Diplomarbeit

Bildregistrierung von 3D-Laserscans und Bildern für Wadi Sura

vorlegt dem

Fachbereich Mathematik und Informatik
der Freien Universität Berlin
Takustr. 9, 14195 Berlin

Diplomand:	Christopher Keiner
Anschrift:	Halbauerweg 19
	12249 Berlin
Betreuer:	Prof. Dr. Raúl Rojas
Datum:	2. April 2011

Hiermit versichere ich, dass ich die von mir vorgelegte Arbeit selbstständig verfasst habe, dass ich die verwendeten Quellen, Internet-Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Berlin, den 2. April 2011

(Unterschrift)

_____

Christopher Keiner

Zusammenfassung

Die verfasste Diplomarbeit thematisiert eine Problemstellung, die in Zusammenarbeit mit dem archäologischen Projekt *Wadi Sura* entstanden ist. Die namensgebende Region *Wadi Sura* liegt in Ägypten und ist durch zahlreiche Fundstellen von Felsmalereien bekannt. Ein Teilziel des Projekts ist, eine bedeutende Entdeckung einer Felswand mit über 8000 Einzelfiguren zu dokumentieren. Die erhobenen Daten zur Dokumentation umfassen unter anderem 3D-Punktwolken einer Laserscan-Vermessung. Bei der Vermessung wurden zeitgleich sphärische Panoramen aus Position des Laserscanners erstellt. Zu einem späteren Zeitpunkt sind hochauflösende Bilder der Felswand aufgenommen worden. Die Position der Kamera der Aufnahmen ist unbekannt.

Die Problemstellung besteht aus der Registrierung der Panoramen und der hochauflösenden Bilder mit den Punktwolken sowie der Weiterverarbeitung der registrierten Daten zu orthographischen Darstellungen.

Die Lösungsansätze zur Registrierung der Bilddaten basieren auf Methoden der Bildregistrierung. Die Bildregistrierung setzt als Eingabe ein Paar aus Bildern voraus. Ein geeignetes Abbild der 3D-Daten muss für die Registrierung der Panoramen sowie der Registrierung der hochauflösenden Bildern erstellt werden. Eine merkmalsbasierte Bildregistrierung ist verwendet worden, um korrespondierende Punkte zwischen den Bildern zu finden. Die korrespondierenden Punkte liefern die Grundlage zur weiteren Berechnung.

Die Registrierung der Panoramen kann durch die Bestimmung einer Translation zwischen einer sphärischen Projektion der 3D-Daten und dem Panorama berechnet werden. Die Registrierung der hochauflösenden Bilder wird durch die Bestimmung der Projektionsmatrix aus der Erweiterung der 2D-2D-Korrespondenzen zu 3D-2D-Korrespondenzen berechnet.

Die Registrierung der Bilddaten liefert texturierte Punktwolken, die durch Methoden der Computergrafik zu orthographischen Darstellungen verarbeitet werden. Die orthographische Darstellung ermöglicht eine entzerrte Darstellung von ausgewählten Figuren oder Bereichen. Die entzerrte Darstellung ist die Grundlage zur weiteren Auswertung für Projektbeteiligte. In der entzerrten Darstellung ist es beispielsweise möglich Figuren zu vergleichen oder zu vermessen.

Danksagung

An dieser Stelle möchte ich allen Personen danken, die mich bei der Ausarbeitung dieser Arbeit unterstützt und einen Beitrag dazu geleistet haben. Die fachübergreifende Arbeit ist ohne diese Unterstützung nicht erfolgreich möglich gewesen.
Ich danke meinen Eltern, die mich während des Studiums unterstützt und einen zügigen Abschluss ermöglicht haben. Die kritische Betrachtung der Arbeit war eine große Hilfe.
Prof. Dr. Raúl Rojas gilt Dank für die Vergabe des interessanten Themas und für den großen Freiraum zur Gestaltung der Arbeit. Er eröffnete die Verbindungen und Möglichkeiten, so dass die Zusammenarbeit mit dem Projekt zustande gekommen ist.
Für die gute Betreuung bedanke ich mich bei Dr. Frank Darius. Als Betreuer und Mitglied des Projekts ermöglichte er den Einstieg und den Kontakt zu den weiteren Mitgliedern. Bei der Gestaltung der Arbeit stand er mit Rat und Tat zur Seite.
Eine große Hilfe waren ebenfalls die zahlreichen Treffen mit Herrn Erik Büttner, der die 3D-Laserscan-Vermessung durchgeführt hat. Sein Wissen über die Vermessung und das Projekt waren überaus hilfreich. Des Weiteren danke ich Ihm für die Koordination zwischen den Bedürfnissen des Projekts und meiner Arbeit.

Inhaltsverzeichnis

Abbildungsverzeichnis **6**

1 Einleitung **7**
1.1 Problemstellung 9
1.2 Aufbau der Arbeit 11

2 Grundlagen **12**
2.1 Daten-Repräsentation 12
2.1.1 Koordinatensysteme 12
2.1.2 Projektion 14
2.2 Verwendete Methoden der Datenerhebung 16
2.2.1 3D-Laserscanning 16
2.2.2 Sphärische Panorama-Bilder 18
2.2.3 Hochauflösende Bilder 18
2.3 Merkmalsbasierte Bildregistrierung 20
2.3.1 SURF 21
2.3.2 Paarung der Merkmale 27
2.3.3 Robuste Methoden 28
2.4 Pixelbasierte Bildregistrierung 29
2.4.1 Metriken 30
2.5 Projektionsmatrix-Bestimmung 33
2.5.1 Lineare Lösung 33
2.5.2 Nichtlineare Optimierung 34
2.5.3 Daten-Normalisierung 35

3 Lösungsansätze **37**
3.1 Erstellung der orthographischen Darstellung 37
3.1.1 Bestimmung der Projektionsrichtung 38
3.1.2 Triangulierung der Punktwolke 38
3.1.3 Verarbeitung durch die Rendering-Pipeline 39
3.2 Bildregistrierung für Texturkoordinatenberechnung 39
3.2.1 Registrierung der Panoramen 40
3.2.2 Registrierung der hochauflösenden Bilder 44
3.3 Lineare Korrektur 47

4 Systementwurf **49**
4.1 Anforderungen 49
4.2 Statischer Entwurf 50
4.3 Dynamischer Entwurf 52
4.3.1 Benutzer-Schnittstelle 54

5 Experimentelle Auswertung und Diskussion **58**
5.1 Auswertung der Bildregistrierung 58
5.1.1 Registrierung der sphärischen Panoramen 59
5.1.2 Registrierung der hochauflösenden Bilder 62
5.2 Analyse der orthographischen Darstellungen 63

6 Abschlussbetrachtung **66**
6.1 Anwendbarkeit und Grenzen 66
6.2 Ausblicke und Erweiterungsmöglichkeiten 67

A Verwendete Software **68**

Literaturverzeichnis **69**

Abbildungsverzeichnis

1.1 Lage der Region Wadi Sura 8
1.2 Untersuchungen in Wadi Sura 10

2.1 Sphärische Koordinaten 13
2.2 Projektion auf die Bildebene 14
2.3 Orthographische Projektion 15
2.4 Unterschied zwischen Intensitäts- und Farbwerten eines Laserscans 17
2.5 Equipment der Datenerhebung 18
2.6 Einzelbilder aus Position des Scanners 19
2.7 Beispiel eines hochauflösendes Bild 19
2.9 Blob-Strukturen 22
2.8 Abfragen im Integralbild 22
2.10 Approximation der LoG-Kerne 23
2.11 Modell des Skalenraums 24
2.12 Vergrößerung der Filterkerne 25
2.13 Filterkerne des Haar-Wavelet 26
2.14 Orientierung der Merkmale 26
2.15 Merkmalsbeschreibung für verschiedene Fälle 27
2.16 Problemstellung des Template-Matchings 30
2.17 Metrik - Euklidische Distanz 31
2.18 Metrik - Kreuzkorrelation 32
2.19 Metrik - Korrelationskoeffizient 33

3.1 Triangulierung der Punktwolke 38
3.2 Algorithmus zum Lückenfüllen 41
3.3 Registrierung eines Panoramas 42
3.4 Abweichung der Drehachse 43
3.5 Registrierung eines hochauflösenden Bilds 45
3.6 Lineare Korrektur der Texturkoordinaten 48

4.1 Kriterium zur Auswahl des aktiven Standpunkts 52
4.2 Arbeitsablauf der Registrierung und Verarbeitung 53
4.3 Darstellungsmöglichkeiten 55
4.4 Selektionsmöglichkeiten 56

5.1 Testdaten: Standpunkt 2 und 3 60

5.2 Testdaten: Hochauflösende Bilder 62
5.3 Schachbrett-Darstellung: Farbwerte 64
5.4 Schachbrett-Darstellung: Intensitätswerte 64
5.5 Schachbrett-Darstellung: Hochauflösenden Fotos 65

Kapitel 1

Einleitung

Die für das Projekt *Wadi Sura* namensgebende Region ist in Ägypten nahe der Grenze zu Libyen und Sudan, im Südwesten des Gilf Kebir Plateaus gelegen. Es existieren in dieser Gegend über 50 Orte, an denen Felsmalereien zu finden sind. Die Felsmalereien stammen von einer dort in der Zeit von 6500 - 4400 v.Chr. lebenden Nomadenkultur.
Die Region wurde im Jahr 1933 durch die Entdeckung der *Höhle der Schwimmer* durch László Almásy bekannt. Der Entdecker Almásy gab zu dieser Zeit ebenfalls der Region den jetzigen Namen *Wadi Sura* (übers. Tal der Bilder). An der entdeckten Höhle ist in Anbetracht der Lage in der Sahara bemerkenswert, dass neben Jägern, Tieren und anderen Figuren schwimmende Menschen wiederzufinden sind.
Im Jahr 2002 wurde etwa 10km westlich der Höhle der Schwimmer eine weitere bedeutende Entdeckung gemacht. Unter einem 18 Meter langen Felsdach fand Massimo Foggini die bislang größte Ansammlung von gemalten und gravierten Figuren in dieser Region. Sowohl die Menge von 8000 Einzelfiguren als auch die Themenvielfalt der Szenen machen diesen Fundort für Archäologen, Prähistoriker und andere Forscher interessant. Der Fundort wird als *Wadi Sura II* oder *Höhle der Monster* bezeichnet. In Abbildung 1.1 sind zur Orientierung und Einordnung die beiden Fundstellen gezeigt.
Das Projekt *Wadi Sura* ist 2009 von der Deutschen Forschungsgemeinschaft gestartet worden, um neue archäologische Untersuchungen an den Fundorten durchzuführen. Zu den Projektbeteiligten zählen unter anderem das Institut für Prähistorische Archäologie der Universität Köln, das Heinrich-Barth-Institut e.V., das Institut für Restaurierungs- und Konservierungswissenschaften der Fachhochschule Köln sowie das Deutsche Archäologische Institut in Kairo.

Das Projekt verfolgt zwei Ziele:

- Ein Dokumentation der Malereien auf der Felswand von Wadi Sura II wird durchgeführt, um den vorgefundenen Zustand festzuhalten. Auf Grund der Vielzahl und Vielfalt der Figuren ist eine systematische Auswertung unerlässlich. Das Ergebnis der Dokumentation wird eine Merkmals-analytische Datenbank sein, in der alle Figuren nach ausgesuchten Merkmalen katalogisiert vorliegen werden.
  Die Dokumentation beschränkt sich nicht nur auf die Erfassung der Felsmalereien, sondern umfasst beispielsweise auch die Materialanalyse. Eine Herausforderung hierbei ist es, die Felswand nicht zu beschädigen. Zu diesem Zweck sind kontaktfreie Verfahren

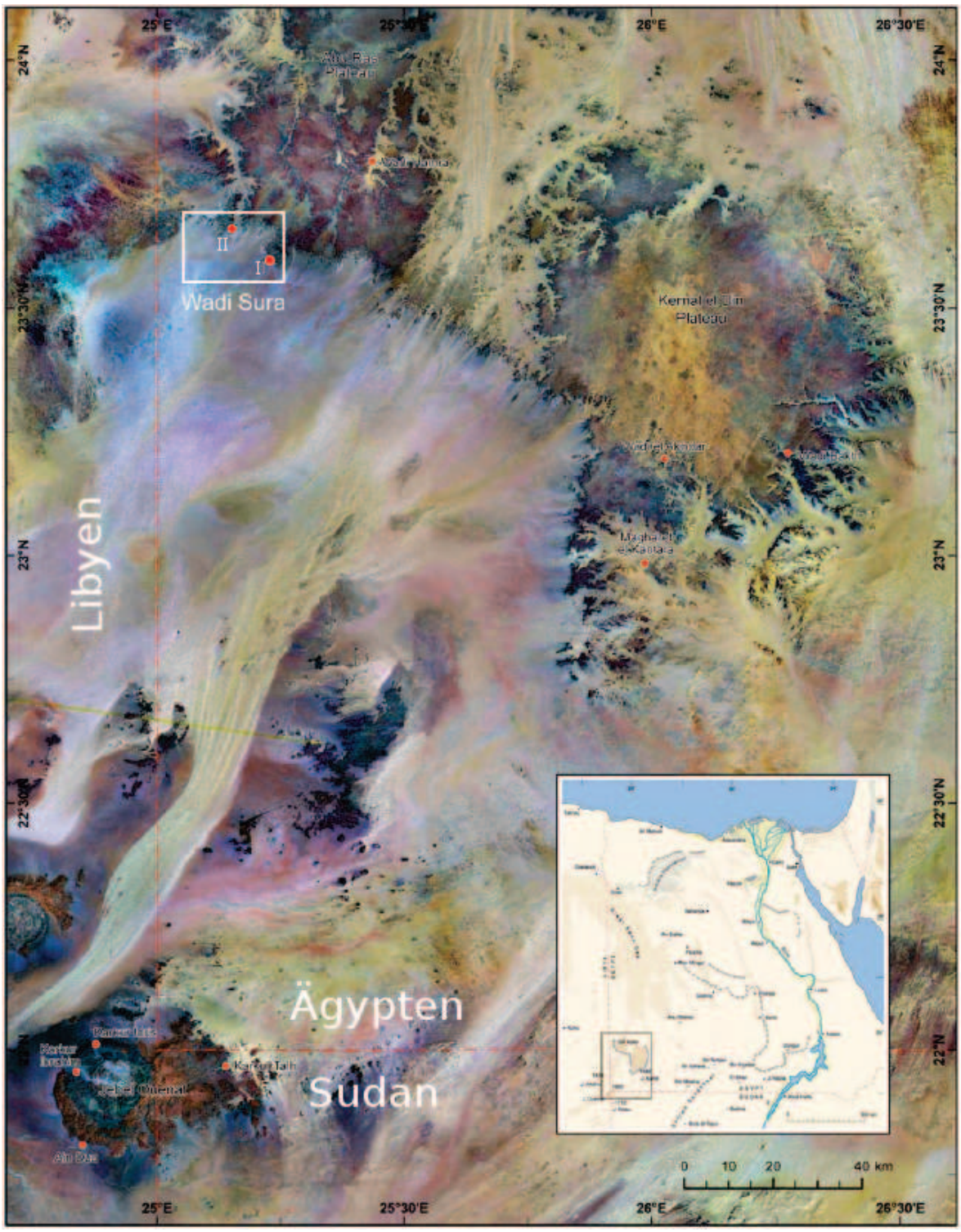

Abbildung 1.1: Auf diesem Satelitenbild lässt sich die Lage von Wadi Sura I und II einordnen. Das Satelitenbild zeigt den Ausschnitt der rechten unteren Karte. Bildquelle: [Kuper 09b]

erprobt worden.

- Archäologische Untersuchungen des Umlandes werden durchgeführt, um die Malereien in einen Kontext zu den dort lebenden Kulturen zu stellen und zu deuten.

Bedingt durch die abgeschiedene Lage und die klimatischen Bedingungen wird die Arbeit in zeitlich begrenzten Expeditionen durchgeführt. Die Expeditionen haben eine Dauer von 4 bis 6 Wochen. Die vollständige Dokumentation der Felswand ist aus diesem Grund vor Ort nicht möglich. Die Datenerhebung steht daher im Vordergrund und die Auswertung der Daten folgt im Anschluss an die Expeditionen. Bisher wurden drei Expeditionen durchgeführt. Die Ergebnisse des Projekts sind in den Field Reports zu den jeweiligen Expeditionen veröffentlicht ([Kuper 09b], [Kuper 09a], [Kuper 10]).

Neben der wissenschaftlichen Arbeit beschäftigt man sich auch mit dem Aspekt, dem öffentlichen Publikum die Erkenntnisse und Ergebnisse zu präsentieren. Ein kurzfristiges Ziel ist dabei die Erstellung eines hochformatigen Buchdruckes. Eine wage, aber interessante Idee ist eine detailgetreue Replik der Felswand, die auf Grund der gesammelten Daten erstellt werden könnte.

1.1 Problemstellung

In Zusammenarbeit mit dem archäologischen Projekt *Wadi Sura* ist die Problemstellung für die vorliegende Arbeit entstanden. Als Grundlage für die Diplomarbeit stehen die erhobenen Daten aus dem Projekt zu Verfügung. Eine 3D-Laserscan-Vermessung ist in der zweiten Expedition von der Felswand durchgeführt worden. Es liegen Punktwolken von 16 verschiedenen Standpunkten vor, die die Geometrie der Felswand flächendeckend und lückenlos dokumentieren. Es wurden gleichzeitig bei der Vermessung mit einer Digitalkamera Fotos aus den Positionen des Laserscanners aufgenommen, die zu sphärischen Panoramen zusammengesetzt worden sind.
Die vorhandene Software zur Verarbeitung der Punktwolke kann diese Bilddaten nur zur Kolorierung der Punktwolke nutzen. Die Auflösung der Punktwolke ist im Vergleich zu den Bilddaten sehr viel geringer. Das Arbeiten mit der kolorierten Punktwolke ist für die archäologischen Anwendungen nicht ausreichend.
Der erste Teil der Problemstellung ist die Registrierung dieser Panoramen mit den 3D-Daten. Die Arbeit mit dem geometrischen 3D-Modell und der hohen Auflösung der vorhandenen Bilddaten soll möglich sein.
Neben diesen Bilddaten stehen hochauflösende Fotos aus der dritten Expedition zu Verfügung. Ein spezielles Scanback-Verfahren ist benutzt worden, um eine wesentlich höhere Auflösung der Aufnahmen zu erreichen. Die Position der Kamera der Aufnahmen ist in diesem Fall unbekannt. Der zweite Teil der Problemstellung ist die Registrierung dieser hochauflösenden Fotos mit der 3D-Punktwolke.
Der dritte Teil der Problemstellung ist die gemeinsame Verarbeitung der 3D-Daten und der registrierten Bilddaten zu orthographischen Darstellungen. In diesen Darstellungen ist es möglich, z.B. die Figuren zu vermessen.

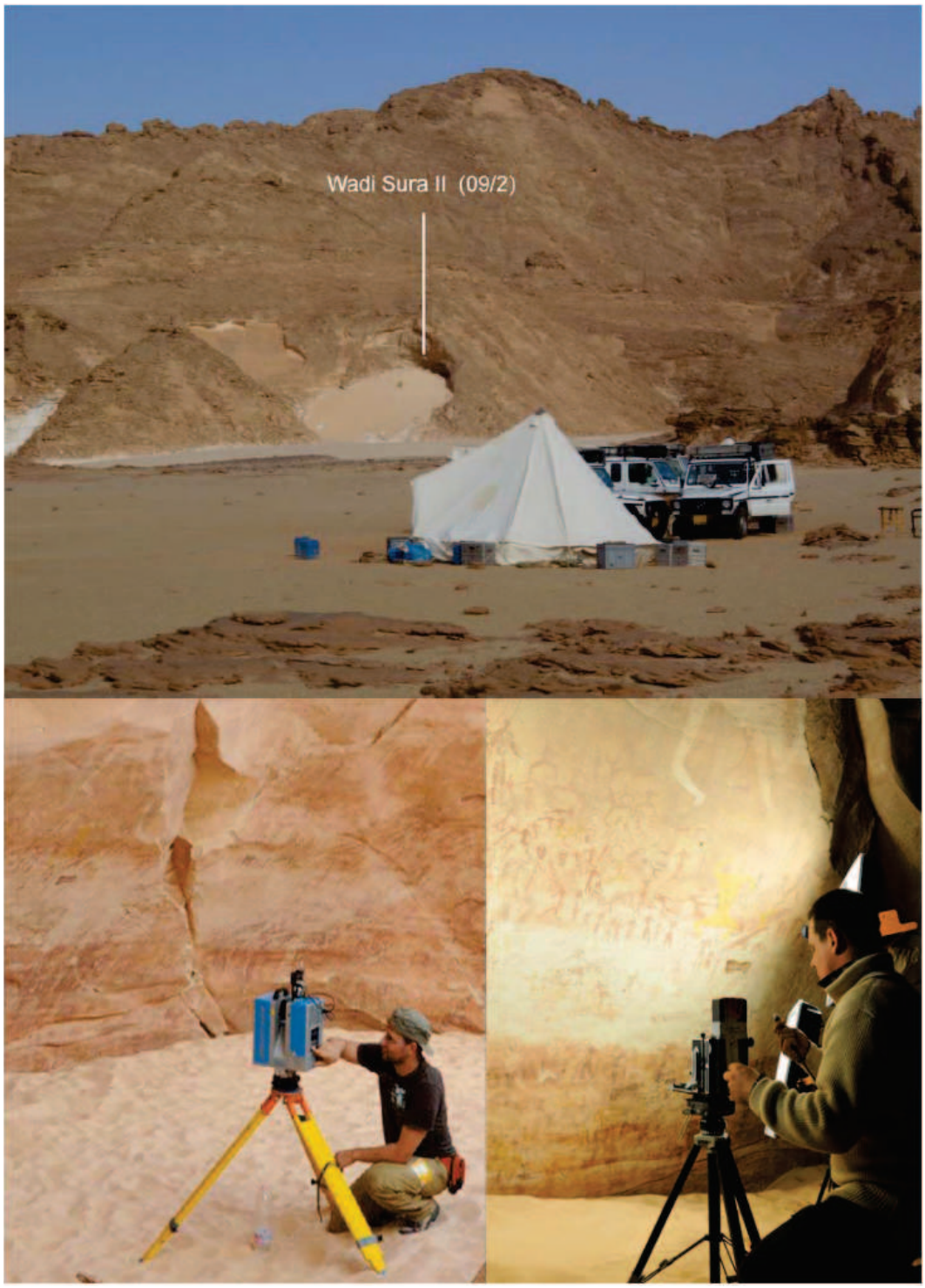

Abbildung 1.2: Diese Bildcollage gibt einen Eindruck über die Umgebung und der Arbeit in der Region Wadi Sura. Auf dem oberen Bild ist die Felswand und das Umland zu sehen. Das linke untere Bild zeigt die Arbeit während der Vermessung der Felswand per Laserscan. Auf dem dritten Bild ist die Dokumentationsarbeit durch die Aufnahme von hochauflösenden Bilder gezeigt. Bildquelle:[Kuper 09a],[Kuper 10]

1.2 Aufbau der Arbeit

Die folgende Arbeit gliedert sich in weitere fünf Kapitel. Im nächsten Kapitel werden die Grundlagen erläutert, die zum weiteren Verständnis notwendig sind. Die mathematischen Grundlagen für die weitere Arbeit werden gelegt. Die verwendeten Methoden der Laserscan-Vermessung und der Aufnahme der hochauflösende Bilder werden vorgestellt, um die Datengrundlage zu präzisieren. Im Anschluss werden merkmalsbasierte und pixelbasierte Methoden zur Bildregistrierung vorgestellt. Die Bestimmung einer Projektionsmatrix bildet den Abschluss des zweiten Kapitels.
Im dritten Kapitel werden die entwickelten Lösungsansätze vorgestellt. Die Erstellung von orthographischen Darstellungen, die Registrierung der Panoramen und die Registrierung der hochauflösenden Bilder sind drei Teilprobleme. Das Kapitel gliedert sich in die Lösungsansätze dieser Teilprobleme.
Die Lösungsansätze sind in einem Software-System zur Benutzung realisiert worden. Das vierte Kapitel stellt den Systementwurf vor. Der statische Entwurf beschreibt die Datenorganisation. Der dynamische Entwurf erläutert die Arbeitsweise der Lösungsansätze in Zusammenhang mit der Datenorganisation.
Das fünfte Kapitel zeigt, inwieweit die Lösungsansätze die Teilprobleme gelöst haben. Dies erfolgt durch eine experimentelle Auswertung der Registrierung sowie der Analyse der orthographischen Darstellungen. Die Ergebnisse der Auswertung werden diskutiert.
Im letzten Kapitel wird die Anwendbarkeit der Lösungsansätze für allgemeinere Anwendungen diskutiert. Kurze Ausblicke für mögliche Erweiterungen schließen die Arbeit ab.

