



UNIVERSITÄT ZU LÜBECK
INSTITUTE OF MATHEMATICS AND
IMAGE COMPUTING

Extraktion von SURF-Merkmalen zur Landmarken-basierten Bildregistrierung

*Extraction of SURF-Features for Landmark-Based
Image Registration*

Bachelorarbeit

im Rahmen des Studiengangs
Medizinische Informatik
der Universität zu Lübeck

vorgelegt von

Dennis Labitzke

ausgegeben und betreut von

Prof. Dr. Jan Modersitzki
Institute of Mathematics and Image Computing

mit Unterstützung von

Nadine Traulsen, M.Sc.
Fraunhofer MEVIS, Lübeck

Lübeck, den 16. Juni 2015

Erklärung

Ich erkläre an Eides statt, die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel und Quellen angefertigt zu haben.

Lübeck, den 16. Juni 2015

Kurzfassung

Die Korrektheit der Ergebnisse Landmarken-basierter parametrischer Registrierungen ist zu großen Teilen von der Qualität der zugrunde liegenden Landmarken abhängig. Die manuelle Detektion solcher Landmarken durch medizinische Experten liefert zwar sehr gute Ergebnisse, jedoch ist dies sehr mühselig und zeitaufwendig. Daher wurde im Rahmen dieser Arbeit ein vollautomatisches, auf der Skalentheorie basierendes Verfahren in Matlab implementiert, das SURF-Landmarken detektiert. Die Implementierung detektiert sehr viele Landmarken in Bilddaten, aus denen zur weiteren Verarbeitung die besten ausgewählt werden. Dies geschieht ebenfalls vollautomatisch. Weiterhin wurde eine Möglichkeit implementiert, Korrespondenzen zwischen gefundenen Landmarken herzustellen.

Die Evaluation der Ergebnisse durchgeführter Experimente, sowie deren beispielhafte Anwendung zeigen, dass die Implementierung sehr gut für die automatische Detektion von Landmarken für die Bildregistrierung geeignet ist.

Abstract

The accuracy of landmark-based parametric registrations is to a large extent based on the quality of the landmarks used. While manual detection of such landmarks yields good results, it is really laborious expensive and time-consuming. A fully automatic algorithm based on scale-space theory, detecting SURF-landmarks, was therefore implemented in Matlab. It detects many landmarks in image data from which the best ones are taken for further processing. This also happens fully automatically. Beyond this, a method for matching landmarks of different images has been implemented.

The Evaluation of performed experiments show, that the provided implementation is suitable for automatic detection of landmarks to be used for image registration.

Inhaltsverzeichnis

1	Einleitung	3
1.1	Gliederung der Arbeit	4
2	Theoretische Grundlagen der Arbeit	7
2.1	Formulierung der Problemstellung	7
2.2	Landmarken in der medizinischen Bildverarbeitung	9
2.3	Landmarken-basierte Bildregistrierung	10
3	Detektion von Landmarken nach der SURF-Methode	17
3.1	Detektion von Landmarken	17
3.2	Beschreibung der detektierten Landmarken	21
3.3	Matching von Landmarken in Bildern	24
3.4	Bestimmung der Parameter für die Bildregistrierung	25
4	Ergebnisse und Auswertung an Beispieldaten	27
4.1	Bilddaten für die Experimente	27
4.2	Ergebnisse der Landmarkendetektion	28
4.3	Ergebnisse des Matchings	32
4.4	Registrierung mit SURF-Landmarken	33
5	Diskussion und Ausblick	37
	Abbildungsverzeichnis	41
	Tabellenverzeichnis	41
	Literaturverzeichnis	43

1 Einleitung

„In image processing one is often interested not only in analyzing one image but in comparing or combining the information given by different images. For this reason, image registration is one of the fundamental task within image processing.“ – J. Modersitzki, 2004

Die Registrierung von Bilddaten ist eine der grundlegenden Anwendungen der Bildverarbeitung [Modersitzki, 2004]. In vielen Anwendungsbereichen werden unterschiedliche Informationen über Objekte in Form von Bilddaten erfasst. Dabei wird häufig eine große Anzahl an Bilddaten erzeugt, um möglichst umfangreich und vollständig über dieses Objekt informiert zu sein. Anschließend müssen die Informationen mehrerer dieser Bilder miteinander kombiniert oder verglichen werden. Die Idee der Registrierung ist es, zwei Bilder einander möglichst ähnlich zu machen, so dass sie optimal miteinander vergleichbar sind. Sie ist oft der erste auszuführende Schritt in einer Abfolge von Verarbeitungsschritten. [Werner et al., 2010] stellt vier Anforderungen an Verfahren zur Landmarkendetektion in Aufnahmen der menschlichen Lunge, welche sich zu Anforderungen an generelle Detektionsverfahren verallgemeinern lassen:

- (A1) Das Verfahren funktioniert vollautomatisch.
- (A2) Es werden ausreichend viele Landmarken detektiert, die verifizierbar charakteristische Punkte beschreiben.
- (A3) Die Landmarken liegen gleichmäßig über das Bild verteilt.
- (A4) Landmarken lassen sich verlässlich auf unterschiedliche Bilder des selben Datensatzes übertragen.

Ein praktisches Beispiel aus der Medizin ist die Kombination von Bildern der Computertomografie (CT) und der Positronenemissionstomografie (PET). Bilder aus einem Computertomographen können besonders gut anatomische Strukturen wie Knochen oder Weichteile darstellen. Dagegen können Bilder, die mit einem Positronenemissionstomographen aufgenommen wurden, Stoffwechselaktivitäten anzeigen, jedoch kaum Informationen über anatomische Strukturen liefern. Um diese unterschiedlichen Informationen miteinander zu kombinieren, kann eine Registrierung der Bilder durchgeführt werden. Da die PET- und CT-Bilder typischerweise aus unterschiedlichen Blickwinkeln, zu unterschiedlichen Zeiten und mit verschiedenen Auflösungen aufgenommen werden, ist eine Registrierung der Bilder notwendig. Danach stehen die kombinierten Informationen aus beiden Bildern zur Verfügung.

Ein prominentes Beispiel für die Registrierung von Bildern, die zu unterschiedlichen Zeiten und aus unterschiedlichen Blickwinkeln aufgenommen wurden, sind die in Abbildung 1.1 dargestellten Röntgenaufnahmen der menschlichen Hand. Hier wurden zwei Hände

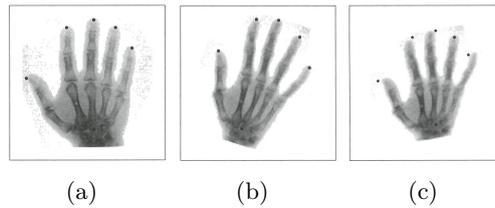


Abbildung 1.1: Aus [Modersitzki, 2004]: Die Hand in (b) wird auf die Hand in (a) registriert. Das Ergebnis ist in (c) dargestellt.

zu verschiedenen Zeitpunkten aufgenommen (Abbildung 1.1 (a) und (b)), wobei in einer Aufnahme die Hand rotiert ist. Nun ist man daran interessiert, dieselben Strukturen in beiden Aufnahmen zu identifizieren, um beispielsweise den Heilungsverlauf von Frakturen zu beobachten. Eine Möglichkeit um Korrespondenzen zwischen den Strukturen in den Bildern herzustellen ist es, die Bilder basierend auf sogenannten Landmarken zu registrieren. In Abbildung 1.1 (c) ist das Ergebnis einer solchen Registrierung dargestellt. Diese Landmarken können nun entweder manuell in einem zeitaufwändigen Prozess von Experten gesetzt werden oder durch halbautomatische oder automatische Methoden erzeugt werden. In dieser Arbeit wird eine Methode betrachtet, die automatisch Landmarken in einem Bild findet und markiert. Automatische Methoden der Landmarkengenerierung sind äußerst wünschenswert, um ein deterministisches und anwenderunabhängiges Verfahren gewährleisten zu können. Das bedeutet, dass bei wiederholter Anwendung eines Verfahrens genau die gleichen Landmarken gefunden werden, die auch schon bei vorherigen Anwendungen generiert wurden. Insbesondere für die Landmarken-basierte Bildregistrierung ist es außerdem wichtig, dass in beiden Bildern tatsächlich korrespondierende Landmarken denselben anatomischen Gegebenheiten entsprechen. Diese wichtigen Aspekte bilden den Grund, aus dem sich die vorliegende Arbeit vor allem mit einem Verfahren zur automatischen Detektion von Landmarken beschäftigt. Dabei ist insbesondere die Implementierung einer Methode zur automatischen Landmarkengenerierung in Matlab wesentlicher Bestandteil dieser Arbeit.

Es gibt zahlreiche Methoden um Landmarken aus einem Bild zu detektieren, so zum Beispiel auch das in [Bay et al., 2008] beschriebene Speeded-Up Robust Features (SURF)-Verfahren. Dabei werden aus einem Bild Landmarken generiert, die neben den reinen Koordinaten auch aus einem beschreibenden Teil besteht, so dass eine Landmarke nicht nur durch den Intensitätswert an ihrer Position definiert ist. Im Rahmen dieser Arbeit wird das SURF-Verfahren vorgestellt und implementiert.

1.1 Gliederung der Arbeit

In Kapitel 2 wird zunächst in Abschnitt 2.1 eine mathematische Formulierung für die Problemstellung der Bildregistrierung eingeführt. Anschließend wird in Abschnitt 2.2 der Begriff der Landmarke definiert und deren Nutzen für die medizinische Bildverarbeitung erläutert. Im darauffolgenden Abschnitt 2.3 wird die auf den zuvor eingeführten Landmarken basierende Registrierung von Bilddaten, sowie deren mathematische Grundlagen

vorgelegt. Im Hauptteil dieser Arbeit wird mit Kapitel 3 ein Verfahren eingeführt, um automatisch Landmarken aus einem Bild zu generieren. Dieses Verfahren ist in zwei Schritte eingeteilt: Die Generierung der Landmarken und deren Beschreibung. Diese Schritte werden in Abschnitt 3.1, beziehungsweise in Abschnitt 3.2 vorgestellt. Im Anschluss wird in Abschnitt 3.3 beschrieben, wie die Landmarken, die in zwei Bildern detektiert wurden, zueinander zugeordnet werden, so dass zwei korrespondierende Landmarken miteinander verpartnert werden. Im darauffolgenden Kapitel 4 werden die Ergebnisse des zuvor vorgestellten Verfahrens zur Landmarkengenerierung anhand von Beispielbilddaten gezeigt und diskutiert. Dafür werden in Abschnitt 4.2 die Ergebnisse der Generierung und in Abschnitt 4.3 die Ergebnisse des Matchings von Landmarken dargestellt. In dem abschließenden Kapitel 5 wird ein Fazit aus den Ergebnissen dieser Arbeit gezogen und ein Ausblick in die Zukunft gegeben.

2 Theoretische Grundlagen der Arbeit

In diesem Kapitel wird der Hintergrund der Arbeit sowie einige ihrer Grundlagen dargestellt. Dabei ist es wie folgt aufgebaut: In Abschnitt 2.1 wird die grundlegende Problemstellung der Bildregistrierung mathematisch formuliert. Anschließend wird in Abschnitt 2.2 der Begriff der Landmarke und deren Nutzen für die Medizinische Bildverarbeitung eingeführt, um zuletzt die Landmarken-basierte Bildregistrierung in Abschnitt 2.3 mit ihren mathematischen Grundlagen zu erklären. Dieser letzte Abschnitt bereitet die Einführung der SURF-Landmarke in Kapitel 3 vor und führt daher insbesondere Integralbilder und den Box Space ein.

2.1 Formulierung der Problemstellung

Die Bildregistrierung beschreibt die Suche nach einer räumlichen Transformation, die ein Bild so deformiert, dass es einem anderen Bild möglichst ähnlich wird. Dabei ist es notwendig, den Begriff des Bildes eindeutig zu definieren.

2.1.1 Formale Definition von Bildern

Jedem Punkt x , der zu einer bestimmten Region $\Omega \subset \mathbb{R}^d$ gehört, wird gerade ein Grauwert $b(x)$ zugeordnet, wobei d die Dimension des Bildes ist. [Modersitzki, 2004] definiert ein kontinuierliches Bild wie folgt:

Definition 2.1.1

Sei $d \in \mathbb{N}$ und $\Omega \subset \mathbb{R}^d$. Das Funktional $b : \Omega \rightarrow \mathbb{R}$ heißt *d-dimensionales Bild*, falls die folgenden Eigenschaften gelten.

1. b hat einen kompakten Träger, d.h. $\text{supp}(b) := \overline{\{x \mid b(x) \neq 0\}} < \infty$,
2. $\forall x \in \Omega: 0 \leq b(x) < \infty$,
3. $\forall k > 0: \int_{\Omega} b(x)^k dx = \infty$.

Die Menge aller Bilder werden mit

$$\text{Img}(d) := \left\{ b : \mathbb{R}^d \rightarrow \mathbb{R} \mid b \text{ ist ein } d\text{-dimensionales Bild} \right\}$$

bezeichnet.

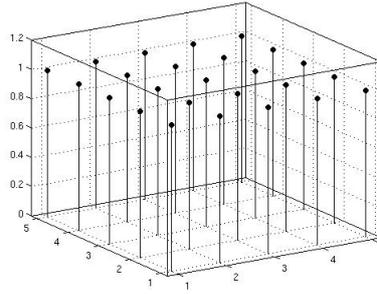


Abbildung 2.1: Abtastfunktion $s(x, y) = \delta(x, y)$.

