    Set $\boldsymbol{P}^k = \mathcal{T}^k \cdot \boldsymbol{P}^{k-1}$
    (2) Correspondence step: compute correspondences to find the closest neighbors after transformation
    Compute the remaining registration error ϵ between $\boldsymbol{P}^k$ and the corresponding points in Q_{cld}
    $k \leftarrow k + 1$
end loop
Compute the final transformation ${}^{P}\mathcal{T}_{Q} = \prod_{k} \mathcal{T}^k$

This optimization is typically non-convex, but guaranteed to converge into a local minimum [243]. The possibility of finding the global optimum depends on the characteristics of the surfaces, the number of points in each surface, and on the initial spatial relation between the two surfaces.
As stated in the purpose of this work, all considerations in the subsequent chapters aim at facilitating the search for the global optimum by providing additional information. Irrespective of this, it should be emphasized that algorithmic improvements may supplement this additional information to achieve optimal registration results.

2.3.2.1 Matching Step

Given a set of correspondences between each point in P_{cld} and one corresponding point in Q_{cld}, the matching step aims at minimizing an error functional between both surfaces. The output will be the optimal transformation matrix. Several error functionals have been proposed with the most popular ones being the so-called point-to-point algorithm [12, 26, 243], the point-to-plane algorithm [19, 22, 59], and point-to-projection algorithms [29, 216]. This work will focus on the first two approaches. The third relies on additional projections of points onto an image plane, and has been shown to yield similar matching

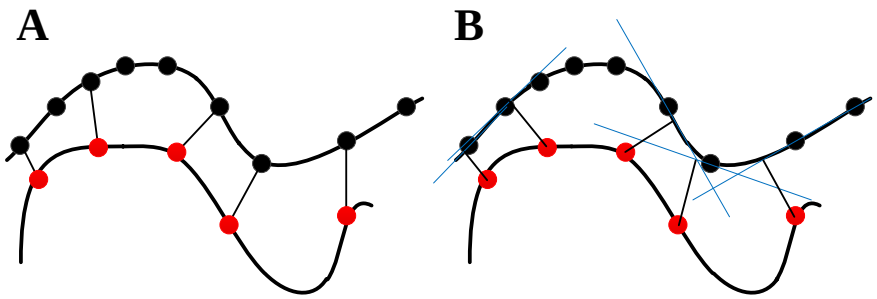


Figure 2.11: The ICP matching step. **A:** The point-to-point algorithm minimizes the Euclidean distances between corresponding point pairs. **B:** The point-to-plane algorithm minimizes the orthogonal distance between a point and the tangent plane of the corresponding point.

results.

Point-to-Point Matching Point-to-point matching takes all points ${}^{P}\boldsymbol{p}_i$ and the corresponding points ${}^{Q}\boldsymbol{q}_i$ and constructs an error functional $\mathfrak{E}$ which minimizes the sum of Euclidean distances $\|\cdot\|_2$ between the point pairs as illustrated in fig. 2.11A. Note that the optimization problem is solved with the N_p correspondence pairs only.

$$\mathfrak{E} = \sum_{i=1}^{N_p} \left\| \boldsymbol{R} \cdot {}^{P}\boldsymbol{p}_i + \boldsymbol{t} - {}^{Q}\boldsymbol{q}_i \right\|^2 \tag{2.55}$$

To get the optimal translational offset $\boldsymbol{t}$, the centroids of both point clouds ($\bar{\boldsymbol{p}}$ and $\bar{\boldsymbol{q}}$) and their shifted versions can be defined:

$$\boldsymbol{p}'_i = \boldsymbol{p}_i - \bar{\boldsymbol{p}} \quad \text{with} \quad \bar{\boldsymbol{p}} = \frac{1}{N_p} \sum_{i=1}^{N_p} \boldsymbol{p}_i \tag{2.56}$$

$$\boldsymbol{q}'_i = \boldsymbol{q}_i - \bar{\boldsymbol{q}} \quad \text{with} \quad \bar{\boldsymbol{q}} = \frac{1}{N_p} \sum_{i=1}^{N_p} \boldsymbol{q}_i \tag{2.57}$$

By inserting the above into eq. 2.55 the optimal translational offset results in:

$$\boxed{\boldsymbol{t} = \boldsymbol{R} \cdot \bar{\boldsymbol{p}} + \bar{\boldsymbol{q}}} \tag{2.58}$$

With the offsets eliminated and the optimal translational solution above, eq. 2.55 yields

$$\mathfrak{E} = \sum_{i=1}^{N} \left\| \boldsymbol{R} \cdot \boldsymbol{p}'_i - \boldsymbol{q}'_i \right\|^2 \tag{2.59}$$

$$= \sum_{i=1}^{N} \left\| \boldsymbol{p}'_i \right\|^2 - 2tr\left(\boldsymbol{R} \sum_{i=1}^{N} \boldsymbol{p}'_i \boldsymbol{q}'^T_i \right) + \sum_{i=1}^{N} \left\| \boldsymbol{q}'_i \right\|^2 \tag{2.60}$$

where $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}$ has been used and $tr(\cdot)$ is the trace of a matrix. The minimization of the term above reduces to maximizing

$$tr\left(\boldsymbol{R} \sum_{i=1}^{N} \boldsymbol{p}'_i \boldsymbol{q}'^T_i \right) = tr\left(\boldsymbol{R}\boldsymbol{H} \right) \tag{2.61}$$

where $\boldsymbol{H} \in \mathbb{R}^{3\times3}$ can be interpreted as the spatial covariance matrix of the points in the cloud. With the singular value decomposition (SVD) of $\boldsymbol{H}$

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T \tag{2.62}$$

and the general matrix lemma $tr(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T) \geq tr(\boldsymbol{R}\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T)$ Arun et al. [12] proved that

$$\boxed{\boldsymbol{R} = \boldsymbol{V}\boldsymbol{U}^T} \tag{2.63}$$

maximizes eq. 2.61 and yields the optimal rotation matrix.

Point-to-Plane Matching Given corresponding point pairs, the point-to-plane algorithm minimizes the sum of orthogonal distances between points ${}^P\boldsymbol{p}_i$ and the tangent planes at the corresponding points ${}^Q\boldsymbol{q}_i$. This is illustrated in fig. 2.11B. This optimization goal makes it possible that reference points by which the ${}^P\boldsymbol{p}_i$ are attracted may slide along the tangent plane and are not restricted to the grid sampling of Q_{cld}. Therefore, point-to-plane optimization reduces the influence of the discrete sampling grid in Q_{cld}. To a minor extent, however, this effect can be still observed, particularly for very coarse sampling of curvy surfaces.

$$\mathfrak{E} = \sum_{i=1}^{N} \left[\left(\boldsymbol{R} \cdot {}^P\boldsymbol{p}_i + \boldsymbol{t} - {}^Q\boldsymbol{q}_i \right) \cdot \boldsymbol{n}_i \right]^2 \tag{2.64}$$

Equation 2.64 lists the error functional for the point-to-plane goal. Essentially, the ${}^P\boldsymbol{p}_i$ are inserted into the Hesse normal plane equation at ${}^Q\boldsymbol{q}_i$, where $\boldsymbol{n}_i = [n_{xi}, n_{yi}, n_{zi}]$ is the normal vector at that point. Finding the error minimum is a hard nonlinear problem which

can be greatly facilitated by the assumption that incremental rotations from iteration to iteration are small. This leads to an approximation of the full general rotation matrix given in sec. 2.3.1 as follows:

$$\boldsymbol{R} \approx \begin{bmatrix} 1 & -r_z & r_y \\ r_z & 1 & -r_x \\ -r_y & r_x & 1 \end{bmatrix} \tag{2.65}$$

Substituting this approximation matrix into eq. 2.64 yields

$$\begin{aligned} \mathfrak{E} &= \sum_{i=1}^{N} [(p_{xi} - r_z p_{yi} + r_y p_{zi} + t_x - q_{xi}) n_{xi} + \\ &= \quad r_x (p_{yi} n_{zi} - p_{zi} n_{yi}) + \\ &= \quad r_y (p_{zi} n_{xi} - p_{xi} n_{zi}) + \\ &= \quad r_z (p_{xi} n_{yi} - p_{yi} n_{xi})]^2 . \end{aligned} \tag{2.66}$$

With $\boldsymbol{c} = \boldsymbol{p}_i \times \boldsymbol{n}_i$, this can be rewritten to

$$\mathfrak{E} = \sum_{i=1}^{N} [(\boldsymbol{p}_i - \boldsymbol{q}_i)\boldsymbol{n}_i + \boldsymbol{t} \cdot \boldsymbol{n}_i + \boldsymbol{r} \cdot \boldsymbol{c}_i] . \tag{2.67}$$

For this expression the partial derivatives with respect to $\boldsymbol{t}$ and $\boldsymbol{r}$, i.e. in total six unknowns, can be obtained and set to zero. These can be re-arranged into a linear matrix equation.

$$\sum_{i=1}^{N} \begin{bmatrix} c_{xi}\boldsymbol{c}_i & c_{yi}\boldsymbol{c}_i & c_{zi}\boldsymbol{c}_i & n_{xi}\boldsymbol{c}_i & n_{yi}\boldsymbol{c}_i & n_{zi}\boldsymbol{c}_i \\ c_{xi}\boldsymbol{n}_i & c_{yi}\boldsymbol{n}_i & c_{zi}\boldsymbol{n}_i & n_{xi}\boldsymbol{n}_i & n_{yi}\boldsymbol{n}_i & n_{zi}\boldsymbol{n}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{t} \end{bmatrix} = -\sum_{i=1}^{N} \begin{bmatrix} \boldsymbol{c}_i \langle \boldsymbol{p}_i - \boldsymbol{q}_i, \boldsymbol{n}_i \rangle \\ \boldsymbol{n}_i \langle \boldsymbol{p}_i - \boldsymbol{q}_i, \boldsymbol{n}_i \rangle \end{bmatrix} \tag{2.68}$$

This equation can be solved by any linear solver, preferably using Cholesky decomposition, since the large matrix on the most left is symmetric. The rotation matrix $\boldsymbol{R}$ and hence the entire transformation matrix can be computed by using $\boldsymbol{r}$.
Finally, the normal vector computation was done based on linear interpolation between the 3D surface samples. Linear interpolation has been found to be sufficient due to the very high resolution of the MR surface sampling. Moreover, spatial variations across the scanned forehead region exhibit rather low frequencies. For the normal at point $\boldsymbol{q}_j$ two orthogonal tangent vectors along the surface have been computed. The normal vector is obtained by computing the outer product of the two vectors such that the normal is always directed away from the forehead. From the perspective of $\boldsymbol{q}_j$ the surface slope may locally depend on the direction along which one would travel on the surface. Therefore,

the tangent vector combinations have been obtained for tangent vectors of different orientation on the surface. By averaging across the slightly varying normal vector outputs, a more robust estimate is achieved. In addition, this procedure eliminates extrapolation problems at the boundary of a surface where $\boldsymbol{q}_j$ may not have neighboring points in all directions.
Finally, note that the approximation in eq. 2.65 is a heuristic which is not always true. Errors may occur if the surfaces have substantially differing poses with respect to each other. In the ICP, a decrease of the registration error for increasing iteration numbers is hence not guaranteed. In many cases, this can entail longer time until convergence, and in some cases the algorithm would not converge to the correct alignment. In these cases a pre-registration using rough surface landmarks or other approaches should be taken into consideration.

2.3.2.2 Correspondence Step

The optimization problem discussed above requires known correspondences. In the general case they are however unknown and need to be estimated. A second optimization problem arises as follows:

$$\forall \boldsymbol{p}_i \in P_{cld} \; : \; \exists \boldsymbol{q}_j \in Q_{cld} \mid \min_j \left\| \boldsymbol{R} \cdot \boldsymbol{p}_i + \boldsymbol{t} - \boldsymbol{q}_j \right\| \tag{2.69}$$

This aims at finding for each point $\boldsymbol{p}_i \in P_{cld}$ the closest point $\boldsymbol{q}_j$ in Q_{cld} such that the distance given the optimal transformation $\boldsymbol{R}$ and $\boldsymbol{t}$ is minimal. Since the optimal transformation is unknown by itself, the iterative ICP process described earlier needs to be adopted.
It should be mentioned that for the general case of this mathematical problem, a corresponding point $\boldsymbol{q}_j$ is not necessarily member of the discrete set Q_{cld}. The problem in eq. 2.69 can be generalized to any point $\boldsymbol{q}_j$ which would lie on the spatially sampled full surface. Various point-to-plane approaches do so by trying to approximate the penetration point of the surface normal with the second surface [19, 59, 216]. Solving this efficiently adds another challenging problem to the framework for which several proposals have been made. The work presented here will only make use of the problem defined in eq. 2.69.
A simple but highly inefficient solution is the brute force approach. This approach would merely go through all possible choices in Q_{cld} and finally select the closest one. A more efficient way, which has been adopted here, is given by so -called k-d trees.

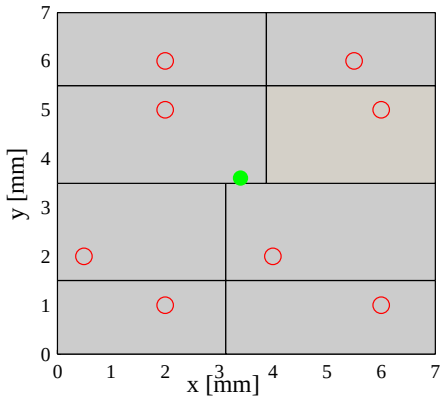


Figure 2.12: Simplified 2D k-d tree example. The red circles denote some data given in 2D space. The line intersections within the plane represent the k-d tree. A new point (green asterisk) can easily be assigned to a corresponding red circle without using a computationally expensive brute force approach.

k-d Trees Similar to Voronoi regions, k-d trees define subregions in a space which are uniquely assigned to given data points [122]. Any new sample residing in that subregion will be assigned to the data point hosting that region. Figure 2.12 gives a simplified illustration of a k-d tree constructed in 2D space.
The partitioning of the full coordinate space can be achieved in various ways. A common approach cycles through all k dimensions, takes the median or mean of all samples with respect to the current dimension and splits the space at the median/mean. In 3D space this split corresponds to a plane where the current coordinate axis is the normal on that plane.
Thus, each node of the binary tree is a decision in one of the k dimensions: If the new sample is smaller than the boundary on this dimension then the sample is assigned to the left child node and otherwise to the right child node. The next decision is then made with respect to a split in the next dimension within the current subregion. The number of cycles through all dimensions depends on the total number of points the k-d tree is built from. The more samples, the finer the space partitioning.
Compared to the brute force search the complexity for finding a corresponding point among N points for a new sample is reduced from $\mathcal{O}(N)$ to $\mathcal{O}(\log(N))$.

2.3.3 Incorporating Additional Knowledge

While the matching step is restricted to the three spatial dimensions, the correspondence step is capable of exploiting additional information. Additional information may be of

any kind – most often including color [71, 105] or directional information locally defined via normal vectors [256]. Extension have also made use of color transformations from RGB to other spaces [194]. This rules out influences of lighting conditions and camera position. The additional information is incorporated into the k-d tree by increasing the search space dimensions. This means that the similarity metric is changed to include more than just spatial information.
This can be achieved in two ways. First, a rejection method can be implemented to eliminate false correspondences. A corresponding pair would be rejected, i.e. not be involved in the matching step, if the similarity in terms of the additional information does not comply with the spatial closeness [71].
As a second option, the most common technique constructs a new distant metric as a weighted sum of different kinds of information [146, 195, 256]. The weighting is an essential normalization to balance the relevance between different pieces of information and to allow for quantities measured in different units.

This work will add scalar tissue thickness information as a fourth dimension to the Euclidean distance. The similarity metric given in eq. 2.69 changes to:

$$\begin{aligned} \left\|\boldsymbol{R}\cdot\boldsymbol{p}_i+\boldsymbol{t}-\boldsymbol{q}_j\right\| &\rightarrow \sqrt{f_{ac1}\cdot\left\|\boldsymbol{R}\cdot\boldsymbol{p}_i+\boldsymbol{t}-\boldsymbol{q}_j\right\|^2+f_{ac2}\cdot\left\|\boldsymbol{d}_{si}-\boldsymbol{d}_{sj}\right\|^2} \\ &\sim \sqrt{\left\|\boldsymbol{R}\cdot\boldsymbol{p}_i+\boldsymbol{t}-\boldsymbol{q}_j\right\|^2+f_{ac}\cdot\left\|\boldsymbol{d}_{si}-\boldsymbol{d}_{sj}\right\|^2} \\ &\quad \text{where } f_{ac}=\frac{f_{ac2}}{f_{ac1}} \end{aligned} \tag{2.70}$$

Here f_{ac} is a weighting factor that balances the relevance between tissue thickness and spatial closeness. It joins the two weighting factors f_{ac1} and f_{ac1} into a single number which can control the entire relevance weighting. An alternative method to incorporate color information into the registration process has been proposed by Men et al. and other groups [127, 130, 194]. They introduced a variant to so-called extended Gaussian images, which correlate surface normal distributions in the Fourier space to get a crude estimate of the rotation matrix [128, 186]. Color information transformed to the Hue space is used to filter the surface information by rejecting points outside a defined subregion of the Hue color space.
Since the ICP algorithm is the most widely used registration algorithm, this work will focus on ICP rather than on alternative methods.

2.3.4 Variants and Extensions

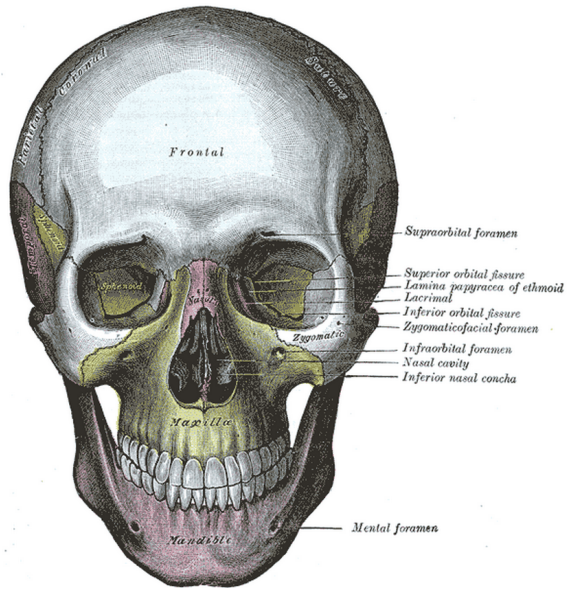
For the ICP algorithm various variants have been proposed to improve speed and precision. A review is given by Rusinkiewicz et al. [243]. The variants include different techniques for point sampling from the surfaces, sample weighting, accelerated matching techniques, or rejection methods. The latter may focus on the rejection of outliers [89, 231, 344] or allow for point cloud registration of limited overlap [243].
Finally, non-rigid ICP extensions should be mentioned as proposed by Amberg et al. [5]. To account for slight deformations in the surface or for noise corruption, the algorithm models locally affine transformations. This means the point clouds are not required to be exactly the same but only similar under certain restrictions. This is achieved by extending the matching objective $\mathfrak{E}$ by stiffness or landmark terms. All terms are weighted and the weights may be changed dynamically. This allows to put a higher weight on a rough landmark based registration at the beginning and to rely on affine transformations only when the optimum is already quite close.
This work will not make use of these non-rigid variants, since benefits of tissue thickness support should be judged in a simple scenario. Using the rigid ICP is also practically more relevant, since it is employed in most clinical devices so far. Therefore, weaknesses of current approaches can be demonstrated and compared to the procedure proposed here. Nonetheless, non-rigid approaches have to mentioned here, since they may constitute the next step. Their dynamic parameter adaptation and their changing methodological states could further serve as an inspiration for a final, universal tracking concept including pre-registration, tissue thickness and non-rigidness.

3 The Simulation of Light-Tissue Interaction

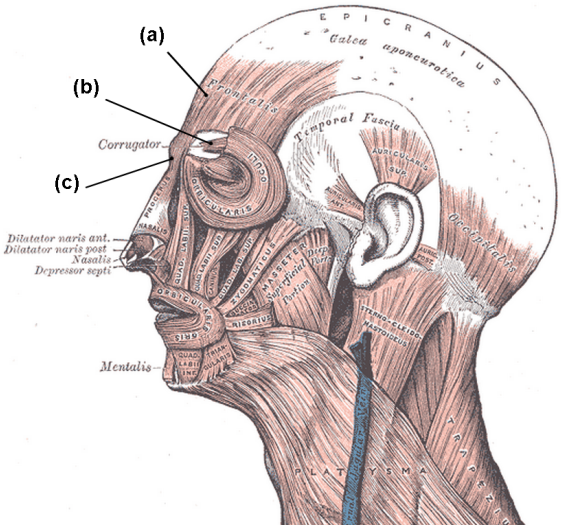
This chapter establishes the theoretical basis for all investigations conducted later on. The main focus is directed to the underlying basis of the proposed concept: the interaction of light with human tissue. The dependency of these interactions on the anatomy is the central requirement for – and actually gives rise to – the novel imaging concept proposed in chapter 1.
After giving general insights into the anatomy and physiology of the human forehead and skin in sec. 3.1, sec. 3.2 formalizes optical properties in terms of a physical skin model and describes how it can be used for simulating the interaction between light and tissue. Section 3.3 then discusses general aspects of how variations in the tissue model result in changes of the backscatter pattern that can be observed by a camera. As the main focus, the effects of changing tissue thickness will be addressed. In sec. 3.4 statistical learning will be presented to model the inverse problem: to predict variations in tissue thickness given its observed effects on optical backscatter. Finally, sec. 3.5 will draw conclusions from the results and discuss how a hardware setup should be specified to optimally measure the relevant effects.
Parts of this chapter have been published in [324] and [325].

3.1 Anatomical and Physiological Background

Knowledge about the anatomy of the head and human skin is vital for a complete understanding of strengths, weaknesses and potential challenges for different head tracking approaches. As indicated in the purpose of this work, the anatomy as well as clinical constraints set up the context a tracking approach has to cope with. Its accuracy and robustness is directly linked to this context and how this context is handled.
Section 1.2.2 outlined that current immobilization devices rely on covering large parts of the head. Access to the head anatomy is mainly given by the patient's face or forehead. The next two subsections are to provide a basic understanding about (1) which parts of



A: Bones of the human skull with the frontal bone accounting for most of the forehead.



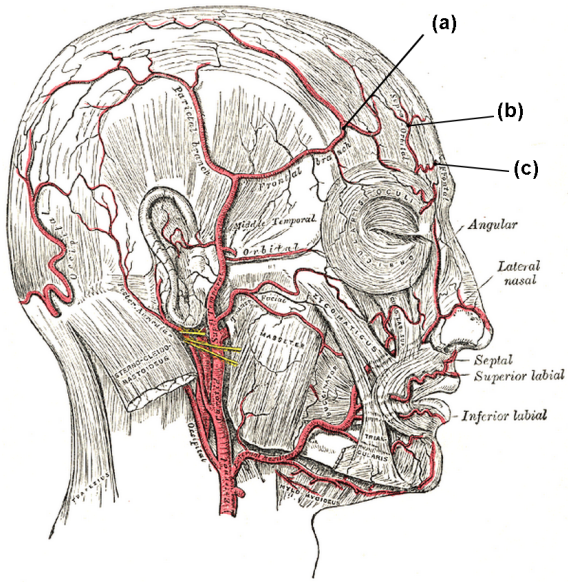
B: Muscles covering the bony structures. For the forehead special focus lies on the (a) Musculus frontalis, (b) Musculus corrugator supercilii, and (c) Musculus procerus.

Figure 3.1: Bony structures and muscles of the human head including the forehead (images from [109]).

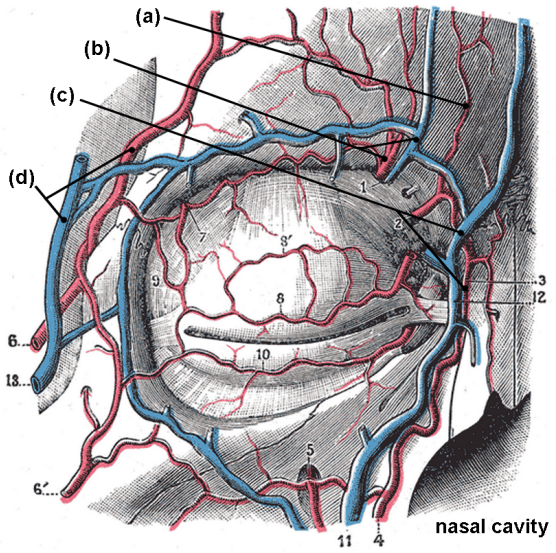
the skull are to which extent covered by elastic soft tissue, (2) which parts of the head are very likely subject to motion or deformation, and (3) the anatomy and physiology of the soft tissue itself.

3.1.1 Anatomy of the Human Forehead

Bone At the forehead region the skull consists of a homogeneous vertical plate – the frontal bone (cf. fig. 3.1A). In caudal direction it connects with the coronal suture to the parietal bone and is mainly bordered by the bony eye orbita in the rostral part of the skull. At the temples, the smooth surface geometry abruptly bends to the back on both sides and only continues with narrow zygomatic processes. On the cheeks, these connect to the zygomatic bones and these finally to the maxilla. The latter further includes the upper jaw and therefore rigidly connects to the upper dentition. In the middle part of the face, the nasofrontal suture connects to the short nasal bone which then ends with the nasal cavity.



A: Extracranial arteries of the human skull. For the forehead the focus is directed to the (a) anterior branch of the superficial temporal artery, (b) supraorbital artery, and (c) supratrochlear (frontal) artery.



B: Detailed frontal circulation: (a) supratrochlear (frontal) artery, (b) supraorbital artery and vein, (c) angular artery and vein, (d) anterior branch of the superficial temporal artery and vein.

Figure 3.2: Arteries (red) and veins (blue) relevant for the forehead region (images from [109] where full labeling is available).

Muscles The frontal bone is widely covered by the musculus frontalis which is part of the musculus epicranius (cf. fig. 3.1B). This muscle is thickest above both eyes and thinner between them. With about 1 mm thickness the muscle is generally rather thin. On the upper part of the forehead it then connects to the galea aponeurotica - a fibrous connective tissue that is connected to the skull. As part of the mimic musculature, the musculus frontalis is used for lifting the eyebrows and causes wrinkles of the skin covering the forehead. Above the root of the nose, the muscle mixes with the musculus procerus, a small muscle covering the nasal bone. It pulls the eye brows in caudal direction. Similar to the musculus frontalis it is grown together with the dermis such that the skin follows the muscle movements. In this specific case the procerus is responsible for the frown and laugh lines between the eyebrows. Finally, wrinkles are also generated by the musculus corrugator supercilii which pulls the eye brows down and inwards.
Both, sensation and muscle movement are achieved by branches of the trigeminal (mainly facial sensation) and facial (facial expressions) nerve [246]. Particularly the supraorbital and supratrochlear nerve have to be mentioned [157] as nerves that run withing the subcutis of the forehead.

Vascular System Various vessels ensure sufficient blood supply for the skin as well as muscles. Figure 3.2 provides an overview of the artery (colored in red) and vein (colored in blue) system as well as their interrelations. The arterial oxygen saturation of a healthy adult is generally above 95 % [346]. The saturation of veins in contrast ranges typically between 60 % and 70 % [116]. The saturation defines the number of oxygenated hemoglobin molecules relative to deoxygenated ones, which is lower for veins. Transporting blood to and from the destination, many of them occur in corresponding pairs (cf. fig. 3.2B).
As one of the larger arteries, the superficial temporal artery runs up the temporal part of the head and bifurcates into a frontal and a parietal part. The former of these reaches across the temples to the forehead. Figure 3.2B illustrates that the corresponding vein behaves in a similar way. Apart from these terminals reaching the forehead from the side, there are two further vessels arising from the bony eye orbit: (1) the supraorbital, and (2) the supratrochlear (also frontal) artery and vein. From both orbits they run more or less symmetrically upwards on the forehead and terminate roughly at the hairline. The latter is located more medially, i.e. closer to the root of the nose. The supraorbital artery springs from the ophthalmic artery at the optic nerve. Besides the dorsal nasal artery, the supratrochlear artery then represents one of two further terminal branches of the ophthalmic artery. The dorsal nasal artery joins the angular artery on the side of the nose. Similar locational considerations can be made for the corresponding veins (cf. fig. 3.2B).

3.1.2 Anatomy, Physiology and Characterization of the Human Skin and its Adjacent Structures

With an area of up to 2 m^2 the skin constitutes the largest human organ [294]. Containing various nerves and vessels, it is responsible for sensation and thermal regulation of the body. By bordering the human body it grants protection with respect to physical, chemical or radiation influences. On average, it can reach thicknesses of up to 2-3 mm [142, 242, 294] and as histological evidences suggest may vary by 30-40 % from body region to body region [193]. Skin found at the head and particularly on the forehead is typically thinner. On the forehead, studies found maximum thicknesses from 1.5-2 mm [114, 219]. The subcutis including adipose tissue adds significantly to the overall thickness down to the cranial bone. It can reach a thickness of several millimeters [193]. All anatomical layers, including skin and subcutis, between the outer surface and the cranial bone will be summarized with the term "tissue" in this work.

The skin can be separated into several layers according to thickness, composition

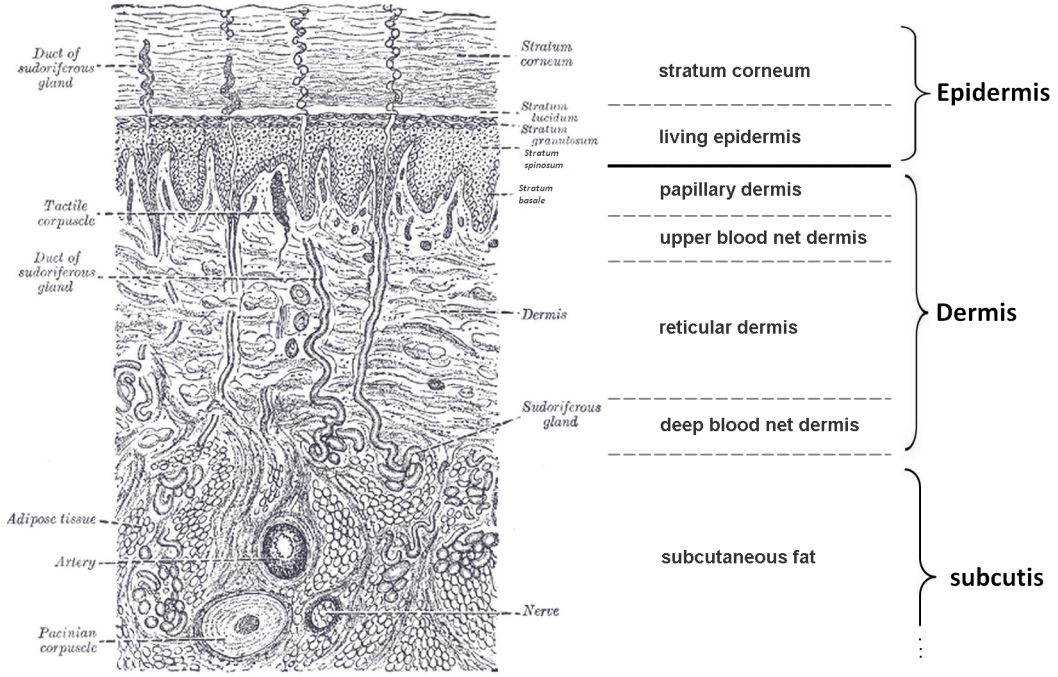


Figure 3.3: Cross section through different layers of the human skin. While the left illustrates content and morphology of the layers and their interfaces, the right subsumes the detailed structure into seven distinct layers (image from [109]).

and morphology. The layering can be defined in different stages of detail, whereas in practise there are only partly sharp transitions between them [294]. The layers of the skin (cutis) can be roughly organized into epidermis (collective term for the upper layers) and dermis (collective term for the lower layers). The subcutis being located below the cutis contains mainly subcutaneous fat, but also muscles and the periosteum, a thin tissue layer covering the bone. For the distinction of further layers, this work will mainly refer to the definitions of Bashkatov and colleagues [17]. Figure 3.3 illustrates the layering which will be discussed in the following.

Epidermis The epidermis is about 100 µm thick and consists of viable and non-viable tissue. The most outer layer, the stratum corneum, consists of flattened, non-living keratinocytes and is about 20 µm thick. The dead cells have a high lipid and protein, but very low water content [17]. It functions as a protective barrier against mechanical injury as well as uncontrolled passage of water, toxic substances or microorganisms. In case of darkly pigmented or tanned individuals, the kertinocytes may contain finely grained melanin dust [142]. Melanin is a high molecular weight polymer that is attached to a

structural protein. It is one of the most dominant pigments or chromophores of the human skin and in parts responsible for the skin color [142]. There are two types of melanin: the brownish-black eumelanin and the reddish-yellow pheomelanin [288]. Both may vary from individual to individual, whereas the proportion of the former is increased in dark skin and the latter in red-haired northern European phenotypes. Due to the nature of melanin production, constitutive (gens, hormones) and facultative pigmentation (tanning, radiation) are distinguished.
Melanin is produced in cytoplasmic organelles of the melanocytes called melanosomes. Melanocytes are dendritic cells whose dendrites reach out to the kertinocytes. Melanosoms are transported upwards through the dendrits and then form a mass or clumps in the stratum corneum.
The melanocytes themselves reside among epithel cells in the living epidermis (stratum germinativum) which consists of the sublayers: stratum lucidum, stratum granulosum, stratum spinosum and startum basale. It is about 80 µm thick. As cell growth happens in the lower epidermal layers, older basal cells migrate upwards to the stratum corneum and meanwhile underlie the process of cornification. At the very surface, dead cells form a rough, plate-like structure and are pushed off.

Dermis The stratum basale forms a rather sharp transition to the underlying dermis. These basal lamina are a relief of contiguous, branching ridges, valleys, crevices, and craters. Oriented in a roughly perpendicular manner, fine elastic fibers from the dermis connect to them [193, 200]. Due to its morphology this 100-150 µm thick layer is called papillary dermis.
A further distinction of layers within the dermis is given by the distribution of blood vessels. The upper blood net plexus is 80-100 µm thick and contains fine capillaries and microcirculation [200]. Underneath the reticular dermis – the main part – follows the deep blood net plexus (100-200 µm). The latter contains larger veins and arteries, while the largest reside in the subcutis. Compared to arteries, veins rather tend to run in upper instead of deeper layers of the skin.
The reticular dermis considerably varies in thickness and can reach between 1 mm and occasionally 4 mm [17]. Typical thicknesses – also for the forehead – region are about 1.5 mm [193]. The dermis is equally thick in black and white individuals [142]. It consists of interwoven bundles of collagen and elastic fibers and hence mainly contributes to maintaining the tensile strength and elasticity of the skin. Collagen constitutes 77 % of the fat free skin weight [142]. Additional components include interfibrillar gel, water, nerves, sweat glands, the pilosebaceous complexes of hair follicles and sebaceous glands.

Subcutaneous fat The adipose tissue mainly contains adipocytes which contain lipids in small droplets. Lipids account for about 95 % of a single fat cell. Arterial and venous capillaries, nerves and reticular fibrils reside in the intercellular spaces. Subcutaneous fat is rich in its water content. As for the dermis, oxygenated and deoxygenated hemoglobin represent the dominating chromophores. At the forehead, the fat layer is typically the thickest layer and can vary from individual to individual.

Cranial Bone Cranial bone consists of a highly mineralized inorganic and organic matrix. The former contains calcium hydroxyapatite and osteocalcium phosphate. This matrix is hence responsible for the physical strength and the rigidness of the bone. The organic matrix consisting of collagen, proteins, blood cells and lipids makes the bone strong. The amount of the bone mineral matrix corresponds to 15 %, the lipid content to 54 %, the protein content to 16 %, and water contribution to 16 % [100].

Aging and Gender Influences on the Skin Various studies confirmed correlations between skin thickness or thickness changes with age and gender [36, 265]. While the thickness of the human skin was reported to be stable between the age of 15 and 70, it increases for individuals younger, and decreases for individuals older than that range [81]. Epidermal degeneration was found to happen faster for men than for women [36]. Males generally have thicker skin than females mainly due to a larger amount of elastin and collagen within the dermis. This also entails better mechanical strength for men. Frenske et al. reported atrophy in the dermis increasing with age [85]. A reduction of the collagen fibers and integrity of the elastic fibers leads to degenerated elasticity, while there was no dependency with age found for skin extensibility [81].
With age, the skin also starts loosing its capability for binding water which leads to a generally dryer skin. The number of hair follicles, sebaceous and sweat glands decreases [294]. Blood microcirculation is destroyed and disorganized, while the endings of sensing elements stay intact. All these changes and moreover also the progressive flattening of the papillary structures due to changes in the elastic fiber connections cause wrinkles on the macroscopic scale [200].
Since the number of melanocytes decreases as well, melanin production is reduced for elderly individuals leading to paler skin [85]. However, particularly for the face, age spots may occur where melanin production is partially increased [294].
Subcutaneous fat also reduces with age. In addition, considerable site to site variations in the thickness of the fat layer have been found. The extent of variation depends on the age and has its highest state of communality during adolescence. More changes in the

subcutaneous fat have been reported for women [266]. Generally, the loss of body weight also mainly affects the fat layer and thickness changes thereof. The physiology of other layers is rather affected by a possibly related reduction of the water content. Furthermore, drastic weight changes can also have indirect, physical effects on tissue layers. For the forehead region this effect is, however, mainly negligible.
Finally, it should be mentioned that aging does not only include the biological age, but also aging by exposure to radiation [142].

Skin Type Classification Older qualitative measures of skin type mainly relied on the color of skin. The simple von Luschan scale [142, 311] matches tiles of different colors and hues to the color of unexposed skin. Going one further step, clinical medicine traditionally and up to date relies on skin phototypes or sun-reactive types for fast skin as well as skin sensitivity classification [106, 218, 245].
This so-called Fitzpatrick scale consists of six general skin types and categorizes according to medical interrogation or questionnaires [93, 94]. Criteria covered in the questioning procedure are related to eye color, natural hair color, freckles, constitutive skin color (unexposed areas), skin sensitivity, tanning habits and tanning history [95]. It thus involves genetic, environmental and habitual influences. Scores result in the following typing.

Skin Type I	Always burns, never tans	(pale white skin)
Skin Type II	Always burns easily, tans minimally	(white skin)
Skin Type III	Burns moderately, tans uniformly	(light brown skin)
Skin Type IV	Burns minimally, always tans well	(moderate brown skin)
Skin Type V	Rarely burns, tans profusely	(dark brown skin)
Skin Type VI	Never burns	(deeply pigmented dark brown to black skin)

For categorizing purposes in ths project, a questionnaire was designed according to [95] and [230]. Although the Fitzpatrick scale is a measure acquired very fast, it is subjective and does not correspond well to constitutive skin color. It has shown limited applicability to moderately or deeply pigmented skin [142]. One further weakness is that it is strongly biased to skin properties in the visible and particularly ultraviolet (UV) light.

3.2 Simulation Approach and Skin Model

3.2.1 Optical Properties and Models of Human Skin

An advantage of qualitative skin typing as described before is that it is fast and does not require any instrumentation. However, it involves subjective measures possibly inducing a high inter-operator and inter-patient variance. To a certain extent there is also intra-operator and intra-patient variance. Therefore, quantitative measures of optical tissue properties are obtained by *in vitro* experiments using spectrophotometers with integrating spheres [17, 100, 295]. The actual parameters are then obtained by applying the inverse version of one of the approaches briefly described in sec. 3.2.2.2.
Although morphology and anatomy of the human skin are quite inhomogeneous, it is still possible to define regions, where the gradient of skin cell structure, chromophores and blood amount roughly equals to zero [17]. Referring to the definitions used by Bashkatov et al. [17], Meglinksi et al. [193] as well as Petrov et al. [224], a seven layer skin model is adopted in this work. It was extended by cranial bone as an eighth layer. Location and order of the tissue layers are illustrated in fig. 3.3. The following wavelength-dependent parameters have been defined for each layer to obtain a full descriptive model of the layered skin [68, 70, 191–193, 224].

Refractive Index

The refractive index n describes a dimensionless number relating the light wave's phase velocity to the speed of light c. It therefore describes how light propagates through a medium and under which angle the light path changes when entering another medium. The latter is described by Snell's law [63]. Here, the relationship of the refractive indices of two adjacent media is of importance. For the skin model, the refractive indices cover a range between 1.3 and 1.6, while that of the ambient medium is 1 (cf. table 3.1). The wavelength dependence of the refractive index for soft tissue was considered negligible as compared to the variation across layers [192].

Absorption Coefficient

The absorption coefficient μ_a [mm^{-1}] equals the average number of absorption events per unit path length of photon travel [17]. When light hits chromophore particles and the photon energy corresponds to the distance between atomic or molecular energy levels, then the photon gets absorbed. After this excitation process of the atom or

Table 3.1: Parameters of the standard Caucasian skin model at 850 nm. This model was used for evaluating general aspects of the backscatter characteristics. Data taken from [68, 192, 193, 224].

layer	n	μ_a [mm^{-1}]	μ_s [mm^{-1}]	g	d [mm]
stratum corneum	1.53	0.031	38.34	0.87	0.02
living epidermis	1.34	0.106	26.72	0.87	0.08
papillary dermis	1.40	0.056	23.13	0.87	0.10
upper blood net dermis	1.39	0.075	19.08	0.88	0.08
reticular dermis	1.40	0.063	15.36	0.87	1.62
deep blood net dermis	1.39	0.105	7.43	0.88	0.20
subcutaneous fat	1.44	0.066	3.85	0.87	2.90
bone	1.56	0.048	16.55	0.90	5.00

molecule, the energy gained during this process is – in most cases – transformed into thermal energy. The re-emission of energy in terms of photons such as in fluorescence is also possible. This will, however, not be modeled by subsequent simulations, since it only constitutes an extremely rare scenario for the problem at hand. Finally, an absorption shadow occurs behind the particle, which is not necessarily the size of the chromophore. Thus, the absorption coefficient can also be viewed as the cross-sectional area of these shadows per unit volume of medium. The coefficient is wavelength dependent.

Scattering Coefficient

The scattering coefficient μ_s [mm^{-1}] equals the average number of scattering events per unit path length of photon travel [17]. When light hits chromophore particles, its path changes due to deflection. Due to this deflection a scattering shadow is formed behind the particle which is not necessarily the same size as the particle. Thus, the scattering coefficient can also be viewed as the cross-sectional area of the shadows originating from scattering per unit volume of medium. The coefficient is dependent on the geometrical particle size (relative to the wavelength), shape and concentration in the medium. It also depends on the wavelength.
Light can be scattered at irregularities within the medium or at boundaries between media. If the nature of the backscatter from a surface is random and single scattering events are averaged out, the process is termed diffuse reflectance R_D. This is in contrast

to specular or regular reflectance R_{sp} which originate from rather smooth surfaces. For skin, regular reflectance is about 4 % to 7 % of the incident light [7].
Scattering in tissue is mainly described by two different types: Rayleigh and Mie scattering [7]. Rayleigh scattering happens with particles sizes less than 10 % of the incident wavelength. It is a weak and isotropic process which decays with the fourth order of the wavelength. Mie scattering, in contrast, happens with larger spherical particles and is strongly directed forward, i.e. non-isotropic. As shown in eq. 3.1, it only weakly depends on the wavelength and hence dominates longer wavelengths.

$$\mu_s^{layer}(\lambda) = F \cdot (\underbrace{2.2 \cdot 10^{11} \cdot \lambda^{-4}}_{Rayleigh} + \underbrace{11.74 \cdot \lambda^{-0.22}}_{Mie}) \tag{3.1}$$

In this equation F denotes a coefficient ranging from 1 to 10 and defines the layer-specific extent of scattering [7, 192, 224].

Anisotropy Factor

The anisotropy factor g (cf. table 3.1) describes the probability distribution p over possible scattering directions. A direction is defined by angle θ_{ph}, whereas $\theta_{ph} = 0$ gives the forward direction of the traveling photon. The functional relationship goes back to Henyey and Greenstein [124] who devised a function $p(\cos(\theta_{ph}))$ that is parametrized by g (cf. eq. 3.2).

$$p(\theta_{ph}) = \frac{1 - g^2}{2\,(1 + g^2 - 2g \cdot \cos\theta_{ph})^{\frac{3}{2}}} \tag{3.2}$$

The function is chosen such that the expected value for $\cos(\theta_{ph})$ according to this probability function is exactly g. It thus expresses dominant directions in the scattering process, where $g = 1$ gives the forward and $g = -1$ the backward direction. For biological tissue such as skin and for bony structures, g has been found to be between 0.7 and 0.9 [100, 295, 301]. It smoothly increases with the wavelength [192]. Together with the anisotropy factor g, the scattering coefficient is often turned into the reduced scattering coefficient $\mu_s' = \mu_s(1 - g)$.

Layer Thickness

The absolute thickness of a skin layer is denoted by d [mm] and the full tissue thickness by d_s [mm]. Typical mean values mentioned in sec. 3.1.2 were taken from various sources (cf. table 3.1).

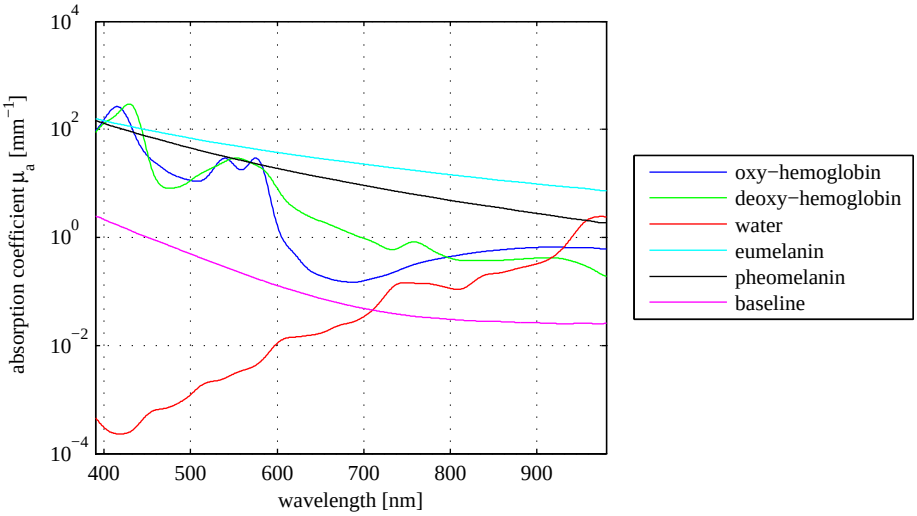


Figure 3.4: Absorption coefficients per wavelength for different chromophores of the human skin. (data according to [224])

Parametrization of Individual Skin Layers The behavior of the aforementioned parameters is substantially different among the skin layers. This is mainly due to a different morphology as well as differences in: the distribution, amount, and concentrations of the most important chromophores contained in each layer.
Anderson et al. [7] state that in the epidermis scattering is due to the refractive surface and – on a smaller scale – particle scattering. Large melanin particles such as melanosomes have diameters of more than 300 nm and cause Mie scattering. Melanin dust with diameters below 30 nm is rather described by Rayleigh scattering. In a darkly pigmented or tanned epidermis the cell compound is more dense and the number of scatterers higher [142]. Overall, the optical properties of this layer are, however, dominated by absorption. While absorption at nucleic acids, peptide bonds or aromatic molecules such as urocanic acid is relevant for low wavelengths (particularly below 400 nm), melanin is the main absorber within the UV and visible light range. It naturally protects the skin from too much UV exposure. As illustrated in fig. 3.4, the absorption coefficient for both melanin types decreases in the visible range. Especially above 1100 nm NIR absorption of melanin is negligible [7]. While most chromophores occur more or less in similar quantities across individuals, there is a substantial individual variation in the distribution and amount for melanin, which is related to the skin type described before.
The dependencies for these and other chromophores are shown in fig. 3.4. They are the result of a comprehensive meta-study conducted by Meglinski et al. [191]. The plot also

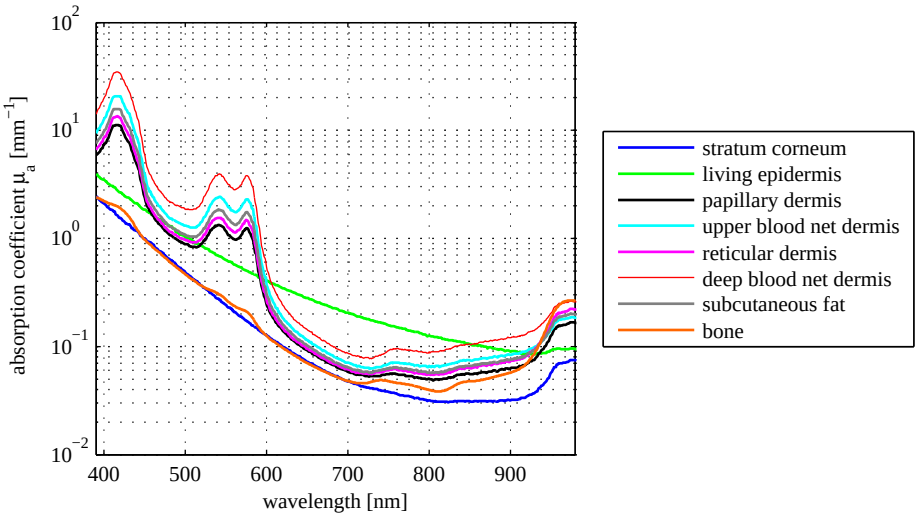


Figure 3.5: Resulting absorption coefficients per wavelength for different layers of the human skin (data according to [224]).

shows that water absorption increases with the wavelength. Above 1000 nm light is mainly blocked by water. This is essential since the water content of the skin is high and moreover sweat, which also contains urocanic acid, will significantly change the optical properties of the skin.

Main absorbers within the dermis are oxy- and deoxy-hemoglobin, β-carotine, bilirubin and water. As illustrated in fig. 3.4, oxy- and deoxy-hemoglobin predominantly absorb light of lower wavelength, whereas the coefficient rapidly decreases with the wavelength. This renders the location and occurrence of blood vessels within different skin layers important and distinguishes their properties from other layers. All aforementioned chromophores, which are not shown in fig. 3.4, are summarized within the term "baseline".

Due to its properties and thickness, the dermis also dominates and determines most of the skin's scattering behavior [7]. It contains very thin collagen and elastin fibrils of size 60-100 nm as well as bundles thereof reaching 1-8 µm in size [7]. Due to the domination of Mie scattering in the dermis – particularly in the NIR and IR range – there is only a weak wavelength dependence for scattering in this layer.

Absorption within the subcutaneous fat is mainly determined by hemoglobin, lipids and water. Spherical droplets of lipids are also the main scatterers. The average size for the adipocytes containing these droplets ranges from 50 to 150 µm, whereas the mean size of adipose scatters within them has a diameter of 0.8 µm [7].

The optical properties of the bone are dominated by its water content, collagen fibers and carbonated apatite nanocrystals of about 30 nm length [100].

All in all, the behavior of the chromophores leaves a therapeutic window from 600 nm to 1300 nm for the clinical context where the penetration depth of light into skin is maximal. Anderson et al. state that around a wavelength of 1000 nm the optical penetration depth (energy decay to 37 % of its incident power) reaches 3.5 mm and that 1 % of the light penetrates the entire chest wall between 605 nm and 850 nm.

Taking all these considerations together, Meglinski et al. [192] were able to approximate functional relationships of the absorption coefficient, scattering coefficient and anisotropy factor with respect to the wavelength. These equations also stated in earlier articles [68, 70, 191, 224] depend on the following parameters: the concentration of blood and melanin, the water content, oxygen saturation, the volume fraction of the blend between the two melanin types, the hematocrit index (volume fraction of cellular content in blood), the volume fraction of hemoglobin and the fraction of erythrocytes. These are used as weighting factors for the graphs shown in fig. 3.4 to result in layer specific graphs for the absorption coefficient presented in fig. 3.5. The anisotropy factor moreover reveals a dependency on correction factors which take the vessel size into account. For large vessels the hemoglobin content in the central vessel part is hardly reached and does hence not affect the scattering as much [192]. The weighting factors F for the general graphs of the scattering coefficient described in eq. 3.1 correspond to [11.04, 7.73, 6.61, 5.51, 4.41, 2.20, 1.10] (skin layers from superficial to deep). The properties of the cranial bone have been taken from references [88, 99, 100, 224, 299].
Finally, note that these properties are valid only for average Caucasian skin of medium age and are different for other skin types. The values found in the literature also underlie certain variations which may be due to the natural dissipation of tissue properties, differences in tissue preparation and storage methods. Further on, temperature conditions during *in vitro* experiments influence the outcome. A variability of properties due to temperature *in vivo* is negligible because of the thermal regulation of the body temperature and homeostasis.

3.2.2 Simulation of Light Transport in Soft Tissue

The theory of radiation transfer provides a rich framework to model the flow of energy in terms of photons in tissue. In this context, Bashkatov et al. [18] distinguish direct and indirect approaches for modeling the light transport through a specimen of tissue. Direct

methods employ fundamental rules and simple analytic expressions such as the Beer-Lambert law. These approaches are however very specific and require a strict fulfillment of the experimental conditions set.
Indirect methods utilize a theoretical model for light propagation in a certain medium. These models are very often based on or related to the radiation transfer equation. This is a differential equation describing the radiance L with respect to the spatial position, its direction and unit time. The general form of the equation has six degrees of freedom and has no general closed-form solution. The following subsection gives an overview of several indirect methods that can be used to simulate light transport through tissue. They all solve the forward problem, i.e. computing measures such as diffuse reflection R_D or total transmittance T_t[1] from given optical parameters of the tissue. They can readily be used to tackle the inverse problem, i.e computing the optical parameters discussed in sec. 3.2.1 from measured diffuse reflection R_D or total transmittance T_t.

3.2.2.1 Approaches for Modeling Light Transport in Tissue

Kubelka-Munk Theory The two-flux Kubelka-Munk model [161] describes the attenuation of a light beam caused by scattering and absorption. It assumes two fluxes that counter-propagate a slab of tissue. As the incident flux, the optical flux propagates downwards, i.e. into the tissue, and is attenuated by absorption and scattering. It superposes with the counter-propagating flux caused by backscatter from the tissue. The model gives rise to differential equations which can be solved with respect to two parameters describing absorption and scattering. Together with an approximation of the radiation transfer equation that results from diffusion theory, measures for the coefficients μ_a and μ_s can be obtained [18].
The results are only valid given that scattering is significantly dominant over absorption, i.e. $\mu_a << \mu_s$ [17, 18]. Further limitations are that all fluxes are assumed to be diffuse and that radiation lost at the edges of the tissue sample is required to be negligible. The model does not account for reflections at boundaries with mismatched refractive index.
To resolve the restriction to diffuse fluxes, extensions to a four, or multi-flux model have been proposed [67, 185]. These also introduce collimated or directional fluxes. The approach was widely used during the last decades [7, 74, 84, 207, 303] before the evolution of more sophisticated methods described below. The Kubelka-Munk model is still used as an initial guess for these iterative methods [18].

[1] Total amount of light passing through a sample or material.

Adding-Doubling Method The adding-doubling method considers the transport equation for plane-parallel layers. It is based on the idea to solve the radiation transfer equation for a very thin tissue slab. This provides a simplified setting for which the solution is relatively easy [300]. This general idea was introduced to tissue optics by Prahl et al. [229]. Mainly used in its inverse adding-doubling version, it can be used to obtain accurate estimates for μ_a, μ_s and g [248]. Internal reflections at sample boundaries are taken into account.
Having the solution for a very thin slab, the term doubling refers to the idea that transmittance and reflectance of a thicker slab – twice as thick as the initial one – can be computed from the transmittance and reflectance of the thin slab. By iteratively superimposing thin layers, reflectance and transmittance are consecutively joined to their final approximation.
The term adding allows for heterogeneous multi-layerd tissue. Taking the variation of refractive indices between the layers into account, the solution for more complex layered tissue is computed by "adding" their contributions. The adding doubling method remains a popular method for solving the one-dimensional transport equation for a slab geometry [17, 100, 295]. The method is limited as it cannot account for a finite beam size or side losses of light at the sample boundaries. It also lacks the model flexibility Monte-Carlo methods can provide [18]. However, this numerical solution to the steady-state transfer equation is still valid for cases in which scattering and absorption are equally important.

Diffusion Approximation Diffusion theory introduces assumptions to approximate the radiation transfer equation and to reduce the number of degrees of freedom. The most essential assumption is, that the medium is highly scattering but only weakly absorbing, i.e. $\mu_a << \mu_s$ [275]. This remains a good approximation for most biological tissue types, but represents a source of inaccuracy. It approximates the temporal properties of light transport well and provides a solution with high computational efficiency.
Apart from being restricted to highly scattering media, studies have argued and shown that diffusion theory is not valid for very thin media [112, 193]. In multi-layer setups, poorest results are obtained for layers close to the surface. Guo et al. showed that inaccuracies occur for samples up to 4 mm thickness. There is still an ongoing discussion about the correct definition of the diffusion coefficient [112]. No ballistic photon transport can be modeled.
Finally, Wang and colleagues [318] identified the conversion from the desired narrow laser beam to an isotropic point source (as required by the theory) as one of the main

factors for inaccuracies.

Monte-Carlo Methods Monte-Carlo methods are capable of simulating settings, where (1) the geometry of the experiment, and (2) the actual tissue structure is complicated. It efficiently joins both, the functional description of light propagation in tissue as well as a specialized object geometry the light is supposed to interact with. This ensures a highly flexible tool for versatile investigations. Interactions with special geometries and distributed optical properties have also been studied using diffusion theory in combination with finite-elements methods [144, 249, 262].
The functional description of light transport is given by the radiation transfer theory as described in more detail in sec. 3.2.2.2. Using this theory, single photon trajectories through the tissue can be simulated. The term Monte-Carlo thereby refers to the probability distributions that govern the path of an individual photon. Simulating a large number of photons (typically $> 10^5$) provides the complete tissue response. The entire simulation considers one point in time only. Therefore, it corresponds to the impulse response for an infinitely short dirac pulse of light. Convolution with a time signal can provide the response for arbitrary light exposure over time. Spatially speaking on the other hand, the simulation of a single pencil beam in laterally homogenous tissue can be seen as the impulse response.
Monte-Carlo simulation of light transport in multi-layered tissue (MCML) was first introduced by Jacques and Wang [317]. It takes several optical properties such as scattering and absorption coefficient, refractive index or anisotropy factor into account, allows for reflections on tissue boundaries, sideways loss from the tissue sample and most important is valid for arbitrary relationships between μ_a and μ_s [18].
The random sampling of many photon trajectories entails one severe drawback: It requires extensive calculations [17]. However, given that each photon is subject to the same functional framework, there is a high potential for parallelization and computational optimization. Recent research directed special focus on outsourcing computational load to the graphics processing unit (GPU) using e.g. the compute unified device architecture (CUDA) [4, 68, 69]. This significantly speeds up the computational part by parallelization and usage of dedicated and specialized hardware components.
Since this relaxes the efficiency problem and time was not essential, the MCML approach was adopted for this work, because (1) it is valid for arbitrary combinations of absorption and scattering coefficients, (2) it straightforwardly allows for modeling complex multi-layered tissue, (3) it is accurate for thin and thick media, (4) boundary effects at interfaces between the layers and between a finite sample and the ambient medium are accounted

for, and finally (5) it provides an ideal framework for customization, extracting various measures of interest and dedicated investigation in general.

3.2.2.2 Monte-Carlo Simulation of Light Transport in Multi-Layered Tissue

This work made use of the functional framework described by Wang et al. and others [68, 193, 317]. The implementation built up on the basic CUDA accelerated framework from Alerstam and coworkers [4] which is provided as an open source code package called `gpumcml`. It was extended and customized for the application at hand as will be outlined in the following. All computations where performed on an Intel® Core™ i5-2500K CPU @ 3.30 GHz, 16 GB RAM with an Nvidia GeForce GTX 470. For similar graphic cards Alerstam et al. [4] showed speed-ups of up to factor 800 compared to normal central processing unit (CPU) execution. The following paragraphs will review the functional framework and discuss the changes and extensions made to the program.

The general MCML framework The Monte-Carlo principle is based on random uniform sampling. The random variable ζ is sampled from a uniform distribution $\mathcal{U}$.

$$\zeta \sim \mathcal{U}(0,1) \quad i.e. \quad \zeta \in [0,1] \tag{3.3}$$

Random sampling is involved in the determination of the photon step size per iteration, the decision whether a photon is reflected at a boundary, azimuth and altitude determination for particle scattering as well as Russian roulette for very low weight photons. The MCML simulation runs in iterations and threads on the GPU. One thread corresponds to one photon packet, which will be treated as one simultaneously acting photon object of specific energy, i.e. photon weight w_{ph}. This weight is dimensionless and is initialized with $w_{ph} = 1$ at photon launch. An iteration on the other hand corresponds to the computation of one random photon step in the tissue and possibly the occurrence of several interaction events discussed hereafter.

A photon is equipped with the following properties:

1. current photon weight w_{ph}
2. 3D photon location $[x_{ph}, y_{ph}, z_{ph}]$
3. photon moving direction in terms of directional cosines $[ux_{ph}, uy_{ph}, uz_{ph}]$
4. photon step size s_{ph}
5. number of steps taken so far$^{(*)}$

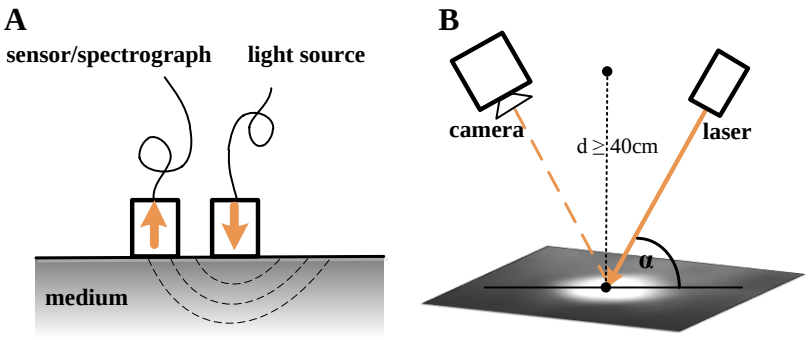


Figure 3.6: Conceptual comparison between two experimental setups. **A:** spectroscopy: light source and detector are probes directly attached to the medium; **B:** laser scanning setup: the setup treated in this work applies a laser beam of certain profile from a distant source to the medium, likewise a camera observes the scene from a distance. (© 2013 OSA. Reprinted, with permission from [325]).