Kapitel 2

Grundlagen

In diesem Kapitel werden die wichtigsten Grundlagen vorgestellt, auf die in späteren Kapiteln Bezug genommen wird. Im ersten Abschnitt werden die mathematischen Modelle als Grundlage zur weiteren Verarbeitung vorgestellt. Nach der genaueren Vorstellung der verwendeten Methoden zur Datenerhebung werden die Verfahren zur Bildregistrierung erläutert. Es wird grob zwischen merkmalsbasierten und pixelbasierten Verfahren unterschieden. Im Anschluss wird auf die Bestimmung einer Projektionsmatrix eingegangen.

2.1 Daten-Repräsentation

In den folgenden Unterabschnitten werden die mathematischen Modelle vorgestellt, mit denen 3D- bzw. 2D-Daten repräsentiert werden können. Es wird erläutert, wie die Beziehung zwischen 3D- und 2D-Daten durch verschiedene Projektionen hergestellt wird.

2.1.1 Koordinatensysteme

Die einfachste Art Punkte im Raum darzustellen ist die Repräsentation in kartesischen Koordinaten. Eine Erweiterung stellen die homogenen Koordinaten dar. Ein Punkt (x, y, z) wird in homogenen Koordinaten durch die Äquivalenzklasse $(x, y, z, 1)$ repräsentiert, der alle Punkte (wx, wy, wz, w) mit $w \neq 0$ angehören. Ein Vorteil dieser Darstellung ist die Möglichkeit, Transformationen, inklusive der Translation, durch eine Matrixmultiplikation auszudrücken. Die Erweiterung existiert analog für die euklidische Ebene.
Neben den kartesischen Koordinaten sind sphärische Koordinaten ebenfalls wichtig. Ein Punkt in sphärischen Koordinaten wird durch den Drehwinkel φ um die z-Achse, den Winkel zur Drehachse θ und den Abstand r zum Ursprung ausgedrückt (siehe Abb.2.1). Eine

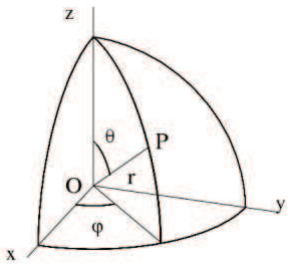

Abbildung 2.1: Ein Punkt P lässt sich in sphärischen Koordinaten durch den Abstand r zum Ursprung, den Winkel θ zur z-Achse und den Winkel φ ausdrücken. Der Winkel φ ist durch den Winkel zwischen der Projektion des Ortsvektors zu P auf die xy-Ebene und der x-Achse definiert. Bildquelle: [Nolting 99]

Umrechnung ist durch die folgenden Beziehungen möglich:

$$r = \sqrt{x^2 + y^2 + z^2} \tag{2.1}$$
$$\varphi = atan2(y, x) \tag{2.2}$$
$$\theta = arccos(\frac{z}{r}) \tag{2.3}$$
$$x = r \sin\theta \cos\varphi \tag{2.4}$$
$$y = r \sin\theta \sin\varphi \tag{2.5}$$
$$z = r \cos\theta \tag{2.6}$$
$$\tag{2.7}$$

Die Sonderfälle $\theta = 0$ oder $r = 0$ sind bei der Umrechnung zu beachten.
Durch diese Darstellungen ist ein Punkt in einem Koordinatensystem eindeutig festgelegt. In der Computergrafik verwendet man oft unterschiedliche Koordinatensysteme, die nun kurz vorgestellt werden. Die Laserscan-Daten dienen als Beispiel zur Verdeutlichung.

Objektkoordinaten Objektkoordinaten sind lokale Koordinaten, die verwendet werden, um ein einzelnes Objekt zu repräsentieren. Eine einzelne Aufnahme des Scanners an einem Standpunkt entspricht den unreferenzierten Objektkoordinaten.

Weltkoordinaten Eine modellierte Szene besteht meist aus einer Menge von Objekten. Die Objektkoordinaten werden in ein gemeinsames Weltkoordinatensystem transformiert. Durch die Transformation wird die Lage der Objekte zueinander festgelegt. Im Blick auf die Laserscan-Daten entspricht dieser Schritt der sog. *Georeferenzierung*, bei der alle Standpunkte in einem gemeinsamen Koordinatensystem vereint werden.

Kamerakoordinaten Das Kamerakoordinatensystem ist eine Transformation des Weltkoordinatensystems in der Weise, dass das Kamerazentrum im Ursprung liegt. Dieses Koordinatensystem kann durch die Translation des Ursprungs eines Standpunkts an die Position des Scanners erreicht werden.

Bildkoordinaten Bilder werden als 2-dimensionale Matrix beschrieben, deren Elemente die Farb- oder Grauwerte enthalten. Als Bildkoordinaten (u, v) werden die Indizes für die Spalte und Reihe benutzt.

2.1.2 Projektion

Die Beziehung zwischen 3D-Punkten $(x, y, z, 1)$ im Raum und 2D-Bildkoordinaten wird durch eine Projektion hergestellt. Die allgemeine Projektion ist durch die Transformation mit einer 3×4-Matrix beschrieben ($\cong$ - projektive Gleichheit).

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \cong \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \tag{2.8}$$

Eine geometrische Interpretation der Projektionsmatrix erfolgt über das Lochkamera-Modell. Das Modell besteht aus einer Bildfläche und einer Einfallsöffnung, die infinitesimal klein ist. Sie bildet das Projektionszentrum C. Alle von Objekten ausgehenden Lichtstrahlen passieren im Modell das Projektionszentrum und fallen anschließend auf die Bildfläche. Die Seiten des projizierten Bilds sind dabei vertauscht. Die Bildfläche besitzt den Abstand f zum Projektionszentrum.
Die Projektion eines Punktes ist in Abb. 2.2 dargestellt. In dieser Abbildung liegt die virtuelle Bildebene vor dem Projektionszentrum, so dass das entstehende Bild aufrecht projiziert wird. Die Kamera schaut bei der Projektion in Richtung der z-Achse. Mit Hilfe des Strahlensatzes ergeben sich die Bildkoordinaten u und v:

$$u = \frac{f \cdot x}{z} \text{ und } v = \frac{f \cdot y}{z} \tag{2.9}$$

In der Praxis ist die Bildebene durch die Unterteilung in Pixel diskretisiert. Die Werte von u und v müssen daher gerundet oder die Farbwerte interpoliert werden.

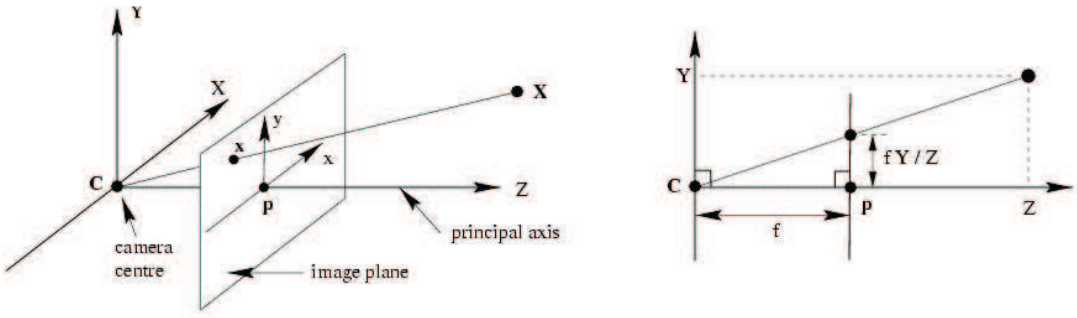

Abbildung 2.2: (links) Die Kamera schaut in Richtung der z-Achse. Die Bildebene liegt vor dem Projektionszentrum C und bildet nach dem Lochkamera-Modell den Punkt x ab. (rechts) Sicht auf die yz-Ebene der gleichen Projektionsskizze. Die Anwendung des Strahlensatzes zur Berechnung der Bildkoordinate v wird ersichtlich. Bildquelle: [Hartley 04]

Die Multiplikation von Punkten in homogenen Koordinaten mit der Projektionsmatrix ist zum obigen Modell äquivalent. Die Projektionsmatrix lässt sich wie folgt in eine Matrix K mit *inneren* Kamera-Parametern und einer Matrix $[R|t]$ mit den *äußeren* Parametern

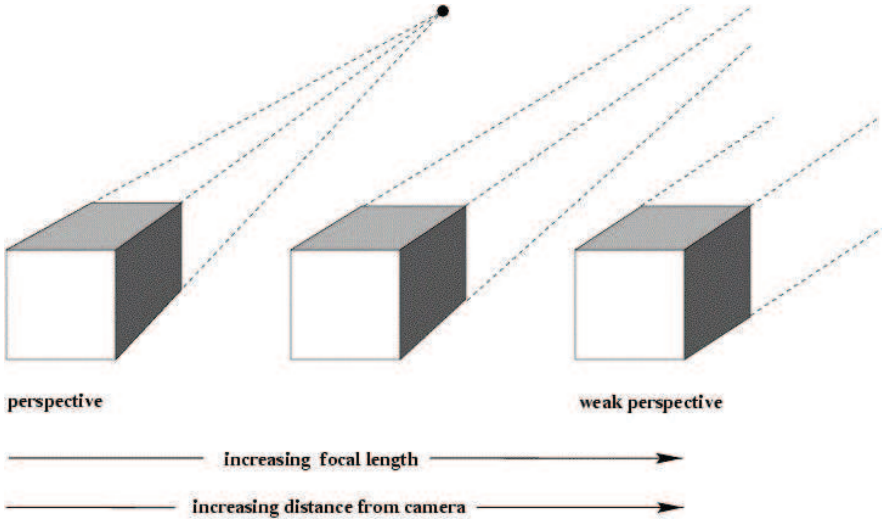

Abbildung 2.3: (links) Eine projektive Abbildung eines Würfels: Die parallelen Kanten laufen bei dieser Projektion in einem Fluchtpunkt zusammen. (rechts) Eine orthographische Projektion des Würfels. Bildquelle: [Hartley 04]

zerlegen:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} = K[R|t] = \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \quad (2.10)$$

Die Zerlegung gibt eine Erklärung für die Äquivalenz. Die Parameter f_x, f_y geben die Brennweite, c_x,c_y eine Translation in der Bildebene und s eine Scherung eines Pixels an. Der Parameter s ist in üblichen Kameramodellen Null. Die äußeren Parameter r_{ij} und t_i werden durch die Zerlegung nur noch von der Orientierung und Position der Kamera bestimmt. Die Parameter f_x und f_y lassen sich im Lochkamera-Modell direkt wiederfinden. Die Parameter c_x und c_y entsprechen einer Verschiebung der Bildebene. Die Normierung der Bildkoordinate auf $(u, v, 1)$ entspricht der Gleichung 2.9.

Im vorigen Teil des Abschnittes ist die allgemeine Projektion erläutert worden. Die *orthographische Projektion* bildet einen Spezialfall. Er entsteht, wenn das Projektionszentrum der Kamera ins Unendliche gelegt wird. Aus Sicht der Mathematik ist die Anwendung der Projektionsmatrix keine projektive Transformation. Die Transformation gehört der affinen Untermenge an.
Bei affinen Transformationen bleiben wichtige Eigenschaften nach der Transformation erhalten. Eine Eigenschaft ist die Erhaltung des Verhältnisses von Abständen. Eine Markierung in der Mitte einer Strecke teilt die transformierte Strecke ebenfalls in der Mitte. Parallele Linien bleiben nach der affinen Transformation ebenfalls parallel. Der typische projektive Effekt entfällt dadurch, dass Parallelen nicht in Fluchtpunkten zusammenlaufen. Zum Vergleich ist in Abb. 2.3 der Unterschied zwischen den Bildern der Projektionen dargestellt.

2.2 Verwendete Methoden der Datenerhebung

Die Grundlage der Arbeit sind die erhobenen Daten aus dem Projekt Wadi Sura. Diese sind die 3D-Laserscan-Daten, die zugehörigen Panoramen und die hochauflösenden Fotos. In diesem Abschnitt werden die Methoden zur Erhebung vorgestellt. Die spezifischen Eigenschaften der einzelnen Daten werden in diesem Zusammenhang erläutert.

2.2.1 3D-Laserscanning

Die 3D-Daten der Felswand wurden für das Projekt Wadi Sura mit einem Laserscanner vom Typ Z+F Imager 5006 aufgenommen. Beim Scanvorgang erzeugt ein rotierender Spiegel eine gleichmäßige Ablenkung des Laserstrahls, der in vertikaler Richtung die Oberfläche abtastet. Aus dem reflektierten Anteil des Laserimpulses lässt sich nun die Entfernung bestimmen. Eine weitere Rotation um die Drehachse des Stativs ergibt schließlich einen kompletten Scan der Umgebung.

Neben der Entfernung ist die reflektierte Intensität von Interesse. Eine Menge von Intensitätswerten repräsentiert das Intensitätsbild der abgetasteten Oberfläche.Ein Laserscanner ist im Unterschied zu einer Fotokamera ein aktives Aufnahmesystem. Eine Fotokamera sammelt die reklektierten Lichtstrahlen, die von verschiedenen Lichtquellen ausgesandt worden sind. Ein Laserscanner sendet aktiv einen Laserimpuls aus und funktioniert dadurch auch in dunklen Räumen.

Die reflektierte Intensität einer Oberfläche, die sowohl der Laserscanner als auch die Fotokamera aufnehmen, ist eine Materialeigenschaft. Die Eigenschaft, wie ein einfallender Lichtstrahl reflektiert wird, wird durch die Bidirektionale Reflektanzverteilungsfunktion (BRDF) beschrieben. Diese Funktion hängt unter anderem vom Einfallswinkel des Lichtstrahls und vom Ausfallsstrahl zum Betrachter ab. Die BRDF ist weiterhin von der Wellenlänge abhängig. Diese Tatsache kann nicht vernachlässigt werden, da der Laser nur ein enges Spektrum emittiert.

Der Vergleich zwischen Intensitätsbild und Grauwertbild einer Kamera ist in Abb. 2.4 zu sehen. Der Hauptunterschied ist die fehlende globale Beleuchtung im Intensitätsbild. Es existieren z.B. keine Schatten. Im Allgemeinen existiert kein linearer Zusammenhang zwischen Intensitätsbild und Grauwertbild. Diese Unterschiede werden im weiteren Verlauf wichtig werden.

Die fehlende Farbinformation eines Laserscans lässt sich mit einer zusätzlichen Kamera erhalten. Zu diesem Zweck wurde eine Z+F M-Cam verwendet, die im Zusammenspiel mit dem Scanner automatisch eine kolorierte Punktwolke liefert. Das komplette Messsystem ist in Abbildung 2.5 dargestellt.

Das Vermessen einer Umgebung erfolgt in der Regel aus verschiedenen Standpunkten. Dieses Vorgehen garantiert, dass keine Daten durch Verdeckungen fehlen. Durch die Verwendung von eindeutig gekennzeichneten Markierungen (sog. Targets) lassen sich die einzelnen Standpunkte automatisch in ein gemeinsames Koordinatensystem überführen.

Die 3D-Daten liegen im Raw-Format *PTX* vor. Neben den Informationen zu den erfassten Punkten ist eine Transformation von Kamerakoordinaten in das festgelegte Referenzkoordinatensystem vorhanden. Die verwendete Auflösung pro Standpunkt beträgt `10000 x 5060` Punkte.

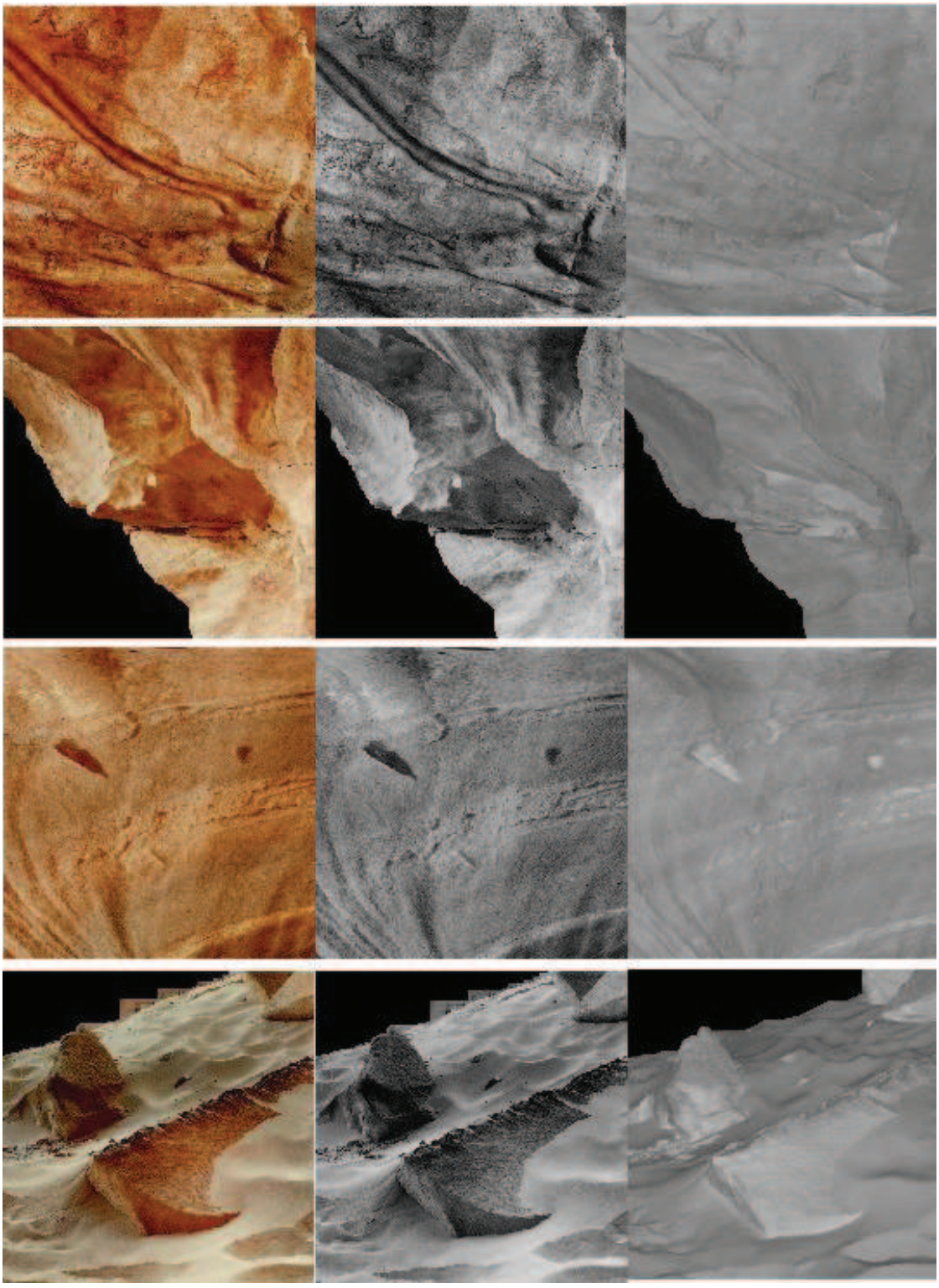

Abbildung 2.4: In dieser Abbildung sind die Unterschiede zwischen den Farbwerten (linke Spalte), den Grauwerten der Farbinformation (mittlere Spalte) und den aufgenommenen Intensitäten des Laserscanners dargestellt(rechte Spalte).

Abbildung 2.5: (Links) Ein Laserscanner von Typ Z+F 5006 mit aufgesetzter M-Cam zur Kolorierung der Punkte; (Mitte) Ein Auto-Target, das die automatische Registrierung von 3D-Punktwolken ermöglicht. (Rechts) Eine Kamera mit digitalem Rückteil, mit dem sehr hochauflösende Bilder im Scanback-Verfahren erstellt werden können.