Für eine Weiterverarbeitung insbesondere auch an einem Rechner sind kontinuierliche Bilder ungeeignet. Stattdessen werden für eine maschinelle Verarbeitung diskrete Bilder betrachtet. Um ein kontinuierliches in ein diskretes Bild umzuwandeln, muss sowohl dessen Definitionsbereich Ω als auch dessen Wertebereich \mathbb{R} diskretisiert werden. Man spricht bei der Diskretisierung des Definitionsbereiches von einer *Abtastung*, beziehungsweise von einer *Quantisierung* bei der Diskretisierung des Wertebereichs.

Abtastung kontinuierlicher Bilder

Ein kontinuierliches Bild wird durch die Abtastung in ein diskretes Bild umgewandelt. Sei $f(x, y)$ ein zu diskretisierendes kontinuierliches Bild definiert auf dem Intervall $\Omega = [1, M] \times [1, N] \subset \mathbb{R}^2$ und $s(x, y)$ die Funktion, mit der das Bild abgetastet werden soll. Ein Bild wird diskretisiert, indem es auf einem regelmäßigen Gitter der Größe $M \times N$ mit Gitterweite $h = 1$ abgetastet wird. Das diskretisierte Bild $f_d(x, y)$ wird berechnet als

$$f_d(x, y) = s(x, y) * f(x, y)$$

mit

$$s(x, y) = \sum_{i=1}^M \sum_{j=1}^N \delta(i, j),$$

wobei δ der Definition der Deltadistribution entspricht. Abbildung 2.1 zeigt eine beispielhafte Abtastfunktion $\delta(x, y)$. Das Ergebnis der mit dieser Funktion durchgeführten Abtastung wäre ein diskretes Bild der Größe 5×5 .

Quantisierung kontinuierlicher Wertebereiche

Um einen Wertebereich zu diskretisieren, wird jeder Intensitätswert $x \in \mathbb{R}$ mit der Funktion $Q(x)$ mit

$$Q(x) = \text{sgn}(x) \left\lfloor |x| + \frac{1}{2} \right\rfloor$$

quantisiert.

Das Problem der Bildregistrierung kann nun wie folgt formal beschrieben werden:

Definition 2.1.2 (*Nach [Modersitzki, 2004]*)

Bezeichne d die Dimension und $\text{Img}(d)$ die Menge aller d -dimensionalen Bilder. Weiter seien ein Ähnlichkeitsmaß $\mathcal{D} : \text{Img}(d)^2 \rightarrow \mathbb{R}$ und zwei Bilder $\mathcal{R}, \mathcal{T} \in \text{Img}(d)$ gegeben. Dabei wird \mathcal{R} als Referenz- und \mathcal{T} als Templatebild bezeichnet. Zu finden ist eine Abbildung $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$, so dass

$$\mathcal{D}(\varphi(\mathcal{T}), \mathcal{R}) = \min$$

gilt.

Betrachtet man ausschließlich die in Definition 2.1.2 gestellten Anforderungen, handelt es sich bei der Registrierung um ein schlecht gestelltes Problem, da es für die geforderte Minimierung mehrere Lösungen geben kann. Zurückgehend auf [Hadamard, 1902] spricht man von einem gut gestellten Problem, wenn für das Problem eine Lösung existiert, diese eindeutig ist und stetig von den Eingangsdaten abhängt. Von einem schlecht gestellten Problem spricht man, sobald mindestens eine dieser Bedingungen nicht erfüllt wird. In den Abschnitten 2.3.3 und 2.3.4 werden Möglichkeiten besprochen, die eine eindeutige Lösung dennoch möglich machen. Um Aussagen über die Ähnlichkeit von Bildern treffen zu können, ist ein präzises Maß notwendig. Ein sogenanntes Ähnlichkeitsmaß wird in Abschnitt 2.3.6 vorgestellt.

2.2 Landmarken in der medizinischen Bildverarbeitung

Der Begriff der Landmarke ist hier angelehnt an die ersten Kapitel des Buches „Landmark-Based Image Analysis: Using Geometric and Intensity Models“ von Karl Rohr ([Rohr, 2001]).

Als *Landmarke* werden auffällige Merkmale in einem Bild bezeichnet, die frei über dieses verteilt liegen. Dabei sind die Landmarken paarweise verschieden. Meist handelt es sich dabei um leicht wiedererkennbare anatomische Strukturen wie beispielsweise bestimmte Punkte oder Windungen. Genauer unterscheidet man zwischen harten, beziehungsweise prospektiven und weichen, beziehungsweise retrospektiven Landmarken ([Modersitzki, 2009]). Harte Landmarken sind solche, die bereits vor der Bildgebung am darzustellenden Objekt angebracht werden, wohingegen weiche Landmarken nachträglich aus dem Bild generiert werden. Diese Arbeit beschäftigt sich mit den sogenannten weichen Landmarken. Formal bezeichne $\mathcal{F}(\mathcal{R}, j)$ die j -te Landmarke im Referenz-Bild \mathcal{R} und $\mathcal{F}(\mathcal{T}, j)$ die j -te Landmarke im Template-Bild \mathcal{T} . Dabei sei $j = 1, \dots, m$, wobei $m \in \mathbb{N}$ die Anzahl der Landmarken in den Bildern \mathcal{R} und \mathcal{T} beschreibt.

Landmarken haben für die medizinische Bildverarbeitung einen großen Nutzen. Sie werden beispielsweise genutzt, um gleiche Strukturen in verschiedenen Bildern wiederzufinden. Im medizinischen Kontext werden Landmarken unter anderem dazu genutzt, um Aufnahmen desselben Patienten zu verschiedenen Zeitpunkten zu vergleichen. Sie nehmen auch einen hohen Stellenwert in der Registrierung von Bilddaten ein. Sie sind die wesentliche Grundlage der parametrischen Bildregistrierung, vergleiche dazu auch

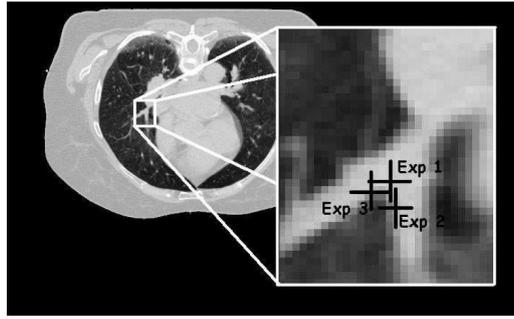


Abbildung 2.2: Aus [Sarrut et al., 2007]: Manuelle Auswahl einer Landmarke durch drei Experten.

[Modersitzki, 2009]. Ein anderes, ebenfalls medizinisches Anwendungsgebiet ist die virtuelle Planung von Hüftoperationen. Dies wird beispielsweise in [Ehrhardt et al., 2003] beschrieben.

Verfahren zur Extraktion von Landmarken werden grundsätzlich in manuelle, halbautomatische und automatische Verfahren unterteilt. Das manuelle Auswählen von Landmarken geschieht meistens durch Experten – im medizinischen Kontext oftmals Ärzte – und ist sehr zeitaufwendig und fehleranfällig ([Werner et al., 2010]). In Abbildung 2.2 sollten drei Experten aus der Medizin in einer CT-Aufnahme der Lunge eine bestimmte, eindeutig identifizierbare Landmarke markieren. In einem Fall wie diesem, in dem die Ergebnisse mehrerer Experten für eine identische Fragestellung stark voneinander abweichen, spricht man von einer hohen Interobserver-Variabilität ([Handels, 2009]). Für viele Fragestellungen ist es wünschenswert, eindeutig reproduzierbare Landmarken erzeugen zu können. Dazu werden oftmals automatische oder semiautomatische Verfahren zur Landmarkengewinnung eingesetzt, da diese schnell viele Landmarke generieren können und somit eine gewisse Robustheit bieten.

2.3 Landmarken-basierte Bildregistrierung

Dieser Abschnitt gibt einen kurzen Überblick über die Landmarken-basierte Bildregistrierung und deren Grundlagen. Orientiert wird sich dabei zu großen Teilen an [Modersitzki, 2004] und [Modersitzki, 2009].

2.3.1 Symmetrie- und Integralbilder

Die Verwendung von Integralbildern spielt für das SURF-Verfahren eine bedeutende Rolle. Das Integralbild zu einem Bild ist definiert als die Summe der Grauwerte des Eingabebildes. Formal lässt sich dies wie folgt beschreiben:

Definition 2.3.1 (*Nach [Viola and Jones, 2001]*)

Sei $\mathcal{I} : [1, \dots, M] \times [1, \dots, N] \rightarrow \mathbb{R}$ ein zweidimensionales Bild der Größe $M \times N$.

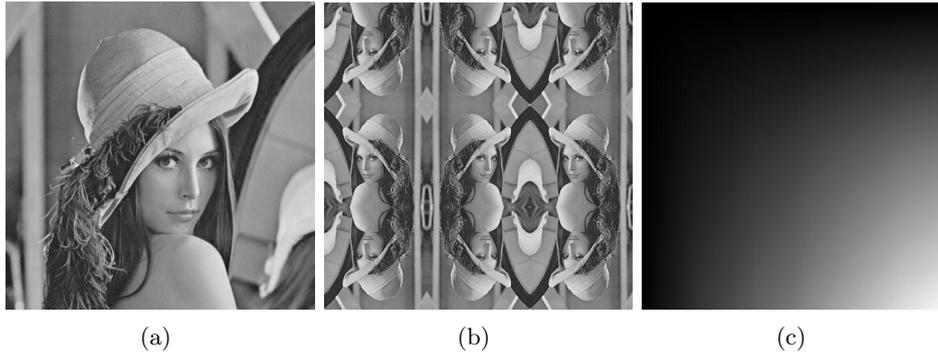


Abbildung 2.3: (a) Das Lena-Bild, sowie dessen (b) Symmetrie- und (c) Integralbild.

Das Integralbild \mathcal{I}_Σ an der Position (x, y) ist definiert als

$$\mathcal{I}_\Sigma(x, y) = \sum_{i=1}^x \sum_{j=1}^y \mathcal{I}(i, j) .$$

In Abbildung 2.3 (c) ist ein beispielhaftes Integralbild dargestellt. Wichtig ist die Definition und die Beachtung von Neumann-Randbedingungen, zum Vergleich siehe [Cheng and Cheng, 2005]. Hierbei wird, wenn für Punkte außerhalb der Bildgröße Werte gesucht werden, der Wert des Punktes gespiegelt am Rand des Bildes genommen. Um dies zu gewährleisten, werden sogenannte Symmetriebilder verwendet. Abbildung 2.3 (b) zeigt beispielhaft ein für die Anwendung von Neumann-Randbedingungen erstelltes Symmetriebild.

2.3.2 Der Box Space

Da das in dieser Arbeit vorgestellte Verfahren seine Vorteile zu großen Teilen aus der Verwendung des Box Space zieht, wird in diesem Abschnitt eine formale Definition für eben diesen geliefert.

Der Box Space kann als eine Anlehnung des in [Lindeberg, 1994] vorgestellten Scale Space gesehen werden. Dieser Scale Space kann anschaulich als eine Pyramide aus übereinander gelegten Varianten des gleichen Bildes interpretiert werden. Dabei wird das Bild je Stufe in der Pyramide in der Größe halbiert und mit einem Gaußfilter geglättet. Ist beispielsweise ein quadratisches Ausgangsbild der Größe $2^{n+1} \times 2^{n+1}$ mit $n \in \mathbb{N}$ gegeben, so ist das Bild auf der zweiten Pyramidenstufe nur noch $2^n \times 2^n$ Pixel groß. Folglich muss um das Bild der n -ten Stufe zu berechnen, das Bild der $n - 1$ -ten Stufe bereits vorliegen. Abbildung 2.4 (a) zeigt den Aufbau der Pyramide des Scale Space. In dieser Arbeit wird genau wie in [Bay et al., 2008] der sogenannte Box Space verwendet. Anders als bei dem Scale Space bleibt bei dem Aufbau des Box Space die Größe des Bildes auf jeder Stufe gleich. Stattdessen wird der Kern des verwendeten Glättungsfilters mit jeder Stufe vergrößert. Dies hat gleich große Bilder auf jeder Stufe zur Folge, die immer stärker

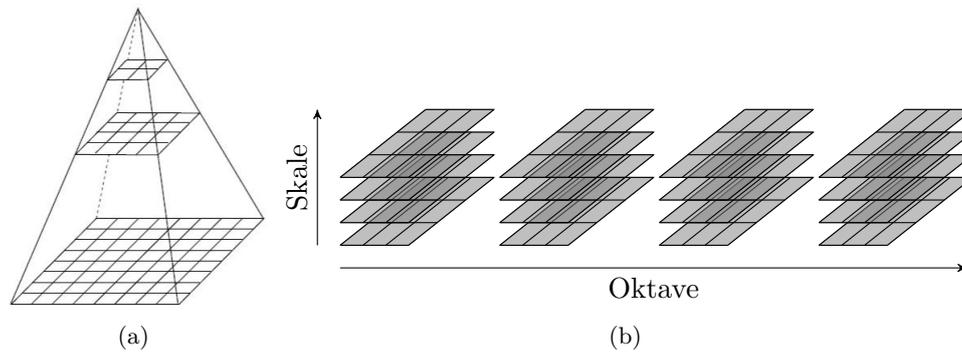


Abbildung 2.4: (a) Aus [Lindeberg, 1994]: Aufbau des Scale Space in Form einer Pyramide mit stetiger Reduktion der Bildgröße. (b) Aufbau des Box Space aus 4 Oktaven mit jeweils 4 Skalen.

geglättet sind. Die grundlegenden Elemente des Box Space sind Skalen und Oktaven. Eine Oktave besteht dabei stets aus vier Skalen, wobei eine Skala dem Begriff der Stufe aus dem Scale Space entspricht. Die Filtergrößen der verschiedenen Skalen werden dabei so gewählt, dass es zwischen den Oktaven zu Überlappungen bezüglich der Filtergrößen kommt. Typischerweise besteht ein Box Space aus drei bis vier Oktaven. Ein beispielhafter Box Space bestehend aus vier Oktaven ist in Abbildung 2.4 (b) dargestellt.