6. distance traveled so far$^{(*)}$
7. time traveled so far taking the refractive index into account$^{(*)}$
8. the current layer the photon resides in
9. the history of layers visited so far$^{(*)}$

Properties such as location, direction, step size and energy are tracked and constantly used by the simulation. The properties labeled with an asterisk have been added in this work for tracking extra properties such as the total number of steps taken, which gives a rough impression of polarization changes while traveling. More sophisticated work on polarization tracking has been published by Ramella-Roman et al. [235, 236]. Further on, it can be measured how far an individual photon has traveled and for how long, before leaving the tissue. The latter accounts for the dependency of photon velocity on the refractive index of the current layer medium. Finally, it can be identified which layers the photon has visited.

Photon Launching The original simulation also assumes cylindrical symmetry [4, 68, 317]. This has been adapted to a Cartesian coordinate frame, which allows simulating customized beam profiles instead of pencil beams only. It also makes it possible to observe changes in the 2D shape of the backscatter distribution on the surface. The beam profile for the simulations was set to a Gaussian beam profile with variance $\sigma^2 = 0.01\,\mathrm{cm}^2$.

At photon launch, i.e. at the start of one thread, the photon was equipped with initial properties being appropriate to model the experimental setup outlined in fig. 3.6B. In contrast to typical setups for spectroscopy (cf. fig. 3.6A), the proposed laser scanning setup does not consist of a detector and a source probe that are directly attached to the tissue surface. Instead, a laser beam emits photons under a certain angle α onto the surface. This happens contactlessly and from a distance of more than 40 cm. The angle is zero for irradiation orthogonal to the surface and 90° parallel to it. In this work $\alpha = 0$ was considered the default case, but generally scenarios covering $\alpha \in [0, 45°]$ were investigated.
The initial start position for a photon is sampled randomly from a 2D Gaussian distribution with mean $\boldsymbol{m}_{ph} = [m_x, m_y]$ and variance $\boldsymbol{\Sigma}_{beam} = [\sigma_x^2, \sigma_y^2]\boldsymbol{I}$, where $\boldsymbol{I}$ is the unit matrix

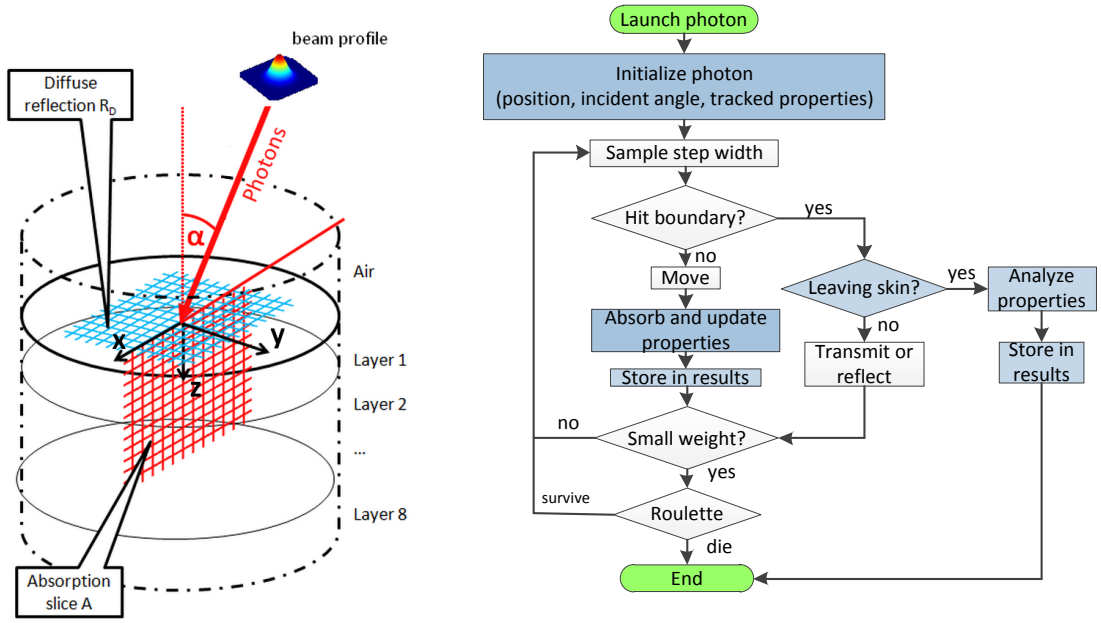
$$[x_{ph}, x_{ph}] \sim \mathcal{N}(\boldsymbol{m}_{ph}, \boldsymbol{\Sigma}_{beam}) \tag{3.4}$$

The directional cosines are initialized according to the chosen α, i.e. $\alpha = 0$ yields $[ux_{ph}, uy_{ph}, uz_{ph}] = [0, 0, 1]$ and an arbitrary α yields $[ux_{ph}, uy_{ph}, uz_{ph}] = [\sin(\alpha), 0, \cos(\alpha)]$. The weight of the photon packet is initially set to one, before a proportion according to specular reflectance is subtracted. This lost proportion equals the specular reflection R_{sp} resulting from the Fresnel equations [4, 317]:

$$R_{sp} = \frac{1}{2} \cdot \left[\frac{\sin^2(\alpha - \alpha_t)}{\sin^2(\alpha + \alpha_t)} + \frac{\tan^2(\alpha - \alpha_t)}{\tan^2(\alpha + \alpha_t)} \right] \tag{3.5}$$

The equation corresponds to the average across the two polarization states, since this MCML implementation does not model polarization states for photon packets. The angle α_t is the transmission angle into the new medium and is defined by Snell's law relating the refractive indices of the first layer and the ambient medium. All other photon properties are reset to zero.
For the laser scanning setup the detector is a camera observing the scene from a similar distance. It has a fixed spatial relationship to the light source.
After the photon launch, the trajectory through the eight layer model starts. A sketch of this simulation space is shown in fig. 3.7A. While the final results are only stored for a finite array on the surface or cross-section through the medium, photon travel during simulation is not limited to the sides. Photons leaving the tissue outside the array bounds *on the surface* are simply not stored.

Photon Step Size At each iteration the free path s_{ph}, i.e. the photon step size is randomly sampled. The probability of traveling s_{ph} or less is given by the definition of



A: Simulation model of multi-layerd soft tissue. The incident beam, the absorption array A_{ph} and diffuse reflection array R_D are labeled

B: MCML program flow chart for the life circle of one photon. A new iteration always starts at sampling a new step width. Blue boxes indicate new contributions to the simulation software

Figure 3.7: Tissue geometry and program flow chart for the MCML simulation software. (© 2013 OSA. Reprinted, with permission from [325]).

$\mu_t = \mu_a + \mu_s$, the total attenuation coefficient [317]. It describes the average free path length before an interaction occurs.

$$P(s < s_{ph}) = 1 - \exp(-\mu_t s_{ph}) \tag{3.6}$$

Setting the area under the probability density curve to a uniformly sampled random number yields the final equation for the step size [317].

$$s_{ph} = -\frac{\ln(\zeta)}{\mu_t} \tag{3.7}$$

Since the medium has several layers of finite thickness, the sampled path length might reach across more than one layer k. The actual path length depends on the attenuation coefficient of the medium. Therefore, the dimensionless step size $s_{ph} \cdot \mu_t$ is typically stored instead. The simulation then checks during photon travel, whether a photon would hit a boundary. This is achieved by comparing the current z coordinate with the stacked layers

depths. This may cause the total path length to be segmented into parts s_{ph}^k:

$$\sum_k s_{ph}^k \mu_t^k = -\ln(\zeta) \tag{3.8}$$

Interaction with Boundaries In case its path length would interfere with a layer transition, the photon travels until it reaches the boundary. Then the remaining path length is stored and two events may happen: reflection or transmittance. The probability for reflection is obtained from eq. 3.5. By sampling another random number ζ, the photon packet is reflected at the boundary if $\zeta \leq R(\alpha_i)$, otherwise it transmits. Here, α_i is the angle under which the photon hits the boundary and can be determined from the directional cosines.
In case of reflection the sign of the directional cosine in z direction is flipped. For transmittance, Snell's law provides the updates of all three directional cosines according to the refractive indices at the boundary [317]. A special case is given, if the photon packet leaves the surface of the medium. In this case, it is stored in terms of a reflected quantity as described later on. After a photon has traveled the sampled path length – possibly by hitting several boundaries – an interaction occurs: (1) energy absorption, and (2) scattering.

Photon Absorption At an interaction site a proportion Δw_{ph} of the photon weight w_{ph} is absorbed and stored at the local element of the absorption array A_{ph}. The extent of absorption is given by

$$\Delta w_{ph} = \frac{\mu_a}{\mu_a + \mu_s} w_{ph} \tag{3.9}$$

Photon Scattering After decreasing the photon weight, the scattering event is described in terms of altitude $\theta_{ph} \in [0, \pi]$ (also deflection angle) and azimuth $\psi_{ph} \in [0, 2\pi]$. The cosine of the altitude follows the Henyey-Greenstein equation in eq. 3.2. Assuming again that the area under this density function equals a uniformly random number yields after integration (see [317] for derivation):

$$\cos\theta_{ph} = \begin{cases} \frac{1}{2g}\left\{1 + g^2 - \left[\frac{1-g^2}{1-g+2g\zeta}\right]^2\right\} & \text{if } \; g \neq 0 \\ 2\zeta - 1 & \text{if } \; g = 0 \end{cases} \tag{3.10}$$

The azimuth is uniformly sampled in its interval with $\psi_{ph} = 2\pi\zeta$. Given azimuth and altitude the directional cosines are updated as follows.

$$\begin{aligned} ux'_{ph} &= \sin\theta_{ph} \cdot (1 - uz_{ph}^2)^{-0.5} \cdot (ux_{ph} uz_{ph} \cos\psi_{ph} - uy_{ph} \sin\psi_{ph}) + ux_{ph} \cos\theta_{ph} \\ uy'_{ph} &= \sin\theta_{ph} \cdot (1 - uz_{ph}^2)^{-0.5} \cdot (uy_{ph} uz_{ph} \cos\psi_{ph} + ux_{ph} \sin\psi_{ph}) + uy_{ph} \cos\theta_{ph} \\ uz'_{ph} &= -\sin\theta_{ph} \cos\psi_{ph} \cdot \sqrt{1 - uz_{ph}^2} + uz_{ph} \cos\theta_{ph} \end{aligned} \tag{3.11}$$

Termination and Data Storage A photon packet would travel along its random path until it either dies or leaves the surface of the medium. In the first case, the thread of the photon packet is terminated because its weight reached zero due to absorption within the tissue. However, due to eq. 3.9, it only slowly approaches, but never reaches zero. Therefore, the way of photon dying is given by Russian roulette. If a photon weight falls below a threshold of 10^{-4}, it is subject to random roulette. It has a 10 % chance of surviving, otherwise it terminates. Below this threshold, photon propagation would add very little new information, but would add unnecessary computational load. In order to *conserve energy* despite terminating photons of non-zero weight, the survival chance needs to be linked to a gain in energy [317]: If a photon should survive by chance, its weight is multiplied by 10 and it continues its trajectory until it once again enters the Russian roulette. Thus, this gain can be understood as the pooled energy of photons eliminated due to the thresholding. The entire program flow is illustrated in fig. 3.7B.
Data storage is accomplished using arrays in the global GPU memory. Therefore, a grid of bins is distributed on the medium surface and along a cross-section of the medium as indicated in fig. 3.7A. The surface grid is located around the origin of the coordinate system and coincides with the $x - y$ plane. The grid on the cross-section is used to store absorption and coincides with the $x - z$ plane.
Each bin has a certain width and height and there is also a maximum number of bins. Each bin corresponds to one global memory address.
If a photon terminates at a location $[x_{ph}, 0_+, z_{ph}]$, the location is discretized into the corresponding bin. The absorption value in that memory cell of the cross-section array is then increased by the weight loss of the photon packet. The term 0_+ indicates the finite bin width around zero. Equally, if a photon leaves the surface at location $[x_{ph}, y_{ph}, 0_+]$, its remaining weight is stored at the corresponding memory address of the surface array. For the surface array the following information can be stored: the total diffuse reflectance R_D, the specular reflectance R_{sp}, or the diffuse reflectance of photons having traveled to layer k R_D^k. The latter records only photons where layer k was the deepest layer they had at least one interaction with. Furthermore, a third array can store photons

which have traveled for time t_{ph}, a total path s_{total}, or had a certain number of scattering events, respectively (first array dimension). The second dimension is provided by the distance $r = \sqrt{x^2 + y^2}$ from the coordinate origin where the photon left the surface.
Since a camera with certain aperture is recording the backscatter from the medium, not all backscattered photons will be sensed. Therefore, the backscatter distribution can be restricted to photons backscattered within a certain altitude and azimuth range only.

All simulations presented hereafter were carried out with a total number of $N_{ph} = 200 \cdot 10^6$ photons. The numerical noise variance of the results relates to the inverse square root of this number. The bin size was 20 µm × 20 µm with a grid of 750×750 bins on the surface and 750×1000 along the cross-section. This corresponds to a field of view of 15 mm×15 mm on the surface and absorption recording down to 20 mm depth. If not denoted otherwise, no restrictions with respect to the aperture were made in default case.

3.3 General Aspects of Light-Tissue Interactions

3.3.1 Determination of General Simulation Parameters

The simulation results of light-skin interactions are subject to some general parameters. These need to be set before effects of a changing skin model on the optical backscatter can be studied. The subsequent paragraphs will present and discuss the initial choices made.

Number of Simulated Photons The MCML simulations do not provide analytic, but iteratively approximated solutions to light-skin interactions. Thus, the number of iterations, which is equal to the number of photons, determines the accuracy of the results. These are subject to random processes which sample from probability distributions. For infinitely many photons, the solution is correct. For any finite number, some numerical noise will remain. The less photons are chosen, the less random effects can be averaged out – the higher is the noise. The computational time increases approximately linear with the number of photons.
Figure 3.8A illustrates different noise levels for 10^6 (blue), $2 \cdot 10^8$ (red), and 10^9 (green) number of photons. In this experiment, the 2D position on a surface patch with 750×750 bins was uniformly sampled at random. Each bin will contain the same amount of photons, if the total number of photons N_{ph} approaches infinity. The variation from bin

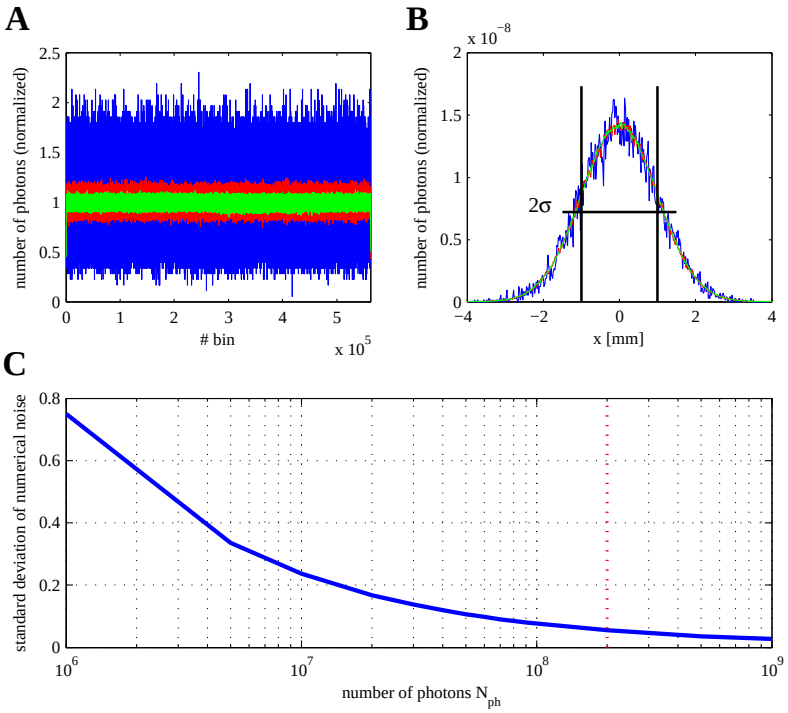


Figure 3.8: Analysis of the numerical stability for different numbers of photons. **A:** Approximated uniform distribution on a patch with 750×750 bins with 10^6 (blue), $2 \cdot 10^8$ (red), and 10^9 (green) number of photons. The number of photons per bin number is normalized by the patch mean and plotted sequentially. **B:** The same three numbers of photons as used in **A** were sampled from a zero-mean 2D normal distribution. The plot visualizes the smoothness of the cross-section along one coordinate axis. The data were normalized by the total number of photons. **C:** The standard deviation (STD) along the approximated uniform distributions was computed for various numbers of photons. A perfectly approximated uniform distribution would have a STD of zero.

to bin will vanish. The figure shows the varying number of photons across bins. For simplicity, the bins are plotted sequentially and not in a patch. To compare different cases, the number of photons per bin has been normalized by the mean number over all bins. It can be seen, that the noise level decreases rapidly for an increasing number of photons. Figure 3.8C plots the STD across all bins for various $N_{ph} \in [10^6, 10^9]$. The red dotted line marks the red case ($N_{ph} = 2 \cdot 10^8$) from fig. 3.8A. Beyond that number the decrease in the noise level does not justify the increase in computational time. For this N_{ph}, which has been chosen for all further experiments, the computational time on an Intel® Core™ i5-2500K CPU @ 3.3GHz, 16GB RAM with a GeForce GTX 470 graphics card took about $1 - 2$ min.

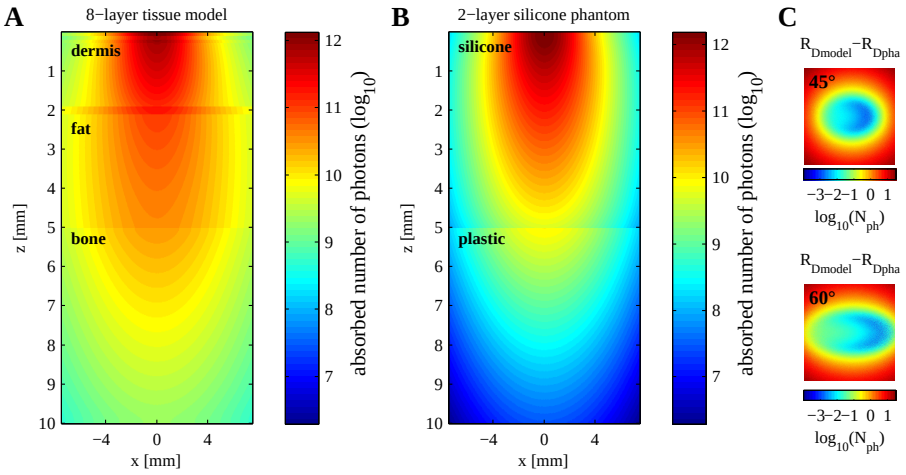


Figure 3.9: Comparison of the absorption profile for the multi-layered tissue model (d_s = 5 mm, **A**) and the two-layer silicone phantom (**B**) at 850 nm[2]. The laser orthogonally hits the target at the surface center (top part of the plot) and penetrates the medium moving downwards. The color indicates the logarithmic number of photons absorbed at a certain coordinate. **C:** Logarithmic difference between diffuse surface reflections of the 8-layer model and the phantom (at $\alpha = 45°$ and $\alpha = 60°$).

Finally, for fig. 3.8B positions in the 2D patch (the $x - y$ plane) were sampled according to a 2D Gaussian distribution ($\sigma_{beam} = 1$ mm). In this way the beam profile of the incoming laser beam is approximated later on. The plotted cross-section of the spot (at $y = 0$) was normalized by N_{ph}. The colors again correspond to the cases in fig. 3.8A. The red case giving a Gaussian of minor noise corruption corresponds to the chosen $N_{ph} = 2 \cdot 10^8$. Note that this only corresponds to one line in the 2D grid. Summation in subregions will further reduce the noise. The grid has 750×750 bins with each bin representing a 20 µm × 20 µm square. This means the complete patch, i.e. the simulated camera image, covers an area of 15 mm × 15 mm.

The previous investigation demonstrated that the input profile is well approximated with $N_{ph} = 2 \cdot 10^8$ photons. To evaluate the numerical approximation of the backscatter image, i.e. the diffuse reflection R_D, a typical 8-layer skin model (d_s= 5 mm, λ=850 nm, cf. fig. 3.9A) has been simulated. Out of all photons, 51.9 % were absorbed in the tissue, 41.7 % diffusely and 4.4 % specularly reflected. The rest left the model at the boundaries of the acquisition grid. Figure 3.8C confirms that even if only half of the $N_{ph} = 2 \cdot 10^8$ photons are reflected, the approximation accuracy is still close to that of the input profile.

[2]This wavelength in the NIR range has been found of high relevance during later experiments. For details see subsequent sections.

Table 3.2: Parameters of silicone-plastic phantom at 850 nm.

layer	n	μ_a [mm^{-1}]	μ_s [mm^{-1}]	g	d [mm]
silicone	1.4	0.12	21.8	0.88	5
plastic	1.56	0.2	20	0.9	5

Multi-Layer Model versus Silicone Phantom The justification for the rather complex 8-layer tissue model is indicated in fig. 3.9. The figure compares the absorbed light energy along the cross-section ($z - x$ plane at $y = 0$) of the tissue between the 8-layer model of skin (fig. 3.9A) and a silicone rubber phantom (fig. 3.9B). The latter mimics human skin with the optical properties given in table 3.2 [178]. The difference is that the layer structure is not as detailed as in the more sophisticated 8-layer model, but rather averaged. The comparison gives an impression of how the backscattered reflection at the surface is influenced even by thin, but highly scattering and absorbing layers. In the skin model particularly the upper layers containing water and the chromophore melanin among other substances, absorb and scatter the medium directly after penetrating the surface. The cone-shaped beam is widened and the distribution of scattered photons is more widely spread – particularly with increasing depth. For the phantom, in contrast, there is a higher proportion of light absorbed at short path lengths and scattering leaves the penetrating beam more narrowly shaped. In comparison with the values in the last section, 59.9 % of the light were absorbed by the phantom, 2.8 % specularly and 37.3 % diffusely reflected. The specular reflection seems to fall below typical values for human skin as they were mentioned earlier. For the diffuse reflection pattern (fig. 3.9C) at different incident angles the surface of the silicone phantom therefore exhibits less photons at the site of incidence and more photons at the exit site of the beam. Thus, it can be expected that for skin the information from deeper layers is rather weak, blurred and superposed by prominent reflections from the upper layers. The silicone rubber model is therefore insufficient for simulating small changes in the surface reflectance pattern.

Wavelength Dependencies The diffuse and specular reflection was simulated for the number of photons chosen above. This reflection can be observed at the surface of the medium as a simulated camera image. The simulation was repeated 13 times with changing optical properties for wavelengths $\lambda \in [400, 980]$ nm. The photons reaching the surface were labeled by the deepest layer they traveled to. This made it possible to decompose the full reflection at the surface into reflection components originating from a particu-

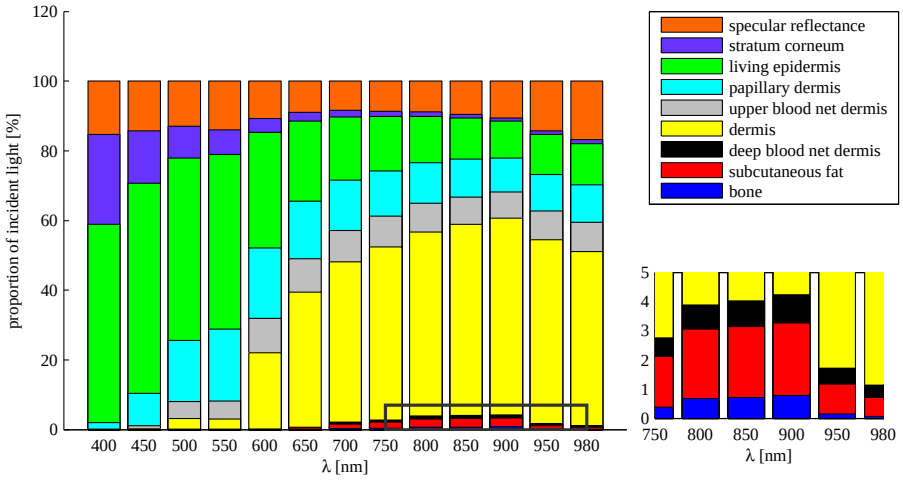


Figure 3.10: Proportion of light returning from the eight model layers for different wavelengths ($d_s = 5$ mm). Photons were counted for a layer if the layer represents the deepest one reached by that photon. A zoom into the region between 750 nm and 980 nm is shown on the right. (© 2013 OSA. Reprinted, with permission from [325]).

lar layer. Figure 3.10 gives a stacked bar diagram which shows the proportions of light (accumulated across the entire 2D area) coming from the different skin layers at certain wavelengths.
In agreement with the theory, most light is reflected at the upper layers for smaller wavelengths being close to the UV. This effect is mainly caused by the optical properties of melanin. This effect quickly weakens above 650 nm, where the structures and contents of the dermis dominate the reflection. In the NIR range, but particularly between 750 nm and 950 nm, light exhibits the deepest penetration depth. This is part of the therapeutic window. As shown in the magnification of that region in fig. 3.10, about 3 % of the reflected light arrive from the subcutaneous fat or bone. As discussed before, these layers provide most information about the tissue thickness, since they traveled along the whole thickness. For the simulation only these 3 % give the information about changes in the thickness, since the fat was the only layer varied in this conservative scenario.
Due to the results for the penetration depth, all further simulations used a wavelength of 850 nm.

3.3.2 Analysis of the Reflected Laser Spot

After setting the wavelength to 850 nm and knowing how much light is reflected from fat and bone, it is important to know *how* the backscattered light is distributed across the

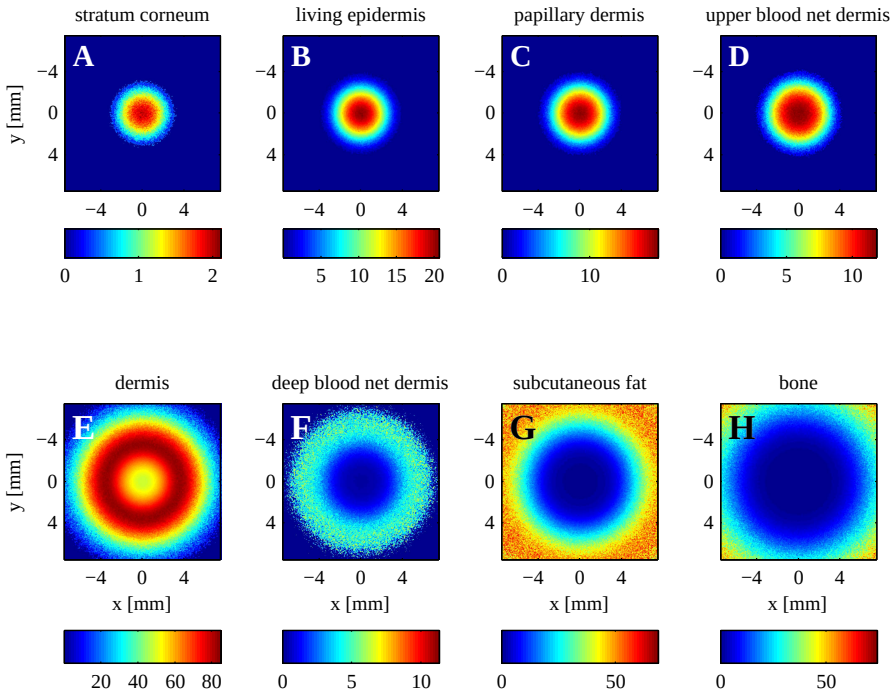


Figure 3.11: Relative light intensities across a simulated camera image. Plots **A** to **H** show the spatial distributions of the light backscattered from each of the eight model layers. Each pixel is normalized by the pixel intensity it has got in the full reflection. The color bars are in percent of the full reflection. (© 2013 OSA. Reprinted, with permission from [325]).

surface. Therefore, the described layer-wise components of the full reflection were normalized by the full reflection. Figure 3.11 plots these normalized layer-wise components. It can be observed that the first four layers account for most of the intensity in the laser spot center. The photons do not travel deep enough to be capable of reaching the outer regions of the spot. In contrast, the depth of the dermis and even lower layers allows photons to travel further out to the spot margins. Since this is not likely for photons backscattered from the upper layers, only the dermis and the deep blood layer reveal ring-shaped relative distributions.

The same argument applies for the subcutaneous fat and the bone. They are responsible for most of the photons reflected from regions further away from the spot center – at the outer borders. This means, for a noise-free environment that the most dominant information about the tissue thickness can be found in the outer regions. Due to the high relative proportions, it is less likely that other effects from other layers superpose this information. In a relative sense the information is purest in these regions

This is also confirmed in fig. 3.12A where the light at each radius r has been accumulated

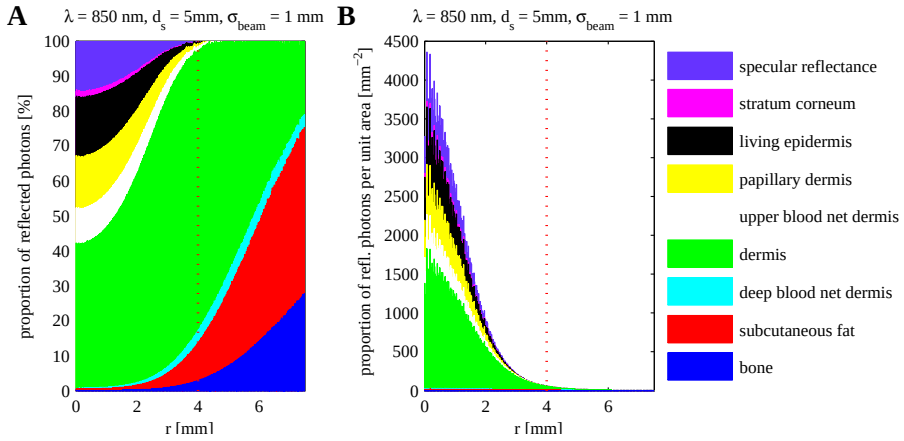


Figure 3.12: Relative (**A**) and absolute backscatter (**B**) from the eight model layers along the spot radius. On the left side, backscattered light reaching the surface from a certain layer is normalized by the full reflection along the radius. (© 2013 OSA. Reprinted, with permission from [325]).

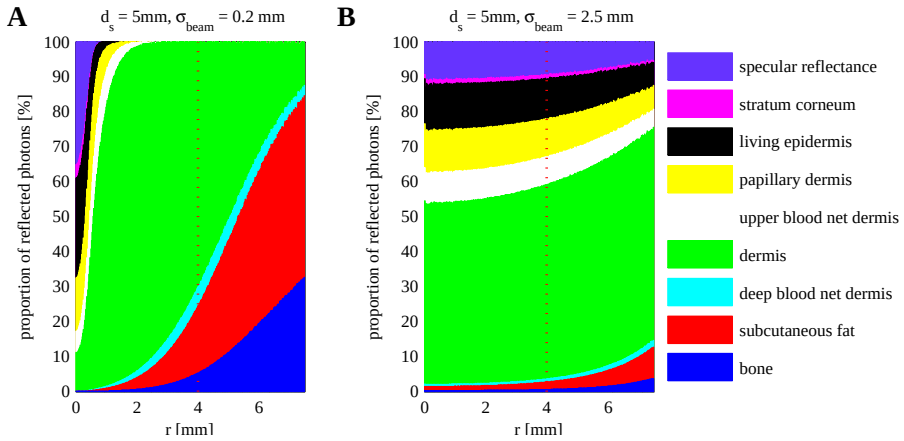


Figure 3.13: Relative proportions of backscattered light for two different spot sizes: $\sigma_{beam} = 0.2\,\text{mm}$ (**A**) and $\sigma_{beam} = 2.5\,\text{mm}$ (**B**).

along a circle concentric around the spot center. The relative proportion of light returning from the last two layers increases with the radius. Particular dominance is given beyond the dotted red line (4 mm from the spot center). The problem with this fact is illustrated in fig. 3.12B which plots the absolute intensity profile along the radius. Beyond the dotted line, there is only very little backscatter in an absolute sense. Although the last two layers are dominant here, the absolute signal is very weak. In a real scenario, noise is superposing this information and may corrupt the purity. The extent needs to be evaluated

practically. Overall, the intensities found at the medium range radius r may provide the best compromise between the dominance of the desired information within effects originating from other layers, and the signal-to-noise ratio (SNR). The radius at which this optimal regions ends will depend on the hardware specifications later on.

Changing Laser Spot Size Although of less importance for the simulation outcome, the size of the laser spot may play an essential role in practice. The spot size is defined by the STD of the isotropic Gaussian beam profile $\sigma_x = \sigma_y = \sigma_{beam}$ (cf. eq. 3.4). For $\sigma_{beam} = 0.2\,\text{mm}$ and $\sigma_{beam} = 2.5\,\text{mm}$, fig. 3.13 illustrates analogously to fig. 3.12A the relative proportions of light returned from each of the eight layers. A comparison of both cases shows that a smaller spot size is preferable over a wide spot diameter. Optical elements such as objective and size of the image sensor should be specified to cover as much new information about the behavior of the deeper layers as possible. Looking at the wide spot reveals that the image plane recorded very similar proportions of light across the area. Segmenting and accumulating pixel intensities in subregions would yield pretty much the same behavior for all subregions. Furthermore, a wider spot would cover a larger area on the skin and it would be difficult to assign the recorded light changes to a specific spot on the surface. The effects of a larger region are averaged.
For the smaller spot size, fig. 3.13A indicates substantially differing light proportions across the area of the image plane. Subregions may hence record different functional behavior making it possibly easier to model a highly predictive functional relationship between tissue thickness and backscatter. This is particularly true if smaller changes that occur in the thickness of the upper layers are correlated to the change in the total tissue thickness to some extent. A change in the dermis or epidermis thickness, for instance, may also indicate that other layers and the total thickness became thicker. For the simulation we omitted these effects. A smaller spot also enables a more precise localization and better lateral resolution of the point grid across the surface. Nonetheless, the SNR argument also applies here. The dotted red line in fig. 3.12B would be shifted to the left and possibly large areas of the sensor image plane might be scarified to the noise. Therefore, a spot size of $\sigma_{beam} = 1\,\text{mm}$ was chosen as a compromise for all further experiments.

Camera and Laser Source Orientation In order to optimally record the optical backscatter from the tissue, the camera is required to highly resolve each spot on the surface. Therefore, only a small patch (here $15\,\text{mm} \times 15\,\text{mm}$) will be visible to spatially distinguish parts of the reflection pattern. This necessitates that the optical axis of the camera will always coincide with the laser beam. This approach can be realized using beamsplitters

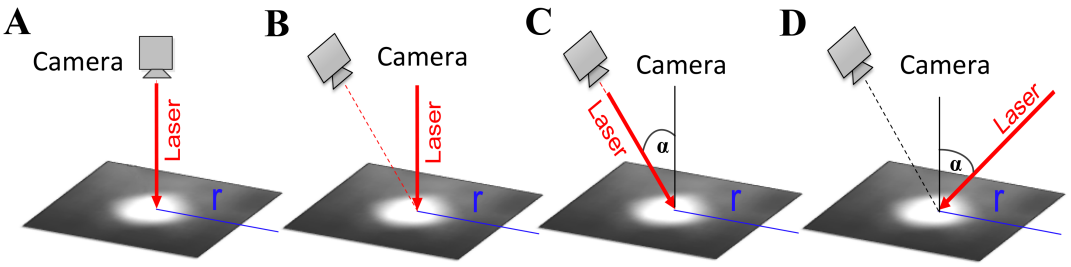


Figure 3.14: Schematic comparison of an in-beam approach (**A** & **C**) and an off-beam setup (**B** & **D**) for two different angles ($\alpha = 0°$ & $\alpha = 45°$).

and mirror-based scanning as discussed later on. Such a setup will be termed in-beam setup and is illustrated in fig. 3.14A & C. For so-called off-beam cases, i.e. where both axes do not coincide, it may happen that the surface spot leaves the FoV of the camera. Changes in the 3D surface geometry, i.e. a differing spot to laser distance, easily yields such a scenario. Off-beam examples are shown in fig. 3.14B & D.
Simulations for all four cases as presented in fig. 3.14 were conducted. In order to model the pose of the camera the azimuth was restricted to the interval $[170°, 190°]$ and the altitude to the interval $[40°, 50°]$[3]. Only photons leaving the surface in this interval were recorded. The angle α was set to either $0°$, $45°$, or $-45°$. The results are plotted in fig. 3.15, where the coloring of the layer proportions is identical to all previous plots.
It can be observed that for the same angles the in-beam and the off-beam approach lead to a very similar outcome. Differences between non-orthogonal and orthogonal irradiation in fig. 3.15B are similar to the ones obtained for a changing spot size. For $\alpha = 0°$ the spot is smaller and not stretched to an ellipse as for $\alpha \neq 0$. For the latter, parts of the reflection pattern, which can be still recorded for the orthogonal case, would leave the image plane. Thus, proportions for the last two layers at the same radius are higher in the $\alpha = 0°$ case. For the orthogonal in-beam case (fig. 3.15A) proportions from the dermis down to the bone are slightly higher ($\sim 1 - 2\,\%$) than in the off-beam case, since the backscatter directions are not uniformly distributed across the unit sphere, but have a bias for the direction of incidence. Photons that penetrate more deeply were subject to more scattering events. This will smooth out this bias.

[3] Except for case A, where the azimuth was within $[0°, 360°]$ and the altitude within $[85°, 90°]$. As described earlier, both are the spherical coordinates corresponding to the Cartesian coordinate system shown in fig. 3.7A.

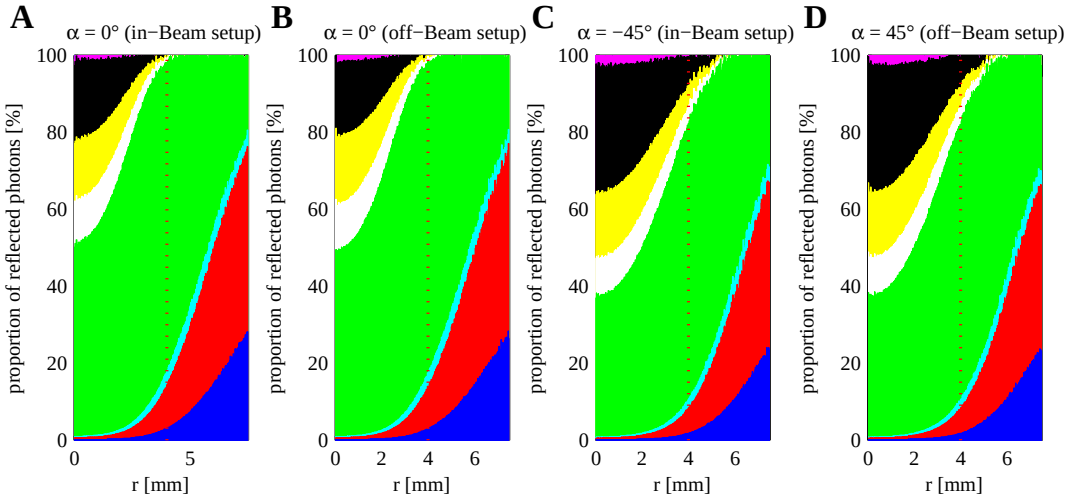


Figure 3.15: Relative proportions of backscattered light for different camera settings: **A:** in-beam setup, $\alpha = 0°$, **B:** off-beam setup, $\alpha = 0°$, **C:** in-beam setup, $\alpha = 45°$, **C:** off-beam setup, $\alpha = -45°$. For the in-beam setup, the laser beam always coincides with the optical axis of the camera, while this is not the case for the off-beam setup. Layer proportions are colored analogously to all previous plots.

3.3.3 Backscattered Light and Changes in Tissue Thickness

To evaluate the effects of changing tissue thicknesses, the thickness of the fat layer d_f was varied from $0\,\text{mm}$ to $5\,\text{mm}$ and therefore the total tissue thickness d_s from $2.1\,\text{mm}$ to $7.1\,\text{mm}$. The changes were investigated in steps of $50\,\mu\text{m}$. Changes in accumulated light intensity at radius r were investigated. Variations of the relative diffuse reflection ΔR_{Dnorm} were computed as follows:

$$\Delta R_{Dnorm}(r, d_s) = \frac{R_D(r, d_s) - R_D(r, 2.1\,\text{mm})}{R_D(r, 2.1\,\text{mm})} \tag{3.12}$$

Figure 3.16 shows $\Delta R_{Dnorm}(r, d_s)$ as a function of the radius for three different tissue thicknesses. A superposition of two effects can be observed. First, the thicker the tissue, the longer are the distances of photon travel and therefore the likelihood for a photon of being absorbed before leaving the medium. Therefore, $\Delta R_{Dnorm}(r, d_s)$ decreases for thicker skin and in particular for larger radii. At these distances the returned photons are more likely to travel longer distances anyway. Second, the thicker the tissue, the more capable are photons of reaching the outer borders of the sensor area after being reflected and scattered in the deeper layers. This scattering effect superposes the absorption effect and even leads to an increase of $\Delta R_{Dnorm}(r, d_s)$ for larger radii r. At very large r and moderately thick skin, the effect causes a positive $\Delta R_{Dnorm}(r, d_s)$ because it dominates

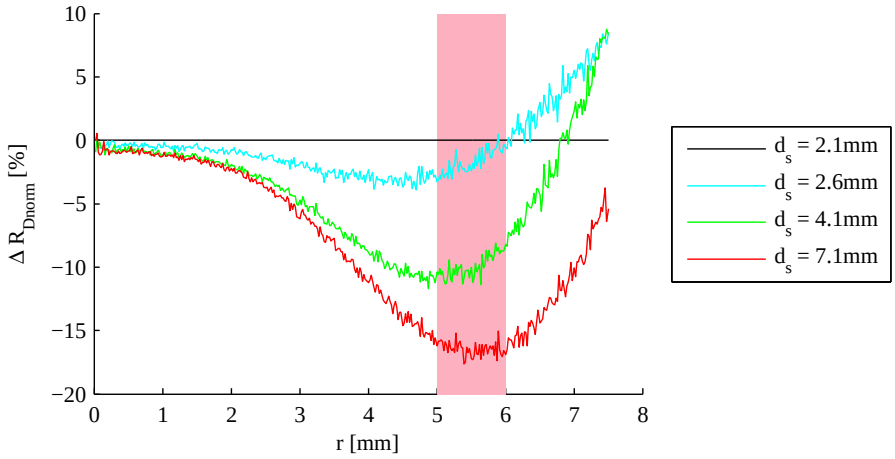


Figure 3.16: Relative changes in the light profile ΔR_{Dnorm} with changing tissue thickness d_s. The concentric accumulation of light intensities at radii r for different tissue thicknesses is normalized by the accumulation at radii r observed for $d_s = 2.1\,\text{mm}$. The red region denotes a bin which can be used for summarizing subregions.

the loss of photons due to absorption. For thinner skin, photons were hardly able to reach these areas and can only do so for thicker skin.
To exploit this information and to convert it into a tissue thickness measure using regression, the information needs to be encoded in a small set of numbers of reduced noise. A small set of features will tackle the so-called curse of dimensionality [309], which makes processing in high dimensional spaces challenging. This can be achieved by accumulating pixel intensities in radial subregions as denoted by the red box in fig. 3.16. Looking at all the backscattered light from the camera's perspective, these ROIs look like ring-shaped concentric circles around the spot center as illustrated in fig. 3.17A. An analytic expression for this set of numbers $\boldsymbol{b} \in \mathbb{R}^D$ – the so-called optical or NIR features – is given by:

$$b_i(d_s) = \sum_{x,y \in \text{ROI}_i} I_{x,y}(d_s) \quad \text{and} \quad \boldsymbol{b} = [b_1, b_2, \ldots, b_D] \tag{3.13}$$

where $I_{x,y}(d_s)$ is the thickness dependent pixel intensity at pixel location (x, y) on the sensor image plane. Each ROI is defined by a radius interval $\Delta r_i = r_{iend} - r_{istart}$ with $r = \sqrt{x^2 + y^2}$. For the simulation, the distance from the spot center to the outer boundary of the image plane was divided in 7 equally spaced intervals with $\Delta r = 2.14\,\text{mm}$.
Figure 3.17B plots the changes of these features across the entire thickness range tested. This is an illustration of the feature space for orthogonal irradiation. Effects similar to

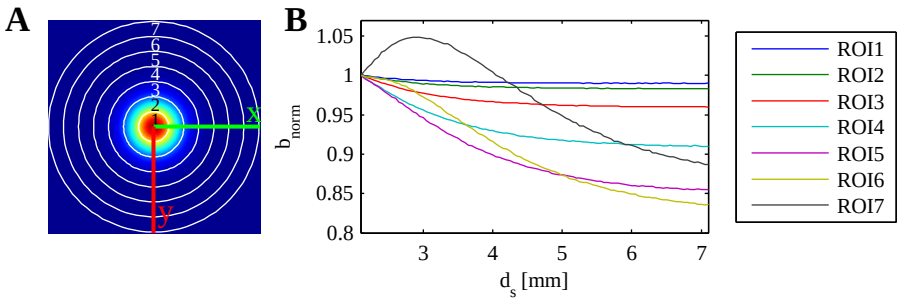


Figure 3.17: Feature space for orthogonal irradiation ($\alpha = 0$). **A:** Definition of seven concentric, ring-shaped bins around the spot center. The accumulation within each bin represents one feature. **B:** Changing feature values for varying tissue thickness: the feature space.

the ones discussed for fig. 3.16 can be observed: the features decrease with increasing d_s resulting in a negative correlation. The last feature (ROI 7) for the outer regions again shows the superposition behavior for the mentioned absorption and scattering effects. The features have again been normalized by the feature vector $\boldsymbol{b}$ at $d_s = 2.1\,\text{mm}$. The feature for ROI 7 gets positive for thicknesses from 2.1 mm to about 4 mm and then decays. It therefore exhibits a nonlinear relationship with d_s.

Gray Value Resolution A camera in an experimental setup needs to be specified such that the light intensity changes per pixel, which are generated by a certain change in tissue thickness, can be resolved by the image sensor. This requires a possibly high gray value resolution per pixel. To investigate this issue, the range between zero and the maximum peak of the expected beam profile was quantized with $N_q = 2^q$ quantization levels, where q denotes the number of bits for a gray value code. Since the absolute intensity change at a certain thickness variation gets smaller with growing radius, the analysis has been performed in dependency of the r.
Figure 3.18 shows an example analysis for $N_{14} = 2^{14} = 16384$ quantization levels for the whole range. Figure 3.18A and Figure 3.18B show how many quantization levels would correspond to a tissue thickness variation of $\Delta d_s = 50\,\mu\text{m}$. The intensity change triggered by Δd_s also depends on the initial thickness, i.e. the same thickness variation at $d_s = 5\,\text{mm}$ will have a weaker response than at $d_s = 2.1\,\text{mm}$. Therefore, graphs are plotted for the minimum, mean and maximum pixel intensity change across the entire range from $2.1\,\text{mm}$ to $7.1\,\text{mm}$. It can be seen, that even for larger r the thickness change of $\Delta d_s = 50\,\mu\text{m}$ can be resolved for some cases. In fact, for radii below $4\,\text{mm}$, most of the changes can be resolved. For comparison, fig. 3.18D illustrates that the entire thickness

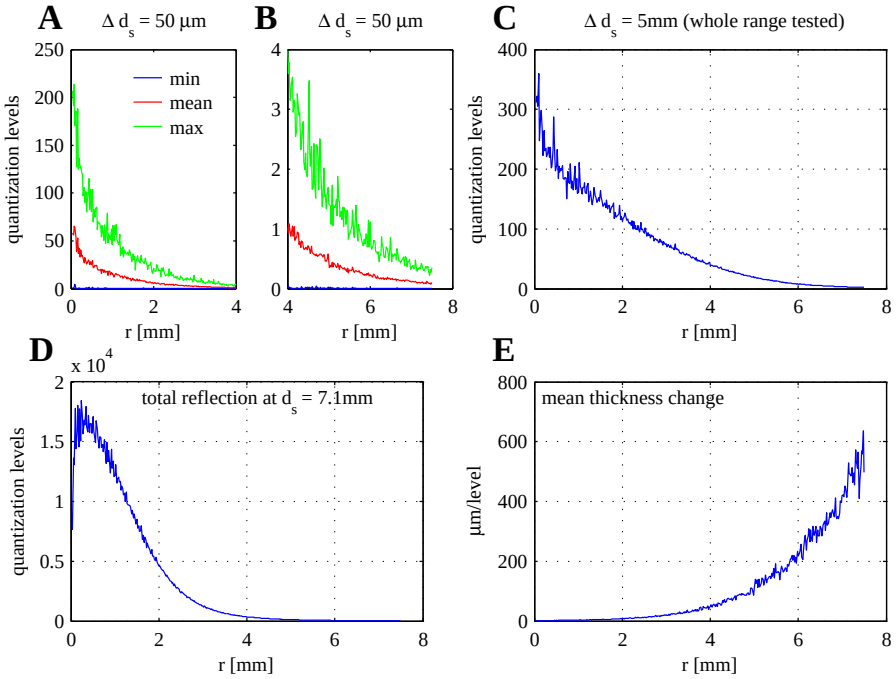


Figure 3.18: Analysis of resolvable light intensity changes at 14 Bit discretization. Quantization is considered for light at radius r along half the spot cross section. **A** & **B:** Quantization levels for minimal, mean and maximal light intensity changes for Δd_s = 50 μm starting from differently thick skin. **C:** Quantization levels for the whole tested thickness interval of length 5 mm (from 2.1 mm to 7.1 mm). **D:** Quantization levels for the full reflection (spot profile). **E:** Tissue thickness change on average required to exceed one quantization level at radii r.

interval of width 5 mm can be resolved at any radius with many or at least a few quantization levels.

This is also confirmed by fig. 3.18E, which shows how much variation Δd_s would be necessary per radius for triggering one quantization level of a 14 bit camera. For the conservative scenario chosen here, it is obvious that at larger distances from the spot center only more coarse changes in tissue thickness can be resolved. A similar observation can be made from fig. 3.18D.

So far, the maximum of the quantization range was set to the expected maximum of the beam profile. Now, driving the laser power beyond the saturation level of the camera sensor would place the quantization range in a better position to resolve smaller variations at the outer boundaries. Due to saturation, some information at the spot center would be lost. The likewise increased sensitivity to the noise level can be tackled by

averaging intensities across many pixels in subregions.

3.3.4 Incident Angle of the Laser Beam

In a real experimental setup it is unrealistic to assume orthogonal irradiation. The laser beam scans an uneven surface which necessarily leads to changing incident angles. As denoted in fig. 3.14, the incident angle α is defined as the angle between the incident laser beam and the surface normal. For orthogonal irradiation, we have $\alpha = 0°$. To investigate the influences of the angle on the backscattered light and finally on the feature space, α was varied between $0°$ and $45°$ in steps of $\Delta\alpha = 3°$. The left column in fig. 3.19 shows the variation of three measures in dependency of a changing incident angle: **A:** the shift of the spot centroid, **C:** the maximum likelihood estimate of the STD of the beam profile, and **E:** the scale of the maximum peak of the beam profile.
As expected, it was observed that the centroid shifts away from the image center with increasing angle. On the other hand, the STD gets larger, i.e. one spot axis stretches out and the spot becomes elliptical. The stretching effect is caused by a larger spread of the photon distribution for higher α. Therefore, less photons reach the same pixel at the surface and the number of photons at the maximum peak drops.
This means there are ambiguities with respect to the tissue thickness. On the one hand, there are more than one possible backscatter pattern for each tissue thickness. On the other hand, the effect may reverse, i.e. for one backscatter pattern several tissue thicknesses could be possible. This renders the problem more difficult for regression models.
There are two ways to possibly tackle the problem: (1) angle compensation from the raw data, or (2) increasing the amount of data to also cover reflections from differently thick tissue at several angles of incidence. The right column of fig. 3.19 presents a compensation approach. Initially, fig. 3.19B plots the beam profile for $\alpha = 0°$ and $\alpha = 45°$. All aforementioned effects are visible. Figure 3.19D corrects for the centroid shift and the increased STD. This $\sigma - \mu$ compensation uses a maximum likelihood approach to fit a Gaussian into the profile and to estimate μ_x and σ_x. Re-interpolation on a new grid using cubic splines leads to the compensated result shown in the figure. Finally, different intensity scales can be corrected as shown in fig. 3.19F. Such compensation approaches aim at fully eliminating angle effects, i.e. mapping the observation for arbitrary angles to its equivalent at orthogonal irradiation.
A possible problem of compensation approaches is illustrated in fig. 3.20. The measures presented in fig. 3.19 depend on the tissue layer. What the camera observes on the surface is not necessarily the case for the backscatter from the bone. Figure 3.20 shows

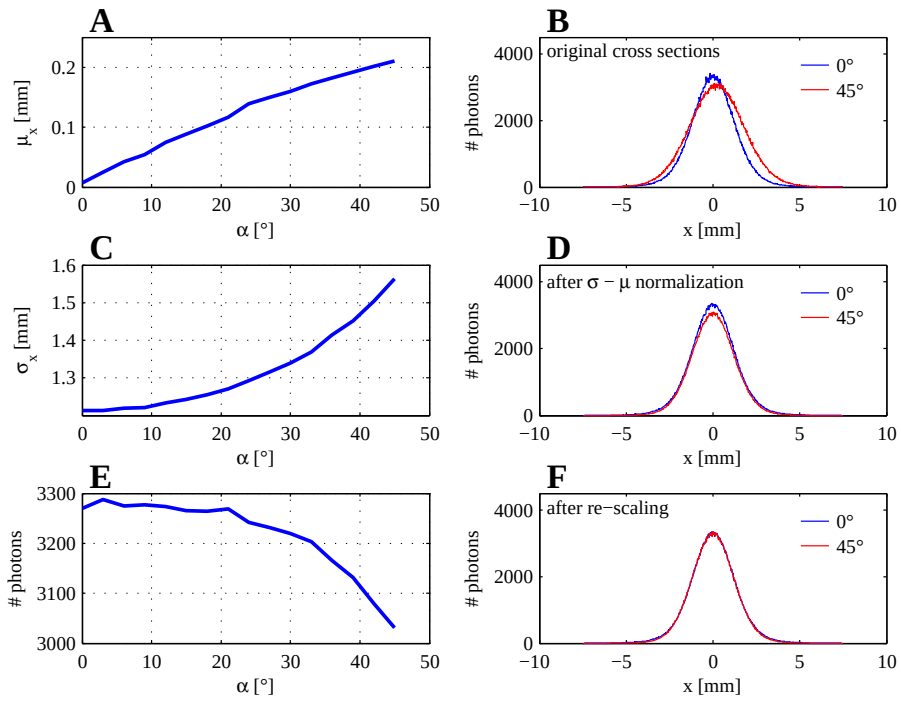


Figure 3.19: Spot properties changing with increasing incident angle. The plots show: **A:** changes in spot centroid, **C:** STD of the beam profile, and **E:** number of photons at the maximum peak. Plot **B** compares the beam profile at $\alpha = 45°$ and $\alpha = 0°$. Plots **D** and **F** illustrate compensation approaches by $\sigma - \mu$ normalization or re-scaling, respectively.

that the centroid shift is more distinct at the bone surface as compared to the skin surface. On the other hand, the ratio between the long and the short half axis of the spot ellipse grows faster on the surface. This indicates that the ellipse observed on the surface is mainly due to scattering in the upper layers, while the reflection from deeper layers tends to be more circular in shape. The compensation result may therefore be misleading.

The reflected beam profile is not precisely Gaussian due to the different optical properties of the eight layers. Therefore, the Gaussian fit is never perfect and information about the changing thickness of one layer will survive compensation. Furthermore, although the profiles in the right column of fig. 3.19 seem to be equal, fig. 3.21 shows that there are still remaining effects caused by the incident angle. The figure plots the cumulative photon energy obtained from the diffuse reflection. After compensation, the intensities are summed up for each radius r. Starting from $r = 0$, the intensity is then stepwise accumulated until $r = 7.5\,\text{mm}$ giving rise to fig. 3.21. The energy is normalized

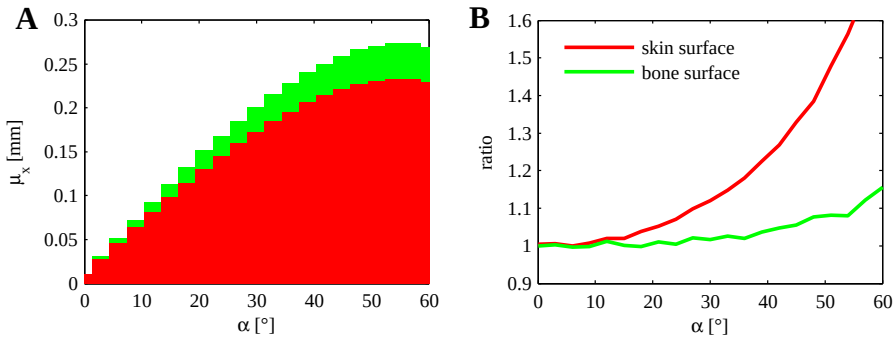


Figure 3.20: Spot properties for the total surface and the bone reflection at certain incident angles. **A:** Centroid shifts of the spot recorded at the bone and skin surface. **B:** Change of the ratio between small and big spot half axis recorded at the bone and skin surface.

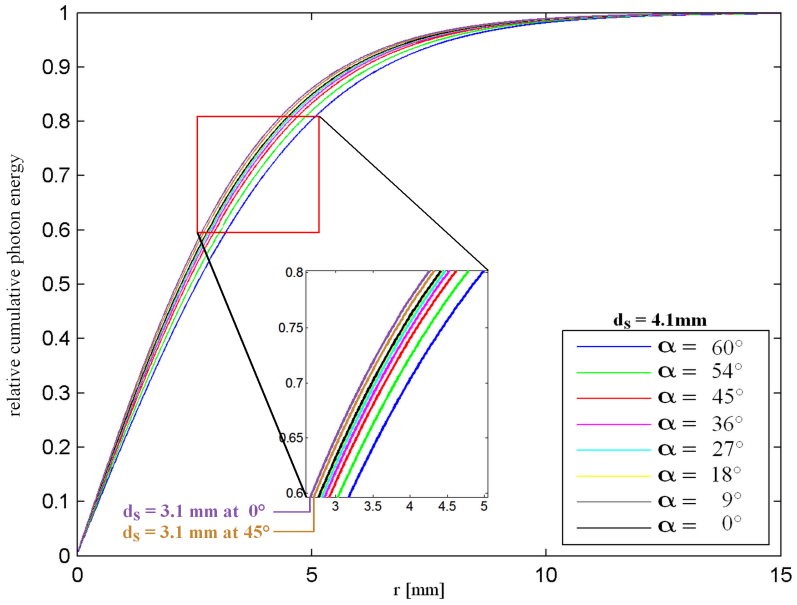


Figure 3.21: Starting from the spot center, the photon energy was accumulated with the growing radius r until all intensities (normalized to "1") are summed up. Images for two tissue thicknesses (d_s = 3.1 mm & d_s = 4.1 mm) and several angles were considered.

by the total energy sum across the entire patch. It follows that the angle as well as the tissue thickness lead to changes in the shape of the quasi-Gaussian backscatter profile. The plot also indicates that this effect is larger for angle changes than for varying tissue thickness.

This is confirmed in fig. 3.22 which plots the full feature space including features

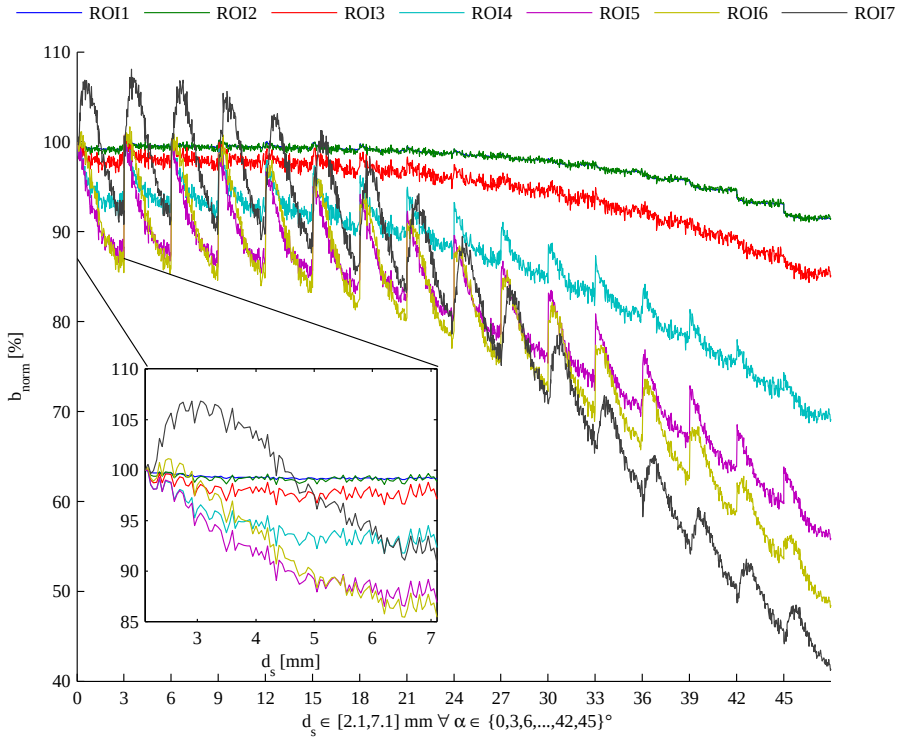


Figure 3.22: Feature space including varying incident angles after applying $\sigma - \mu$ normalization. Features are normalized by the values recorded at $\alpha = 0$ and $d_s = 2.1\,\text{mm}$. The horizontal axis is labeled with the incident angles, but between two angle labels at this axis, features are repeatedly plotted for tissue thicknesses from 2.1 mm to 7.1 mm.