2.2.2 Sphärische Panorama-Bilder

Im letzten Unterabschnitt ist bereits die Kolorierung der Punktwolke durch eine externe Kamera erwähnt worden. Zusätzlich sind mit einer digitalen Kamera des Typs Nikon D80 und einem Nodalpunktadapter für jeden Standpunkt systematisch 128 Bilder gemacht worden. Die Verwendung des Nodalpunktadapters stellt sicher, dass um das optische Zentrum der Kamera rotiert wird und so keine Parallaxe zwischen Einzelfotos existiert. Eine Sequenz der Einzelbilder ist in Abbildung 2.6 zu sehen.
Die Sequenz von Bildern ist mit Hilfe einer Stitching-Software zu einem sphärischen Panorama zusammengesetzt worden. Der Vorteil des Nodalpunktadapters zeigt sich in diesem Arbeitsschritt. Um die relative Position von aufeinander folgenden Bildern zu berechnen, sind ausschließlich die Parameter der Rotation zu berechnen. Die relative Orientierung wird in diesem Schritt für jeweils zwei benachbarte Bilder berechnet. Die internen Parameter der Kamera sind in allen Bildern unverändert. Die Parameter können mittels Selbstkalibrierung [Hartley 04] bei der Berechnung des Panoramas bestimmt werden. Nachdem eine erste Lösung der internen und externen Kameraparameter gefunden worden ist, folgt der Bündelausgleich [Triggs 00]. In diesem Schritt wird eine Optimierung durchgeführt, die alle Parameter und Datenpunkte zugleich betrachtet.
Die erstellten Panoramen haben eine Auflösung von 25000 x 12500 Punkten und sind für diese Arbeit bereit gestellt worden.

2.2.3 Hochauflösende Bilder

Unabhängig von den zuletzt beschriebenen Aufnahmen liegen besonders hochauflösende Fotos vor, die mit einem speziellen Scanback-Verfahren aufgenommen worden sind. Als optisches System ist eine Linhof Master Technika verwendet worden, an dem ein digitales Rückteil des Typs David2 der Firma ANAGRAM zur Aufnahme verwendet wurde. Dieses System ist in

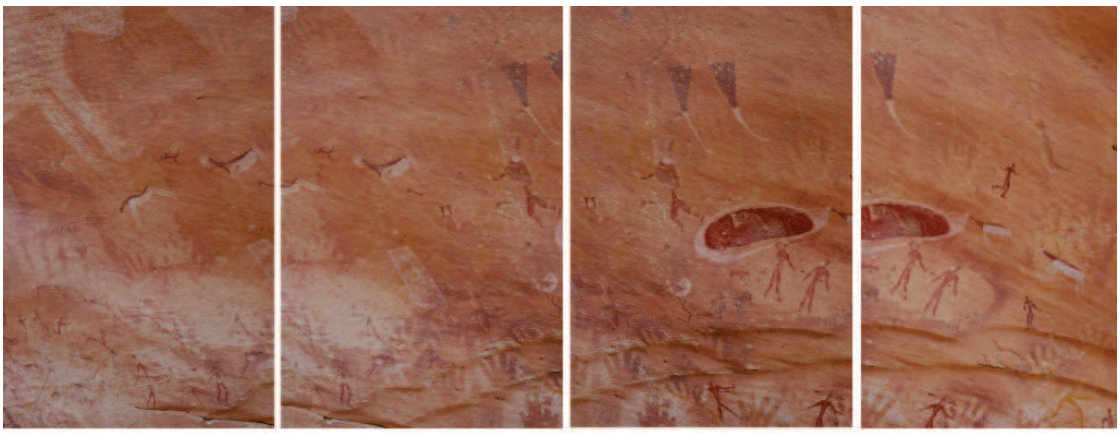

Abbildung 2.6: Sequenz von Einzelbildern, die mit einer digitalen Kamera und einem Nodalpunktadapter erstellt worden sind. Es existiert jeweils ein Überlappungsbereich zwischen den Bildern. Anhand der Überlappung können die Bilder mit einer Stitching-Software zu einem Panorama zusammengesetzt werden.

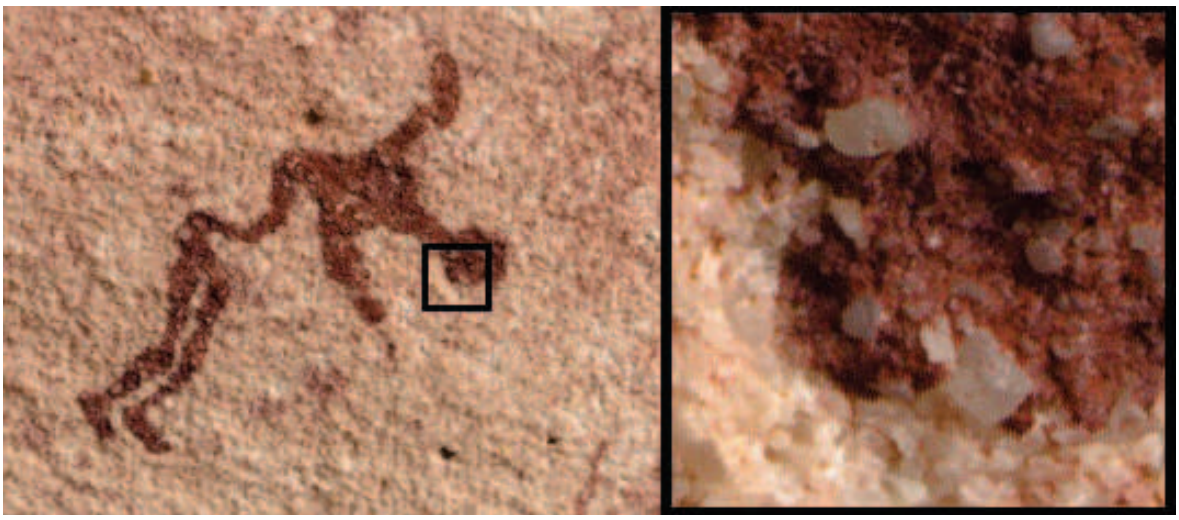

Abbildung 2.7: Das Bild zeigt die Verkleinerung eines hochauflösenden Detail-Bildes. Auf der rechten Seite ist ein Ausschnitt in voller Auflösung dargestellt.

Abb. 2.5 zu sehen. Die Felswand wurde durch 100 Fotos dokumentiert, die eine maximale Auflösung von `25000 x 16000` Punkte haben. Zusätzlich sind von besonders interessanten Teilen der Felswand Detailaufnahmen aus kürzerer Entfernung erstellt worden. Ein Eindruck über die Qualität dieser Fotos lässt sich durch Abb. 2.7 gewinnen.
Informationen über Orientierung oder Kalibrierung sind nicht bekannt.

2.3 Merkmalsbasierte Bildregistrierung

In diesem Abschnitt werden die Grundlagen der merkmalsbasierten Bildregistrierung erläutert. Das Ziel der Bildregistrierung ist, die Ausrichtung zwischen zwei gegebenen Bildern einer Szene zu finden. Die merkmalsbasierte Bildregistrierung erreicht das Ziel durch die Erkennung von korrespondierenden Merkmalen in beiden Bildern. Aus den Merkmalskorrespondenzen lässt sich eine relative Ausrichtung der Bilder zueinander bestimmen. Die Informationen über die Merkmalskorrespondenzen und der Bildregistrierung können für die folgende Auswahl der Anwendungen benutzt werden, die oft aus der Computer Vision oder Bildverarbeitung stammen.

- Bei der 3D-Rekonstruktion einer Szene aus einer Menge von Bildern werden die Merkmalskorrespondenzen zur Berechnung der Fundamentalen Matrix benutzt. Für einen beliebigen Punkt eines Bilds kann aus der Fundamentalen Matrix eine Gerade bestimmt werden, auf der der korrespondierende Punkt auf dem anderen Bild liegen muss. Diese Erkenntnis der Epipolar-Geometrie wird oft ausgenutzt, um 3D-Informationen per Triangulierung aus der Bildmenge zu gewinnen.
- Beim Stitching liegen eine Menge von Bildern vor, die mindestens paarweise eine Überlappung besitzen. Die Merkmale der einzelnen Bilder können zum einen benutzt werden, um eine statistische Überprüfung durchzuführen, ob zwei Bilder benachbart sind. Zum anderen lässt sich die relative Orientierung der Kamera zwischen den Bildern bestimmen. Mit diesem Wissen können die Bilder in einem größeren Bild zusammengefügt werden.
- In der medizinischen Bildverarbeitung ist die Bildregistrierung wichtig. Verschiedene bildgebende Verfahren, wie z.B. eine Computertomographie oder eine Röntgenaufnahme zeigen verschiedene Strukturen des Körpers. Durch eine Bildregistrierung können die Informationen von verschiedenen Quellen zusammen dargestellt werden.
- Eine andere Anwendung ist die Objekterkennung. Es wird dabei eine Datenbank aus bekannten Objekten angelegt, die durch eine Beschreibung der korrespondierenden Punkte abgespeichert werden. Ein Bild eines unbekannten Objekts kann mit einer Paarung in der Datenbank klassifiziert werden.

Die merkmalsbasierte Bildregistrierung besteht aus vier Grundschritten. Der erste Schritt ist die Detektion der Merkmale. Der zweite Schritt ist die Beschreibung der Merkmale durch einen Deskriptor, der im dritten Schritt zur Paarung der Merkmale auf Grund der Beschreibung dient. Mit den erhaltenen Korrespondenz-Paaren kann im vierten Schritte die Berechnung der gesuchten Ausrichtung durchgeführt werden, wobei falsche Paarungen meist nicht ausgeschlossen werden können. In diesem werden *robuste* Algorithmen benötigt, die Ausreißer tolerieren.
Es existiert eine Vielzahl von Merkmalsdetektoren, die verschiedene Eigenschaften besitzen.

Die Merkmalsdetektoren unterscheiden sich stark in der zu suchenden Struktur der Merkmale. Es existieren Detektoren z.B. für Ecken, Blobs, Linien oder Sternpunkte [Bradski 00].
Die Beschreibung der gefundenen Punkte erfolgt meist auf Grundlage der Pixel- oder Helligkeitswerte in den lokalen Nachbarschaft. Ein wichtiges Merkmal der Beschreibung ist die Invarianz gegen bestimmte Typen von Transformationen. Durch diese Invarianzen ist es z.B. möglich, Merkmale von skalierten oder rotierten Bilder zu vergleichen.
Die oben erwähnten Grundschritte der merkmalsbasierten Bildregistrierung werden in den nächsten Unterabschnitten erläutert. In Unterabschnitt 2.3.1 werden die Detektion und Beschreibung der sog. SURF-Merkmale betrachtet. Die Paarung der SURF-Merkmale ist Thema in Unterabschnitt 2.3.2. Eine robuste Methode zur Berechnung von beliebigen Transformationen wird in Unterabschnitt 2.3.3 vorgestellt.

2.3.1 SURF

Die Bezeichnung Speeded-Up Robust Features (SURF) ist von Bay et al. [Bay 06] eingeführt. Die SURF-Merkmale sind eine Abwandlung der bekannten SIFT-Merkmale [Lowe 04]. Die SURF-Merkmale decken die Detektion und Deskription durch eigene entwickelte Verfahren (im Folgenden SURF-Verfahren genannt) ab. Im Gegensatz zu den SIFT-Merkmalen sind die Verfahren durch die Verwendung von Integralbildern beschleunigt worden. SURF verwendet als Eingabe ein Grauwertbild der Intensitäten, das gegebenenfalls aus den Farbwerten erstellt werden muss.

Integralbilder

Ein Integral-Bild [Crow 84] lässt sich aus einem Eingabebild I durch ein Aufsummieren von Intensitäten berechnen. Für ein Pixel an Position (x, y) werden alle Intensitäten $I(x', y')$ summiert, deren Koordinaten $x' \leq x$ und $y' \leq y$ erfüllen. Ein Integral-Bild I_Σ ist formal definiert als:

$$I_\Sigma(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \tag{2.11}$$

Es wird bei der Verwendung der Integral-Bilder notwendig sein, die Intensitätssumme eines beliebigen Achsen-parallelen Rechtecks zu bestimmen. Solche Anfragen lassen sich durch ein zuvor berechnetes Integralbild effektiv berechnen. Seien die Punkte des Rechtecks A,B,C und D (siehe Abb. 2.8), dann lässt sich die Intensitätssumme Σ wie folgt berechnen:

$$\Sigma = I_\Sigma(D) - I_\Sigma(C) - I_\Sigma(B) + I_\Sigma(A) \tag{2.12}$$

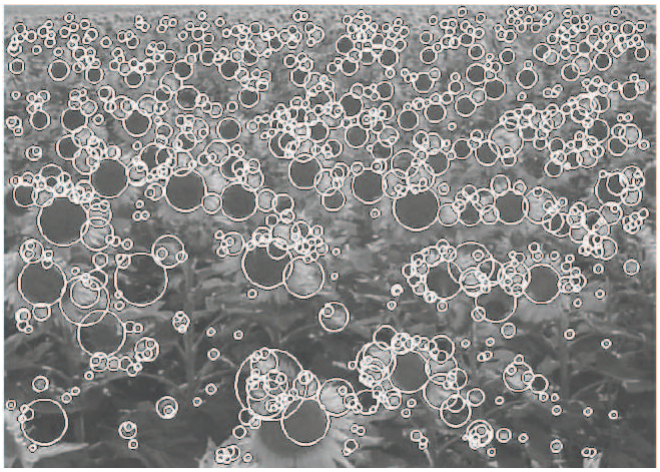

Abbildung 2.9: Der SURF-Detektor entdeckt *Blob*-Strukturen. Diese detektieren Strukturen sind im Bild umkreist. Bildquelle: [Bay 06]

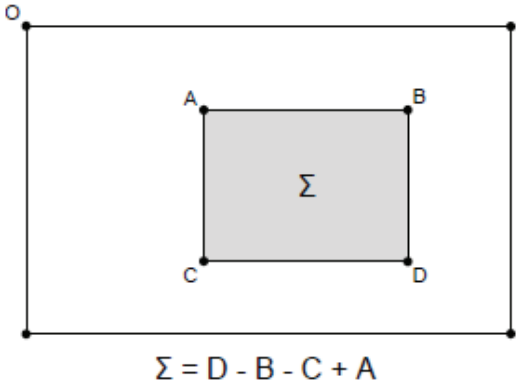

Abbildung 2.8: Die Intensitätssumme eines Achsen-parallelen Rechtecks lässt sich effizient durch Addition und Subtraktion im Integralbild berechnen. Bildquelle:[Bay 06]

Detektor

SURF verwendet zum Detektieren von Merkmalspositionen ein Kriterium basierend auf der Determinante der Hesse-Matrix eines Bilds. Es wird nach Positionen im Bild gesucht, an denen die Determinante ein lokales Maximum erreicht. Dieses Kriterium liefert Maxima an den Positionen, an denen im Bild eine sog. *Blob-Struktur* zu finden ist. Der Begriff wird in Abb. 2.9 deutlich. Die Einträge der Hesse-Matrix ergeben sich aus der Faltung des Bilds mit den partiellen Ableitungen zweiter Ordnung $\frac{\partial^2 g(\sigma)}{\partial x^2}$, $\frac{\partial^2 g(\sigma)}{\partial x \partial y}$, $\frac{\partial^2 g(\sigma)}{\partial y^2}$ einer Gauss-Funktion $g(\sigma)$ mit Varianz σ. Die Funktionen sind auch als *Laplacian of Gaussians (LoG)* bekannt. Der Parameter σ wird später im Skalenraum aufgegriffen.

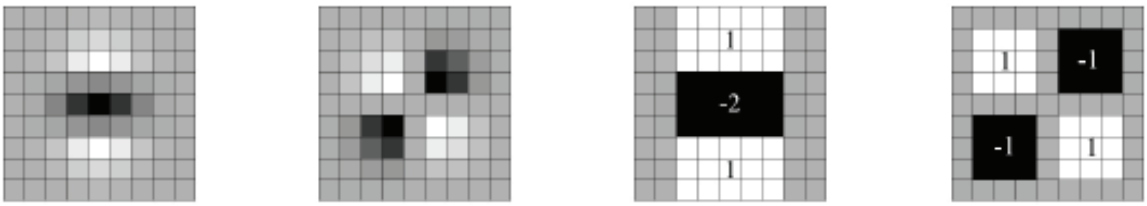

Abbildung 2.10: Auf der linken Seite sind die beiden diskretisierten Filter-Kerne L_{yy} und L_{xy} im Vergleich zu den approximierten Varianten dargestellt. Bildquelle: [Bay 06]

Für einen gegebenen Punkt $\hat{x}$ im Eingabebild I ist die Hesse-Matrix definiert durch:

$$H(\hat{x},\sigma) = \begin{bmatrix} L_{xx}(\hat{x},\sigma) & L_{xy}(\hat{x},\sigma) \\ L_{xy}(\hat{x},\sigma) & L_{yy}(\hat{x},\sigma) \end{bmatrix} \text{ mit (Konvolution als } * \text{)} \tag{2.13}$$

$$L_{xx}(\hat{x},\sigma) = (\frac{\partial^2 g(\sigma)}{\partial x^2} * I)(\hat{x}) \tag{2.14}$$

$$L_{xy}(\hat{x},\sigma) = \tag{2.15}$$

$$L_{yx}(\hat{x},\sigma) = (\frac{\partial^2 g(\sigma)}{\partial x \partial y} * I)(\hat{x}) \tag{2.16}$$

$$L_{yy}(\hat{x},\sigma) = (\frac{\partial^2 g(\sigma)}{\partial y^2} * I)(\hat{x}) \tag{2.17}$$

Die Einträge der Hesse-Matrix könnten durch eine Faltung des Bilds mit diskretisierten LoG-Kernen berechnet werden. Dieses Vorgehen benutzt z.B. SIFT. Die diskretisierten Faltungskerne sind in Abb. 2.10 auf der linken Seite zu sehen. Das SURF-Verfahren benutzt hingegen die Faltungskerne auf der rechten Seite. Diese Kerne approximieren die Kerne mit Hilfe von Mittelwert-Filtern.
An dieser Stelle lässt sich zum ersten Mal das Integralbild verwenden. Die Berechnung der approximierten Werten L_{xx}, L_{xy} und L_{yy} der Hesse-Matrix an der Stelle $\hat{x}$ beginnt mit der Abfrage der drei/vier Regionen des Kerns (siehe Abb. 2.10) im Integralbild. Die Intensitätssummen der Regionen werden mit den eingezeichneten Gewichten multipliziert. Auf diese Weise lässt sich die Faltung für eine Position in konstantem Zeitaufwand durchführen. Die Determinante der Hesse-Matrix H_{approx} mit den approximierten Werten L_{xx}, L_{xy} und L_{yy} lässt sich als

$$det\ H_{approx} = L_{xx}L_{yy} - (wL_{xy})^2 \tag{2.18}$$

schreiben. Das eingeführte Gewicht w gleicht einen Unterschied zwischen Approximation des Kern und dem ursprünglichen Kern aus. Dieser Wert wird auf den konstanten Wert $w = 0.9$ festgelegt. Die Berechnung der Determinanten an allen Pixelkoordinaten $\hat{x}$ liefert ein Ergebnisbild, in dem die lokalen Minima und Maxima gefunden werden können. Die Berechnung erfolgt in Abhängigkeit des Parameters σ, der die Größe des Filterkerns bestimmt. Im nächsten Abschnitt wird vorgestellt, wie durch Variation dieses Parameters der Skalenraum erzeugt wird.

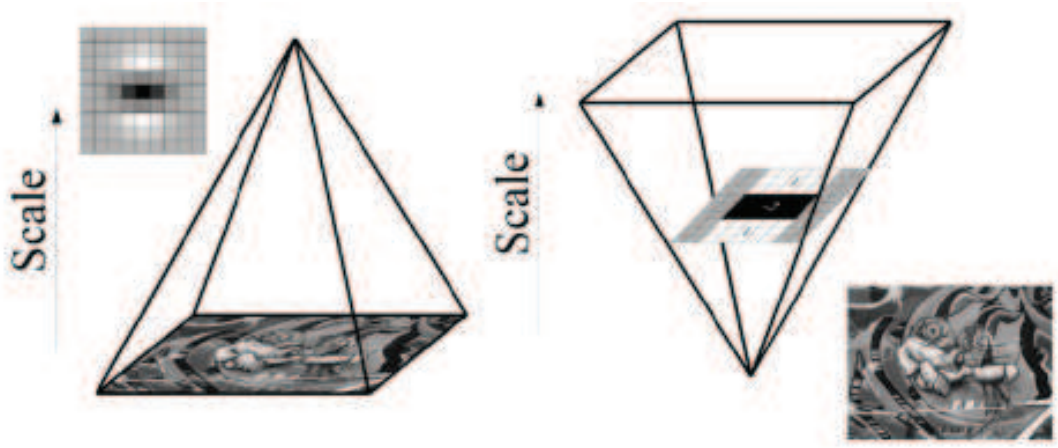

Abbildung 2.11: Der Skalenraum erweitert die 2D-Bilddaten um eine dritte Dimension durch die Skalierung eines Bilds. Auf der linken Seite wird der Skalenraum durch sukzessives Skalieren des Bild erreicht. Auf der rechten Seite wird der Skalenraum durch die Vergrößerung der Filterkerne aufgebaut. Bildquelle: [Bay 06]

Skalenraum

Das Grundkonzept des Skalenraums bildet eine Bildpyramide. Ein Eingabebild wird sukzessive in der Größe reduziert. Die Stufen der Skalierung bilden eine dritte Dimension im Skalenraum. Dieses Prinzip ist Abb. 2.11 veranschaulicht. Es wäre bereits im vorigen Schritt möglich gewesen, die Merkmale auf Grund der berechneten Maxima für eine fixe Skalierung zu extrahieren. Das SURF Verfahren extrahiert hingegen die Merkmale im Skalenraum und fügt eine weitere Information über die Größe hinzu. Damit wird es bei der Erstellung des Deskriptors möglich sein, eine Invarianz gegen Skalierung zu erreichen.

Das SURF-Verfahren berechnet im Unterschied zu anderen Verfahren, wie z.B. SIFT, keine explizite Bildpyramide durch sukzessive Skalierung. Der Parameter σ wird stattdessen genutzt um die Filterkerne zu vergrößern. Eine Gegenüberstellung ist in Abb. 2.11 zu sehen.
Die Vergrößerung kann nur nach bestimmten Regeln erfolgen, damit die Form der approximierten Kerne erhalten bleibt. SURF unterscheidet dabei die Vergrößerung der Kerne in *Oktaven*, die eine Reihe von Skalierungen (*Intervalle*) beinhalten.
In Abb. 2.10 sind bereits die kleinsten Kerne mit einer Größe von 9×9 Pixeln zu sehen gewesen. Die kleinste Kante *(lobe)* der weißen und schwarzen Fläche wird in der ersten Oktave um zwei Pixel vergrößert. Bei der Vergrößerung sind die Bedingungen, dass der mittlere Pixel erhalten bleibt, das Verhältnis der Flächen nicht verändert wird und der Filter eine gerade Größe behält, einzuhalten. Für die erste Oktave vergrößert sich die Seitenlänge der Kerne damit für jedes Intervall um 6 Pixel. Die Vergrößerung des Kerns ist in Abb. 2.12 aufgezeichnet.
Der Unterschied zu den anderen Oktaven besteht in der Vergrößerung um 12, 24 oder 48 Pixel pro Intervall.