2.3.3 Transformation von Bildern

Die in Definition 2.1.2 bereits erwähnte Abbildung φ beschreibt eine Transformation oder Deformation eines Bildes. Würden keine Anforderungen an die zu findende Transformation φ gestellt, wäre das zuvor beschriebene Minimierungsproblem schlecht gestellt. Durch zusätzliche Anforderungen an die Transformation kann die Schlechtgestellttheit des Problems regularisiert werden. Dabei kann man Anforderungen an die zulässigen Transformationen stellen. Beispielsweise können parametrische Transformationen als Lösung gefordert werden. Genauso können Anforderungen an die Glattheit der Transformation gestellt werden. Da die in dieser Arbeit betrachteten Landmarken wichtiger Bestandteil der parametrischen Bildregistrierung sind, wird im Folgenden noch einmal genauer auf parametrische Transformationen eingegangen. Eine besondere Form parametrischer Transformationen sind rigide Transformationen. Sie erlauben Translationen und Rotationen. Man spricht dabei von Längen- und Winkelerhalt.

Definition 2.3.2 (Nach [Modersitzki, 2009])

- Eine *Transformation* φ ist eine Funktion $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$
- Eine *parametrische Transformation* ist eine Linearkombinationen bestimmter Basisfunktionen q^l , die über die jeweiligen Parameter w^l gewichtet werden.

Ein Beispiel aus [Modersitzki, 2009] ist die eindimensionale, lineare Funktion $y = w_1x + w_2$ mit den Parametern w_1 und w_2 und den Basisfunktionen $q^1(x) = x$ und $q^2(x) = 1$.

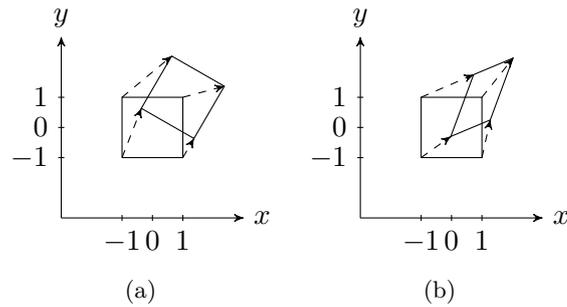


Abbildung 2.5: Beispielhafte rigide und affine 2D-Transformationen. (a) Rigide Transformation, Rotation um 30° und Translation um $(1,1)^T$. (b) Affine Transformation, Rotation um 30° , Scherung entlang der x -Achse um den Faktor 1, Skalierung um den Faktor 0,75 in x - und y -Richtung, Translation um $(1,1)^T$.

Allgemein kann es sehr viele Möglichkeiten geben, ein Bild zu deformieren. Daher werden oft explizite Anforderungen an eine Deformation gestellt.

Definition 2.3.3 (*Nach [Modersitzki, 2004]*)

Eine Deformation φ heißt *rigide*, wenn $\varphi(x) = Qx + b$, wobei $Q \in \mathbb{R}^{d \times d}$ eine orthogonale Matrix mit $\det(Q) = 1$ und $b \in \mathbb{R}^d$ ist.

Anders als rigide Transformationen sind affine Transformationen Kombinationen aus Rotationen, Translationen, Skalierungen und Scherungen. Durch solche Transformationen bleiben Parallelverhältnisse, anders als bei rigiden Transformationen Längen und Winkel aber nicht zwingend erhalten. Siehe dazu auch [van den Elsen et al., 1993].

Definition 2.3.4 (*[Modersitzki, 2004]*)

Eine Deformation φ heißt *affin linear*, falls $\varphi(x) = Ax + b$, wobei $A \in \mathbb{R}^{d \times d}$, $\det(A) \neq 0$ und $b \in \mathbb{R}^d$ gilt.

Abbildung 2.5 zeigt beispielhaft je eine rigide und eine affine Transformation.

2.3.4 Regularisierung

Wie in Abschnitt 2.1 beschrieben, handelt es sich nach Hadamard bei der Registrierung im Allgemeinen um ein schlecht gestelltes Problem. Um ein schlecht gestelltes Problem in ein eindeutig lösbares umzuwandeln, werden weitere Anforderungen an die Transformation in Form eines als Strafterm fungierenden Regularisierers \mathcal{S} eingebracht. Das Registrierungsproblem kann nun einschließlich der Regularisierung wie folgt formuliert werden:

Definition 2.3.5 (*[Modersitzki, 2009]*)

Gegeben sei ein Ähnlichkeitsmaß $D : \text{Img}(d)^2 \rightarrow \mathbb{R}$, zwei Bilder $\mathcal{R}, \mathcal{T} \in \text{Img}(d)$ und der Regularisierer $\mathcal{S} : \text{Img}(d)^2 \rightarrow \mathbb{R}$. Zu finden ist eine Abbildung $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$, so

dass

$$\mathcal{D}(\varphi(\mathcal{T}), \mathcal{R}) + \alpha \mathcal{S}(\varphi(\mathcal{T}), \mathcal{R}) = \min$$

gilt. α bezeichnet den Regularisierungsparameter und gewichtet den Einfluss der Regularisierung gegenüber dem Distanzmaß.

2.3.5 Interpolation von Grauwerten

Im Allgemeinen sind Bilder definiert als eine endliche Menge von Pixelwerten, die auf einem gleichmäßigen Raster angeordnet sind (vgl. Abschnitt 2.1.1). Um korrekt mit Werten umzugehen, die nach der Transformation eines Bildes zwischen dem Pixelgitter liegen, wird eine Interpolationsfunktion $I : \mathbb{R}^d \rightarrow \mathbb{R}$ angewandt. Nach [Bredies and Lorenz, 2011] soll I außerdem folgende Anforderungen erfüllen:

- Die interpolierten Werte liegen innerhalb eines vorgegebenen Intervalls
- I ist einfach zu bestimmen
- $I \in \mathbb{C}^1(\Omega)$, damit eine ableitungsorientierte Optimierung möglich ist

Aufgrund der Schnelligkeit und der Güte ihrer Ergebnisse wird in dieser Arbeit für I die bilineare Interpolation verwendet:

Definition 2.3.6 ([Modersitzki, 2009])

Geben sei ein Bild $\mathcal{I} \in \text{Img}(d)$ dessen Grauwert am Punkt (x, y) als $\mathcal{I}(x, y)$ bezeichnet werde. Sei $p = (x^*, y^*)$ der zu interpolierende Punkt. Sei $\tilde{p}_1 = (x_1, y_1) = (\lfloor x^* \rfloor, \lfloor y^* \rfloor)$ und $\tilde{p}_2 = (x_2, y_2) = (\lfloor x^* \rfloor + 1, \lfloor y^* \rfloor + 1)$. Weiter sei $\xi = (\xi_1, \xi_2) = p - \tilde{p}_1$. Der bilinear interpolierte Grauwert ist definiert als

$$\begin{aligned} \mathcal{I}_I(x^*, y^*) = & \mathcal{I}(x_1, y_1)(1 - \xi_1)(1 - \xi_2) + \mathcal{I}(x_2, y_1)\xi_1(1 - \xi_2) \\ & + \mathcal{I}(x_1, y_2)(1 - \xi_1)\xi_2 + \mathcal{I}(x_2, y_2)\xi_1\xi_2 . \end{aligned}$$

Abbildung 2.6 veranschaulicht das Schema der bilinearen Interpolation.

2.3.6 Distanzmaße für die Bildregistrierung

Nach Definition 2.1.2 soll durch die Registrierung zweier Bilder \mathcal{R} und \mathcal{T} die Distanz zwischen diesen minimiert werden. Bezeichne $\mathcal{F}(\mathcal{R}, j)$, bzw. $\mathcal{F}(\mathcal{T}, j)$ die j -te Landmarke im jeweiligen Bild mit $j = 1, \dots, m$. Die gesuchte Transformation φ , die beide Bilder aufeinander registriert, lässt sich beschreiben als

$$\mathcal{F}(\mathcal{R}, j) = \varphi(\mathcal{F}(\mathcal{T}, j)) .$$

Um nun die Distanz zwischen zwei Landmarkenmengen, beziehungsweise zwischen zwei Bildern mathematisch evaluieren zu können, sind sogenannte Distanzmaße notwendig. Für das in Abschnitt 3.3 beschriebene Verfahren zum Matching von Landmarken ist ebenfalls ein Distanzmaß notwendig. In dieser Arbeit wird als Distanzmaß die Summe

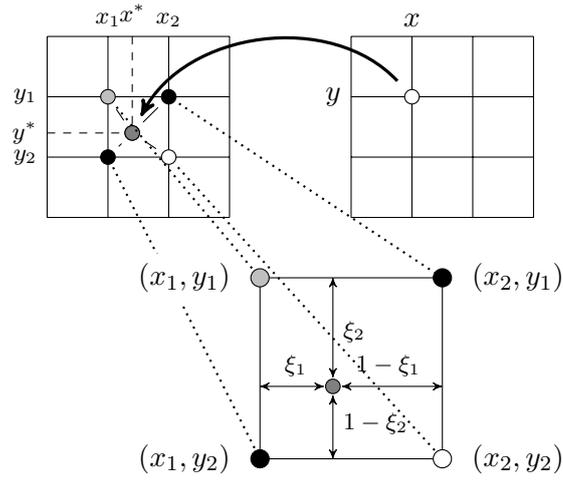


Abbildung 2.6: Nach [Modersitzki, 2009] und [Handels, 2009]: Bilineare Interpolation des Punktes $(x^*, y^*) = \varphi((x, y))$.

der quadratischen Differenzen verwendet. Zwischen zwei Landmarkenmengen kann diese berechnet werden als

$$D(\varphi) = \sum_{j=1}^m \|\mathcal{F}(\mathcal{R}, j) - \varphi(\mathcal{F}(\mathcal{T}, j))\|_2^2. \quad (2.1)$$

Weitere Distanzmaße werden beispielsweise in [Modersitzki, 2009] beschrieben.

3 Detektion von Landmarken nach der SURF-Methode

Nachdem in dem vorherigen Kapitel der Nutzen und die Verwendungsmöglichkeiten von Landmarken in der Medizin veranschaulicht und mathematische Grundlagen erläutert wurden, soll es nun darum gehen, wie man Landmarken automatisch in Bildern detektieren kann. Dabei wird das Verfahren *Speeded-Up Robust Features* (SURF) aus [Bay et al., 2008] vorgestellt.

Dieses Verfahren eignet sich im Kontext der Landmarken-basierten Bildregistrierung besonders für die Detektion von Landmarken, da die SURF-Landmarke nach [Bay et al., 2008] aus zwei Teilen besteht: der Position im Bild und einem Teil, der Bildeigenschaften um die lokale Position der Nachbarschaft beschreibt. Dieser zweite Teil wird nachfolgend wie in [Bay et al., 2008] als Descriptor bezeichnet und kann bei der Registrierung für das Ähnlichkeitsmaß verwendet werden. Im Vergleich zu ähnlichen Methoden, wie beispielsweise das in [Lowe, 2004] vorgestellte SIFT-Verfahren, ist SURF sehr zeiteffizient und die Implementierung invariant gegenüber Rotation und Skalierung, siehe auch [Bay et al., 2008].

In Abschnitt 3.1 wird beschrieben, wie die Position von Landmarken in einem Bild ermittelt werden kann. Anschließend beschreibt Abschnitt 3.2 den Aufbau und die Zusammensetzung des Descriptors. Zuletzt wird in Abschnitt 3.3 eine Methode vorgestellt, wie zwischen den Landmarken zweier Bilder Korrespondenzen hergestellt werden können.

3.1 Detektion von Landmarken

Der in dieser Arbeit beschriebene Algorithmus detektiert Landmarken in Grauwertbildern, die für ein Bild besonders charakteristisch sind. Dabei werden insbesondere Kanten, Krümmungen und prägnante Grauwertunterschiede untersucht. Werden ihm zur Verarbeitung Farbbilder gegeben, so werden diese im ersten Schritt zu Graustufenbildern konvertiert. Erst danach beginnt die eigentliche Detektion der Landmarken im Bild mit dem Aufbau des Box Space.

3.1.1 Aufbau des Boxspace

Die Detektion von *SURF*-Landmarken basiert auf der Hesse-Matrix $\mathcal{H}(x, \sigma)$ mit

$$\mathcal{H}(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix}. \quad (3.1)$$

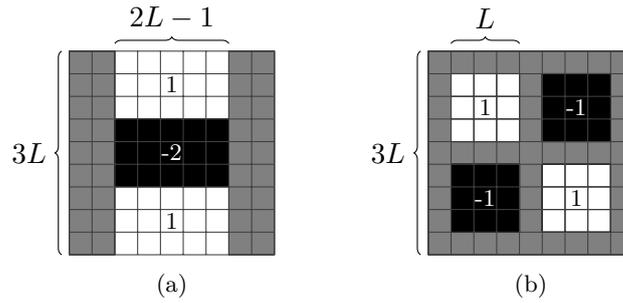


Abbildung 3.1: Aus [Bay et al., 2008]: Boxfilter um Gaußfilter zweiter Ordnung zu approximieren. (a) $L_{yy}(x, \sigma)$. (b) $L_{xy}(x, \sigma)$.