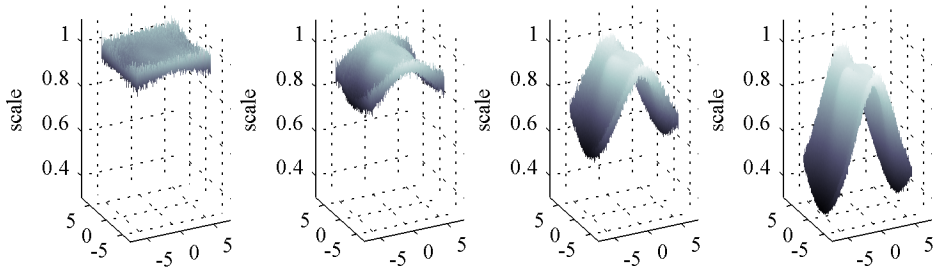


Figure 3.23: 2D intensity look-up tables for $\alpha = \{12°, 24°, 36°, 45°\}$ (from left to right). Each look-up table is the pixel-wise ratio between images at the aforementioned angles and the image for $\alpha = 0$. The ratios were averaged across several thicknesses (from 2.1 mm to 7.1 mm).

originating from varying angles. To generate the plot, the tissue thickness was varied for each of the tested angles from $d_s = 2.1\,\text{mm}$ to $d_s = 7.1\,\text{mm}$ in steps of $\Delta d_s = 50\,\mu\text{m}$. Strictly speaking, fig. 3.22 is a concatenation of feature spaces like the one in fig. 3.17B for different angles. The individual feature space sections look similar, but globally viewed, a strong angle influence is still visible despite compensation. Features are again normalized by the features $\boldsymbol{b}$ at $d_s = 2.1\,\text{mm}$ and $\alpha = 0°$. Up to angles around $10°$ the behavior is rather stable, before it increasingly reveals impact of the angle on the features.

By pixel-wise relating the diffuse reflection of arbitrary angles to the one obtained for orthogonal irradiation one can visualize the remaining effects in a 2D look-up table (LUT). Averaged across the tested tissue thickness range, the LUTs for four angles are shown in fig. 3.23. Generally, all values are below one, since no re-scaling has been applied. The ratio in the LUT drops in the spot axis where the angle affected the backscatter pattern. This means, that the compensated backscatter image had lower pixel intensities than the one at orthogonal irradiation. This is in line with observations in previous plots. It indicates that the estimation of the STD σ'_x tends to overestimate the spread of the backscatter profile and squeezes the stretched spot too much during normalization.

Compensation factors and techniques beyond the described scheme do not seem very reasonable in the context of other disturbance factors probably faced in a real-world scenario. The underlying interactions are simply too complex. Surely, two of these are the uneven nature of the surface or lateral skin heterogeneity.

3.3.5 Impulse Response and Time Shift

Another perspective on light-tissue interactions is given by temporal considerations. These may provide a completely different approach for retrieving information about the tissue thickness. A brief insight is given in the following.

All simulations so far did not take the time into account. Although photons are simulated sequentially by the hardware, the results add up to an observation integrated over time. Therefore, the investigated reflections can be interpreted as the integrated tissue response on an infinitely narrow light pulse, where all photons hit the surface at the same time.

Due to probabilistic scattering and reflections, each photon travels another distance. Together with the refractive index and the speed of light it is possible to compute the time a photon needs for its traveling. Figure 3.24A plots this time for all photons which have reached and left the surface at a certain radius r from the spot center. Most photons

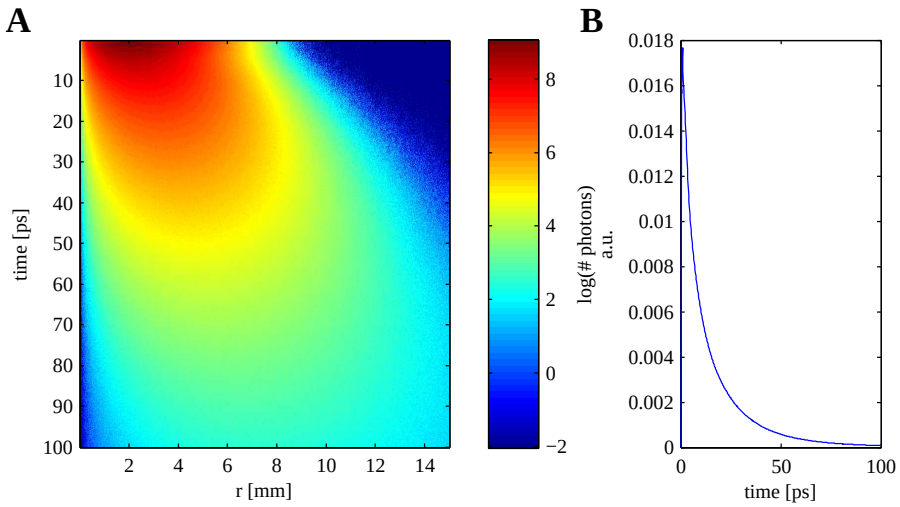


Figure 3.24: Travel time for backscattered photons. **A:** Travel time per photon is plotted against the radius r where they leave the tissue surface. The color codes the logarithmic number of photon for a certain time-versus-radius bin. **B:** Impulse response of the multi-layered skin model for a Dirac excitation.

left the surface close to the spot center having traveled only a short time in the range of a few ps. Nevertheless, there are photons which traveled longer, particularly at larger r.
The assumption that a sensor like a photodiode captures all the photons without spatial information gives rise to fig. 3.24B. This normalized plot shows how many photons are delayed to which extent. In signal processing this is called the (Dirac) impulse response of the system. The response is equivalent to a low-pass filter which smears the input signal, i.e. attenuates and delays its harmonics.
However, the lowpass filter only affects high frequencies as the time scale of the impulse response indicates. An example signal is given in fig. 3.25A. The red input signal $u[k]$ was selected to be the first five harmonics of a $315\,\mathrm{MHz}$ rectangular signal, i.e. a pulsed laser beam. The sampling rate is $50\,\mathrm{GHz}$. The system response $y[k]$ for this input signal can be computed by convolving the input with the impulse response $h[k]$.

$$y[k] = \sum_{i=1}^{\infty} h[i] \cdot u[k-i] \tag{3.14}$$

The system response is the green triangularly shaped signal in fig. 3.25A. The magnitudes of the harmonics of the rectangular signal follow a $sinc$ function. This function has a slower decay compared to the magnitudes of the harmonics from the triangular function. Therefore, the harmonics of the system response are expected to be smaller than those of the input signal. Figure 3.25B shows this attenuation for the first harmonic. Changes in

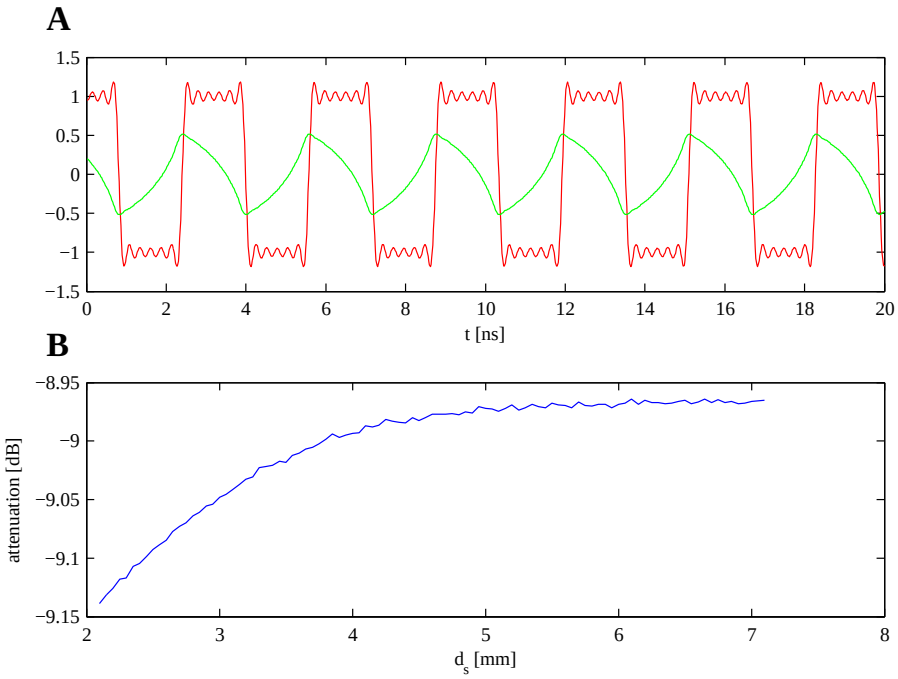


Figure 3.25: System response of the skin for a pulsed laser. The incident light beam is modeled by the rectangular function plotted at the top with its first five harmonics. The laser is pulsed with 315 MHz and the signal sampled with 50 GHz. The same plot also shows the system response. The bottom plot illustrates the ratio (in dB) changes for varying tissue thickness between the magnitudes for the first harmonic of the system response and the exciting rectangular signal.

the tissue thickness cause also changes in the impulse response. Thus, a dependency of the attenuation coefficient on the tissue thickness d_s is observed. This finding could also be used to estimate the thickness at a given spot. Challenges of this approach however include: (1) working with very high frequencies in the MHz and GHz range, (2) reproducibly generating the pulse sequence and (3) precisely measuring the system response.

3.4 Estimating Tissue Thickness using Statistical Learning

All findings presented in sec. 3.3 can be used to prepare the raw data for the statistical learning. They give rise to the processing chain shown in fig. 3.26. After acquiring the raw images of the optical backscatter, pre-processing will be applied to compensate for angle effects. The compensation module may be skipped if alternative approaches for handling the incident angle are chosen. As discussed before, the alternative to pre-

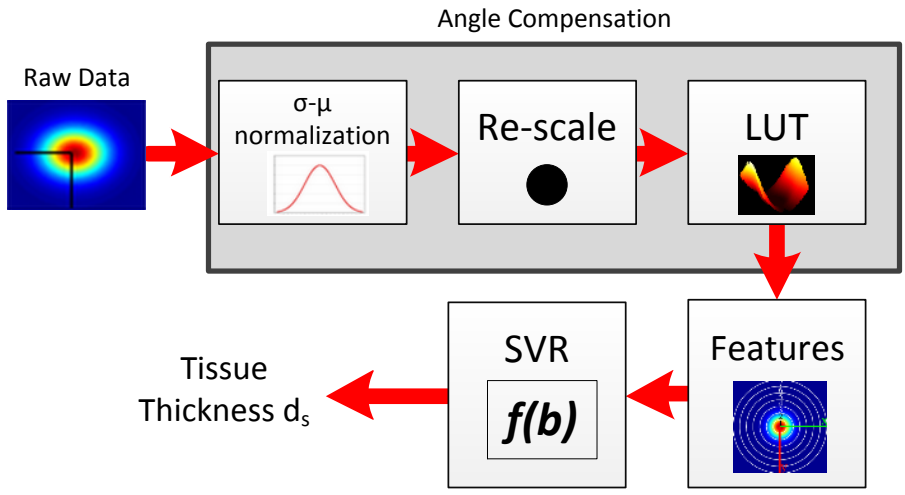


Figure 3.26: Processing chain for a camera image of backscatterd light. After acquiring the raw images, different modules for compensating angle effects can be applied. Subsequently, optical NIR features are extracted from ring-shaped ROIs, before statistical learning namely SVR converts them into a tissue thickness measure.

processing-based compensation is given by increasing the amount of training data, i.e. to incorporate the angle effects into the feature space and learn from more examples. This may be accompanied by adding the incident angle for an individual spot as an additional feature space dimension.
In case of compensation, the following sub-modules will be evaluated: (1) $\sigma - \mu$ normalization, (2) re-scaling, and (3) LUT-based corrections. The latter divides the final image pixel-wise by the reference pixels of a LUT, which was obtained as a pixel-wise ratio between images for arbitrary angles and $\alpha = 0$. The ratios were averaged across a range of different tissue thicknesses. The following regression analysis will evaluate different scenarios for handling the incident angle.
After pre-processing, the features are extracted according to eq. 3.13. In a further pre-processing step the mean of the features is subtracted and they are scaled to unit variance. In terms of unbiased testing, the necessary pre-processing parameters (mean and scale) are only computed from the training data and applied to the test data. The resulting data are used by SVR as input data. The following considerations will only use the RBF kernel and evaluate the MAE and RMSE after 5-times-10-fold CV. Parameters are obtained via grid search.

Table 3.3: SVR estimation results for different pre-processing and feature spaces.

Data	Angle Range	RMSE [mm]	MAE [mm]	STD [mm]
• no compensation	$\alpha = 0°$	**0.019**	**0.013**	**0.013**
• $\sigma - \mu$ normalization	$\alpha = 0°$	0.038	0.026	0.024
• $\sigma - \mu$ normalization (nested CV)	$\alpha = 0°$	0.043	0.030	0.029
• re-scaling	$\alpha = 0°$	0.060	0.039	0.037
• $\sigma - \mu$ normalization	$\alpha \in [0°, 45°]$	0.057	0.042	0.040
• no compensation + angle feature	$\alpha \in [0°, 45°]$	**0.016**	**0.010**	**0.012**
• $\sigma - \mu$ normalization + angle feature	$\alpha \in [0°, 45°]$	0.034	0.023	0.024
• LUT	$\alpha \in [0°, 45°]$	0.033	0.023	0.023
• LUT + angle feature	$\alpha \in [0°, 45°]$	0.027	0.018	0.020

Orthogonal Irradiation Table 3.3 lists the generalization errors made by SVR during CV. The upper part compares different angle compensation techniques on data recorded for orthogonal irradiation. Since there is no angle influence to compensate, this gives an impression how much information about the tissue thickness is destroyed by applying the pre-processing techniques. The baseline for all comparisons is given by features that were not subject to compensation. With an error of 19 µm the SVR yielded an accurate and promising prediction accuracy. Feature changes for thickness variations at thicker skin are generally smaller. Thus, slightly higher errors were observed when predicting these thicknesses rather than thin skin. After $\sigma - \mu$ normalization, the error worsens due to two effects. First, the Gaussian fit is not perfectly reproducible. Numerical variations in the simulations due to the finite number of photons and a profile which is not exactly Gaussian lead to minor uncertainties in the fit. This explains the noise showing up in fig. 3.22 after normalization. Second, information of the tissue thickness is also encoded into a changing variance of the backscattered light. This information is destroyed by normalization.

Further on, re-scaling the maximum peak leads to a further increase of the error[4]. Since there is more absorption for thicker skin, the maximum peak will decrease for growing d_s. Since the changing angle may have the same effect, one effect may mimic the other

[4]Each new compensation module builds up on all previous ones.

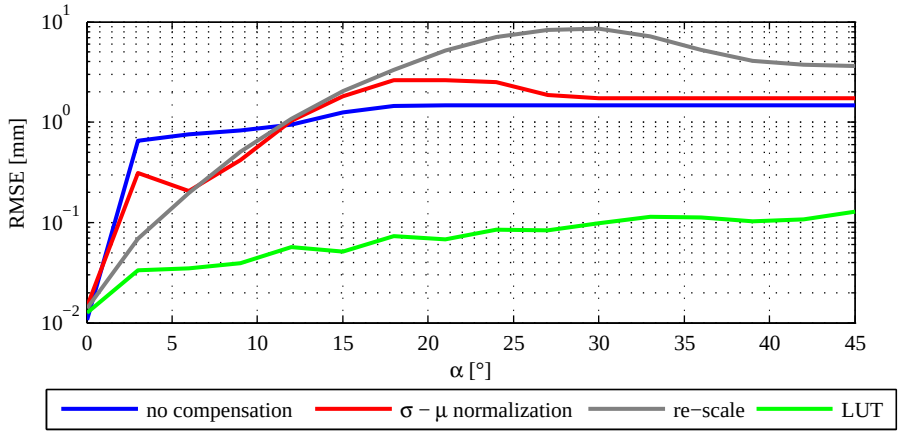


Figure 3.27: Drop of the SVR tissue thickness prediction error for growing incident angles. After training the SVR model with data originating from orthogonal irradiation, the learned model has been applied to data of growing incident angle. The analysis was carried out for different pre-processing options. Each legend entry for a compensation approach also includes the methods listed on its left as pre-steps. The estimation result at $\alpha = 0°$ corresponds to a biased test on the training data.

and ambiguities may arise. Eliminating the angle influence on the intensity of the maximum peak hence also destroys information about the thickness. SVR can now only retrieve information from a changing shape of the backscattered profile. This is caused by the distinct contributions different tissue layers will add too an overall change in tissue thickness. Nevertheless, it also prevents ambiguities.
For completeness, a result for nested CV at $\sigma - \mu$ normalization is added. Computing this error is very time consuming, but is capable of finding the parameters of SVR without the test data. They are obtained by an additional (nested) CV loop on the training data. The error value for nested testing shall show that the other results will slightly underestimate the real generalization error. The impact is however small and the discussions made remain still valid.

Generalization to Changing Incident Angles The compensation approaches aim at making the SVR model applicable to data from other angles. Figure 3.27 illustrates how this aim is achieved for the proposed compensation modules. Note that again each new module always includes all previous ones. This means the LUT approach gets a $\sigma - \mu$ normalized and re-scaled image as an input.
In Figure 3.27, the prediction error on data of higher angles was computed, after training a model with data from $\alpha = 0$ only. SVR makes predictions according to eq. 2.16, i.e.

it exploits a weighted sum of similarities between the test and the training data plus a constant offset. For angles up to $10°$, all compensation methods outperformed cases of no compensation. For higher angles, the kernel then outputs very low similarities such that the weighted sum does not predict the true thickness well. In fact, no compensation showed better generalization properties than other methods. However, for $\alpha > 18°$ the features become so dissimilar, that the kernel outputs zero and only the constant offset remains. For $\sigma - \mu$ normalization or re-scaling this happens only for higher angles. While these two methods keep the data in the length scale range γ of the kernel, they do have no overall positive effect on the prediction error. Only the LUT approach generalizes well to other angles (with an RMSE roughly below 0.1 mm). It nicely maps the feature space for higher angles into the space obtained for orthogonal irradiation. Nevertheless, this approach is not realistic and feasible in practice. It would require recording images for a fine grid of defined angles and average them across different skin thicknesses. This is tedious or just infeasible in terms of measurement time, hardware effort and accuracy.

Learning from Data with Arbitrary Incident Angles A more practical solution is given in the lower part of table 3.3 [325]. Data for angles $\alpha \in [0°, 45°]$ have been simulated for different thicknesses and all the data were passed to the SVR. This case is practically simple, since just any data needs to be acquired and then passed to the regression step for training. A drawback is that more data will be required, because the interaction between features, thickness and incident angle is more complex. Again, it was found that the LUT approach works best among all compensation methods. Apart from that, two observations were made. First, data not subject to angle compensation ("no compensation") yielded better results. This is because the effect of destroying information by pre-processing is more harmful than the increase of training data for arbitrary angles is positive. As a result, refraining from compensation techniques seems more promising. Second, adding the incident angle as an eight feature, i.e. generating a new vector $\boldsymbol{b'} \in \mathbb{R}^8$, leads to improvements. It adds a new dimension to the feature space, which makes it easier for SVR to tell ambiguities apart (e.g. where different thicknesses look similar due to differing incident angles). Finally, it has to be mentioned that the mean difference between "no compensation" ($\alpha = 0$) and "no compensation + angle feature" ($\alpha \in [0°, 45°]$) is hard to evaluate. The corresponding STD values show that the error distributions substantially overlap. However, a reason for a possible RMSE difference is that the second approach had only one more dimension, but 16 times as much data to sample the feature space. For the best result (no compensation with additional angle feature) the optimal SVR parameters were obtained as $\varepsilon = 0.0022$, $\gamma = 0.5$ and $C = 600$.

3.5 Conclusions and Hardware Specification

This section elaborated on three parts. First, background about the anatomy and physiology of skin, and the simulation of light-tissue interactions was discussed. Further steps were justified. Second, general considerations were made about the nature of the expected backscatter patterns from human skin. These gave rise to recommendations for possible hardware specifications of an experimental setup. Third, a theoretical proof of concept was given for predicting tissue thickness from optical NIR feature by means of statistical learning.
Therefore, the section answered RQ 1. A concise conclusion on the four main challenges raised by RQ 1 is given in the following.

- RQ 1.1: **What are the most suitable hardware parameters for an optical setup?**

  The results demonstrated that due to its penetration depth, NIR light, e.g. wavelengths around 850 nm, are most suitable for getting information about the tissue thickness. To sufficiently resolve the brightness changes triggered by small changes of the tissue thickness, an high dynamic range (HDR) with at least 14 bit gray value resolution and high sensitivity in the NIR range is recommended. Lenses and optical elements should be chosen such that an at least 15 mm $\times$ 15 mm large patch of the spot is visible to also extract backscatter further away from the spot center. An image of this spot needs to be of sufficiently high resolution, since accumulating many pixels will average out random pixel noise particularly for low intensities of the informative backscatter signal. In fact, this is a tradeoff which needs to be evaluated in practice for a concrete camera device. A Gaussian beam profile with 1 mm STD was found to be a good compromise. Nevertheless, a finer spot size with a higher laser power might also be a promising option. Adjusting laser power and aperture in a way that drives the spot center into the saturation limit of the camera, could provide a better dynamic range coverage for light changes at the spot margins.
  No objections with respect to an in-beam setup were found.

- RQ 1.2: **How is information coded in backscattered light and how can it be optimally translated into informative features?**

  For a Gaussian profile the relative proportion of light reaching the surface from deeper tissue layers was found to grow with the radius r from the spot center.

This includes the bone surface and the subcutaneous fat layer. Light from both layers will provide valuable information about the tissue thickness. Extracting features as accumulated luminance from ring-shaped ROIs, concentric around the spot center would be a natural choice. Other options may be possible as well. Many effects – also from the upper tissue layers – are superposing each other at the spot center. The acquisition noise however will have a considerable effect on the weak information signal recorded further away from the spot center. The latter implies that the SNR decreases with the distance from the spot center. Thus, light from regions of medium distance from the spot ($\sim 2-4\,\mathrm{mm}$) is the most promising compromise between the two effects. However, the upper bound of this interval will depend on the hardware. A low pixel noise of the camera and an accumulation of many pixel intensities from a high resolution image of the area could make information further away from the spot center well extractable. Noise reduction may also be achieved by trading the number of ROIs for their size, i.e. the number of pixels used for averaging the noise out.

- RQ 1.3: **What are possible disturbance quantities?**

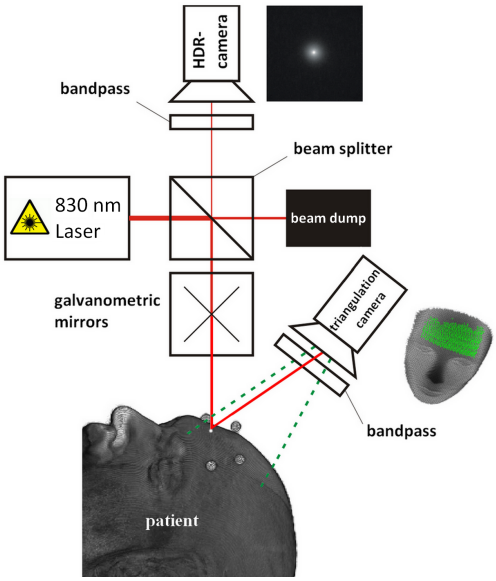
  The incident angle of the laser beam was found to be one highly relevant disturbance. It may generate ambiguities and mimic misleading changes of the tissue thickness. While compensation approaches were found to handle angles up to 10° well, they are not the recommended choice. The preferable choice is given by measuring the actual incident angle and adding it as an additional dimension to the feature space.
  Nevertheless, many questions about disturbances remain and need to be evaluated with real-world data from an experimental setup. This includes the impact of issues that were not modeled by the current MCML approach: non-planar boundaries between tissue layers, photon-photon interactions, muscle tissue, changing perfusion, oxygenation or pulse, or lateral tissue inhomogeneities such as freckles, moles, sweat, hair etc. Finally, the influence of skin type, gender and age have to be evaluated, since the simulation is only valid for average Caucasian skin.

- RQ 1.4: **How can informative features be used to retrieve a reproducible pattern to support surface tracking?**

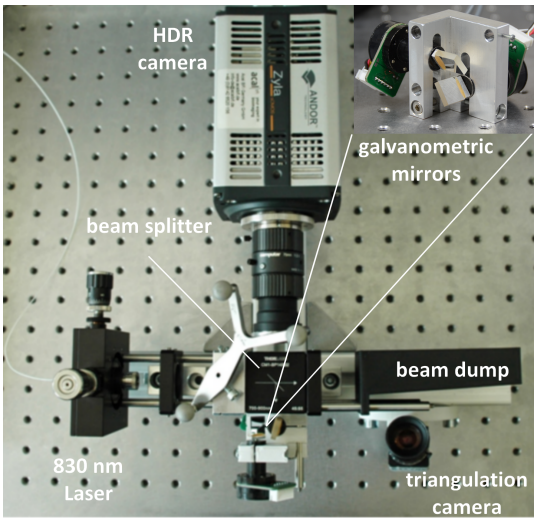
  Optical backscatter patterns may be influenced by several factors such as the tissue thickness, but also by disturbance quantities such as the incident angle of the laser beam. Therefore, these patterns cannot directly be used to support surface tracking as additional landmarks. Instead, it is proposed to make use of statistical learning. This will build models that relate the backscatter features to a tissue thickness measure and separate this information from other influences. Based on the simulations, a proof of concept was presented that SVR can predict the tissue thickness with very high accuracy. Nevertheless, to tackle external disturbances, large data sets may be required. This is even more relevant with respect to the aforementioned disturbances in a real scenario. These have not been covered by the simulation. Further on, supervised learning requires a ground truth for the tissue thickness. Several options for that will be presented, but the effect of having a limited accuracy for the ground truth needs to be evaluated with real experimental data.

4 Experimental Validation - Tissue Thickness Estimation on Real-World Data

The previous chapter has used Monte-Carlo simulations to demonstrate that tissue thickness can theoretically be predicted from optical backscatter. The key finding was that light with wavelengths around 850 nm penetrates the skin deepest. The reflected spot can be analyzed by a camera to extract additional information. The analysis evaluates intensity changes in subregions of the backscatter images and aims at estimating the local tissue thickness from that information.
This subsequent chapter shall experimentally validate this approach on real-world data. The first section describes the experimental setup. It summarizes necessary work included in the PhD theses of Patrick Stüber who designed the optical hardware setup [281, 283, 285], and Benjamin Wagner who contributed a software approach for laser triangulation using a galvanometric deflection unit [312–314]. Section 4.2 then outlines general aspects of the volunteer study. Special focus is directed to the ground truth acquisition for the tissue thickness as well as the registration thereof to the optical surface scans. A general overview about all data characteristics and its fusion across several modalities and coordinate systems is given. Linking back to chapter 3, sec. 4.3 discusses general aspects of the light-skin interactions, before sec. 4.4 evaluates the results of the statistical learning used to estimate the tissue thickness. Parts of this work have been published in [327, 329, 330, 337]. The section will particularly address the incorporation of prior knowledge such as the incident angle of the laser beam [328] or the exploitation of local neighborhoods in the scanning grid [330, 331] to improve the estimation quality. Finally, sec. 4.5 will give first insights into sparse approximation techniques for GPs to enhance the learning and estimation efficiency for larger data sets [332]. Section 4.6 will draw final conclusions from the findings presented.



A: Simplified sketch of the optical hardware design.



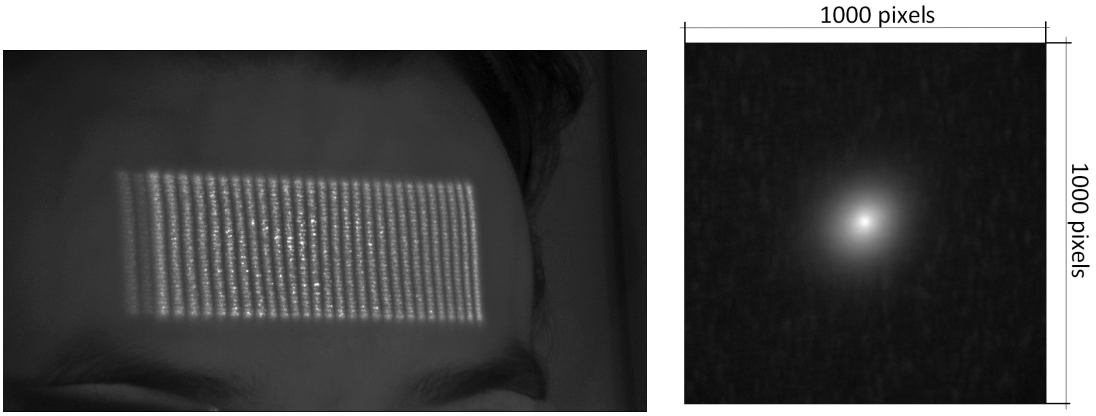
B: Image of the hardware implementation of the design.

Figure 4.1: Experimental setup for acquiring NIR surface scans. A galvanometric deflection unit deflects a laser beam onto the target. Reflections are then observed by a triangulation camera, and – coupled into the laser path – an HDR camera to evaluate the NIR optical backscatter (© 2015 IEEE. Reprinted, with permission from [335]).

4.1 Experimental Setup

4.1.1 Optical Hardware Design

The experimental setup for the laser scanning hardware consists of two parts. The general design is shown in fig. 4.1A. First, surface information is obtained by laser triangulation. Therefore, an 830 nm fiber-coupled NIR laser (LPS-830-FC with CFC-2X-B collimator, Thorlabs, Inc. [289]) was directed onto a cage cube-mounted pellicle beam-splitter (CM1-BP145B2, Thorlabs, Inc. [289]). The laser was driven with a bias current of 27.5 mA generating an output power of 1.27 mW. The beamsplitter directed 45 % of the light onto a galvanometric, mirror-based deflection unit (ASX-V20, Laserwinkel [165]). The remaining 55 % of light were absorbed by a customized beam dump. The motors of the galvanometric unit were equipped with two silver-coated square mirrors (ME05S-P01, 96.5 % reflectance at 830 nm, Thorlabs, Inc. [289]) as illustrated in the upper right corner of fig. 4.1B. Each of these mirrors either controlled the horizontal or vertical direction in the rectangular 32×32 scanning grid. Light with a power of about 0.5 mW finally reached the object. Requirements for laser safety Class 1 were fulfilled.



A: Image of a subjects forehead recorded by the triangulation camera. A grid of 1024 points was projected onto the surface for illustration purpose. True scanning was done interleaved and row-wise to avoid overlapping spots.

B: Laser spot image recorded by the HDR camera.

Figure 4.2: Images from the two main cameras of the system: the triangulation camera and the HDR camera. (Left image reprinted from [331], © (2015), with permission from Springer).

The triangulation camera (IDS UI-3340CP-NIR-GL, CMOS, 60 fps, 1280 × 1024 [135, 136]) used a Pentax C2514-M objective [239] with 25 mm focal length. The latter was chosen such that the FoV of the camera at 40 cm distance would cover a subject's forehead (cf. fig. 4.2A). The exposure time of the camera was chosen to cover half a row of the laser spot grid. Furthermore, the sequential scanning process was adjusted such that first all odd and then all even spot numbers in one row were projected (interleaved scanning). Thus, each camera image contained 16 spots. One row of the grid required two, and a full grid 64 images. This procedure avoids overlapping laser spots and ensures a precise spot localization in the image. Projected spots were typically 2-3 mm apart from each other when projected onto a typical forehead.
The second part of the hardware design is dedicated to measuring the NIR backscatter. An HDR camera (ANDOR Zyla 5.5 [8, 9]) was coupled into the beam path to record the light which is backscattered from the forehead – through both deflecting mirrors and the beamsplitter. Such a design is called in-beam setup as already discussed in chapter 3. This scientific CMOS (sCMOS) camera joins advantages from standard CMOS cameras (fast acquisition, no blooming) and CCD cameras (high resolution and dynamic range). The list below summarizes the most important key facts.

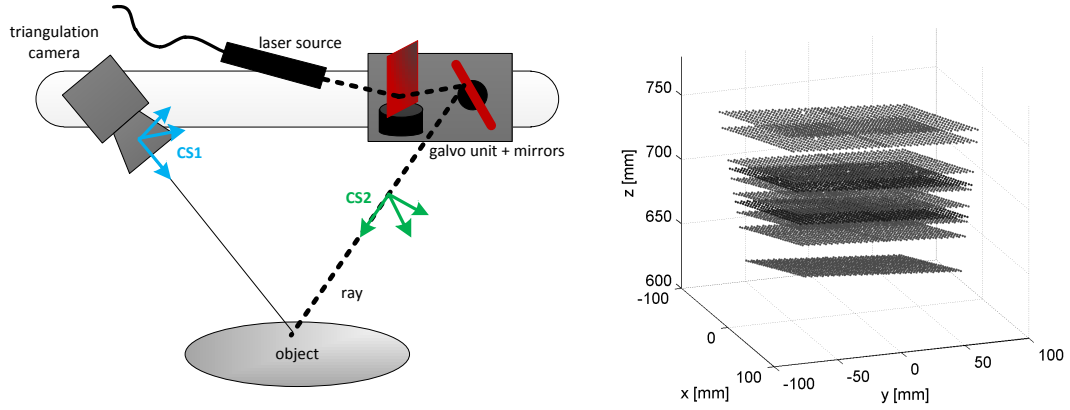
- 2560 × 2160 pixel (5.5 Mpix) resolution

- 16.6 × 14.0 mm (21.8 mm diagonal) sensor with pixel size 6.5 µm
- max. 30 fps scanning speed @ full resolution
- 1.2 e^- readout noise (RMS) and 30,000 e^- well depth
- 16 Bit resolution for analog-to-digital conversion, ∼ 25 % quantum efficiency at 830 nm

The camera was equipped with a Computar M7528-MP objective having 75 mm focal length [46]. The width of one pixel then corresponds to 28.2 µm on the object at 40 cm distance to the target. A 1000 px × 1000 px area-of-interest was selected from the full image size (cf. fig. 4.2B). The laser spot center was aligned with the center of this area. Finally, the aperture of the objective was tuned such that the pixel intensities of the laser spot center drove the 16 bit quantization into saturation (at 5 ms exposure time on skin). In this way, the image sensor could exploit the entire dynamic range for resolving the beam profile as suggested in sec. 3.3.3.
Analogous to the simulation study (cf. fig. 3.17), features were extracted as pixel intensities accumulated within concentric, ring-shaped ROIs around the laser spot center. For each ring a width of 90 px, i.e. 2.54 mm, has been chosen. Defining five of such ROIs results in a total diameter of $2 \times 5 \times 90\,\text{px} = 900\,\text{px}$, i.e. 22.8 mm. The remaining 100 px were left as a buffer for cases where the spot center of gravity is not exactly in the middle of the 1000×1000 px area-of-interest. Later on, angle compensation methods may also require this buffer after re-scaling. Note that the number of ROIs has been reduced from seven to five with respect to the simulation study. Each ROI also covers a larger area and also larger number of pixels. For the experimental validation, this provides a better SNR even for weak backscatter. A lower dimensional feature space is also beneficial for the performance of a machine learning algorithm as will be discussed in more detail later on. As a result, the ROI indices from simulation and experimental evaluation do not directly correspond to each other. For the experimental evaluation, ROI i covers locations at distances $r \in [(i-1) \cdot 2.54\,\text{mm}, i \cdot 2.54\,\text{mm}]$ from the spot center.

4.1.2 Laser Triangulation

Laser triangulation is used to compute positions in 3D space given a laser spot in a 2D camera image. This requires a known spatial relationship between the laser source and the triangulation camera. For the optical setup described in sec. 4.1.1, however, the orientation of the laser beam with respect to the camera is controlled by a galvanometric deflection unit. This enables a fast scanning speed, but also renders the triangulation



A: Illustration of the triangulation problem. Coordinate systems for the camera and one laser beam are indicated in blue and green respectively.

B: A pre-defined grid of 32 × 32 points is projected onto planes at different distances from the camera for calibration.

Figure 4.3: General concept of the laser triangulation used for 3D surface imaging (© 2015 IEEE. Reprinted, with permission from [336]).

problem very complex [336]. A simplified sketch of the problem is presented in fig. 4.3A. Although the author of this work has shown in a separate study that data-driven learning can tackle this challenge in a generalized manner [336], the data acquired for all subsequent experiments was recorded using the LUT approach outlined by Wagner et al. [312, 314]. The approach is briefly described in the following.

Calibration of the Triangulation Camera First, the camera itself is calibrated according to the pinhole model to obtain a projection matrix from the 3D camera space onto the imaging plane [120]. This is achieved by recording images of a checkerboard in several different poses within the calibration space. The calibration optimizes parameters of a model which relate the known spacings of the checkerboard squares to their measured pixel coordinates in the camera images. The parameters consist of intrinsics (pixel coordinates of the optical camera axis, the focal length as well as radial and tangential distortion coefficients) and extrinsics (translational and rotational parameters of each checkerboard pose with respect to 3D camera coordinate system). The origin of the camera coordinate system coincides with the location of the aperture.

Calibration of the Mirror System and Triangulation Now, the LUT approach for calibrating the scanner setup relies on the idea of calibrating each ray of a fixed 32 × 32 grid. This means 1024 fixed spatial relationships between laser beam and camera

are considered. The system is only capable of triangulating along these pre-defined rays. The calibration data is recorded by stepping a checkerboard board through the calibration space and projecting the laser spot grid onto each of them. By using the camera calibration above, the extrinsic parameters of these poses can be obtained. Finally, 3D coordinates of the projected grids are computed by exploiting homography matrices [314, 336]. The result is illustrated in fig. 4.3B.
For each ray a line is fitted through the 3D points (in a least-squares sense), resulting in an estimate of a 3D description for all 1024 rays. Each ray is considered as the optical axis of a virtual camera coordinate system, where the z-axis coincides with the ray pointing to the target. This orthonormal system is known in coordinates of the triangulation camera, since the 3D laser points were estimated from the extrinsics of the camera. Therefore, the orthonormal system of each ray corresponds to a projection matrix. The inverse of this matrix projects from the 3D coordinate system of the triangulation camera into that of the virtual camera. The output of the scanner calibration yields 1024 projection matrices linked to specific rays representing certain mirror positions.

For triangulation, the mirror position i.e. the ray number needs to be linked to the pixel coordinates of the laser spot in the triangulation camera image. Two systems of equations can be set up: (1) Projecting the unknown 3D location into the virtual imaging plane yields the origin of the plane $[0, 0]$, and (2) projecting the unknown 3D location with the intrinsic parameters to the imaging plane of the triangulation camera results in the measured pixel coordinates. To obtain the 3D location, the equations are solved with the direct linear transformation (DLT) algorithm [2]: Since both equation systems result in the same 3D vector, the cross product of the equations' left hand sides has to yield the null vector. After joining the equations this way, different optimization techniques, e.g. involving SVD, can be used to get the solution.
Since one image of the triangulation camera contains more than one spot, the correct correspondence between a projection matrix and the imaged spot needs to be identified. Therefore, a sequence of images is acquired from the same grid, where each spot is switched "on" and "off". The sequence of these two states in the images provides a binary code being unique to each projection matrix. For 1024 spots a 10 bit, and for 16 spots a 4 bit code is required. After a "burn-in" phase for this sequence, a unique spot identification can be done after each incoming bit, or image [312, 314].
The process of acquiring one full 3D surface scan à 1024 spots including all NIR backscatter information took about 20 s. This includes the exposure time for both cameras, driving the mirrors, laser switching times and safety delays to ensure a stable

Table 4.1: General characteristics of the subject cohort (total & skin type specific): number of subjects, gender, age and questionnaire score (Fitzpatrick scale, cf. sec. 3.1.2 and [95]).

	total	type II	type III	type IV	type V
subjects	30	6	8	13	3
male	14	5	4	5	2
female	16	1	4	8	1
age (mean)	31.7	28.2	32.8	33.7	27.0
age (range)	24-65	24-34	24-53	27-65	24-29
score (mean)	19.4	9.0	16.3	23.7	30.0
score (range)	7-31	7-13	14-20	21-26	29-31

measurement situation.

4.2 Data Acquisition

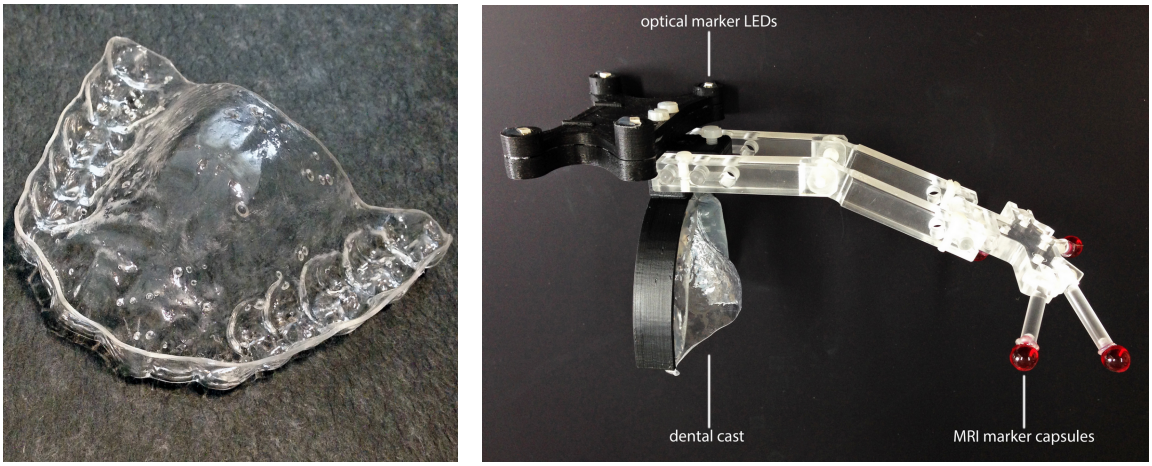
4.2.1 Volunteer Study

Experimental data was acquired from 30 healthy volunteers. The cohort comprised 14 male and 16 female subjects aged between 24 and 65. Full characteristics are listed in table 4.1. Using the aforementioned skin typing questionnaire, each subject was assigned a score. This score classifies into one of the in total 6 different skin types of the Fitzpatrick scale (cf. sec. 3.1.2) [95]. The mean score among female subjects was 18 (range: 7-30), and male subjects 21 (13-31).
All procedures were in accordance with the ethical standards of the responsible committee on human experimentation (institutional[1] and national) and with the Helsinki Declaration. Informed consent was obtained from all subjects for being included in the study.
Per subject the collection of the experimental data was conducted in three steps which are outlined below. They will be detailed in the following subsections.

Fabrication of a Dental Cast First, an informed consent form was filled by each volunteer[2]. Afterward, an individual dental cast was fabricated for the upper dentition of

[1]Ethics committee of the University of Lübeck, Germany, file number 13-112, `https://www.uni-luebeck.de/forschung/kommissionen/ethikkommission.html`

[2]See sec. 7.1. The author acknowledges the help of Dr. med. Benjamin Sack, Department of Neurology, University Hospital Schleswig-Holstein, Lübeck for informing all volunteers about the process of MR



A: Individual dental cast for subject one (S1) fabricated from PMMA material.

B: Dental cast attached to a marker construction with MRI visible capsules and an active optical marker.

Figure 4.4: Subject specific marker construction for experimental steps two and three. The NIR scans were transformed into the reference space of the MR ground truth.

each subject (cf. fig. 4.4A)[3]. The cast was made from Poly(methyl methacrylate) (PMMA) and was used as a mouth bite in all subsequent steps. Through the upper jaw, the cast was rigidly linked to the cranium.
The cast was then attached to an Acrylonitrile butadiene styrene (ABS) base plate produced with a 3D printer. The base plate was equipped with an active optical marker geometry. This geometry contained four 850 nm LEDs. These can be tracked by a marker-based tracking system (accutrack 250, Atracsys LLC [13, 14]). A marker extension as illustrated in fig. 4.4B was only attached to the ABS part for MRI. It was made from PMMA and was equipped with MR visible capsules (Nitrolingual®). Section 4.2.3 will describe how the marker was used to pre-register the NIR scans to the MR ground truth.

Tissue Thickness Ground Truth Recording from MRI In a clinical scenario, CT imaging (e.g. from planning) constitutes the easiest way to record and segment a ground truth for the tissue thickness from the forehead. However, due to the required exposure to radiation, this was not feasible for the group of volunteers in this study. Optical coherence tomography, on the other hand, is an accurate approach, but difficult to handle for

scanning.

[3]The author appreciates the help of Ms. Annika Klaus, Dr. Christin-Sophie Deutz, and Dr. Dr. Hans Christian Jacobsen from the Clinic of Oral and Maxillofacial Surgery, University Hospital Schleswig-Holstein, Lübeck.

a rather large surface. The skin penetration depth is also roughly limited to 2 mm [199]. The usage of pressure-compensated US results only in moderate accuracy [80]. Here, main challenges and error sources occur when registering single measurements into the same space. In addition, skin deformation caused by the contact force of the US probe corrupts the outcome.
After compensating for this force, US results have been shown to agree with the segmentation from MRI. The latter has been chosen as a gold standard for obtaining a highly precise tissue thickness measure[4]. Section 4.2.2 will detail the developed procedure (processing chain and validation published in [333, 334]).

Optical NIR Scans Scans were acquired with the subject in supine position. For stable positioning a head rest (F35758-MD, Orfit Industries n.v.[212]) as typically used in SRT was placed on the couch. The projected laser grid was positioned on the subject's forehead. The laser was exclusively switched on when the subject's eyes were covered by protective goggles. The latter as well as the scanning setup were shielded from the accutrack system and the flashing LEDs by a black curtain. In this way, the HDR images were not corrupted by their strong NIR light capable of passing the bandpass of the HDR camera. Note that the optical tracking system is only required for the experimental validation of the concept. It matches the MR ground truth to any of the optical scans (cf. sec. 4.2.3) and is also used to compensate head motion from spot to spot for this rather slow functional prototype. For the clinical application (where a model is only built once during planning) as well as faster clinical prototypes, the tracking system will not be necessary.
Three subsequent scans in each of three different head poses (nine scans in total) were recorded for every subject. Photographs of the skin on the forehead were taken.

4.2.2 Tissue Segmentation from MR-Scans

To serve as a ground truth, high resolution MR images (0.1025 mm $\times$ 0.1025 mm $\times$ 1 mm) of the forehead region were acquired with an Ingenia 3.0 T MR scanner (Philips Healthcare [158]). Figure 4.5 indicates the volume-of-interest (VOI) in a low resolution T1 scan of the head. The imaging sequence was a gradient echo (FFE-T1) with a 15° flip angle to rapidly record the VOI (210 mm $\times$ 210 mm $\times$ 70 mm). The echo time TE was turned as low as possible (TE/TR = 5/17 ms) to minimize susceptibility artifacts at tissue-air/bone

[4]The author is grateful for the support of Dr. Uwe Melchert, Christian Erdmann, Armin Herzog, and Dr. Georg Schramm, Institute for Neuroradiology, University Hospital Schleswig-Holstein, Lübeck.

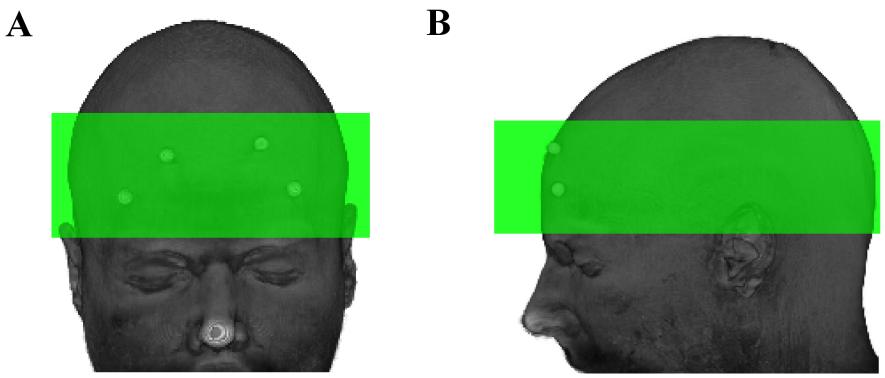


Figure 4.5: Clinical T1 contrast scan of a subject's head. The VOI is marked in green. It contains the forehead region and MR-visible marker spheres for registering the volume into other coordinate spaces. **A:** frontal view (Reprinted from [334]), **B:** view from the left.

interfaces [82]. The volume was aligned in parallel to the AC-PC line and then individually adjusted to be approximately orthogonal to the forehead surface [284].

Tissue Segmentation Pipeline The segmentation chain had nine major steps. It was performed slice-wise. First, the software package SPM8 was used to initially segment five main components of the anatomy (gray matter, white matter, cerebrospinal fluid (CSF), meninges and skull/skin) [96]. The output corresponds to probability maps, which indicate the likelihood for each voxel of belonging to a particular component. Second, the first four components were joined into a negative and the last component (extra-cranial tissue and skull) into a positive mask. The term positive mask refers to an image which, by pixel-wise multiplication with the data image, sets pixels to zero which do not belong to the object of interest (i.e. the skin). A negative mask is an approximations of the inverse counterpart. In order to cut out the interior of the skull, the original image was masked with both of them. Voxel intensities were reset to zero, when their probability of belonging to component five fell below 25 % and the chance of belonging to one of the other components exceeded 68 % (cf. fig. 4.6A-D). The VOI was then selected to contain the forehead only (fig. 4.6E).
Manual inspection in a third step allowed to correct for rare drops in local contrast or large vessels. In rare cases the latter could lead to holes in the tissue region which cannot be closed by previously applied morphological closing operators or a median low-pass filter.
Fourth, the largest, connected tissue segment was extracted using a 2D region growing algorithm (fig. 4.6F). Fifth, 2D snakes were applied to obtain a smoother tissue-bone

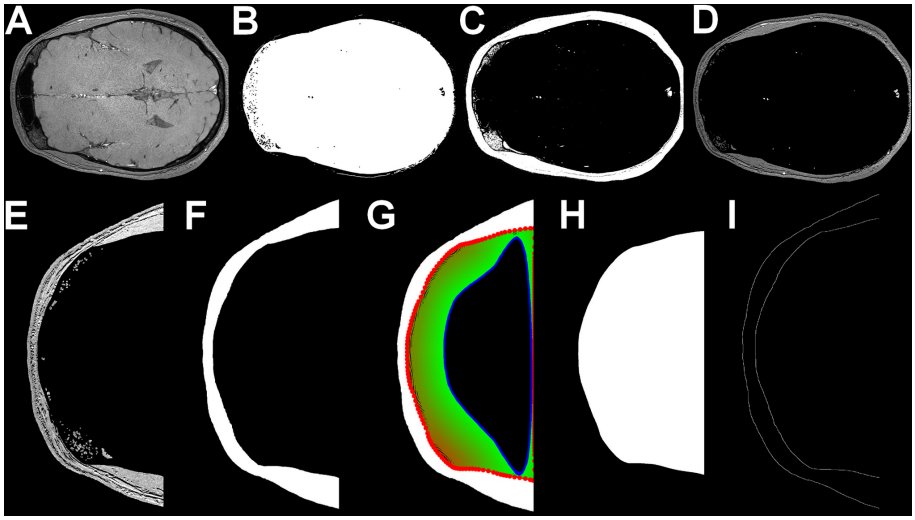


Figure 4.6: Illustration of the segmentation pipeline. **A:** raw volume slice, **B:** negative mask of intra-cranial sites, **C:** positive mask of likely extra-cranial tissue sites, **D:** raw slice after masking, **E:** forehead region from **D**, **F:** region growing output, **G:** active bone contour detection using snakes (blue: initial contour, red: output contour), **H:** snake output segment, **I:** output from **F** and **H** after Canny edge detection (Reprinted from [334]).

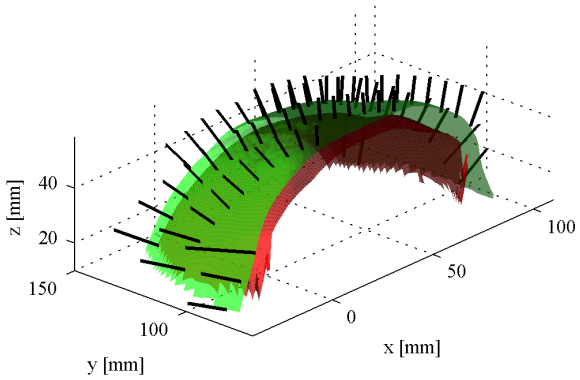


Figure 4.7: 3D point clouds for the skin (green) and bone (red) surface. Tissue thickness is computed by **normal vector (black)** penetration through both surfaces. The orthogonality condition holds for the skin surface (Reprinted from [334]).

boundary [147] (200 support points, 150 gradient descent steps, weights for tension and stiffness were 0.03 and 0.005, respectively). The active contour also employed a balloon force as well as gradient vector flow as an external force. The initial contour and the optimized result are illustrated in fig. 4.6G for one slice. Figure 4.6H shows the segment enclosed in this optimized contour. Sixth, Canny edge detection [45] yields the air-tissue

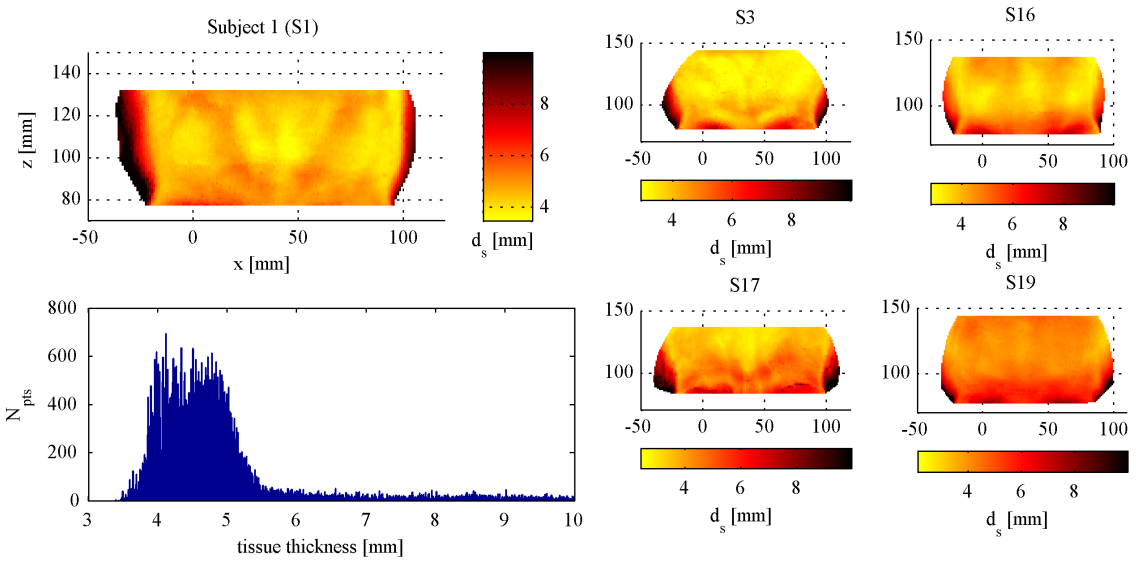


Figure 4.8: Segmentation result and histogram of the tissue thickness for subject 1 (left, reprinted from [334]). The 3D skin surface is overlaid with color-coded tissue thickness. Surface plots for the remaining subjects are given on the right (color bars were set to the same range).

boundary (cf. fig. 4.6I). Finally, agglomerative clustering rejects smaller contour fragments (e.g. surrounding vessels) and collects the point clouds for the skin and bone surface.

The actual tissue thickness was computed from these two clouds as the distance between each point on the skin surface and the penetration point of the corresponding normal vector at the bone surface (cf. fig. 4.7).

Expert Segmentation The algorithmic segmentation was validated based on the segmentation results of five skilled human experts. Each of them manually delineated the skin and bone surfaces for one slice per subject. As a representative subset, data from five subjects (S1, S3, S16, S17 & S19: 3 male, 2 female, aged 25-64) were selected. Representative slices for a forehead region were chosen, which is most likely scanned by the marker-less tracking system. To evaluate the segmentation error propagation to the skin thickness measure, the experts were also asked to segment five slices for S1. The normal vector penetration procedure was applied to these five slices. The resulting thickness measures were compared with the tissue thickness output of the segmentation algorithm. The deviation of the algorithmic from the expert segmentation was computed as the average 3D distance between point-to-point correspondences. These were identified via Euclidean nearest-neighbor search. Finally, expert five segmented one slice of S1 five

Table 4.2: Mean and standard deviation of the absolute differences (in mm) (Data taken from [334]):

A: Between expert and algorithmic segmentation for skin and bone contours.

Subject	Skin [mm]	Bone [mm]
S1	0.101 ± 0.082	0.154 ± 0.126
S3	0.095 ± 0.078	0.185 ± 0.118
S16	0.102 ± 0.071	0.155 ± 0.095
S17	0.082 ± 0.072	0.199 ± 0.128
S19	0.093 ± 0.081	0.168 ± 0.098

B: Between the computed tissue thickness from the expert and algorithmic segmentation (in mm). Tissue thickness was computed from five volume slices.

Subject	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
S1	0.161	0.183	0.162	0.154	0.205
	± 0.123	± 0.151	± 0.145	± 0.128	± 0.159

times. This gave an impression of the intra-operator reproducibility.

Validation Results Figure 4.8 shows the algorithmic segmentation result for all five subjects. The top left part shows the 3D skin surface for S1, overlaid with color-coded tissue thickness. The surface shows smooth subcutaneous and cutaneous variations which originate from subcutaneous vessels, facial muscles or changes in the subcutaneous fat layer. The histogram at the bottom left reveals that tissue thickness on the forehead of S1 mainly ranged between 3.4 mm and 5.6 mm. This interval corresponds to 80 % of the histogram area. Corresponding ranges for the other subjects were 3.35 mm (S19, 70 %), 3.87 mm (S17, 70 %), 2.60 mm (S16, 80 %), and 1.98 mm (S3, 80 %).
The first investigation was dedicated to the accurate determination of the skin or bone contour as such. The deviation of the algorithmic from the expert segmentation is listed in table 4.2A. Figure 4.9 illustrates that the tissue/air boundary was always segmented with an average error of less than 0.1 mm and the bone with less than 0.2 mm. This is on average less than twice the in-plane voxel size. With regard to the theoretical limit of half the voxel size and the thickness ranges across the forehead (here the structures were varying in ranges of 1.98 mm or more), this is acceptable. From the standard deviations

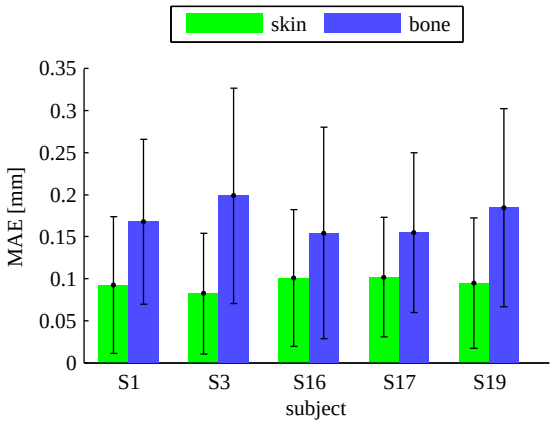


Figure 4.9: Deviations of the algorithmic from the expert segmentation for all subjects averaged across all experts.

shown in table 4.2A, an average inter-operator variability of 0.077 mm for the skin and 0.114 mm for the bone surface was computed (root-mean-square across all subjects).
The error values for individual subjects may vary due to motion during the acquisition process (as it was the case for S3 and S17). The acquisition lasted approximately 16 min. The motion causes a lower SNR and blurs tissue boundaries. Intense motion entails halo-like noise patterns around the head. Generally, segmentation of the bone was more prone to errors due to a poorer contrast with respect to the adjacent cranium or meninges structures. Therefore, errors and variability were higher for the bone than for the skin contour across all subjects and experts.
In the second experiment, the stack of five slices taken from S1 was used to evaluate how the segmentation error on the skin and bone boundaries translate to the tissue thickness measure. The results are given in table 4.2B. The MAE on the tissue thickness in that region was found to be 0.173 mm and for all experts less than 0.21 mm. Thus, the mean boundary deviations do not additively translate to the skin thickness measure. One possible reason for this is that the thickness was extracted along the forehead normal direction, while the slice orientation was not precisely orthogonal.
In the third investigation part, expert five reproduced his results with mean deviations (MAE) of 0.084 mm for the skin and 0.086 mm for the bone. This suggests that significant parts of the aforementioned deviations between algorithmic and expert segmentation may be due to intra-operator variability.

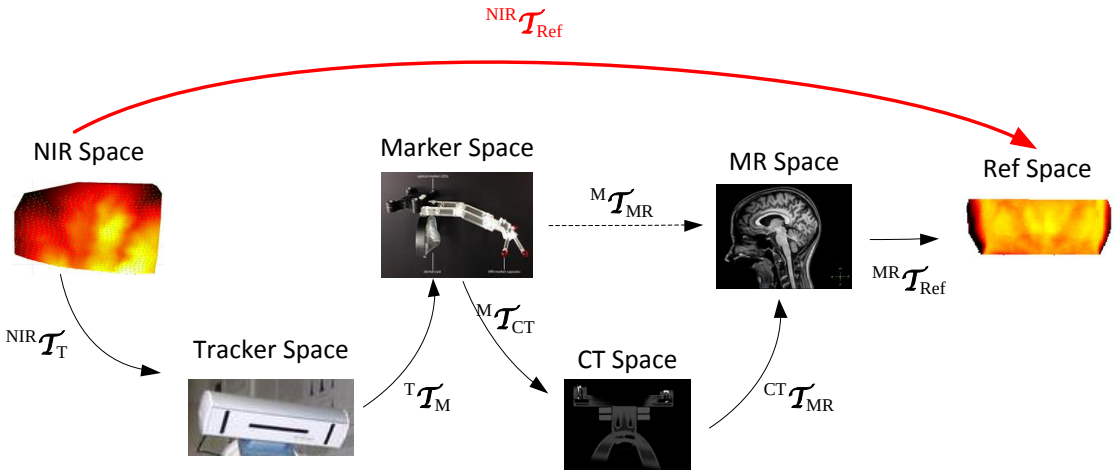


Figure 4.10: Calibration chain of the marker based pre-registration between the MR reference and the triangulation camera space. Single calibrations were fused to obtain the overall transformation denoted in red color.