Die lokalen Maxima der Determinante werden nun im Skalenraum gesucht. Dazu werden für einen Punkt (x, y, σ) die 27 benachbarten Punkte untersucht. Zu den Nachbarn zählen die

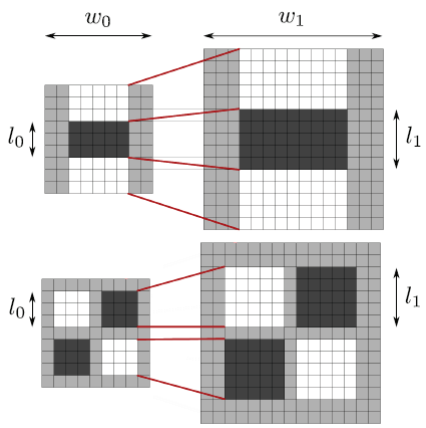

Abbildung 2.12: Bei der Vergrößerung des Kerns muss die Seitenkante der schwarzen oder weißen Flächen *(lobe)* l_0 um mindestens 2 Pixel vergrößert werden, damit der zentrale Pixel erhalten bleibt, die Filtergröße w_1 gerade ist und das Verhältnis der Flächen beibehalten werden kann. Bildquelle:[Bay 06]

jeweils 9 Punkte aus der größeren und kleineren Skalierung sowie die 8 angrenzenden Punkte bei Skalierung σ. Ein Maximum ist gefunden, wenn alle Nachbarn kleiner als der eigene Wert sind. Eine Schwellwertfilterung verhindert, dass jedes lokale Maximum zu einer Merkmalsposition erklärt wird. Zur genauen Lokalisierung des Maximums wird eine Interpolation im Skalenraum und im Bild durchgeführt. Eine genauere Beschreibung der verwendeten Verfahren ist in [Lowe 04] zu finden. Am Ende der Berechnung ist eine Subpixel-genaue Lokalisierung des Interessenpunkts im Skalenraum (x, y, σ) bekannt.

Deskriptor

Nachdem ein Merkmal im Bild lokalisiert ist, folgt nun dessen Beschreibung. Die Beschreibung basiert auf den Intensitäten in der Nachbarschaft der Merkmalsposition. Im Falle von SURF werden Intensitätsänderungen in der Umgebung berechnet.
Zu diesem Zweck wird das Haar-Wavelet benutzt. Der rechteckige Filterkern des Haar-Wavelet ist in Abb. 2.13 dargestellt. Bei der Faltung mit diesem Kern wird die Intensitätsdifferenz der beiden Flächenteile berechnet. Zur schnelleren Auswertung werden die Integralbilder verwendet.

Die Berechnung der Beschreibung erfolgt in zwei Schritten. Die weiteren Schritte sind abhängig von der Skalierung σ. Der erste Schritt der Berechnung besteht aus der Auszeichnung einer reproduzierbaren Richtung. Die Umgebung des Merkmals wird in vertikaler und horizontaler Richtung mit dem Haar-Wavelet ausgewertet. Die erhaltenen Werte (dx, dy) für die beiden Filterrichtungen werden in ein Koordinatensystem aufgetragen (siehe Abb. 2.14). Die Ebene wird in 6 Winkelbereiche um den Ursprung aufgeteilt. Die Werte (dx, dy) der enthaltenen Punkte werden jeweils komponentenweise aufsummiert. Die beiden Summen werden zu einem Vektor zusammengefasst. Aus den 6 Vektoren der Segmente wird der längste Vektor als dominante Richtung gewählt. Zur weiteren Berechnung wird eine Rechteckregion um den Interessenpunkt betrachtet, die in der berechneten Richtung ausgerichtet ist. Dieser Schritt ermöglicht eine invariante Beschreibung der Merkmale gegen Rotationen.
Im zweiten Schritt wird die definierte Rechteckregion in 4×4 Unterregionen aufgeteilt, in

Abbildung 2.13: In der Abbildung sind die Filterkerne des Haar-Wavelets dargestellt. Die schwarze Fläche ist mit -1, die weiße Fläche mit $+1$ gewichtet. Durch Konvolution kann dadurch die Intensitätsdifferenz zwischen weißer und schwarzer Fläche berechnet werden. Der linke Kern berechnet die horizontale, der rechte Kern die vertikale Differenz. Bildquelle:[Bay 06]

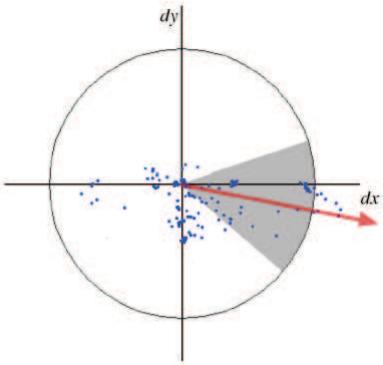

Abbildung 2.14: Die Auswertung des Haar-Wavelets in horizontaler und vertikalen Richtung wird in einem Koordinatensystem aufgetragen. Für verschiedene Kreissegmente werden die Koordinaten komponentenweise zu einem Orientierungsvektor aufsummiert. Bildquelle:[Bay 06]

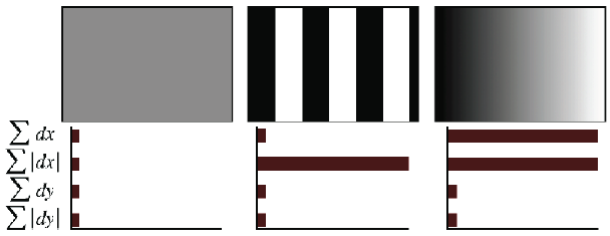

Abbildung 2.15: Die Bildmuster unterscheiden sich in den verschiedenen Summen der Haar-Wavelet-Auswertungen. Bildquelle:[Bay 06]

denen jeweils 5×5 Stichproben der Intensitäten ermittelt werden. Über diese Werte einer Unterregion wird erneut das Haar-Wavelet ausgewertet. Die Werte dx und dy der Auswertung werden in einem Vektor wie folgt aufsummiert:

$$\left(\sum dx, \sum |dx|, \sum dy, \sum |dy|\right) \tag{2.19}$$

Die Bedeutung der verschiedenen Komponenten ist Abb. 2.15 für verschiedene Extremfälle illustriert.
Die Menge der Vektoren der 4×4 Unterregionen bilden schließlich die Beschreibung des Merkmals. Durch die Verwendung des Haar-Wavelets, das nur die Änderungen der Intensitäten berechnet ist die Beschreibung unabhängig von der Grundhelligkeit. Durch eine Normierung des Vektors wird eine Invarianz gegenüber Kontrast erreicht.

2.3.2 Paarung der Merkmale

In diesem Abschnitt werden Strategien zur Paarung von Merkmalen besprochen. Die einzelnen Merkmale sind durch ihre *Merkmalsvektoren* beschrieben. Die Algorithmen zur Paarung bekommen meist eine große Menge an gefundenen Merkmalen der Bilder als Eingabe. Auf Grund der Merkmalsvektoren werden nun Korrespondenzen in den beiden Bildern gesucht. Eine Anforderung an die Algorithmen ist die Effizienz, damit die Paarung keine Engstelle im Ablauf der Bildregistrierung bildet. Auf der anderen Seite sollen möglichst wenige falsche Paarungen gebildet werden, damit sie im nächsten Schritt gut weiterverarbeitet werden können.

Ein Ähnlichkeitsmaß für Merkmalsvektoren u und v ist durch die euklidischen Distanz über die Einträge der Vektoren gegeben. Ein Standard-Ansatz um Paarungen zu bilden, ist die *Suche der nächsten Nachbarn*. In der klassischen Variante hat dieser Algorithmus jedoch eine schlechte Laufzeit. Eine approximierte Berechnung der nächsten Nachbarn durch randomisierte kd-Bäume [Muja 09] ist in ausreichender Geschwindigkeit möglich.
Eine Paarung nur auf Grund des nächsten Nachbarn erzeugt eine Vielzahl von falschen Paarungen. Das fehlende Auftreten des Partner-Merkmals kann ein Grund dafür sein. Ebenso kann durch veränderte Darstellungen der Merkmale in den Bildern ein anderer Partner gewählt werden. Eine Heuristik von Lowe [Lowe 04] baut auf der Wahl des nächsten Nachbars auf. Zur Paarung werden die zwei nächsten Nachbarn eines Merkmals betrachtet. Zwei Merkmale werden gepaart, wenn der Abstand d_1 zum ersten Nachbarn und der Abstand d_2 zum zweiten Nachbarn durch die Bedingung

$$d_1 < k \cdot d_2 \text{ mit Schwellwert } k < 1$$

erfüllt ist. Die Heuristik zieht die Eindeutigkeit der Paarung in Betracht, in dem die zweitbeste Paarung einen Mindestabstand einhalten muss. Die Eindeutigkeit lässt sich durch Wahl des Parameter k variieren.
Die Paarungen beruhen auf der Beschreibung der lokalen Umgebung der Merkmale. Dies ist eine Charakteristik bei merkmalsbasierten Verfahren. Anwendungsspezifisches Wissen kann aber oft benutzt werden, um globalere Eigenschaft bei der Paarung einzubringen.

2.3.3 Robuste Methoden

Die bestimmten Merkmals-Paarungen sind die Grundlage zur Berechnung einer Transformation zwischen den beiden Bildern. Bei den beschriebenen Strategien zur Paarung muss in Betracht gezogen werden, dass Ausreißer in Form von falschen oder ungenauen Paarungen auftreten. Diese Ausreißer führen bei Berechnungen zu Instabilitäten oder Abweichungen der Lösung. Es werden *robuste* Methoden benötigt, die die Möglichkeit von Ausreißern berücksichtigen.
Eine mögliche Methode, um eine robuste Berechnung durchzuführen, ist *Random Sample Consensus* (RANSAC) [Fischler 81]. Als Eingabe ist eine Datenmenge gegeben. Die Paare aus Ursprungs- und Zielpunkt einer unbekannten Transformation sind die Datenpunkte der Datenmenge.
Die Methode arbeitet iterativ und nimmt immer nur eine Untermenge *(Consensus-Menge)* der Datenpunkte zur Berechnung. Die Methode basiert auf der Idee, dass nach genügend Iterationen n_{it} eine Berechnung mit einer Ausreißer-freien Untermenge durchgeführt wird. Die Größe der Untermenge ist durch die kleinste Anzahl n_{cs}, der zur Lösung benötigten Punkte, bestimmt. Beliebige Berechnungen können nach folgendem Schema durch RANSAC gelöst werden:

1. Ziehe zufällig n_{cs} Stichproben aus der Datenmenge. Die gezogenen Datenpunkte bilden die Untermenge.
2. Führe die Berechnung der Transformation mit der Untermenge durch. Die gefundene Lösung ist für die verwendeten Datenpunkte gültig.
3. Wende die gefundene Lösung auf die gesamte Datenmenge an und prüfe, wie viele Datenpunkte innerhalb einer festgelegten Abweichung ϵ liegen. Speichere die Lösung, falls sie bislang die meisten Datenpunkte aus der Datenmenge erfasst.
4. Wiederhole n_{it}-mal Schritt 1. - 3.

Nach n_{it} Iterationen ist eine Lösung bekannt, die die größte Anzahl an Datenpunkten aus der Datenmenge erfasst. Die gefundene Lösung ist unter zwei Voraussetzung mit einer festgelegten Wahrscheinlichkeit p die gesuchte Lösung.
Erstens muss der Wert ϵ so gewählt sein, dass die Abweichung von richtigen Datenpunkten innerhalb der ϵ-Abweichung liegt. Zweitens müssen mehr richtige Datenpunkte als Ausreißer vorhanden sein. Die relative Häufigkeit von Ausreißern muss $r_A < 0.5$ erfüllen. Die zweite Voraussetzung ist keine zwingende Bedingung. Eine korrekte Lösung ist möglich, aber kann im Allgemeinen nicht vorausgesetzt werden.
Der Parameter n_{it} für die Anzahl der Iterationen hängt von der relativen Häufigkeit r_A von Ausreißern und der Größe der Untermenge ab. Es wird solange iteriert, bis eine große Wahrscheinlichkeit p für eine Ausreißer-freie Untermenge erreicht ist. Die Wahrscheinlichkeit,

dass bei n_{it} immer mindestens ein Ausreißer enthalten ist beträgt:

$$(1 - (1 - r_A)^{n_{cs}})^{n_{it}} = 1 - p$$

Damit lässt sich eine Ungleichung schreiben mit p als Mindest-Wahrscheinlichkeit für eine Ausreißer-freie Menge und nach n_{it} auflösen:

$$\begin{aligned} p \leq & 1 - (1 - (1 - r_A)^{n_{cs}})^{n_{it}} \\ (1 - (1 - r_A)^{n_{cs}})^{n_{it}} \leq & 1 - p \\ n_{it} \geq & \frac{log(1-p)}{log(1 - (1 - r_A)^{n_{cs}})} \end{aligned}$$

Die folgende Tabelle zeigt wie sich die Parameter n_{cs} und die Fehlerhäufigkeit r_A für verschiedene gesuchte Transformationen auf die Anzahl der Iterationen auswirken. Es wird $p = 0.99$ als Schranke vorgegeben.

		$r_A = 0.1$	$r_A = 0.2$	$r_A = 0.3$	$r_A = 0.4$
Translation	$(n_{cs} = 1)$	2	3	4	6
Affine Transformation	$(n_{cs} = 3)$	4	7	11	19
Projektionsmatrix	$(n_{cs} = 6)$	7	16	37	97

Aus der Tabelle ist ersichtlich, dass mit größer werdender Häufigkeit der Ausreißer und steigender Größe des Untermenge die Anzahl der benötigten Iterationen stark steigt. Für Berechnungen von Lösungen mit langer Dauer existieren aus diesem Grund weitere Varianten von RANSAC. Ein präemptives RANSAC [Nister 03] bricht zum Beispiel die Iteration vorzeitig ab, wenn genügend Datenpunkte die gefundene Lösung unterstützen.

2.4 Pixelbasierte Bildregistrierung

Dieser Abschnitt thematisiert die pixelbasierte Bildregistrierung am Beispiel des Template-Matchings. Das Ziel des Template-Matchings ist die Suche nach der Ausrichtung eines Template-Bilds in einem größeren Bild. Die Problemstellung ist in Abb. 2.16 illustriert. Die Parametrisierung der möglichen Ausrichtungen bildet einen Suchraum. Jeder Punkt im Suchraum kann mit einer Transformation des Templates assoziiert werden. Die Übereinstimmung des Templates und des Zielbilds kann auf Grund einer Metrik bewertet werden. Die Metrik betrachte die Pixelwerte an den gemeinsamen Pixelkoordinaten. Mögliche Metriken sind im nächsten Unterabschnitt vorgestellt [Bradski 00]. Das Template-Matching lässt sich als Suche eines Optimums im Suchraum auffassen.

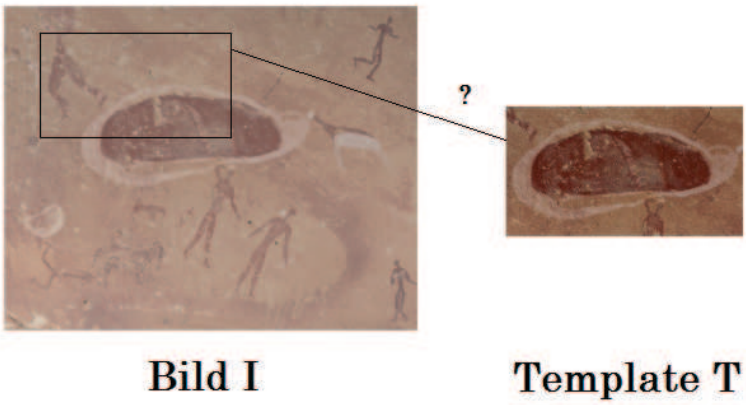

Abbildung 2.16: Die Problemstellung des Template-Matchings ist die Suche nach der besten Ausrichtung des Templates im Bild. Eine Bewertung zwischen Template und der Region des Bilds ist nötig.

2.4.1 Metriken

Es werden im folgenden verschiedene Metriken zur Messung der Ähnlichkeit eines Bildes I und eines Templates R vorgestellt. Zur jeder vorgestellten Norm wird eine Auswertung für das in Abb. 2.16 gezeigte Bild und Template durchgeführt. Das Template ist an jede mögliche Stelle im Bild verschoben worden. Die Ergebnisse sind als Grauwert-Bild dargestellt. Jeder Pixel (x, y) stellt den Wert bei einer Verschiebung des Templates um (x, y) dar. Die erhaltenen Werte der Metrik sind auf die Grauwerte [0...255] abgebildet. Das Bild wurde bewusst im Vergleich zum Template aufgehellt.

Distanz

Die erste Definitionsmöglichkeit einer Metrik ist die quadratische euklidische Norm (2.20). Die Abstände der einzelnen Pixelwerte werden quadratisch aufsummiert. Eine Variante der Metrik (2.21) ergibt sich durch Normierung der Summe mit der Gesamtintensität des Bildausschnittes. Diese Norm gewichtet abweichende Werte stark.

$$R(x,y) = \sum_{(i,j)\in T} (T(i,j) - I(x+i, y+j))^2 \tag{2.20}$$

$$R(x,y) = \frac{\sum_{(i,j)\in T} (T(i,j) - I(x+i, y+j))^2}{\sqrt{\sum_{(i,j)\in T} (T(i,j))^2 \sum_{(i,j)\in T} (I(x+i, y+j))^2}} \tag{2.21}$$

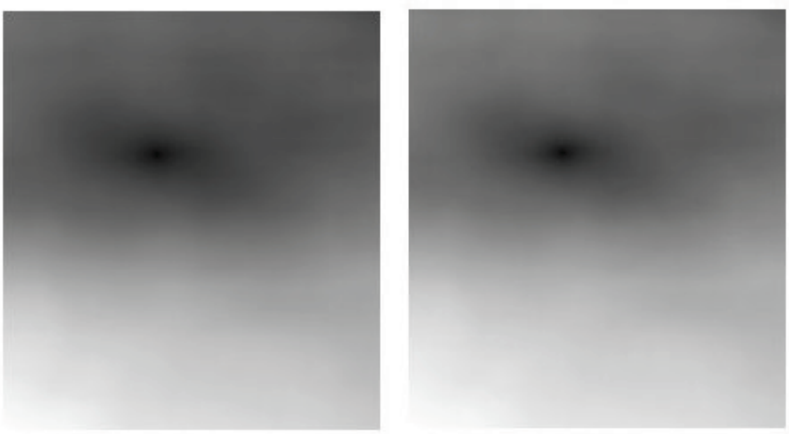

Abbildung 2.17: Das Bild und Template aus Abb. 2.16 sind mit den Metriken 2.20 (links) und 2.21 (rechts) verglichen worden. Das Minimum ist deutlich zu erkennen. Durch die Aufhellung des Referenzbilds ist das Bild der nicht normierten Metrik heller, da die Grundabweichung größer ist.

Kreuz-Korrelation

Die Kreuz-Korrelation ergibt sich aus der elementweisen Multiplikation der Pixelwerte. Die Korrelation betrachtet nicht im Vergleich zur euklidischen Norm die Abweichung. Die Bilder werden als diskrete Werte eines 2D-Signals betrachtet. Die Korrelation drückt die Ähnlichkeit zwischen den Signalen aus. Eine große Ähnlichkeit wird durch einen großen Wert sichtbar. Die Definitionen der Korrelation und der normierten Variante sind in 2.22 und 2.23 zu sehen. Der Effekt der Normierung ist in der Auswertung in Abb. 2.18 zu erkennen. Die Aufhellung des Bildes hat in der Auswertung zur Folge, dass das das Maximum nur schwach zu erkennen ist.

$$R(x,y) = \sum_{(i,j)\in T} (T(i,j) \cdot I(x+i, y+j)) \tag{2.22}$$

$$R(x,y) = \frac{\sum_{(i,j)\in T} (T(i,j) \cdot I(x+i, y+j))}{\sqrt{\sum_{(i,j)\in T} (T(i,j))^2 \sum_{(i,j)\in T} (I(x+i, y+j))^2}} \tag{2.23}$$

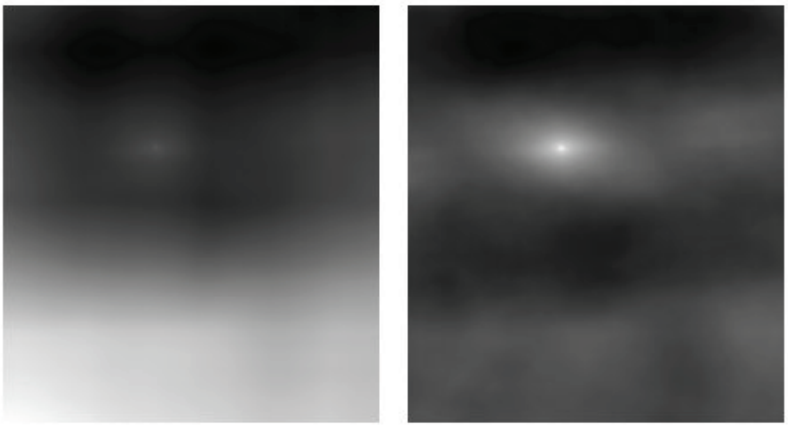

Abbildung 2.18: Das Bild und Template aus Abb. 2.16 sind mit den Metriken 2.22 (links) und 2.23 (rechts) verglichen worden. Die Aufhellung des Grundbildes hat starken Einfluss auf die nicht-normierte Variante.

Korrelationskoeffizient

Zur Berechnung des Korrelationskoeffizient werden zuerst die Mittelwerte der Intensitäten für die Bilder bestimmt. Die Anzahl der Pixel im Template sei dabei M. Die Bilder T' und I' sind die Differenzbilder zum Mittelwert. Der Korrelationskoeffizient wird nun auf Grund der Bilder T' und I' durch Korrelation berechnet. Die Metrik hat einen statistischen Ansatz, der die Kovarianz im Nenner und die Varianz im Zähler von 2.27 betrachtet.

$$T'(i,j) = T(i,j) - \frac{1}{M} \sum_{(x',y') \in T} T(x',y') \tag{2.24}$$

$$I'(x+i,y+j) = I(x+i,y+j) - \frac{1}{M} \sum_{(x',y') \in T} I(x+x',y+y') \tag{2.25}$$

$$R(x,y) = \sum_{(i,j) \in T'} (T'(i,j) \cdot I'(x+i,y+j)) \tag{2.26}$$

$$R(x,y) = \frac{\sum_{(i,j) \in T'} (T'(i,j) \cdot I'(x+i,y+j))}{\sqrt{\sum_{(i,j) \in T'} (T'(i,j))^2 \sum_{(i,j) \in T'} (I'(x+i,y+j))^2}} \tag{2.27}$$

Die Aufhellung des Bildes I hat in der Auswertung des Korrelationskoeffizients nach (2.26) und des normierten Korrelationskoeffizients (2.27) keine Gewichtung, da dieser Unterschied durch die Entfernung des Mittelwerts entfällt.