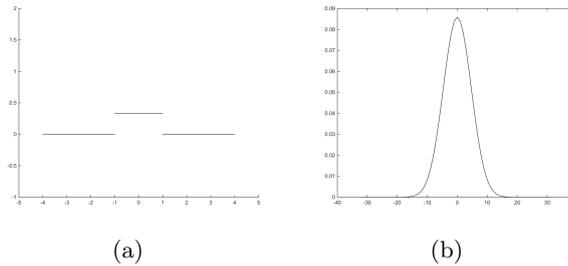


Abbildung 3.2: Annäherung der Boxfilter an die Form der Gaußfilter. (a) Originaler Boxfilter. (b) Form des Filters, nachdem er fünf mal mit sich selbst gefaltet wurde.

Dabei bezeichnen $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$ und $L_{yy}(x, \sigma)$ die Ergebnisse der Faltung von Gauß-Filtern zweiter Ordnung $\frac{\delta^2}{\delta x^2} g(\sigma)$ mit der Standardabweichung σ mit einem Bild I an der Position x in die entsprechende Richtung. Diese Gaußfilter können sehr effizient durch Boxfilter approximiert werden, wie sie in Abbildung 3.1 gezeigt sind. Dies ergibt sich aus der Eigenschaft der Boxfilter, dass diese sich, je öfter sie mit sich selbst gefaltet werden, mit ihrer Form immer stärker an die Form von Gaußfiltern annähern. Abbildung 3.2 zeigt, wie sich mit einem beispielhaften eindimensionalen Boxfilter eine Gaußfilterung approximieren lässt. Dieser Effekt wird in [Gwosdek et al., 2012] weiter beschrieben. Die Ergebnisse der Boxfilterungen werden künftig als D_{xx} , D_{yy} und D_{xy} . Die Suche nach geeigneten Landmarken findet im Box Space statt, der bereits in Abschnitt 2.3.2 definiert wurde. Dieser wird durch die Determinanten der in Gleichung 3.1 definierten Hesse-Matrix berechnet, welche wie folgt durch unterschiedliche Boxfilterungen approximiert werden:

$$\det(\mathcal{H}_{\text{approx}}) = D_{xx}D_{yy} - (0,9D_{xy})^2 .$$

Dabei werden die Filterergebnisse jeweils durch die Frobeniusnorm normiert. Sei $A \in \mathbb{R}^{m \times n}$ eine Matrix der Größe $m \times n$ und a_{ij} das Element dieser Matrix an der Position (i, j) mit $i, j \in \mathbb{N}$, so ist die Frobeniusnorm $\|\cdot\|_F$ der Matrix A definiert als

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} .$$

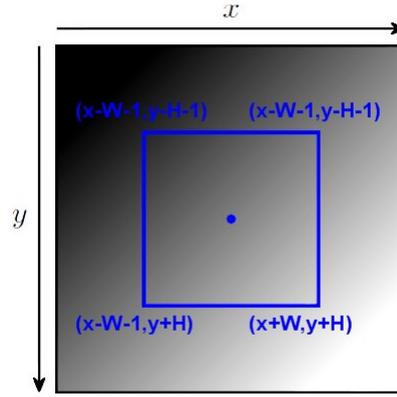


Abbildung 3.3: Effiziente Nutzung eines Boxfilters über Integralbilder.

Der Box Space ist aus sogenannten Oktaven und Skalen aufgebaut, wobei eine Oktave aus vier Skalen besteht. Siehe dazu auch Abschnitt 2.3.2. Je nachdem, in welcher Oktave und in welcher Skala man sich befindet, wird das Bild mit einem Filterkern entsprechender Größe gefaltet. Sei o die aktuelle Oktave und s die aktuelle Skala. Basierend auf dem in Abbildung 3.1 angegebenen L wird die Größe des Filterkerns berechnet als $L = 2^o s + 1$. Die zuvor beschriebenen Boxfilter lassen sich effizient durch die Benutzung von Integralbildern anwenden. Abbildung 3.3 zeigt schematisch, wie die Faltung mit Boxfiltern effizient über Integralbilder berechnet werden können. Bezeichne \mathcal{I} das Eingabebild, \mathcal{I}_Σ das zugehörige Integralbild und D der anzuwendende Boxfilter. Dann lässt sich das Faltungsergebnis $(\mathcal{I} * D)$ an der Position (x, y) berechnen als

$$\begin{aligned} (\mathcal{I} * D)(x, y) &= \mathcal{I}_\Sigma(x - W - 1, y - H - 1) + \mathcal{I}_\Sigma(x + W, y + H) \\ &\quad - \mathcal{I}_\Sigma(x + W, y - H - 1) - \mathcal{I}_\Sigma(x - W - 1, y + H) . \end{aligned} \quad (3.2)$$

Dabei definieren W und H die aus dem Integralbild zu verwendenden Pixel, so dass W und H gerade die Größe des approximierten Filters beschreiben. Hierbei muss zwischen zwei Fällen unterschieden werden: Der erste Fall behandelt die Boxfilter D_{xx} und D_{yy} , der zweite Fall die Boxfilter D_{xy} und D_{yx} . Sei L das in Abbildung 3.1 spezifizierte Maß zur Definition der Größe der Boxfilter. Um die Faltung über Integralbilder zu definieren, muss zunächst einmal geklärt werden, welche Punkte wie in Gleichung 3.2 aus dem Integralbild verwendet werden. Dafür bezeichne $\Gamma = [-W, W] \times [-H, H]$ den in Abbildung 3.3 schematisch dargestellten Bereich in einem Integralbild. Somit lässt sich D_{yy} berechnen als

$$D_{yy} = (\mathcal{I} * D)_{\Gamma_1} - 3(\mathcal{I} * D)_{\Gamma_2} \quad (3.3)$$

mit

$$\Gamma_1 = [-L + 1, L - 1] \times \left[\frac{-L + 1}{2} - L, \frac{L - 1}{2} + L \right]$$

und

$$\Gamma_2 = [-L + 1, L - 1] \times \left[\frac{-L + 1}{2}, \frac{L - 1}{2} \right] .$$

Im Gegensatz dazu kann D_{xy} berechnet werden als

$$D_{xy} = -(\mathcal{I} * D)_{\Gamma_1} - (\mathcal{I} * D)_{\Gamma_2} + (\mathcal{I} * D)_{\Gamma_3} + (\mathcal{I} * D)_{\Gamma_4} \quad (3.4)$$

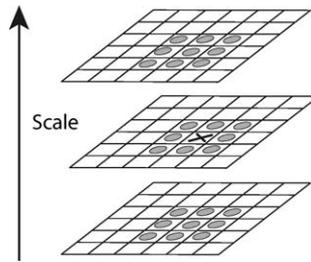


Abbildung 3.4: Aus [Lowe, 2004]: 26 Nachbarn zur Maximabestimmung.

mit

$$\Gamma_1 = [-L, -1] \times [1, L], \Gamma_2 = [1, L] \times [-L, -1], \Gamma_3 = [1, L] \times [1, L]$$

und

$$\Gamma_4 = [-L, -1] \times [-L, -1] .$$

Der soeben aufgebaute Box Space besteht nun aus beliebig vielen Oktaven, die jeweils aus 4 Skalen bestehen. Üblicherweise besteht ein Box Space aus 3 bis 4 Oktaven. In Abbildung 2.4 (b) aus Kapitel 2 ist der grundlegende Aufbau des Box Space gezeigt. Alle weiteren Schritte zur Merkmalsdetektion finden in dem Box Space statt.

3.1.2 Bestimmung der Landmarkenpositionen

Um nun die Position von Landmarken in einem Bild zu ermitteln, wird auf jeder Skala und jeder Oktave in einem abhängig von der aktuellen Oktave gewählten Bildausschnitt

$$\left[1, \dots, \frac{M}{2^{o-1}} - 1\right] \times \left[1, \dots, \frac{N}{2^{o-1}} - 1\right] \quad (3.5)$$

nach lokalen Maxima in den Grauwerten des Bildes im Box Space gesucht. Dabei sei $M \times N$ die Größe des Bildes, in dem nach Merkmalen gesucht wird. Dadurch, dass dies auf unterschiedlichen Oktaven und Skalen geschieht, wird die Anwendung der Glättungsfilter verschiedener Größen approximiert. Zunächst einmal wird jeder Punkt der aktuellen Oktave und Skala im entsprechenden Bildbereich überprüft, ob er über einem bestimmten Schwellwert liegt. Dieser wird berechnet als das arithmetische Mittel aller Intensitätswerte des entsprechenden Eingabebildes. Ist dies der Fall, so wird untersucht, ob es sich bei dem aktuellen Punkt um ein lokales Maximum handelt. Dazu werden alle umliegenden 26 Nachbarn betrachtet, wobei die Nachbarschaft über verschiedene Skalen hinweg definiert ist. Abbildung 3.4 zeigt die Verteilung der 26 betrachteten Nachbarn. Ist der Grauwert des aktuell betrachteten Punktes größer als die aller seiner 26 Nachbarn, so wird er als lokales Maximum betrachtet. Wurde ein Punkt als lokales Maximum markiert, so wird seine Position auf die ursprüngliche Größe des Bildes interpoliert, um die Position des gefundenen Punktes in dem Bild der eigentlichen Größe zu ermitteln.

Bezeichne x und y die Koordinaten eines Punktes im Bild und L die verwendete Größe eines Boxfilters. Um einen Punkt zu interpolieren, wird im ersten Schritt ein dreidimensionaler Vektor $\nabla = (\nabla_x, \nabla_y, \nabla_L)^T$ berechnet, der die drei Gradienten in x -, y - und

L -Richtung repräsentiert. Die Vektorelemente werden dabei wie folgt berechnet, wenn o die zu betrachtende Oktave und i die zu betrachtende Skala des Box Space bezeichnet:

$$\begin{aligned}\nabla_x &= \frac{1}{2} \left(\text{BoxSpace}_{o,i}(x+1, y) - \text{BoxSpace}_{o,i}(x-1, y) \right) \\ \nabla_y &= \frac{1}{2} \left(\text{BoxSpace}_{o,i}(x, y+1) - \text{BoxSpace}_{o,i}(x, y-1) \right) \\ \nabla_L &= 2^{-o-1} \left(\text{BoxSpace}_{o,i+1}(x, y) - \text{BoxSpace}_{o,i}(x, y) \right) .\end{aligned}$$

Im zweiten Schritt wird eine Hesse-Matrix \mathcal{H} aufgestellt mit

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_{xx} & \mathcal{H}_{xy} & \mathcal{H}_{xL} \\ \mathcal{H}_{xy} & \mathcal{H}_{yy} & \mathcal{H}_{yL} \\ \mathcal{H}_{xL} & \mathcal{H}_{yL} & \mathcal{H}_{LL} \end{pmatrix} .$$

Dabei werden die Matrixelemente als

$$\begin{aligned}\mathcal{H}_{xx} &= \text{BoxSpace}_{o,i}(x+1, y) + \text{BoxSpace}_{o,i}(x-1, y) - 2 \cdot \text{BoxSpace}_{o,i}(x, y) \\ \mathcal{H}_{xy} &= \frac{1}{4} \left(\text{BoxSpace}_{o,i}(x+1, y+1) + \text{BoxSpace}_{o,i}(x-1, y-1) \right. \\ &\quad \left. - \text{BoxSpace}_{o,i}(x-1, y+1) - \text{BoxSpace}_{o,i}(x+1, y-1) \right)\end{aligned}$$

berechnet. Schließlich wird der dreidimensionale Vektor δ mit $\delta = (\delta_x, \delta_y, \delta_L)^T = -\mathcal{H}^{-1}\nabla$ berechnet, auf dessen Basis die Interpolation statt findet. Dabei ist \mathcal{H}^{-1} die Inverse der Matrix \mathcal{H} . Nun wird überprüft, ob das Maximum der Beträge der Elemente von δ kleiner eins ist, wobei δ_L zuvor mit dem Faktor 2^o multipliziert wird. Ist dies nicht der Fall, so wird der gefundene Punkt als Merkmal verworfen. Sind die genannten Elemente jedoch alle betragsmäßig kleiner eins, so wird die Position des Punktes folgendermaßen auf die Originalgröße des Bildes interpoliert:

$$\begin{aligned}x_{\text{neu}} &= 2^{o-1} (x + \delta_x) \\ y_{\text{neu}} &= 2^{o-1} (y + \delta_y) \\ L_{\text{neu}} &= L + \delta_L .\end{aligned}$$

Um zu verhindern, dass gefundene Merkmalspunkte im Bild zu dicht beieinander liegen, wird abschließend zur Detektion überprüft, ob in der direkten Nachbarschaft der ermittelten Merkmalsposition bereits Punkte als Merkmal markiert wurden. Ist dies nicht der Fall, wird der aktuelle Punkt zu der Liste der detektierten Merkmale hinzugefügt, ansonsten wird er verworfen.