4.2.3 Ground Truth Registration

After recording the MR tissue thickness ground truth, a registration with the NIR scans from the optical setup is required. Only by mapping NIR features to a corresponding tissue thickness, training data $\mathcal{D}$ can be generated. This will then be used for building a regression model. Exclusively applying ICP to obtain this matching is challenging, since the iterative algorithm may converge into local minima, i.e. spatial similarities. Misregistration may originate from different sources such as: general ambiguities across the reference surface, deformation of the soft tissue surface, the fact that the surfaces were recorded with different modalities, measurement noise or limited spatial resolution. Some of these points were somewhat kept under control since both scans were recorded in temporal proximity. The matching was performed in two steps, to ensure accuracy.

1. marker-based pre-registration (seed)
2. point-to-plane ICP refinement

The calibration chain for the pre-registration process is shown in fig. 4.10. This chain and the reusable bite marker allow for repeated NIR scans and, due to the tracking system, for spot-wise motion compensation during each scan. This is reasonable for the experimental evaluation of the functional prototype, but not required for the clinical application later

Table 4.3: Listing of the mean and STD for different measures across all 30 subjects: subject motion (STD from mean position; averaged across all scans; translations listed as the length of a 3D vector, and rotations as the angle from an axis-angle representation of the rotation matrices), pixel re-projection errors for the NIR-to-tracker calibration, registration error of the optical marker as seen by the accutrack 250 system and its optical geometry file, registration error between optical geometry file and segmented geometry from CT, registration error between segmented geometry from MR and segmented geometry from CT, marker-based pre-registration and refined registration error between NIR scans and MR ground truth.

	total		male		female	
	mean	STD	mean	STD	mean	STD
motion $\|\boldsymbol{t}\|$ [mm]	0.14	0.08	0.16	0.10	0.12	0.05
motion r [°]	0.06	0.05	0.07	0.06	0.05	0.03
reprojection x [px]	0.18	0.14	0.22	0.17	0.14	0.10
reprojection y [px]	0.18	0.10	0.19	0.10	0.18	0.10
$\bar{E}_{T-to-M}$ [mm]	0.19	0.10	0.22	0.12	0.17	0.08
$\bar{E}_{M-to-CT}$ [mm]	0.26	0.09	0.27	0.11	0.24	0.08
$\bar{E}_{CT-to-MR}$ [mm]	0.29	0.06	0.29	0.06	0.30	0.07
$\bar{E}_{NIR-to-Ref}$ (marker) [mm]	3.42	2.22	3.42	2.30	3.41	2.23
$\bar{E}_{NIR-to-Ref}$ (ICP refined) [mm]	0.23	0.04	0.24	0.04	0.22	0.04

on. Chapter 6 will propose disposable and non-reusable alternatives for initial landmark-based pre-registration. They would only be used once during planning.

Calibrating NIR– to Tracker–Space The triangulation of the NIR scan has a nominal accuracy of 0.16 mm [314] and is obtained in the coordinate system of the triangulation camera.

The NIR point cloud was transformed into the coordinate space of the accutrack 250 system [13, 14]. To compute the necessary transformation matrix ${}^{NIR}\mathcal{T}_T$, the marker was tracked for 30 s by the accutrack system, while simultaneously recording images from the triangulation camera. The transformation matrix was then obtained as the product of the average tracked marker pose and the extrinsic parameters of the marker geometry with respect to the camera. For the latter, the camera calibration computed earlier and the geometry stored in the optical marker reference file were used. Re-projection errors of the marker pose onto the imaging plane of the camera were on average less than 0.23 px (cf. table 4.3).

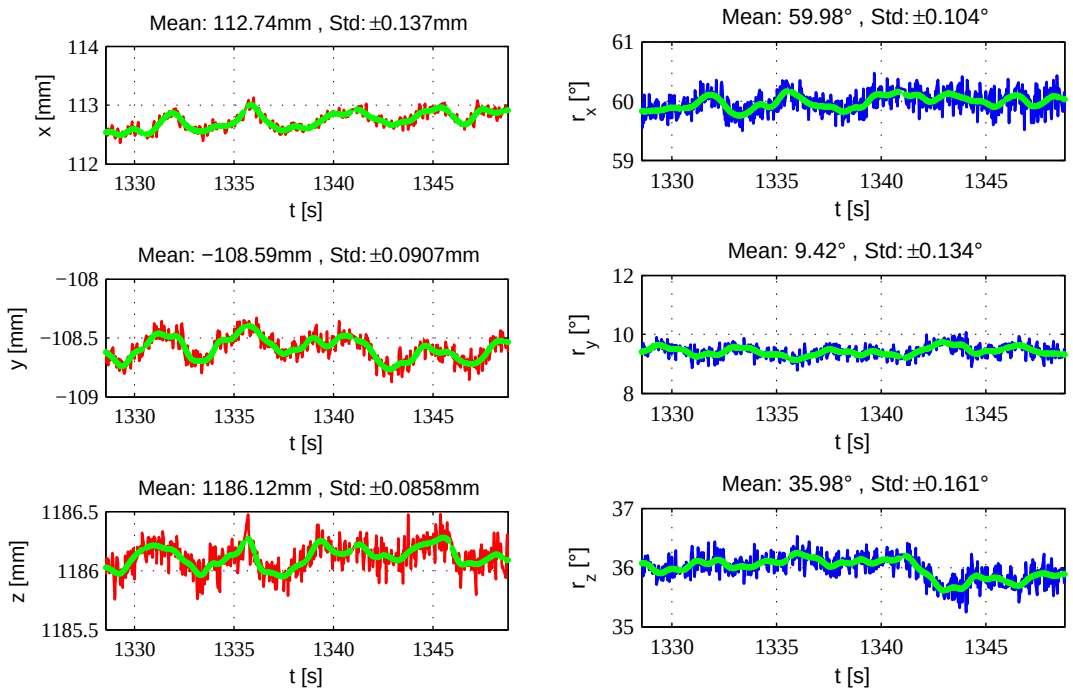
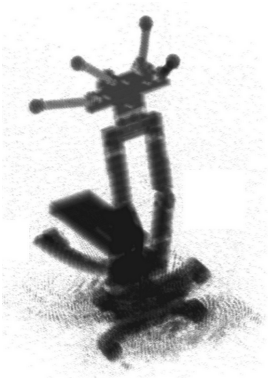


Figure 4.11: Head motion traces for subject one, head pose one, scan one. Temporally filtered position (green) with respect to the tracking camera and the unfiltered deviation from this mean are shown. These deviations include all six degrees of freedom: three translational (red) and three rotational (blue) degrees. Rough orientations of the shown directions: x : LAT, y : SI, z : AP.

Tracker– to Marker–Space and Subject Motion The forehead surface is assumed to be rigidly linked to the cranium, the upper dentition, and therefore to the optical marker shown in fig. 4.4. The active optical marker part is tracked with the mentioned marker-based tracking system, which has a 3D marker identification RMSE of 0.14 mm for targets in a distance of 1.05 m to 2 m [13]. The optical geometry file for each marker is internally used to track all six degrees of freedom. On average, the registration error between this geometry model and the online recording was 0.19 mm. Variations across the subject cohort arose from a re-positioning of the marker-based tracking system. This was done for each subject to guarantee an optimal FoV. The marker pose from the accutrack system was recorded with 80 Hz during the experiment and was hence available for each triangulated spot. Such a pose corresponds to $\left({}^{T}\mathcal{T}_{M}\right)^{-1}$ in fig. 4.10, i.e. the inverse of the transformation from the tracking system into the coordinate space of the stored marker geometry file. This was used to compensate for subject motion at each triangulated 3D



A: 3D rendering of the marker geometry.

B: Slice 250 of the CT scan rendered on the left.

Figure 4.12: CT scan (0.31 mm $\times$ 0.31 mm $\times$ 0.2 mm resolution) of the marker geometry which was used to link the coordinate space of the accutrack 250 system to that of the MR-space. Copper wires in the optical marker produced only minor artifacts.

point. Before compensation, the motion traces were smoothed with a fourth order Butterworth lowpass filter with a cut-off frequency at 1 Hz (cf. fig. 4.11). Filtering and averaging of rotations was done in the tensor space suggested by Brun et al. [40].
Table 4.3 lists the subject motion recorded with the marker during the experiment. The values give the STD around the mean head position during a scan. Translations are given as the length of the corresponding 3D shift and rotations as the angle from the axis-angle representation of the rotation matrix. With the head rest, subjects were able to move in a very limited range only (20 s of approximate scanning time). Significant indications (translation: $p = 0.039$, rotation: $p = 0.076$) were found that male subjects tended to move more than female subjects.

Calibrating Marker– to MR–Space While the optical geometry file only stores the spatial relationships within the LED part of the marker, CT imaging (0.31 mm $\times$ 0.31 mm $\times$ 0.2 mm resolution) was used to measure the geometric relations between the active LEDs and the MR visible capsules. An example CT scan is shown in fig. 4.12. Both, the LEDs and the center of the capsules were manually segmented. The capsules, for instance visible in fig. 4.5, were also segmented within the MR scan. With known marker correspondences between the modalities and the SVD step from the ICP described in sec. 2.3.2.1, the transformations ${}^{M}\mathcal{T}_{CT}$ and ${}^{CT}\mathcal{T}_{MR}$ were computed. These yield ${}^{M}\mathcal{T}_{MR}$. The registration errors for both SVD steps are listed in table 4.3.

Final Pre-registration and Refinement The transformation ${}^{MR}\mathcal{T}_{Ref}$ arises from the VOI definition and further pre-set transformations in the segmentation procedure. Here, *Ref* denotes the reference space into which the MR scan was transformed before tissue thickness segmentation (cf. sec. 4.2.2). The transformation is not corrupted by any errors. Finally, the desired transformation ${}^{NIR}\mathcal{T}_{Ref}$ can be obtained from the calibrations above as shown in fig. 4.10 and eq. 4.1.

$$ {}^{NIR}\mathcal{T}_{Ref} = {}^{NIR}\mathcal{T}_{T} \times {}^{T}\mathcal{T}_{M} \times \underbrace{{}^{M}\mathcal{T}_{CT} \times {}^{CT}\mathcal{T}_{MR}}_{{}^{M}\mathcal{T}_{MR}} \times {}^{MR}\mathcal{T}_{Ref} \tag{4.1} $$

This matrix is used as a seed for a subsequent point-to-plane ICP refinement. The registration errors before and after ICP refinement can be seen from table 4.3. The ICP refinement transformations were averaged across all scans from all head poses to obtain a more robust estimate. During all these scans, the subjects did not remove the dental cast from their upper dentition, i.e. the refinement would ideally result in the same matrix. The listing shows, that marker-based pre-registration had an accuracy within the low millimeter range. This partly originates from the calibration errors discussed above, but was also found to be due to a limited insertion accuracy of the dental cast. The cast was fabricated very tight and stiff. Therefore, it was experienced as challenging or worrisome for some subjects to reproducibly attach the marker to their upper dentition.

4.3 General Aspects of Light-Tissue Interactions

NIR scans have been recorded for all 30 subjects. A typical result for the backscatter features is shown for subject one (S1) in fig. 4.13. The features correspond to intensities accumulated within one of the five ROIs. ROI 1 generates only weak structures. Partly, this is influenced by the laser spot power being tuned to drive the spot center into the saturation level of the HDR camera. Therefore, some pixels of ROI 1 are saturated and do not contribute to variations across the surface. This is weaker at the boundaries of the grid, where the reflected laser power drops since the laser hits the surface under relatively flat angles (cf. fig. 4.13F). This increases the amount of non-saturated pixels. The saturation however only affects very few pixels (for S1 $<1\,\%$). Another reason for the weak patterns in ROI 1 was indicated by the MCML simulations: the proportion of light reflected from upper tissue layers, e.g. from the epidermis, is substantially higher close to the spot center. Therefore, many photons in this ROI will have a low penetration depth. The experimental observation supports this finding.
All ROIs show dark red regions for spots at the grid boundary (in fig. 4.13A-E in the upper and left grid part). This drop in intensity is caused by flatter incident angles, where

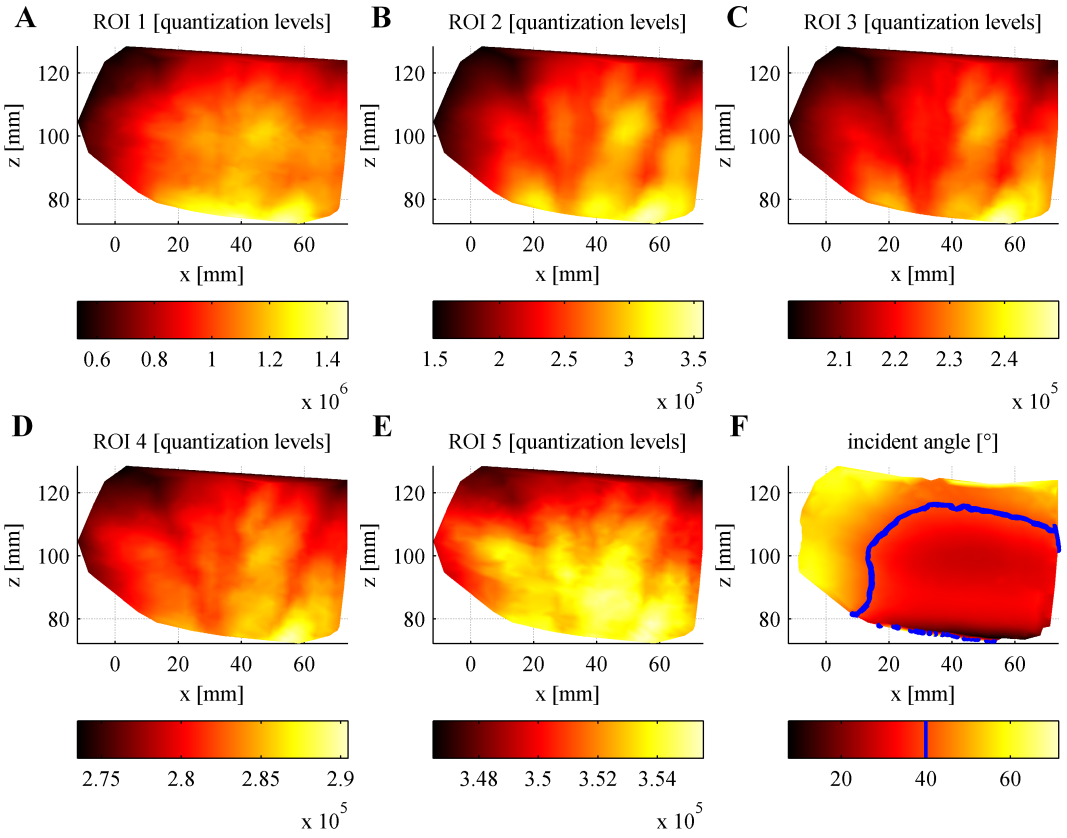
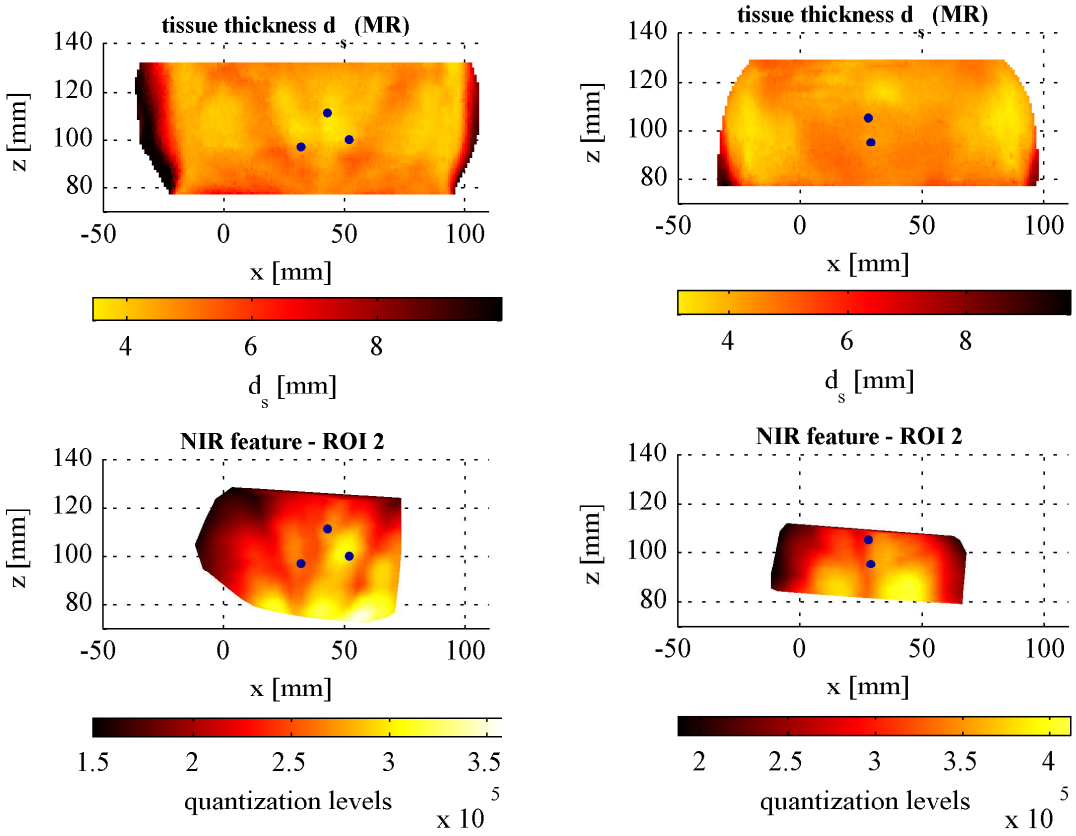


Figure 4.13: Backscatter patterns from the forehead patch of subject one (S1). The plots illustrate variations in NIR backscatter recorded from the five ROIs (**A-E:** ROIs 1-5) across the scanned surface area. **F:** The distribution of the incident angle for this optical scan.

the circular spot shape turns into an ellipse. The behavior is in line with simulations conducted in sec. 3.3.4. Besides, regions of smaller angles exhibit also other patterns. These originate from changing optical properties of the skin, which are linked to the tissue thickness. Figure 4.14 confirms this impression and sets the scanned NIR pattern in relation to the MR tissue thickness ground truth after transforming both into the *Ref*-space. In these plots, the eyes are at lower, and the hairline at higher z-coordinates of the *Ref*-space. Figure 4.14 compares typical cases of a male and a female subjects with Caucasian skin type. It can be seen that cutaneous structures are more prominent for the male subject, while tissue thickness varies smoothly, with smaller gradients, for the female subject. This is, for instance, caused by variations in the fat and muscle tissue and, most important, by subcutaneous vessels. The latter comprise the supraorbital and supratrochlear arteries and veins as discussed in sec. 3.1. It was observed that they are

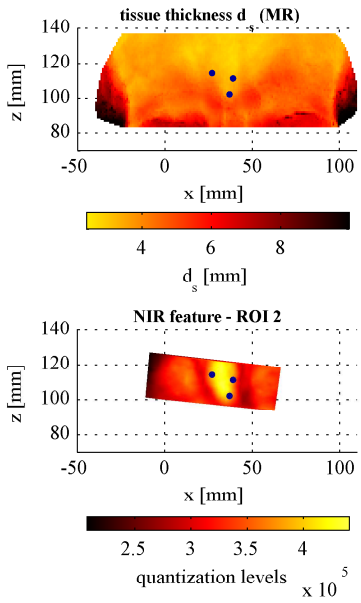


A: Subject 1 (male, aged 27, skin type IV).

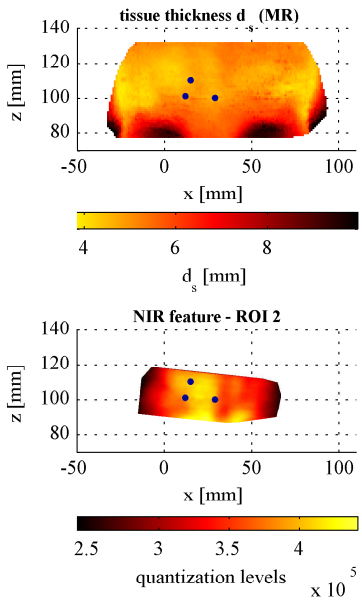
B: Subject 5 (female, aged 33, skin type III).

Figure 4.14: Tissue thickness patterns segmented from the MR ground truth (top) are compared with patterns of NIR backscatter features from ROI 2 (bottom). Both surfaces reside in the *Ref*-space as introduced in fig. 4.10. The blue dots mark interesting locations within the forehead patterns. They are the identically located in both plots and provide therefore better orientation.

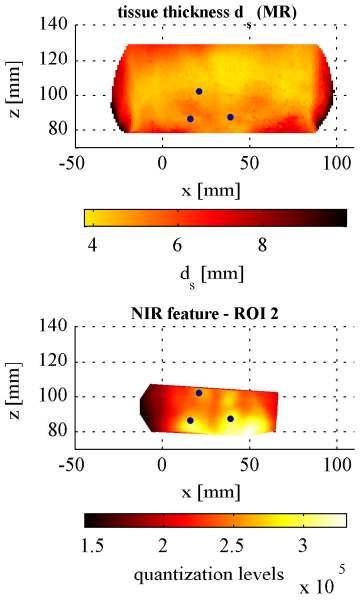
more distinct for many male subjects and hence have more impact on the tissue thickness for them. Unlike the pattern in fig. 4.14B, the NIR backscatter in fig. 4.14A shows this effect. Apart from increased thickness, higher absorption and scattering caused by the increased blood content lead to an even higher drop in backscatter intensity. In the MCML simulations of sec. 3.2.2, a corresponding skin model which contains a large vessel would be subject to a substantial increase in oxy- and deoxy-hemoglobin and most relevant: water. This would cause a high absorption according to fig. 3.4. This is what the reduced backscatter intensity along vessels in fig. 4.14 and fig. 4.15 also reflects.



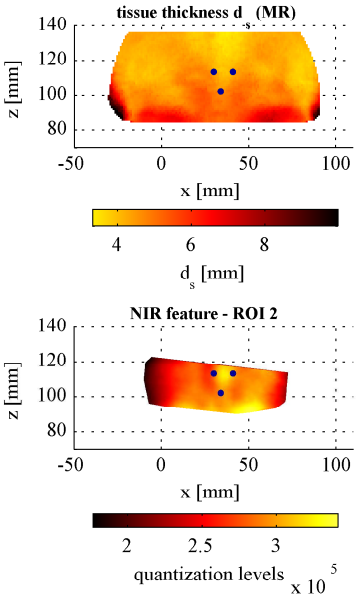
A: Subject 17 (male, aged 65, skin type IV).



B: Subject 24 (female, aged 54, skin type IV).



C: Subject 21 (male, aged 24, skin type V, Indian).



D: Subject 27 (male, aged 29, skin type IV, Asian).

Figure 4.15: MR-to-NIR comparison for elderly subjects and particular skin types. As in fig. 4.14, blue dots again mark individually chosen and interesting locations within the patterns to better identify similarities.

The smooth patterns are best visible for ROIs 2, 3 and to some extent ROI 4. ROIs 2 and 3 cover distances from 2.54 mm to 7.61 mm from the center. This supports the conclusions from MCML simulations which indicated that backscatter at medium distances from the spot center is characterized by a good SNR. For higher ROIs, particularly ROI 5, the tissue-related effects are much smaller and more prone to noise corruption. This comprises measurement noise from the hardware such as the quantization process, and external disturbances (cf. the shiny effects in ROI 5 in fig. 4.13). This suggests that information retrieval becomes more and more challenging and in fact critical within ROI 4, i.e. at 7.61 mm to 12.7 mm distance from the spot center. The scales of the color bars also confirm the decreasing magnitude of the recorded effects: The average number of quantization levels per pixel can be obtained by normalizing the ranges in fig. 4.13 by the total number of pixels in each ROI. The decrease of the normalized values and the width of the min-max variation range reflect this trend for S1.

ROI 1: 1900 - 5265 $\frac{levels}{px}$ (25285 px in total)
ROI 2: 176 - 423 $\frac{levels}{px}$ (76052 px in total)
ROI 3: 142 - 177 $\frac{levels}{px}$ (126976 px in total)
ROI 4: 138 - 147 $\frac{levels}{px}$ (177824 px in total)
ROI 5: 136 - 140 $\frac{levels}{px}$ (228780 px in total)

Distinct subcutaneous structures and prominent NIR patterns where also observed for elderly subjects as shown in fig. 4.15A and fig. 4.15B. Wrinkles and vessels tend to augment the structural information for aged skin. Finally, fig. 4.15C and fig. 4.15D show that even for dark and Asian skin types such structures were discovered. Visual inspection did not reveal differences between these types and typical Caucasian skin. This is in line with earlier simulations and the optical properties of human skin. The Fitzpatrick skin type and its visual appearance is mainly influenced by pigmentation and hence the melanin content. Melanin, however, plays only a very minor role for wavelengths in the NIR range. A higher impact is given by the skin water content, which is hardly related with the skin type according to the Fitzpatrick scale.

Feature Space Analysis The ROI feature space is a five dimensional coordinate space which maps to a scalar value, namely the tissue thickness. Each laser spot in the scanning grid corresponds to one data sample in that space. For illustration purposes, fig. 4.16 projects all the data onto one dimension of the space. This is done for all five ROIs resulting in plots similar to fig. 3.17 obtained from MCML simulations. The plots were generated for S4 (male, aged 35, skin type III). His rather large head size and a relatively wide and

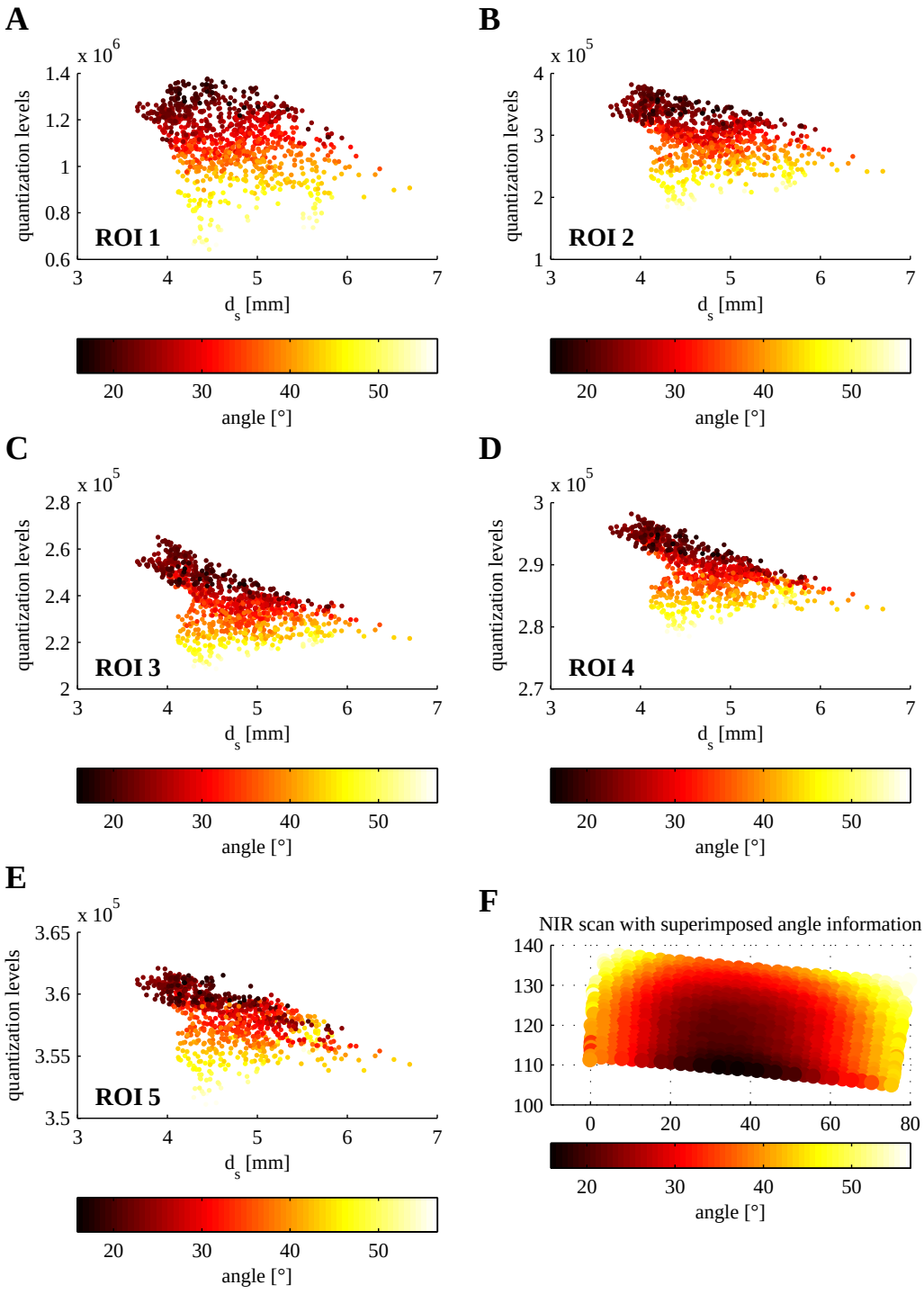


Figure 4.16: Functional relationship between features and target labels (tissue thickness) projected onto one dimension of the feature space for plotting (**A-E:** ROIs 1-5, subject 4). **F:** The distribution of the incident angle is shown with the forehead in *Ref*-space coordinates.

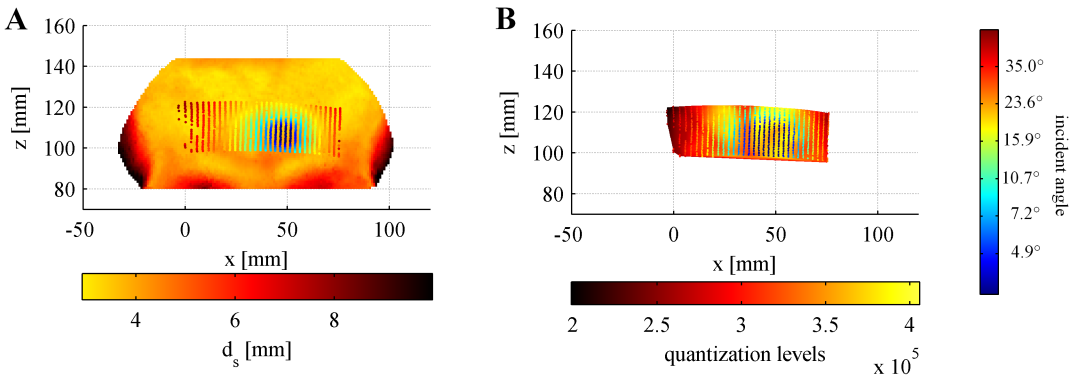


Figure 4.17: Angle influences within the feature space illustrated on the scanned surface patch. **A:** The scanned laser spot grid is overlaid on top of the MR surface with colored incident angle information. **B:** The angle influence introduces (dark) red regions at the grid margins due to flat angles. This mimics thicker skin which is in fact not there as it can be seen on the left plot (data from S3).

flat forehead limited the impact of the incident angle. This data set therefore suggests itself for an analysis of tissue related features. Compared to the simulated feature space, the functional relationship between optical backscatter and tissue thickness is much more complex and noisier for the real case in fig. 4.16. This is due to many effects in the real data, which also have an influence on the reflected light intensity. Lateral heterogeneity across the forehead is caused by freckles, vessels, moles and others. However, one of the main disturbances is given by the incident angle of the laser beam. Its effect on the data is encoded in the color of each data sample.

The following can be observed: Both, the incident angle, as well as the tissue thickness are negatively correlated with the ROI intensities. A decrease in optical backscatter may, therefore, be caused by a flatter incident angle (departure from orthogonal irradiation at $\alpha = 0°$), or and increase in tissue thickness. The plots also suggest, that the angle effect is stronger, i.e. it decreases more rapidly for increasing ROI intensities than the tissue thickness. Finally, this also means that one effect may mimic the other. Figure 4.17 illustrates this effect for a forehead scan of subject 3.

The behavior of the backscatter (i.e. the extent of its increase or decrease) for a defined step in tissue thickness mainly depends on two effects: the absolute tissue thickness at which this change happens or local skin characteristics, respectively. In this context, the effect on the backscatter can be smaller, if (1) the skin is already very thick at the location of interest, (2) there is a vessel underneath the surface, or (3) other things such as hair follicles occur. On the other hand, if two spots were recorded at locations with a slight

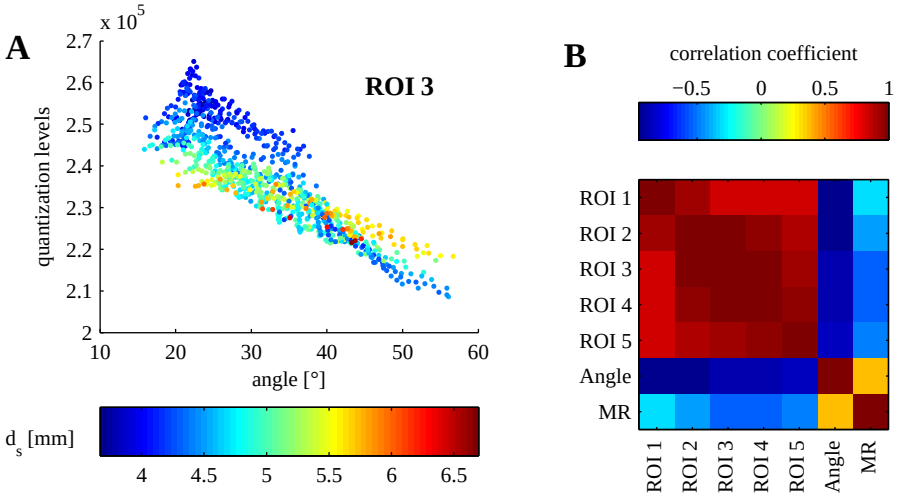


Figure 4.18: A: Relationship between incident angle and the features for ROI 3 of S4. **B:** The linear correlation coefficient between all ROI features, the incident angle, and the tissue thickness ground truth for S4 is illustrated as a color-coded matrix.

difference in thickness, but one is located on a vessel and one is not, then the related change in backscatter intensity will be amplified.

Angle Influence Figure 4.18A addresses the problem of angle influences on variation in the data for S4. Now, the color coding represents the tissue thickness. Clearly, a strong negative correlation can be observed. The dependency on the tissue thickness is also apparent, however less strict in its negative correlation. Nonlinear effects, as discussed above, influence the behavior. Figure 4.18B illustrates the mutual Pearson correlation coefficients between NIR features, the incident angle and the tissue thickness ground truth from MRI. This coefficient captures only the linear correlation, and is related to the slope of a linear function fitted to the data. The plot confirms that both, tissue thickness and incident angle are negatively correlated to the optical backscatter features, whereas the coefficients for the angle are higher in an absolute sense. As expected, there is a high positive correlation among the optical features. Their joined behavior in the 5D feature space will provide much clearer information about the changing tissue thickness than the single 1D projections in fig. 4.16 do.
To evaluate the correlation of both, the angle and the tissue thickness with the optical backscatter, the following analysis has been conducted (across a grid of N_p spots): Let $\boldsymbol{b}_i$ be a vector joining all entries from the i-th dimension of all available feature vectors $\boldsymbol{B}$. First, the mean was removed from each ROI feature $\boldsymbol{b}_i$, the tissue thicknesses $\boldsymbol{d}_s$, and the

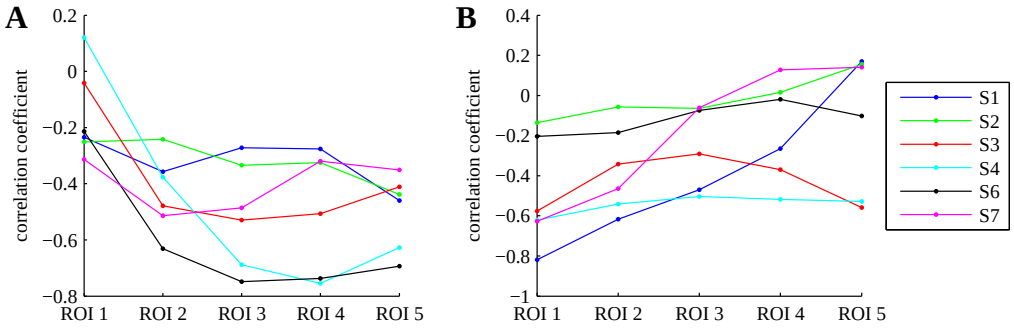


Figure 4.19: Pearson correlation coefficient for the relationships between the five ROI features and tissue thickness (**A**) or incident angle (**B**), respectively. Data shown for S1, S2, S3, S4, S6, and S7.

incident angles $\boldsymbol{\alpha}$. Afterward, the tissue thickness was fitted (in a least-squares sense, cf. eq. 4.2) to each feature:

$$\boldsymbol{b}_i = k_{reg} \cdot \boldsymbol{d}_s + \varepsilon_{res} \quad \text{with} \quad k_{reg} = \left(\boldsymbol{d}_s^T \boldsymbol{d}_s\right)^{-1} \boldsymbol{d}_s^T \boldsymbol{b}_i \tag{4.2}$$

Here, ε_{res} is the residual after a linear least-squares fit. Then, the fitted proportion has been regressed out from the feature signals. Finally, the Pearson correlation coefficient was computed between the residual ε_{res} and the angle α. This ensures a conservative estimate of the correlation between angle and features. It is conservative, because the proportion of the feature signal which correlates with the tissue thickness has been removed. This also removes parts of the feature signal which may be explained by either angle or tissue thickness simultaneously. The latter originates from unwanted correlation between angle and thickness. The extent of this correlation depends on (1) the nature of the cutaneous structures across the forehead, (2) the head shape, and (3) the head pose under the laser scanner. For example, imagine the following scenario: For a subject the tissue thickness is lowest at the center of the forehead. From there it increases with a small gradient to all sides. Now, when the laser scanner is placed above the head such that the ray in the center of the scanning grid hits the tissue thickness minimum orthogonally, then the incident angle would also grow to all sides due to the deflection-based scanning procedure. This scenario entails high positive correlation between angle and thickness by coincidence. There are practical cases where this scenario becomes true to a certain extent (cf. fig. 4.18).
A similar procedure leads to a conservative estimate for the correlation between tissue thickness and features after regressing out the angle. The results for subjects S1-S4, S6, and S7 are presented in fig. 4.19. This analysis assumes that a linear model would

Table 4.4: Pearson correlation coefficient between incident angle and features. The table lists the mean correlation computed with the Fisher transform [90], as well as the minimum and maximum value across all 30 subjects.

measure	ROI 1	ROI 2	ROI 3	ROI 4	ROI 5
mean	-0.783	-0.721	-0.661	-0.700	-0.676
min	-0.894	-0.837	-0.812	-0.814	-0.831
max	-0.525	-0.468	-0.385	-0.361	-0.251

fully explain the relationship between the correlates and the features. It does not account for nonlinear behavior. The simulations in fig. 3.17 show that this is only a rough approximation. Practically, this is even more true, since vessels, for instance, observed in the NIR signal do not necessarily affect the tissue thickness obtained from MRI. The same applies to other heterogeneity effects or the incident angle. Therefore, a thorough analysis across all subjects is challenging, since nonlinearities would need to be accounted for. However, the subset of subjects mentioned above has been found suitable for getting an impression of the underlying processes. Figure 4.19A shows several effects: (1) the correlation between thickness and backscattered intensity is negative. Thicker skin means less reflected light. Further on, this correlation tends to be stronger for higher ROIs. The findings agree with results obtained in an early case study with a robotized laser scanner [327]. Depending on the subject, correlation may get slightly weaker for ROIs 4 and 5, while it is strongest for ROI 3. Irrespective of a concrete ROI number, this general trend is in line with the expectations arising from earlier MCML simulations: The relative proportion of light deeply penetrating the tissue increases for ROIs further away from the spot center. This entails a stronger correlation coefficient. However, the information signal also gets weaker with lower SNR and may also exhibit a more nonlinear relationship to the thickness. This weakens the linear correlation coefficient, such that practically the clearest correlation can be obtained from ROIs at medium distance from the spot center. The saturation of some pixels in ROI 1 and a high proportion of light from upper tissue layers also affect the correlation coefficient.
On the other hand, fig. 4.19B reveals that the negative correlation between angle and features is highest in the inner ROIs and decreases for the outer ones. This is in agreement with the simulations (cf. fig. 3.23) which gave rise to the conclusion that scattering in the tissue reduces the impact of the beam direction (”blurring” effect). Interestingly, the correlation coefficients between angle and features are higher in an absolute sense if the thickness effects are not regressed out beforehand (cf. table 4.4). This indicates that parts

Table 4.5: Proportions of the feature variance which (in a linear sense) account for different correlates. For a scan of S3, the signal variations for all five ROIs have been analysed. The table lists the proportion of the total signal variance that can be linearly explained by the different correlates. Variations of the incident angle, which linearly correlate with the tissue thickness have been removed.

measure	ROI 1	ROI 2	ROI 3	ROI 4	ROI 5
prop. tissue [%]	2.7	17.1	24.7	18.9	4.4
prop. angle [%]	40.0	12.0	7.9	13.7	36.9
residual [%]	61.3	72.2	68.2	68.8	62.5

of the variation in the feature signals could be (linearly) explained by either the angle *or* the thickness. Given the data, there is no indication to distinguish between the two. The correlation between angle and thickness leads to covariation effects between the two: For the subjects in this case study 1.5% (S1), 0.3% (S2), 10.1% (S3), 0.4% (S4), 0.9% (S6), 18.4% (S7) of the tissue thickness variance could be linearly explained by variations of the angle. It is straightforward to see that this specific conclusion is flawed, but the uncertainty about the origin of variations in the backscatter features on the other hand remains. To which extent the angle mimics effects that may also arise from thickness changes is not clear from the data. Misinterpretations from the machine learning side can only be fully avoided by acquiring more data from different views. Finally, it can be expected that for nonlinear models this ambiguity becomes more challenging.
Table 4.5 assumes that this uncertainty is clarified in favor of the thickness. It lists the proportions of the total feature variance which can be explained by thickness or angle variations. The thickness is regressed out first, and then the angle. In ROI 3 the proportion of thickness-related variation is with 25% highest. For the inner and outer ROIs this proportion drops, while the proportion for the angle increases up to 40%. This is in agreement with fig. 4.19. While trends are similar for the other subjects, absolute values for the tissue thickness proportion partially reach more then 50%. High proportions for the residual can be explained by the restrictions to a linear model not fully describing the functional relationship.

Finally, the procedure of regressing out the angle shall be mentioned as an option to partly get rid of the angle influence. It will be investigated in more detail later on. Figure 4.20 shows the remaining feature variation when regressing the angle out. This regression is conservative: In contrast to the previous processing, all ambiguous

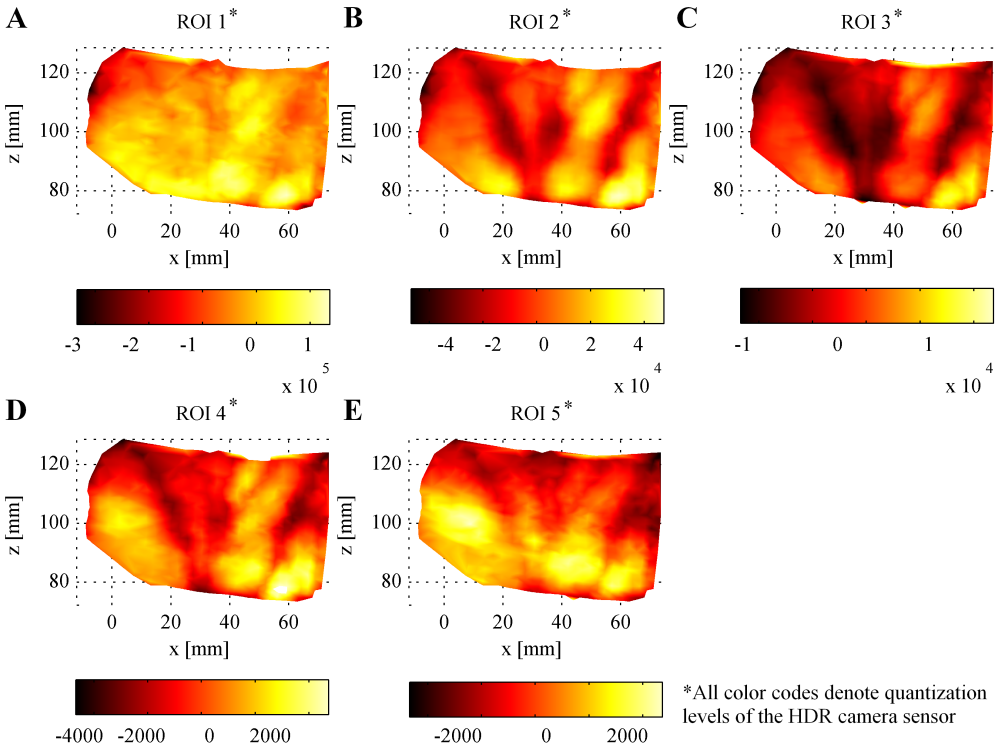


Figure 4.20: Backscatter patterns from the forehead patch of S1 with the angle regressed out from the features. The plots **A-E** illustrate variations in NIR backscatter of ROIs 1-5 which have no correlation with the incident angle, but still with the tissue thickness (cf. with original data in fig. 4.13).

variations are implicitly postulated to be due to the angle influence. In the worst case some of the actual information is removed from the features. It can be seen, that the red shadings at the margins (cf. fig. 4.13) disappear and structural information also found in the MR is emphasized.

4.4 Tissue Thickness Estimation and Prior Knowledge

The simulation studies in chapter 3 were conducted to identify optimal conditions for retrieving information about the tissue thickness from optical backscatter. Although a direct, quantitative comparison between the average Caucasian skin model and experimental in-vivo measurements is hardly feasible, general trends and a qualitative evaluation give valuable information for specifying the setup. The simulation model always remains a simplified approximation of the real world which cannot take all impact factors

or variations within a subject cohort or even within the restricted forehead region into account.
The previous subsections found evidences in the real data that are in agreement with the aforementioned simulation results and therefore support the earlier conclusions:

- Features and tissue thickness are negatively correlated.
- Backscatter features at medium distance from the spot exhibit the best SNR (roughly 4-8 mm from the spot center).
- The mapping between features and tissue thickness is a nonlinear one.
- The incident angle correlates negatively with the features and is the main confounding factor.
- Angle influence is highest close to the spot center and decreases with increasing distance to it.
- The impact of the angle tends to be higher than that of the tissue thickness on the backscatter characteristics.

Based on these findings, this section will now investigate the prediction of tissue thickness from the ROI features via machine learning. As a general goal, it has to be evaluated how the more challenging real-world conditions and the limited accuracy of the ground truth affect the prediction errors. Therefore, the next two subsections will start with initially discussing two ideas for incorporating prior knowledge: Based on the simulation study, sec. 4.4.1 conducts a case study to investigate promising approaches of how to handle the incident angle. The goal is to get a beneficial effect on the prediction performance. Then sec. 4.4.2 discusses an approach that aims at incorporating spatial neighborhood information into the learning problem. Finally, sec. 4.4.3 takes the core findings of both subsections and evaluates them on the entire subject cohort. A comparison to merely using backscatter features is made.

4.4.1 Handling Changes of the Incident Angle

Section 4.3 suggests that separating relevant information from disturbances is expected to be challenging. The following paragraphs will evaluate two general approaches for tackling the major disturbance factor: changes in the incident angle of the laser beam. These are:

1. ***Compensation*** of angle effects by pre-processing the backscatter data.

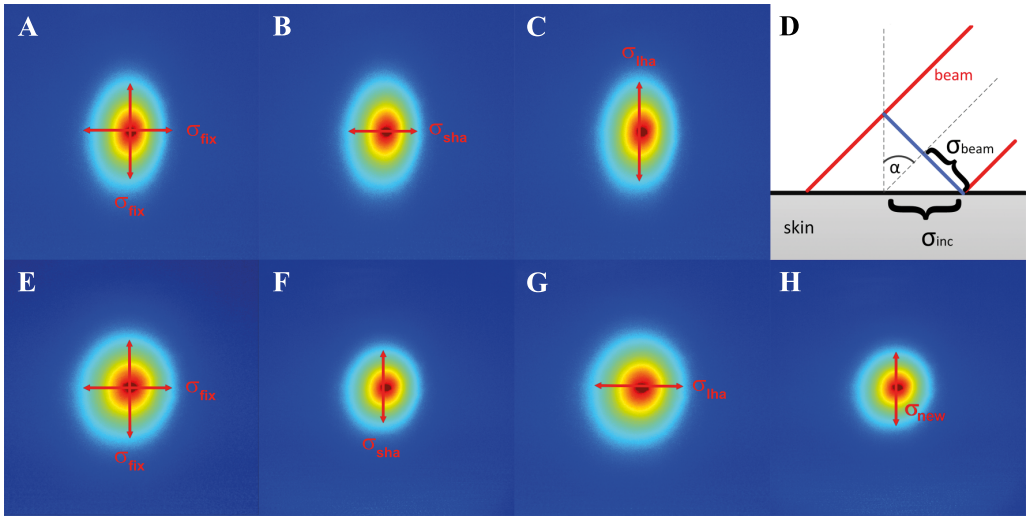


Figure 4.21: Visualization of different approaches to compensate for effects of the incident angle. The top row marks the reference measure for the spot size on a raw spot image ($\alpha = 45°$): **A:** fixed values irrespective of the spot profile (*fix*), **B:** the short half axis of the spot (*sha*), **C:** the long half axis of the spot (*lha*), **D:** the actual beam radius estimated with eq. 4.3 from the incident angle and the length of the long half axis (*inc*). The bottom row (**E-H**) shows the spot image after compensation (scaled image dimension marked with red arrows).

2. ***Extension*** of the feature space by a measure of the incident angle as an additional sixth feature. The originating feature space will be labeled with A.

Figure 4.21 introduces the first four of the in total five compensation approaches under consideration. These rely on pre-processing the images from the HDR camera. This is done independently for each image. First, the spread of the logarithmic spot profile along both half axes of the elliptical contours is estimated by the STD of a fitted Gaussian. The orientation of the main axes in the camera image is determined using weighted Principal Component Analysis (PCA).
Second, a reference value for the spot profile spread is selected. This distinguishes the first four approaches: The reference is either (1) some pre-defined fixed value σ_{fix}[5] (*fix*), (2) the length of the shorter half axis σ_{sha} (*sha*), (3) the length of the longer half axis σ_{lha} (*lha*), or (4) a STD estimate σ_{new} of the original spot STD σ_{beam} (*inc*). The latter estimate is computed by eq. 4.3 which can be trigonometrically derived as illustrated in fig. 4.21 on the top right.

[5]The choice of this fixed constant was based on the STD values of logarithmic spot images which have only been subject to minor or none angle influence. A value of $\sigma_{fix} = 340$ px was selected.

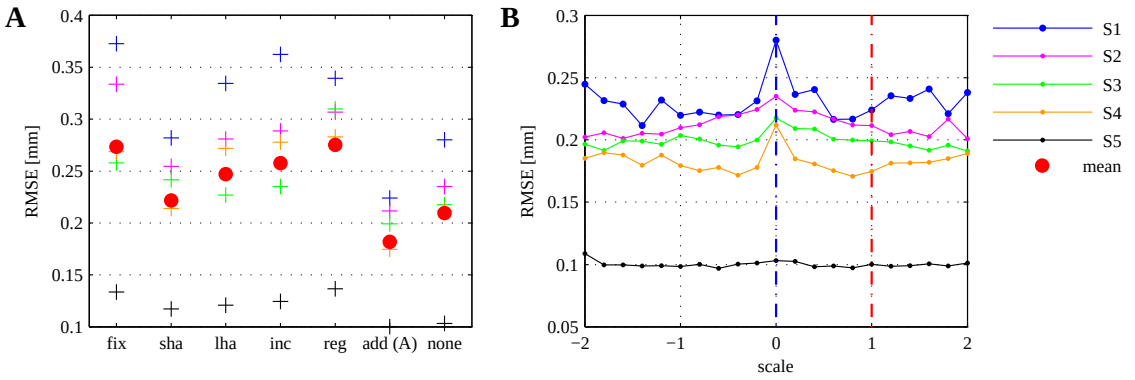


Figure 4.22: Prediction results for different compensation methods (*fix*, *sha*, *lha*, *inc*, and *reg*) and the feature space extension approach (*add*) (yielding space *A*). **A:** All approaches are compared with the outcome obtained with no compensation and no extension (*none*). They are compared using the RMSE of the tissue thickness prediction by SVR (data from subjects S1-S5, mean value in red). **B:** RMSE for feature space extension by the incident angle as a sixth dimension: The impact of the angle feature relevance is changed by varying its scaling factor. The RMSE at scale 0 marks the *none* case and the red line at scale 1 the *add* case from the left plot.

$$\sigma_{new} = \sigma_{beam} \approx \sigma_{inc} \cdot \sin(90 - \alpha) \tag{4.3}$$

The reference is then used to scale one or both of the half axes to this reference value (cf. bottom row in fig. 4.21). This yields a circular shape for the pre-processed spot from which the ROI features can then be extracted. A fifth compensation approach (*reg*) follows the procedure illustrated in fig. 4.20. As for the *inc* estimate, the fifth approach also exploits a measure of the incident angle. The latter is computed as the angle between the normal on the triangulated surface and the incident laser beam. The beam position in triangulation coordinates is known from the calibrations described in sec. 4.2.3. Using a least-squares fit, the coefficients for a linear regression model are estimated. With this model, the linear angle effect is subtracted ("regressed out") from all ROI features. Therefore, the approach does not work in an image-wise manner like the alternatives, but considers the entire scan at once. This yields the *reg* estimate.

All five approaches aim at removing all effects possibly caused by the angle, irrespective of whether it may actually originate from thickness changes or not. In contrast, the extension approach termed *add* uses the computed angle estimate mentioned above and adds it as a sixth feature to the five ROI features. Note that it is pre-processed like the other features by subtracting the mean and scaling it to unit variance. The approach *add*

yields feature space A as a result.

All approaches have been evaluated on the data of S1 to S5 using SVR. SVR used the isotropic RBF kernel and was evaluated analogously to the simulations in a 5-times-10-fold CV scheme. The main results are shown in fig. 4.22A and have been published in [328]. The *fix* approach which was also used in the simulation chapter yielded only poor results. One reason is, that scaling to fixed values will also destroy changes which affect the entire spot and not only one half axis.
Similarly, the *lha* scales the short half axis up to the long one. The latter axis is the orientation actually affected by the angle. This may accidentally amplify the angle effects. Thus, the RMSE of the thickness prediction was poor for this approach as well. This is in contrast to *sha* which tries to reverse the angle effect by down-scaling the long half axis.
Generally, all approaches are affected by the fact that the reflected beam profile is not perfectly Gaussian. Changes in thickness and angle may influence the nature of this non-Gaussianity. This is because thickness changes may have different reasons and may therefore affect distinct skin layers differently. This could provide an explanation why the *inc* and *reg* approaches failed. The backscattered beam profile may just be subject to much more complex deformations.
Adding the incident angle as an additional feature achieved the best results on average. The RMSE was found to be better than for the *sha* approach, which was the best compensation method. However, the angle may be differently relevant relative to the backscatter features. Therefore, different scaling factors have been applied after standard pre-processing and before SVR testing. Indeed, fig. 4.22B shows that the RMSE is worst for a scaling factor of zero (no angle included), but also not best for a scaling factor of one. Since the RBF kernel for SVR treats the feature space isotropically, an individual scaling factor for the angle would need to be determined externally by grid search as done for fig. 4.22B.

The findings give rise to the following conclusions: Compensation approaches involving additional pre-processing gave poor results. Promising approaches such as *sha* were not consistent in their results across the subjects of this case study. The best approach is given by extending the feature space by an additional sixth dimension being the appropriately scaled angle. This is in agreement with the simulation results presented in table 3.3, where this approach also outperformed its alternatives.
Finally, SVR is not optimal for meeting the requirements of e.g. efficiently incorporating

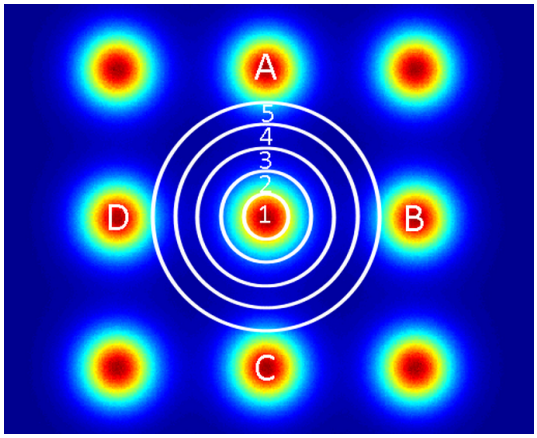


Figure 4.23: Definition of a local neighborhood. Apart from the ROI backscatter features of the central spot, the *NBH* approach also uses the ROI features from four of its neighbors (*A*, *B*, *C* and *D*). (Reprinted from [331], Copyright (2015), with permission from Springer).

the relevance of the angle feature. In contrast, GPs and their more efficient ARD capabilities are expected to be more performant.

4.4.2 Local Neighborhoods

All feature spaces considered so far map from a space of backscatter intensities to the scalar tissue thickness d_s. They do not take into account that tissue thickness varies smoothly across the forehead. Abrupt changes and steep gradients between spatial neighbors in the spot grid are not expected. The current solution ignores this fact and may result in noisy reconstructions when plotting the thickness estimates across the surface.

Therefore, prior knowledge about the local neighborhood of a spot may improve the estimation accuracy. A local neighborhood is defined as illustrated in fig. 4.23. In addition to the five ROI features of the central spot, ROI features from the four closest spatial neighbors (the upper, lower, left and right) are added to the feature space. Depending on which and how many ROIs are selected from them, the feature space dimension may, however, increase dramatically. Assuming that a ROI feature is only added for all four closest neighbors together, the dimension D of the space for $N_{ROI} = 5$ ROIs is given as:

$$D = N_{ROI} + 1 + N_{ROI} \cdot 4 = 5 \cdot N_{ROI} + 1 \tag{4.4}$$

Adding all ROIs from all neighbors and the incident angle yields $D = 26$ dimensions. Investigations of the neighborhood effect based on a case study with subjects S1-S5 have been published in [330, 331].

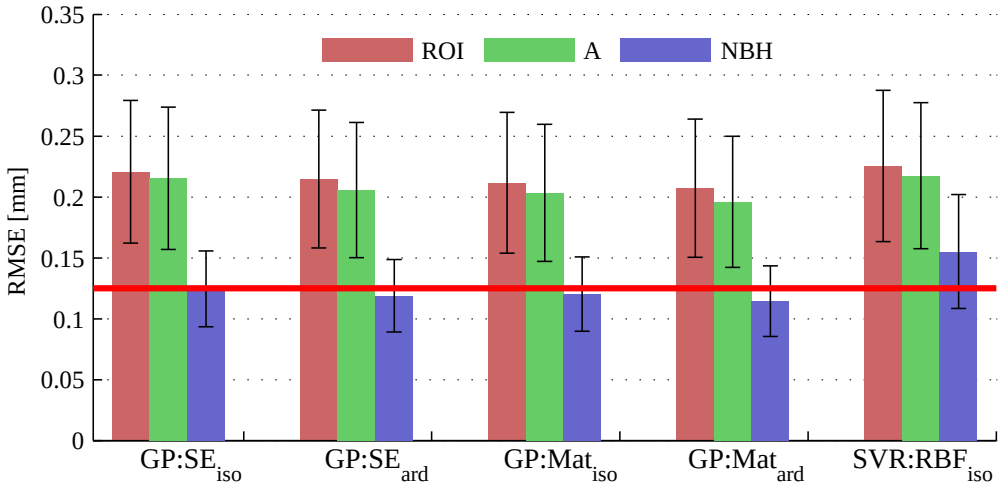


Figure 4.24: Tissue thickness estimation results for CV testing. CV has been performed in a scan-wise manner. Results are shown with mean and STD across all 30 subjects for all considered machine learning approaches (SVR & GPs).

4.4.3 Overall Results

The simulations as well as the considerations and findings in the last two subsections give rise to three different feature spaces. These will be investigated for the entire subject cohort throughout the remaining parts of this chapter:

1. **Space *ROI*:** A feature space which only contains five backscatter features computed from accumulated pixel intensities in five ROIs of the HDR image ($D = 5$).

2. **Space *A*:** A feature space which contains the five backscatter features and the incident angle as an additional sixth feature (corresponds to the *ang* approach, $D = 6$).

3. **Space *NBH*:** Apart from all aforementioned features, this space also contains *all* ROIs from all four neighbors ($D = 26$).

All 30 subjects got scanned in three different head poses à 3 scans each (i.e. $3\times$ 3 = 9 scans in total). These form the basis for the three testing schemes defined in sec. 2.2.3. They will be evaluated in the following: (1) CV (unbiased testing within each scan), (2) AM1 (training on frames of **one** head pose and testing on **one** of the other poses), (3) AM2 (training on frames of **all but one** head pose and testing on **the remaining one**). The RMSE will be given as the average across all possible combinations for a subject. The testing was done for all three feature spaces.