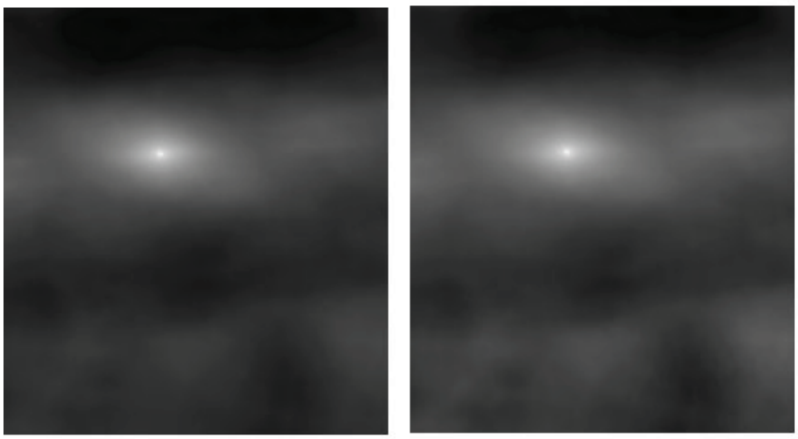

Abbildung 2.19: Das Bild und Template aus Abb. 2.16 sind mit den Metriken 2.26 (links) und 2.27 (rechts) verglichen worden. Die Aufhellung des Grundbildes hat keinen Einfluss, da die Metrik eine Kreuzkorrelation auf Grund des dynamischen Teils des Bildes berechnet.

2.5 Projektionsmatrix-Bestimmung

In diesem Abschnitt wird die Berechnung einer Projektionsmatrix [Hartley 04] vorgestellt. Die Erläuterung der Eigenschaften der Projektionsmatrix erfolgte bereits in Abschnitt 2.1.2. Der Algorithmus erwartet als Eingabe eine Menge von 3D-2D-Korrespondenzen $(\hat{X}_i, \hat{x}_i)$ mit $i = 1, ..., N$.
Die Projektionsmatrix wird die Bedingung

$$\lambda \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} \cong \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \tag{2.28}$$

für die 3D-2D-Korrespondenzen bestmöglich erfüllen.
Der Algorithmus arbeitet in zwei Schritten. Im ersten Schritt(Kap. 2.5.1) wird eine lineare Lösung bestimmt. Die lineare Lösung kann spezifische Beschränkungen, wie z.B. Vorwissen über das Kameramodell nicht berücksichtigen. Deshalb wird im zweiten Schritt (Kap. 2.5.2) durch ein Optimierungsverfahren die lineare Lösung verbessert und die Beschränkungen werden berücksichtigt. Die Stabilität der Berechnung wird durch eine Datennormalisierung (Kap. 2.5.3) verbessert.

2.5.1 Lineare Lösung

Die Gleichung 2.28 beschreibt die projektive Äquivalenz zwischen 3D-2D-Korrespondenzen. Die Punkte der Korrespondenzen können frei skaliert werden, ohne die Äquivalenz zu verletzen. Bei der Berechnung wird die Freiheit der Skalierung durch die Formulierung mit Hilfe

des Kreuzprodukts eliminiert. Eine Umformung liefert drei lineare Gleichungen.

$$\hat{x}_i \times P\hat{X}_i = 0$$

$$(x_i, y_i, 1)^T \times P(X_i, Y_i, Z_i, 1)^T = 0$$

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} \times \begin{pmatrix} p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14} \\ p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24} \\ p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34} \end{pmatrix} = 0$$

$$\begin{pmatrix} y_i(p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}) & - & (p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24}) \\ -x_i(p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}) & + & (p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14}) \\ x_i(p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24}) & - & y_i(p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14}) \end{pmatrix} = 0$$

Die Umformung liefert zwei unabhängige Gleichungen für jedes Korrespondenzpaar. Die Projektionsmatrix besitzt 11 unbekannte Einträge. Ein Eintrag kann wegen der freien Skalierung festgelegt werden. Zur Lösung der unbekannten Einträge werden somit mindestens 6 Paare benötigt, die in folgender Form als Koeffizientenmatrix A geschrieben werden.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -X_1 & -Y_1 & -Z_1 & -1 & y_1X_1 & y_1Y_1 & y_1Z_1 & y_1 \\ X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 & -x_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & -X_N & -Y_N & -Z_N & -1 & y_NX_N & y_NY_N & y_NZ_N & y_N \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -x_NX_N & -x_NY_N & -x_NZ_N & -x_N \end{pmatrix} \tag{2.29}$$

Die Koeffizientenmatrix ist Grundlage für das lineare Gleichungssystem.

$$A \cdot p = 0$$

Der Vektor p besitzt die 12 Einträge der Projektionsmatrix

$$p = (p_{11}, p_{12}, p_{13}, p_{14}, p_{21}, p_{22}, p_{23}, p_{24}, p_{31}, p_{32}, p_{33}, p_{34})^T$$

und kann nach der Lösung des Gleichungssystems in die bekannte 3×4-Matrix umgeschrieben werden. Das Gleichungssystem lässt sich durch Anwendung einer Singulärwert-Zerlegung (SVD) lösen.

2.5.2 Nichtlineare Optimierung

Die lineare Lösung ist meist keine exakte Lösung. Ein überbestimmtes Gleichungssystem liefert eine Least-Square-Lösung. Beschränkungen, wie ein Vorwissen über das Kameramodell, können mit der linearen Lösung nicht erzwungen werden. Die lineare Lösung wird als Startwert für ein Optimierungsverfahren benutzt. Ein Optimierungsverfahren benötigt eine Parametrisierung der Projektionsmatrix und eine zu minimierende Fehlerfunktion.
Die Parametrisierung erfasst die inneren und äußeren Parameter. Die äußeren Parameter

setzen sich aus der Rotation und der Translation zusammen, die die Orientierung und Position der Kamera beschreiben. Die Rotation lässt sich auf verschiedene Weisen darstellen. Es gibt z.B die Möglichkeiten Quaternionen, eine Drehachse mit Winkel oder Euler-Winkel [Szeliski 06] zu benutzen. Die Translation lässt ebenfalls die Wahl zwischen einer Darstellung in kartesischen oder homogenen Koordinaten. Der Vorteil bei homogenen Koordinaten ist die Möglichkeit, durch kleine Änderungen der w-Komponente große Veränderungen in den kartesischen Koordinaten auszudrücken. Die Wahl der inneren Parameter hängt vom Vorwissen oder einer Vermutung des Kameramodells ab.
Die Fehlerfunktion F kann als die Summe der euklidischen Differenzen zwischen den projizierten Punkten (x'_i, y'_i) und den korrespondieren Bildpunkten (x_i, y_i) gewählt werden. Die Fehlerfunktion hängt von der gewählten Parametrisierung $\alpha_1, ..., \alpha_m$ ab.

$$F(\alpha_1, ..., \alpha_m) = \sum_{i=1...N} \sqrt{(x'_i - x_i)^2 + (y'_i - y_i)^2}$$

Ein Optimierungsverfahren hat das Ziel, die definierte Fehlerfunktion schrittweise zu minimieren. Es wird ein sog. Gradientenabstieg durchgeführt. Der Gradientenvektor setzt sich aus den partiellen Ableitungen der Fehlerfunktion nach den einzelnen Parametern zusammen. Der Vektor gibt die Richtung des steilsten Anstieges der Fehlerfunktion an.

$$\nabla F = \left(\frac{\partial F}{\partial \alpha_1}, ..., \frac{\partial F}{\partial \alpha_m} \right)^T$$

In jeder Iteration folgt das Verfahren eine Schrittweite der Richtung des negativen Gradienten. Das Verfahren iteriert solange, bis ein Minimum gefunden ist und sich die Fehlerdifferenz nicht mehr ändert. Das gefundene Minimum ist nicht zwangsläufig das globales Minimum. Eine gute Startbelegung der Parameter ist für den Erfolg des Verfahren erforderlich, damit das Verfahren das globale Minimum erreicht und die Parameter der gesuchten Lösung entsprechen.
Das Levenberg-Marquardt-Verfahren [Levenberg 44] ist ein häufig genutztes Optimierungsverfahren, das nach dem beschriebenen Schema arbeitet.

2.5.3 Daten-Normalisierung

Im Vorfeld der Berechnung wird eine Daten-Normalisierung durchgeführt. Sowohl die 3D-Koordinaten als auch die Bildkoordinaten können in verschiedenen Anwendungen sehr unterschiedliche Wertebereiche annehmen. Daher ist eine Normalisierung notwendig, um eine Stabilität des Algorithmus zu gewährleisten. Die Daten werden in neue Koordinatensysteme gelegt.
Für die Punkte $\hat{X}_i$ im 3D-Raum und die Bildpunkte wird der Schwerpunkt berechnet. Der Schwerpunkt bildet den Ursprung des neuen Koordinatensystems. Nach der Translation der 3D-Daten in das neue Koordinatensystem wird die durchschnittliche Abweichung in jede Achse auf den Wert 1 skaliert. Die Normalisierung entspricht einer Transformation der Form:

$$T = \begin{pmatrix} s_x & 0 & 0 & t_x \\ 0 & s_y & 0 & t_y \\ 0 & 0 & s_z & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Die Bildkoordinaten werden analog in ein Koordinatensystem mit den beschriebenen Eigenschaften transformiert. Die Transformation lässt sich in diesem Fall als 3×3-Matrix U schreiben.
Die Berechnung der Projektionsmatrix findet auf Grund der normalisierten Daten statt und ergibt die normalisierte Projektionsmatrix P_N. Die Normalisierung eines Punktes, Anwendung der normalisierten Projektionsmatrix und anschließender Denormalisierung kann einfach durch Matrixmultiplikation erreicht werden.

$$P = U^{-1} \cdot P_N \cdot T$$

Kapitel 3

Lösungsansätze

Das Thema dieses Kapitels ist die Lösung der Problemstellung. Die Problemstellung umfasst drei Teilprobleme. Die beiden ersten Teilprobleme sind die Registrierung der Panoramen und die Registrierung der hochauflösenden Bilder mit den 3D-Daten. Die folgende Lösung der beiden Teilprobleme liefert Texturkoordinaten, mit denen eine Zuordnung zwischen 3D-Daten und Bilddaten hergestellt wird. Das dritte Teilproblem ist die Erstellung von orthographischen Projektionen, die auf Grund der berechneten Texturkoordinaten erstellt werden können.
Der Lösungsansatz dieses Teilproblems wird im ersten Abschnitt erläutert. Die Berechnung der Texturkoordinaten durch die Registrierung der Panoramen und die Registrierung der hochauflösenden Bildern wird in Abschnitt 3.2 besprochen. In Abschnitt 3.3 wird eine Möglichkeit der nachträglichen Korrektur der Texturkoordinaten vorgestellt.

3.1 Erstellung der orthographischen Darstellung

In diesem Abschnitt wird die Erstellung der orthographischen Projektionen erläutert. Die orthographische Projektion ist eine Methode um ein Bild einer 3D-Szene zu erzeugen. Im Gegensatz zur perspektivischen Projektion hat der Abstand eines Objekts zur Kamera keinen Einfluss. Es entstehen keine perspektivischen Längenverzerrungen des Objekts. Die orthographische Darstellung ist lediglich von der Blickrichtung abhängig. Das Objekt wird in Blickrichtung auf die Bildebene projiziert.
Die orthographischen Darstellungen werden von den Archäologen von kleinen Ausschnitten der Felswand erstellt. Die Blickrichtung wird senkrecht zur Felswand gewählt. Der erste Vorteil ist eine einheitliche Darstellung der Ausschnitte. Der zweite Vorteil ist die Vermeidung der projektiven Verzerrung. In der orthographischen Darstellung wird senkrecht auf die Felswand geschaut, so dass z.B. eine Vermessung möglich ist.
Die Erstellung der orthographischen Darstellung auf Grund der 3D-Daten und der Bilddaten wird in den nächsten Abschnitten erklärt. Es wird vorausgesetzt, dass die Bilddaten bereits mit den 3D-Daten registriert worden sind. Die Erstellung besteht aus drei Schritten. Die Bestimmung der Projektionsrichtung wird im ersten Unterabschnitt erklärt. Der zweite Unterabschnitt beschreibt den Schritt der Triangulierung. Im dritten Unterabschnitt wird die Verwendung der Rendering-Pipeline zur Erzeugung der orthographischen Darstellung erläutert.

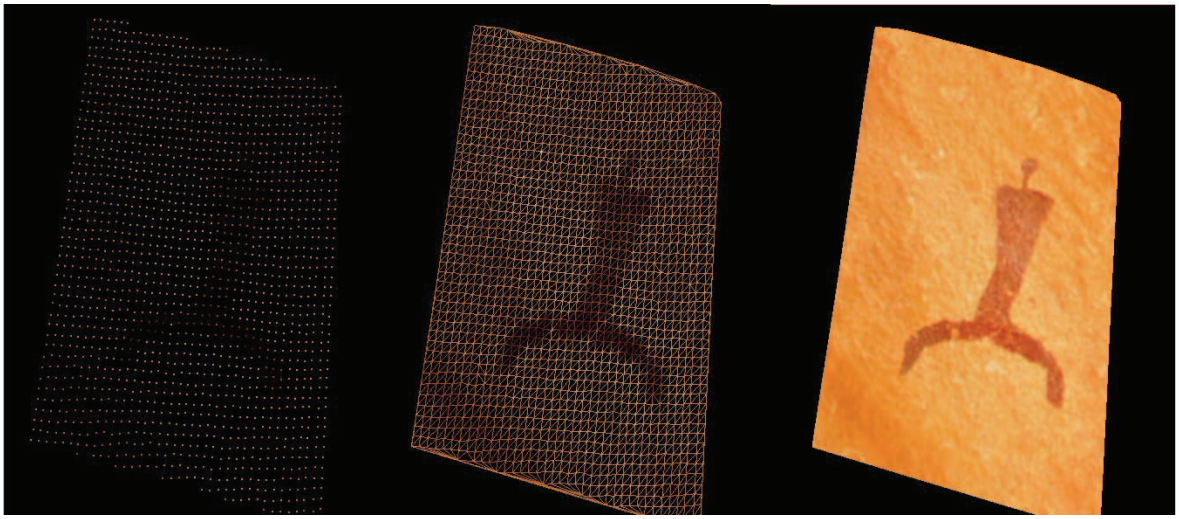

Abbildung 3.1: Auf der linken Seite ist eine Auswahl der Punktwolke gezeigt. Die ursprüngliche Auflösung ist um Faktor 5 in jede Richtung reduziert. In der mittleren Abbildung ist die Triangulierung der Punkte zu sehen. Das Gitternetz bildet die Kanten der Flächen. Auf der rechten Seite wurden diese Flächen durch die Textur interpoliert.

3.1.1 Bestimmung der Projektionsrichtung

Die orthographischen Darstellungen werden von einem Ausschnitt der Felswand erstellt. Die Felswand zeigt in kleinen Ausschnitten eine geringe Krümmung, so dass die Ausschnitte als planar angenommen werden können. Die Projektionsrichtung kann als Normale einer eingepassten Ebene berechnet werden.
Zuerst wird der Schwerpunkt μ der Punkte des Ausschnitts bestimmt. Der Schwerpunkt ist notwendig, um die Kovarianzmatrix Σ zu berechnen. Für eine Menge M der Punkte X_i des Ausschnitts mit $1 \leq i \leq n = |M|$ lässt sich die Kovarianzmatrix wie folgt berechnen:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{3.1}$$

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu) \cdot (X_i - \mu)^T \tag{3.2}$$

Die Normale der eingepassten Ebene lässt sich als Eigenvektor zum kleinsten Eigenwert λ finden. Der Wert λ gibt die Streuung in Richtung der Normale an. Er lässt sich als Kontrollwert einsetzen, um die Planarität des Ausschnitts zu kontrollieren. Der berechnete Schwerpunkt wird als Kameraposition benutzt. Die Projektion aus dieser Position muss sowohl die Punkte vor als auch hinter der Kameraposition berücksichtigen, da die Kamera im Mittelpunkt der Streuung liegt.

3.1.2 Triangulierung der Punktwolke

Die 3D-Daten sind bislang immer als Punktwolke betrachtet worden. Die Auflösung der Punktwolke reicht für die Erstellung von hochauflösenden Darstellungen nicht aus. In diesem

Abschnitt wird die Punktwolke durch eine Triangulierung zu einer Oberfläche erweitert, damit der Zwischenraum mit zusätzlichen Farbinformationen gefüllt werden kann (siehe Abb. 3.1). Die Triangulierung generiert eine Menge von Dreiecken, die durch drei benachbarte Punkte aufgespannt werden.
Die Triangulierung muss nicht im 3D-Raum durchgeführt werden, sondern kann in der Ebene gelöst werden. Die Position des Laserscanners ist bekannt. Jeder Punkt der Punktwolke ist aus dieser Position sichtbar. Die Punkte lassen sich damit eindeutig durch die Winkelkoordinaten φ und θ der sphärischen Koordinaten beschreiben. Die Punkte werden auf Grund der Koordinaten (φ, θ) in der Ebene trianguliert.
Ein Sonderfall tritt auf, wenn der gewählte Ausschnitt den Übergang zwischen 0° und 360° enthält. Die Punkte sind zwischen 0° und 360° benachbart, besitzen aber den größten Abstand. Die Koordinaten müssen in diesem Fall verschoben werden. Die Erkennung des Sonderfalls erfolgt über die Berechnung der Punktverteilung im Intervall $[0°, 360°]$. Das Maximum der Verteilung wird an die Position 180° verschoben. In der Praxis beträgt der Winkelumfang wenige Grad, so dass Problemfälle mit großem Winkelumfang nicht in Erscheinung treten.
Ein weiterer Sonderfall ist ein Ausschnitt, der $\theta = 0$ oder $\theta = 180°$ enthält. Dieser Fall hat in der Praxis auch keine sinnvolle Anwendung.

3.1.3 Verarbeitung durch die Rendering-Pipeline

Der dritte Schritt der Erstellung der orthographischen Darstellung besteht aus der Projektion der triangulierten Fläche. Die Projektion wird durch die Rendering-Pipeline ausgeführt. Die Bilddaten werden als Textur benutzt und beim Rendern automatisch auf die Fläche aufgebracht. Die Rendering-Pipeline benötigt für diesen Vorgang die Definition von Texturkoordinaten.
Die Texturkoordinaten definieren für einen 3D-Punkt eine Position in der Textur. Die drei Eckpunkte einer Dreiecksfläche bestimmen durch die zugehörigen Texturkoordinaten ebenfalls ein Dreiecksfläche in der Textur. Die Dreiecksfläche im Raum wird mit dieser Region der Textur koloriert.
Die Verbindung zwischen 3D-Daten und Bilddaten ist durch die Texturkoordinaten bestimmt. Das Ziel der Registrierung der Bilddaten im nächsten Abschnitt ist die Generierung dieser Texturkoordinaten.
Die Rendering-Pipeline ermöglicht eine beschleunigte Berechnung durch Unterstützung der Grafikkarte. Im Unterschied zur normalen Benutzung der Rendering-Pipeline wird die Ausgabe in den Grafikpuffer gerendert. Dieses Verfahren ist als *Offscreen-Rendering* bekannt. Der Grafikpuffer wird als Ausgabe abgespeichert.

3.2 Bildregistrierung für Texturkoordinatenberechnung

Im vorigen Abschnitt ist bereits erläutert worden, dass die 3D-Daten und die Bilddaten über Texturkoordinaten zusammengeführt werden. Das Thema in diesem Abschnitt ist die Berechnung der Texturkoordinaten mit Hilfe einer Bildregistrierung. Das Vorgehen und die auftretenden Probleme sind bei den Panoramen und den hoch auflösenden Fotos unterschiedlich, so dass diese Themen getrennt in den Unterabschnitten vorgestellt werden.

3.2.1 Registrierung der Panoramen

Als Eingabe der Bildregistrierung werden im Allgemeinen immer zwei Bilder vorausgesetzt, deren Ausrichtung als Transformationsvorschrift gesucht wird. Die Farbwerte der 3D-Daten werden aus diesem Grund im ersten Schritt in eine sphärische Darstellung gebracht. Eine sphärische Darstellung ist eine Projektion der 3D-Punkte auf eine Kugel um den Betrachter. Die horizontale Richtung in der Darstellung beschreibt eine Drehung des Betrachters um eine definierte Drehachse. Die Panoramen liegen ebenfalls in der sphärischen Darstellung vor. Unter der Annahme, dass beide Drehachsen identisch sind, muss im zweiten Schritt eine Translation zwischen den Bildern berechnet werden. Die berechnete Ausrichtung wird im dritten Schritt verbessert, in dem eine Korrektur der Drehachse durchgeführt wird.
Die nächsten Unterabschnitte beschreiben die drei Schritte genauer.

Sphärische Projektion

Zur Berechnung einer sphärischen Projektion werden die Kartesischen Koordinaten (x, y, z) in sphärische Koordinaten (r, φ, θ) umgerechnet. Eine Normierung der Koordinaten φ und θ auf das Intervall $[0, 1]$ liefert die relativen Bildkoordinaten. Die sphärische Projektion bildet die Umgebung auf eine Kugel um den Betrachter ab.
Die Umgebung besteht aus der Menge an einzelnen Punkten. Das Ergebnis des Vorgehens ist eine Abbildung mit Lücken, wenn für bestimmte Winkelbereiche keine 3D-Daten vorliegen. Kleinere Lücken lassen sich jedoch durch den folgenden Algorithmus schließen.
Der Algorithmus interpoliert das Bild zuerst in vertikaler und in horizontaler Richtung. Die beiden Interpolationen werden unabhängig voneinander auf dem Eingabebild durchgeführt. In vertikaler Richtung arbeitet der Algorithmus spaltenweise. In horizontaler Richtung wird zeilenweise das Bild durchlaufen. Die Start- und Endpunkte vor einer Lücke werden für eine Spalte/Zeile ermittelt und die fehlenden Werte durch Interpolation berechnet. Die Interpolationsbilder der beiden Richtungen werden im nächsten Schritt vereint. Ein Pixel einer Lücke kann einen interpolierten Wert in einer oder in beiden Richtungen besitzen. In ersten Fall wird der interpolierte Wert der einen Richtung in das Eingabebild übernommen. Im zweiten Fall werden die beiden interpolierten Werte gemittelt. Das Verfahren wird iteriert, bis die Lücken ausreichend gefüllt wurden. Der Algorithmus ist in Abb. 3.2 illustriert.
Die sinnvolle Bildgröße der sphärischen Projektion hängt von der Anzahl der 3D-Punkte ab, da ansonsten sehr viele Freiräume zwischen den projizierten Punkten entstehen. Die verwendeten 3D-Daten haben eine Auflösung von 2024×861 abgetasteten Punkten, die auch zur Erstellung der sphärischen Darstellungen benutzt wird.