3.2 Beschreibung der detektierten Landmarken

Um die im vorhergehenden Kapitel extrahierten Merkmale eindeutig identifizieren und verarbeiten zu können, wird ihnen nun ein Merkmalspezifischer, beschreibender Bestandteil zugeordnet, nachfolgend als *Descriptor* bezeichnet.

Der Aufbau des Descriptors besteht im Wesentlichen aus zwei Schritten: Zunächst einmal wird jedem Merkmalspunkt die für ihn dominante Gradientenorientierung berechnet.

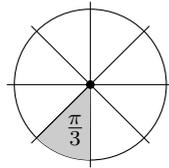


Abbildung 3.5: Das runde Fenster um einen Merkmalspunkt zur Berechnung dessen dominanter Orientierung, nicht maßstäblich.

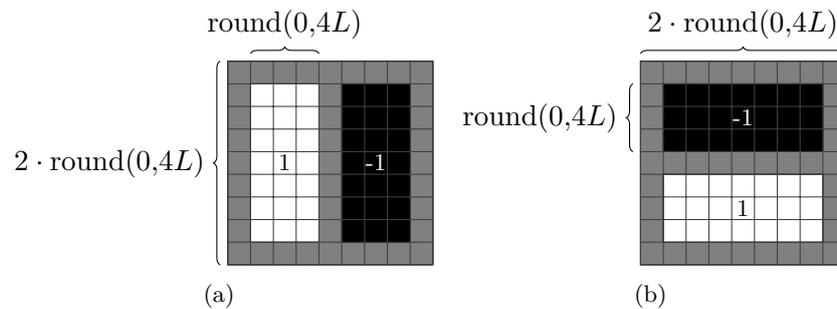


Abbildung 3.6: Nach [Bay et al., 2008]: Boxfilter erster Ordnung in (a) x -Richtung und (b) y -Richtung.

Dies wird in Abschnitt 3.2.1 beschrieben und führt zu einer Rotationsinvarianz der detektierten Landmarken. Anschließend wird in Abschnitt 3.2.2 erläutert, wie der eigentliche Descriptor zusammengesetzt ist und berechnet wird.

3.2.1 Berechnung der Orientierung

Um die Orientierung für einen Merkmalspunkt zu berechnen, wird ein kreisrundes Fenster um diesen Punkt betrachtet, dessen Größe von der Skala abhängt, auf der dieser Punkt detektiert wurde. Dieses Fenster wird wie in Abbildung 3.5 in Kreisabschnitte der Größe $\pi/3$ unterteilt. In jedem dieser Fenster wird für jeden darin liegenden Punkt die Ergebnisse der Faltung mit den in Abbildung 3.6 dargestellten Boxfiltern berechnet. Um eine Gauß-Gewichtung entsprechend der Position des Punktes im Fenster zu erreichen, werden die Ergebnisse mit der Gleichung

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3.6)$$

multipliziert. Dadurch gehen Pixel, die nah am aktuell betrachteten Punkt liegen, stärker gewichtet in die Berechnung ein als solche Pixel, die weiter am Rand der runden Umgebung liegen. Anschließend werden die Faltungsergebnisse innerhalb eines Teilbereiches des Fensters jeweils aufaddiert. Die aufaddierten Faltungsergebnisse werden nachfolgend als

$$\Psi = \begin{pmatrix} \Psi_x \\ \Psi_y \end{pmatrix}$$

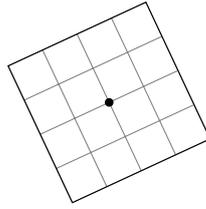


Abbildung 3.7: Umgebung mit 16 Unterfenstern um den Merkmalspunkt mit der dominanten Orientierung 25° , aus dem die Descriptorelemente extrahiert werden.

bezeichnet. Sei

$$|\Psi| = \Psi_x^2 + \Psi_y^2$$

definiert als der Betrag von Ψ . Die Größe der Summen der Faltungsergebnisse eines Bereiches werde als

$$\Psi_{\max} = \begin{pmatrix} \Psi_{x_{\max}} \\ \Psi_{y_{\max}} \end{pmatrix}$$

bezeichnet. Die dominante Orientierung θ eines Merkmalspunktes wird nun berechnet als der zweidimensionale Arcustanges von Ψ_{\max} :

$$\theta = \arctan2(\Psi_{y_{\max}}, \Psi_{x_{\max}})$$

mit

$$\arctan2(y, x) = \begin{cases} \arctan \frac{y}{x} & y > 0 \\ \arctan \frac{y}{x} + \pi & y \leq 0, x < 0 \\ \arctan \frac{y}{x} - \pi & y < 0, x < 0 \\ +\frac{\pi}{2} & y > 0, x = 0 \\ -\frac{\pi}{2} & y < 0, x = 0 \\ \text{undefiniert} & x = 0, y = 0 \end{cases} .$$

3.2.2 Berechnen der Descriptorelemente

Ein Descriptor für einen SURF-Merkmalspunkt besteht aus je 16 mal vier Elementen, die normalisierte Gradienten darstellen und den Merkmalspunkt eindeutig beschreiben.

Um den Descriptor zu berechnen, wird eine Umgebung bestehend aus vier mal vier Unterregionen um den Merkmalspunkt gelegt. Dabei wird die Region entsprechend der dominanten Orientierung des Merkmalspunktes rotiert. Abbildung 3.7 zeigt eine beispielhafte Umgebung eines Merkmalspunktes. Um nun die Descriptorelemente zu berechnen, werden aus jeder dieser Unterregionen fünf gleichmäßig über dieses verteilt liegende Faltungsergebnisse extrahiert, welche über die Faltung mit den in Abbildung 3.6 dargestellten und bereits zur Berechnung der dominanten Orientierung verwendeten Boxfilter ermittelt werden. Das Ergebnis dieser Faltungen ist ein zweidimensionaler Vektor bestehend aus dem Gradienten in x -Richtung und dem Gradienten in y -Richtung. Der Descriptor setzt sich nun so zusammen, dass diese Ergebnisse für alle 25 betrachteten Punkte eines Unterfensters aufsummiert und anschließend mit der bereits in Gleichung 3.6 vorgestellten Funktion

gewichtet werden. Weiterhin werden die Beträge der Faltungsergebnisse aufsummiert und auf die gleiche Weise gewichtet. Sei $\xi = (\xi_x, \xi_y, \xi_{|x|}, \xi_{|y|})^T$ das Descriptorelement eines Unterfensters und d_i die Faltungsergebnisse innerhalb eines Unterfensters mit $i = 1, \dots, 25$. g_i bezeichne den zu verwendenden Gewichtungsfaktor. Das Descriptorelement lässt sich also als

$$\begin{pmatrix} \xi_x \\ \xi_y \\ \xi_{|x|} \\ \xi_{|y|} \end{pmatrix} = \begin{pmatrix} \sum_i g_i d_{xi} \\ \sum_i g_i d_{yi} \\ \sum_i g_i |d_{xi}| \\ \sum_i g_i |d_{yi}| \end{pmatrix}$$

berechnen. Um den berechneten Descriptor vergleichbar zu machen, wird er abschließend noch mit dem Faktor α

$$\alpha = \xi_x^2 + \xi_y^2 + \xi_{|x|}^2 + \xi_{|y|}^2$$

normiert.

3.3 Matching von Landmarken in Bildern

Zwischen den in den beiden vorhergehenden Kapiteln detektierten Merkmalspunkten lassen sich Korrespondenzen herstellen, so dass Merkmale zweier Bilder zu Zweierpaaren gruppiert werden können. Dadurch wird eine Weiterverarbeitung für beispielsweise die parametrische Bildregistrierung möglich. Der Vorgang des Findens von Korrespondenzen wird auch als *Matching* von Landmarken verschiedener Bilder bezeichnet. In dieser Arbeit wird als Matchingmethode das aufwendige Suchen verwendet. Nachdem in zwei Bildern \mathcal{I}_1 und \mathcal{I}_2 m -, beziehungsweise n -viele Landmarken detektiert wurden, werden zwischen den Descriptoren der beiden Landmarkenmengen $\mathcal{F}(\mathcal{I}_1)$ und $\mathcal{F}(\mathcal{I}_2)$ alle möglichen Descriptordifferenzen berechnet, indem die Adjazenzmatrix zwischen den Mengen $\mathcal{F}(\mathcal{I}_1)$ und $\mathcal{F}(\mathcal{I}_2)$ aufgebaut wird. Als Distanzmaß wird hierfür die in Abschnitt 2.3.6, Gleichung 2.1 vorgestellte Summe der quadratischen Differenzen verwendet. Sei $\Delta_{a,b}$ der Abstand zwischen den Landmarken $\mathcal{F}(\mathcal{I}_1, a)$ und $\mathcal{F}(\mathcal{I}_2, b)$ mit $a \in [1, \dots, m]$ und $b \in [1, \dots, n]$. Dann ist die Adjazenzmatrix $\mathcal{D}_{\mathcal{F}(\mathcal{I}_1), \mathcal{F}(\mathcal{I}_2)}$ zwischen den beiden Landmarkenmengen $\mathcal{F}(\mathcal{I}_1)$ und $\mathcal{F}(\mathcal{I}_2)$ wie folgt aufgebaut:

$$\mathcal{D}_{\mathcal{F}(\mathcal{I}_1), \mathcal{F}(\mathcal{I}_2)} = \begin{pmatrix} \Delta_{1,1} & \cdots & \Delta_{1,n} \\ \vdots & \ddots & \vdots \\ \Delta_{m,1} & \cdots & \Delta_{m,n} \end{pmatrix}.$$

Beim Matching wird einer Landmarke $\mathcal{F}(\mathcal{I}_1, a)$ nun genau die Landmarke $\mathcal{F}(\mathcal{I}_2, b)$ als Partner zugewiesen, zu der die kleinste Descriptordifferenz besteht, dass also gerade

$$\Delta_{a,b} = \min \left(\mathcal{D}_{\mathcal{F}(\mathcal{I}_1), \mathcal{F}(\mathcal{I}_2)} \right)$$

gilt, wobei \mathcal{D} das in Abschnitt 2.3.6 beschriebene Distanzmaß bezeichnet.

3.4 Bestimmung der Parameter einer affin-linearen Abbildung für die Bildregistrierung

Um basierend auf den detektierten und gematchten Landmarken zwei Bilder parametrisch aufeinander zu registrieren, muss eine affin-lineare Transformation $\varphi(x)$ ermittelt werden. Nach Definition 2.3.4 wird eine affin-lineare Abbildung beschrieben als $\varphi(x) = Ax + b$ mit $A \in \mathbb{R}^{d \times d}$ und $b \in \mathbb{R}^d$, wobei d die Dimension der Abbildung bezeichnet. Für die in dieser Arbeit betrachteten zweidimensionalen Fall kann eine affin-lineare Abbildung entsprechend mit $d = 2$ als

$$\varphi(x) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (3.7)$$

beschrieben werden. Wenn $m \in \mathbb{N}$ Landmarkenpaare $x^j = (x_1^j, x_2^j)^T$ und $y^j = \varphi(x^j) = (y_1^j, y_2^j)^T$ mit $j \in \{1, \dots, m\}$ gegeben sind, müssen die Parameter in (3.7) so bestimmt werden, dass $Ax^j + b = y^j$ für alle $j \in \{1, \dots, m\}$ gilt. Die Punktepaare und die zu bestimmenden Parameter hängen durch das lineare Gleichungssystem

$$\underbrace{\begin{pmatrix} x_1^1 & x_2^1 & 1 & 0 & 0 & 0 \\ x_1^2 & x_2^2 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^m & x_2^m & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1^1 & x_2^1 & 1 \\ 0 & 0 & 0 & x_1^2 & x_2^2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & x_1^m & x_2^m & 1 \end{pmatrix}}_{=K} \underbrace{\begin{pmatrix} a_{11} \\ a_{12} \\ b_1 \\ a_{21} \\ a_{22} \\ b_2 \end{pmatrix}}_{=u} = \underbrace{\begin{pmatrix} y_1^1 \\ y_1^2 \\ \vdots \\ y_1^m \\ y_2^1 \\ y_2^2 \\ \vdots \\ y_2^m \end{pmatrix}}_{=y} \quad (3.8)$$

zusammen. Da dieses Gleichungssystem für $m > 3$ im Allgemeinen nicht lösbar ist, werden stattdessen die Parameter im Sinne der 2-Norm berechnet, also gerade

$$\hat{u} = \arg \min_u \|Ku - y\|_2 . \quad (3.9)$$

Nach [Modersitzki, 2009] lässt sich (3.9) nun lösen als

$$\hat{u} = (K^T K)^{-1} (K^T y) .$$

4 Ergebnisse und Auswertung an Beispieldaten

In Kapitel 3 wurde das im Rahmen dieser Arbeit implementierte Verfahren zur automatischen Detektion von Landmarken vorgestellt und erklärt. Nun soll dieses Verfahren und die hier entstandene Implementierung durch entsprechende praktische Experimente getestet und die Ergebnisse auf Korrektheit untersucht werden. Abschnitt 4.1 stellt die in den Experimenten verwendeten Bilder dar. Abschnitt 4.2 evaluiert im Anschluss die Ergebnisse der Landmarkendetektion. Die Ergebnisse des in dieser Arbeit in Kapitel 3 vorgestellten Matching-Verfahrens werden in Abschnitt 4.3 besprochen. Abschließend werden in Abschnitt 4.4 die in dieser Arbeit erreichten Ergebnisse exemplarisch für eine parametrische Bildregistrierung angewendet.