Table 4.6: Tissue thickness prediction RMSE for all considered machine learning approaches (SVR & GPs) under CV testing scheme. Results are listed for all three feature spaces (*ROI*, *A*, *NBH*) with the best as well as best compromise **marked in bold**. The first three rows mark the number of data points N_p in each feature space (averaged across all NIR scans used).

		gender		skin type			
	total	male	female	II	III	IV	V
N_p (*ROI*)	870.6	923.6	824.1	804.5	908.8	888.6	822.6
N_p (*A*)	866.9	920.4	820.1	800.4	905.4	885.0	819.0
N_p (*NBH*)	679.6	749.9	618.1	594.7	728.4	703.8	614.8
$RMSE_{GP:SE_{iso}}$ (*ROI*) [mm]	0.221	0.218	0.223	0.223	0.205	0.224	0.246
$RMSE_{GP:SE_{iso}}$ (*A*) [mm]	0.215	0.201	0.228	0.224	0.196	0.214	0.255
$RMSE_{GP:SE_{iso}}$ (*NBH*) [mm]	0.125	0.120	0.129	0.128	0.114	0.121	0.159
$RMSE_{GP:SE_{ard}}$ (*ROI*) [mm]	0.215	0.213	0.216	0.216	0.201	0.221	0.223
$RMSE_{GP:SE_{ard}}$ (*A*) [mm]	0.206	0.197	0.213	0.213	0.191	0.208	0.219
$RMSE_{GP:SE_{ard}}$ (*NBH*) [mm]	0.119	0.114	0.123	0.122	0.109	0.118	0.147
$RMSE_{GP:Mat_{iso}}$ (*ROI*) [mm]	0.212	0.211	0.212	0.215	0.198	0.217	0.218
$RMSE_{GP:Mat_{iso}}$ (*A*) [mm]	0.203	0.194	0.212	0.210	0.186	0.205	0.228
$RMSE_{GP:Mat_{iso}}$ (*NBH*) [mm]	**0.120**	**0.116**	**0.124**	**0.124**	**0.109**	**0.117**	**0.154**
$RMSE_{GP:Mat_{ard}}$ (*ROI*) [mm]	0.207	0.206	0.208	0.209	0.194	0.213	0.215
$RMSE_{GP:Mat_{ard}}$ (*A*) [mm]	0.196	0.189	0.202	0.202	0.182	0.199	0.211
$RMSE_{GP:Mat_{ard}}$ (*NBH*) [mm]	**0.114**	**0.110**	**0.118**	**0.117**	**0.103**	**0.114**	**0.141**
$RMSE_{SVR:RBF_{iso}}$ (*ROI*) [mm]	0.225	0.224	0.227	0.227	0.211	0.232	0.230
$RMSE_{SVR:RBF_{iso}}$ (*A*) [mm]	0.217	0.206	0.228	0.224	0.201	0.219	0.242
$RMSE_{SVR:RBF_{iso}}$ (*NBH*) [mm]	0.155	0.159	0.152	0.172	0.130	0.160	0.167

Cross-Validation within Single Scans An unbiased estimate for the generalization RMSE was obtained with the CV testing scheme (averaged across scan-wise results). This scheme has been applied to all three scans, for all head poses. Figure 4.24 plots the mean and STD for both, GPs and SVR. The GPs were tested for the isotropic and ARD variants of the SE and Matérn kernel. SVR was only used with the isotropic RBF kernel. The STD includes the variance across samples within single scans, the variance across scans and head poses for one subject, and finally inter-subject variability.
The plot as well as table 4.6 reveal that the prediction accuracy increases the more additional information is added. Adding the incident angle led to moderate improvements.

This was the case for all subjects and was statistically significant ($p<0.005$) for all approaches except for GPs with the isotropic SE kernel ($p=0.04$). Most decrease in RMSE was observed when including local neighborhood information, where the mean error fell below 0.125 mm for all methods (significant for all approaches with $p<0.001$). Nevertheless, table 4.6 shows in its first rows that several data samples were dropped. A sample was discarded if the incident angle could not be computed reliably (e.g. at the grid corners), or some neighbors were missing (e.g. at the grid boundary). The latter effect was considerably larger. In these cases a full description of this sample in the chosen feature space was not possible.
Interestingly, the GPs outperformed SVR. This could be caused by a better optimization procedure, in which hyperparameters are directly optimized by gradient descent within a probabilistic framework. SVR optimizes its parameters one after the other using grid search. This search has only a very limited resolution in the parameter space.
Nevertheless, this finding remains remarkable, since SVR was optimizing the generalization error estimate from CV directly. The results might hence be slightly biased, i.e. too optimistic (cf. sec. 2.2.3). However, very often this bias was less influential than the limited grid resolution as well as the sequential parameter optimization. For instance, the compromising procedure outlined in sec. 2.2.3 achieved an RMSE of 0.262 mm for head pose one of S1 for the *ROI* space. In contrast, nested CV which optimized all parameters simultaneously with a step-wise grid refinement (5 steps) yielded an even better RMSE of 0.257 mm. The computation time, however, amounted to several hours compared to minutes for the other scheme. GPs optimize the NLML which is computed only from the training data.
Finally, SVR is more tuned toward sparsity. The optimization of ε-insensitive loss is not optimal for achieving a minimum RMSE which is rather linked to a quadratic loss function where outliers would more severely affect the result.
Generally, ARD approaches outperformed the isotropic kernel versions. This indicates that features are differently relevant for predicting the thickness. However, the rather small difference in RMSE hardly justifies preferring ARD, due to their high computation time on the other hand. Furthermore, the functional relationship between features and thickness is rather rough than smooth, since the Matérn kernel always outperformed the SE kernel.
No significant correlation of the RMSE was found with respect to age ($p > 0.14$) or Fitzpatrick score ($p > 0.42$) for any feature space. Thus, the results indicate that tissue thickness prediction works equally well for any skin type. The Fitzpatrick score seems not to group according to skin types which are relevant for the application: Since this

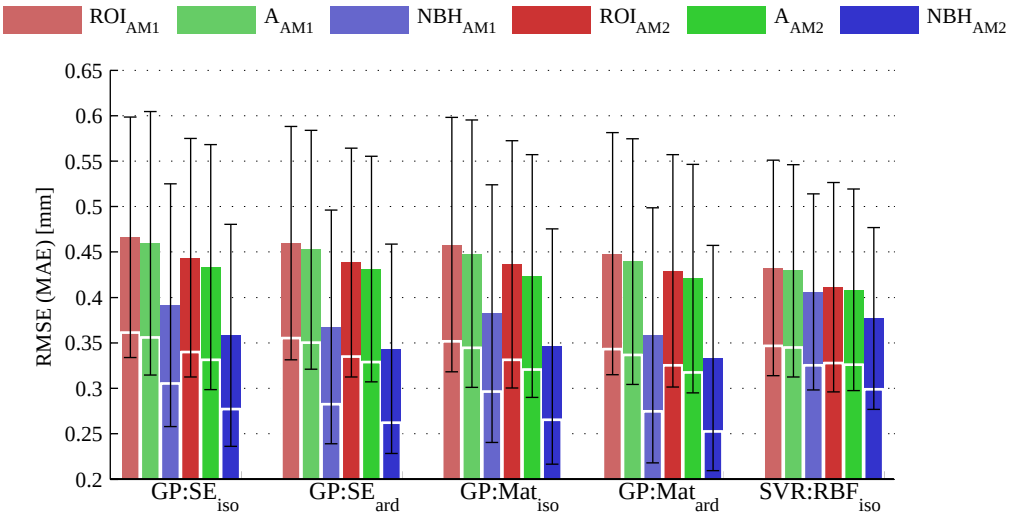


Figure 4.25: Tissue thickness estimation results for AM1 and AM2 testing scheme. Results are shown with RMSE and STD across all 30 subjects for all considered machine learning approaches. The white mark in each bar denotes the MAE error level, which is generally lower than the RMSE.

score and the visual appearance of skin is highly related to the Melanin content (the tan), it seems to hardly affect the tissue interaction of the NIR laser used here. This highly absorbing chromophore primarily affects wavelengths below 600 nm or in the ultraviolet (UV). The 830 nm NIR laser lies within the therapeutic spectral window and is hardly influenced by Melanin.
Moreover, no generally significant difference in RMSE was found between males and females. However, the RMSE tended to be higher for subjects exhibiting more prominent structures due to muscles, wrinkles or just more thicker vessels which locally increased the total tissue thickness. This entailed steeper gradients, i.e. more feature change at a defined change in thickness, and possibly a higher SNR. In particular, young females with brighter skin had slowly varying thickness gradients and hence slightly higher RMSEs (cf. table 4.6, skin type II). This was not the case for all female subjects, and thus not causing any significant gender difference.

Validation across head poses (AM1 & AM2) Training on one head pose and testing on another is a realistic scenario for tracking. The perspective on the scanned surface changes due to motion. Motion between the three recorded head poses ranged between a few millimeters and several centimeters. This led to a varying extent of surface overlap

Table 4.7: Tissue thickness prediction accuracies for the isotropic Matérn GP kernel under AM1 and AM2 testing scheme. The results are listed for all feature spaces. The first three rows for each testing scheme list the statistics for the mutual overlap (ov) of point clouds corresponding to the different head poses (samples from the scan remaining for the *ROI* feature space). For each scheme a test for the entire cloud and one considering only points in the mutual overlap set is presented. The latter results are labeled with the overlap mark ov. The best results are **marked in bold**.

		gender		skin type			
	total	male	female	II	III	IV	V
Across Measurements 1 (AM1)							
Mean Overlap$_{AM1}$ [%] (*ROI*)	86.2	85.1	87.3	90.0	79.6	89.0	84.6
Min Overlap$_{AM1}$ [%] (*ROI*)	25.8	25.8	28.4	64.5	25.8	59.8	67.0
Max Overlap$_{AM1}$ [%] (*ROI*)	100.0	100.0	100.0	100.0	99.4	100.0	99.6
$RMSE_{GP:Mat_{iso}}$ (*ROI*) [mm]	0.463	0.457	0.469	0.444	0.471	0.453	0.525
$RMSE_{GP:Mat_{iso}}$ (*A*) [mm]	0.454	0.443	0.464	0.438	0.460	0.440	0.533
$RMSE_{GP:Mat_{iso}}$ (*NBH*) [mm]	0.392	0.388	0.396	0.354	0.416	0.385	0.437
$RMSE^{ov}_{GP:Mat_{iso}}$ (*ROI*) [mm]	0.411	0.410	0.413	0.391	0.389	0.437	0.400
$RMSE^{ov}_{GP:Mat_{iso}}$ (*A*) [mm]	0.408	0.403	0.413	0.391	0.384	0.433	0.400
$RMSE^{ov}_{GP:Mat_{iso}}$ (*NBH*) [mm]	**0.355**	**0.348**	**0.360**	**0.323**	**0.342**	**0.382**	**0.333**
Across Measurements 2 (AM2)							
Mean Overlap$_{AM2}$ [%] (*ROI*)	93.1	92.1	94.0	96.0	89.1	94.4	92.8
Min Overlap$_{AM2}$ [%] (*ROI*)	32.3	32.3	54.3	85.8	32.3	66.3	79.6
Max Overlap$_{AM2}$ [%] (*ROI*)	100.0	100.0	100.0	100.0	99.4	100.0	99.6
$RMSE_{GP:Mat_{iso}}$ (*ROI*) [mm]	0.440	0.446	0.435	0.395	0.456	0.438	0.503
$RMSE_{GP:Mat_{iso}}$ (*A*) [mm]	0.428	0.426	0.429	0.384	0.445	0.422	0.493
$RMSE_{GP:Mat_{iso}}$ (*NBH*) [mm]	0.354	0.357	0.351	0.313	0.369	0.357	0.379
$RMSE^{ov}_{GP:Mat_{iso}}$ (*ROI*) [mm]	0.411	0.408	0.413	0.387	0.401	0.424	0.426
$RMSE^{ov}_{GP:Mat_{iso}}$ (*A*) [mm]	0.403	0.396	0.410	0.379	0.392	0.419	0.414
$RMSE^{ov}_{GP:Mat_{iso}}$ (*NBH*) [mm]	**0.338**	**0.337**	**0.339**	**0.307**	**0.337**	**0.355**	**0.326**

between the scanned grids. The general problem is illustrated in fig. 4.26A. This overlap was on average 86.2 %±14.5 % and ranged between 25.7 % and 100 % of the data samples remaining for the *ROI* feature space. A sample was defined a member of the overlapping set of samples, if there is a sample from the corresponding second point cloud within at most 2 mm distance. The proportion of overlap was then computed as the number of

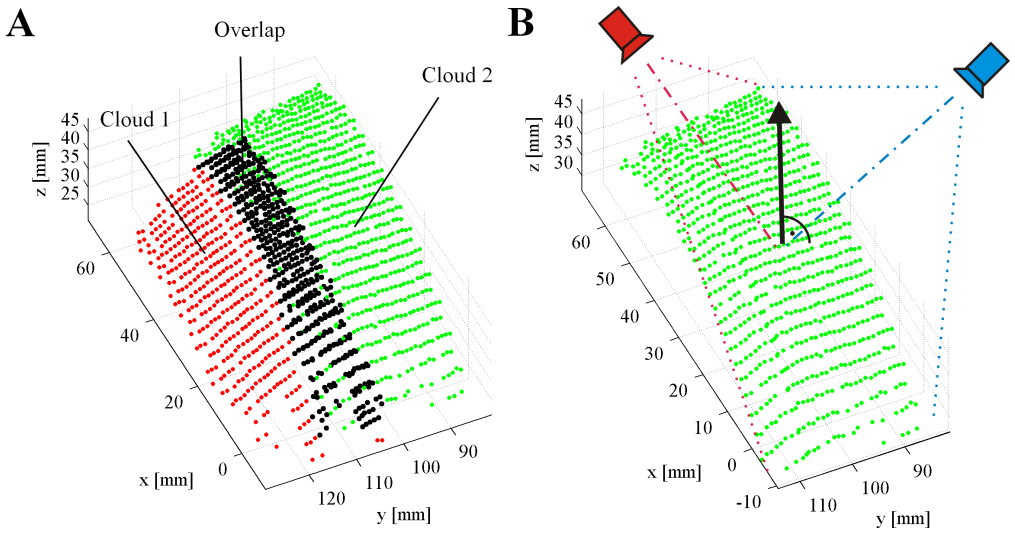


Figure 4.26: Illustration of the main two problems for prediction across different head poses: **A:** Surface overlap. When different parts of the forehead are scanned, the triangulated surfaces may only partially overlap (marked in black). **B:** Scanner-to-forehead perspective. Changing NIR backscatter patterns arise from different perspectives under which a surface is scanned. The influence of the incident angle on the recorded features changes, while the thickness related content remains rather stable.

points in the overlap set divided by the total number of points in the test set. The average overlap between all mutual head pose combinations is listed in table 4.7. As expected, the overlap is generally higher for the AM2 than for the AM1 set. The average point cloud overlap for the other feature spaces was similar to the values shown in table 4.7 (AM1: 86.1 % for *A*, 85.0 % for *NBH*, and AM2: 92.2 % for *A*, 93.2 % for *NBH*).

The RMSEs for both testing schemes (AM1 & AM2), all feature spaces and machine learning techniques are plotted in fig. 4.25 and listed in table 4.7 (for completeness, full detail is given in sec. 7.2). Similar trends to fig. 4.24 can be observed. The *NBH* space yields the best performance, with the Matérn kernels outperforming the others. The improvements observed from the *A* to the *NBH* space were statistically significant for all approaches ($p<0.01$). The *A* space has significantly lower RMSEs than the *ROI* space only for the isotropic GP kernels ($p<0.05$). The ARD kernels may have adapted to well to the training set and hence did not generalize as well as the isotropic kernels.

Note that SVR outperformed the GPs for the feature spaces *ROI* & *A* (cf. sec. 7.2). This may be due to the fact that the regression problem is harder than for CV. Often, however, less knowledge is available about important regions in the feature space. This is caused

by smaller point cloud overlaps. The testing set may request predictions from regions of the feature space which are poorly sampled during training. Due to its objective function with the smoothness term, SVR tends towards less flexible and more biased models. This may be beneficial for extrapolating into poorly sampled regions of the feature space. The chosen GP models tend to adapt to close to the training data in some cases.

Further observations from fig. 4.25 are:

- Larger mean errors in general due to a more complex regression problem
- Larger STD on the mean RMSE due to effects like varying point cloud overlap or forehead-to-scanner positions
- Larger differences between MAE and RMSE indicating high errors from a few outliers, but moderate ones for the majority of samples
- Errors for the AM1 scheme are larger than for the AM2 one

The last point supports the reasoning that larger errors and error variations are mainly due to different forehead-to-scanner positions. While the AM1 scheme used only a scan from a single head pose for training the model, AM2 included scans from an additional head pose. Thus, the likelihood of sufficiently sampling regions of the feature space which might be requested during testing is generally higher. Therefore, the errors are lower. Indeed, the correlation between RMSE and overlap was -0.246 ($p = 0.0009$) across all subjects. On a subject-wise level, six subjects had correlations below -0.85 ($p < 0.05$). These results suggest that the smaller the spatial overlap between two point clouds, the larger is the RMSE. This is in agreement with the arguments above.
The RMSEs for the isotropic Matérn kernel were also computed when discarding all samples from the data that were not in the overlap set (superscript ov in table 4.7). With $p < 0.003$ the RMSE was significantly lower than for cases where testing was allowed to request samples outside the overlap set (on average Δ RMSE < -0.04, minimum Δ RMSE $= -0.228$ for all feature spaces and AM1). The same, yet to a weaker extent, is true for the results from the AM2 scheme. This again suggests that having more head poses in the training set, increases the likelihood of a more representative sampling of the feature space and finally a good performance on the test set. This is supported by the fact, that the more additional information is added to the feature space (e.g. local neighborhoods), the smaller is the RMSE difference between (1) testing on all data and (2) testing on the overlap set only.
Incomplete point cloud overlap is only one factor for worse RMSE prediction errors. A second factor is due to the fact that the laser scanner may look at the target surface from

a different perspective. Thus, the same area may be scanned, but from a different angle. Although the area is contained in the training set, an attempt to make correct predictions may fail, since the features for these thicknesses look different from another perspective. The problem is illustrated in fig. 4.26B. The model lacks training examples for this angle case. On average, the absolute angle difference in the overlap sets was $|\Delta\alpha| = 11.6° \pm 3.6°$ ranging between $5.4°$ and $27.4°$. Overall, the correlation between this angle difference and the RMSE of the AM1 scheme was 0.23 ($p = 0.002$). The larger the absolute angle difference in the overlap set, the larger the RMSE. For four subjects this correlation was higher than 0.80 ($p < 0.05$) on a subject-wise level.
These two relationships and the results in fig. 4.25 suggest, that training data in a couple of representative poses is necessary to achieve errors in the AM1/AM2 set which are comparable to those in the CV set.

Tissue Thickness Reconstruction under CV and AM2 Scheme Finally, fig. 4.27 compares the NIR features for ROI 2, the tissue thickness ground truth from MR and thickness reconstructions from different feature spaces. Comparing the reconstruction from the *A* and *NBH* space, shows that introducing local neighborhood knowledge has a smoothing effect on the reconstructed structures. This is reasonable since the feature space ties local neighbors together. Local neighborhoods have a regularization effect, since they enforce that spatial neighbors should have similar thickness values. That this is not the case for the *A* space, can be seen from its rather noisy reconstruction.
Finally, fig. 4.27F shows a reconstruction under the AM2 scheme with *NBH* feature space (95.0 % surface overlap, $0.76°$ average angle difference). Although the RMSE is with 0.269 mm higher than for the *A* case (CV scheme) in fig. 4.27C (0.230 mm), the reconstruction looks qualitatively smoother. Having a large surface overlap, a small average angle difference, and a small RMSE supports the above argument and shows that decent prediction accuracies can be obtained across scans, if the training data is selected carefully.

Sequential Forward Selection (SFS) Sequential forward selection (SFS) was used to evaluate the importance of certain features for the prediction accuracy. Results are partly published in [331]. SFS is a so-called wrapper method for feature selection. An introduction to feature selection is given in [113]. Starting from an empty set, features are added step-wise. In each step the feature which causes the highest increase in prediction accuracy is added to the set. This increase is determined with the desired machine learning method (here GPs with isotropic Matérn kernel). The usage of a particular

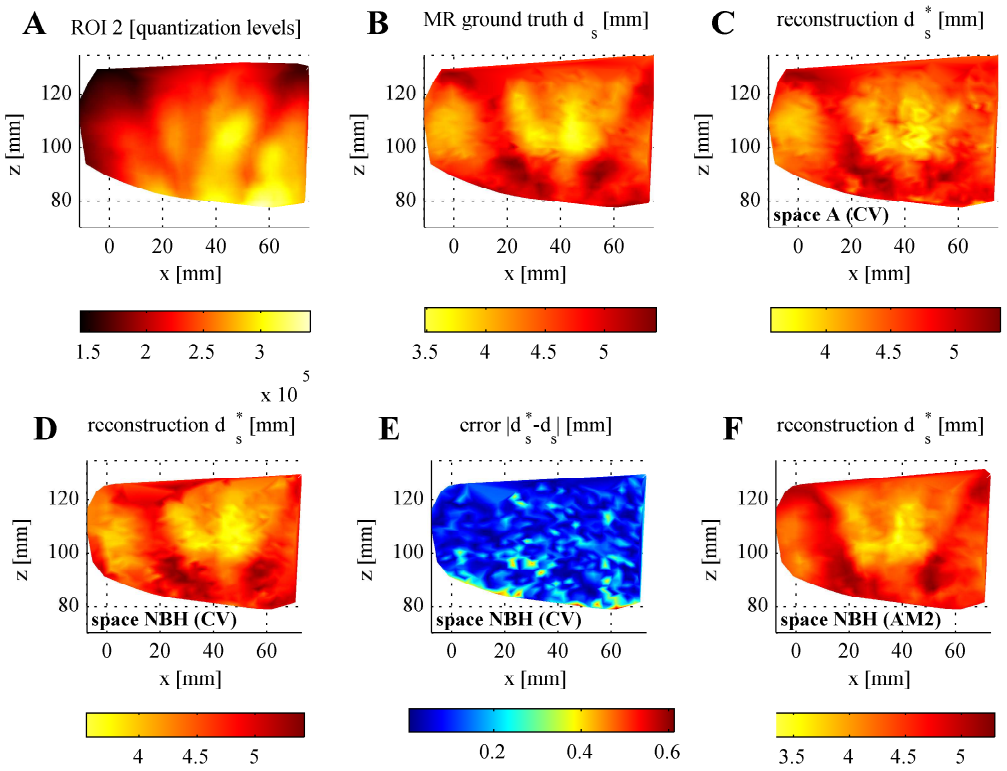


Figure 4.27: Tissue thickness reconstruction for different feature spaces compared to the MR ground truth of S1. **A:** backscatter feature for ROI 2, **B:** MR tissue thickness ground truth, **C:** tissue reconstruction from feature space *A*, **D:** tissue reconstruction from feature space *NBH*, **E:** thickness error for feature space *NBH* (all previous plots under CV scheme), **F:** tissue reconstruction from feature space *NBH* under AM2 scheme (all data from S1, head pose one).

machine learning technique distinguishes wrapper from so-called filter methods which do not rely on a specific learning technique [113].

Two different tests were conducted:

1. **SFS: ROI** - Treats only NIR backscatter features and performed SFS only on the five ROIs. It tests whether certain ROIs are given higher priorities than others.

2. **SFS: NBH** - Assumes a space of all five ROIs and the incident angle and performs from there SFS on the neighborhood features. It tests whether certain ROIs from the neighborhood are given higher priorities than others.

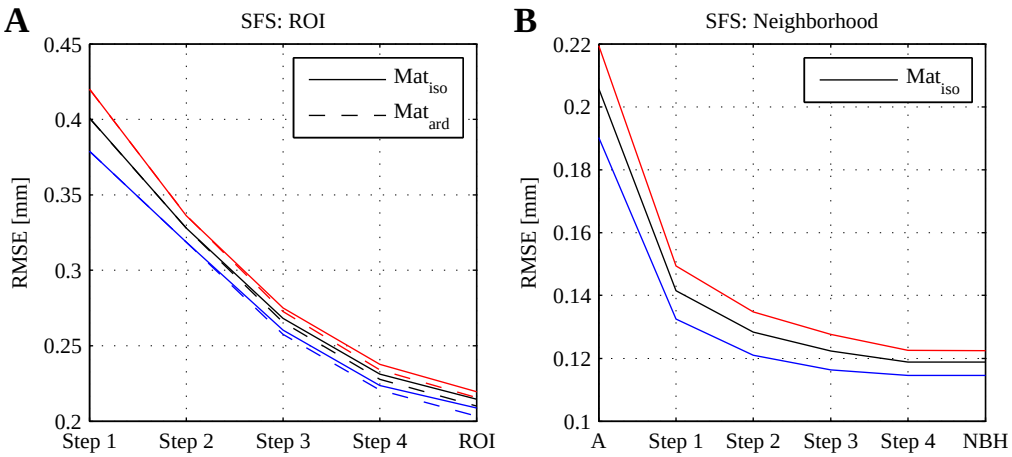


Figure 4.28: Mean RMSE decrease for tissue thickness prediction after step-wise adding backscatter and neighborhood features by sequential forward selection. **Black lines refer to the entire subject cohort**, red lines to female, and blue lines to male subjects. **A:** SFS results for adding NIR backscatter features from the central spot. **B:** SFS results for adding NIR backscatter features from neighboring spots.

Figure 4.28 illustrates the results for both experiments. Generally, the RMSE decreases the more features are added to the system. This happens more steeply for the ROI features and reaches almost saturation when adding too many neighborhood features. For some subjects the RMSE even increases when adding the full neighborhood (SFS step one here corresponds to feature space *A* and step five to space *NBH*). When averaging across many subjects such as in fig. 4.28 the effect is not as prominent. Nonetheless, the error decrease levels off and remains stable before it would probably increase again for more features. This is due to the curse of dimensionality [309]. The neighborhood search already starts with a six dimensional space and adds four dimensions in each step (cf. eq. 4.4). With each new dimension it is harder to have a sampling of the space which is sufficient for the learning algorithm to robustly model a function from the same data. The amount of data required to ensure the same sampling density within the space grows exponentially with its dimension D. Example cases for individual subjects are provided in the appendix (fig. 7.1), which illustrate a differently strong influence of the curse of dimensionality.

Therefore, there is a tradeoff between the increasing dimension and additional information added with a feature. At some point the negative effect of higher dimensions dominates the benefit from new features. When this happens depends on the amount of training data, the complexity of the functional relationship to be learned, and the information con-

Table 4.8: Evaluating the relevance of features by SFS. The upper part lists the mean and STD of the step number in which the backscatter feature, corresponding to ROI i, was added to the feature space. Moreover, the prediction error $\text{RMSE}_{\text{ROI i}}$ for single ROI features is given. The lower part shows the same step information for the ROIs from the neighboring laser spots. In addition, mean and STD of the RMSE gain for adding a neighborhood ROI and the $\text{RMSE}_{\text{NBH i}}$ for a single neighborhood ROI are shown.

	ROI 1	ROI 2	ROI 3	ROI 4	ROI 5
SFS: ROI					
selection step (mean)	2.5	2.7	3.2	3.2	3.3
selection step (STD)	1.3	1.4	1.5	1.6	1.2
$\text{RMSE}_{\text{ROI i}}$ [mm]	0.425	0.419	0.419	0.424	0.431
SFS: NBH					
selection step (mean)	3.0	2.1	2.3	3.1	4.5
selection step (STD)	0.9	1.2	1.5	1.2	0.9
ΔRMSE [mm] (mean)	-0.007	-0.030	-0.037	-0.012	-0.001
ΔRMSE [mm] (STD)	0.007	0.031	0.041	0.019	0.005
$\text{RMSE}_{\text{NBH i}}$ [mm]	0.156	0.144	0.143	0.148	0.162

tent provided by the new feature.
It is important to note that the decrease in RMSE is highest when the first neighborhood feature is added. This indicates that it is more important whether this novel information is added to the system at all, rather than how much of it is provided. Generally, the graphs look similar for female and male subjects, whereas the RMSE for females is slightly higher. Further on, ARD behaves similarly to the isotropic case, but tends to achieve increasingly better results for higher dimensions. This was expected from its algorithmic nature: The more features exist in total, the more likely is it that a weighting for relevance reveals an impact on the prediction errors. Nevertheless, each feature dimension adds new parameters and therefore increases the computational load and complexity. More degrees of freedom entail a harder optimization problem. As stated before, the low decrease of the prediction error does not justify all these difficulties. This is also why the neighborhood SFS has not been computed for the ARD.
Finally, SFS on the ROI features showed that all ROIs have similar accuracies in the first step (cf. table 4.8). This is reasonable since they are highly correlated as it was shown in fig. 4.18. Therefore, table 4.8 shows no clear pattern across subjects concerning the order in which the ROI are added. The high STD confirms this. ROI 1 has a slight tendency to

be selected earlier. This may be caused by the fact, that it is less correlated to the other ROIs.
In contrast, ROIs 2 and 3 are clearly preferred in the neighborhood SFS. They provide the highest gain in accuracy and the lowest rank for the step in which they were selected on average (cf. table 4.8). This is strictly in agreement with previous findings on real data as well as simulations. These two ROIs provide the best SNR of all ROIs. At the same time they achieve higher proportions of photons from deeper tissue layers and also have a higher magnitude in their information signal with respect to the acquisition noise. They are less prone to external disturbances. Thus, spatially speaking, they give rise to the clearest and smoothest spatial pattern. They are therefore preferred as neighborhood features.

4.5 Enhanced Learning using Sparse Approximation Techniques

The results presented in the last section suggest that the required machine learning technique ideally needs to handle data from many different scanning perspectives. This entails an increase in N, i.e. the total number of samples in the training set. GPs provide a couple of very promising properties for the regression problem at hand, but as discussed in sec. 2.2.2.7, suffer from an increasing computational complexity (by $\mathcal{O}(N^3)$) if the training data exceeds a few thousand samples.
The following paragraphs evaluate the SoD and FITC sparse approximation techniques to tackle this problem. The results have been published in [332]. The evaluation will be done in terms of a case study with subjects S1 to S5. From each subject two subsequent scans from the first head pose are taken. The number of data samples amounts to 1447 (S1), 1764 (S2), 1612 (S3), 1790 (S4), and 1557 (S5). For selecting the inducing variables $\boldsymbol{u}$ from the training set, SoD and FITC methods will be evaluated with random choice (subscript $_R$) and k-means clustering (subscript $_k$). The tested subset sizes were $M = \{25\,\%, 33\,\%, 50\,\%, 66\,\%, 75\,\%, 100\,\%\} \cdot N$ inducing points. All timings were measured on an Intel® Core™ i7-4770S CPU @ 3.1 GHz, 32 GB RAM. Implementations were done in MATLAB® (Mathworks, Inc. [188]).

Figure 4.29A shows all results for S1. From the top plot it can be seen that FITC outperforms SoD when restricting the methods to the same M. FITC maintains reasonably low errors even for only 50 % of the data. Both converge to the same RMSE for $N = M$, which is the full GP. For $N < M$ the error worsens faster for SoD than for

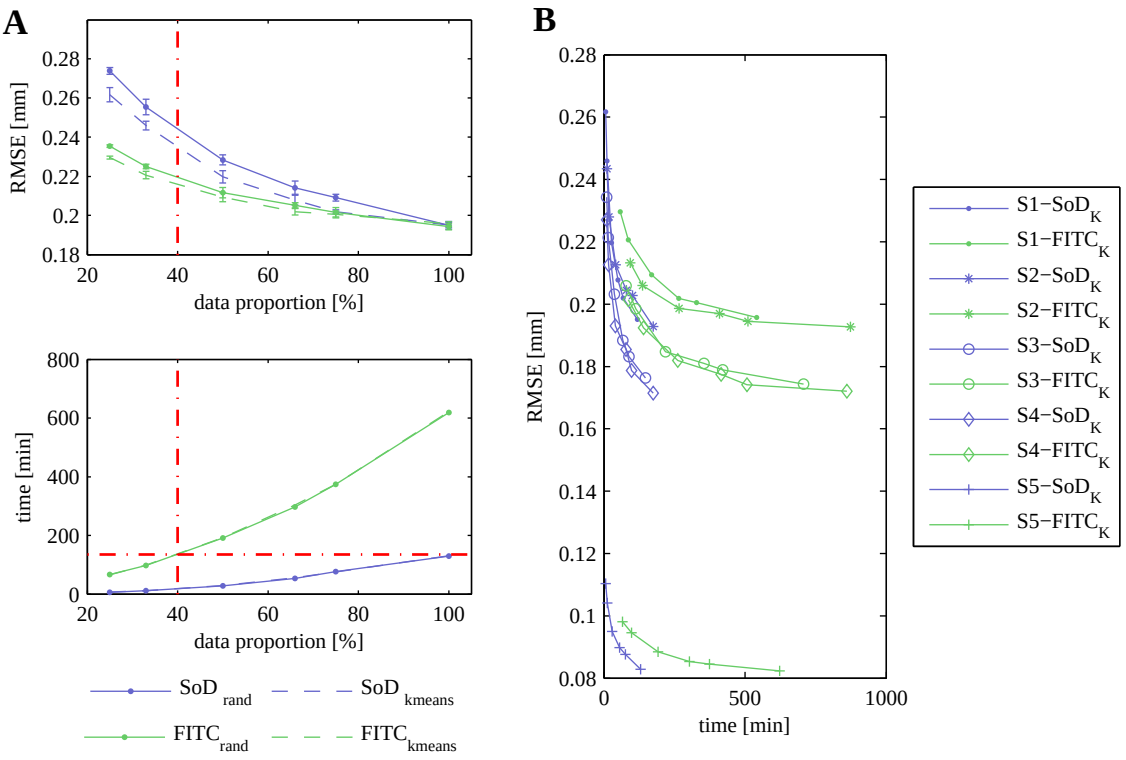


Figure 4.29: Tradeoff between computation time and prediction error for different sparse approximation techniques. **A:** Results for S1: RMSE (top) and computation time (bottom) are shown for all four methods. The horizontal and vertical lines denote the time for the full GP and the FITC result requiring the same time. **B:** Efficiency plot comparing all five subjects. RMSE versus time is compared for SoD_K (blue) and FITC_K (green). Subject associations are indicated by the same marker. Each mark on a graph corresponds to one specific setting for M. (© 2015 IEEE. Reprinted, with permission from [336]).

FITC. Similar behavior was observed for the other subjects (plots not shown). This is not surprising, since SoD simply throws data away, whereas FITC keeps all the information and approximates the GP model assumptions. It just assumes the training samples to be statistically independent, given the inducing points $\boldsymbol{u}$. Nevertheless, there is still an RMSE increase – yet less steep than for SoD – since this assumption is not fully true.
The full GP model yielded an RMSE of 0.195 mm in 118 min. The FITC equivalent for the full model at $M = N$ requires longer computation times. It involves a higher number of operations, e.g. matrix multiplications, which entail an increased overhead. Within the same time window, FITC is able to handle at most 40 % of the data. In this setting the RMSE reaches approximately 0.215 mm or more (i.e. worsening by $> 10\,\%$, setting

Table 4.9: Computation time and RMSE for different sparse GP approximations. Results are shown for the first five subjects. The GP was trained with either 100 % or 33 % of the available data. Time is given as a proportion of the time required by the full GP. (© 2015 IEEE. Reprinted, with permission from [336]).

	S1	S2	S3	S4	S5	
Full GP	t [min]	118	175	148	175	129
	RMSE [mm]	0.195	0.193	0.176	0.172	0.084
$\text{SoD}^{b}_{\text{R33}}$	t [%]	9.3	9.1	9.7	9.0	8.9
	RMSE [mm]	0.256	0.233	0.2257	0.223	0.108
$\text{SoD}^{b}_{\text{K33}}$	t [%]	9.4	10.1	10.6	9.3	9.1
	RMSE [mm]	0.246	0.228	0.221	0.213	0.104
$\text{FITC}^{b}_{\text{R33}}$	t [%]	71.4	78.5	78.1	79.9	75.6
	RMSE [mm]	0.225	0.211	0.202	0.200	0.095
$\text{FITC}^{b}_{\text{K33}}$	t [%]	73.1	77.5	75.5	78.9	76.3
	RMSE [mm]	0.221	0.206	0.199	0.192	0.095

b Values given for 33 % of the data.

marked with dash-dotted red lines in fig. 4.29A). If the computation time for the full GP model needs to be reduced, the horizontal line in fig. 4.29A (bottom) will be shifted downward. Now, to prefer FITC over SoD, the RMSE for FITC at e.g. 30-40 % would need to be lower than that of SoD at 80-90 %. This was not the case for any subject given the characteristics of the experimental data treated in this work.

An example case is listed in table 4.9 for 33 % of the data. The values indicate that FITC_K is capable of maintaining an RMSE which comes closest to the full model. The worsening amounts to 13.3 % (S1), 6.7 % (S2), 13.1 % (S3), 11.7 % (S4), and 13.1 % (S5) of the original RMSE only. However, at the same time the savings in computational cost are only about 30 %, while SoD approaches are by factor 7-9 faster for this M. The computation time for testing on unseen data (excluding training and optimization) followed a similar qualitative trend. The full GP, for instance, required $1.32 \pm 0.02\,s$ for predicting thickness values of the test set, while FITC (with 50 % of the data being inducing variables) required $1.76 \pm 0.04\,s$.

The latter argument concerns efficiency, i.e. the accuracy-vs.-time tradeoff. A more detailed statement can be made from fig. 4.29B. The plot illustrates this tradeoff for all subjects and k-means-based techniques. Efficient settings tend towards the lower left corner of the plot, where neither computation time, nor the error are reaching extreme

values. Irrespective of the M chosen (points on the graph), SoD approaches can be considered more efficient on subject level, and for certain cases even when comparing across subjects. When reducing the data to the most relevant 66 % of all samples, the RMSE drops only by 6.5 % (S1), 6.1 % (S2), 6.9 % (S3), 8.2 % (S4), and 8.4 % (S5). As before, the RMSE for S5 (female) was lower than for the others. For her, the tissue thickness variance across the forehead was smaller than for subjects. In an absolute sense, using FITC with small subset sizes (33-50 %) was more justifiable for her than for the other subjects.

However, generally speaking, the computational effort of FITC does not justify favoring it due to the slower increase in RMSE. These results are in line with similar findings of Chalupka et al. on completely different data [51]. Confirming theory, the results also show that the computational complexity of FITC has a scaling which is worse than that of SoD ($\mathcal{O}(NM^2)$ vs. $\mathcal{O}(M^3)$ for $M < N$). Therefore, the difference in computation time between both approaches grows with $\frac{M}{N}$ the more data is taken. Likewise, the argument applies to the storage demands.

In particular for FITC, the RMSE does not increase rapidly enough for our data when dropping parts of it. This indicates that similar information is shared between training samples. Spatial proximity of the laser spot images on the forehead may be one explanation. Each spot has eight neighbors (cf. fig. 4.23). Their feature vectors share information about their corresponding tissue thickness values, because spatial frequencies for thickness variations across the forehead are rather low. The thickness varies rather slowly compared to the sampling step width of the grid (a few millimeters). Thus, discarding a few samples will still keep related information. This is in agreement with the findings about local neighborhoods in sec. 4.4.3. Possibly noisy data can be recovered from the neighbors, since tissue thickness is required to be similar among them.

Finally, an informed choice of the inducing variables via k-means clustering was found of advantage. The computational overhead with respect to a random selection is negligible. Clustering induces a more homogenous coverage of the feature space. K-means was used such that, if a cluster would loose all its members, a new cluster is created with the data sample farthest away from the existing centroids. It therefore avoids cases where all inducing variables are massed in one place. More sophisticated unsupervised learning methods such as farthest point clustering [51] may extend the aforementioned principle. Other promising approaches may look at similarities in the kernel space, since this is the space in which the actual GP will model dependencies. A greedy algorithm was proposed by Smola et al. based on this idea [270]. SoR methods work along similar lines.

4.6 Conclusions

This chapter treated four main parts. First, it was described how findings from the simulation studies were implemented in an experimental optical setup. Special focus was directed to the generation of 3D surface information via laser triangulation, and to how a tissue thickness ground truth is obtained from high resolution MR images. To test the concept of tissue thickness estimation, a volunteer study with 30 subjects was prepared and conducted. Second, general aspects of light-tissue interactions were evaluated on these data and set into the simulated context discussed earlier. Third, the prediction accuracy for tissue thickness from different feature spaces was tested. In particular, disturbances and the incorporation of prior knowledge was addressed. Fourth, sparse approximations for GPs were investigated to tackle the increasing computational complexity – a technical limitation of GPs for future applications.
Thus, this chapter answered RQ 2. Conclusions on the four main challenges raised by RQ 2 will be given in the following.

- RQ 2.1: **How are the simulation results reflected in real data?**

  The negative correlation between NIR backscatter and tissue thickness was confirmed. Moreover, influences of changing incident angles were similarly found to superpose the information signal. Particularly for smaller incident angles subcutaneous structures were qualitatively visible. This is in agreement with earlier simulations. For increasing angles the disturbance effects then dominate the signal variations. Since separating both effects is challenging, only a limited analysis of the information signal was possible for some subjects. Due to measurement noise and the heterogeneity of the human skin (e.g. due to freckles, blood vessels, or moles) the feature spaces exhibited much nosier and more complex behavior compared to MCML simulations. This again indicates that simulations will simplify the real world, especially when a complex system such as human skin is concerned and the information is carried in only weak effects. Nonetheless, the average Caucasian model was vital to derive qualitative trends to be expected from the data, which led the design of an experimental setup to optimal acquisition conditions. The mentioned discrepancy between the simplified model and the real world is also apparent from the prediction accuracies. They were indeed substantially higher for the simulated case.
  Finally, as speculated based on the simulations, ROIs 2 and 3 were found to

provide information most robustly (about 2.5 mm to 7.6 mm from the spot center). Patterns were most prominently visible in them, wherefore they were preferably chosen during SFS of neighborhood features. Nevertheless, high linear correlation was found among the NIR features. This fact could render methods for dimensionality reduction by e.g. matrix transformation interesting for future research. Particular focus should be directed to this extension when also including many of the NIR features from spots in the local neighborhood.

- RQ 2.2: **Which statistical learning approach is most suitable for retrieving information about tissue thickness from optical features?**

  While SVR again achieved reasonable results, GPs were found to be more promising for future applications. Hyperparameter optimization can be efficiently carried out by minimizing the NLML instead of time consuming grid search. This also facilitates explicit relevance weighting for individual features. This is of particular interest when incorporating prior knowledge in terms of additional features. Their characteristics and importance may deviate from the highly correlated optical features. Best accuracies were achieved with the ARD Matérn kernel, which has a higher functional complexity than exponential kernels and can model rougher functions. Nevertheless, the isotropic Matérn kernel along with a GP constitutes the recommended choice. The prediction errors of on average 0.21 mm are only slightly higher than for its ARD variant. However, particularly for model building, computation time is a lot lower.
  A technical challenge for GPs remains the handling of larger data sets with a few thousand samples. A simple pre-selection of a subset of data (SoD$_K$) with k-means clustering was found to be the most efficient solution to tackle this problem. It outperformed more sophisticated methods such as FITC and still achieves reasonable results when cutting the data from two scans in half. For very large data sets, however, the time for training a model may still become prohibitive.

- RQ 2.3: **Which disturbances have an influence on the estimation accuracy and how can they be handled?**

  The main disturbing influence is given by changing incident angles of the laser beam. All compensation methods (relying on pre-processing the raw HDR images) did not yield consistent and therefore satisfying results. This includes the

approach termed *fix* which was also tested on simulated data earlier. While regressing the angle effect out from the data improved the qualitative appearance of the subcutaneous pattern, it did not reveal promising effects on the prediction accuracy. The recommended approach adds the incident angle as an additional sixth feature to the set of ROI values. Further improvements on the prediction error were obtained when weighting the angle according to its relevance in the current data set.

Further on, skin heterogeneity across the forehead surface was found to influence the appearance of the feature space. Measurement noise and variability of the tissue thickness ground truth (on average $< 0.2\,\mathrm{mm}$) add to this rather complex behavior of the data in the NIR feature space. Considerable improvements of the prediction error were achieved by including backscatter information from the local neighborhood. This means one spot will be also labeled with information of the backscatter behavior from its neighboring spots. This has a regularizing effect on the regression problem and helps tackling the aforementioned disturbances. Linking a spot to its spatial neighbors implicitly enforces the target label to be similar as well. This rules out effects which are not conform with this prior knowledge. The smoothing effect on the reconstructed thickness pattern is expected to have a positive effect on the surface registration performance.

Finally, investigations of prediction accuracies across different head poses revealed that a comprehensive training data set is required to sufficiently cover the feature space. Otherwise, incomplete overlap between point clouds of the training data and the requested test data would leave certain regions of the feature space for extrapolation. This negatively affects the prediction accuracy. Moreover, training scans ideally need to be recorded from different perspectives of the laser scanner with respect to the forehead. Angle differences between training and testing – even when the same surface patch is scanned – have been found to be positively correlated with the prediction error.

The latter effect may require larger sets of training data. This makes sparse approximation techniques for GPs relevant for future research. Alternatively, other regression techniques such as classical neural networks, deep belief networks or autoencoder networks could be taken into consideration. Special concern should be directed to this field when aiming at universal regression models which are required to predict tissue thickness across subjects or different skin types.

- RQ 2.4: **Are there indications of gender, age or skin type affecting the learning outcome?**

  The data set of 30 volunteers did not show any evidences for a correlation with the prediction error of the tissue thickness. A gap between the RMSE of male and female subjects was consistently identified, but no significance at $p = 0.05$ was found. The gap may also arise from the fact that most female subjects were young and had a tendency towards slowly varying and low thickness gradients across the forehead. Therefore, the informative part of the signal was weaker in magnitude and harder to learn from data corrupted with the mentioned external disturbances. Many male and elderly subjects exhibited rather prominent structures which stood out more clearly. Final validation of this reasoning can only be obtained based on a more comprehensive clinical study. This especially applies to statements about dependencies with respect to the age. The presented subject cohort mainly focused on subjects in their twenties or beginning of their thirties. Furthermore, it should be considered that skin age and chronological age of a person may not coincide [37, 38].
  No correlation was also found with the skin type as defined by the Fitzpatrick scale. Thus, the concept seems to work equally fine for bright and dark skin colors. This is reasonable, since this skin typing mainly relies on the Melanin content and UV sensitivity of skin. Therefore, this kind of grouping may be rather irrelevant for light-tissue interactions in the NIR range. The presented results cannot exclude that there is a relevant typing for the proposed concept. Further investigation is required based on a more comprehensive study combined with a careful characterization of the skin.
  The results of this volunteer study imply two things for the clinical application. First, the results suggest that the proposed procedure of generating patterns on a scanned forehead surface works for individuals irrespective of skin type, age or gender. No volunteer exhibited a behavior which is prohibitive for this approach. Therefore, the preliminary outcome does not exclude any specific group of patients. This is of special importance for countries of high ethnical diversity such as the United States. Second, the lack of correlation between prediction error and Fitzpatrick score does not preclude the possibility that there are groups of patients, who could share similar models after all. Larger clinical validation will have to provide more detailed insight into this. Grouping and model usage across subjects has a great potential for facilitating the work load within the clinical work flow.

5 Tissue-Supported Head-Tracking – A Proof of Concept

All previous chapters were mainly concerned with the estimation of the tissue thickness measure from NIR backscatter. The solution to this problem ensures high quality of the structural information which is available for tracking. High structural quality is the crucial foundation for higher tracking accuracy. This chapter will take the next step to link back to the original motivation of this work. It will exploit the findings for predicting tissue thickness and explore the benefits for tracking. It treats the central question of whether structural information leads to a better registration of two surfaces for the application at hand.
Note that these considerations only aim at a general proof of concept. Standard ICP without any sophisticated variants will be used. The development and optimization of a dedicated tracking algorithm is beyond the scope of this work. With simple means it will be evaluated whether the original claim of this work is true: that surface registration is more accurate and robust against outliers when additional information in the form of cutaneous structures is used. Parts of this chapter have been published in [335].
Therefore, sec. 5.1 will first define the experimental setting, the testing procedure as well as the comparative quality measures. Next, sec. 5.2 will first of all consider the optimal case. This optimal case corresponds to a perfect prediction of tissue thickness from NIR backscatter. Therefore, pure surface matching will be compared with structure-supported registration, where the structures are taken from the ideal ground truth. The second part of this section will then comment on how the tissue thickness prediction error will influence these optimal results, i.e. how it will translate to tracking accuracy. Finally, sec. 5.3 will provide conclusions based on these findings.

5.1 Point Cloud Registration and Evaluation Procedure

A core question of this work is, whether the use of tissue structures yields any improvement for surface registration. To investigate this issue, a ground truth for motion between

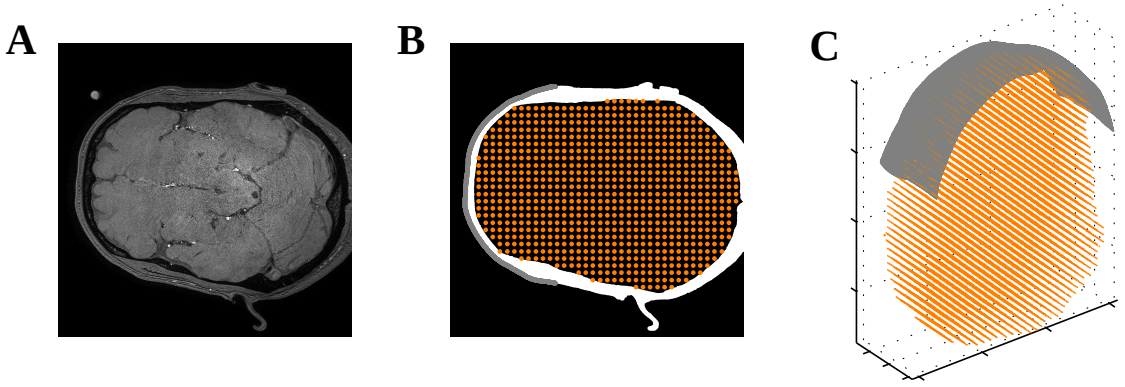


Figure 5.1: Determining possible tumor targets: segmentation of the brain. **A:** High resolution MR scans are used to get possible CNS tumor sites. **B:** Therefore, the space withing the cranial bone is segmented (orange). The skin-air boundary (gray) on the forehead is used as a reference surface for registration. **C:** The segmentation results for all slices show the reference for surface registration (gray) and possible tumor targets (orange) which need to be tracked.

two poses is required. The transformation matrix output from surface registration can then be compared against this ground truth. There are two possibilities. Either known motion is applied to an object experimentally, or motion is simulated offline based on a single pair of scans. The first option entails two substantial drawbacks.
First, motion cannot be applied to an object without any errors. The ground truth itself will comprise errors, since the desired motion and the one actually applied will differ from each other. Hardware uncertainties in robot kinematics constitute only one example. This is unfortunate, since the registration errors under consideration are expected to be rather small. They may be easily corrupted by errors of the ground truth.
Second, the extent of possible motion that can be tested experimentally is limited. A small number of trails will have difficulties to provide generalizing conclusions.
Therefore, all subsequent tests will make use of simulated motion. Here, the ground truth is exactly known and an exhaustive number of random transformations can sufficiently cover the motion space. Details will be described in the following.

Tumor Target Sites The registration error can be computed for the forehead surface, where the registration takes place. However, for clinical applications the localization of possible targets is more relevant. These targets are typically tumor sites within the cranium. For surface registration, the tracking errors of these targets can be higher than the one on the surface. Only considering surface errors would hence underestimate the

matching error. To compute so-called target registration errors (TREs)[1], possible tumor sites within the cranium have been defined as shown in fig. 5.1: A volume of possible targets within the cranium has been segmented using morphological operators followed by a region growing algorithm. The area originally covered by the brain has then been sampled with a 3D grid (5 mm spacing) of points (cf. fig. 5.1B). For each subject, this results in a known spatial relationship between the skin-air boundary at the forehead (reference surface to which all NIR scans will be registered) and a volume of possible tumor targets (cf. fig. 5.1C). For each of these targets the TRE can be computed.

Simulated Motion and Testing Scheme Section 4.2.3 described how a NIR scan is registered to the MR ground truth. This is achieved by an ICP-refined, marker-based matching. This initial registration is assumed to be exact and will be used for (1) learning a supervised model between NIR features and tissue thickness, and (2) the registration ground truth for any simulated motion. The setting is illustrated in fig. 5.2A.
Under the assumption that the ground truth registration is exact, it follows that the targets (segmented from the MR scan) are also the target locations with respect to the NIR patch. Motion is therefore simulated by applying random transformations $\mathcal{T}_{rnd}$ to the registered NIR patch as well as the targets (fig. 5.2B). This simulates subject head motion away from this ground truth position. The transformation matrix is generated from the three translational and three rotational degrees of freedom. These are randomly sampled from a uniform distribution.

$$t_{x,y,z} \sim \mathcal{U}[-20\,\text{mm}, 20\,\text{mm}] \tag{5.1}$$

$$r_{x,y,z} \sim \mathcal{U}[-20^\circ, 20^\circ] \tag{5.2}$$

Now, the point-to-plane ICP algorithm with (W) and without (WO) tissue thickness support is required to re-register the NIR to the MR reference surface. This means an estimate $\mathcal{T}_{ICP-W}/\mathcal{T}_{ICP-WO}$ for $\mathcal{T}_{rnd}^{-1}$ has to be computed. The tissue supported ICP has been used as described in sec. 2.3.3. In total, $N_{rnd} = 5,000$ random transformations are applied and re-registered with both approaches. The surface registration error $\text{RMSE}_{ICP-W/WO}$ is computed as the mean point-to-point distance between the original NIR patch $\left\{\boldsymbol{p}_j^{gt}\right\}_{j=1\ldots N_p}$ (red, fig. 5.2) and the re-registered version of the randomly transformed NIR patch $\left\{\boldsymbol{p}_j^{re-reg}\right\}_{j=1\ldots N_p}$ (cyan, fig. 5.2). The errors are also averaged across

[1] Definition in the glossary according to [115]. Note that the target region is not required to be involved in the registration process. For surface tracking, the transformation matrix is always computed without any knowledge about internal targets.

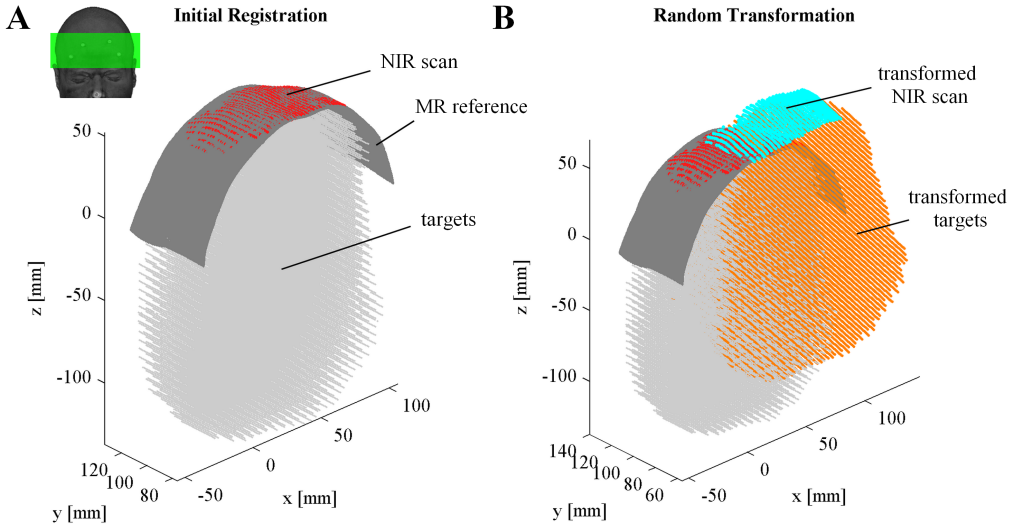


Figure 5.2: Testing scheme for comparing surface registration with and without tissue support. **A:** Initial registration between MR surface reference (dark gray) and an optical NIR scan (red). This registration is achieved using the aforementioned ICP-refined, marker-based matching. This matching establishes a link between the NIR scan red) and the targets from fig. 5.1 (light gray). **B:** Starting from this registration, random subject motion is simulated. Therefore, random transformations are applied to the NIR scan (red to cyan) and the targets (light gray to orange). The transformed targets yield an accurate estimate of how the rest of the cranium will have moved with the NIR scan. The algorithm under test needs to re-register the cyan NIR patch to the dark gray MR reference. The error is computed as the mean point-to-point distance between the original red, and the re-registered cyan NIR patch.

all random transformations.

$$RMSE_{ICP-W/WO} = \sum_{i=1}^{N_{rnd}} \sum_{j=1}^{N_p} \left\| \boldsymbol{p}_j^{gt} - \boldsymbol{p}_{ij}^{re-reg} \right\|_2 \tag{5.3}$$

The TRE is computed point-wise for all target locations $\left\{ \boldsymbol{\tau}_j^{re-reg} \right\}_{j=1...N_{targets}}$. The average is taken across all N_{rnd} random transformations.

$$TRE_{ICP-W/WO}^{j} = \sum_{i=1}^{N_{rnd}} \left\| \boldsymbol{\tau}_j^{gt} - \boldsymbol{\tau}_{ij}^{re-reg} \right\|_2 \tag{5.4}$$

Note that there are random and systematic error sources. Random error sources involve the random jitter around a 3D point. They describe the uncertainty of how well a single point on a surface can be identified. The correct location is typically blurred by a random noise distribution. This identification error links to the total surface registration error as

well as to the TRE. However, there is no statistical relationship between surface registration error and TRE directly as discussed by Fitzpatrick and colleagues [91, 92]. This means, the random aspect of the surface registration error cannot be used to infer the uncertainty on the TRE.
The testing scheme described above, excludes these random identification errors from investigation. The simulated motion always uses the same NIR patch. There is no random variation from registration to registration in terms of 3D point identification. It is always the same sample from the uncertainty distribution. The probability distribution in terms of [91] therefore corresponds to a Dirac pulse – all the probability mass is gathered at one instance.
The testing scheme rather investigates systematic registration errors. These can be described as centers of accumulation. This refers to areas where misregistrations tend to end up due to prevailing circumstances rather than random effects. For these, the statistical dependency statement between surface registration error and TRE does not apply [91]. Both errors are directly linked by the connecting lever between the two locations under consideration. The errors are not subject to any random acquisition effects, but rather due to local registration minima, geometric deformations, ambiguities and spatial similarities across the surface. These are the challenges which should be tackled by exploiting supportive cutaneous structures. Therefore, the tests will focus on how robust the registration algorithms can handle them.

5.2 Tissue-Supported Registration

Table 5.1 lists the test results which were obtained under the scheme described above. The evaluation involves the entire subject cohort. The rows labeled with MR denote tests that used the MR ground truth as supportive structures on the surface. Comparing registration with (W) and without (WO) these structures, reveals that additional structural information had a considerably positive effect. On average, conventional surface registration was outperformed by more than a factor of 29.5. The 90 % error bound I90 was almost twice the average registration error for the conventional method. For tissue supported registration, this bound was quite close to the average value. This implies that the latter method is more robust against outliers. It is less prone to getting stuck in local registration minima on the surface.
Both, the spatial coordinates as well as the tissue thickness are given in millimeters. Nevertheless, the spread of the data in each of the four dimensions is different. Point-to-point correspondences between two surfaces are identified via Euclidean distances in this 4D

Table 5.1: Comparison of ICP registration errors with (W) or without (WO) tissue thickness support. RMSE and I90 are given as an average across all 30 subjects and also distinguished according to gender and skin type. Tissue support was generated from (1) the MR tissue ground truth, (2) predictions from feature space *A*, and (3) predictions from feature space *NBH*.

		gender		skin type			
	total	male	female	II	III	IV	V
RMSE_{ICP-WO} (MR) [mm]	2.889	0.978	4.561	3.592	6.084	0.932	1.444
I90_{ICP-WO} (MR) [mm]	4.551	1.482	7.236	4.907	9.955	1.626	2.100
RMSE_{ICP-WO} (A) [mm]	2.884	0.950	4.577	3.768	5.982	0.970	1.153
I90_{ICP-WO} (A) [mm]	4.556	1.378	7.337	4.875	9.967	1.645	2.105
RMSE_{ICP-WO} (NBH) [mm]	3.380	1.200	5.287	4.435	6.209	1.609	1.401
I90_{ICP-WO} (NBH) [mm]	4.437	1.521	6.988	5.335	8.119	2.352	1.853
RMSE_{ICP-W} (MR) [mm]	0.098	0.075	0.118	0.154	0.162	0.050	0.022
I90_{ICP-W} (MR) [mm]	0.101	0.085	0.114	0.245	0.143	0.025	0.027
RMSE_{ICP-W} (A) [mm]	0.510	0.429	0.582	0.673	0.363	0.491	0.661
I90_{ICP-W} (A) [mm]	0.577	0.521	0.625	0.678	0.486	0.564	0.670
RMSE_{ICP-W} (NBH) [mm]	0.478	0.431	0.520	0.559	0.465	0.402	0.683
I90_{ICP-W} (NBH) [mm]	0.521	0.487	0.551	0.511	0.563	0.455	0.718

space. Thus, dimensions with higher data variance have a higher impact on these point-to-point assignments. The median value of the tissue thickness STD across the forehead scan was 0.465 mm (minimum 0.264 mm, maximum 0.774 mm). To analyze the spread within the spatial coordinates, PCA was used, since the subject pose under the scanner varied across subjects and measurements. The median STDs along the three principal components of the scanned patches were [3.4 mm, 7.69 mm, 21.9 mm] (minimum [2.4 mm, 6.0 mm, 17.1 mm], maximum [4.6 mm, 13.9 mm, 23.9 mm]). For finding point-to-point correspondences, the tissue thickness values were given a higher relevance. Scaling it generally by $f_{ac} = 15$ lifts its magnitude approximately to the order of the second spatial principal component.

It is interesting to note that pure surface registration was harder for female than for male subjects. Average errors (WO) for females were found to be dramatically higher than for men. Reasons are related to the smaller forehead area for female volunteers. The likelihood is higher for discarding triangulated points due to exceeding the hairline or sampling the eye brows or laser goggles. Available forehead regions for males were ob-

served to be larger on average. They also exhibited more distinct surface variations, while scanned surfaces from females tended to have rather bowl-shaped characteristics. The risk for local registration minima was thus higher. No significant correlation ($p > 0.1$) between the registration error and the skin type score have been found.

Registration Errors versus Prediction Errors Table 5.1 further shows the impact of tissue thickness prediction on the re-registration results. Experiments analogous to the ones above were conducted after replacing the tissue ground truth by predictions from backscatter data. Predictions were generated by a single run of 10-fold CV on the scanned patch. This generated unbiased predictions from (1) feature space *A*, and (2) feature space *NBH*.

These predictions deviate from the ground truth as they contain the prediction errors discussed in sec. 4.4. Thus, the reference for registration is still the MR ground truth, but the triangulated NIR patch is labeled by the prediction output.

For both feature spaces the registration errors increase compared to the earlier case, where labels from the ground truth were used. This is reasonable, since prediction errors on the tissue thickness prevent reaching an optimal agreement with the ground truth. The tissue label on the reference surface and the corresponding label on the NIR patch may differ from each other. This effect, observed across the whole scanning patch, could lead to misleading or globally contradicting point-to-point correspondences with the reference surface. Nevertheless, table 5.1 shows that registration errors finally remained below 0.6 mm for both feature spaces. Qualitative statements with respect to gender and skin type are equivalent to the ones made above.