Translation

Nach der Erstellung der sphärischen Projektion kann eine Bildregistrierung mit den Panoramen erfolgen. Die Panoramen sind auf die Größe der sphärischen Projektion reduziert worden. Eine Detektion der Merkmale in voller Auflösung ist nicht möglich. Die Laufzeit und der Speicherverbrauch sind zu groß. Die Merkmale werden jedoch auf Subpixel-Genauigkeit bestimmt, sodass eine Extrapolation keinen großen Fehler bringen sollte.
Unter der Voraussetzung, dass beide Drehachsen der Darstellung identisch sind, ist eine Translation in x-Richtung zu berechnen. Die berechnete Translation wird die gesuchten Texturkoordinaten liefern. Für jeden 3D-Punkt beschreibt die Texturkoordinate eine korrespondierende

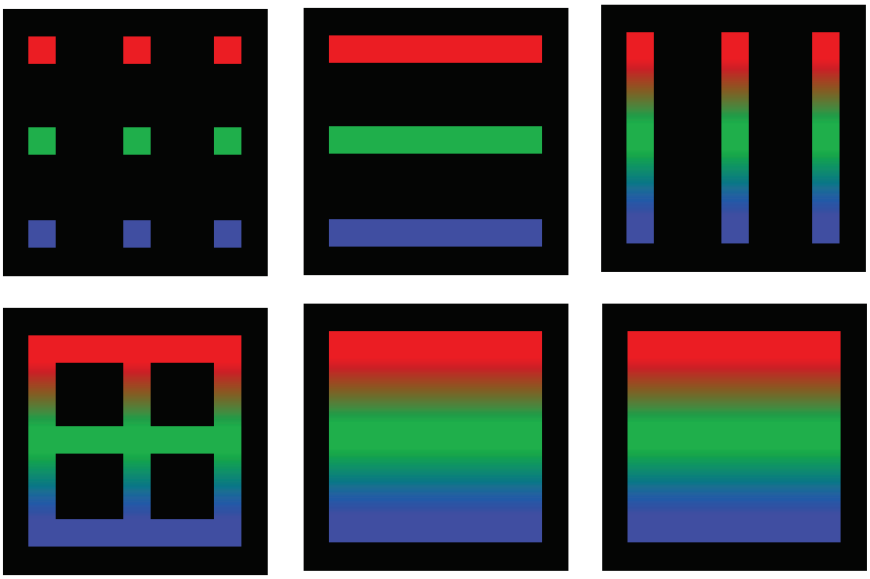

Abbildung 3.2: Die Abbildung zeigt die Arbeitsweise des Algorithmus zum Füllen von Lücken. Jede Zeile entspricht einer Iteration des Algorithmus. In der linken Spalte ist das Eingabebild vor jeder Iteration zu sehen. In der mittleren Spalte ist die Interpolation in horizontaler Richtung, in der rechten Spalte die Interpolation in vertikale Richtung gezeigt. Die Werte der beiden Interpolationen werden vereint und bilden die Eingabe der nächsten Iteration.

Position im Panorama. Die Berechnung der Texturkoordinaten unter der Annahme von identischen Drehachsen ist mit dem folgenden Vorgehen möglich:

1. Detektiere SURF-Merkmale in der sphärischen Projektion und dem Panorama. Um die folgende Paarung zu vereinfachen, kann die Detektion der Merkmale auf die obere Region des Bildes beschränkt werden.Der untere Teil des Bildes wird wegen der Uneindeutigkeit des Sandbodens ignoriert.

2. Führe eine Paarung der Merkmale durch (siehe Abb. 3.3). Bei der Paarung der Merkmale kann in Betracht gezogen werden, dass die Merkmale auf gleicher relativer Höhe gefunden werden müssen. Zwischen den beiden Bildern existiert keine Rotation, so dass ebenfalls die bestimmte dominante Richtung des Merkmals nur eine kleine Abweichung haben kann.

3. Berechne mit Hilfe von RANSAC eine Lösung der Translation in x-Richtung. Es ist zu beachten, dass die Translation $d\varphi$ einen Winkel repräsentiert. Daher gilt für die berechneten Werte

$$d\varphi < 0^\circ : d\varphi \mathrel{\hat{=}} 360^\circ - d\varphi \tag{3.3}$$
$$d\varphi \geq 360^\circ : d\varphi \mathrel{\hat{=}} d\varphi - 360^\circ \tag{3.4}$$

4. Die Texturkoordinate für einen Punkt $(x, y, z)^T$ lässt sich über die sphärische Koordinate (r, φ, θ) bestimmen, da das Bild der sphärischen Projektion ebenfalls auf Grund der Winkel φ und θ erzeugt worden ist. Die Texturkoordinate $(u, v) \in [0, 1] \times [0, 1]$

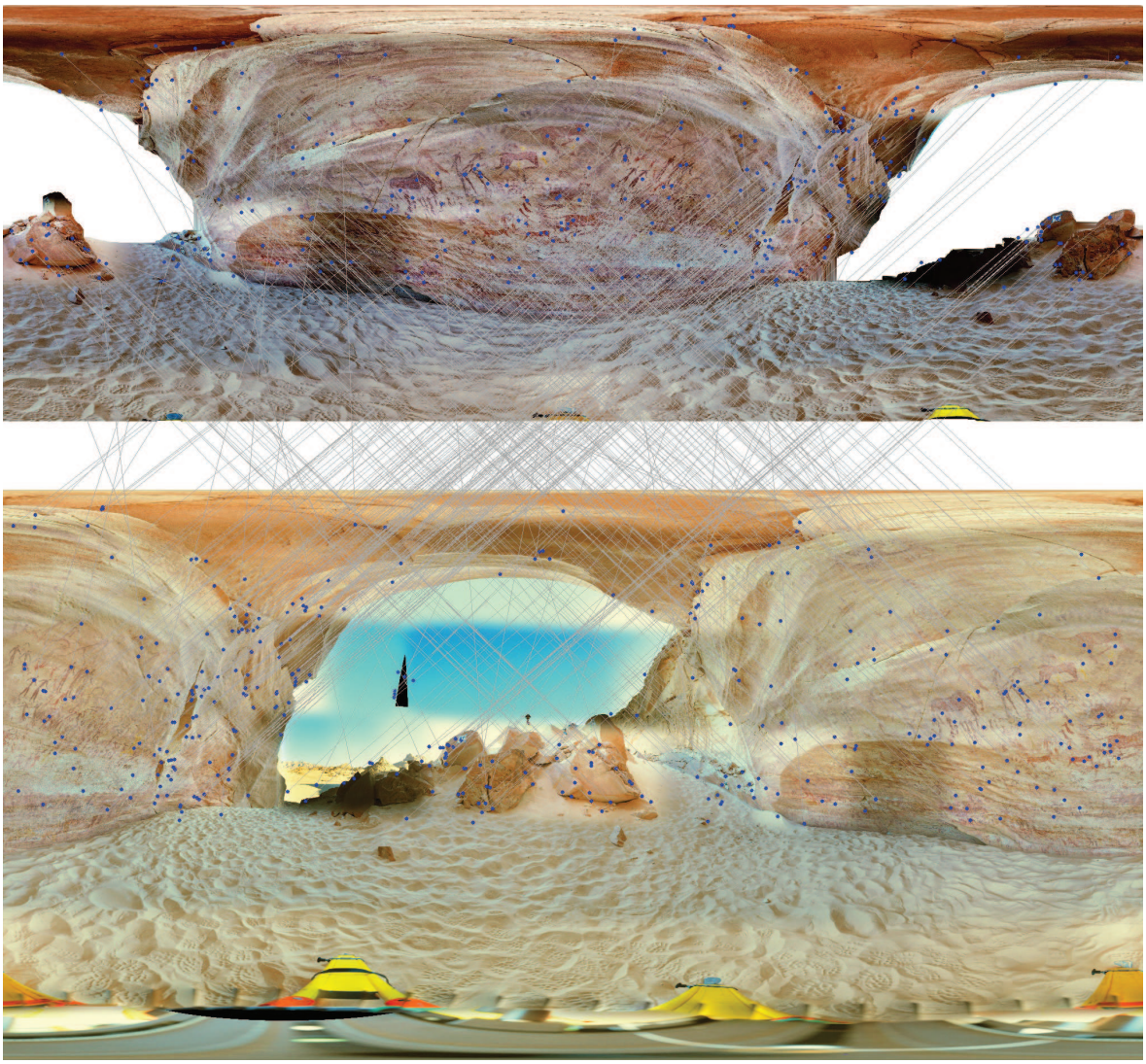

Abbildung 3.3: Auf dem oberen Bild ist eine sphärische Projektion der 3D-Daten zu sehen. Das untere Bild zeigt das zugehörige Panorama. In beiden Bilder sind SURF-Merkmal detektiert und gepaart worden. Die Merkmale sind als blaue Punkte markiert und mit dem korrespondieren den Partner verbunden.

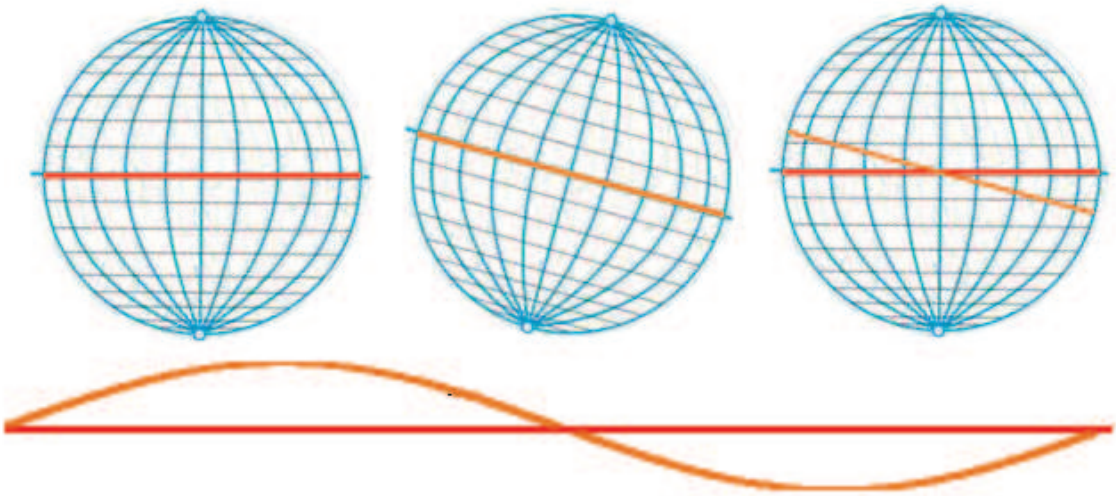

Abbildung 3.4: In der linken und mittleren Kugeloberfläche ist der Äquator markiert. Die Drehachse ist verkippt. Die beiden Kugeln sind in der rechten Kugel übereinander gelegt. Die sphärische Projektion unter Verwendung der Drehachse der linken Kugel liefert die untere Darstellung der Linien.

ergibt sich somit als:

$$(u, v) = (\frac{\varphi + d\varphi}{360°}, \frac{\theta}{180°})$$

Korrektur der Drehachse

Die Annahme, dass die Drehachsen der sphärischen Darstellungen gleich sind, ist in den vorliegenden Daten nicht erfüllt. Es existiert eine kleine Abweichung. Die wahrscheinlichste Ursache für die Verkippung der Drehachse ist der Austausch der Geräte auf dem Stativ. Zur Aufnahme der Einzelbilder für die Panoramen musste der Laserscanner durch die Kamera mit Nodalpunktadapter ausgetauscht werden. Durch diesen Umbau kann eine minimale Veränderung der Drehachse nicht ausgeschlossen werden.
Der Effekt der Verkippung ist im Abb. 3.4 gezeigt. In dieser Abbildung wird eine Linie entlang des Äquators auf der Kugeloberfläche betrachtet. Die Kugeln haben eine unterschiedliche Drehachse. In einem gemeinsamen Koordinatensystem wird bei der sphärischen Projektion die Linie auf der verkippten Kugel nicht mehr als Linie abgebildet. Die Linie beschreibt eine nicht lineare Kurve.
Die linear berechneten Texturkoordinaten berücksichtigen diesen Effekt nicht. Die Textur weist eine unterschiedliche Abweichung in vertikaler Richtung auf. Die beobachtete Änderung des Winkels zwischen den Drehachsen hat eine Größenordnung von 0.1°. Da die Panoramen eine sehr hohe Auflösung besitzen, wirkt sich der Fehler auf eine Verschiebung um bis zu 20 Pixel in vertikaler Richtung aus.
Das Problem kann durch eine Anpassung der Texturkoordinaten gelöst werden. Die detektierten Merkmale werden dazu durch die Darstellung auf der Einheitskugel in den 3D-Raum transformiert. Es folgt eine nichtlineare Optimierung, um eine Rotationsmatrix zu bestimmen, die die Merkmale auf den Einheitskugeln bestmöglich zur Deckung bringt. Die Rotationsma-

trix ist mit den Yaw-, Pitch- und Roll-Winkeln parametrisiert. Es wurde an dieser Stelle die Bibliothek *levmar* [Lourakis l] verwendet, die den Levenberg-Marquardt-Algorthmus zur nichtlinearen Optimierung bereit stellt.
Die Texturkoordinaten lassen sich nun auf der Einheitskugel im Raum durch Anwendung der Rotationsmatrix korrigieren.

3.2.2 Registrierung der hochauflösenden Bilder

Die Registrierung der hochauflösenden Bilder beginnt mit der Erstellung einer geeigneten Abbildung der 3D-Daten. Eine Bildregistrierung wird mit dem Abbild und dem hochauflösenden Bild durchgeführt. Die gefundenen 2D-2D-Korrespondenzen der Bilder werden im Anschluss in 3D-2D-Paarungen umgerechnet. Dadurch ist es möglich, eine Projektionsmatrix zu berechnen. Die Projektionsmatrix kann direkt zur Bestimmung der Texturkoordinaten verwendet werden. Zur Verbesserung der Stabilität der Berechnung der Projektionsmatrix wird eine Ausreißer-Filterung vor der Erstellung der 3D-2D-Korrespondenzen ausgeführt. In den folgenden Unterabschnitten wird das Vorgehen genauer beschrieben.

Bild der 3D-Daten

Die Registrierung eines Bilds mit 3D-Daten ist ohne weitere Einschränkungen schwierig. Die verwendeten SURF-Merkmale sind gegen projektive Verzerrung nur begrenzt invariant, so dass eine sphärische Projektion nicht verwendet werden kann. Das Abbild der 3D-Daten muss eine zum Bild vergleichbare Darstellung haben, damit die Bildregistrierung durchgeführt werden kann. An dieser Stelle kann die Funktionalität des Programms selbst genutzt werden, um eine orthographische Darstellung zu erzeugen. Für die Erstellung der Darstellung ist der Benutzer eingebunden, um einen groben Interessenbereich in den Daten auszuwählen, der für die Darstellung verwendet werden soll.
Die Position und Blickrichtung der Kamera, mit der die hochauflösenden Bilder aufgenommen wurden, sind unbekannt. Die Verwendung von orthographischen Darstellungen für die Bildregistrierung hat den Vorteil, dass die Darstellungen nicht von der Position der Kamera abhängen. Die unbekannte Blickrichtung wird senkrecht zur Felswand gewählt. Die hochauflösenden Bilder haben den Zweck, die Figuren auf der Felswand zu dokumentieren. Aus diesem Grund ist die Felswand relativ frontal aufgenommen worden. Die Wahl der Blickrichtung liefert daher vergleichbare Darstellungen zu den hochauflösenden Bildern.

3D-2D-Korrespondenzen

Nachdem ein Bild der 3D-Daten generiert ist, kann eine Bildregistrierung analog zum vorigen Fall durchgeführt werden. Die SURF-Merkmale werden aus beiden Bildern extrahiert. Die anschließende Paarung der Merkmale liefert 2D-2D-Korrespondenzen zwischen den Bildern. Abbildung 3.5 zeigt diese Bildregistrierung für ein hoch auflösendes Bild.
Das generierte Bild basiert auf den bekannten 3D-Daten. Jedes Pixel im Bild ist eine Abbildung der 3D-Daten. Eine Umkehrabbildung der Projektion ist nicht verfügbar. Die Zuordnung von Bildpunkt zu 3D-Punkt ist durch einen Vorverarbeitungsschritt möglich. Die Erstellung der orthographischen Darstellung ist eine Transformation der 3D-Punkte. Die Transformation ist bekannt und kann benutzt werden, um die 3D-Punkte außerhalb der Rendering-Pipeline

Abbildung 3.5: Das obere Bild zeigt eine orthographische Darstellung des gewählten Ausschnitts. Das untere Bild ist das zugehörige hoch auflösende Bild. Die gefundenen SURF-Merkmale sind in beiden Bildern als blaue Punkte eingezeichnet und mit ihrem korrespondierenden Partner verbunden.

zu projizieren. Die 3D-Punkte werden mit ihren projizierten Bildkoordinaten abgespeichert. Ein gefundenes Merkmal im Bild kann durch die Suche des nächsten Nachbarn an Hand der Bildkoordinate mit einem 3D-Punkt assoziiert werden. Auf diese Weise werden die 2D-2D-Korrespondenzen zu 3D-2D-Korrespondenzen erweitert.
Eine effiziente Möglichkeit der Abfrage ist durch die Verwendung eines *Quad-Tree* möglich. Ein Quad-Tree teilt einen festgelegten Bereich hierarchisch auf. Der Mittelpunkt eines Bereichs wird zur Aufspaltung in vier Unterbereiche verwendet. Diese Bereiche bilden die Kinder des Baumknotens. Der Baum hat eine festgelegte Tiefe. Die Blätter speichern die eingefügten Schlüssel-Wert-Paare.

Berechnung der Texturkoordinaten

Auf Grund der berechneten 3D-2D-Korrespondenz kann nun eine Projektionsmatrix P bestimmt werden. Die gefundenen Korrespondenzen sind durch die Projektionsmatrix als allgemeine Abbildung beschrieben. Für einen beliebigen Punkt $\hat{X}$ lässt sich die Bildkoordinate $\hat{x}$ durch $\hat{x} \cong P \cdot \hat{X}$ bestimmen. Es sind jedoch zwei Aspekte zu beachten:

1. Ein 3D-Punkt $\hat{X}$ ist auf dem Bild zu sehen, wenn die Bildkoordinate $\hat{x} = \lambda \cdot (u, v, 1)^T$ die Bedingungen $0 \leq u < w$ und $0 \leq v < h$ erfüllt. Die Werte w und h geben die Breite und die Höhe des Bilds an.

2. Ein 3D-Punkt muss vor der Kamera liegen, damit er auf dem Bild sichtbar ist. Jeder 3D-Punkt der auf der Geraden durch das Projektionszentrum und der Bildkoordinate liegt, wird bei der Projektion auf die Bildkoordinate abgebildet. Ein Punkt hinter der Kamera wird im mathematischen Sinne ebenfalls auf diese Bildkoordinate abgebildet. Die Projektion eines Punktes vor der Kamera und eines Punktes hinter der Kamera lässt sich am Vorzeichen von λ erkennen. Diese Unterscheidung reicht alleine nicht aus, da die Projektionsmatrix frei skaliert werden kann. Als einheitliche Bedingung für Punkte vor der Kamera ist

$$0 < \lambda \cdot det \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \tag{3.5}$$

   verwendet worden. Die Variablen r_{ij} sind die ersten drei Spalten von P.

Eine Umrechnung der absoluten Bildkoordinaten in relative Bildkoordinaten liefert die gesuchten Texturkoordinaten. Die relativen Bildkoordinaten drücken die Position im Bild unabhängig von der Bildgröße im Intervall $[0, 1] \times [0, 1]$ aus.

Ausreißer-Erkennung

Zur Verbesserung der Stabilität des Algorithmus zur Bestimmung der Projektionsmatrix wurde eine gezielte Suche nach Ausreißern durchgeführt. Die Suche findet bereits zwischen den 2D-2D-Korrespondenzen statt, bevor die Korrespondenzen weiterverarbeitet werden. Zur Bestimmung von Ausreißern wird die 3×3-Homographie-Matrix H zwischen den 2D-2D-Korrespondenzen berechnet. Die Homographie-Matrix beschreibt die projektive Transformation zwischen korrespondierenden 2D-Bildpunkten.

Die geometrische Interpretation erfolgt über die Projektionsmatrizen P_i und P_j, die die jeweiligen Bilder durch Projektionen erzeugt haben. Die Homographie-Matrix beschreibt die Rückprojektion eines Bildpunkts in den 3D-Raum durch die inverse Projektionsmatrix P_j^{-1} und der anschließenden Projektion durch die Matrix P_i. Die Projektionsmatrizen sind dabei nicht explizit bekannt und dienen an dieser Stelle der Erklärung.
Die berechnete Homographie-Matrix kann benutzt werden um 2D-2D-Korrespondenzen zu filtern, die nicht dieser Matrix entsprechen.

3.3 Lineare Korrektur

In Abschnitt 3.2.1 ist erläutert worden, dass durch eine verkippte Drehachse möglicherweise die Texturkoordinaten eine Abweichung aufweisen. Eine Diskussion über weitere Fehlerquellen von Abweichungen und deren Ausmaß ist in Kapitel 5 zu finden.
Im Arbeitsablauf lässt sich eine Korrektur der Texturkoordinaten vor dem Rendern realisieren. Die Korrektur erfolgt durch eine weitere Registrierung. Die Registrierung wird zwischen den Farbinformationen der Punktwolke und dem Ausschnitt der verwendeten Textur durchgeführt. Die Texturausschnitte sind im Vergleich zur Registrierung in Abschnitt 3.2 sehr viel kleiner. Der nichtlineare Fehler aus 3.2.1 lässt sich auf kleinen Bereichen gut mit einer linearen Korrektur beheben.
Ein grobes Referenzbild wird aus den Farbwerten und den Texturkoordinaten der 3D-Punkte erstellt. Die 3D-Punkte werden auf Grundlage der Texturkoordinaten in eine Ebene gelegt, trianguliert und die Farbwerte aus der Punktwolke interpoliert. Diese Vorgehen ist im Abb. 3.6 dargestellt.
In Abhängigkeit von der Anzahl der ausgewählten Punkte, bzw. von der Auflösung der Punktwolke werden unterschiedliche Methoden der Bildregistrierung verwendet. Im Fall eines gut aufgelösten Referenzbilds kann eine merkmalsbasierte Registrierung verwendet werden, um eine affine Transformation zwischen Referenzbild und verwendeter Textur zu finden.
Bei kleineren Ausschnitten sind nur grobe Farbbereiche im Referenzbild zu erkennen. Diesen Unterschied zeigt Abb. 3.6. In diesem Fall wird eine pixelbasierte Bildregistrierung durchgeführt. Die verwendete Metrik ist die normalisierte Korrelation.
Die pixelbasierte Lösung korrigiert eine Translation in x- und y- Richtung. Der Fehler durch eine Verschiebung ist im Allgemeinen am stärksten ausgeprägt und wird durch die Translation beseitigt.