4.1 Bilddaten für die Experimente

Neben dem bereits in Abbildung 2.3 gezeigten Lena-Bild¹ werden für die Evaluation des in dieser Arbeit vorgestellten Verfahrens sowohl eigene Bilder, als auch Bilder aus [Mikolajczyk et al., 2005] verwendet. Letztere sind unter [Mikolajczyk et al., 2007] zu finden. Diese Bilder wurden ausgewählt, um verschiedene Eigenschaften des für diese Arbeit implementierten SURF-Verfahrens untersuchen zu können und eine Auswertung unter verschiedenen Bedingungen und mit bestimmten Voraussetzungen zu ermöglichen. Die Abbildungen 4.1 (a) und (b) zeigen das in der Bildverarbeitung bekannte und oft verwendete Lena-Bild einmal im Original und einmal um 20 Grad im Uhrzeigersinn rotiert. Das mit diesem Datensatz durchgeführte Experiment soll die Rotationsinvarianz des implementierten Algorithmus untersuchen. Die Bilder in Abbildung 4.1 (c) und (d) zeigen einen Parkplatz in der belgischen Stadt Löwen. Es handelt sich in beiden Fällen um das gleiche Motiv, bei Bild (d) wurden jedoch Intensitätswerte gleichmäßig reduziert, so dass dieses Bild dunkler erscheint als Bild (c). Hiermit soll die Fähigkeit des Algorithmus, Landmarken in Abhängigkeit des Bildkontextes nicht nur auf absoluten Intensitätswerten basierend zu detektieren, analysiert werden. In Abbildung 4.1 (e) und (f) die Vorderansicht eines Gebäudes der University of British Columbia zu sehen. Bild (f) weist deutliche Artefakte der JPEG-Kompression auf. Es soll gezeigt werden, dass der implementierte Algorithmus auch invariant gegenüber derartiger Artefakte ist. Die mit dem SURF-Verfahren detektierten Landmarken können ebenfalls dazu verwendet werden, Objekte in Bildern zu erkennen und zu markieren. Die in Abbildung 4.1 (g) und (h) gezeigten Bilder eines KFZ-Kennzeichens und der darauf befindlichen Plakette werden beispielhaft für eine solche Objekterkennung verwendet. Zusätzlich zu den in Abbildung 4.1 gezeigten

¹<http://sipi.usc.edu/database/database.php?volume=misc&image=12#top>

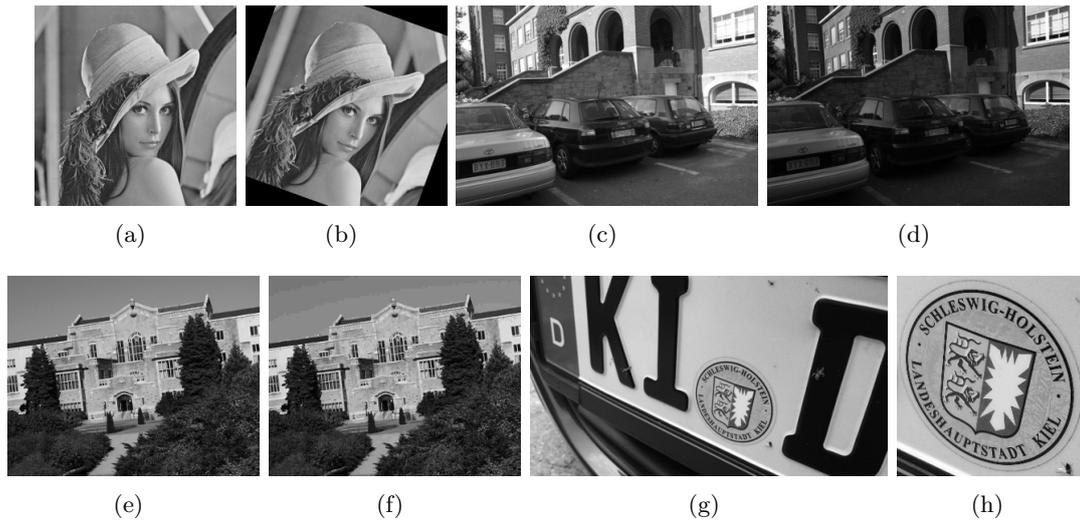


Abbildung 4.1: Die für die Experimente verwendeten Bilddaten.

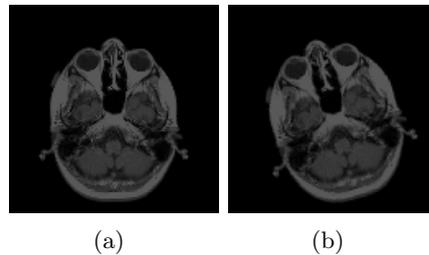


Abbildung 4.2: (a) MRT-Bild des Kopfes und (b) die gleiche Schichtaufnahme rotiert.

Bildern wurde der in dieser Arbeit implementierte Algorithmus auch mit einem medizinischen Bild getestet. Abbildung 4.2 (a) zeigt eine in einem Magnetresonanztomografen (MRT) entstandene Abbildung des menschlichen Kopfes, während in Abbildung 4.2 (b) die gleiche Aufnahme um 15 Grad entgegen dem Uhrzeigersinn rotiert dargestellt ist. Dieser Datensatz kann in MATLAB[®] mit dem Befehl `load mri` geladen werden. Siehe dazu auch [The Mathworks Inc., 2015]. Diese Bilder werden ebenfalls verwendet, um den implementierten Algorithmus zu testen.

4.2 Ergebnisse der Landmarkendetektion

Hier werden nun die Ergebnisse der mit den im vorherigen Abschnitt vorgestellten Bilddaten durchgeführten Experimente vorgestellt. Jedes Experiment analysiert dabei einen Anwendungsfall mit bestimmten Bedingungen und Voraussetzungen, siehe dazu auch Abschnitt 4.1. Zudem soll ein zuletzt durchgeführtes Experiment den Umgang des SURF-Verfahrens mit praxisnahen Bildern zeigen. Um die Ergebnisse der Landmarkendetektion zu visualisieren, werden für jedes Experiment beide in Abschnitt 4.1 gezeigten Bilder mit den Positionen der detektierten Landmarken dargestellt. Weiterhin wird um jede

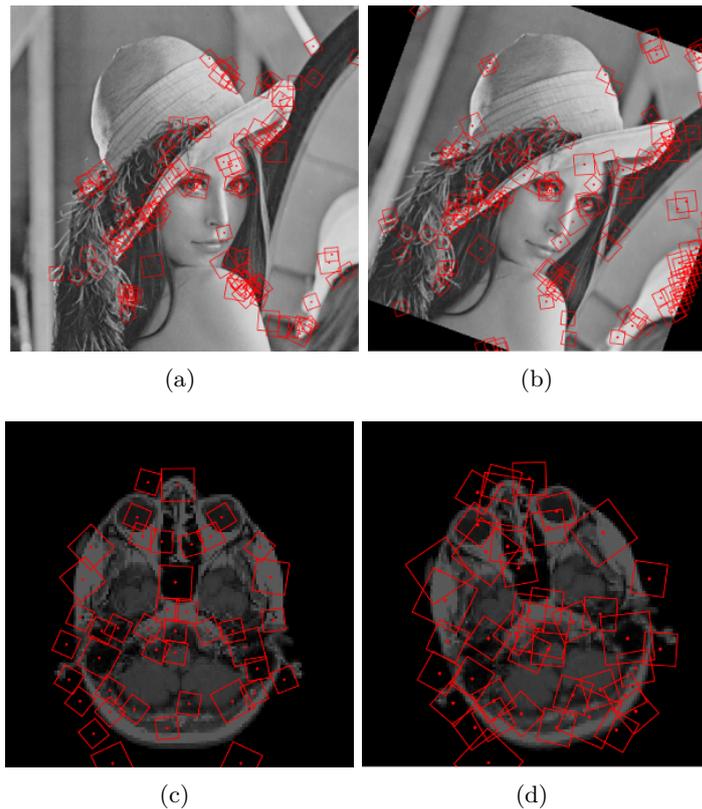


Abbildung 4.3: Der Lena- und der MRT-Datensatz mit detektierten Landmarken. (a) Das originale Lena-Bild mit 114 Landmarken, (b) das rotierte Lena-Bild mit 139 Landmarken, (c),(d) Die Bilder des MRT-Datensatzes mit je 42 Landmarken.

Landmarke ein Quadrat bestimmter Größe und Rotation gezeichnet, wobei die Größe des Quadrates gerade die Skala repräsentiert, auf der die entsprechende Landmarke extrahiert wurde. Dabei entspricht die Seitenlänge des Quadrates dem zehnfachen des Index der Skala, auf der die Landmarke detektiert wurde. Die Winkel, um welche die Quadrate rotiert wurden, repräsentieren die dominante Orientierung der entsprechenden Landmarke, vgl. Abschnitt 3.2.1.

4.2.1 Experiment 1: Rotationsinvaranz

Abbildung 4.3 zeigt das Lena-Bild im Original, sowie eine um 20° im Uhrzeigersinn rotierte Variante, jeweils mit den in dem entsprechenden Bild detektierten Landmarken. Nachdem das originale Lena-Bild rotiert wurde, kommt es in den Ecken des gedrehten Bildes zu schwarzen Dreiecken, die den Intensitätswert Null annehmen. Dadurch kommt es zur Detektion von Landmarken an den Übergängen zwischen dem eigentlichen Bild und den schwarzen Rändern, welche nicht weiter beachtet und in die nachfolgenden Berechnungen zur Evaluation nicht eingehen. Abbildung 4.4 (a) zeigt die detektierten Landmarken beider Lena-Bilder in einer Darstellung. Die Landmarken des originalen Bildes sind

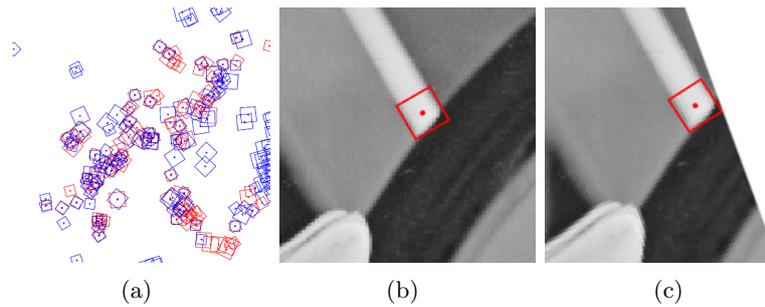


Abbildung 4.4: (a) Überlappung der in dem originalen und dem rotierten Lena-Bild detektierten Landmarken. (b), (c) Der gleiche Ausschnitt in beiden Bildern mit einer hervorgehobenen Landmarke.



Abbildung 4.5: Ergebnisse des Experimentes zur Intensitätsinvarianz. Die Bilder aus der Stadt Löwen mit detektierten Landmarken.

dabei in rot und die des rotierten Bildes in blau dargestellt. Dabei wurden die Position der in dem rotierten Bild detektierten Landmarken um 20° entgegen dem Uhrzeigersinn gedreht, um die Rotation auszugleichen, durch die das Bild generiert wurde und die Landmarkenmengen entsprechend miteinander vergleichen zu können. Es lässt sich sehr gut sehen, dass zwischen beiden Landmarken-Mengen viele gemeinsame Pixel markiert wurden, deren Orientierungen zueinander sehr ähnlich sind. Um dies noch genauer beurteilen zu können, ist in Abbildung 4.4 (b) und (c) je der gleiche Ausschnitt beider Lena-Bilder dargestellt. Betrachtet man in diesen Bildausschnitten die Landmarke, die sich an dem unteren Ende der dicken weißen Linie befindet, wird deutlich, dass beide Landmarken sowohl auf der gleichen Skala, als auch mit der gleichen Orientierung detektiert wurden. Die Abbildungen 4.3 (c) und (d) zeigen die Schicht aus der MRT Aufnahme im Original und um 15° entgegen dem Uhrzeigersinn rotiert. Hier wurden genau wie bei dem Lena-Datensatz annähernd komplett die gleichen Landmarken mit sehr ähnlichen Orientierungen detektiert.

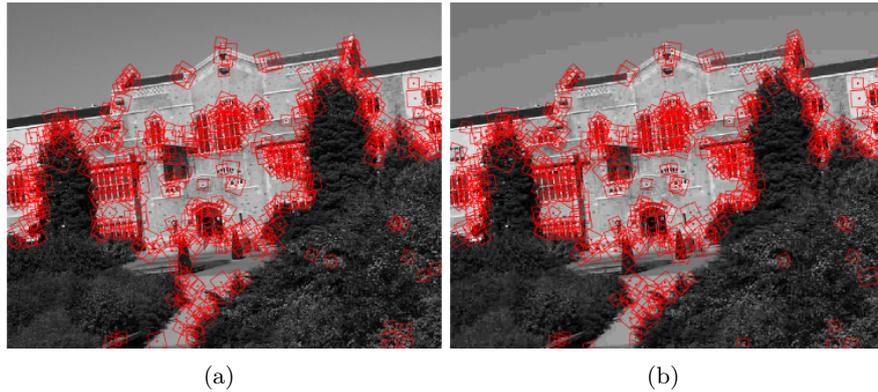


Abbildung 4.6: Ergebnisse des Experiments zum Einfluss von Kompressionsartefakten.