With these findings on the re-registration error, conclusions about the tissue thickness prediction error can be drawn. Since re-registration errors were satisfying for both, the *A* and the *NBH* space, the corresponding prediction errors can be assessed as sufficient. For the *A* space, the majority of subjects had a prediction error below 0.25 mm, and for the *NBH* space below 0.15 mm. Thus, both values can be considered as a rule of thumb for desirable prediction errors. The latter value is of particular interest for small scanning grids. The prediction error bound recommended for a specific subject may, however, depend on the STD of the corresponding target labels. Figure 5.3 shows a re-registration case study for the first five subjects, where the MR ground truth thicknesses have been superimposed by a Gaussian noise distribution of increasing STD. The noise shall simulate the prediction error of a regression algorithm. The green line denotes the STD for the tissue thickness at $\sigma_n = 0$ (MR-label STD). Satisfactory re-registration errors were achieved for the 0.15 mm and the 0.25 mm threshold as well as noise STDs close to the

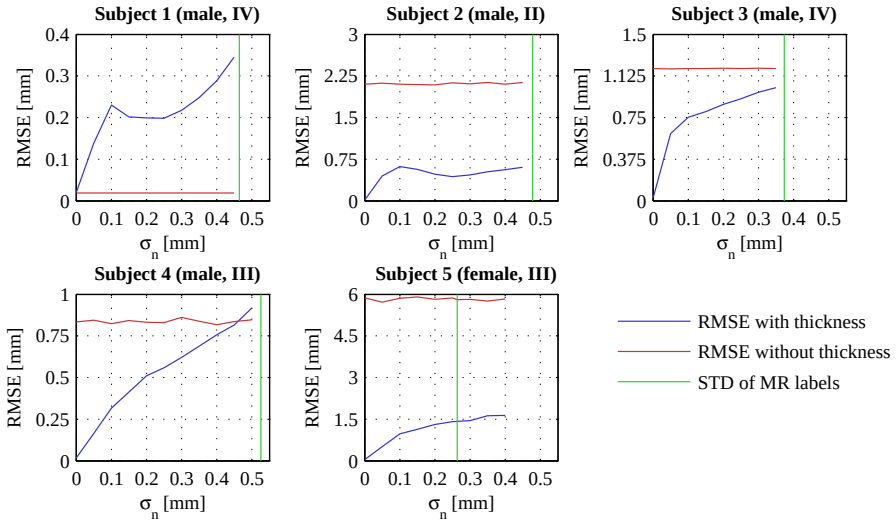


Figure 5.3: Behavior of the re-registration error for an increasing STD of random noise superimposing the tissue thickness labels. The re-registration errors averaged across 5,000 simulated movements have been evaluated for the first five subjects. Tissue thickness labels were taken from the MR ground truth, where artificial Gaussian noise of increasing STD σ_n was added.

MR-label STD. To achieve sub-millimeter accuracy for S5, however, the noise STD was required to fall below 0.1 mm (37.8% of the MR-label STD). For her, the MR-label STD was smaller than for the others. Therefore, it is reasonable to express the rule of thumb in terms of a proportion of the subject-specific MR-label STD. With the isotropic Matérn kernel, median prediction errors were 43.0% of the MR-label STD (min. 29.0%, max. 54.0%) for space *A* and 27.3% of the MR-label STD (min. 15.9%, max. 36.1%) for space *NBH*.
Figure 5.4 illustrates a typical result for subject 2 with data from the *A* feature space. This space discards less 3D points than space *NBH* during tissue thickness estimation and also constitutes a more conservative scenario in terms of the prediction error. Therefore, it has been chosen for illustration here. Each mark in the scatter plot indicates one of the 5,000 random movements. The axes are labeled by re-registration errors with (W) and without (WO) tissue support. Apart from the fact that re-registration errors were smaller when using structural information, it can be seen that the marks are not randomly scattered. There are certain areas where they pile up. This suggests the existence of local registration minima. The locations of these minima differ between cases which do or do not use tissue thickness patterns. Using the 4D instead of the 3D space for finding point-to-point correspondences, narrows the range for local minima down to a tighter error interval. Moreover, this interval also contains smaller errors. This can be seen as evidence for

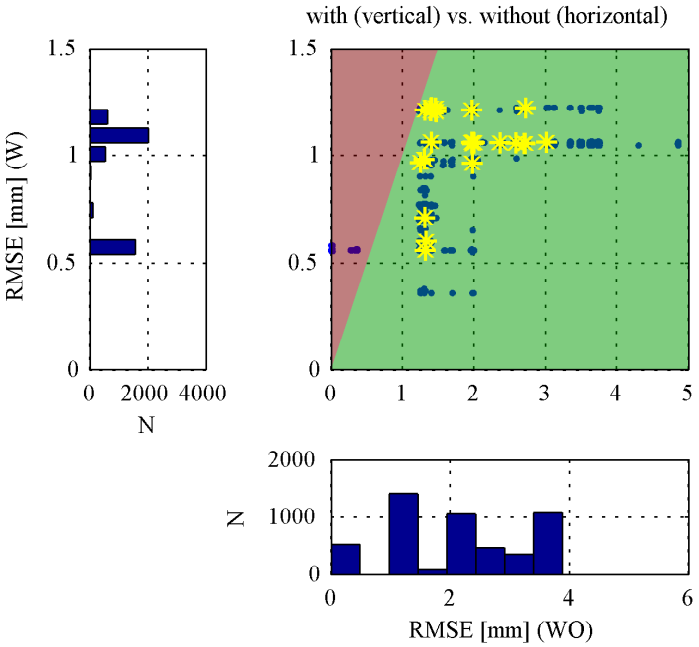


Figure 5.4: Scatter plot comparing registration with and without tissue support (subject 2, feature space A). The RMSE with (vertical axis) and without (horizontal axis) tissue thickness is plotted for 5,000 random transformations. Yellow asterisks denote cases of little motion ($t_{x,y,z} \leq 10\,\text{mm}$, $t_{x,y,z} \leq 10°$), and blue dots larger motion ($t_{x,y,z} > 10\,\text{mm}$, $r_{x,y,z} > 10°$). The histograms show accumulated case counts across one of the two coordinate axes. Regions where the surface supported method outperformed the conventional one are colored in green, the opposite is colored in red.

more robust registration.
Interestingly, the location of local minima is hardly influenced by the extent of motion. The yellow asterisks denote re-registration errors, where translational motion was less than 10 mm, and rotational motion less than 10° in each dimension. The scattering pattern across the plotted plane remains quite stable. The number of points within the accumulation centers did not generally change in favor of smaller re-registration errors. Similar effects were observed for other subjects. This suggests that less motion does not provide a guarantee for avoiding local registration minima.

Overall Discussion of the Results The interaction of different effects influencing the registration performance can be seen from the box plots in fig. 5.5. Note that subjects 8 and 19 were not included in the plot for illustration purposes. Both exceeded an average error of 10 mm for pure surface registration. Both were female.

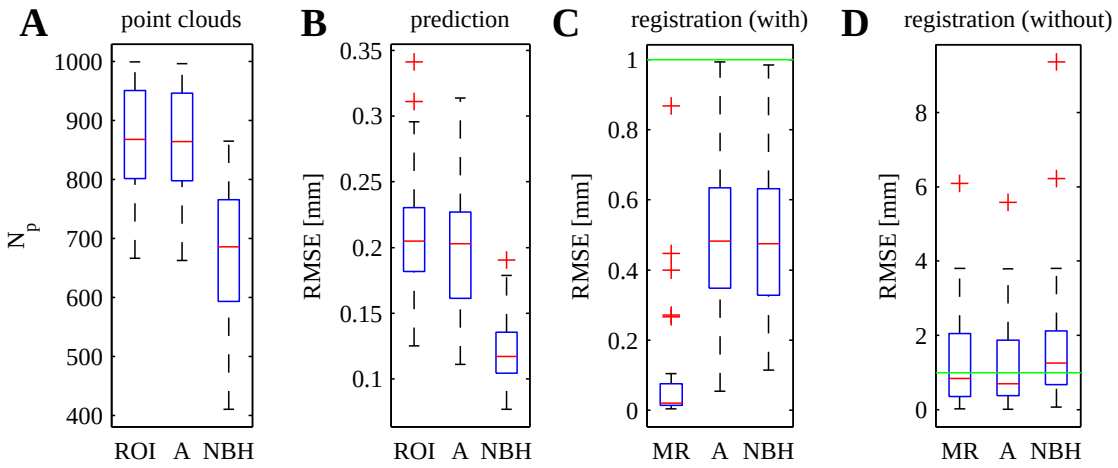


Figure 5.5: Box plots comparing the main results for tissue thickness prediction and surface registration across all 30 volunteers. **A:** number of points N_p within the tested point cloud, **B:** tissue thickness prediction error (both plots for all three feature spaces), **C:** ICP registration error with, and **D:** without tissue support (both plots for structures from the MR ground truth or predicted from the last two feature spaces). The green line denotes the 1 mm RMSE limit.

The number of 3D points for registration depends on the feature space used. When using the MR ground truth as a fourth dimension, all points that hit the forehead were used. Triangulation may have failed for some points in the grid. Therefore, there are rare cases where a point in the grid is missing all its neighbors. This predominantly happens at the grid borders. For these, no reliable estimate of the incident angle could be computed. This led to a minimal drop in the number of points for case *A*, since its feature space requires an angle estimate. Nevertheless, re-registration errors without tissue information were hardly affected by this drop and were very similar for the *MR* and *A* case. For the *NBH* case, this reduction in the total number of available samples was larger. Samples were discarded if no complete local neighborhood as defined in fig. 4.23 was available. While this may be handled differently in the future, it also constitutes an interesting case for now.

The re-registration errors for pure surface registration increased due to this loss of samples. This effect may have been intensified by the fact that loosing 3D samples at the grid margins also reduces the surface coverage in regions of higher spatial gradients. Areas around the temples constitute a typical example.

The corresponding errors with tissue support from the *NBH* space were on average slightly, but not thoroughly, better than those of the *A* space. This demonstrates, how-

ever, that the worsening effect due to the reduction of surface samples was compensated by additional information about cutaneous structures. The registration error is hence less sensitive to the size of the scanned patch.
The *NBH* case shall be discussed in more detail. For 90 % of the simulated head movements, re-registration was possible with an error of 0.52 mm or less. This bound increased to 4.56 mm if no tissue support was provided. On average, 46.1 % of all movements were re-registered with an error of less than 1 mm for both registration approaches *simultaneously*. With cutaneous features 0.52 % and without 53.8 % of all cases exceeded the error limit of 1 mm. Generally, 98.9 % of all cases that had an re-registration error above 1 mm for pure surface registration, fell below an error of 1 mm after adding support from the tissue thickness reconstruction of the *NBH* space. When using the MR ground truth instead, this value increases to even 99.96 %. In fact, 35.3 % fell even below 0.5 mm afterward. Again, no clear trend of the registration error with respect to the skin type was observed. However, the re-registration error without tissue thickness tended to be higher for females as compared to males (p = 0.06). When adding structural support, this was only the case for the A space reconstructions (p = 0.041). This space was also subject to the lowest prediction RMSEs among the cases tested for tracking. The trend was least significant when using the MR ground truth as tissue thickness label (p = 0.266).

Convergence If the correspondence problem can be resolved more efficiently, the convergence rate of the algorithm may be affected as well. All experiments used a point-to-plane surface registration. In contrast to the point-to-point alternative, a strict decay of the registration error, i.e. the evaluation of the objective function, is therefore not guaranteed. During the iterative process, the registration error can increase or decrease. One reason for this is the validity of the linearized transformation matrix. The required adjustments of the six motion parameters may not be small enough for every iteration step – particularly at the beginning of the iterative process. Thus, the fundamental assumption of the algorithm may be violated. Moreover, the algorithm occasionally oscillates between two sets of correspondence assignments when it is already quite close to an optimization minimum. This is due to the limited sampling resolution of the grids.
For the presented data, this kind of ripple was found to be only a minor problem. For most cases the registration error decreases and converges to its final value. The remaining ripple then typically stays within 10 % of this final value.
The ICP convergence rate with and without for all 30 volunteers was compared for an exemplary head movement ($t_x = 10\,\text{mm}$, $t_y = -10\,\text{mm}$, $t_z = 10\,\text{mm}$, $r_x = -8°$, $r_y = 8°$,

Table 5.2: ICP convergence rate with and without tissue thickness support for an exemplary head movement. The median, minimum and maximum ICP iteration number are given for the group of 30 subjects. Tissue thickness was either taken from the MR ground truth, or the prediction of one of the two feature spaces *A* or *NBH*.

	Median		Minimum		Maximum	
space	WO	W	WO	W	WO	W
MR	12	17	3	2	-	64
A	11	7	4	5	-	64
NBH	13	9.5	3	6	-	77

$r_z = -8°$). Each registration was done with 1,000 ICP iterations. Then, the last 100 evaluations of the objective function, i.e. the registration error, have been averaged. This mean was multiplied by 1.1 to obtain a rough threshold for defining an optimization state as settled (when 110 % of the final value is reached). The resulting statistics for the resulting iteration numbers across all 30 subjects are given in table 5.2 (without tissue thickness (WO) and with tissue thickness (W)). Except for the *MR* case, tissue thickness support led to faster convergence for most subjects. Maximum values are not given for pure surface registration. This is because the error did not settle well for subjects 5, 19, 22 and 25 (all female) resulting in a high number of iterations for them. Nevertheless they constitute interesting case studies. An example shall be given for subject 19 and the *NBH* feature space: After optimization, the evaluation of the ICP objective function yields an error of 0.60 mm without, and 0.92 mm with tissue thickness support. However, looking at the error measure defined in eq. 5.3 gives 28.96 mm without and 0.70 mm with tissue support. The other subjects show similar behavior. This indicates that these were subjects where pure surface registration had substantial problems to identify the correct alignment. The optimization was switching back and forth between several local minima. It also shows that alignments with good spatial fit are not necessarily the best guess for actually corresponding surface sites.

In terms of the median across subjects, thickness-based registration settled approximately at its final result at less than 10 iterations. Across all subjects, less than 100 iterations were sufficient. This is not the case when using the MR ground truth for the thickness overlay. Since there is no prediction error on the thickness in this case, very accurate registration was possible. After getting the rough alignment right, the refinement tended to require more iterations than pure surface-based registration. This is also reflected when looking at the number of subjects for which the thickness supported registration was faster: 13 of

30 for the *MR*, 26 of 30 for the *A*, and 22 of 30 for the *NBH* case. It seems that the higher the prediction accuracy, the more iterations are required to reach the final solution.
This suggests that there are two effects. First, tissue thickness information helps to get the rough alignment faster, and, second enables a more accurate registration when admitting more iterations for refinement. It seems reasonable that the latter effect can only have an effect if the prediction error is low enough and does allow for this.

Relevance and Scaling of the Tissue Thickness Tissue thickness was weighted by a factor of $f_{ac} = 15$ for registration exploiting the MR ground truth. For the *A* and *NBH* cases, factors $f_{ac} = 8$ and $f_{ac} = 15$ were tested first. If they did not achieve satisfactory results, f_{ac} was then increased in steps of 10 (i.e. $25, 35, 45 \ldots$) until the average error on a small subset of transformations fell below 1 mm. If this was the case, f_{ac} was set to this number for final evaluation. Then the full evaluation across 5,000 random movements was carried out. The average weighting factor for the *A* case was 24.2 (minimum 8, maximum 125), and for the *NBH* case 44.8 (minimum 8, maximum 125). This indicates that for the *NBH* case a higher weight on the tissue thickness was required to achieve similar errors. This supports the earlier argument. The higher prediction quality for this space allowed for a higher weight on the prediction output to achieve better registration results even on a smaller patch.
For the first five subjects, fig. 5.6 shows how the registration error for the feature spaces behaved when changing the scaling factor. Several observations can be made. First of all, the reduction in the total number of data samples is visible for pure surface registration with the remaining points from the *NBH* space. Second, when directly using the MR segmentation for tissue support, the registration error converges to zero for higher scaling factors. This is not necessarily the case at $f_{ac} = 15$ already. The predicted thickness does not show this behavior due to the prediction errors. For all subjects there are larger scaling factor intervals, for which the registration errors fall below 1 mm. For four of them, there are also regions where the *NBH* error undercuts that of the *A* space. This happens due to the more accurate thickness estimate and in spite of having less points in the 3D cloud. For subjects 2, 4 and 5, tissue support outperformed pure surface registration. For the other two, the error was at least similar and in any case was found to fall below 1 mm for a wide interval of scaling factors.
On average, tissue support led to an improved registration error for 29, 23, or 28 of all 30 subjects (cases *MR*, *A*, and *NBH*). For all subjects with a mean registration error of more than 1 mm before, tissue support pushed the error below this limit (*MR*: 14 , *A*: 13, and *NBH*: 20 subjects). Note that there is a higher number of subjects for the *NBH* case,

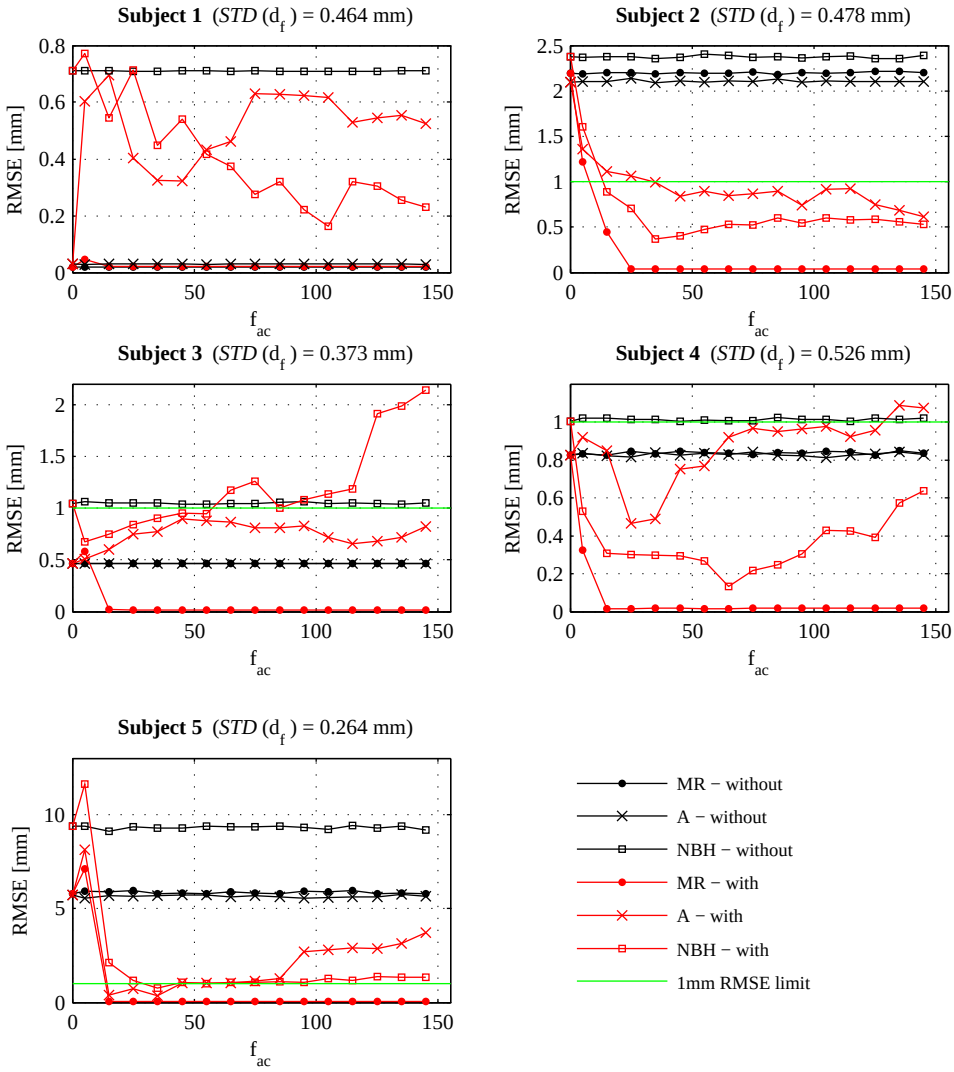


Figure 5.6: Scaling the tissue thickness for changing its relevance in the registration process – case study on five subjects. For each subject the scaling factor f_{ac} was varied between 0 and 150. The behavior of the RMSE for these different factors is presented in one plot per subject. The green line denotes the 1 mm RMSE limit and illustrates for which scaling intervals the red graphs fall below this threshold.

because the feature extraction method required discarding many points from the point cloud. This makes surface registration more difficult as mentioned before. As a result, more volunteers exhibited mean registration errors above 1 mm, which were however all pushed below this limit by using tissue thickness patterns.
Overall, satisfactory results were achieved with scaling factors between 30 and 50 in most

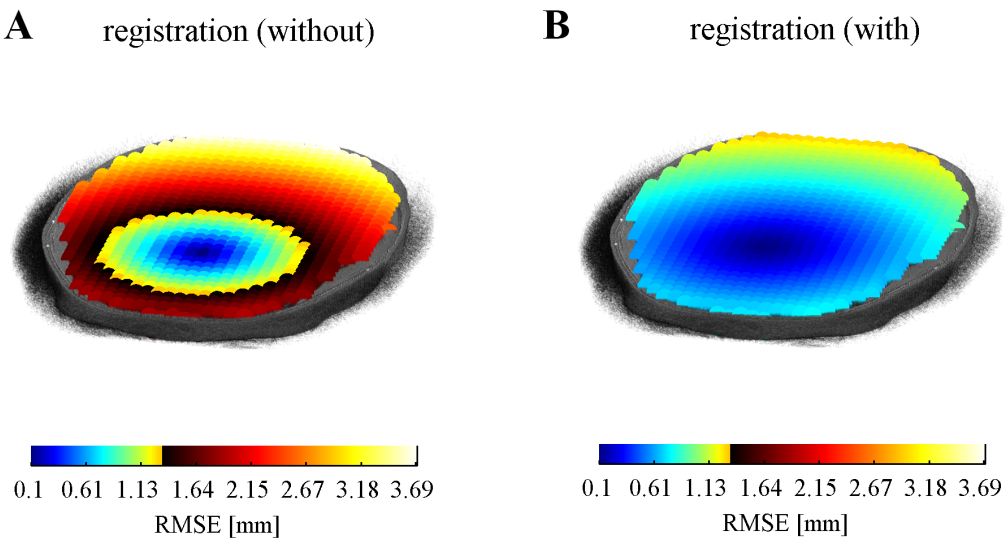


Figure 5.7: TRE comparison for subject two (male, skin type II). **A:** surface registration without, and **B:** surface registration with tissue support.

cases. The choice, however, should depend on the prediction errors for the tissue thickness. For most subjects higher scaling factors (up to 100 or more) can be considered a good choice, once a low prediction error can be achieved as for the *NBH* space. The more 3D points are available, the more the requirements for the prediction error can be relaxed.
Finding the optimal scaling for each subject can be time consuming, but can be considered as a question of computational power available.

Target Registration Errors (TREs) Finally it will be investigated how the surface registration error translates to targets in the brain. Figure 5.7 and fig. 5.8 illustrate the TREs for subject 2 and subject 5 (male and female, respectively). The registration errors were computed for the *A* feature space and equivalently for pure surface registration. Space *A* has been used for the same reason as before: It constitutes a conservative scenario in terms of tissue thickness prediction error, and does not reduce the number of points in the point cloud due to the way features are extracted as much.
The registration process happens at the forehead, which is the bottom part of each plot. The TREs depend on the distance from this forehead site (acting as a "lever"). In addition, misregistrations occur along the frontal skin-to-air transition. This frontal part describes a curved surface. It therefore spans a circle in each slice resulting in something like a blurred rotational center in the middle of the brain (in fact it has a cylindrical shape

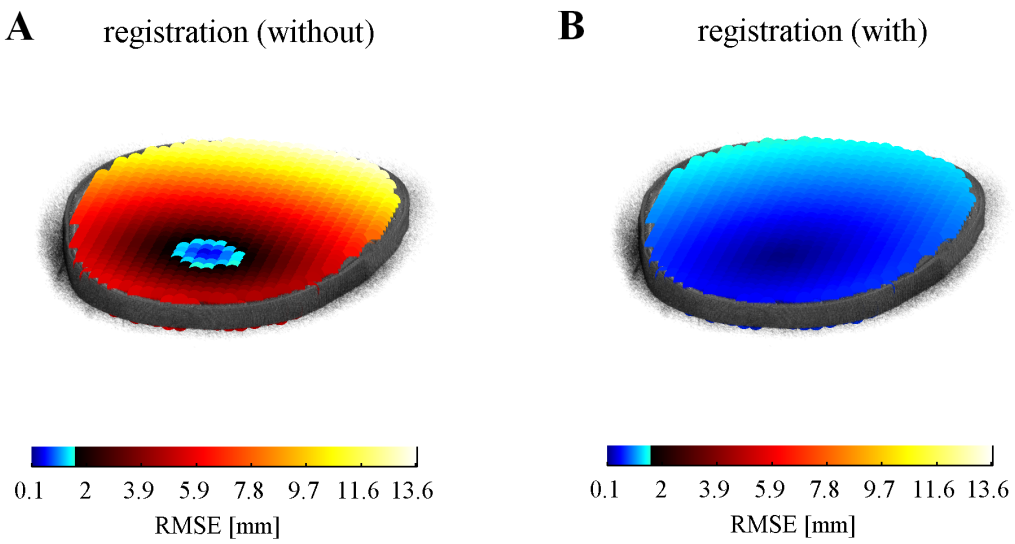


Figure 5.8: TRE comparison for subject five (female, skin type III). **A:** surface registration without, and **B:** surface registration with tissue support.

along the longitudinal body axis). All connecting lines between the registration site at the forehead and the intracranial target (the levers) roughly intersect here. The location of this center depends on the shape of the head. It is the location where the lowest TREs reside. These can in fact be lower than the actual registration error on the surface. Isolines for the TRE form ellipses around this area. Irrespective of misregistrations on the surface, voxels in the center would not suffer from large deviations from their true position. The opposite is the case for targets at the back of the head, where misregistrations are supposed to have the highest impact. This is apparent in the plots.
When comparing the results with and without tissue thickness, it is obvious that the error interval for the former, covers only a limited area in the plot of the latter. Most regions suffer from by far larger errors the extent of which depends on the subject.

5.3 Conclusions

This chapter was concerned with the interaction between the acquisition of supportive structural information, one the one hand, and its benefits for surface registration, on the other hand. The findings of the two previous chapters were used to provide a first proof of concept for the central motivation of this work: Accuracy and robustness of surface registration can be improved by providing additional tissue thickness information.
In the first subsection, a testing concept was introduced. By using simulated motion on

real world data, an exhaustive analysis for the space of possible motion trajectories was made. The actual motion parameters for each random movement were exactly known for reference. Thus, the registration performance can be compared by average point-to-point distances between the original, i.e. the desired, and the re-registered position of the scanned patch with respect to the reference surface. Note that this overcomes the requirement of somehow comparing entire transformation matrices without decoupling translational and rotational effects. Errors in one of them may entail deviations in the other, since they are all optimized simultaneously. The chosen RMSE and TRE errors are capable of covering both, translational and rotational errors in one measure.
Finally, pure surface registration was compared to structure-supported registration. Benefits were evaluated and discussed. Special concern was directed to the interaction between imperfect tissue thickness prediction and the registration error. This was investigated based on registration experiments which exploited predicted tissue thickness measures from the spaces A and NBH. The results were compared to the ideal case, where prediction would be perfect, i.e. directly yields the MR ground truth. Discussions were extended to possible intracranial tumor sites by introducing the TRE.
Overall, the findings of this chapter answer RQ 3: Does head tracking gain from incorporating tissue thickness information? The answer is yes. Detailed conclusions on the two sub-challenges raised by RQ 3 will be given in the following.

- RQ 3.1: **Where are concrete benefits for standard matching algorithms?**

  Most importantly, additional structural information helps to resolve the correspondence problem in the registration process. This is to identify corresponding pairs of points between two surfaces before performing the least-squares optimization on the transformation matrix. In the 3D space of spatial coordinates, ambiguities may arise if the surfaces are lacking prominent landmarks which can be resolved with sufficient detail. Points with the smallest Euclidean distance are not necessarily the best guess for corresponding points.
  Tissue thickness was found to tackle this problem. It adds a fourth dimension to the 3D space. Therefore, spatial similarity as a criterion is qualified by additionally considering the conformity in terms of local patterns of tissue thickness. Two points can only be considered as corresponding if they also agree in terms of their thickness value. The presented results confirm this effect. For pure surface registration, the registration error has not been found to be randomly scattered in an interval. Instead, centers of accumulation, i.e. local minima, were identified over a certain interval. This interval may cover several milli-

meters and for some subjects even errors in the low centimeter range. This was mainly the case for female subjects, whose scans had a tendency to exhibit less prominent spatial characteristics.
For tissue supported registration, these effects were less dominant. The registration errors were smaller and the aforementioned error intervals were narrowed down. Local minima do still exist, but tend to be closer to the correct match. This is also explicit in the quantitative results of table 5.1. Registration errors were lower with than without tissue thickness. Moreover, the I90 was always close to the average error. This was not the case for pure surface registration. This confirms the discussion about the error interval. Pure surface registration is more prone to outliers in the registration error, while outliers are reduced to a minimum when providing thickness information. It was also shown, that less head motion does not necessarily protect from these outliers.
These evidences show that tissue support not only leads to smaller errors in an absolute sense, but also improves the robustness of the registration process. Elaborations on the TRE demonstrated that these positive effects are even more important for tumor sites at the back of the head or generally far away from the registration site at the forehead. Nevertheless, intracranial regions have been found where misregistrations have minimal impact. Registration errors may partly even undercut those on the surface.
Finally, the convergence rate for a rough alignment was in most cases found to be faster when considering tissue thickness rather than only spatial information. If the accuracy of the thickness prediction error also allows a more precise alignment, extra iterations will lead to a refined registration. The number of iterations may then slightly exceed these for pure surface registration as it was the case for roughly half of the volunteers here.

- RQ 3.2: **Which impact has imperfect data on these benefits?**

  It was shown that tissue thickness predictions always generate imperfect reconstructions of the thickness patterns. These prediction errors could – in the worst case - lead to misleading correspondence assignments. These, in turn, could have delicate effects on the optimization of the best-fit transformation matrix. The registration process is typically iterative in its nature. Flaws in the optimized transformation matrix of previous steps may therefore corrupt the entire process of convergence. This is particularly the case for the rotational parameters.

Nevertheless, thickness reconstructions from the two feature spaces *A* and *NBH* were still found to provide sufficient support for the registration error. The average of the error was always below 1 mm. For all subjects with an error of more than 1 mm before using tissue support, the average was pushed below that limit after using it. On the level of all random transformations, more than 98 % of all cases were on average pushed below 1 mm.
The MR ground truth case improved the mean error of pure surface registration by a factor of 29. Compared to that, the reconstruction of space *A* achieved a factor of 5.6 and that of space *NBH* a factor of 7. This leads to the conclusion, that the benefit is still there but less pronounced. To achieve optimal results, the scaling factor was required to adjust the relevance of the thickness measure within the registration process. Particularly for space *A*, a smaller range of scaling factors led to improvements. They tended to fall below those identified for the *NBH* space data.
Depending on the given thickness gradients of a subject, prediction errors of less than 0.25 mm would be recommended as a rule of thumb. Ideally, errors would fall below 0.15 mm as for the *NBH* space, where registration also benefits from a smoother reconstruction of the thickness pattern. Nevertheless, subjects exhibiting only little thickness variation on the forehead, may require more accurate predictions than subjects with substantial variation within the pattern. This dependency is somewhat covered when relating the prediction errors to the STD of the corresponding MR target labels. For the isotropic Matérn kernel, the median value of this relative prediction error across all subjects was 43.0% for the *A* space, and 27.3% for the *NBH* space.
Neighborhood features led to very promising results when considering that the total number of points available on the surface was substantially reduced. For pure surface registration, on the contrary, this led to a decrease of the registration performance.
A direct comparison to the state-of-art scanning systems described in sec. 1.2.2 has not been made yet. This direct comparison would be essential and necessary, once the functional characteristics of the prototype used here have been transferred into a faster clinical prototype. Furthermore, the commercial device from a clinical cooperation partner is required. Nonetheless, the literature review in sec. 1.2.2 demonstrated that surface tracking can be reliable on average. However, also registration outliers have been reported. As a matter of fact, this has also been observed and reproduced for pure surface tracking here. For these

outliers, it was demonstrated that tissue thickness support provided a remedy. This promising indirect comparison suggests that the method proposed in this work could hence also tackle the outliers reported in those studies.
Overall, tissue thickness support is most promising when (1) unambiguous spatial characteristics of the scanned surface are rare, (2) available landmarks cannot be sufficiently resolved to achieve the desired registration performance, or (3) the scanned patch is rather small with only very little points available. Although this has not been tested, the findings suggest that registration may also be more robust, if facial movements slightly deform the surface. The best case scenario for registration involves many points per surface and a smoothly reconstructed pattern of tissue thickness, which was predicted with an error of less than 0.15 mm.
Tissue thickness support has benefits for several clinical scenarios. It can either provide a cross check or a refinement for pure surface matching, even if the scanning approach used a full head scan in first place. Moreover, the proposed method would be ideal for cases where limited access to the patient's head surface is given. Many scenarios are imaginable here, whereas the most relevant is probably given by different preferences for thermoplastic mask systems from different clinics. Many hospitals and guidelines still prefer the conservative approach of having as much of the patient's head as possible covered by the mask.

6 Conclusions

This final section will conclude the main outcome of the presented work. It will link back to the conceptional proposal described in sec. 1.3 and assess the findings in the context of this overall goal.

The proposal suggested to extend common surface registration from purely spatial matching by additional information about the surface. Since the surface is scanned optically with collimated laser light, backscattered light is an obvious source for this additional information. Tissue thickness on the forehead can serve as such information under the assumption that changes in tissue thickness will influence the local characteristics of reflected light. Given this relationship, the local tissue thickness can be inferred from the reflected light. As a label for each 3D point, it can support registration between scans. The key is that an alignment has not only to agree in terms of spatial geometry, but also in terms of the tissue labels.
Thus, the goal was to identify optimal conditions and a processing framework which can generate these labels for incoming scans during a treatment session. Label prediction will be based on a model obtained from data in the treatment planning phase (cf. fig. 1.13). Tissue thickness could be extracted from the planning modalities such as CT or MR. By mapping online scans in the planning data, the enhanced tracking concept will hence help to ensure that irradiation during the treatment is always conform to the planned dose distribution.

The three previous chapters studied the feasibility of this proposed concept in terms of research questions defined in sec. 1.3. Chapter 3 presented simulations of interactions between laser light and skin of varying thickness. Optimal parameters of an optical systems were investigated which can be used for this purpose. Chapter 4 then evaluated data which was recorded with an accordingly designed first functional prototype. The study encompassed 30 volunteers of different age, gender and skin type. Processing methods suggested by the simulations were validated on real data. The final achievement of this fourth chapter was a framework which uses feature extraction and

statistical learning to accurately predict tissue thickness from optical backscatter. Finally, chapter 5 investigated how this thickness affects the tracking performance when using state-of-the-art ICP. The ideal benefit in case of "perfect" prediction was addressed, and based on that, how prediction errors would affect this result.

The following three sections will address conclusions for each of the general three research questions. Afterward, an overall assessment within a broader context will be given. Finally, ideas and challenges for future work will be discussed in the last section. This includes suggestions for alternative fields of application for this method.

6.1 Optimal Conditions for Backscatter Analysis – Findings from Simulation (RQ 1)

Wavelength The most important parameter is the wavelength of the laser light. To receive a maximum of information about the thickness of tissue, light is required to deeply penetrate the skin. In this way, even changes in deeper tissue layers will affect a certain proportion of photons. This includes substantial changes which may occur in the subcutaneous fat. In agreement with evidences in the literature, MCML simulation confirmed that light in the NIR range is well suited for this purpose. This spectral band avoids the high water absorption at higher wavelengths and constitutes a good compromise with respect to the absorption caused by melanin, oxy- and deoxy-hemoglobin. Therefore, a wavelength of 830 nm was chosen for experimental validation.

Tissue Thickness and Optical Backscatter Changes in tissue thickness were found to have two main effects on the optical reflection. First, thicker tissue will reflect less light than thinner tissue. This was also observed when changing the incident angle of the laser beam, which due to the anisotropy factor similarly prolongs the path length of photons in the tissue until they hit the bone. Second, a global change in tissue thickness typically originates from a superposition of different changes in different layers. Therefore photons with dissimilar penetration depths are differently affected relative to some reference thickness. This will change the shape of the reflected beam profile on the surface. The simulation results also reflect that. In the conservative approximation, only the fat layer, but no other layer changed in thickness.
This gives rise to different behavior of backscatter intensities at different areas in a camera image when tissue thickness is changing. It suggests that for orthogonal irradiation, pixel intensities in ROIs concentric around the spot center show different behavior along

the radius from the spot center to the sides of the image. Contributions to this effect also arise from the fact that light which does not deeply penetrate the tissue will hardly reach areas far away from the spot center. In contrast, photons reaching the bone surface are more likely to be scattered to these areas.

Informative Features Thus, seven or later five ROIs in the form of concentric rings have been defined around the spot center. Features have been defined as the accumulated pixel intensities within each region. Their behavior for changing thickness is highly correlated due to the aforementioned global intensity changes and similar trends. Nevertheless, the feature space in fig. 3.17 also shows that the detailed functional relationship with respect to the thickness is different. This supports the argument about the second effect from the last paragraph. Nonlinear behavior is particularly seen for ROIs further away from the spot center. This was found to be a superposition of two effects. First, the thicker the skin, the more light is likely to be scattered into the outer regions. At the same time, the likelihood for absorption grows, since the traveling distance of the photon is longer. The first effect is dominant for ranges of thinner tissue. The second effect seems stronger for tissue which is already quite thick (thicknesses starting from 2-3 mm). The expected tissue thickness range on the forehead suggests that the second effect will be more dominant.
A general conclusion from this is, that nonlinear models are required to predict thickness from backscatter. This is even more true when effects of changing incident angles have to be considered as well. Features may of course be extracted from differently shaped ROIs such as stripe-shaped regions or when only taking half of the image due to artifacts. Nevertheless, it should be considered that this will tend to mix up effects from light reflected from upper as well as deeper layers. So, although the behavior is strongly correlated, the function in the feature space may turn into something more complicated. This may favor slightly different learning models and/or more training data. Nonetheless, such a consideration may be relevant when changing the light source to e.g. line lasers.
The relative proportion of photons with large tissue penetration depth is higher in the outer ROIs (cf. fig. 3.11). This means, there is also more information about a global change in tissue thickness available. Other ROIs are expected to contain more effects actually originating from regions near the surface. This finding may be compromised in practice. While the relative information content is higher further out, the signal also rapidly becomes weaker and more sensitive to noise. Considering both factors, SNR and relative information content, suggests that ROIs at intermediate distance will show the most promising tradeoff.

Imaging Sensors and Laser Power To fully exploit the information content and to optimize the SNR, laser power should be adjusted such that it drives the camera sensor into saturation at the center peak of the spot. This will allow the analog-to-digital converters of the sensor to assign as many quantization levels as possible also to small intensity changes in the spot profile. High NIR sensitivity of the sensor element is recommended. In the conservative scenario of having only thickness variance caused by the subcutaneous fat layer, it was shown that 14 bit converter resolution can resolve most of the expected thickness changes. To observe these changes with low noise, averaging across many pixels in one ROI is also beneficial. That has to be taken into account when defining the number of ROIs. Practically, it was one reason for selecting five instead of the simulated finer ROI spacing. It is always a tradeoff between averaging out information and doing the same with noise. This also implies that a highly resolved spot image is preferable to a large image that maybe contains several spots. This can be achieved with an in-beam setup, where the optical camera axis and the laser beam coincide. Simulations have not shown any prohibitive disadvantages for using in-beam instead of off-beam assemblies.

Limitations Simulations are valuable in that general disturbances such as the incident angle can be studied one by one. It has been found that this superposing effect will require compensation and a carefully chosen, most likely multivariate, feature space which possibly exploits prior knowledge in addition to optical features.
Nonetheless, simulations never perfectly cover the real-world scenario. As a shortcomings of the performed simulations, the following things have not been considered: non-planar boundaries between tissue layers, photon-photon interactions, muscle tissue, changing perfusion, oxygenation or pulse, or lateral tissue inhomogeneities such as freckles, moles, sweat, hair etc. Strictly speaking, the simulation is also only valid for average Caucasian skin. No influences arising from skin type, age or gender have been modeled. Some of these points can surely be integrated in more sophisticated simulations. However, directly proceeding to experimental validation seemed more reasonable, given that many central intuitions were already obtained from the chosen simulation approach.

6.2 Tissue Thickness Estimation (RQ 2)

Ground Truth The supervised learning problem of inferring tissue thickness from optical features requires a reliable ground truth. For this purpose, MRI scans have been

obtained from all 30 volunteers in the study. The scans had an in-plane resolution of 0.1025 mm, which is also approximately the resolution of the tissue thickness as the slices were chosen to provide cross sections through the skin. Prediction errors of the machine learning can only be reliably estimated down to this threshold. Smaller quantities might be obtained, but should be understood such that the error is indistinguishable from the resolution of the ground truth.
Furthermore, the segmentation for the thickness measure accuracy was estimated to fall below 0.2 mm from manual expert delineation. However, a large proportion from that may arise from intra-operator variability. Moreover, systematic instead of completely random segmentation errors may not be noticed by the machine learning. Instead, systematic errors might be included into the model in case of any spurious correlation with the features. Since the segmentation algorithm is based on rules, those systematics cannot be ruled out.
This may particularly have relevance when including neighborhood features as they rely on smooth changes within the tissue. Strictly speaking, these may also originate from systematic errors. Nevertheless, surface matching to the planning ground truth will rely on the same segmentation which definitely relativizes this argument.
Finally, ethical standards did not allow using CT in this volunteer study. Nevertheless, CT images are much more common and available for treatment planning in FSRT. Therefore, it is worth exploring the possibility of employing clinically used CT or MR sequences as a ground truth. Technically speaking, automatic segmentation of CT data may be easier then processing MR data which is typically richer in textures. Lower voxel resolutions could however act as a restriction. Note that lower resolutions are not necessarily a problem. They lead to a more coarse, smoothed view on the thickness patterns. As long as a reproducible pattern can be observed, it may serve as an additional landmark. A predictive model might even require a less complex functional hypothesis.

Backscatter Features and Disturbances The negative correlation seen between features and tissue thickness in the simulations was experimentally confirmed. Skin heterogeneity caused by blood vessels or disturbances such as the incident angle were found to influence this correlation. Thus, the correlation could only be shown as long as the forehead anatomy of the volunteer permitted it. Generally, the incident angle was also found to be negatively correlated with the backscatter features. This implies lower features at higher angles, i.e. the further the incidence deviates from the orthogonal case. This changes the feature space appearance in that thin skin could still correspond to low backscatter if the angle was far from orthogonal. This blurs the correlation coefficient

between features and thickness.
In agreement with the argument before, ROIs 2 and 3 (about 2.5-7.6 mm from spot center) tended to exhibit the most prominent thickness patterns and were preferably chosen by SFS. Signals in ROI 5 were weaker and more prone to external influences such as light sources or artifacts. In contrast, ROI 1 provided only minor information, since pixels were driven into sensor saturation and huge parts from the backscatter originate from upper tissue layers.

Investigations on angle compensation techniques led to the conclusion that its effects cannot be fully removed by pre-processing methods. The interaction is too complex and pre-processing options did not result in consistent improvements. Most promising was the extension of the feature space to six dimensions. This adds an estimate of the incident angle as prior knowledge to the backscatter features.

Prediction Error and Prior Knowledge The error for predicting tissue thickness was tested for three feature spaces: (1) only backscatter features (*ROI*), (2) *ROI* features plus the incident angle (*A*), and (3) all of before plus the backscatter information from neighboring spots (*NBH*). For all of them, SVR worked well as a learning algorithm and confirmed the impression obtained from earlier simulations. However, as a conclusion from the results in chapter 4, GPs should be recommended as an alternative learning algorithm. They showed superior performance in terms of prediction error and exhibit several convenient properties:

- Unbiased hyperparameter determination using maximum a posteriori estimates from the NLML.
- Avoidance of time consuming grid search.
- straightforward ARD by introducing scaling factors as additional hyperparameters.
- As for SVR, there is a kernel-based tuning of the functional complexity to fight overfitting.
- Uncertainty measures for every prediction given by the probabilistic output.

While indeed the ARD approaches yielded the lowest errors, isotropic alternatives of the same kernel did not stay far behind. Since the computational complexity is lower for isotropic approaches, the isotropic Matérn kernel was identified as the best compromise overall. On average, tissue thickness was predicted with an RMSE of 0.21 mm (*ROI*),

0.20 mm (*A*), and 0.12 mm (*NBH*). In a pair-wise comparison, the incident angle led to better prediction errors than only using ROIs. There were only very few exceptions such as subject 30 (female, skin type V), whose RMSE worsened for some frames. This was caused by rather poor triangulation results leading to a corrupted angle estimate.
Substantial improvements were then achieved when adding neighborhood features. This adds spatial information of the scanning grid to the feature space. With only ROI features, the GP maps from an intensity space to the thickness and is not aware of spatial proximity among the triangulated spots. The reconstructed tissue thickness pattern from the *NBH* space appears smoother and without abrupt changes between close by samples.
Prediction accuracy did not significantly depend on the skin type. Most likely this is due to the skin type not being tied to the underlying problem. The classical Fitzpatrick scale rather aims at UV light and therefore distinguishes between tissue characteristics relevant for lower wavelengths. There was also no significant gender difference. Nevertheless, an existing gap between errors of male and female subjects still suggests that investigations on a larger subject group are needed to further explore this.
Finally, it was found that predicting tissue thickness across different head poses is a harder problem. Errors are typically higher than for CV validation on the same scan. Evidence was found for two possible reasons. First, the overlap between the two surface patches may be too small in some cases. So, model training on one patch does not result in a sufficient coverage of the feature space to properly predict the second patch. The data from the second patch would partly be perceived as unknown behavior. Second, even if the same area on the forehead was scanned, errors can still be elevated. Different views of the scanning unit onto the surface were found to possibly cause this difference. Differences of the incident angle were correlated to increased errors. Both evidences require more training examples – from different scanning perspectives – to sufficiently sample the feature space. Investigations of this kind were not possible with the functional prototype used in this work (one scan required 20 s with a couple of minutes ”burn-in” phase preceding for initializing triangulation [312]). They will be feasible with the currently developed clinical prototype (4 s per scan, without triangulation ”burn-in”, status sept. 2015).

6.3 Surface Registration using Tissue Thickness (RQ 3)

Testing Scheme The testing scheme started from marker-based registration between NIR scan and MR ground truth. The same matching was used for GP model building. Therefore, it constitutes a reasonable choice as the reference position for tests on registra-

tion performance (cf. fig. 5.2). From there, 5,000 head movements were simulated within intervals of ± 20 mm translation and $\pm 20°$ rotation in each axis. The re-registration performance with and without tissue thickness then constitutes a good measure to evaluate a proof of concept for registration exploiting tissue thickness. Since the design of sophisticated tracking algorithms was beyond the scope of this work, state-of-the-art point-to-plane ICP was used for this. Advantages of the simulated motion concept include that:

- The ground truth is known by means of the marker-based reference registration.
- A comprehensive space of possible movements can be easily covered.
- Both, translational and rotational errors of the transformation matrix can be covered in a single error measure as defined by eq. 5.3.

Enhanced Surface Registration Results of the registration experiments supported the claim that surface registration is prone to local minima, i.e. geometric ambiguities due to similarities on the surface. The iterative process of registration was found to get trapped in them – even when the upper bounds for the random motion were reduced. As a matter of fact, this is in line with relevant background literature, where studies – even when using full-face scans – also reported registration outliers. Spatially, these local minima were more widely spread for pure surface registration than for the tissue enhanced alternative. This indicated better robustness against outliers when tissue structures are available. This is also supported by the corresponding I90 values, which were relatively close to the mean registration error. This is not the case for pure surface registration.
The tissue thickness used for registration was generated from either (1) the MR segmentation, or cross-validated predictions arising from (2) the *A*, or (3) the *NBH* space. Averaged across all transformations and subjects, an improvement of factor 29, 5.6, and 7, respectively, was achieved. With regard to the results for pure surface registration, this implies on average sub-millimeter accuracy on subject level (cf. fig. 5.5). This is in agreement with the overall goal of this work. While outliers with average errors above 1 mm were observed for pure surface registration, in fact not a single subject had average registration errors above this threshold when switching to the enhanced concept proposed here.
Setting the registration errors into the context of the prediction error, permits recommendations of how accurate the statistical learning should ideally be. Generally, prediction errors of less than 0.25 mm are recommended as a rule of thumb. Depending on the prominence of the forehead patterns, this may of course slightly vary. For a subject having very prominent gradients due to muscles, head shape or vessels, higher errors

can be acceptable. This dependency can be somewhat included into the rule of thumb by expressing the prediction error as a proportion of the corresponding MR-label STD. The median of this measure was 43.0%, when applying the isotropic Matérn kernel to data from the *A* space. Equally, it is also important how the error is spatially distributed. The rule of thumb is based on randomly scattered errors across the forehead. In case of very clustered errors, i.e. some systematic misinterpretation, the impact on the registration performance could be more negative. Therefore, lower prediction errors are always desirable, particularly if the scanned area or the number of scanned points is small. This can be seen in the *NBH* case where most errors were below 0.15 mm. Here, the isotropic Matérn kernel yielded RMSEs corresponding on average to only 27.3% of the MR-label STD.
Conditions in which tissue thickness provides particularly valuable benefits are:

- The lack of unambiguous spatial characteristics of the scanned surface.
- An insufficient capability to resolve available landmarks within the desired tracking accuracy.
- A small area or only a few hundred points are available from the scan.

Although not directly investigated here, one may speculate that especially the compensation of rotational head motion may benefit from these additional landmarks. For the head, rotations belong to the rather poorly fixed degrees of freedom, particularly when only a forehead patch is scanned.

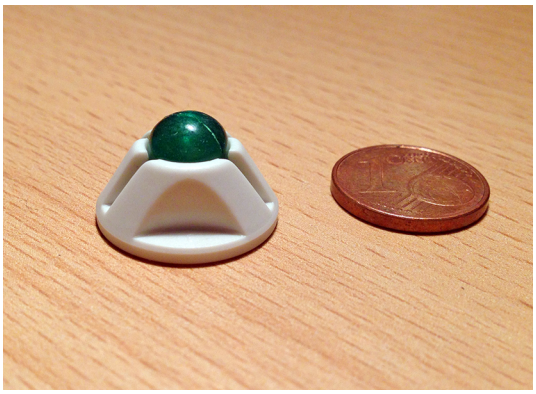
6.4 Overall Assessment of this Work

The last three sections explicitly discussed the results obtained in this work in detail and with respect to the research questions. The next paragraphs will provide an overall assessment in a broader context. A central question is to which extent the presented results already fulfill the requirements of the proposal illustrated in fig. 1.13. Which criteria of a clinically applicable concept are already met and what are open issues. Naming these will then lead to the next section presenting future work.

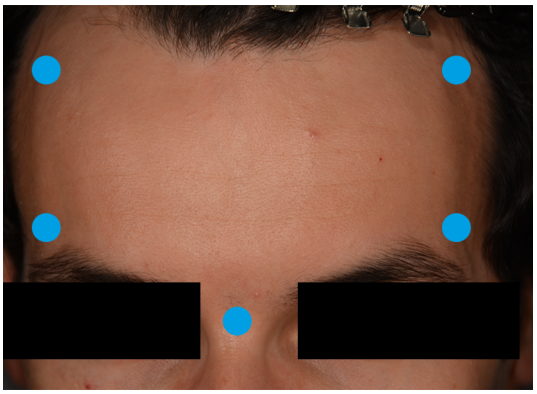
The goal was to develop a concept for monitoring head motion during treatment – marker-less and based on surface registration. This surface registration should be made more robust and accurate by measuring additional landmarks from the optical backscatter to support the spatial information. This has been achieved. As a concept,

it was shown, that NIR optical backscatter from tissue can be used to train a statistical model. This model can then be used to predict patterns of tissue thickness which have a positive effect on the tracking performance. With prediction errors of mainly less than 0.25 mm, it was shown that the number of outliers is reduced and surface registration errors fall into the sub-millimeter range. Convincing results were obtained for CV where the training data was capable of sufficiently modeling the function within the feature space. This demonstrates the feasibility for small deviations, i.e. little motion, from the head pose in which the model was trained.
In a clinical scenario the training data will be obtained during planning and maybe within the simulator. The model is then applied during treatment later on to estimate thickness patterns and to use them for tracking. The agreement between planning and treatment in terms of patient and device alignment would ideally render the tested CV scenario as realistic. One can imagine a setting where the patient and the optical tracking device are aligned to the isocenter of the LINAC as defined during planning. Then deviations from these positions need to be detected.
Nevertheless, alignment may not be perfect in all cases and larger motion may also be relevant. An example would be the initial repositioning of the patient during treatment. Therefore, prediction errors between head poses were also explored in this work. Between the three head poses of each volunteer, head motion mostly in the centimeter range and rotations of several degrees occurred. The results suggest, that insufficient overlap of the data within the feature space makes it harder to keep the prediction errors in the aforementioned range.
Nonetheless, the results obtained were very promising as they indicate that including more relevant samples into the training set might resolve the problem. This was indicated by prediction errors across poses which improved, when the same area under a similar scanning angle was covered within the training data.
As a conclusion, strong evidences were given that the concept works under little deviation from the training pose (a few millimeters and degrees). For larger deviations, promising results and recommendations were developed. To further strengthen the evidence (1) a large scale clinical study with real patients will be necessary, and (2) more training data has to be acquired for model building on patient level. Both points will be possible with the currently developed clinical prototype having a faster scanning speed.

Another important aspect is given by the MR ground truth. The tracking concept relies on supervised learning and requires target labels, i.e. a tissue thickness reference. In this work, a possible approach was presented, implemented and validated. It consists



A: Example marker spheres that contain liquid and can be observed by volume imaging and the scanning prototype.



B: Five possible placements for at least 3 markers which are unlikely to interfere with the scanning process.

Figure 6.1: Suggestion for replacing the initial registration between MR and NIR based on bite markers. Since the marker is only needed once, sticky spheres could be used instead.

of high resolution MR scans and a bite-marker-based matching thereof to optical scans. This approach is feasible, but more practical solutions may exist. The MR sequence is not clinical standard and would entail more costs and additional time effort in the clinic. Possible alternatives include the clinical CT or in some cases T1 MR scans used for treatment planning. For model adaptation the On-Board Imager® CBCT can also be considered. The resolution is generally lower, but can be increased – particularly in-plane. As for CT scans, this always entails a tradeoff between accuracy, radiation dose and time for acquisition. Nevertheless, the usage of lower resolution images is not precluded and needs to be evaluated.

The bite marker was used for initial registration to match thicknesses and NIR features. Head motion could also be tracked and removed from the data. This bite block marker and the tracking camera (Atracsys accuTrack 250) are not necessary at faster scanning speeds. They are of course not part of the marker-less tracking concept in any way. Here, it was only used for better experimental validation. Therefore, the bite marker can be replaced by simple sticky marker spheres as shown in fig. 6.1A. These are chosen such that they are visible within the reference volume scan and can be seen by the triangulation camera of the scanning prototype. This allows for an easy and less complex initial registration. The transformation pipeline in fig. 4.10 and an additional CT scan of the marker would be obsolete. The sticky markers are only required once for model building. The patient would be equipped with at least three of them (cf. example

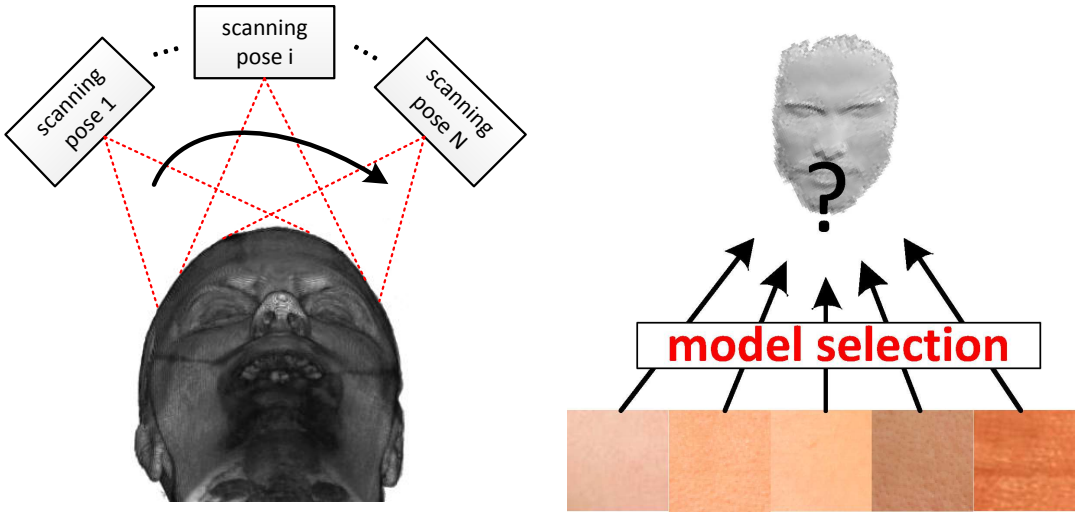
Table 6.1: Comparative overview of different target localization approaches updated by including the concept presented in this work.

	CBCT & stereoscopic X-ray	marker-less optical tracking	enhanced marker-less optical tracking
Imaging Modality	kV X-rays	optical	optical
Imaging Dose	low-medium	none	none
Detection Capability	inter- and limited intra-fractional	inter- and intra-fractional	inter- and intra-fractional
Registration	volumetric	surface	structured surface
Patient Comfort	mask immobilization	potentially without immobilization	potentially only head rest
Accuracy	high	medium	medium-high
Imaging speed	slow-medium	real-time	real-time
Operating Expense	costly, but minimal impact on workflow	inexpensive and low impact on workflow	inexpensive and low impact on workflow[a]

a Potentially. Further evaluation on that has to be included in future work. See next section.

in fig. 6.1B) before he undergoes the planning CT. After CT and/or maybe another reference image, the NIR scans for model building are acquired. After this, the markers can be removed. There is no need to wear them again.

Overall, the feasibility of a novel tracking concept in radiotherapy was successfully demonstrated. To the author's knowledge, it is the first of its kind. Similar to table 1.3 at the very beginning, table 6.1 gives a comparative overview, now updated by including the new concept. Main concern and also major contribution of this work was the key feature of this concept: the exploitation of optical information to obtain supportive patterns of tissue thickness for enhanced tracking. A proof of concept has been presented. As a main conclusion, it can be said that the concept is promising for related applications like gating and/or online motion compensation. It is capable of reducing outliers and to generally decrease the surface registration error.



A: Data acquisition for building individual patient models.

B: Patient model assignment according to skin type related data pool.

Figure 6.2: Two future recommendations involve the process of model building. On patient level data should be recorded by sweeping the setup around the head while scanning. A workaround for individual models would involve unsupervised learning and subject classification into one set of available models.

6.5 Future Work

To fully validate the concept such that it efficiently adapts to a typical clinical setting, more detailed insights are required. First of all, the study conducted here comprises only 30 volunteers. More comprehensive investigations are only possible on a larger and more realistic data basis. Thus, a larger study in a typical clinical environment would be required. Apart from the advantage of realistic surroundings, data will also be recorded from real patients who underwent treatment planning. Further on, the proposed prototype could be compared with commercially available surface tracking systems in the clinic. This would also include a comparison between the tissue supported forehead scanning and the classical full-face scan. Immobilization using an open-face mask would serve this purpose. Only such an investigation, where both systems run simultaneously, allows for a definitive statement about whether outliers of the classical approach (as reported by peer studies) can be reduced with the proposed concept. For this investigation, scanning speed and data acquisition have to be faster. Issues discussed in the previous section need to be taken into account. This also involves a prototype to isocenter calibration.

Data Acquisition First, the results for predicting tissue thickness across different head poses have shown, that data for building the statistical model needs to be carefully acquired. More data from different scanning perspectives is required. Figure 6.2A illustrates a possible workflow where the scanner is swept around the forehead and N scans are acquired: A patient would be equipped with sticky markers as described before. Then he would undergo CT or possibly MR imaging. Before removing the markers, the patient is placed on the simulator couch and the scanning device is swept around the head while acquiring optical scans. These are then registered to the volume reference and the statistical model between tissue thickness and NIR features can be built. For this model building, sparse approximation or clustering techniques will be of importance to reduce computational complexity.
Second and beyond patient level, one may think about using statistical models across patients. This would save time and costs in the clinic, since no individual model is required for a patient. As illustrated in fig. 6.2B, the patient will be classified into one existing model class and the most suitable model will be selected and can be employed off-the-shelf. Reference scans are not required from new patients, once the data base has been generated. No marker equipment of any kind would be necessary.
Different strategies are imaginable. The data pool can be viewed as a dictionary, where each patient model in the data base is one class. A new patient is assigned the model from another patient which is closest to the characteristics of the current patient. The other extreme would be a unified model. One model is trained across all patients in the data base. To cover all possible variations across patients this would require large amounts of data. Finally, a compromise can be found. The data base can be clustered into groups of similar models, e.g. according to their relevant skin type. Then, the patient is classified into one of these groups and the group model will be used. This would require less data, but might be still challenging for GPs.
Depending on the amount of data, artificial neural networks would also constitute an option [28]. They are capable of more efficiently dealing with large amounts of data. Only a few weights instead of the entire covariance matrix need to be stored. The idea could then be extended to deep learning architectures which require larger amounts of data. These would also provide ways to optimize feature extraction and to identify more appropriate data representations. Autoencoders constitute one example here [20]. Generally, the field of deep learning has encountered several breakthroughs in the last decade [125], which makes it usable and promising for this kind of application.