Abbildung 3.6: Auf der linken Seite sind die Abbilder der 3D-Daten zu sehen. Die Punkte sind entsprechend der Texturkoordinaten mit Farbinformation der 3D-Daten gerendert. Eine nachträgliche Korrektur der Ausrichtung der Textur kann durch eine Bildregistrierung mit den Original-Texturen (rechts) erfolgen. Im oberen Fall ist eine merkmalsbasierte Registrierung noch möglich. Der untere Fall kann durch die Verwendung der pixelbasierten Registrierung gelöst werden.

Kapitel 4

Systementwurf

Die entwickelten Lösungsansätze aus dem letzten Kapitel sind in einem Programm realisiert worden. Das Programm hat zum einen den Zweck die Korrektheit der Lösungsansätze zu zeigen, sondern zum anderen ein brauchbares Werkzeug für die Anwender darstellen.
In diesem Kapitel wird der Entwurf des Programms vorgestellt. In Abschnitt 4.1 werden die ermittelten Anforderungen vorgestellt, auf denen der Entwurf des Programms basiert. In Abschnitt 4.2 wird der statische Entwurf vorgestellt, der die Konzepte umfasst, die zur Repräsentation der Daten im Programm verwendet werden. Der dynamische Entwurf (Kap. 4.3) beschreibt den Arbeitsablauf des Programms.

4.1 Anforderungen

Die Entwicklung des Programms erfolgte in Etappen in Zusammenarbeit mit den Projektbeteiligten von Wadi Sura. Bei gemeinsamen Treffen sind eine Menge an Anforderungen und Zielen festgelegt worden. Auf Grund der Realisierung der Anforderungen einer Etappe sind Änderungen, Verbesserungen oder neue Anforderungen entstanden.
Die ermittelten Anforderungen werden im Folgenden als kurze Zusammenfassung aufgelistet:

Unterstützung der Austauschformate: Die ersten Anforderungen betreffen die Unterstützung der Austausch-Formate für die Daten. Als Eingabe-Format der 3D-Daten ist das PTX-Format gewählt worden, weil sowohl Informationen über die Kamera- als auch Weltkoordinaten gespeichert sind. Eine Repräsentation der Punkte mit Grauwert-Intensität (x, y, z, I) oder zusätzlich mit Farbinformationen (x, y, z, I, r, g, b) wird unterstützt. Die Punktwolken sind vor der Übergabe bereits gefiltert und referenziert worden. Die Bilddaten zu jedem Standpunkt werden als sphärische Panoramen bereitgestellt. Die Panoramen bilden den Winkelbereich von $360° \times 180°$ ab. Bei den hochauflösenden Bildern sind keine Einschränkungen getroffen worden.

Gleichzeitige Ansicht von mehreren Standpunkten: Die Vermessung der Felswand ist aus verschiedenen Standpunkten erfolgt, die als verschiedene Datensätze vorliegen. Jeder Standpunkt bildet einen bestimmten Bereich der Felswand ab. Eine gleichzeitige Betrachtung von mehreren Standpunkten soll möglich sein. Die Verarbeitung erfolgt auf Grund der Daten des geeigneten Standpunkts.

Genauigkeit der Registrierung: Die Registrierung der Bilddaten mit der Punktwolke dient zur Erstellung von entzerrten Darstellung. Die Archäologen sind für kleine Bereiche der Felswand an einer orthographischen Darstellung interessiert. Für planare Ebenen entspricht diese Darstellung einer senkrechten Sicht auf die Felswand. Für diese Anwendung spielt die Genauigkeit der Anwendung nur eine untergeordnete Rolle, denn die Geometriedaten werden nur zum Berechnung einer ausgleichenden Ebene benutzt. Eine Abweichung von wenigen Pixel wäre in diesem Fall akzeptabel.

Kontrollmechanismus bei der Selektierung: Die entzerrten Darstellungen sind die Grundlage für die Archäologen zur weiteren Auswertung. Eine Vermessung auf Grund der entzerrten Darstellung und statistische Untersuchung der Größe von einzelnen Figuren ist eine mögliche Anwendung. Eine Vermessung in der orthographischen Darstellung setzt voraus, dass der betrachtete Bereich planar ist. Ein Kontrollmechanismus soll dieses Kriterium bei der Erstellung überprüfen. Mögliche Ausreißer in der Auswahl zur Ebene sollen ebenfalls entdeckt werden können, da sie die Lage einer ausgleichenden Ebene beeinflussen.

Selektierungsmöglichkeiten: Der gewählte Bereich der orthographischen Darstellungen hängt von den Figuren ab. Die Figuren haben unregelmäßige Formen. Eine Auswahl der Figuren soll mit einem Rechtecksbereich und einem beliebigen Polygon möglich sein.

Ansprechende Darstellung: Eine Anwendung ist die Erstellung eines Lageplans, der aus entzerrten Darstellungen von Figuren zusammengesetzt wird. Diese Anwendung fordert eine hoch aufgelöste und ansprechende Darstellung.

Orientierung am Raster: Die Felswand ist durch ein gedachtes Raster in $60cm \times 40cm$ große Bereiche unterteilt worden. Dieses Raster erstreckt sich entlang der xz-Ebene im festgelegten Koordinatensystem der 3D-Daten. Ein Bereich beinhaltet alle Punkte, deren x- und z-Koordinate in den Bereich fallen. Die Orientierung an diesem Rastern soll möglich sein.

Referenz im Ausgabebild: In den erstellten Darstellungen muss ebenfalls eine Information über die Orientierung und Lage enthalten sein, um die Darstellung auf der Felswand einordnen zu können. Ein Referenzpunkt, sowie ein Maßstab sind aus diesem Grund gefordert worden.

Benutzerfreundlichkeit: Die Benutzung des Programms soll möglichst einfach sein, damit ein Benutzer ohne Kenntnisse der Methoden der Informatik und der Vermessungstechnik mit dem Programm arbeiten kann.

4.2 Statischer Entwurf

Der statische Entwurf beschreibt die konzeptionellen Elemente, auf denen das Programm arbeitet. Das wichtigste Konzept ist als *Standpunkt* bezeichnet. Der Begriff und das Konzept werden in diesem Abschnitt definiert. Das Programm verwaltet eine Menge von Standpunkten, die als Projekt zusammengefasst werden können. Das Programm nutzt aber nur den geeignetsten Standpunkt zur Verarbeitung. Dem Programm liegt somit ein standpunktorientierter Entwurf zu Grunde.

Der Begriff *Standpunkt* ist abgeleitet aus der Organisation der 3D-Daten, die aus verschiedenen Standpunkten aufgenommen wurden. Der Begriff ist erweitert und verallgemeinert worden. Zusätzlich zu den 3D-Daten gehören die Bilddaten ebenfalls zu einem Standpunkt. Die Bilddaten können entweder die sphärischen Panoramen oder die hochauflösenden Fotos sein. Im letzten Fall bezeichnet der Standpunkt die Pose der Kamera, die das Foto aufgenommen hat. Die 3D-Daten sind auf die Punkte des sichtbaren Bereichs des Bilds reduziert. Mit dieser Verallgemeinerung ist keine Unterscheidung zwischen den verschiedenen Bilddaten mehr nötig. Die Verarbeitung kann für beide Fälle in gleicher Form durchgeführt werden.
Jeder Standpunkt wird in einer separaten Datei abgespeichert. Für die Speicherung ist ein eigenes Dateiformat eingeführt worden. Diese Entscheidung ergibt sich aus den folgenden Gründen:

- Die 3D-Daten sind im PTX-Format übergeben worden. Das Format speichert die 3D-Daten als Klartext ab. Die Punkte werden zeilenweise abgespeichert. Jede Zeile besteht aus den x-, y-, z-Koordinaten, der Intensität und optional aus den RGB-Werten eines Punktes. Der größte Nachteil besteht darin, dass bei jedem Laden die Punkte erneut von Ascii- in Binär-Darstellung umgewandelt werden müssen. Das eigene Dateiformat ist binär und liefert eine erhebliche Beschleunigung, da die Anzahl der Punkte in der Größenordnung von 10^8 bis 10^9 liegen.

- Die Beziehung zwischen 3D-Daten und den Bilddaten eines Standpunktes besteht durch die Erweiterung der Punkte mit Texturkoordinaten. Eine Texturkoordinate (u, v) eines Punkts beschreibt die zugehörige Position im Bild. Die Texturkoordinaten lassen sich durch die Registrierung berechnen. Die Berechnung ist zeitaufwändig, so dass die Texturkoordinaten mit Zuordnung zu den 3D-Daten abgespeichert werden müssen. Ein neues Format ist unumgänglich.

- Die Verwendung des vorhandenen VRML-Formats ist in Betracht gezogen worden. Das Format ist geeignet, um ein 3D-Modell mit Textur abzuspeichern. Der Nachteil besteht darin, dass dieses Format die Daten als Klartext abspeichert. Das Format eignet sich daher nur als Übergabeformat.

Eine Menge von Standpunkten wird in einem *Projekt* zusammengefasst. Ein Projekt speichert die Referenzen in Form der Dateinamen auf die Standpunkte, so dass ein Standpunkt alle Hauptinformationen beinhaltet.
Bei der Laserscan-Vermessung sind die Bereiche der Felswand von verschiedenen Standpunkten aus aufgenommen worden, damit keine Informationen durch verdeckte Stellen fehlen. Das Programm ermöglicht aus diesem Grund ebenfalls das Laden von mehreren Standpunkten, damit eine flächendeckende Darstellung der Felswand möglich ist.
Die Verarbeitung erfolgt aber mit den Daten eines einzelnen Standpunkts. Die Lösungsansätze des vorigen Kapitels basieren auf diesem Vorgehen, da einige Vereinfachungen möglich sind. Eine Vereinfachung betrifft die Triangulierung (Kap. 3.1.2) eines Ausschnitts der Punktwolke. Die Triangulierung für eine allgemeinere Punktmenge im Raum ist schwierig.
Ein weiterer Vorteil bei der Verwendung eines einzelnen Standpunkts ist die Vermeidung von zwei Texturen für einen Bereich. Eine Abweichung in der Registrierung oder ein Helligkeitsunterschied in den verwendeten Texturen würde zu Artefakten führen, die eine unästhetische Darstellung zur Folge hätten.
Die Verarbeitung der 3D-Daten im Programm besteht aus der Selektion eines Ausschnitts durch den Benutzer. Eine orthographische Projektion des Ausschnitts wird erstellt. Ein Problem ist die Wahl eines Standpunkts zur Verarbeitung. Es wird der Standpunkt gewählt, der

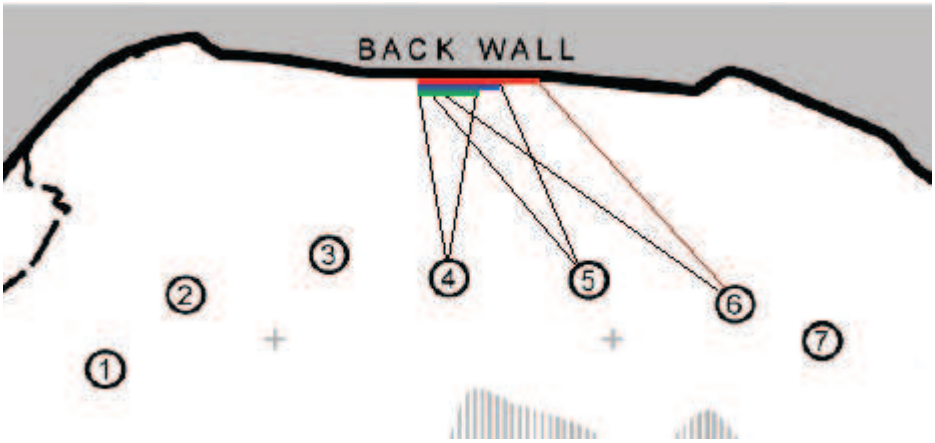

Abbildung 4.1: Die Skizze zeigt die 7 inneren Scan-Standpunkt entlang der Felswand. An Hand der Standpunkte 4, 5 und 6 wird das Kriterium zur Auswahl des aktiven Standpunkts verdeutlicht. Die Abtastpunkte liegen im gleichen Winkelabstand. Je näher ein Standpunkt am Interessenbereich liegt, desto mehr Abtastpunkte befinden sich im Interessenbereich. Bildquelle: [Kuper 09a]

die größte Punktanzahl im Bereich der Auswahl besitzt. Dieses Kriterium ist gewählt worden, weil der Scanner die Felswand in gleichen Winkelabständen abtastet. Ein Standpunkt besitzt die meisten Abtastpunkte des Bereichs, wenn er dem Bereich am nächsten liegt. Das zugehörige Panorama besitzt in diesem Fall auch die geringste Verzerrung. Das Kriterium ist in Abb. 4.1 illustriert.

4.3 Dynamischer Entwurf

In diesem Abschnitt wird der Arbeitsablauf im Programm vorgestellt, der in Abb. 4.2 schematisch dargestellt ist. Der Arbeitsablauf lässt sich in drei Prozesse aufspalten, die den Teilproblemen entsprechen, die durch die Lösungsansätze im vorigen Kapitel gelöst worden sind.

- Der erste Prozess ist das Registrieren der Panoramen. Als Eingabe stehen die 3D-Daten und das zugehörige Panorama zur Verfügung. Eine Bildregistrierung zwischen einer sphärischen Projektion der 3D-Daten und dem Panorama liefert die Texturkoordinaten der 3D-Punkte. Das Ergebnis ist eine Standpunkt-Datei, die die kompletten 3D-Daten mit Texturkoordinaten beinhaltet.

- Der zweite Prozess berechnet die Texturkoordinaten für die hochauflösenden Fotos. Als Eingabe wird ein Standpunkt benötigt, der den Bereich des zu registrierenden Fotos am besten zeigt. Das Ergebnis dieses Schrittes ist wieder ein Standpunkt, der als Bilddaten das hochauflösende Foto und die zugehörigen 3D-Punkte enthält.

- Der letzte Prozess ist die Verarbeitung zu einem orthographischen Bild. Der Benutzer wählt dazu einen Bereich in den Daten aus, der anschließend unter Verwendung der

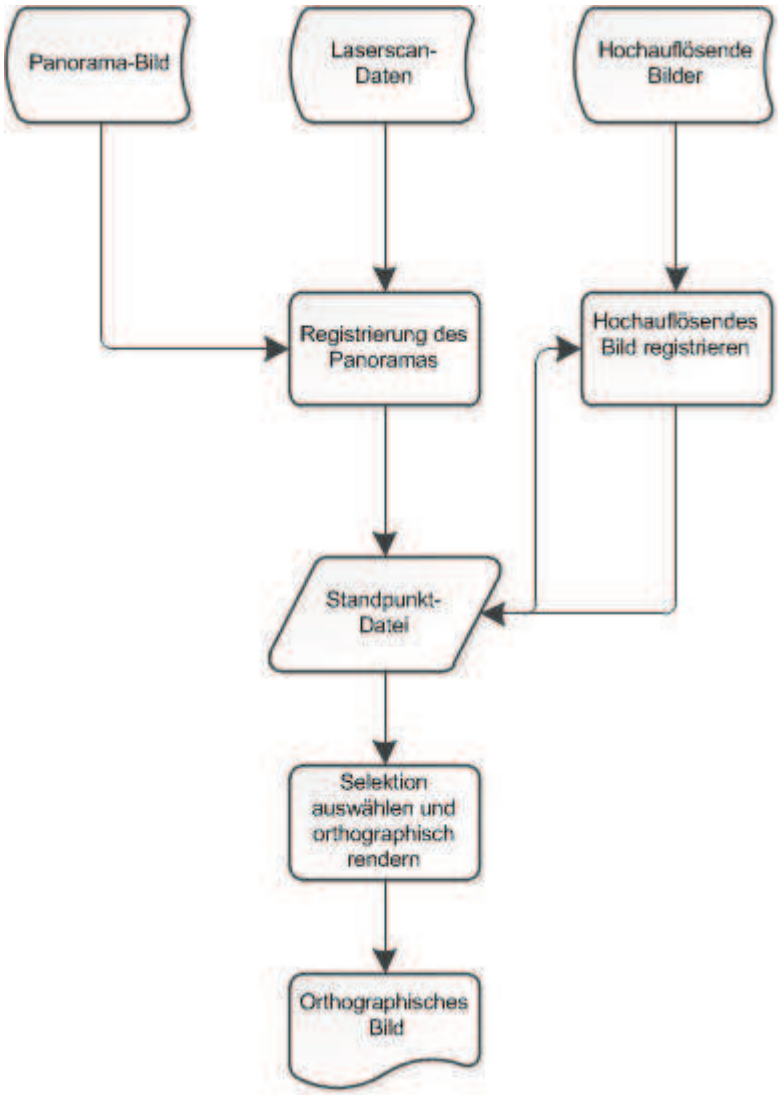

Abbildung 4.2: Der Arbeitsablauf zur Registrierung der Bilddaten ist in dieser Abbildung dargestellt. Das Ergebnis der Registrierung ist jeweils eine Standpunkt-Datei, die die Grundlage der weiteren Verarbeitung ist.

Textur zu einem Bild verarbeitet wird. Die Möglichkeiten der Auswahl eines Bereichs werden im nächsten Abschnitt vorgestellt.

4.3.1 Benutzer-Schnittstelle

In den vorigen Kapiteln ist bereits erwähnt, dass die orthographischen Darstellungen aus einer Auswahl der Daten erstellt werden. Die Registrierung der hochauflösenden Bilder setzt ebenfalls eine Auswahl eines Bereichs aus den Daten voraus, in dem das Bild registriert werden soll. Die Auswahl erfolgt durch den Benutzer. Eine Benutzer-Schnittstelle ist aus diesem Grund notwendig.

Die Auswahl eines Bereichs setzt eine geeignete Darstellung der Daten voraus. Die Darstellung erfolgt auf zwei unterschiedlichen Weisen, in denen die Auswahlen getroffen werden können.

- Die erste Darstellungsart ist die Ansicht der Punktwolken im 3D-Raum. Die Visualisierung von mehreren Punktwolken ist möglich. Die Transformation der unterschiedlichen Punktwolken in ein gemeinsames Koordinatensystem ist im Übergabeformat der 3D-Daten enthalten und resultiert aus dem Schritt der sog. Georeferenzierung, die durch die Laserscan-Software möglich ist.
  Die Navigation innerhalb des 3D-Raums ist durch das Konzept eines *Ankerpunkts* realisiert worden. Der Ankerpunkt ist eine ausgezeichnete Position im 3D-Raum. In Abb. 4.3 ist er als roter Punkt erkennbar. Die Szene kann durch den Benutzer um diesen Ankerpunkt gedreht werden. Ein Picking-Mechanismus ermöglicht es, den Ankerpunkt an einen beliebigen Punkt der Punktwolke zu setzen.

- Die zweite Darstellungsart ist die Ansicht der Projektion der Punktwolken auf die xz-Ebene. Das gemeinsame Koordinatensystem ist in der Art festgelegt worden, dass die y-Richtung zur Felswand zeigt. Die Archäologen haben auf Grundlage des gemeinsamen Koordinatensystems ein Raster entwickelt, um einzelne Bereiche benennen zu können. Das Raster liegt in der xz-Ebene und besteht aus Rechtecken der Größe $60cm \times 40cm$. Das Raster ist in die Projektion der Punktwolke eingezeichnet worden, damit eine Orientierung an Hand des Rasters möglich wird. In Abb. 4.3 ist diese Darstellungsart gezeigt.
  Die Punktwolken zeigen die komplette Umgebung inklusive des Sandbodens und der Objekten im Vordergrund der Umgebung. In der projizierten Darstellung der Punktwolke sollen diese Objekte die Felswand nicht verdecken. Das folgende Verfahren löst dieses Problem, ohne die 3D-Daten explizit verändern zu müssen.
  Die Projektion verwendet eine Modifikation des Z-Buffer-Algorithmus. Der Z-Buffer-Algorithmus hat den Zweck die Tiefeninformation der projizierten Bildpunkte zu verwalten. In der ursprünglichen Form werden in Projektionsrichtung weiter entfernte Punkte durch näher gelegene Punkte überschrieben. Diese Vorgehensweise wurde umgekehrt. Die weit entfernten Punkte der Felswand überschreiben die im Vordergrund liegenden Objekte.

Die Festlegung einer Auswahl durch den Benutzer erfolgt in beiden Darstellungsarten auf gleiche Weise. Die Festlegung erfolgt dabei in zwei Schritten.