4.2.2 Experiment 2: Invarianz der Intensitätswerte

Mit diesem Experiment soll das Verhalten der Implementierung des SURF-Verfahrens auf Änderungen der Intensitätswerte innerhalb der zu registrierenden Bilder untersucht werden. In Abbildung 4.5 ist das Ergebnis des Experiments abgebildet. Die Abbildung (a) zeigt 462 detektierte Landmarken auf dem helleren der beiden Bildern. In Bild (b), dessen Intensitätswerte im Vergleich zu Bild (a) gleichmäßig reduziert wurden, sind 508 Landmarken detektiert worden. Es wird deutlich, dass in beiden Bildern sehr viele identische Landmarken detektiert wurden, da der Algorithmus als Landmarken insbesondere Kanten und Grauwertübergänge in Betracht zieht, welche sich bei einer gleichmäßigen Reduzierung beziehungsweise Erhöhung aller Grauwerte nicht verändern.

4.2.3 Experiment 3: Einflüsse von Artefakten

Mit diesem Experiment sollten die Auswirkungen von Artefakten, wie sie bei der JPEG-Kompression, vgl. [Golestaneh and Chandler, 2014], vorkommen. Die Ergebnisse sind in Abbildung 4.6 gezeigt. Dabei handelt es sich bei Bild (a) um das unkomprimierte Foto mit 613 und bei Bild (b) um das stark komprimierte Foto mit 595 detektierten Landmarken. Das komprimierte Foto weist entsprechend starke Artefakte auf. Die Abbildung zeigt, dass auch in diesem Experiment zwei nahezu identische Landmarkenmengen detektiert wurden.

4.2.4 Experiment 4: Objekterkennung mit SURF-Landmarken

Mit dem zuletzt durchgeführten Experiment sollte der potentielle Nutzen von SURF-Landmarken für die Objekterkennung als ein weiteres Anwendungsgebiet neben der Bildregistrierung untersucht werden. Abbildung 4.7 (a) und (b) zeigt die für dieses Experiment verwendeten Bilddaten sowie die als Ergebnis detektierten Landmarken. Die stark unterschiedliche Anzahl an Landmarken in beiden Bildern resultiert aus den unterschiedlichen für die Detektion berechneten Grenzwerten und der sich stark unterscheidenden

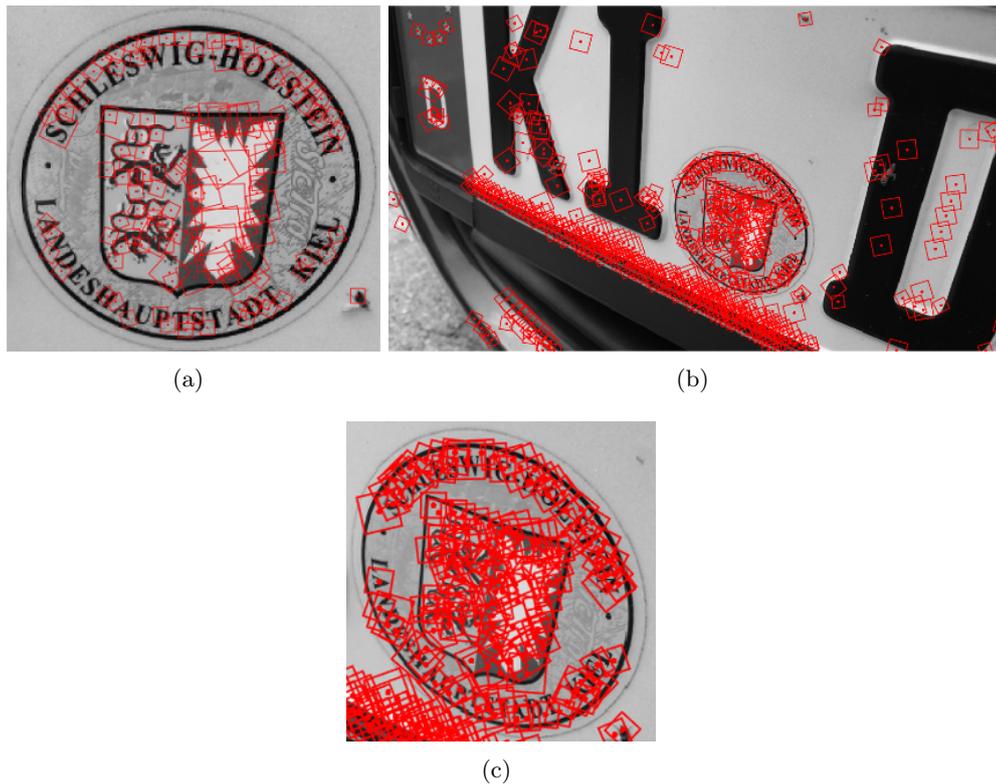


Abbildung 4.7: Ergebnisse der Landmarkendetektion im Experiment zur Objekterkennung mit SURF-Landmarken.

Bildgrößen. In Abbildung 4.7 (c) ist ein Ausschnitt aus Abbildung 4.7 (b) zu sehen, der gerade das zu erkennende Objekt aus Bild (a) enthält. Gerade bei dem oberen und unteren Schriftzug wird deutlich, dass in den Bildern die gleichen Punkte als Landmarke markiert worden sind.

4.3 Ergebnisse des Matchings

In diesem Abschnitt werden die Ergebnisse des Matchings der zuvor detektierten Landmarken vorgestellt. Abbildung 4.8 zeigt die Ergebnisse des Matchings aus allen verwendeten Datensätzen, während in Tabelle 4.1 statistische Informationen über diese Ergebnisse aufgeführt sind. Dabei werden für alle fünf Datensätze, die für die Experimente verwendet wurden, die Anzahl der in beiden Bildern jeweils detektierten Landmarken, die Anzahl der verpartnerten Landmarken und der prozentuale Anteil korrekt zugeordneter Landmarkenpaare angegeben. Am besten hat das Matching mit dem Datensatz der Bilder (e) und (f) funktioniert. Hier kam es zu keinen Fehlern beim Matching, 100% der Verpartnerungen waren korrekt. Die schlechtesten Ergebnisse lieferte das Matching der in den Bildern (g) und (h) detektierten Landmarken. Mit fünf Fehlern ergibt sich für diesen Datensatz eine Matching-Korrektheit von 94%. Gemittelt über alle fünf durchgeführten

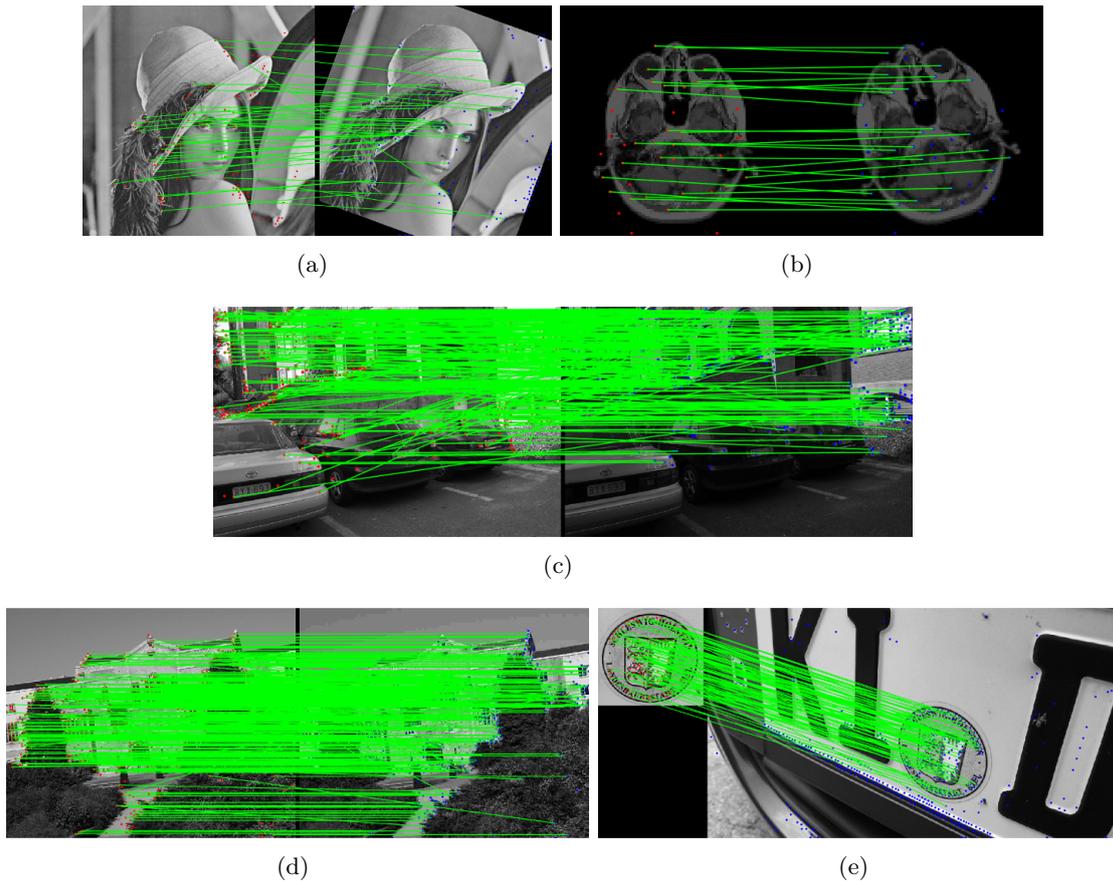


Abbildung 4.8: Ergebnisse des Matchings für alle vier Datensätze.

Experimente ergibt sich eine durchschnittliche Korrektheit der Matchingergebnisse von 97%.

4.4 Registrierung mit SURF-Landmarken

Die Ergebnisse der im Rahmen dieser Arbeit entstandenen Implementierung der SURF-Methode können nun dazu verwendet werden, zwei Bilder zu registrieren. Dafür werden aus der Menge der detektierten Landmarkenpaare zunächst die fünf besten Matches ausgewählt. Dazu werden die Paare anhand der bereits für das Matching berechneten Distanzen zwischen zwei Landmarken aufsteigend angeordnet um anschließend die fünf ersten Paare zu bestimmen. Die so ausgewählten Punktepaare werden benutzt, um mit der in Abschnitt 3.4 vorgestellten Methode die Parameter der affin-linearen Transformation zu berechnen, mit der die Registrierung durchgeführt wird. Dabei gilt in Gleichung 3.8 mit fünf ausgewählten Landmarkenpaaren $m = 5$. In Abbildung 4.9 ist der Vorgang der Registrierung des in Abbildung 4.1 (a) und (b) gezeigten Lena-Datensatzes gezeigt. In (a) sind beide Bilder mit den jeweiligen Landmarken, die zu den fünf am besten gematchten Paaren gehören, dargestellt. Bei Abbildung (b) handelt es sich um das originale

Tabelle 4.1: Evaluation des Matchingverfahrens

Datensatz aus Abb. 4.1 bzw. 4.2	(a), (b)	(c), (d)	(e), (f)	(g), (h)	Kopf-MRT
Landmarken Bild 1	114	462	613	138	42
Landmarken Bild 2	139	508	595	439	42
Paare	45	45	298	90	24
Korrekte Paare	98%	98%	100%	94%	96%
Durchschnittliche Korrektheit				97%	

Lena-Bild überlagert von dem Ergebnis der durchgeführten Registrierung. Hiermit soll die Genauigkeit der Registrierung, sowie die Korrektheit des Ergebnisses ausgewertet werden. Zusätzlich dazu zeigt (c) das Differenzbild, berechnet aus dem Lena-Bild als Referenzbild und dem Registrierungsergebnis. In Abbildung (d) wurden nur die 5 verwendeten Landmarkenpaare transformiert und auf dem originalen Lena-Bild dargestellt. Mit den gezeigten 5 Landmarkenpaaren ergab sich folgende durchzuführende affin-lineare Transformation:

$$\varphi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0,9360 & 0,3404 \\ -0,3394 & 0,9367 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -69,7424 \\ 102,4414 \end{pmatrix}.$$

Da der Winkel, um den das originale Lena-Bild rotiert wurde um das zweite Bild zu generieren 20° beträgt, ist die perfekte Rotationsmatrix als

$$\begin{pmatrix} \cos(20^\circ) & \sin(20^\circ) \\ -\sin(20^\circ) & \cos(20^\circ) \end{pmatrix} = \begin{pmatrix} 0,9397 & 0,3420 \\ -0,3420 & 0,9397 \end{pmatrix}$$

gegeben. Um die Güte der errechneten Rotationsmatrix zu untersuchen, wird deren absolute Differenz mit der Referenzrotation berechnet:

$$\left| \begin{pmatrix} 0,9360 & 0,3404 \\ -0,3394 & 0,9367 \end{pmatrix} - \begin{pmatrix} 0,9397 & 0,3420 \\ -0,3420 & 0,9397 \end{pmatrix} \right| = \begin{pmatrix} 0,0037 & 0,0016 \\ 0,0026 & 0,0030 \end{pmatrix}.$$



(a)



(b)

(c)

(d)

Abbildung 4.9: Parametrische Registrierung des Lena-Bildes.

5 Diskussion und Ausblick

In diesem Kapitel wird diskutiert, wie die im Rahmen dieser Arbeit entstandene Implementierung des SURF-Verfahrens die zu Anfang in Kapitel 1 gestellten Anforderungen an Detektionsverfahren erfüllt. Weiterhin werden die in Kapitel 4 beschriebenen Ergebnisse interpretiert.