Subject Classification Findings in this work have provided evidence that the Fitzpatrick scale is not a suitable, yet very subjective measure for a relevant skin type. Relevance is linked to the goal of distinguishing model characteristics in the context of the work presented here. Other skin types and characteristics need to be explored and to be assessed with respect to their relevance. This may involve optical, mechanical or other physiological properties of the skin. Weyrich et al. built a huge skin reflectance data base including 149 subjects of different skin types [322]. They further presented a setup [321] which, in a similar or simpler way, could be used along with standard dermatological means to characterize the human skin. Corresponding data based on skin reflectance could be mapped to the statistical models which were acquired here. Correlations and relationships can be evaluated to enable subject classification or to define a "closeness measure" with respect to existing models. Several studies in many different fields have already investigated possibilities to model skin [10] or how to extract physiologically relevant features for classification [286, 297, 342]. This includes spectral features and a decomposition into components which partly resemble physiological characteristics such as hemoglobin or melanin content. This may be interesting for monitoring quantities other than tissue thickness – in other application fields – with the presented setup.

Feature Extraction Autoencoders have already been mentioned as a tool to optimize feature extraction from the camera images. They are capable of taking a high dimensional data input and of learning a compression to encode the most relevant information. Thus, they can learn an optimal transformation between raw data and extracted features. Other improvements may be closer to the current processing chain and can tackle its existing weaknesses. One of them is given by the fact that the neighborhood features are not computable for every point. For some points of the grid, some neighbors are missing which would require them to be removed from feature set as well as the point cloud. This gives away a lot of information and there is surely potential for improvement.
As one possibility, standard imputation methods from statistical learning can be evaluated [66, 83, 267]. These tackle cases where feature entries in some feature vectors are missing. Essentially, it is aimed at guessing optimal replacement entries to be still able to use the entire vector. A large number of the approaches relies on various ways of interpolation.
A second alternative is given by extending GPs to a so-called multi-task Gaussian process (MTGP) [32, 73]. So far, the GPs treated a single stream of multivariate feature vectors to predict the tissue thickness. This is a problem once feature entries are missing. As a workaround, MTGPs define several tasks, which are less strongly tied together but still

correlated. The covariance matrix is extended by a correlation matrix. The tasks share similar parameters for their properties, but are not required to be sampled at the same feature instances. Therefore, missing values are not a problem. This would replace a single feature vector with synchronous entries of ROI and NBH features by two or more tasks. One task is given by the variation of the tissue thickness according to the ROI features of the central spot. The second or the other tasks model the variation of the neighborhood features in dependency of the ROI features at the central spot.
This is the analogue to recordings of two non-synchronized sensors over time. Time is replaced by the ROI features of the central spot and the sensor outputs are the tissue thickness and the neighborhood features (e.g. one task per neighbor location). Thus, correlation between tissue thickness and neighborhood is modeled by the task interaction and no strict availability of a neighborhood for each spot is required. In fact, the correlation between tasks could be seen as a way of imputation as well.

Enhanced Tracking Concept This work did not aim at developing a dedicated tracking algorithm. This will be a matter of future work. It has to be investigated whether there are better ways for incorporating the tissue thickness into the registration problem. The elimination of the thickness weighting factor would be one desirable goal, for example. Further on, a multi-step algorithm is imaginable. This could, for instance, allow for using vessels or moles as artificial landmarks for pre-registration, if available. Enhancements can also be attained by admitting further portions of the face with properties similar to the forehead: Including e.g. the cheek bones could add more spatial information and may also be eligible for the tissue thickness concept. Furthermore, the feasibility of hybrid approaches could also be explored. These may use both, surface information with and without thickness information in a single registration problem. This would make full face scans applicable.
Beside the exploitation of different options on the algorithmic side (e.g. non-rigid registration as done by Amberg and colleagues [5]), the concept as such could also be adjusted. Motion monitoring over weeks may involve challenges induced by changes of the patient anatomy. These changes may include long-term effects such as extreme weight loss or instantaneous effects such as facial expressions during treatment. The first could be tackled by online adaptation of the tissue model. This would also tie the model closer to the patient in case a general model was used. A possible modality, which can act as a reference for the tissue thickness in this context, is the CBCT if used for re-positioning check. These checks are not required in each session, since marker-less tracking could take over. They are, however, useful as occasional, e.g. weekly, checks as some clinics suggest [202, 220].

Furthermore, the tracking concept can be designed such that the bone surface is computed from skin surface and tissue thickness. The thickness patterns can then be projected onto the bone surface, which is then, in turn, used for tracking. This can be more robust against slight surface deformations and in parts also against facial expressions. Comparisons between the bone and skin surface, possibly over time, could also detect facial expressions such as frowning etc. A detection of these strong deformations can be used for gating or to switch between bone and skin surface. A minor disadvantage of the bone surface is that triangulation errors and tissue thickness prediction errors can add up and compromise the accuracy.

Further Applications The idea presented in this work can be adopted in other fields of application. First, the recovery of thickness patterns on the forehead could be used in biometry. Benefits are imaginable in many fields. First of all, this would make it possible to identify the FSRT patient on the treatment couch by comparing the thickness patterns of the NIR scan and the tissue thickness from the planning CT or MR. This can at least make the doctor aware of any discrepancy and help to avoid severe mistakes.
Second, the principle can be used for general contact-less patient monitoring. The backscatter does not only contain information about the tissue thickness, but also about other vital parameters. These include heart rate, blood oxygenation, blood pressure all the way to more abstract parameters such as arousal. Similar work has already been done by other groups, e.g. for contact-less monitoring in the MR scanner [184] or in other scenarios [261, 308].
Another similar application is given by neonatal monitoring of newborns. In particular for premature babies, contact-less monitoring is of high importance. First applications for infrared thermography exist and there is a strong need for remotely recording other parameters such as the aforementioned ones [1, 58]. Studies on measuring the blood oxygenation of newborns via optical reflection have been shown to be remote but not yet contact-less [57]. In this context, the developed data processing chain may be a good start. The 3D surface reconstruction would not necessarily be required, but can have advantages when dealing with motion. Contact-less monitoring is always sensitive to motion, since it is not ensured that data comes from the same spot. Therefore, information retrieval is aggravated. The knowledge about tracked 3D surfaces could solve this challenging problem (known as bulk motion sensitivity [184]).
Finally, a last application is given in material testing. There are several applications where products have to be checked with respect to cracks, pores etc. Among others, this is done optically using e.g. optical coherence tomography or other means [131, 251, 280]. Ana-

lyzing reflectance patterns could serve as a promising complement. This particularly applies to the food industry where non-destructive and contact-less testing is required [141]. In this context one can imagine that a manufacturer wants to detect irregularities within his production line. Products can be classified into classes, quality parameters can be predicted, but also one-class classification is imaginable. This would be related to the field of novelty detection, where anomalies are detected if a set of features is unlikely to be conform to the ”normal” standards [226].

7 Appendix

7.1 Volunteer Study

Table 7.1 lists the detailed subject information from all 30 volunteers. Subject numbers agree with the ones used within the rest of this work.

Table 7.1: Summary of the subject information. Age and skin type (I-VI) are given for the time of the study.

subject	age	gender	score[a]	skin type	eye color[b]	hair color	comment
1	27	m	24	IV	3	brown	Caucasian
2	34	m	13	II	1	blond	Caucasian
3	30	m	22	IV	2	brown	Caucasian
4	35	m	14	III	1	brown	Caucasian
5	34	f	14	III	0	blond	Caucasian
6	34	m	14	III	2	dark blond	Caucasian
7	31	m	23	IV	1	dark blond	Caucasian
8	30	f	8	II	2	red	Caucasian
9	32	m	26	IV	1	brown	Mexican
10	29	m	16	III	2	dark blond	Caucasian
11	29	m	22	IV	2	dark blond	Caucasian
12	53	m	18	III	1	dark blond	Caucasian
13	27	f	18	III	3	dark blond	Caucasian
14	28	f	9	II	3	red	Caucasian
15	29	f	7	II	0	red	Caucasian
16	26	f	7	II	1	blond	Caucasian
17	65	m	24	IV	2	brown	Caucasian
18	28	f	29	IV	3	brown	Caucasian
19	27	f	16	III	0	dark blond	Caucasian
20	27	f	24	IV	4	black	Chinese
21	24	m	31	V	4	black	Indian
22	28	f	26	IV	3	brown	Indonesian
23	25	f	10	II	0	blond	Caucasian
24	54	f	21	IV	2	dark blond	Caucasian
25	28	f	26	V	4	black	Nepalese
26	30	f	23	IV	3	brown	Arabic
27	30	m	21	IV	2	black	Chinese
28	25	f	20	III	3	light brown	Caucasian
29	29	m	26	IV	3	black	Thai
30	29	f	30	V	4	black	South African
mean	31.9	14/16	19	0/6/8/13/3/0			

a Skin typing score according to the Fitzpatrick scale (cf. [95, 230]).

b Eye color as grouped within the questionnaire according to the Fitzpatrick scale (cf. [95, 230]).

7.2 Tissue Thickness Prediction Accuracy

7.2.1 Statistical Learning

The following table supplements the data shown in sec. 4.4.3. For real data obtained during the subject study, the tables list the accuracies for the remaining machine learning approaches.

Table 7.2: Tissue thickness prediction accuracies for the SVR and GP approaches under AM1 testing scheme. The results are listed for all feature spaces. For SVR a test for the entire cloud and one considering only points in the mutual overlap set is presented. The latter results are labeled with the overlap mark ov.

		gender		skin type			
	total	male	female	II	III	IV	V
$RMSE_{GP:SE_{iso}}$ (*ROI*) [mm]	0.471	0.459	0.481	0.459	0.475	0.463	0.512
$RMSE_{GP:SE_{iso}}$ (*A*) [mm]	0.465	0.449	0.480	0.453	0.468	0.451	0.548
$RMSE_{GP:SE_{iso}}$ (*NBH*) [mm]	0.400	0.392	0.408	0.372	0.411	0.392	0.463
$RMSE_{GP:SE_{ard}}$ (*ROI*) [mm]	0.464	0.451	0.475	0.456	0.467	0.457	0.503
$RMSE_{GP:SE_{ard}}$ (*A*) [mm]	0.457	0.442	0.471	0.451	0.462	0.450	0.488
$RMSE_{GP:SE_{ard}}$ (*NBH*) [mm]	0.376	0.361	0.389	0.350	0.395	0.369	0.411
$RMSE_{GP:Mat_{ard}}$ (*ROI*) [mm]	0.453	0.444	0.461	0.437	0.460	0.444	0.503
$RMSE_{GP:Mat_{ard}}$ (*A*) [mm]	0.445	0.432	0.456	0.433	0.453	0.436	0.484
$RMSE_{GP:Mat_{ard}}$ (*NBH*) [mm]	**0.368**	**0.354**	**0.381**	**0.332**	**0.407**	**0.356**	**0.393**
$RMSE_{SVR:RBF_{iso}}$ (*ROI*) [mm]	0.436	0.428	0.442	0.435	0.428	0.435	0.458
$RMSE_{SVR:RBF_{iso}}$ (*A*) [mm]	0.432	0.424	0.440	0.432	0.431	0.428	0.453
$RMSE_{SVR:RBF_{iso}}$ (*NBH*) [mm]	0.411	0.404	0.417	0.414	0.406	0.412	0.411
$RMSE^{ov}_{SVR:RBF_{iso}}$ (*ROI*) [mm]	0.396	0.395	0.398	0.391	0.373	0.417	0.380
$RMSE^{ov}_{SVR:RBF_{iso}}$ (*A*) [mm]	0.396	0.391	0.400	0.395	0.376	0.413	0.378
$RMSE^{ov}_{SVR:RBF_{iso}}$ (*NBH*) [mm]	0.375	0.377	0.373	0.376	0.353	0.394	0.346

Table 7.3: Tissue thickness prediction accuracies for the SVR and GP approaches under AM2 testing scheme. The results are listed for all feature spaces. For SVR a test for the entire cloud and one considering only points in the mutual overlap set is presented. The latter results are labeled with the overlap mark ov.

		gender		skin type			
	total	male	female	II	III	IV	V
$RMSE_{GP:SE_{iso}}$ (*ROI*) [mm]	0.447	0.450	0.445	0.409	0.458	0.446	0.498
$RMSE_{GP:SE_{iso}}$ (*A*) [mm]	0.437	0.432	0.442	0.400	0.450	0.431	0.505
$RMSE_{GP:SE_{iso}}$ (*NBH*) [mm]	0.365	0.366	0.364	0.329	0.378	0.366	0.398
$RMSE_{GP:SE_{ard}}$ (*ROI*) [mm]	0.441	0.442	0.441	0.402	0.454	0.441	0.486
$RMSE_{GP:SE_{ard}}$ (*A*) [mm]	0.434	0.429	0.438	0.403	0.449	0.430	0.470
$RMSE_{GP:SE_{ard}}$ (*NBH*) [mm]	0.349	0.355	0.344	0.306	0.377	0.347	0.369
$RMSE_{GP:Mat_{ard}}$ (*ROI*) [mm]	0.433	0.436	0.430	0.392	0.448	0.431	0.479
$RMSE_{GP:Mat_{ard}}$ (*A*) [mm]	0.424	0.421	0.426	0.387	0.443	0.419	0.466
$RMSE_{GP:Mat_{ard}}$ (*NBH*) [mm]	**0.340**	**0.345**	**0.336**	**0.292**	**0.379**	**0.336**	**0.349**
$RMSE_{SVR:RBF_{iso}}$ (*ROI*) [mm]	0.413	0.416	0.411	0.389	0.414	0.421	0.429
$RMSE_{SVR:RBF_{iso}}$ (*A*) [mm]	0.410	0.407	0.413	0.390	0.418	0.414	0.417
$RMSE_{SVR:RBF_{iso}}$ (*NBH*) [mm]	0.380	0.380	0.381	0.384	0.379	0.382	0.368
$RMSE^{ov}_{SVR:RBF_{iso}}$ (*ROI*) [mm]	0.396	0.394	0.397	0.379	0.385	0.409	0.397
$RMSE^{ov}_{SVR:RBF_{iso}}$ (*A*) [mm]	0.397	0.392	0.401	0.385	0.387	0.409	0.394
$RMSE^{ov}_{SVR:RBF_{iso}}$ (*NBH*) [mm]	0.371	0.374	0.368	0.372	0.361	0.382	0.346

7.2.2 Sequential Forward Selection

The plots below present different RMSE characteristics after SFS for different subjects. They provide example cases for the averaged behavior shown in fig. 4.28. From left to right the influence of the curse of dimensionality has an increasing impact. The extent of this impact depends on the complexity of the manifold in the subject's feature space as well as the number of data points.

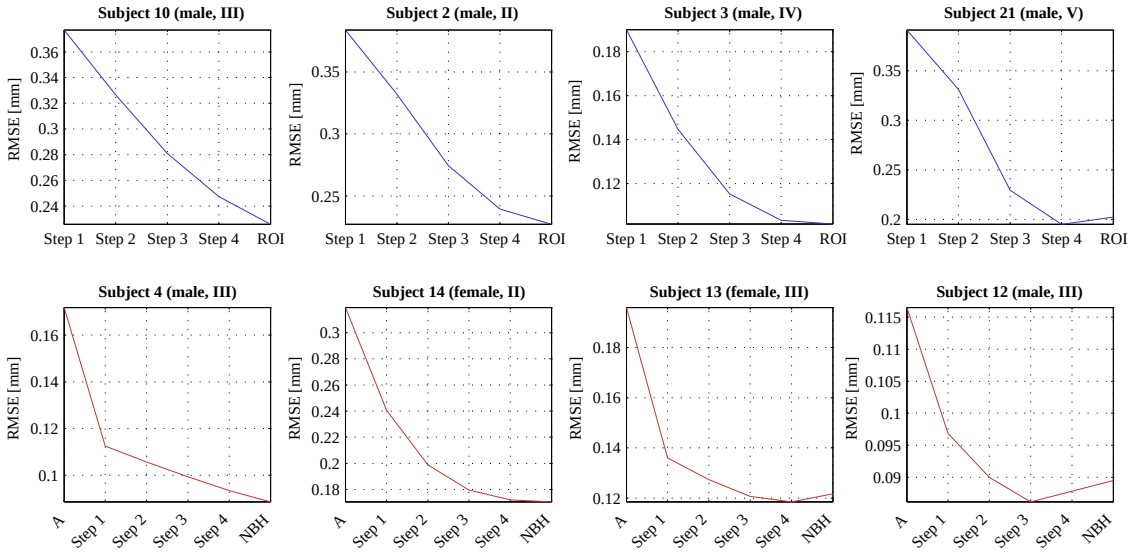


Figure 7.1: Increasing the dimension of the feature space corresponds to a tradeoff between an information gain and the curse of dimensionality [309]. Depending on the feature space characteristics, one of the two factors dominates. This leads to a decrease, or particularly for higher dimensions, an increase of the prediction error. The plots show SFS examples for different subjects, whereas the negative impact of the number of feature space dimensions is increasingly visible from left to right. Subject gender and skin type are given in parentheses. **Top:** SFS on the ROI backscatter features. **Bottom:** SFS on the NBH features from neighboring spots. Plots are analogous to fig. 4.28.

Bibliography

[1] A. K. Abbas and S. Leonhardt. Intelligent neonatal monitoring based on a virtual thermal sensor. *BMC Med Imaging*, **14**:9, 2014.

[2] Y. Abdel-Aziz and D. H. Karara. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proceedings of the Symposium on Close-Range Photogrammetry*, pp. 1–18, Falls Church, VA: American Society of Photogrammetry, 1971.

[3] Accuray Inc., Sunnyvale, CA, USA, 2015. `http://www.accuray.com/`. (accessed on 2015-04-09).

[4] E. Alerstam, W. C. Y. Lo, T. D. Han, J. Rose, S. Andersson-Engels, and L. Lilge. Next-generation acceleration and code optimization for light transport in turbid media using gpus. *Biomed. Opt. Express*, **1** (2):658–675, 2010. DOI: 10.1364/BOE.1.000658.

[5] B. Amberg, S. Romdhani, and T. Vetter. Optimal step nonrigid ICP algorithms for surface registration. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1–8, 2007. DOI: 10.1109/CVPR.2007.383165.

[6] American Society for Radiation Oncology (ASTRO), 2015. `www.astro.org`. (accessed on 2015-04-13).

[7] R. R. Anderson and J. A. Parrish. The optics of human skin. *Journal of Investigative Dermatology*, **77**(1): 13–19, 1981. DOI: 10.1111/1523-1747.ep12479191.

[8] L. Andor Technology. *Zyla sCMOS Hardware Guide v1.6*, 15 edition, 2014.

[9] Andor Technology Ltd., Belfast, UK, 2015. `http://www.andor.com/`. (accessed on 2015-07-06).

[10] E. Angelopoulo, R. Molana, and K. Daniilidis. Multispectral skin color modeling. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, pp. II–635–II–642 vol.2, 2001. DOI: 10.1109/CVPR.2001.991023.

[11] L. E. Antonuk. Electronic portal imaging devices: a review and historical perspective of contemporary technologies and research. *Physics in Medicine and Biology*, **47**(6):R31, 2002.

[12] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D point sets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **9**(5):698–700, 1987. DOI: 10.1109/TPAMI.1987.4767965.

[13] Atracsys, LLC. *accuTrack 250 & 500 Datasheet*. Le Mont-sur-Lausanne, Switzerland, 2013. http://www.atracsys.com/web/documents/upload/accuTrack.pdf.

[14] Atracsys LLC, Le Mont-sur-Lausanne, Switzerland, 2015. `www.atracsys.com/`. (accessed on 2015-07-06).

[15] G. H. Barnett, M. E. Linskey, J. R. Adler, J. W. Cozzens, W. A. Friedman, M. P. Heilbrun, L. D. Lunsford, M. Schulder, and A. E. Sloan. Stereotactic radiosurgery–an organized neurosurgery-sanctioned definition. *J. Neurosurg.*, **106**(1):1–5, 2007. DOI: 10.3171/jns.2007.106.1.1.

[16] G. Baroni, M. Riboldi, M. F. Spadea, B. Tagaste, C. Garibaldi, R. Orecchia, and A. Pedotti. Integration of enhanced optical tracking techniques and imaging in igrt. *Journal of Radiation Research*, **48**(Suppl A): A61–A74, 2007. DOI: 10.1269/jrr.48.A61.

[17] A. N. Bashkatov, E. A. Genina, V. I. Kochubey, and V. V. Tuchin. Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *Journal of Physics D Applied Physics*, **38**:2543–2555, 2005. DOI: 10.1088/0022-3727/38/15/004.

[18] A. N. Bashkatov, E. A. Genina, and V. V. Tuchin. Optical properties of skin, subcutaneous, and muscle tissues: A review. *Journal of Innovative Optical Health Sciences*, **04**(01):9–38, 2011. DOI: 10.1142/S1793545811001319.

[19] B. Bellekens, V. Spruyt, and R. B. Maarten Weyn. A survey of rigid 3D pointcloud registration algorithms. In *Fourth International Conference on Ambient Computing, Applications, Services and Technologies, Proceedings*, pp. 8–13. IARA, 2014. ISBN 9781612083568.

[20] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, **2**(1):1–127, 2009. DOI: 10.1561/2200000006.

[21] R. I. Berbeco, H. Mostafavi, G. C. Sharp, and S. B. Jiang. Towards fluoroscopic respiratory gating for lung tumours without radiopaque markers. *Physics in Medicine and Biology*, **50**(19):4481, 2005.

[22] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau. Towards a general multi-view registration technique. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **18**(5):540–547, 1996. DOI: 10.1109/34.494643.

[23] J. Bernier, E. J. Hall, and A. Giaccia. Radiation oncology: a century of achievements. *Nat. Rev. Cancer*, **4**(9):737–747, 2004. DOI: 10.1038/nrc1451.

[24] C. Bert, K. G. Metheany, K. Doppke, and G. T. Y. Chen. A phantom evaluation of a stereo-vision surface imaging system for radiotherapy patient setup. *Medical Physics*, **32**(9):2753–2762, 2005. DOI: http://dx.doi.org/10.1118/1.1984263.

[25] C. Bert, K. G. Metheany, K. P. Doppke, A. G. Taghian, S. N. Powell, and G. T. Chen. Clinical experience with a 3d surface patient setup system for alignment of partial-breast irradiation patients. *International Journal of Radiation Oncology*Biology*Physics*, **64**(4):1265 – 1274, 2006. DOI: http://dx.doi.org/10.1016/j.ijrobp.2005.11.008.

[26] P. Besl and N. D. McKay. A method for registration of 3-D shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **14**(2):239–256, 1992. DOI: 10.1109/34.121791.

[27] O. O. Betti, C. Munari, and R. Rosler. Stereotactic radiosurgery with the linear accelerator: treatment of arteriovenous malformations. *Neurosurgery*, **24**(3):311–321, 1989.

[28] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995. ISBN 0198538642.

[29] G. Blais and M. Levine. Registering multiview range data to create 3D computer objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **17**(8):820–824, 1995. DOI: 10.1109/34.400574.

[30] P. Bloomfield and W. Steiger. Least absolute deviations curve-fitting. *SIAM Journal on Scientific and Statistical Computing*, **1**(2):290–301, 1980. DOI: 10.1137/0901019.

[31] J. Boda-Heggemann, C. Walter, A. Rahn, H. Wertz, I. Loeb, F. Lohr, and F. Wenz. Repositioning accuracy of two different mask systems 3d revisited: Comparison using true 3D/3D matching with cone-beam CT. *International Journal of Radiation Oncology*Biology*Physics*, **66**(5):1568 – 1575, 2006. DOI: http://dx.doi.org/10.1016/j.ijrobp.2006.08.054.

[32] E. Bonilla, K. Chai, and C. Williams. pp. 153–160 in , *Multi-task Gaussian Process Prediction*. NIPS Foundation, 2008.

[33] F. J. Bova, J. M. Buatti, W. A. Friedman, W. M. Mendenhall, C.-C. Yang, and C. Liu. The university of florida frameless high-precision stereotactic radiotherapy system. *International Journal of Radiation Oncology*Biology*Physics*, **38**(4):875 – 882, 1997. DOI: http://dx.doi.org/10.1016/S0360-3016(97)00055-2.

[34] A. Brahme, P. Nyman, and B. Skatt. 4D laser camera for accurate patient positioning, collision avoidance, image fusion and adaptive approaches during diagnostic and therapeutic procedures. *Medical Physics*, **35**(5):1670–1681, 2008. DOI: http://dx.doi.org/10.1118/1.2889720.

[35] Brainlab AG, Feldkirchen,Germany, 2015. `https://www.brainlab.com/`. (accessed on 2015-04-09).

[36] M. C. Branchet, S. Boisnic, C. Frances, and A. M. Robert. Skin thickness changes in normal aging skin. *Gerontology*, **36**(1):28–35, 1990.

[37] I. M. Braverman and E. Fonferko. Studies in cutaneous aging: I. The elastic fiber network. *J. Invest. Dermatol.*, **78**(5):434–443, 1982.

[38] I. M. Braverman and E. Fonferko. Studies in cutaneous aging: II. The microvasculature. *J. Invest. Dermatol.*, **78**(5):444–448, 1982.

[39] R. A. Brown. A computerized tomography-computer graphics approach to stereotaxic localization. *Journal of Neurosurgery*, **50**(6):715–720, 1979. DOI: 10.3171/jns.1979.50.6.0715.

[40] A. Brun, C.-F. Westin, S. Haker, and H. Knutsson. A tensor-like representation for averaging, filtering and interpolation of 3-d object orientation data. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, pp. III–1092–5, 2005. DOI: 10.1109/ICIP.2005.1530586.

[41] J. M. Buatti, F. J. Bova, W. A. Friedman, S. L. Meeks, R. B. M. Jr., J. Mickle, T. L. Ellis, and W. M. Mendenhall. Preliminary experience with frameless stereotactic radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, **42**(3):591 – 599, 1998. DOI: 10.1016/S0360-3016(98)00276-4.

[42] C-Cam Technologies, Vector international, bvba, Leuven, Belgium, 2015. `http://www.c-cam.be/`. (accessed on 2015-04-30).

[43] A. C-Rad. Year end report 2012, Dec, 2012. `http://www.c-rad.info/document/Year_End_2012.pdf`.

[44] C-Rad, AB, Uppsala, Sweden, 2015. `http://www.c-rad.info/`. (accessed on 2015-04-28).

[45] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transaction on*, **PAMI-8**(6):679–698, 1986. DOI: 10.1109/TPAMI.1986.4767851.

[46] CBC AMERICAS Corp., Commack, NY,USA, 2015. `http://www.cbcamerica.com/`. (accessed on 2015-07-06).

[47] Center for Medicare and Medicaid Services. CMS manual system: Publication # 100-20. one-time notification. transmittal 32, Department of Health and Human Services, Washington, DC, 2003. Department of Health and Human Services, Washington, DC.

[48] Centers for Disease Control and Prevention, Atlanta, GA, USA, 2015. `http://www.cdc.gov/`. (accessed on 2015-04-09).

[49] L. I. Cervino, T. Pawlicki, J. D. Lawson, and S. B. Jiang. Frame-less and mask-less cranial stereotactic radiosurgery: a feasibility study. *Physics in Medicine and Biology*, **55**(7):1863, 2010.

[50] L. I. Cervino, N. Detorie, M. Taylor, J. D. Lawson, T. Harry, K. T. Murphy, A. J. Mundt, S. B. Jiang, and T. A. Pawlicki. Initial clinical experience with a frameless and maskless stereotactic radiosurgery treatment. *Practical Radiation Oncology*, **2**(1):54 – 62, 2012. DOI: http://dx.doi.org/10.1016/j.prro.2011.04.005.

[51] K. Chalupka, C. K. I. Williams, and I. Murray. A framework for evaluating approximation methods for gaussian process regression. *J. Mach. Learn. Res.*, **14**(1):333–350, 2013.

[52] M. Chan, J. Yang, Y. Song, C. Burman, P. Chan, and S. Li. Evaluation of imaging performance of major image guidance systems. *Biomed Imaging Interv J*, **7**(2):e11, 2011.

[53] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, **2**:27:1–27:27, 2011.

[54] J. Chang, K. M. Yenice, A. Narayana, and P. H. Gutin. Accuracy and feasibility of cone-beam computed tomography for stereotactic radiosurgery setup. *Medical Physics*, **34**(6):2077–2084, 2007. DOI: http://dx.doi.org/10.1118/1.2731031.

[55] J. Chavaudra and A. Bridier. Définition des volumes en radiothérapie externe : rapports {ICRU} 50 et 62. *Cancer/Radiothérapie*, **5**(5):472–478, 2001. DOI: http://dx.doi.org/10.1016/S1278-3218(01)00117-2.

[56] F. Chen, G. M. Brown, and M. Song. Overview of three-dimensional shape measurement using optical methods. *Optical Engineering*, **39**(1):10–22, 2000. DOI: 10.1117/1.602438.

[57] W. Chen, I. Ayoola, S. Oetomo, and L. Feijs. Non-invasive blood oxygen saturation monitoring for neonates using reflectance pulse oximeter. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, pp. 1530–1535, 2010. DOI: 10.1109/DATE.2010.5457054.

[58] W. Chen, S. B. Oetomo, and L. Feijs. *Neonatal Monitoring Technologies: Design for Integrated Solutions*. IGI Global, 2012. DOI: 10.4018/978-1-4666-0975-4.

[59] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. In *Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on*, pp. 2724–2729, 1991. DOI: 10.1109/ROBOT.1991.132043.

[60] CIVCO Medical Solutions, Inc., Coralville, Iowa, USA, 2013. `http://www.civco.com/`. (accessed on 2015-04-24).

[61] D. J. D'Ambrosio, J. Bayouth, I. J. Chetty, M. K. Buyyounouski, R. A. P. Jr, C. R. Correa, T. J. Dilling, G. E. Franklin, P. Xia, E. E. Harris, and A. Konski. Continuous localization technologies for radiotherapy delivery: Report of the american society for radiation oncology emerging technology committee. *Practical Radiation Oncology*, **2**(2):145 – 150, 2012. DOI: http://dx.doi.org/10.1016/j.prro.2011.10.005.

[62] M. Delannes, N. Daly, J. Bonnet, J. Sabatier, and M. Tremoulet. Fractionated radiotherapy of small inoperable lesions of the brain using a non-invasive stereotactic frame. *International Journal of Radiation Oncology*Biology*Physics*, **21**(3):749 – 755, 1991. DOI: http://dx.doi.org/10.1016/0360-3016(91)90695-Z.

[63] W. Demtroeder. *Experimentalphysik 2*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-33794-2. DOI: 10.1007/978-3-540-33795-9_8.

[64] L. Dietrich, S. Jetter, T. Tuecking, S. Nill, and U. Oelfke. Linac-integrated 4D cone beam CT: first experimental results. *Physics in Medicine and Biology*, **51**(11):2939, 2006.

[65] D. Djajaputra and S. Li. Real-time 3D surface-image-guided beam setup in radiotherapy of breast cancer. *Medical Physics*, **32**(1):65–75, 2005. DOI: http://dx.doi.org/10.1118/1.1828251.

[66] A. R. T. Donders, G. J. van der Heijden, T. Stijnen, and K. G. Moons. Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, **59**(10):1087 – 1091, 2006. DOI: http://dx.doi.org/10.1016/j.jclinepi.2006.01.014.

[67] N. Dong, J. Ge, and Y. Zhang. Four-flux kubelka-munk model of the light reflectance for printing of rough substrate. *Proc. SPIE*, **7241**:72411I–72411I–10, 2009. DOI: 10.1117/12.804402.

[68] A. Doronin. *The unified Monte Carlo model of photon migration in scattering tissue-like media for the needs of Biomedical Optics*. PhD thesis, University of Otago, New Zealand, 2014.

[69] A. Doronin and I. Meglinski. Online object oriented monte carlo computational tool for the needs of biomedical optics. *Biomed. Opt. Express*, **2**(9):2461–2469, 2011. DOI: 10.1364/BOE.2.002461.

[70] A. Doronin, I. Fine, and I. Meglinski. Assessment of the calibration curve for transmittance pulse-oximetry. *Laser Physics*, **21**:1972–1977, 2011.

[71] L. Douadi, M.-J. Aldon, and A. Crosnier. Pair-wise registration of 3D/color data sets with ICP. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pp. 663–668, 2006. DOI: 10.1109/IROS.2006.282551.

[72] S. F. Dunbar, N. J. Tarbell, H. M. Kooy, E. A. III, P. M. Black, P. D. Barnes, L. Goumnerova, R. M. Scott, S. L. Pomeroy, B. L. Vally, S. E. Sallan, and J. S. Loeffler. Stereotactic radiotherapy for pediatric and adult brain tumors: Preliminary report. *International Journal of Radiation Oncology*Biology*Physics*, **30**(3):531–539, 1994. DOI: http://dx.doi.org/10.1016/0360-3016(92)90938-E.

[73] R. Dürichen, M. Pimentel, L. Clifton, A. Schweikard, and D. Clifton. Multitask gaussian processes for multivariate physiological time-series analysis. *Biomedical Engineering, IEEE Transactions on*, **62**(1):314–322, 2015. DOI: 10.1109/TBME.2014.2351376.

[74] D. W. Ebert, C. Roberts, S. K. Farrar, W. M. Johnston, A. S. Litsky, and A. L. Bertone. Articular cartilage optical properties in the spectral range 300-850. *Journal of Biomedical Optics*, **3**(3):326–333, 1998. DOI: 10.1117/1.429893.

[75] A. A. El-Gayed, A. Bel, R. Vijlbrief, H. Bartelink, and J. V. Lebesque. Time trend of patient setup deviations during pelvic irradiation using electronic portal imaging. *Radiotherapy and Oncology*, **26**(2):162 – 171, 1993. DOI: http://dx.doi.org/10.1016/0167-8140(93)90098-S.

[76] Elekta AB (publ)., Stockholm, Sweden, 2015. `http://www.elekta.com`. (visited on 2015-04-09).

[77] Elekta AB (publ.). C-RAD sentinel selected as elekta preferred solution for intra fraction motion detection [press release], Palo Alto, CA, USA, July 21, 2009. `http://www.elekta.com/pressreleases/438204/`. (accessed on 2015-04-25).

[78] M. Engelsman, S. J. Rosenthal, S. L. Michaud, J. A. Adams, R. J. Schneider, S. G. Bradley, J. B. Flanz, and H. M. Kooy. Intra- and interfractional patient motion for a variety of immobilization devices. *Medical Physics*, **32**(11):3468–3474, 2005. DOI: http://dx.doi.org/10.1118/1.2089507.

[79] F. Ernst, R. Bruder, T. Wissel, P. Stüber, B. Wagner, and A. Schweikard. Real time contact-free and non-invasive tracking of the human skull—first light and initial validation. In *Applications of Digital Image Processing XXXVI*, pp. 88561G–1 – 88561G–8. Published in *Proceedings of SPIE*, **8856**:88561G–1 – 88561G–8, 2013. DOI: 10.1117/12.2024851.

[80] F. Ernst, R. Bruder, T. Wissel, P. Stüber, B. Wagner, and A. Schweikard. Measuring cranial soft tissue thickness with MRI or pressure-compensated tracked ultrasound. *British Journal of Medicine and Medical Research*, **4**(4):937–948, 2014.

[81] C. Escoffier, J. de Rigal, A. Rochefort, R. Vasselet, J. L. Leveque, and P. G. Agache. Age-related mechanical properties of human skin: an in vivo study. *J. Invest. Dermatol.*, **93**(3):353–357, 1989.

[82] K. Farahani, U. Sinha, S. Sinha, L. C.-L. Chiu, and R. B. Lufkin. Effect of field strength on susceptibility artifacts in magnetic resonance imaging. *Computerized Medical Imaging and Graphics*, **14**(6):409 – 413, 1990. DOI: 10.1016/0895-6111(90)90040-I.

[83] A. Farhangfar, L. Kurgan, and J. Dy. Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, **41**(12):3692 – 3705, 2008. DOI: http://dx.doi.org/10.1016/j.patcog.2008.05.019.

[84] S. K. Farrar, C. Roberts, W. M. Johnston, and P. A. Weber. Optical properties of human trabecular meshwork in the visible and near-infrared region. *Lasers in Surgery and Medicine*, **25**(4):348–362, 1999. DOI: 10.1002/(SICI)1096-9101(1999)25:4⟨348::AID-LSM9⟩3.0.CO;2-2.

[85] N. A. Fenske and C. W. Lober. Structural and functional changes of normal aging skin. *Journal of the American Academy of Dermatology*, **15**(4, Part 1):571 – 585, 1986. DOI: http://dx.doi.org/10.1016/S0190-9622(86)70208-9.

[86] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. Parkin, D. Forman, and F. Bray. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancerbase No. 11, Lyon, France: International Agency for Research on Cancer, 2013. `http://globocan.iarc.fr`. (accessed on 2015-04-13).

[87] V. Feygelman, L. Walker, P. Chinnaiyan, and K. Forster. Simulation of intrafraction motion and overall geometrical accuracy of a frameless intracranial radiosurgery process. *J Appl Clin Med Phys*, **9**(4):2828, 2008.

[88] M. Firbank, M. Hiraoka, M. Essenpreis, and D. T. Delpy. Measurement of the optical properties of the skull in the wavelength range 650-950 nm. *Physics in Medicine and Biology*, **38**(4):503, 1993.

[89] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, **24**(6):381–395, 1981. DOI: 10.1145/358669.358692.

[90] R. A. Fisher. On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**:3–32, 1921.

[91] J. M. Fitzpatrick. Fiducial registration error and target registration error are uncorrelated. *Proc. SPIE*, **7261**:726102–726102–12, 2009. DOI: 10.1117/12.813601.

[92] J. M. Fitzpatrick and J. B. West. The distribution of target registration error in rigid-body point-based registration. *IEEE Trans Med Imaging*, **20**(9):917–927, 2001.

[93] T. B. Fitzpatrick. Soleil et peau. *Journal de Médecine Esthétique*, **2**:33–34, 1975.

[94] T. B. Fitzpatrick. Ultraviolet-induced pigmentary changes: Benefits and hazards, pp. 25–38 in H. Hönigsmann and G. Stingl, editors, *Therapeutic Photomedicine*, volume 15 of *Current Problems in Dermatology*. Karger, Basel, 1986. ISBN 978-3-8055-4151-0.

[95] T. B. Fitzpatrick. The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatology*, **124**(6):869–871, 1988. DOI: 10.1001/archderm.1988.01670060015008.

[96] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, London, 2006.

[97] M. Fuss, B. J. Salter, D. Cheek, A. Sadeghi, J. M. Hevezi, and T. S. Herman. Repositioning accuracy of a commercially available thermoplastic mask system. *Radiotherapy and Oncology*, **71**(3):339 – 345, 2004. DOI: 10.1016/j.radonc.2004.03.003.

[98] S. Gao, W. Du, P. Balter, P. Munro, and A. Jeung. Evaluation of IsoCal geometric calibration system for Varian linacs equipped with on-board imager and electronic portal imaging device imaging systems. *J Appl Clin Med Phys*, **15**(3):4688, 2014. DOI: 10.1120/jacmp.v15i3.4688.

[99] E. A. Genina, A. N. Bashkatov, and V. V. Tuchin. Optical clearing of human cranial bone by administration of immersion agents. *Advances in Optical Technologies*, pp. 616311–616311–9, 2006. DOI: 10.1117/12.697308.

[100] E. A. Genina, A. N. Bashkatov, and V. V. Tuchin. Optical Clearing of Cranial Bone. *Advances in Optical Technologies*, **2008**:1–9, 2008. DOI: 10.1155/2008/267867.

[101] German Cancer Society e.V., Berlin, Germany, 2015. `http://www.krebsgesellschaft.de`. (accessed on 2015-04-13).

[102] T. Gevaert, D. Verellen, B. Engels, T. Depuydt, K. Heuninckx, K. Tournel, M. Duchateau, T. Reynders, and M. D. Ridder. Clinical evaluation of a robotic 6-degree of freedom treatment couch for frameless radiosurgery. *International Journal of Radiation Oncology*Biology*Physics*, **83**(1):467–474, 2012. DOI: http://dx.doi.org/10.1016/j.ijrobp.2011.05.048.

[103] L. Gilbeau, M. Octave-Prignot, T. Loncol, L. Renard, P. Scalliet, and V. Gregoire. Comparison of setup accuracy of three different thermoplastic masks for the treatment of brain and head and neck tumors. *Radiotherapy and Oncology*, **58**(2):155 – 162, 2001. DOI: http://dx.doi.org/10.1016/S0167-8140(00)00280-2.

[104] S. Gill, D. Thomas, A. Warrington, and M. Brada. Relocatable frame for stereotactic external beam radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, **20**(3):599 – 603, 1991. DOI: http://dx.doi.org/10.1016/0360-3016(91)90076-G.

[105] G. Godin, M. Rioux, and R. Baribeau. Three-dimensional registration using range and intensity information. *Proc. SPIE*, **2350**:279–290, 1994. DOI: 10.1117/12.189139.

[106] L. A. Goldsmith, S. I. Katz, B. A. Gilchrest, A. S. Paller, D. J. Leffell, and K. Wolff. *Fitzpatrick's Dermatology in General Medicine*. McGraw-Hill Professional, New York, 8th edition, 2012. ISBN 978-0071669047.

[107] O. Gopan and Q. Wu. Evaluation of the accuracy of a 3D surface imaging system for patient setup in head and neck cancer radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, **84**(2):547 – 552, 2012. DOI: http://dx.doi.org/10.1016/j.ijrobp.2011.12.004.

[108] J. D. Graham, A. P. Warrington, S. S. Gill, and M. Brada. A non-invasive, relocatable stereotactic frame for fractionated radiotherapy and multiple imaging. *Radiotherapy and Oncology*, **21**(1):60–62, 1991. DOI: http://dx.doi.org/10.1016/0167-8140(91)90342-E.

[109] H. Gray and W. H. Lewis. Anatomy of the human body, Philadelphia, PA, 1918. `http://www.bartleby.com/107/`. (visited on 2015-04-09).

[110] E. Grusell, A. Montelius, K. R. Russell, E. Blomquist, L. Pellettieri, A. Lilja, U. Moström, and P. Jakobsson. Patient positioning for fractionated precision radiation treatment of targets in the head using fiducial markers. *Radiotherapy and Oncology*, **33**(1):68 – 72, 1994. DOI: http://dx.doi.org/10.1016/0167-8140(94)90088-4.

[111] M. Guckenberger, K. Baier, I. Guenther, A. Richter, J. Wilbert, O. Sauer, D. Vordermark, and M. Flentje. Reliability of the bony anatomy in image-guided stereotactic radiotherapy of brain metastases. *International Journal of Radiation Oncology*Biology*Physics*, **69**(1):294 – 301, 2007. DOI: http://dx.doi.org/10.1016/j.ijrobp.2007.05.030.

[112] Z. Guo, S. Wan, K. Kim, and C. Kosaraju. Comparing diffusion approximation with radiation transfer analysis for light transport in tissues. *Optical Review*, **10**(5):415–421, 2003. DOI: 10.1007/s10043-003-0415-y.

[113] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, **3**: 1157–1182, 2003.

[114] R. Y. Ha, K. Nojima, W. P. Adams, and S. A. Brown. Analysis of facial skin thickness: defining the relative thickness index. *Plast. Reconstr. Surg.*, **115**(6):1769–1773, 2005.

[115] J. V. Hajnal, D. L. Hill, and D. J. Hawkes. *Medical Image Registration*. The Biomedical Engineering Series. CRC Press, 1st edition, 2001.

[116] T. Hamaoka, T. Katsumura, N. Murase, S. Nishio, T. Osada, T. Sako, H. Higuchi, Y. Kurosawa, T. Shimomitsu, M. Miwa, and B. Chance. Quantification of ischemic muscle deoxygenation by near infrared time-resolved spectroscopy. *J Biomed Opt*, **5**(1):102–105, 2000.

[117] M. I. Hariz and A. T. Eriksson. Reproducibility of repeated mountings of a noninvasive CT/MRI stereoadapter. *Appl Neurophysiol*, **49**(6):336–347, 1986.

[118] M. I. Hariz and H. Fodstad. Stereotactic localization of small subcortical brain tumors for open surgery. *Surgical Neurology*, **28**(5):345 – 350, 1987. DOI: http://dx.doi.org/10.1016/0090-3019(87)90056-5.

[119] M. I. Hariz, R. Henriksson, P.-O. Laefroth, L. V. Laitinen, and N.-E. Soeterborg. A non-invasive method for fractionated stereotactic irradiation of brain tumors with linear accelerator. *Radiotherapy and Oncology*, **17**(1):57–72, 1990. DOI: http://dx.doi.org/10.1016/0167-8140(90)90049-3.

[120] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2nd edition, 2003. ISBN 0521540518.

[121] G. H. Hartmann, W. Schlegel, V. Sturm, B. Kober, O. Pastyr, and W. J. Lorenz. Cerebral radiation surgery using moving field irradiation at a linear accelerator facility. *International Journal of Radiation Oncology*Biology*Physics*, **11**(6):1185 – 1192, 1985. DOI: http://dx.doi.org/10.1016/0360-3016(85)90068-9.

[122] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning*. Springer, 2nd edition, 2003.

[123] M. P. Heilbrun, T. S. Roberts, M. L. J. Apuzzo, J. Trent H. Wells, and J. K. Sabshin. Preliminary experience with brown-roberts-wells (brw) computerized tomography stereotaxic guidance system. *Journal of Neurosurgery*, **59**(2):217–222, 1983. DOI: 10.3171/jns.1983.59.2.0217.

[124] L. Henyey and J. L. Greenstein. Diffuse radiation in the Galaxy. *Astrophysical Journal*, **93**:70–83, 1941. DOI: 10.1086/144246.

[125] G. E. Hinton, S. Osindero, and Y. W. Teh. A fast learning algorithm for deep belief networks. *Neural Computation*, **18**(7):1527–1554, 2006. DOI: doi:10.1162/neco.2006.18.7.1527.

[126] A. E. Hoerl and R. W. Kennard. Ridge regression: Applications to nonorthogonal problems. *Technometrics*, **12**(1):69–82, 1970.

[127] H. Hong and B. Lee. Colored point cloud registration with improved hue-assisted normal distributions transform. *International Journal of Computer Theory and Engineering*, **8**(1):63–68, 2014.

[128] B. Horn. Extended gaussian images. *Proceedings of the IEEE*, **72**(12):1671–1686, 1984. DOI: 10.1109/PROC.1984.13073.

[129] P. V. Houdek, J. V. Fayos, J. M. Van Buren, and M. S. Ginsberg. Stereotaxic radiotherapy technique for small intracranial lesions. *Medical Physics*, **12**(4):469–472, 1985. DOI: http://dx.doi.org/10.1118/1.595673.

[130] B. Huhle, M. Magnusson, W. Strasser, and A. Lilienthal. Registration of colored 3D point clouds with a kernel-based extension to the normal distributions transform. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pp. 4025–4030, 2008. DOI: 10.1109/ROBOT.2008.4543829.

[131] P. Huke, R. Klattenhoff, C. von Kopylow, and R. Bergmann. Novel trends in optical non-destructive testing methods. *Journal of the European Optical Society - Rapid publications*, **8**(0), 2013.

[132] D. Hyde, F. Lochray, R. Korol, M. Davidson, C. S. Wong, L. Ma, and A. Sahgal. Spine stereotactic body radiotherapy utilizing cone-beam CT image-guidance with a robotic couch: Intrafraction motion analysis accounting for all six degrees of freedom. *International Journal of Radiation Oncology*Biology*Physics*, **82**(3):e555 – e562, 2012. DOI: http://dx.doi.org/10.1016/j.ijrobp.2011.06.1980.

[133] ICRU-50. ICRU Report 50 International Commission on Radiation Units and Measurements. Technical report, Bethesda, Md, USA, 1993.

[134] ICRU-62. ICRU Report 62 International Commission on Radiation Units and Measurements. Technical report, Bethesda, Md, USA, 1999.

[135] IDS Imaging Development Systems GmbH, Obersulm, Germany, 2015. `http://de.ids-imaging.com/`. (accessed on 2015-07-06).

[136] IDS Imaging Development Systems GmbH. *uEye-Handbuch 4.60*. Obersulm, Germany, 2015.

[137] IOP Publishing. First order of C-RAD catalyst from UK customer [press release], Bristol, UK, April 15, 2014. `http://medicalphysicsweb.org/cws/article/newsfeed/58254`. (accessed on 2015-04-25).

[138] IOP Publishing. C-RAD receives estonian order for two catalyst and two sentinel systems [press release], Bristol, UK, April 2, 2015. `http://medicalphysicsweb.org/cws/article/newsfeed/60760`. (accessed on 2015-04-25).

[139] IOP Publishing. German radiation oncology clinic orders three C-RAD systems [press release], Bristol, UK, Dec 16, 2014. `http://medicalphysicsweb.org/cws/article/newsfeed/59658`. (accessed on 2015-04-25).

[140] IOP Publishing. C-RAD secures order for two catalyst systems and a sentinel system [press release], Bristol, UK, March 24, 2015. `http://medicalphysicsweb.org/cws/article/newsfeed/60643`. (accessed on 2015-04-25).

[141] J. Irudayaraj and C. Reh. *Nondestructive Testing of Food Quality*. Wiley-Blackwell, 2007.

[142] N. G. Jablonski and G. Chaplin. The evolution of human skin coloration. *Journal of Human Evolution*, **39**(1):57 – 106, 2000. DOI: http://dx.doi.org/10.1006/jhev.2000.0403.

[143] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *International Journal of Radiation Oncology*Biology*Physics*, **53**(5):1337 – 1349, 2002. DOI: http://dx.doi.org/10.1016/S0360-3016(02)02884-5.

[144] H. Jiang and K. D. Paulsen. Finite-element-based higher order diffusion approximation of light propagation in tissues. *Proc. SPIE*, **2389**:608–614, 1995. DOI: 10.1117/12.210007.

[145] J.-Y. Jin, F.-F. Yin, S. E. Tenn, P. M. Medin, and T. D. Solberg. Use of the BrainLAB ExacTrac X-ray 6D system in image-guided radiotherapy. *Medical Dosimetry*, **33**(2):124 – 134, 2008. DOI: http://dx.doi.org/10.1016/j.meddos.2008.02.005.

[146] A. E. Johnson and S. B. Kang. Registration and integration of textured 3D data. *Image and Vision Computing*, **17**(2):135 – 147, 1999. DOI: http://dx.doi.org/10.1016/S0262-8856(98)00117-6.

[147] M. Kaas, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, **1**:321–331, 1988. DOI: 10.1007/BF00133570.

[148] J. A. Kalapurakal, Z. Ilahi, A. G. Kepka, T. Bista, S. Goldman, T. Tomita, and M. H. Marymont. Repositioning accuracy with the Laitinen frame for fractionated stereotactic radiation therapy in adult and pediatric brain tumors: preliminary report. *Radiology*, **218**(1):157–161, 2001.

[149] R. Kamath, T. C. Ryken, S. L. Meeks, E. C. Pennington, J. Ritchie, and J. M. Buatti. Initial clinical experience with frameless radiosurgery for patients with intracranial metastases. *International Journal of Radiation Oncology*Biology*Physics*, **61**(5):1467 – 1472, 2005. DOI: http://dx.doi.org/10.1016/j.ijrobp.2004.08.021.

[150] H. Kang, D. M. Lovelock, E. D. Yorke, S. Kriminski, N. Lee, and H. I. Amols. Accurate positioning for head and neck cancer patients using 2D and 3D image guidance. *J Appl Clin Med Phys*, **12**(1):3270, 2011.

[151] K. I. Kauweloa, D. Ruan, J. C. Park, A. Sandhu, G. Ya Kim, T. Pawlicki, W. Tyler Watkins, B. Song, and W. Y. Song. GateCT surface tracking system for respiratory signal reconstruction in 4DCT imaging. *Medical Physics*, **39**(1):492–502, 2012. DOI: http://dx.doi.org/10.1118/1.3671941.

[152] P. J. Keall, H. Cattell, D. Pokhrel, S. Dieterich, K. H. Wong, M. J. Murphy, S. S. Vedam, K. Wijesooriya, and R. Mohan. Geometric accuracy of a real-time target tracking system with dynamic multileaf collimator tracking system. *International Journal of Radiation Oncology*Biology*Physics*, **65**(5):1579 –1584, 2006. DOI: http://dx.doi.org/10.1016/j.ijrobp.2006.04.038.

[153] W. Kilby, J. R. Dooley, G. Kuduvalli, S. Sayeh, and C. R. Maurer. The cyberknife robotic radiosurgery system in 2010. *Technology in Cancer Research and Treatment*, **9**(5):433–452, 2010.

[154] K.-H. Kim, M.-J. Cho, J.-S. Kim, J.-S. Kim, C.-J. Song, S.-H. Song, S.-H. Kim, L. Myers, and Y.-E. Kim. Isocenter accuracy in frameless stereotactic radiotherapy using implanted fiducials. *International Journal of Radiation Oncology*Biology*Physics*, **56**(1):266 – 273, 2003. DOI: http://dx.doi.org/10.1016/S0360-3016(03)00088-9.

[155] S. Kim, H. C. Akpati, J. G. Li, C. R. Liu, R. J. Amdur, and J. R. Palta. An immobilization system for claustrophobic patients in head-and-neck intensity-modulated radiation therapy. *International Journal of Radiation Oncology*Biology*Physics*, **59**(5):1531 – 1539, 2004. DOI: http://dx.doi.org/10.1016/j.ijrobp.2004.01.025.

[156] Y. Kim, R. Li, Y. H. Na, R. Lee, and L. Xing. Accuracy of surface registration compared to conventional volumetric registration in patient positioning for head-and-neck radiotherapy: A simulation study using patient data. *Medical Physics*, **41**(12):121701, 2014. DOI: http://dx.doi.org/10.1118/1.4898103.

[157] D. M. Knize. *The Forehead and Temporal Fossa: Anatomy and Technique*. Lippincott Williams & Wilkins, 1st edition, 2001.

[158] Koninklijke Philips N.V., Amsterdam, The Netherlands, 2015. `http://www.philips.de/`. (accessed on 2015-07-31).

[159] H. M. Kooy, L. A. Nedzi, J. S. Loeffler, E. A. III, C.-W. Cheng, E. G. Mannarino, E. J. Holupka, and R. L. Siddon. Treatment planning for stereotactic radiosurgery of intra-cranial lesions. *International Journal of Radiation Oncology*Biology*Physics*, **21**(3):683–693, 1991. DOI: http://dx.doi.org/10.1016/0360-3016(91)90687-Y.

[160] H. M. Kooy, S. F. Dunbar, N. J. Tarbell, E. Mannarino, N. Ferarro, S. Shusterman, M. Bellerive, L. Finn, C. V. McDonough, and J. S. Loeffler. Adaptation and verification of the relocatable gill-thomas-cosman frame in stereotactic radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, **30**(3):685 – 691, 1994. DOI: 10.1016/0360-3016(92)90956-I.

[161] P. Kubelka and F. Munk. Ein Beitrag zur Optik der Farbanstriche. *Zeitschrift für technische Physik*, **12**: 593–601, 1931.

[162] P. Kurzweil, B. Frenzel, and F. Gebhard. *Physik Formelsammlung*. Springer Vieweg, 3rd edition, 2014.

[163] L. V. Laitinen, B. Liliequist, M. Fagerlund, and A. Eriksson. An adapter for computed tomography-guided stereotaxis. *Surgical Neurology*, **23**(6):559 – 566, 1985. DOI: http://dx.doi.org/10.1016/0090-3019(85)90003-5.

[164] LAP GmbH Laser Applikationen, Lueneburg, Germany, 2015. `http://www.lap-laser.com/`. (accessed on 2015-04-30).

[165] Laserwinkel, Sliedrecht, Netherlands, 2015. `http://www.laserwinkel.nl/`. (accessed on 2015-07-06).

[166] C. Lee, K. M. Langen, W. Lu, J. Haimerl, E. Schnarr, K. J. Ruchala, G. H. Olivera, S. L. Meeks, P. A. Kupelian, T. D. Shellenberger, and R. R. Manon. Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily megavoltage computed tomography and deformable image registration. *International Journal of Radiation Oncology*Biology*Physics*, **71**(5):1563 – 1571, 2008. DOI: http://dx.doi.org/10.1016/j.ijrobp.2008.04.013.

[167] N. Lee, C. Chuang, J. M. Quivey, T. L. Phillips, P. Akazawa, L. J. Verhey, and P. Xia. Skin toxicity due to intensity-modulated radiotherapy for head-and-neck carcinoma. *International Journal of Radiation Oncology*Biology*Physics*, **53**(3):630 – 637, 2002. DOI: http://dx.doi.org/10.1016/S0360-3016(02)02756-6.

[168] L. Leksell. The stereotaxic method and radiosurgery of the brain. *Acta Chir Scand*, **102**(4):316–319, 1951.

[169] L. Leksell, C. Lindquist, J. R. Adler, D. Leksell, B. Jernberg, and L. Steiner. A new fixation device for the Leksell stereotaxic system. *Journal of Neurosurgery*, **66**(4):626–629, 1987. DOI: 10.3171/jns.1987.66.4.0626.

[170] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller. Introduction to machine learning for brain imaging. *NeuroImage*, **56**(2):387 – 399, 2011. DOI: http://dx.doi.org/10.1016/j.neuroimage.2010.11.004.

[171] L. B. Leybovich, A. Sethi, N. Dogan, E. Melian, M. Krasin, and B. Emami. An immobilization and localization technique for SRT and IMRT of intracranial tumors. *J Appl Clin Med Phys*, **3**(4):317–322, 2002.

[172] G. Li, A. Ballangrud, L. C. Kuo, H. Kang, A. Kirov, M. Lovelock, Y. Yamada, J. Mechalakos, and H. Amols. Motion monitoring for cranial frameless stereotactic radiosurgery using video-based three-dimensional optical surface imaging. *Medical Physics*, **38**(7):3981–3994, 2011. DOI: http://dx.doi.org/10.1118/1.3596526.

[173] G. Li, D. M. Lovelock, J. Mechalakos, S. Rao, C. Della-Biancia, H. Amols, and N. Lee. Migration from full-head mask to "open-face" mask for immobilization of patients with head and neck cancer. *Journal of Applied Clinical Medical Physics*, **14**(5), 2013.

[174] S. Li, D. Liu, G. Yin, P. Zhuang, and J. Geng. Real-time 3D-surface-guided head refixation useful for fractionated stereotactic radiotherapy. *Medical Physics*, **33**(2):492–503, 2006. DOI: http://dx.doi.org/10.1118/1.2150778.

[175] S. Li, L. R. Kleinberg, D. Rigamonti, M. D. Wharam, A. Rashid, J. Jackson, D. Djajaputra, S. He, T. Creasey, and T. L. DeWeese. Clinical results of a pilot study on stereovision-guided stereotactic radiotherapy and intensity modulated radiotherapy. *Technol. Cancer Res. Treat.*, **9**(6):603–617, 2010.

[176] N. Linthout, D. Verellen, K. Tournel, and G. Storme. Six dimensional analysis with daily stereoscopic x-ray imaging of intrafraction patient motion in head and neck treatments using five points fixation masks. *Medical Physics*, **33**(2):504–513, 2006. DOI: http://dx.doi.org/10.1118/1.2165417.

[177] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.*, **114**(2):97–109, 2007. DOI: 10.1007/s00401-007-0243-4.

[178] M. Lualdi, A. Colombo, B. Farina, S. Tomatis, and R. Marchesini. A phantom with tissue-like optical properties in the visible and near infrared for use in photomedicine. *Lasers Surg Med*, **28**(3):237–243, 2001.

[179] W. Lutz, K. R. Winston, and N. Maleki. A system for stereotactic radiosurgery with a linear accelerator. *International Journal of Radiation Oncology*Biology*Physics*, **14**(2):373 – 381, 1988. DOI: http://dx.doi.org/10.1016/0360-3016(88)90446-4.

[180] C.-M. C. Ma and K. Paskalev. In-room CT techniques for image-guided radiation therapy. *Medical Dosimetry*, **31**(1):30 – 39, 2006. DOI: http://dx.doi.org/10.1016/j.meddos.2005.12.010.

[181] J. Ma, J. Theiler, and S. Perkins. Accurate On-line Support Vector Regression. *Neural Computation*, **15** (11):2683–2703, 2003.

[182] J. Ma, Z. Chang, Z. Wang, Q. J. Wu, J. P. Kirkpatrick, and F.-F. Yin. ExacTrac X-ray 6 degree-of-freedom image-guidance for intracranial non-invasive stereotactic radiotherapy: Comparison with kilo-voltage cone-beam CT. *Radiotherapy and Oncology*, **93**(3):602 – 608, 2009. DOI: http://dx.doi.org/10.1016/j.radonc.2009.09.009.

[183] J. T. MacKenzie, M. B. Podgorsak, and D. Moreland. Validity of stereotactic frame localization during radiosurgery after one fixation pin removal. *Journal of Neurosurgery*, **97**(supplement 5):539–541, 2002. DOI: 10.3171/jns.2002.97.supplement5.0539.

[184] J. Maclaren, M. Aksoy, and R. Bammer. Contact-free physiological monitoring using a markerless optical system. *Magnetic Resonance in Medicine*, **74**(2):571–577, 2015. DOI: 10.1002/mrm.25781.

[185] B. Maheu, J. N. Letoulouzan, and G. Gouesbet. Four-flux models to solve the scattering transfer equation in terms of Lorenz-Mie parameters. *Appl. Opt.*, **23**(19):3353–3362, 1984. DOI: 10.1364/AO.23.003353.

[186] A. Makadia, A. I. Patterson, and K. Daniilidis. Fully automatic registration of 3D point clouds. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, pp. 1297–1304, Washington, DC, USA. IEEE Computer Society. `http://dx.doi.org/10.1109/CVPR.2006.122`. Published in *CVPR '06*, pp. 1297–1304, 2006. ISBN 0-7695-2597-0. DOI: 10.1109/CVPR.2006.122.

[187] L. Masi, F. Casamassima, C. Polli, C. Menichelli, I. Bonucci, and C. Cavedon. Cone beam CT image guidance for intracranial stereotactic treatments: Comparison with a frame guided set-up. *International Journal of Radiation Oncology*Biology*Physics*, **71**(3):926 – 933, 2008. DOI: http://dx.doi.org/10.1016/j.ijrobp.2008.03.006.

[188] Mathworks, Inc., Natick, MA, USA, 2015. `http://www.mathworks.de`. (accessed on 2015-09-23).