Die in dieser Arbeit vorgestellte Implementierung des SURF-Verfahrens kann in beliebigen Eingabebildern Landmarken detektieren und diesen eine Descriptor genannte beschreibende Komponente zuweisen. Um Landmarken zu detektieren und zu beschreiben, benötigt das Verfahren eine Reihe von Parametern, die geeignet zu wählen sind. So besteht beispielsweise die Möglichkeit, die Anzahl der zu verarbeitenden Oktaven und Skalen, sowie die zu verwendende Filtergröße angepasst werden. Dabei wurde sich für die entstandene Implementierung an die Parameterwerte gehalten, die in [Bay et al., 2008] vorgeschlagen wurden, welche sich für alle getesteten Fälle als geeignet erwiesen haben. Der einzige variable Parameter ist ein Grenzwert, der die Auswahl potentieller Landmarkenpositionen reguliert. In der Implementierung in dieser Arbeit wurde dieser Wert als der durchschnittliche Intensitätswert des Eingabebildes gewählt, was sich empirisch als geeignet erwiesen hat. In Zukunft könnte man auf großen Datenmengen untersuchen, ob die Parameter auch für die Praxis geeignet gewählt sind. Außerdem könnte der Grenzwert so gewählt werden, dass er nicht von den Werten eines Eingabebildes abhängt, sondern für alle Anwendungsfälle gleich ist. Insgesamt erfüllt das hier vorgestellte Verfahren die Anforderung (A1) der vollständig automatischen Detektion von Landmarken. Zur Ausführung dieses Algorithmus muss kein medizinischer Experte anwesend sein, um etwaige Einstellungen vorzunehmen oder sogar Landmarken manuell zu setzen. Daraus resultiert zum einen eine enorme Zeitersparnis, da eine manuelle Wahl von Landmarken wie in Kapitel 1 erwähnt sehr mühsam und zeitaufwendig ist. Zum anderen resultiert aus der vollautomatischen Detektion, dass der Algorithmus mit den gleichen Eingabedaten die gleichen Landmarken detektiert. Entsprechend ist das Ergebnis reproduzierbar.

Anforderung (A2) erwartet die Detektion ausreichend vieler Landmarken, bei denen es sich um verifizierbar charakteristische Punkte des Eingabebildes handelt. Die Anzahl der detektierten Landmarken hängt neben der Größe des Eingangsbildes ebenfalls von dem bereits im vorherigen Absatz genannten veränderlichen Parameter ab. Dieser hat Einfluss darauf, ob ein Punkt als potentielle Landmarke markiert wird. Dies geschieht anhand der einfachen Überprüfung, ob der Wert der potentiellen Landmarke im Integralbild (vgl. Abschnitt 2.3.1) über dem Grenzwert liegt. Bei den durchgeführten Experimenten, vgl. Tabelle 4.1, wurden je Bild durchschnittlich 274 Landmarken detektiert. Da für das in Abschnitt 3.4 beschriebene Verfahren zur Berechnung der Registrierungsparameter mindestens drei Landmarken je Bild notwendig sind, detektiert die Registrierung in jedem Fall ausreichend viele SURF-Landmarken um eine affin-lineare Abbildungsvorschrift zu bestimmen. Die Abbildungen 4.3, 4.5, 4.6 und 4.7 zeigen, dass in jedem Bild Punkte als

Landmarken markiert wurden, die in Abschnitt 3.1 als charakteristisch für ein Bild bezeichnet wurden. Damit ist auch diese Anforderung erfüllt. Diese Abbildungen bestätigen ebenfalls, dass die detektierten Landmarken gleichmäßig über dem Eingangsbild verteilt liegen. Der in Kapitel 3 eingeführte Mindestabstand zwischen Landmarken begünstigt dies zudem, indem er dazu führt, dass nicht zwei dicht benachbarte Punkte als Landmarken erkannt werden. In Zukunft ist zu untersuchen, ob dieser einzuhaltende Mindestabstand nicht auch dazu führen kann, dass in zwei Bildern, die beispielsweise zueinander verschoben oder rotiert sind, nicht exakt die gleichen Landmarken detektiert werden, wenn in zwei benachbarte Punkte in den beiden in unterschiedlicher Reihenfolge betrachtet werden. Da die in Abschnitt 3.1 genannte Charakteristika von Bildern in der Regel gleichmäßig über das Bild verteilt liegen, sind auch die detektierten Landmarken gleichmäßig über das Bild verteilt. Dementsprechend ist auch die Anforderung (A3) erfüllt. Mit der Anforderung (A4) wird verlangt, dass Landmarken auch auf andere Bilder übertragen werden können. Dies wird durch das in Abschnitt 3.3 vorgestellte Verfahren zur Ermittlung von Korrespondenzen zwischen Landmarkenmengen möglich. In Tabelle 4.1 ist für dieses Matchingverfahren eine durchschnittliche Korrektheit von 97% angegeben. Durch diese gut funktionierende Implementierung der Detektion von Korrespondenzen in Landmarkenmengen wird auch die Anforderung (A4) erfüllt.

Die durch die entstandene Implementierung generierten Ergebnisse wurden exemplarisch in Form einer parametrischen Registrierung eines Datensatzes angewendet, siehe dazu auch Abschnitt 4.4. Wie zuvor erwähnt, detektiert die entstandene Implementierung deutlich mehr Landmarken als für die Registrierung notwendig wären. Dadurch, dass die Auswahl der für die Registrierung zu verwendenden Landmarken automatisch basierend auf einem Distanzmaß erfolgt, liefert die Registrierung ein gutes Ergebnis. Eine perfekte Registrierung hätte zur Folge, dass das Differenzbild aus Referenz- und Ergebnisbild an jeder Stelle den Intensitätswert Null annimmt, bei der Visualisierung also komplett schwarz ist. Für das gezeigte Beispiel bedeutet eine perfekte Registrierung die Berechnung eines Rotationswinkels von 20° bedeuten. Dass das in Abbildung 4.9 (c) dargestellte Differenzbild nicht komplett schwarz ist, obwohl offensichtlich in Folge der Registrierung eine Rotation um den richtigen Winkel durchgeführt wurde, ist auf die Verarbeitung diskreter Bilder und die damit verbundene Ungenauigkeit während der notwendigen Abtastung (vgl. Abschnitt 2.1.1) zurückzuführen. Zu derartigen Fehlern kommt es insbesondere wenn für die Registrierung eine Rotation notwendig ist. Die dafür nötigen Winkelberechnungen lassen sich maschinell nur mit einer endlichen Genauigkeit durchführen.

In Anforderung (A2) wird unter anderem eine Verifizierung gefordert, wie charakteristisch die detektierten Landmarken für die entsprechenden Eingangsdaten sind. Zum aktuellen Zeitpunkt kann dies nur durch manuelles Betrachten der Ergebnisse ausgewertet werden. Für die Zukunft wäre es wünschenswert, diese Beurteilung automatisch möglich zu machen. So könnte beispielsweise die unter anderem in [Handels, 2009] beschriebene Maßzahl des Target-Registration-Errors auf ein Ergebnis angewendet werden, sofern bereits eine Referenzsegmentierung für das jeweilige Bild vorliegt. Weiterhin funktioniert die hier entstandene Implementierung ausschließlich für zweidimensionale Bilddaten. Für drei- oder höherdimensionale Daten gibt es derzeit kaum Ansätze, die sich mit der Detektion SURF-ähnlicher Landmarken beschäftigen. Eine Erweiterung der des Ansatzes und der Implementierung in Zukunft wäre wünschenswert.

Mit Blick auf die in Kapitel 4 vorgestellten Ergebnisse der durchgeführten Experimente lässt sich abschließend sagen, dass die im Rahmen dieser Arbeit gute und wie in Abschnitt 4.4 gezeigt, auch praktisch anwendbare Ergebnisse liefert.

Abbildungsverzeichnis

1.1	Beispielhafte Bildregistrierung menschlicher Hände	4
2.1	Abtastfunktion $s(x, y) = \delta(x, y)$	8
2.2	Manuelle Landmarkenbestimmung	10
2.3	Lena-Bild: Original, Symmetrie- und Integralbild	11
2.4	Scale- und Box Space	12
2.5	Rigide und affine Transformationen	13
2.6	Bilineare Interpolation	15
3.1	Approximation von Gaußfiltern zweiter Ordnung	18
3.2	Annäherung der Boxfilter an die Form der Gaußfilter	18
3.3	Effiziente Nutzung eines Boxfilters über Integralbilder	19
3.4	Maximabestimmung in einer Nachbarschaft	20
3.5	Rundes Fenster um einen Merkmalspunkt	22
3.6	Boxfilter erster Ordnung	22
3.7	Umgebung mit Unterfenstern	23
4.1	Die für die Experimente verwendeten Bilddaten	28
4.2	MRT-Bilder des Kopfes	28
4.3	Ergebnisse der Landmarkendetektion im Lena- und MRT-Bild	29
4.4	Landmarken in den Lena-Bildern	30
4.5	Ergebnisse der Landmarkendetektion im Experiment zur Intensitätswertinvarianz	30
4.6	Ergebnisse der Landmarkendetektion im Experiment zum Einfluss von Kompressionsartefakten	31
4.7	Ergebnisse der Landmarkendetektion im Experiment zur Objekterkennung mit SURF-Landmarken	32
4.8	Matching-Ergebnisse	33
4.9	Parametrische Registrierung des Lena-Bildes	35

Tabellenverzeichnis

4.1	Evaluation des Matchingverfahrens	34
-----	---	----

Literaturverzeichnis

- [Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.*, 110(3):346–359.
- [Bredies and Lorenz, 2011] Bredies, K. and Lorenz, D. A. (2011). *Mathematische Bildverarbeitung - Einführung in Grundlagen und moderne Theorie*. Vieweg+Teubner Verlag.
- [Cheng and Cheng, 2005] Cheng, A. H.-D. and Cheng, D. T. (2005). Heritage and early history of the boundary element method. *Engineering Analysis with Boundary Elements*, 29(3):268 – 302.
- [Ehrhardt et al., 2003] Ehrhardt, J., Handels, H., Malina, T., Strathmann, B., Plötz, W., and Pöppel, S. J. (2003). Atlas-based recognition of anatomical structures and landmarks to support the virtual three-dimensional planning of hip operations. In Ellis, R. E. and Peters, T. M., editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*, volume 2878 of *Lecture Notes in Computer Science*, pages 17–24, Montréal, Canada. Springer Berlin Heidelberg, Springer Berlin Heidelberg.
- [Golestaneh and Chandler, 2014] Golestaneh, A. S. and Chandler, D. M. (2014). Algorithm for jpeg artifact reduction via local edge regeneration. *Journal of Electronic Imaging*, 23(1).
- [Gwosdek et al., 2012] Gwosdek, P., Grewenig, S., Bruhn, A., and Weickert, J. (2012). Theoretical Foundations of Gaussian Convolution by Extended Box Filtering. In *Proceedings of the Third International Conference on Scale Space and Variational Methods in Computer Vision*, SSVM'11, pages 447–458, Berlin, Heidelberg. Springer-Verlag.
- [Hadamard, 1902] Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. pages 49–52.
- [Handels, 2009] Handels, H. (2009). *Medizinische Bildverarbeitung*. Reihe Medizinische Informatik. Vieweg+Teubner Verlag, Wiesbaden.
- [Lindeberg, 1994] Lindeberg, T. (1994). Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, pages 224–270.
- [Lowe, 2004] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60(2):91–110.
- [Mikolajczyk et al., 2005] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. V. (2005). A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72.

- [Mikolajczyk et al., 2007] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. V. (2007). Affine covariant features. <http://www.robots.ox.ac.uk/~vgg/research/affine/>. [Online; Abgerufen am 17.05.2015].
- [Modersitzki, 2004] Modersitzki, J. (2004). *Numerical Methods for Image Registration*. Oxford University Press.
- [Modersitzki, 2009] Modersitzki, J. (2009). *FAIR: Flexible Algorithms for Image Registration*. SIAM, Philadelphia.
- [Rohr, 2001] Rohr, K. (2001). *Landmark-Based Image Analysis: Using Geometric and Intensity Models*. Computational imaging and vision". Kluwer Academic Publ. cop., Dordrecht, Boston, London.
- [Sarrut et al., 2007] Sarrut, D., Delhay, B., Villard, P.-F., Boldea, V., Beuve, M., and Clarysse, P. (2007). A Comparison Framework for Breathing Motion Estimation Methods from 4D Imaging. *EEE Transactions on Medical Imaging*, (-).
- [The Mathworks Inc., 2015] The Mathworks Inc. (2015). Exploring Slices from a 3-Dimensional MRI Data Set. <http://de.mathworks.com/help/images/examples/exploring-slices-from-a-3-dimensional-mri-data-set.html>. [Online; Abgerufen am 17.05.2015].
- [van den Elsen et al., 1993] van den Elsen, P. A., Pol, E. J. D., and Viergever, M. A. (1993). Medical image matching-a review with classification. *Engineering in Medicine and Biology Magazine, IEEE*, 12(1):26-39.
- [Viola and Jones, 2001] Viola, P. A. and Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. In *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), with CD-ROM, 8-14 December 2001, Kauai, HI, USA*, pages 511-518.
- [Werner et al., 2010] Werner, R., Wolf, J.-C., Ehrhardt, J., Schmidt-Richberg, A., and Handels, H. (2010). Automatische Landmarkendetektion und -übertragung zur Evaluation der Registrierung von thorakalen CT-Daten. In Deserno, T. M., Handels, H., Meinzer, H.-P., and Tolxdorff, T., editors, *Bildverarbeitung für die Medizin*, volume 574 of *CEUR Workshop Proceedings*, pages 31-35. CEUR-WS.org.