[189] S. L. Meeks, F. J. Bova, T. H. Wagner, J. M. Buatti, W. A. Friedman, and K. D. Foote. Image localization for frameless stereotactic radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, **46** (5):1291 – 1299, 2000. DOI: http://dx.doi.org/10.1016/S0360-3016(99)00536-2.

[190] S. L. Meeks, W. A. Tome, T. R. Willoughby, P. A. Kupelian, T. H. Wagner, J. M. Buatti, and F. J. Bova. Optically guided patient positioning techniques. *Seminars in Radiation Oncology*, **15**(3):192 – 201, 2005. DOI: http://dx.doi.org/10.1016/j.semradonc.2005.01.004.

[191] I. Meglinski and S. Matcher. Computer simulation of the skin reflectance spectra. *Computer Methods and Programs in Biomedicine*, **70**(2):179 – 186, 2003. DOI: 10.1016/S0169-2607(02)00099-8.

[192] I. Meglinski, A. Doronin, A. N. Bashkatov, E. A. Genina, and V. V. Tuchin. Dermal component-based optical modeling of skin translucency: Impact on skin. *Computational Biophysics of the Skin, CRC Press Book*, art. 1, 2014. DOI: 10.1201/b17205-4.

[193] I. V. Meglinski and S. J. Matcher. Modelling the sampling volume for skin blood oxygenation measurements. *Med Biol Eng Comput*, **39**(1):44–50, 2001.

[194] H. Men and K. Pochiraju. Hue-assisted automatic registration of color point clouds. *Journal of Computational Design and Engineering*, **1**(4):223–232, 2014. DOI: http://dx.doi.org/10.7315/JCDE.2014.022.

[195] H. Men, B. Gebre, and K. Pochiraju. Color point cloud registration with 4D ICP algorithm. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pp. 1511–1516, 2011. DOI: 10.1109/ICRA.2011.5980407.

[196] J. Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **209**(441-458):415–446, 1909. DOI: 10.1098/rsta.1909.0016.

[197] J. Meyer, J. Wilbert, K. Baier, M. Guckenberger, A. Richter, O. Sauer, and M. Flentje. Positioning accuracy of cone-beam computed tomography in combination with a hexapod robot treatment table. *International Journal of Radiation Oncology*Biology*Physics*, **67**(4):1220 – 1228, 2007. DOI: http://dx.doi.org/10.1016/j.ijrobp.2006.11.010.

[198] G. Minniti, M. Valeriani, E. Clarke, M. D'Arienzo, M. Ciotti, R. Montagnoli, F. Saporetti, and R. M. Enrici. Fractionated stereotactic radiotherapy for skull base tumors: analysis of treatment accuracy using a stereotactic mask fixation system. *Radiat Oncol*, **5**:1, 2010.

[199] M. Mogensen, L. Thrane, T. M. Joergensen, P. E. Andersen, and G. B. Jemec. Optical coherence tomography for imaging of skin and skin diseases. *Semin Cutan Med Surg*, **28**(3):196–202, 2009. DOI: 10.1016/j.sder.2009.07.002.

[200] W. Montagna and K. Carlisle. Structural changes in aging human skin. *J. Invest. Dermatol.*, **73**(1):47–53, 1979.

[201] T. Moser, S. Fleischhacker, K. Schubert, G. Sroka-Perez, and C. P. Karger. Technical performance of a commercial laser surface scanning system for patient setup correction in radiotherapy. *Physica Medica*, **27**(4):224 – 232, 2011. DOI: http://dx.doi.org/10.1016/j.ejmp.2010.10.005.

[202] T. Moser, G. Habl, M. Uhl, K. Schubert, G. Sroka-Perez, J. Debus, K. Herfarth, and C. P. Karger. Clinical evaluation of a laser surface scanning system in 120 patients for improving daily setup accuracy in fractionated radiation therapy. *International Journal of Radiation Oncology*Biology*Physics*, **85**(3):846 – 853, 2013. DOI: http://dx.doi.org/10.1016/j.ijrobp.2012.05.026.

[203] F. Mundinger, W. Birg, and C. Ostertag. Treatment of small cerebral gliomas with CT-aided stereotaxic curietherapy. In S. Wende, editor, *Proceedings of the XI. Symposium Neuroradiologicum*, pp. 564–567. Springer Berlin Heidelberg, 1978. ISBN 978-3-540-08782-3. DOI: 10.1007/978-3-642-66959-0_178.

[204] M. J. Murphy, J. Balter, S. Balter, J. A. BenComo, I. J. Das, S. B. Jiang, C. M. Ma, G. H. Olivera, R. F. Rodebaugh, K. J. Ruchala, H. Shirato, and F. F. Yin. The management of imaging dose during image-guided radiotherapy: report of the AAPM Task Group 75. *Med Phys*, **34**(10):4041–4063, 2007.

[205] National Cancer Institute (NCI) as the National Institutes of Health, Bethesda, MD, USA, 2015. `http://www.cancer.gov/`. (accessed on 2015-04-09).

[206] NDI, Inc., Waterloo, Canada, 2015. `http://www.ndigital.com`. (accessed on 2015-04-29).

[207] J. H. Nobbs. Kubelka-munk theory and the prediction of reflectance. *Review of Progress in Coloration and Related Topics*, **15**(1):66–75, 1985. DOI: 10.1111/j.1478-4408.1985.tb03737.x.

[208] J. Nocedal and S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2nd edition, 2006.

[209] J. Q. nonero Candela, C. Rasmussen, and C. Williams. Approximation methods for Gaussian process regression, pp. 203–223 in , *Large Scale Learning Machines*. MIT Press, 2007.

[210] U. Oelfke, T. Tücking, S. Nill, A. Seeber, B. Hesse, P. Huber, and C. Thilmann. Linac-integrated kV-cone beam CT: Technical features and first applications. *Medical Dosimetry*, **31**(1):62 – 70, 2006. DOI: http://dx.doi.org/10.1016/j.meddos.2005.12.008.

[211] L. A. Olsen, C. G. Robinson, G. R. He, H. O. Wooten, S. Yaddanapudi, S. Mutic, D. Yang, and K. L. Moore. Automated radiation therapy treatment plan workflow using a commercial application programming interface. *Practical Radiation Oncology*, **4**(6):358 – 367, 2014. DOI: http://dx.doi.org/10.1016/j.prro.2013.11.007.

[212] Orfit Industries n.v., Wijnegem, Belgium, 2015. `http://www.orfit.com/`. (accessed on 2015-04-23).

[213] S. Pallotta, L. Marrazzo, M. Ceroti, P. Silli, and M. Bucciolini. A phantom evaluation of sentinel, a commercial laser/camera surface imaging system for patient setup verification in radiotherapy. *Medical Physics*, **39**(2):706–712, 2012. DOI: http://dx.doi.org/10.1118/1.3675973.

[214] S. Pallotta, G. Simontacchi, L. Marrazzo, M. Ceroti, F. Paiar, G. Biti, and M. Bucciolini. Accuracy of a 3D laser/camera surface imaging system for setup verification of the pelvic and thoracic regions in radiotherapy treatments. *Medical Physics*, **40**(1):011710, 2013. DOI: http://dx.doi.org/10.1118/1.4769428.

[215] H. Pan, L. I. Cervino, T. Pawlicki, S. B. Jiang, J. Alksne, N. Detorie, M. Russell, B. S. Carter, K. T. Murphy, A. J. Mundt, C. Chen, and J. D. Lawson. Frameless, real-time, surface imaging-guided radiosurgery: clinical outcomes for brain metastases. *Neurosurgery*, **71**(4):844–851, 2012.

[216] S.-Y. Park and M. Subbarao. An accurate and fast point-to-plane registration technique. *Pattern Recogn. Lett.*, **24**(16):2967–2976, 2003. DOI: 10.1016/S0167-8655(03)00157-0.

[217] R. A. Patchell. The management of brain metastases. *Cancer Treat. Rev.*, **29**(6):533–540, 2003. DOI: 10.1016/S0305-7372(03)00105-1.

[218] M. A. Pathak, K. Jimbow, G. Szabo, and T. B. Fitzpatrick. Sunlight and melanin pigmentation, pp. 211–239 in K. C. Smith, editor, *Photochemical and photobiological reviews*, volume 1. Plenum Press, New York, 1976.

[219] G. Pellacani and S. Seidenari. Variations in facial skin thickness and echogenicity with site and age. *Acta Derm. Venereol.*, **79**(5):366–369, 1999.

[220] J. L. Peng, D. Kahler, J. G. Li, S. Samant, G. Yan, R. Amdur, and C. Liu. Characterization of a real-time surface image-guided stereotactic positioning system. *Medical Physics*, **37**(10):5421–5433, 2010. DOI: http://dx.doi.org/10.1118/1.3483783.

[221] L.-C. Peng, D. Kahler, S. Samant, J. Li, R. Amdur, J. R. Palta, and C. Liu. Quality assessment of frameless fractionated stereotactic radiotherapy using cone beam computed tomography. *International Journal of Radiation Oncology*Biology*Physics*, **78**(5):1586 – 1593, 2010. DOI: http://dx.doi.org/10.1016/j.ijrobp.2010.02.011.

[222] G. Penney, J. Weese, J. Little, P. Desmedt, D. Hill, and D. Hawkes. A comparison of similarity measures for use in 2-d-3-d medical image registration. *Medical Imaging, IEEE Transactions on*, **17**(4):586–595, 1998. DOI: 10.1109/42.730403.

[223] G. C. Pereira, M. Traughber, and R. F. Muzic. The role of imaging in radiation therapy planning: past, present, and future. *Biomed Res Int*, **2014**:231090, 2014. DOI: http://dx.doi.org/10.1155/2014/231090.

[224] G. I. Petrov, A. Doronin, H. T. Whelan, I. Meglinski, and V. V. Yakovlev. Human tissue color as viewed in high dynamic range optical spectral transmission measurements. *Biomed Opt Express*, **3**(9):2154–2161, 2012.

[225] J. P. Pignon, J. Bourhis, C. Domenge, and L. Designe. Chemotherapy added to locoregional treatment for head and neck squamous-cell carcinoma: three meta-analyses of updated individual data. MACH-NC Collaborative Group. Meta-Analysis of Chemotherapy on Head and Neck Cancer. *Lancet*, **355**(9208):949–955, 2000. DOI: http://dx.doi.org/10.1016/S0140-6736(00)90011-4.

[226] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko. Review: A review of novelty detection. *Signal Process.*, **99**:215–249, 2014. DOI: 10.1016/j.sigpro.2013.12.026.

[227] S. Placht, J. Stancanello, C. Schaller, M. Balda, and E. Angelopoulou. Fast time-of-flight camera based surface registration for radiotherapy patient positioning. *Medical Physics*, **39**(1):4–17, 2012. DOI: http://dx.doi.org/10.1118/1.3664006.

[228] J. C. Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998.

[229] S. A. Prahl. The adding-doubling method. In A. Welch and M. Van Gemert, editors, *Optical-Thermal Response of Laser-Irradiated Tissue*, Lasers, Photonics, and Electro-Optics, pp. 101–129. Springer US, 1995. ISBN 978-1-4757-6094-1. DOI: 10.1007/978-1-4757-6092-7_5.

[230] A. R. Protection and A. G. Nuclear Safety Agency, Yallambie, Australia, 2015. `http://www.arpansa.gov.au/pubs/RadiationProtection/FitzpatrickSkinType.pdf`. (accessed on 2015-05-07).

[231] K. Pulli. Multiview registration for large data sets. In *Proceedings of the 2nd International Conference on 3-D Digital Imaging and Modeling*, pp. 160–168, Washington, DC, USA. IEEE Computer Society. `http://dl.acm.org/citation.cfm?id=1889712.1889738`. Published in *3DIM'99*, pp. 160–168, 1999. ISBN 0-7695-0062-5.

[232] J. Quiñonero Candela and C. E. Rasmussen. A unifying view of sparse approximate gaussian process regression. *J. Mach. Learn. Res.*, **6**:1939–1959, 2005.

[233] T. S. Ram, P. B. Ravindran, F. R. Viswanathan, P. N. Viswanathan, and S. P. Pavamani. Extracranial doses in stereotactic and conventional radiotherapy for pituitary adenomas. *J Appl Clin Med Phys*, **7**(2):96–100, 2006.

[234] N. Ramakrishna, F. Rosca, S. Friesen, E. Tezcanli, P. Zygmanszki, and F. Hacker. A clinical comparison of patient setup and intra-fraction motion using frame-based radiosurgery versus a frameless image-guided radiosurgery system for intracranial lesions. *Radiotherapy and Oncology*, **95**(1):109 – 115, 2010. DOI: http://dx.doi.org/10.1016/j.radonc.2009.12.030.

[235] J. Ramella-Roman, S. Prahl, and S. Jacques. Three Monte Carlo programs of polarized light transport into scattering media: part I. *Opt. Express*, **13**(12):4420–4438, 2005. DOI: 10.1364/OPEX.13.004420.

[236] J. C. Ramella-Roman, S. A. Prahl, and S. L. Jacques. Three Monte Carlo programs of polarized light transport into scattering media: part II. *Opt. Express*, **13**(25):10392–10405, 2005. DOI: 10.1364/OPEX.13.010392.

[237] C. E. Rasmussen. *Gaussian processes for machine learning*. MIT Press, 2006.

[238] C. E. Rasmussen and H. Nickisch. Gaussian processes for machine learning (gpml) toolbox. *J. Mach. Learn. Res.*, **11**:3011–3015, 2010.

[239] RICOH IMAGING DEUTSCHLAND GmbH, Hamburg, Germany, 2015. `http://www.ricoh-imaging.de/`. (accessed on 2015-04-23).

[240] J. Robar, B. Clark, J. Schella, and C. S. Kim. Analysis of patient repositioning accuracy in precision radiation therapy using automated image fusion. *Journal of Applied Clinical Medical Physics*, **6**(1):71–83, 2005.

[241] R. L. Rotondo, K. Sultanem, I. Lavoie, J. Skelly, and L. Raymond. Comparison of repositioning accuracy of two commercially available immobilization systems for treatment of head-and-neck tumors using simulation computed tomography imaging. *International Journal of Radiation Oncology*Biology*Physics*, **70**(5):1389 – 1396, 2008. DOI: http://dx.doi.org/10.1016/j.ijrobp.2007.08.035.

[242] R. F. Rushmer, K. J. K. Buettner, J. M. Short, and G. F. Odland. The skin. *Science*, **154**(3747):343–348, 1966. DOI: DOI:10.1126/science.154.3747.343.

[243] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *3rd International Conference on 3D Digital Imaging and Modeling (3DIM 2001), 28 May - 1 June 2001, Quebec City, Canada*, pp. 145–152, 2001. DOI: 10.1109/IM.2001.924423.

[244] T. C. Ryken, S. L. Meeks, E. C. Pennington, P. Hitchon, V. Traynelis, N. A. Mayr, F. J. Bova, W. A. Friedman, and J. M. Buatti. Initial clinical experience with frameless stereotactic radiosurgery: analysis of accuracy and feasibility. *International Journal of Radiation Oncology*Biology*Physics*, **51**(4):1152 – 1158, 2001. DOI: http://dx.doi.org/10.1016/S0360-3016(01)01756-4.

[245] S. Sachdeva. Fitzpatrick skin typing: Applications in dermatology. *Indian Journal of Dermatology, Venereology, and Leprology*, **75**(1):93–96, 2009. DOI: 10.4103/0378-6323.45238.

[246] R. D. Sanders. The Trigeminal (V) and Facial (VII) Cranial Nerves: Head and Face Sensation and Movement. *Psychiatry (Edgmont)*, **7**(1):13–16, 2010.

[247] A. Santhanam, D. Low, and P. Kupelian. TH-C-BRC: 3D tracking of interfraction and intrafraction head and neck anatomy during radiotherapy using multiple Kinect sensors. *Medical Physics*, **38**(6): 3858–3858, 2011. DOI: http://dx.doi.org/10.1118/1.3613528.

[248] D. K. Sardar, B. G. Yust, F. J. Barrera, L. C. Mimun, and A. T. Tsin. Optical absorption and scattering of bovine cornea, lens and retina in the visible region. *Lasers Med Sci*, **24**(6):839–847, 2009.

[249] D. K. Sardar, B. G. Yust, F. J. Barrera, L. C. Mimun, and A. T. Tsin. Optical absorption and scattering of bovine cornea, lens and retina in the visible region. *Lasers Med Sci*, **24**(6):839–847, 2009.

[250] A. Sawant, R. L. Smith, R. B. Venkat, L. Santanam, B. Cho, P. Poulsen, H. Cattell, L. J. Newell, P. Parikh, and P. J. Keall. Toward submillimeter accuracy in the management of intrafraction motion: The integration of real-time internal position monitoring and multileaf collimator target tracking. *International Journal of Radiation Oncology*Biology*Physics*, **74**(2):575 – 582, 2009. DOI: http://dx.doi.org/10.1016/j.ijrobp.2008.12.057.

[251] S. Sayil, J. Kerns, D.V., and S. Kerns. Comparison of contactless measurement and testing techniques to a all-silicon optical test and characterization method. *Instrumentation and Measurement, IEEE Transactions on*, **54**(5):2082–2089, 2005. DOI: 10.1109/TIM.2005.854253.

[252] J. Schaerer, A. Fassi, M. Riboldi, P. Cerveri, G. Baroni, and D. Sarrut. Multi-dimensional respiratory motion tracking from markerless optical surface imaging based on deformable mesh registration. *Physics in Medicine and Biology*, **57**(2):357, 2012.

[253] D. Schmidhalter, M. K. Fix, M. Wyss, N. Schaer, P. Munro, S. Scheib, P. Kunz, and P. Manser. Evaluation of a new six degrees of freedom couch for radiation therapy. *Medical Physics*, **40**(11):111710, 2013. DOI: http://dx.doi.org/10.1118/1.4823789.

[254] P. J. Schoeffel, W. Harms, G. Sroka-Perez, W. Schlegel, and C. P. Karger. Accuracy of a commercial optical 3D surface imaging system for realignment of patients for radiotherapy of the thorax. *Physics in Medicine and Biology*, **52**(13):3949, 2007.

[255] B. Schölkopf and A. J. Smola. *Learning with Kernels - Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning series. MIT Press, Cambridge, Massachusetts, USA, 1st edition, 2002.

[256] C. Schütz, T. Jost, and H. Hügli. Multi-feature matching algorithm for free-form 3D surface registration. In *Fourteenth International Conference on Pattern Recognition, ICPR 1998, Brisbane, Australia, 16-20 August, 1998*, pp. 982–984, 1998. DOI: 10.1109/ICPR.1998.711852.

[257] J. G. Schwade, P. V. Houdek, H. J. Landy, J. L. Bujnoski, A. A. Lewin, A. A. Abitol, C. F. Serago, and V. J. Pisciotta. Small-field stereotactic external-beam radiation therapy of intracranial lesions: fractionated treatment with a fixed-halo immobilization device. *Radiology*, **176**(2):563–565, 1990. DOI: http://dx.doi.org/10.1148/radiology.176.2.2195597.

[258] A. Schwaighofer and V. Tresp. Transductive and inductive methods for approximate gaussian process regression. In S. Thrun and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pp. 953–960. MIT Press, Cambridge, MA, 2002.

[259] M. Schwarz, K. Giske, A. Stoll, S. Nill, P. E. Huber, J. Debus, R. Bendl, and E. M. Stoiber. IGRT versus non-IGRT for postoperative head-and-neck IMRT patients: dosimetric consequences arising from a PTV margin reduction. *Radiat Oncol*, **7**:133, 2012. DOI: 10.1186/1748-717X-7-133.

[260] T. W. Scott, J. L. Beach, and O. A. Mendiondo. A precision repeat localization head frame for fractionated stereotactic radiotherapy. *Medical Dosimetry*, **22**(1):5 – 8, 1997. DOI: http://dx.doi.org/10.1016/S0958-3947(96)00148-3.

[261] C. Scully, J. Lee, J. Meyer, A. Gorbach, D. Granquist-Fraser, Y. Mendelson, and K. Chon. Physiological parameter monitoring from optical recordings with a mobile phone. *Biomedical Engineering, IEEE Transactions on*, **59**(2):303–306, 2012. DOI: 10.1109/TBME.2011.2163157.

[262] G. Shafirstein, W. Bäumler, M. Lapidoth, S. Ferguson, P. E. North, and M. Waner. A new mathematical approach to the diffusion approximation theory for selective photothermolysis modeling and its implication in laser treatment of port-wine stains. *Lasers in Surgery and Medicine*, **34**(4):335–347, 2004. DOI: 10.1002/lsm.20028.

[263] L. Sharp, F. Lewin, H. Johansson, D. Payne, A. Gerhardsson, and L. E. Rutqvist. Randomized trial on two types of thermoplastic masks for patient immobilization during radiation therapy for head-and-neck cancer. *International Journal of Radiation Oncology*Biology*Physics*, **61**(1):250 – 256, 2005. DOI: http://dx.doi.org/10.1016/j.ijrobp.2004.04.047.

[264] H. Shirato, M. Oita, K. Fujita, Y. Watanabe, and K. Miyasaka. Feasibility of synchronization of real-time tumor-tracking radiotherapy and intensity-modulated radiotherapy from viewpoint of excessive dose from fluoroscopy. *International Journal of Radiation Oncology*Biology*Physics*, **60**(1):335 – 341, 2004. DOI: http://dx.doi.org/10.1016/j.ijrobp.2004.04.028.

[265] S. Shuster, M. M. Black, and E. McVitie. The influence of age and sex on skin thickness, skin collagen and density. *British Journal of Dermatology*, **93**(6):639–643, 1975. DOI: 10.1111/j.1365-2133.1975.tb05113.x.

[266] R. M. Siervogel, A. F. Roche, J. H. Himes, W. C. Chumlea, and R. McCammon. Subcutaneous fat distribution in males and females from 1 to 39 years of age. *Am. J. Clin. Nutr.*, **36**(1):162–171, 1982.

[267] J. Sim, J. S. Lee, and O. Kwon. Missing values and optimal selection of an imputation method and classification algorithm to improve the accuracy of ubiquitous computing applications. *Mathematical Problems in Engineering*, **2015**:1–15, 2015. DOI: doi:10.1155/2015/538613.

[268] A. Smola and B. Schölkopf. A tutorial on Support Vector regression. *Statistics and Computing*, **14**(3): 199–222, 2004. DOI: 10.1023/B:STCO.0000035301.49549.88.

[269] A. J. Smola and P. Bartlett. Sparse greedy Gaussian process regression. In *Advances in Neural Information Processing Systems 13*, pp. 619–625. MIT Press, 2001.

[270] A. J. Smola and B. Schölkopf. Sparse greedy matrix approximation for machine learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 911–918, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc. `http://dl.acm.org/citation.cfm?id=645529.657980`. Published in *ICML '00*, pp. 911–918, 2000. ISBN 1-55860-707-2.

[271] E. Snelson. *Flexible and efficient Gaussian process models for machine learning*. PhD thesis, Gatsby Computational Neuroscience Unit, University College London, 2007.

[272] W. Y. Song, S. Kamath, S. Ozawa, S. A. Ani, A. Chvetsov, N. Bhandare, J. R. Palta, C. Liu, and J. G. Li. A dose comparison study between XVI and OBI CBCT systems. *Med Phys*, **35**(2):480–486, 2008.

[273] M. F. Spadea, B. Tagaste, M. Riboldi, E. Preve, D. Alterio, G. Piperno, C. Garibaldi, R. Orecchia, A. Pedotti, and G. Baroni. Intra-fraction setup variability: IR optical localization vs. X-ray imaging in a hypofractionated patient population. *Radiat Oncol*, **6**:38, 2011.

[274] Staat & Gesellschaft - Todesursachen, Statistisches Bundesamt, Wiesbaden, Germany, 2015. `https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/Todesursachen/Todesursachen.html`. (accessed on 2015-04-09).

[275] W. Star. Diffusion theory of light transport. In A. J. Welch and M. J. van Gemert, editors, *Optical-Thermal Response of Laser-Irradiated Tissue*, pp. 145–201. Springer Netherlands, 2011. ISBN 978-90-481-8830-7. DOI: 10.1007/978-90-481-8831-4_6.

[276] M. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer New York, 1999. ISBN 9780387986296.

[277] J. A. Stephenson and A. L. Wiley. Current techniques in three-dimensional CT simulation and radiation treatment planning. *Oncology (Williston Park, N.Y.)*, **9**(11):1225–1232, 1995.

[278] F. Stieler, F. Wenz, D. Scherrer, M. Bernhardt, and F. Lohr. Clinical evaluation of a commercial surface-imaging system for patient positioning in radiotherapy. *Strahlentherapie und Onkologie*, **188**(12):1080–1084, 2012. DOI: 10.1007/s00066-012-0244-7.

[279] F. Stieler, F. Wenz, M. Shi, and F. Lohr. A novel surface imaging system for patient positioning and surveillance during radiotherapy. *Strahlentherapie und Onkologie*, **189**(11):938–944, 2013. DOI: 10.1007/s00066-013-0441-z.

[280] D. Stifter, E. Leiss-Holzinger, Z. Major, B. Baumann, M. Pircher, E. Götzinger, C. K. Hitzenberger, and B. Heise. Dynamic optical studies in materials testing with spectral-domain polarization-sensitive optical coherence tomography. *Opt. Express*, **18**(25):25712–25725, 2010. DOI: 10.1364/OE.18.025712.

[281] P. Stüber, T. Wissel, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst. An improvement for the scanning process in high accuracy head tracking. In *12. Jahrestagung der Deutschen Gesellschaft für Computer- und Roboterassistierte Chirurgie (CURAC)*, pp. 179–182. CURAC, 2013.

[282] P. Stüber, T. Wissel, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst. Design and evaluation of a highly accurate optical setup for backscatter analysis. In *44. Jahrestagung der DGMP*, pp. 181–186. DGMP, 2013.

[283] P. Stüber, T. Wissel, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst. A comparison of different hardware design approaches for feature-supported optical tracking with respect to angular dependencies. In *56th Annual Meeting of the AAPM*, pp. 204. Published in *Medical Physics*, **41**(6):204, 2014. DOI: 10.1118/1.4888259.

[284] P. Stüber, B. Wagner, T. Wissel, R. Bruder, A. Schweikard, and F. Ernst. An approach to improve accuracy of optical tracking systems in cranial radiation therapy. *Cureus*, **7**(1):e239, 2015. DOI: doi: 10.7759/cureus.239.

[285] P. Stüber, T. Wissel, B. Wagner, A. Schweikard, and F. Ernst. Feasibility test of line sensors for optical tissue thickness estimation. In *Optical Methods for Inspection, Characterization, and Imaging of Biomaterials II*, pp. 95291C. Published in *Proceedings of SPIE Optical Metrology*, **9529**:95291C, 2015. DOI: 10.1117/12.2024851.

[286] J. Sun and M. Smith. Multidimensional imaging for skin tissue surface characterization. *Computers in Industry*, **64**(9):1383 – 1389, 2013. DOI: http://dx.doi.org/10.1016/j.compind.2013.06.004.

[287] T. Takakura, T. Mizowaki, M. Nakata, S. Yano, T. Fujimoto, Y. Miyabe, M. Nakamura, and M. Hiraoka. The geometric accuracy of frameless stereotactic radiosurgery using a 6D robotic couch system. *Physics in Medicine and Biology*, **55**(1):1, 2010. DOI: doi:10.1088/0031-9155/55/1/001.

[288] A. J. Thody, E. M. Higgins, K. Wakamatsu, S. Ito, S. A. Burchill, and J. M. Marks. Pheomelanin as well as eumelanin is present in human epidermis. *J. Invest. Dermatol.*, **97**(2):340–344, 1991.

[289] Thorlabs Inc, Newton,NJ,USA, 2015. `http://www.thorlabs.de/`. (accessed on 2015-07-06).

[290] R. Tibshirani. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(3):273–282, 2011. DOI: 10.1111/j.1467-9868.2011.00771.x.

[291] M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.*, **1**: 211–244, 2001. DOI: 10.1162/15324430152748236.

[292] W. A. Tome, S. L. Meeks, J. M. Buatti, F. J. Bova, W. A. Friedman, and Z. Li. A high-precision system for conformal intracranial radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, **47**(4):1137 – 1143, 2000. DOI: http://dx.doi.org/10.1016/S0360-3016(00)00502-2.

[293] J.-C. Tonn, M. Westphal, and J. T. Rutka. *Oncology of CNS Tumors*. Springer, Berlin, 2010. ISBN 9783642028731. DOI: 10.1007/978-3-642-02874-8.

[294] M. Trebsdorf. *Biologie, Anatomie, Physiologie: Lehrbuch und Atlas*. Europa-Lehrmittel, 13th edition, 2013.

[295] T. L. Troy and S. N. Thennadil. Optical properties of human skin in the near infrared wavelength range of 1000 to 2200 nm. *Journal of Biomedical Optics*, **6**(2):167–176, 2001. DOI: 10.1117/1.1344191.

[296] E. Tryggestad, M. Christian, E. Ford, C. Kut, Y. Le, G. Sanguineti, D. Y. Song, and L. Kleinberg. Inter- and intrafraction patient positioning uncertainties for intracranial radiotherapy: A study of four frameless, thermoplastic mask-based immobilization strategies using daily cone-beam CT. *International Journal of Radiation Oncology*Biology*Physics*, **80**(1):281 – 290, 2011. DOI: 10.1016/j.ijrobp.2010.06.022.

[297] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake. Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. *ACM Trans. Graph.*, **22**(3):770–779, 2003. DOI: 10.1145/882262.882344.

[298] M. Uematsu, M. Sonderegger, A. Shioda, K. Tahara, T. Fukui, Y. Hama, T. Kojima, J. R. Wong, and S. Kusano. Daily positioning accuracy of frameless stereotactic radiation therapy with a fusion of computed tomography and linear accelerator (focal) unit: evaluation of z-axis with a z-marker. *Radiotherapy and Oncology*, **50**(3):337 – 339, 1999. DOI: http://dx.doi.org/10.1016/S0167-8140(99)00037-7.

[299] N. Ugryumova, S. J. Matcher, and D. P. Attenburrow. Measurement of bone mineral density via light scattering. *Physics in Medicine and Biology*, **49**(3):469, 2004.

[300] H. C. van de Hulst. Asymptotic fitting, a method for solving anisotropic transfer problems in thick layers. *J. Comput. Phys.*, **3**(2):291–306, 1968. DOI: 10.1016/0021-9991(68)90023-5.

[301] M. J. van Gemert, S. L. Jacques, H. J. Sterenborg, and W. M. Star. Skin optics. *IEEE Trans Biomed Eng*, **36**(12):1146–1154, 1989.

[302] M. van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, **47**(4):1121–1135, 2000. DOI: http://dx.doi.org/10.1016/S0360-3016(00)00518-6.

[303] W. E. Vargas and G. A. Niklasson. Applicability conditions of the Kubelka-Munk theory. *Appl Opt*, **36**(22):5580–5586, 1997.

[304] Varian Medical Systems, Inc., Palo Alto, CA, USA, 2015. `http://www.varian.com/`. (accessed on 2015-04-09).

[305] M. Velec, J. N. Waldron, B. O'Sullivan, A. Bayley, B. Cummings, J. J. Kim, J. Ringash, S. L. Breen, G. A. Lockwood, and L. A. Dawson. Cone-beam CT assessment of interfraction and intrafraction setup error of two head-and-neck cancer thermoplastic masks. *International Journal of Radiation Oncology*Biology*Physics*, **76**(3):949 – 955, 2010. DOI: http://dx.doi.org/10.1016/j.ijrobp.2009.07.004.

[306] D. Verellen, N. Linthout, and G. Storme. Target localization and treatment verification for intensity modulated conformal radiation therapy of the head and neck region. The AZ-VUB experience. Akademisch Ziekenhuis-Vrije Universiteit Brussel. *Strahlenther Onkol*, **174 Suppl 2**:19–27, 1998.

[307] D. Verellen, M. De Ridder, N. Linthout, K. Tournel, G. Soete, and G. Storme. Innovations in image-guided radiotherapy. *Nat. Rev. Cancer*, **7**(12):949–960, 2007. DOI: 10.1038/nrc2288.

[308] W. Verkruysse, L. O. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, **16**(26):21434–21445, 2008. DOI: 10.1364/OE.16.021434.

[309] M. Verleysen and D. François. The curse of dimensionality in data mining and time series prediction. In J. Cabestany, A. Prieto, and F. Sandoval, editors, *Computational Intelligence and Bioinspired Systems*, volume 3512 of *Lecture Notes in Computer Science*, pp. 758–770. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-26208-4. DOI: 10.1007/11494669_93.

[310] Vision RT Ltd., London, UK, 2015. `http://www.visionrt.com/`. (accessed on 2015-04-23).

[311] F. von Luschan. *Beitraege zur Voelkerkunde der deutschen Schutzgebiete*. Reimer, Berlin, 1897.

[312] B. Wagner. *Structured Light for Head Tracking in Radiation Therapy*. PhD thesis, University of Lübeck, 2015.

[313] B. Wagner, P. Stüber, T. Wissel, R. Bruder, A. Schweikard, and F. Ernst. Accuracy of object tracking based on time-multiplexed structured light. In *12. Jahrestagung der Deutschen Gesellschaft für Computer- und Roboterassistierte Chirurgie (CURAC)*, pp. 139–142. CURAC, 2013.

[314] B. Wagner, P. Stüber, T. Wissel, R. Bruder, A. Schweikard, and F. Ernst. Accuracy analysis for triangulation and tracking based on time-multiplexed structured light. *Medical Physics*, **41**(8):082701, 2014. DOI: 10.1118/1.4890093.

[315] T. H. Wagner, S. L. Meeks, F. J. Bova, W. A. Friedman, T. R. Willoughby, P. A. Kupelian, and W. Tome. Optical tracking technology in stereotactic radiation therapy. *Med Dosim*, **32**(2):111–120, 2007. DOI: doi:10.1016/j.meddos.2007.01.008.

[316] J. Z. Wang, R. Rice, T. Pawlicki, A. J. Mundt, A. Sandhu, J. Lawson, and K. T. Murphy. Evaluation of patient setup uncertainty of optical guided frameless system for intracranial stereotactic radiosurgery. *J Appl Clin Med Phys*, **11**(2):3181, 2010.

[317] L. Wang, S. L. Jacques, and L. Zheng. Mcml - monte carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedicine*, **47**(2):131 – 146, 1995. DOI: 10.1016/0169-2607(95)01640-F.

[318] L. V. Wang and S. L. Jacques. Source of error in calculation of optical diffuse reflectance from turbid media using diffusion theory. *Computer Methods and Programs in Biomedicine*, **61**(3):163 – 170, 2000. DOI: http://dx.doi.org/10.1016/S0169-2607(99)00041-3.

[319] M. Wannenmacher, J. Debus, and F. Wenz. *Strahlentherapie*. Springer, Berlin, 2006. ISBN 9783540228127.

[320] C. Weltens, K. Kesteloot, G. Vandevelde, and W. V. den Bogaert. Comparison of plastic and orfit masks for patient head fixation during radiotherapy: Precision and costs. *International Journal of Radiation Oncology*Biology*Physics*, **33**(2):499 – 507, 1995. DOI: http://dx.doi.org/10.1016/0360-3016(95)00178-2.

[321] T. Weyrich, W. Matusik, H. Pfister, A. Ngan, M. Gross, T. Weyrich, W. Matusik, H. Pfister, A. Ngan, and M. Gross. Measuring skin reflectance and subsurface scattering. Technical report, Mitsubishi Electric Research Laboratories (MERL), 2005.

[322] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, and M. Gross. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph.*, **25**(3):1013–1024, 2006. DOI: 10.1145/1141911.1141987.

[323] R. D. Wiersma, Z. Wen, M. Sadinski, K. Farrey, and K. M. Yenice. Development of a frameless stereotactic radiosurgery system based on real-time 6D position monitoring and adaptive head motion compensation. *Physics in Medicine and Biology*, **55**(2):389, 2010.

[324] T. Wissel, R. Bruder, and A. Schweikard. Skin thickness estimation for high precision optical head tracking during cranial radiation therapy—a simulation study. In *Proceedings of the 55th Annual Meeting of the ASTRO*, pp. S669, Atlanta, GA, USA. ASTRO. Published in *International Journal of Radiation Oncology, Biology, Physics*, **87**(2):S669, 2013. DOI: 10.1016/j.ijrobp.2013.06.1775.

[325] T. Wissel, R. Bruder, A. Schweikard, and F. Ernst. Estimating soft tissue thickness from light-tissue interactions - a simulation study. *Biomed. Opt. Express*, **4**(7):1176–1187, 2013. DOI: 10.1364/BOE.4.001176.

[326] T. Wissel, T. Pfeiffer, R. Frysch, R. T. Knight, E. F. Chang, H. Hinrichs, J. W. Rieger, and G. Rose. Hidden markov model and support vector machine based decoding of finger movements using electrocorticography. *Journal of Neural Engineering*, **10**(5):056020, 2013.

[327] T. Wissel, P. Stuber, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst. Preliminary study on optical feature detection for head tracking in radiation therapy. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pp. 1–5, Chania, Greece, 2013. DOI: 10.1109/BIBE.2013.6701632.

[328] T. Wissel, P. Stüber, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst. Angle influence and compensation for marker-less head tracking based on laser scanners. In *Proceedings of the 28th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS'14)*, pp. 62–63. Published in *International Journal of Computer Assisted Radiology and Surgery*, **9**(S1):62–63, 2014.

[329] T. Wissel, P. Stüber, B. Wagner, R. Dürichen, R. Bruder, A. Schweikard, and F. Ernst. Tissue thickness estimation for high precision head-tracking using a galvanometric laser scanner - a case study. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '14)*, pp. 3106–3109, Chicago, USA, 2014. EMBS, IEEE. DOI: 10.1109/EMBC.2014.6944280.

[330] T. Wissel, P. Stüber, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst. Enhanced tissue thickness computation by exploiting local neighborhoods. In *Proceedings of the 29th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS'15)*. Published in *International Journal of Computer Assisted Radiology and Surgery*, 2015.

[331] T. Wissel, P. Stüber, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst. Enriching 3D optical surface scans with prior knowledge: tissue thickness computation by exploiting local neighborhoods. *International Journal of Computer Assisted Radiology and Surgery*, **11**(4):569–579, 2015. DOI: 10.1007/s11548-015-1246-6.

[332] T. Wissel, P. Stüber, B. Wagner, A. Schweikard, and F. Ernst. Efficient estimation of tissue thicknesses using sparse approximation for gaussian processes. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '15)*, pp. 7015–7018, Milan, Italy, 2015. EMBS, IEEE. DOI: 10.1109/EMBC.2015.7320007.

[333] T. Wissel, P. Stüber, B. Wagner, A. Schweikard, and F. Ernst. Tissue segmentation from head mri: A ground truth validation for feature-enhanced tracking. In *Biomedizinische Technik*, pp. S184, Lübeck, Germany. Walter de Gruyter, Berlin, Boston. Published in *49th DGBMT Annual Conference*, **60**(s1):S184, 2015. DOI: 10.1515/bmt-2015-5008.

[334] T. Wissel, P. Stüber, B. Wagner, A. Schweikard, and F. Ernst. Tissue segmentation from head mri: A ground truth validation for feature-enhanced tracking. *Current Directions of Biomedical Engineering*, **1**(1):228–231, 2015. DOI: 10.1515/CDBME-2015-0057.

[335] T. Wissel, P. Süber, J. Manit, R. Bruder, A. Schweikard, and F. Ernst. Predicting cutaneous features for marker-less optical head-tracking in cranial radiotherapy. In *Bioinformatics and Bioengineering (BIBE), 2015 IEEE 15th International Conference on*, pp. 1–5, Belgrade, Serbia, 2015. DOI: 10.1109/BIBE.2015.7367695.

[336] T. Wissel, B. Wagner, P. Stüber, A. Schweikard, and F. Ernst. Data-driven learning for calibrating galvanometric laser scanners. *IEEE Sensors Journal*, **15**(10):5709 – 5717, 2015. DOI: 10.1109/JSEN.2015.2447835.

[337] T. Wissel, P. Stüber, B. Wagner, R. Bruder, C. Erdmann, C.-S. Deutz, B. Sack, J. Manit, A. Schweikard, and F. Ernst. Enhanced optical head tracking for cranial radiotherapy: Supporting surface registration by cutaneous structures. *International Journal of Radiation Oncology*Biology*Physics*, pp. [accepted], 2016. DOI: 10.1016/j.ijrobp.2016.01.041.

[338] J. Xia and R. A. Siochi. A real-time respiratory motion monitoring system using KINECT: proof of concept. *Med Phys*, **39**(5):2682–2685, 2012.

[339] M. Yamamoto, M. Ide, and M. Jimbo. Leksell frame fixation using zygomatic bone for stereotactic radiosurgery. *Journal of Radiosurgery*, **2**(1):43–45, 1999. DOI: 10.1023/A:1022964318986.

[340] H. Yan, F.-F. Yin, and J. H. Kim. A phantom study on the positioning accuracy of the novalis body system. *Medical Physics*, **30**(12):3052–3060, 2003. DOI: http://dx.doi.org/10.1118/1.1626122.

[341] Y. Yan, Y. Song, and A. L. Boyer. An investigation of a video-based patient repositioning technique. *International Journal of Radiation Oncology*Biology*Physics*, **54**(2):606 – 614, 2002. DOI: http://dx.doi.org/10.1016/S0360-3016(02)02934-6.

[342] K. Yoshida, I. Nishidate, N. Ojima, and K. Iwata. Reduction of shading-derived artifacts in skin chromophore imaging without measurements or assumptions about the shape of the subject. *Journal of Biomedical Optics*, **19**(1):016009, 2014. DOI: 10.1117/1.JBO.19.1.016009.

[343] Zentrum für Krebsregisterdaten, Robert Koch Institute, Berlin, Germany, 2015. `http://www.krebsdaten.de`. (accessed on 2015-04-10).

[344] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *Int. J. Comput. Vision*, **13**(2):119–152, 1994. DOI: 10.1007/BF01427149.

[345] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, **67**:301–320, 2005. DOI: 10.1111/j.1467-9868.2005.00503.x.

[346] A. Zourabian, A. Siegel, B. Chance, N. Ramanujan, M. Rode, and D. A. Boas. Trans-abdominal monitoring of fetal arterial blood oxygenation using pulse oximetry. *J Biomed Opt*, **5**(4):391–405, 2000.

List of Figures

List of Tables

Glossary and List of Abbreviations

Symbols

γ-ray

Electromagnetic radiation with a wavelength $\lambda \leq 0.01\,\text{nm}$. In the medical context used for radiating tumors or MV imaging and often inconsistently referred to as (high energy) X-rays in the medical context.

A

A

Label for items that are related to a feature space with additional incident angle feature. In addition to five NIR backscatter features, the incident angle is added as a sixth feature.

ABS

A common thermoplastic terpolymer with chemical formula $(C_8H_8)_x \cdot (C_4H_6)_y \cdot (C_3H_3N)_z$.

AC

The anterior commissure in the brain consists of a nerve fiber bundle which connects the two cerebral hemispheres. It is located in front of the columns of the fornix.

AE

Absolute value of the difference between the estimated and true result.

AM1

Testing scenario "Across Measurements 1". Three measurements corresponding to three different head poses are used for testing. With one measurement à two frames in the training set and equivalently one measurement in the test set, there are six mutual combinations. The overall error is averaged across all combinations.

AM2

Testing scenario ”Across Measurements 2”. Three measurements corresponding to three different head poses are used for testing. With two measurements à one frame in the training set and one measurement à two frames in the test set, there are three combinations. The overall error is averaged across all combinations.

Anterior-posterior (AP)

Axis along the human body - from the front surface to the back.

Automatic Relevance Detection (ARD)

A property of kernel functions, which may assign and learn different weights per feature dimension according to their relevance. This is achieved by introducing one length scale parameter per feature space dimension.

C

Central Nervous System (CNS)

Brain region and in the narrow sense an anatomical term to classify cancers including brain stem glioma, craniopharyngioma, medulloblastoma, and meningioma.

Central Processing Unit (CPU)

Main microprocessor of a computer.

Cerebrospinal Fluid (CSF)

A clear and colorless liquid surrounding the brain. It is produced in the choroid plexuses of the ventricles which are also filled with the fluid. It acts as a protective buffer.

Charge-coupled Device (CCD)

A semiconductor device used as a camera sensor in which charges accumulate on each pixel due to light incidence. Charges are shifted along these capacitive bins (”bucket chains”) to sequentially read out the pixel intensities.

Clinical Target Volume (CTV)

Tissue volume that contains a GTV and/or subclinical microscopic malignant disease, which has to be eliminated. The CTV has to be treated adequately in order to achieve the aim of therapy: cure or palliation. It is a purely anatomical concept and

needs to be defined *before* the choice of treatment techniques is made. (definition cf. [133]).

Complementary Metal-Oxide-Semiconductor (CMOS)

Semiconductor element where n- and p-channel semiconductors are both attached to a substrate. In a light-sensitive version it can be used as a camera sensor. In contrast to CCD pixels are typically read out in parallel.

Compute Unified Device Architecture (CUDA)

A programming technique introduced by the Nvidia Corporation (Santa Clara, CA, USA). Code can be distributed on the CPU as well as GPU. The latter provides possibilities for parallelizing work load and to exploit additional, dedicated and specialized hardware and processing power.

Computed Tomography (CT)

Anatomical imaging method to reconstruct 3D volumes using multiple X-ray images.

Cone Beam Computed Tomography (CBCT)

Medical imaging technique based on X-ray CT and often used for onboard imaging during radiotherapy. Emitted X-rays diverge and therefore form a cone.

Conformal Radiation Therapy (CRT)

Extension of conventional RT that aims at dose distributions which are spatially conform with the tumor. Using different irradiation directions and a superposition of multiple beams, irregular shaped volumes can be treated and safety margins can be reduced.

cross-validation (CV)

Testing scheme for supervised learning. The data is separated into n_{fld} folds. A model is trained on $n_{fld} - 1$ folds and tested on the remaining one.

D

Desoxyribonucleic Acid (DNA)

Nuclei acid in form of a double helix that carries genetic information.

Digitally Reconstructed Radiograph (DRR)

Artificial 2D images similar to standard X-ray images, but digitally reconstructed from acquired 3D CT volumes by summing up voxel intensities along a well-defined spatial direction.

Direct Linear Transformation (DLT)

The algorithm solves a system of similarity relations where left and right hand side are equal up to an unknown multiplicative scaling factor.

E

External Beam Radiation Therapy (EBRT)

In contrast to brachytherapy radiation is percutaneously delivered in form of an external treatment beam from a certain distance using a LINAC.

F

Field of View (FoV)

Area or volume in the real world that can be accessed and imaged by a particular imaging modality, e.g. a camera in the simplest case.

Fractionated Stereotactic Radiation Therapy (FSRT)

Special type of SRT in which low dose is delivered temporally distributed across several treatment sessions (fractions). The accumulated dose controls the growth of malignant cells.

Fully Independent Training Conditional (FITC)

Sparse approximation method for GPs. The covariance matrix of training conditional (labels $\boldsymbol{d}_s$ given the inducing variables $\boldsymbol{u}$) is approximated by a diagonal matrix. This means the training data are mutually independent given $\boldsymbol{u}$ (see sec. 2.2.2.7).

G

Gaussian process

A Gaussian process is probability distribution over functions where each subset follows a multivariate Gaussian distribution. Here, the term is used in the context of GP models. These constitute a supervised learning method for regression problems.

Graphics Processing Unit (GPU)

Main microprocessor located on a graphics card. In contrast to a CPU, it is optimized for matrix, vector and floating point operation. Workload can be parallelized.

Gross Tumor Volume (GTV)

Gross palpable or visible/demonstrable extent and location of malignant growth. (definition cf. [133]).

H

Head and Neck (HN)

Anatomical term used to classify tumors of the mouth, lips, nasal cavity, sinuses, salivary glands, throat, larynx, and lymph nodes in the neck.

High Dynamic Range (HDR)

The dynamic range is the relation between the brightest and darkest luminance of a scene. For a camera it refers to the luminance range which can be resolved with the available quantization levels, i.e. gray value resolution. High dynamic range cameras typically have a depth resolution of more than 8 bit.

I

I90

For a given set of absolute errors only 10 % or all errors are larger than this bound (definition sec. sec. 2.2.3).

iid

Independent and identically distributed means that a random variable has the same probability distribution as the others, while they are all mutually independent.

Image Guided Radiation Therapy (IGRT)

Extension of conventional RT that uses image guidance for target localization, patient positioning, verification or motion compensation. OAR can be identified and the high nominal precision of state-of-the-art treatment machines can be brought to the actual target. Common modalities are X-ray based (CBCT, stereo X-ray) or optical techniques (marker-based, marker-less). Additional modalities such as US or MRI have also been investigated.

Infrared (IR)

Light of the electromagnetic spectrum between 780 nm and 1 mm [162].

Intensity-Modulated Radiation Therapy (IMRT)

Extension of conventional RT without a constant dose distribution across the target volume. In contrast, dose intensity is spatially modulated by defining sub-fields of differing dose intensity. This technique is also called dose painting (in 2D or 3D space) and allows better sparing of normal tissue and increased conformity to the tumor shape.

Internal Margin (IM)

Additional safety margin that takes into account variations in size, shape and position of the CTV in relation to anatomical reference points (e.g. bladder/stomach filling, respiratory deformations). The variation sources are exclusively physiological processes. (definition cf. [134]).

Iterative Closest Point (ICP)

The iterative closest point algorithm aims at finding a rigid transformation to bring two sets of 3D points (point clouds or surfaces) into coincidence. This is called registration and can be done using several variants of the algorithm. (for details see sec. 2.3.1).

K

Karush-Kuhn-Tucker (KKT)

Karush-Kuhn-Tucker multipliers extend the concept of Lagrange multipliers, which are only valid for equality constraints. If the optimization problem has inequality constraints the (necessary) Karush-Kuhn-Tucker conditions have to be met, which state that the product of dual variables and primal constraints has to vanish. This is due to the fact that the problem may have a solution not located at the constraint boundary (the equality case). The problem is locally unconstrained in these regions and the multipliers need to vanish.

kilo Voltage (kV)

Voltage in the range of 10^3 Volts that is used to accelerate electrons (also keV).

L

Lateral (LAT)

Axis along the human body - from left to right.

Light Emitting Diode (LED)

Semiconductor element that emits light when driving a current through it.

Linear Accelerator (LINAC)

Device used in RT to accelerate electrons onto a target to generate γ-radiation.

Look-up Table (LUT)

A table which hard-codes the assignments of a specific output value to a fixed set of input values.

M

MAE

Mean absolute value of the differences between the estimated and true results (definition sec. sec. 2.2.3).

Magnetic Resonance (MR)

See MRI.

Magnetic Resonance Imaging (MRI)

Anatomical imaging method based on principles of nuclear magnetic resonance.

Matérn

Function which only depends on the absolute distance r between a point $\mathbf{b}$ and a reference point $\mathbf{b}'$, with $r = \|\mathbf{b} - \mathbf{b}'\|$. Strictly speaking, it is hence an RBF. The function depends on the Gamma function and modified Bessel functions (definition sec. sec. 2.2.2).

maximum a posteriori (MAP)

An estimation technique which computes the maximum of the negative logarithm of a posterior distribution. In Bayesian regression it can be used to derive approaches such as ridge regression or least squares from a probabilistic framework.

ME

Mean of the signed differences between the estimated and true results (definition sec. sec. 2.2.3).

Mega Voltage (MV)

Voltage in the range of 10^6 Volts that is used to accelerate electrons (also MeV).

Monte-Carlo Simulation of light Transport in Multi-Layered Tissue (MCML)

Light transport in a multi-layered tissue model is accomplished by radnom sampling of photon trajectories. The functional framework underlies the radiation transport theory (see sec. 3.2.2.2).

Multi-Task Gaussian process (MTGP)

Multi-task extension of a GP which can be used to model multiple feature-series simultaneously.

Multileaf Collimator (MLC)

Device used to shape the beam of an RT treatment device. It consists of multiple moveable lead leaves for beam shadowing.

N

NBH

Label for items that are related to a feature space with additional incident angle feature and neighborhood information. In addition to five NIR backscatter features, the incident angle is added as a sixth feature. The full neighborhood (all ROIs from all four closest neighbors) results in a feature space dimension of $D = 26$.

Near-Infrared (NIR)

Light of the electromagnetic spectrum between 800 nm and 2500 nm [162].

Negative Log Marginal Likelihood (NLML)

The negative logarithm of the probability $p(\boldsymbol{d}_s|\boldsymbol{B})$, i.e. of the labels given the features. Since this probability depends on the hyperparameters it can be minimized to obtain optimal estimates for these parameters.

O

Organs at Risk (OAR)

Normal tissue with increased radiation sensitivity. It may significantly influence the planning and definition of the prescribed dose distribution. (definition cf. [133]).

P

PC

The posterior commissure in the brain consists of a nerve fiber bundle which connects the two cerebral hemispheres. It is located at the dorsal part of the upper end of the cerebral aqueduct. The AC-PC line is the connecting line between both commissures in midline sagittal images.

Planning Target Volume (PTV)

Volume that takes the net effect of all possible geometric variations and inaccuracies into account. It is defined to select beam size and beam arrangements and ensures that the prescribed dose is actually delivered to the CTV. The overall volume size includes the CTV and also depends on the treatment technique, in order to compensate effects of organ and patient motion, as well as patient setup inaccuracies. It is computed by the union set of CTV, SM and IM (definition cf. [134]).

PMMA

A transparent thermoplastic which can be used as a casting resin. Among others it is known under the trademark Plexiglas.

portal imaging

Imaging techniques that uses MV photons (γ-rays) to generate a 2D image directly through the eye of the therapeutic beam.

Positron Emission Tomography (PET)

Functional imaging method which is based on nuclear injected tracers that emit positrons.

principal component analysis (PCA)

A component analysis which computes the covariance matrix across all dimensions of the data space. From that it estimates the main orientations of scatter within the data. These orientations are required to be mutually orthogonal and are computed by an eigenvalue analysis.

Q

quadratic programming (QP)

A mathematical optimization problem with linear inequality constraints. The objective function contains first and second order terms of the variables to be optimized.

R

radial basis function (RBF)

Function which only depends on the absolute distance r between a point $\mathbf{b}$ and a reference point $\mathbf{b}'$, with $r = \|\mathbf{b} - \mathbf{b}'\|$. In the context of SVR it refers to the special case of the Gaussian kernel function (definition sec. sec. 2.2.2).

Radiation Therapy (RT)

General term for a therapy using ionizing radiation (γ-rays) to treat cancer by killing or controlling the growth of malignant cells. Often also named radiotherapy.

region-of-interest

Subregion within a larger area to which a higher interest and the main focus is directed. The acronym is also used as a label for the basic feature space in sec. 4.4 which uses only NIR features accumulated from different regions-of-interest in a high resolution camera image of backscattered light.

RGB

The RGB color space defines any color by a superposition of three values: a red, green and blue component.

RMSE

Root of the mean of the squared errors (definition sec. sec. 2.2.3).

S

Scientific CMOS (sCMOS)

A CMOS sensor element with a special pixel structure and read-out circuits. High resolution, high dynamic range as well as fast imaging speeds are achieved by simultaneously converting the pixel signal with two 11 Bit analog-to-digital converters (one rough and one fine scale conversion. Joining the information results in high dynamic range sensors.

sequential forward selection (SFS)

A wrapper method for feature selection that sequentially selects the next most informative feature (see sec. 4.4.3).

sequential minimal optimization (SMO)

An algorithm introduced by Platt [228] to efficiently solve the dual problem for SVR optimization. The algorithm sequentially iterates through pairs of Lagrange multipliers and solves the optimization problem by assuming that the others are constant.

Set-up Margin (SM)

Additional safety margin that takes into account uncertainties in patient positioning and beam alignment during planning and all treatment sessions. Uncertainties include mechanical inaccuracies of the equipment (e.g. gantry, mask etc.), human errors, beam geometry selection and expected variations in patient positioning (definition cf. [134]).

Signal-to-Noise ratio (SNR)

The ratio of the information signal power and the power of the overlaying noise. The higher the SNR the more prominent is the signal in the measurement.

Single-photon emission computed tomography (SPECT)

Functional imaging method which is based on nuclear injected tracers that emit γ-rays.

Singular Value Decomposition (SVD)

The singular value decomposition factorizes a real matrix $\boldsymbol{H} \in \mathbb{R}^{m \times n}$ into $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T$ where $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ are real unitary matrices, and $\boldsymbol{\Lambda}$ is an $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal.

SPM8

Statistical Parametric Mapping 8 - a software package for Matlab to analyze brain imaging data sequences.

squared exponential (SE)

Strictly speaking, a special case for an RBF, but practically sometimes used synonymously with the RBF. In the SVR community, the latter is similarly used for

a function that depends on a squared argument in the exponent of a exponential function (definition sec. sec. 2.2.2).

STD

Square root of the average deviation from the mean value of a data sequence.

Stereotactic Body Radiation Therapy (SBRT)

EBRT that uses stereotaxy with the focus on tumors outside of the CNS.

Stereotactic Radiation Therapy (SRT)

Special type of RT which uses the principle of stereotaxy aiming at high precision treatment. It is used to treat HN and CNS cancer. Stereotaxy relies on a coordinate frame in which the position of the tumor is fixed. The frame can be defined by different means such as stereotactic frames, thermoplastic masks or patient anatomy (e.g. skull bone) and can be located externally using medical imaging techniques.

Stereotactic Radiosurgery (SRS)

Special type of SRT which destroys target tissue while preserving adjacent normal tissue. In contrast to general SRT or FSRT, high dose is delivered in one or, in the modern sense, up to five fractions.

Subset of Data (SoD)

One of the simplest sparse approximation methods for GPs. It relies on discarding a proportion of the training data and thus reducing the number from N to $M < N$ (see sec. 2.2.2.7).

Subset of Regressors (SoR)

A sparse approximation method for GPs. It assumes that each test label $f_\star$ can be fully described by a weighted sum of similarities between the inducing variable input $\boldsymbol{B_u}$ and the test input $\boldsymbol{b}_\star$. Similarities are expressed by the kernel function k. (see sec. 2.2.2.7).

Superior-inferior (SI)

Longitudinal axis along the human body - from the soles to the vertex of the head.

Support Vector (SV)

In SVR samples $\boldsymbol{b}_i$ for which $\alpha_i^{(*)} \neq 0$ are called Support Vectors. Only these contribute to the final regression output f.

Support Vector Regression (SVR)

A supervised learning method which can be used for regression problems. It is based on an ε-insensitive loss function leading to a convex optimization problem.

T

T1

In MR imaging a T1 weighted sequence demonstrates the differences in the T1 (spin-lattice) relaxation times of tissues in the voxels of the corresponding output image.

Target Registration Error (TRE)

Apart from registration errors directly at the registration site (e.g. the surface used for surface registration), the target registration error denotes the alignment error for a defined target location after applying the estimated transformation matrix to it. This target might be far away from the registration site and its localization error can be substantially different from that of the registration site (definition according to [115]). For surface tracking, internal targets are typically never involved in the process of registration, i.e. the computation of the transformation matrix itself.

TE

In MR imaging the echo time refers to the time between the radiofrequency excitation pulse and the peak of the response signal induced into the coil.

TR

In MR imaging the repetition time refers to the time between two subsequent radiofrequency excitation pulses.

U

Ultrasound (US)

Imaging method based on ultrasound.

Ultraviolet (UV)

Light of the electromagnetic spectrum between 100 nm and 400 nm [162].

V

Volume-of-Interest (VOI)

3D subregion within a larger volume to which a higher interest and the main focus is directed.

Volumetric Intensity Modulated Arc Therapy (VMAT)

Special type of IMRT that increases treatment efficiency by generating different irradiation directions in a step and shoot procedure. The gantry of the treatment device goes through a minimal number of rotations (arcs) around the tumor isocenter. In each arc the gantry stops at optimized positions and activates the beam. Most efficient irradiation is achieved by RapidArc® (Varian Medical Inc.) which sculpts a 3D dose distribution in a single 360° rotation.

W

World Health Organisation (WHO)

Agency of the United Nations that is specialized in international public health.

X

X-ray

Electromagnetic radiation used for medical imaging with a wavelength $\lambda \in [0.001, 10]$ nm. In the medical context kV and MV imaging is distinguished depending on the energy used to generate the radiation. From a strict physical point of view, the medical term MV X-rays actually refers to γ-rays.

Acknowledgements

Im Rahmen dieser Arbeit möchte ich an erster Stelle meiner Familie und insbesondere meinen Eltern danken. Mit ihrer uneingeschränkten Unterstützung, auch in schwierigen Zeiten, sind sie stets ein unverzichtbarer Rückhalt bei allen meinen Entscheidungen. Diese Arbeit wäre ohne sie nie entstanden.

I also want to express my gratitude to my supervisor Prof. Dr.-Ing. Achim Schweikard for his trust in my work and the valuable experiences I gained in the last four years under his supervision. I would like to thank my colleagues from the Institute for Robotics and Cognitive Systems in Lübeck for the inspiring working environment as well as their thorough support. In particular, I want to mention Dr.-Ing. Robert Dürichen, Patrick Stüber, Dr. rer. nat. Floris Ernst and the remaining members of Varian research group. Many ideas arose from very fruitful and sometimes vigorous discussions with them. I also appreciate their permanent willingness to give critical feedback and constructive input.
All investigations would not have been possible without the work of Patrick Stüber who developed the optical scanning prototype and who was a great support in all the uncountable sessions of data acquisition and during our initial trials. I also acknowledge the work of Benjamin Wagner who developed the triangulation framework being indispensable for the higher level data analyzes. Furthermore, I am particularly grateful for the thoughtful engagement of Cornelia Rieckhoff, who always had a solution for all the administrative challenges.
I further want to express my gratitude to Christian Erdmann, Dr. rer. nat. Uwe Melchert, Dr. med. dent. Christin-Sophie Deutz, Dr. med. Benjamin Sack and Annika Klaus for their time and support during the subject study. In this context I also want thank all volunteers who were willing to participate in this study.
Finally, I want to acknowledge the help of all friends and colleagues who supported me during the writing process of this thesis and provided valuable feedback after proof reading.

Erklärung an Eides Statt

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbständig sowie ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.
Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.
Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.
Ich versichere an Eides Statt, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

(Ort, Datum)	(Unterschrift)