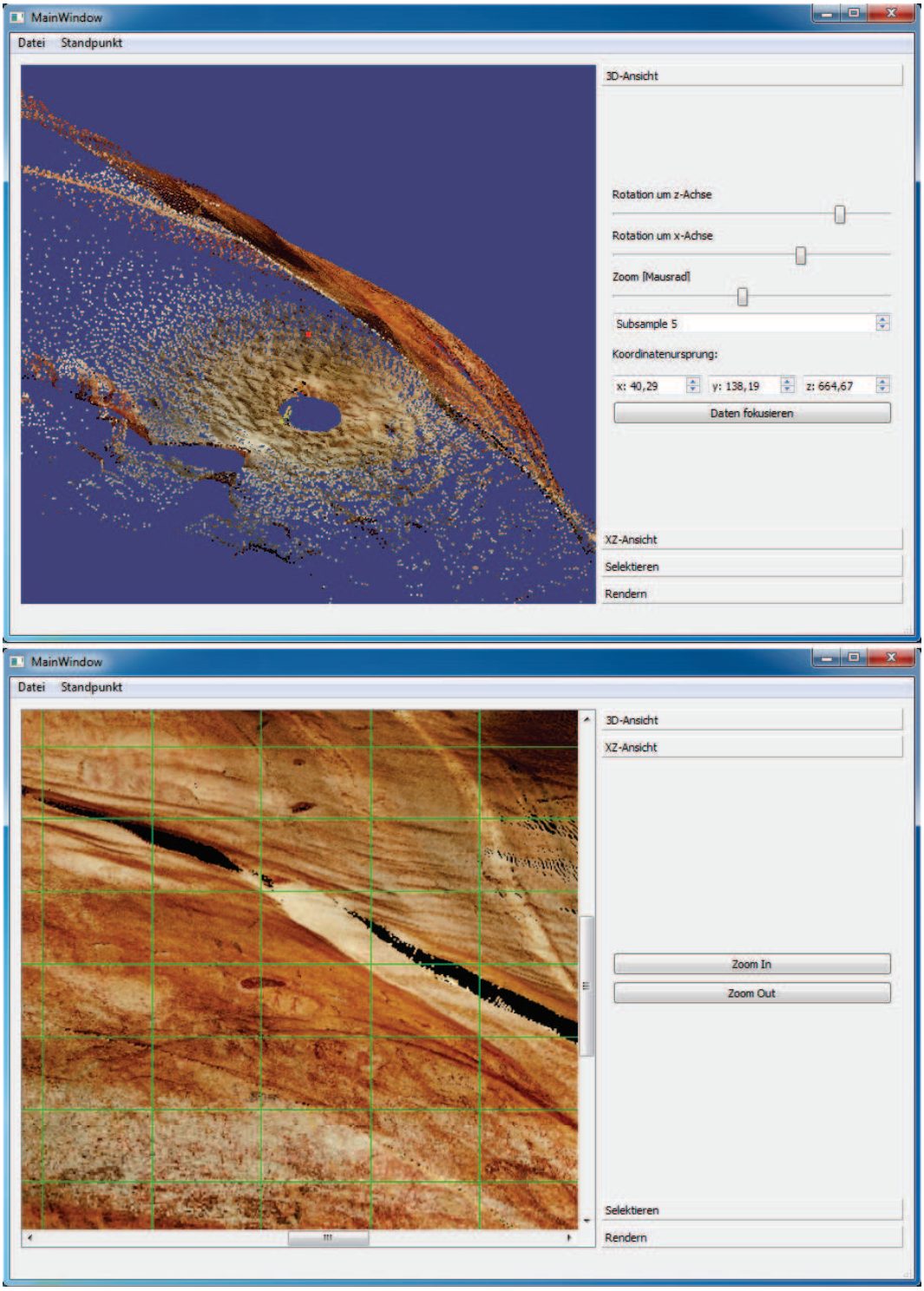

Abbildung 4.3: Das obere Bild zeigt die Darstellungsmöglichkeit als Punktwolke im 3D-Raum. Der rote Punkt bildet den Ankerpunkt, um den die Szene durch den Benutzer gedreht werden kann. Das untere Bild zeigt die Ansicht der Projektion auf die xz-Ebene. Das Raster ermöglicht eine Orientierung auf der Felswand.

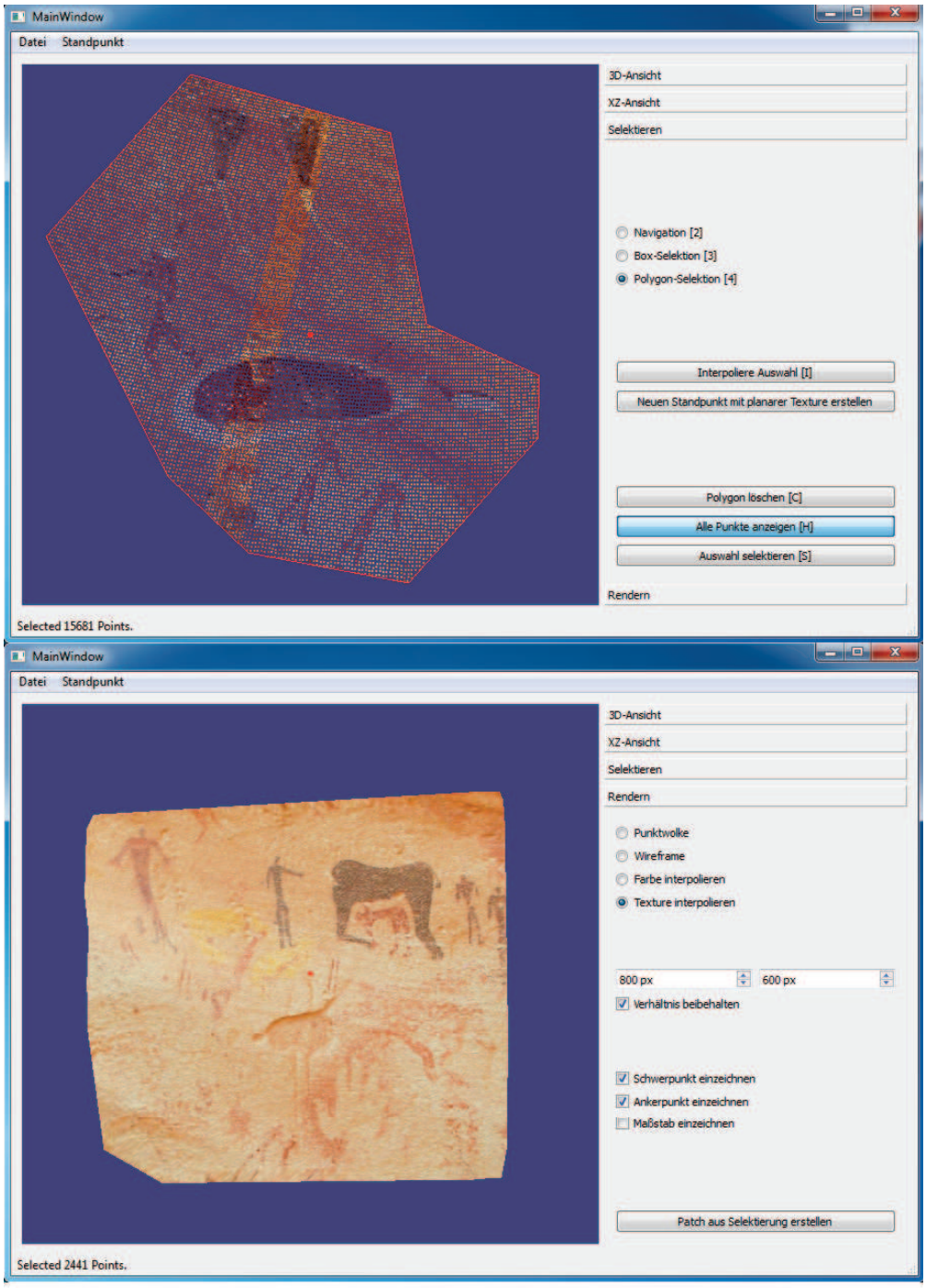

Abbildung 4.4: Das obere Bild zeigt die Möglichkeit der Polygon-Selektion im 3D-Raum durch Definition eines Polygons auf der Bildebene. Das untere Bild zeigt eine texturierte Punktmenge, die auf Grund der Auswahl eines Rasterbereichs in der xz-Ansicht gewählt worden ist.

1. Der Benutzer navigiert zum auszuwählenden Bereich, so dass der Bereich für den Benutzer sichtbar ist.
2. Der Benutzer zeichnet ein Rechteck oder Polygon auf die Bildebene. Die sichtbaren Punkte, die im Rechteck oder Polygon liegen, werden ausgewählt (siehe Abb. 4.4).

Die Bestimmung der Punkte, die durch das Rechteck oder Polygon festgelegt worden sind, erfolgt für die beiden Ansichten auf unterschiedliche Weise:

- Die Darstellungsart der Punktwolke im 3D-Raum zeigt dem Benutzer eine orthographische Projektion der Punktwolke. Die Position des Ankerpunkts und die Blickrichtung des Benutzers sind bekannt. Die Projektionsmatrix kann bestimmt werden. Bei der Auswahl müssen alle Punkte der Punktwolke projiziert werden, um überprüfen zu können, welche Punkte in das Rechteck oder Polygon fallen. Dieser Schritt ist zeitaufwändig.
- Die Bestimmung der Punkte in einem Polygon oder Rechteck, das in der xz-Ansicht festgelegt worden ist, lässt sich effizient durchführen. In einem Vorverarbeitungsschritt werden die 3D-Punkte entsprechend ihrer x- und z-Koordinate in einem Quad-Tree abgespeichert. Der Quad-Tree teilt einen Bereich am Mittelpunkt in vier Unterbereiche auf. Das Polygon wird durch ein Umrandungsrechteck vereinfacht, das das Polygon komplett umschließt. Der Quad-Tree muss somit lediglich ein Rechtecksabfrage unterstützen. Ein nachträglicher Schritt prüft, ob die Punkte, die im Umrandungsrechteck liegen, auch im Polygon liegen.
  Die Rechtecksabfrage im Quad-Tree kann Unterbereiche zur Abfrage ausschließen, wenn keine Ecke des Rechtecks in den Unterbereich fällt. Eine große Menge an Punkten muss nicht weiter überprüft werden.

Eine Auswahl in einer bereits ausgewählten Punktmenge wird unterstützt. Eine effiziente Arbeit ist möglich, wenn in der xz-Ansicht ein grobe Auswahl getroffen wird, die in der 3D-Ansicht verfeinert werden kann. Diese Arbeitsweise wird gefördert, in dem ein Bereich im Raster durch einen Doppelklick ausgewählt und in der 3D-Ansicht bereitgestellt wird. In Abb. 4.4 ist dieser Fall gezeigt.

Kapitel 5

Experimentelle Auswertung und Diskussion

In diesem Kapitel werden die Ergebnisse der Lösungsansätze gezeigt, ausgewertet und diskutiert. Im ersten Abschnitt folgt die Auswertung und Betrachtung der Fehlerquellen der Bildregistrierung. Es wird dazu eine Statistik über die Häufigkeit der Fehlpaarungen bei der Bildregistrierung vorgestellt sowie die Genauigkeit der daraus bestimmten Transformationsvorschrift ausgewertet. Im zweiten Abschnitt werden einige Beispiele an erstellten orthographischen Darstellungen analytisch betrachtet.

5.1 Auswertung der Bildregistrierung

Die möglichen Fehler können an verschiedenen Stellen im Verfahren der merkmalsbasierten Bildregistrierung auftreten. Die folgende Liste zeigt die Fehlerquellen und die Fortpflanzung der Fehler bei der Registrierung der Bilddaten:

- Die Grundlage für die merkmalsbasierte Bildregistrierung ist die Verwendung des SURF-Verfahrens. Die SURF-Merkmale sind der Ausgangspunkt des Fehlers, der sich in den weiteren Verfahren fortpflanzt. Kleine Darstellungsunterschiede auf Detailebene können bei der Detektion zu Abweichungen in der Lokalisierung einer Merkmalsposition führen. Die abweichenden Darstellungen sind durch perspektivische Effekte oder Beleuchtungsveränderungen begründet.
- Ein Partner-Merkmal muss an der korrespondierenden Stelle wieder detektiert werden, damit eine korrekte Paarung möglich wird. Beim Fehlen des Partners kann eine ungünstige Fehlpaarung auftreten, wenn ein Merkmal in direkter Nähe gefunden wird. Die Beschreibung dieses Merkmals ist sehr ähnlich zum fehlenden Partner. Es kann eine *ungenaue* Paarung entstehen. Das fehlende Auftreten eines Merkmals kann durch die Schwellwertfilterung der Maxima im Skalenraum auftreten.
- Die Verwendung des RANSAC-Verfahrens kann die ungenauen Paarungen nur bedingt verhindern, da das Verfahren einen Toleranzwert ε zur Erkennung von korrekten Paarungen verwendet. Dieser Wert kann nicht zu klein gewählt werden, da ansonsten eine

falsche Lösung auf Grund einer Ausreißer-Menge berechnet wird. Der tolerierte Fehler fließt in die Berechnung der Transformation ein.

- Die Registrierung ist mit verkleinerten Darstellungen der Panoramen durchgeführt. Die Extrapolation der Merkmalspunkte auf die Originalgröße der Panoramen überbrückt eine Skalierung um den Faktor 10. Der Fehler der Registrierung wird durch die Extrapolation ebenfalls um das 10fache vergrößert. Kleinste Abweichungen werden somit sichtbar.

- Die Verwendung der Nächsten-Nachbarn-Suche zur Bestimmung der 3D-2D-Korrespondenzen erzeugt einen weiteren Fehler. Ein Merkmal auf dem generierten Abbild der 3D-Daten muss einer 3D-Koordinate zuordnet werden, welche den Bildpunkt des Merkmals erzeugt hat. Die korrekte 3D-Koordinate ist nicht bekannt, da das Bild durch Interpolation der Farbwerte beim Rendern erzeugt worden ist.

Die Größe und das Ausmaß der Fehler werden in den nächsten Unterabschnitten für die Registrierung der Panoramen und der hochauflösenden Bilder genauer betrachtet.

5.1.1 Registrierung der sphärischen Panoramen

Die untersuchten Aspekte bei der Registrierung der Panoramen sind die Häufigkeit von Fehlpaarungen und die daraus folgende Genauigkeit der Transformationsvorschrift. Das Verhältnis zwischen korrekten und falschen Paarungen ist für die Arbeit mit RANSAC entscheidend, damit eine korrekte Lösung gefunden wird. Die Genauigkeit der Paarungen ist entscheidend für die Genauigkeit der berechneten Translation.
Die Panoramen haben eine sehr ähnliche Erscheinung. Die Auswertung ist für Standpunkt 2 und 3 durchgeführt worden, da zumindest im Bezug auf die Helligkeit ein Unterschied existiert (siehe Abb. 5.1). Die untersuchten Panoramen sind mit der Punktwolke mit Farbinformationen[1] (RGB) oder mit Intensitäten (I) des Laserscanners registriert worden.
Die *korrekte* Translation t_u zwischen Panorama und Abbild der 3D-Daten ist aus einer Menge von Korrespondenzen bestimmt worden, die per Hand auf Originalgröße des Panoramas ausgewählt worden sind. Die Translation ist auf $[0, 1]$ normiert, um die Skalierung der Bilder vernachlässigen zu können.

$$\begin{aligned} &\text{Standpunkt 2 (per Hand):} \quad t_u = 0.1652 \pm 0.0006 \\ &\text{Standpunkt 3 (per Hand):} \quad t_u = 0.6679 \pm 0.0011 \end{aligned}$$

Die bestimmte Translation t_u dient zur Messung der relativen Abweichung zwischen den Merkmals-Paaren, die durch das Verfahren der Bildregistrierung gepaart worden sind. Für jede Paarung der Merkmale muss im korrekten Fall diese Translation zwischen den Merkmalspositionen zu finden sein. Sei eine Paarung an Merkmalspositionen (u_i, v_i) und (u_j, v_j) gegeben, da ergibt sich die relative Abweichung zur korrekten Lösung durch:

$$du = |u_j - u_i - t_u| \tag{5.1}$$

[1] Die Registrierung mit SURF-Merkmalen setzt ein Grauwertbild als Eingabe voraus. Die Bezeichnung *Farbinformation* dient zur Abgrenzung zur *Intensität* des Laserscanners, die ebenfalls ein Grauwert darstellt.

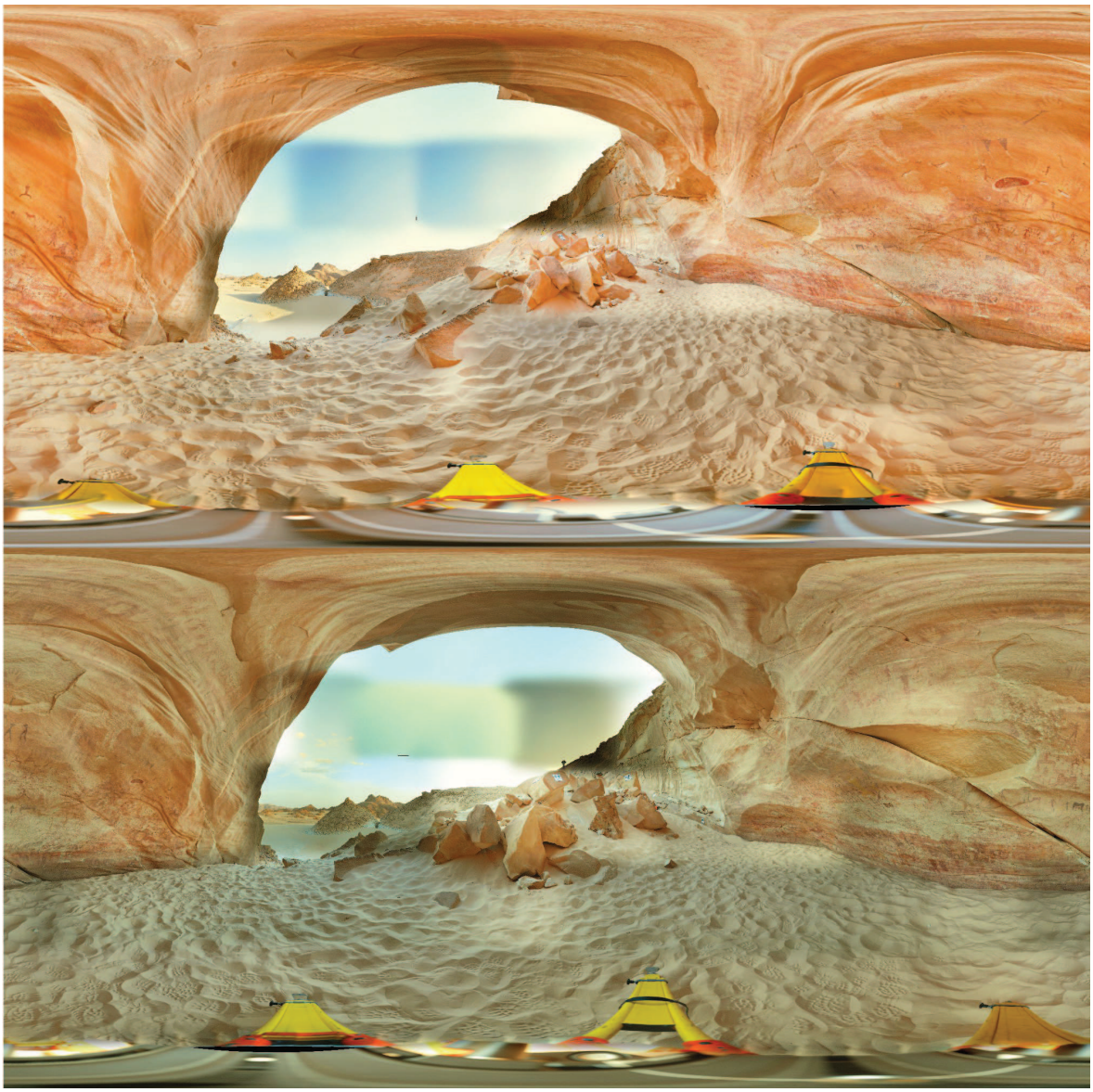

Abbildung 5.1: Diese Abbildung zeigt die Panoramen der Standpunkte 2 (oben) und 3 (unten), die zur Auswertung benutzt worden sind.

In der folgenden Tabelle ist das Verhältnis an Merkmals-Paaren eingetragen, die eine kleinere relative Abweichung als ε erfüllen.

	$\varepsilon = 0.001$	$\varepsilon = 0.01$	$\varepsilon = 0.1$	$\varepsilon = 0.2$
Standpunkt 2 (RGB)	320/946	764/946	897/946	920/946
Standpunkt 2 (I)	11/191	44/191	102/191	129/191
Standpunkt 3 (RGB)	75/634	493/634	586/634	602/634
Standpunkt 3 (I)	1/127	16/127	60/127	79/127

In dieser Tabelle fällt ein markanter Unterschied zwischen der Registrierung mit Farbinformationen und Intensitäten auf. Bei der Registrierung mit Farbinformationen haben mehr als die Hälfte der gefundenen Korrespondenzen eine Abweichung von weniger als 1%. Die Registrierung mit den Intensitäten liefert für diese Fehlerklasse ein Verhältnis von 23%, bzw. 12%. Der Grund für das schlechte Ergebnis wurde in Kapitel 2 bereits angesprochen. Die Mermals-Deskriptoren speichern die lokale Umgebung auf Grund der Intensitätsänderungen ab. Die Intensität und die Farbinformation haben im Allgemeinen einen unbekannten nichtlinearen Zusammenhang. In gleichen Bildausschnitten kann die Intensitätsänderung in den Farbinformationen und den Laser-Intensitäten abweichen. Dies führt zu falschen Paarungen auf Grund der unterschiedlichen Beschreibungen der Merkmale.
Die Translation ist mit den gleichen Korrespondenzen durch die robuste Methode RANSAC berechnet worden. Als Ausreißer sind Merkmals-Paare mit einer relativen Abweichung von $\geq \varepsilon = 0.03$ gewertet worden.

Standpunkt 2	(RGB)	$t_u = 0,1669$
Standpunkt 2	(I)	$t_u = 0,1687$
Standpunkt 3	(RGB)	$t_u = 0,6665$
Standpunkt 3	(I)	$t_u = 0,6631$

Die Abweichung zum Ergebnis der korrekten Lösung liegt für die Registrierung mit Farbinformationen im Bereich von 0.1-0.2%. Die Registrierung mit den Intensitätswerten erreicht eine Abweichung von maximal 0.5%. Das Ergebnis der Registrierung mit Intensitätswerten ist dabei kein reiner Zufall, sondern beruht auf einem spezifischen Phänomen der Daten.
Das angesprochene Phänomen bezieht sich auf den Übergang zwischen Felswand und freiem Gelände in den Panoramen. Die Intensitätsänderung ist für diesen Übergang oft eindeutig wiederzufinden. Die Merkmals-Paarungen am Übergang bilden eine Grundmenge für die Lösung.
Das RANSAC-Verfahren kann bei zu vielen Ausreißern eine falsche Lösung finden, wenn eine Ausreißer-Gruppe eine mögliche Lösung bildet und genügend weitere Ausreißer diese Lösung unterstützen. Die gefundene Grundmenge an Korrespondenzen verhindert diesen Fall. Die Fehlpaarungen sind zufällig verteilt und gefährden die Lösung nicht.
Der vorgestellte Fehler wird durch die anschließende Korrektur der Drehachse weiter minimiert. Im Fall der farbigen Punktwolke kann eine Pixel-genaue Registrierung erreicht werden. Im Fall der Punktwolke mit Intensitätswerten schlägt die Korrektur der Drehachse wegen der hohen Anzahl von Fehlpaarungen meist fehl. Es bleibt eine Abweichung von einigen Pixeln, die durch die lineare Korrektur vor der Erstellung der orthographischen Darstellungen beseitigt werden kann. Ein Beleg für diese Aussage ist in Abschnitt 5.2 zu finden.

Abbildung 5.2: Die Abbildung zeigt die drei Fotos, die zur Auswertung benutzt worden sind. Das obere linke wird als *Bild 1*, das untere als *Bild 2* und das rechte Foto als *Bild 3* bezeichnet.

5.1.2 Registrierung der hochauflösenden Bilder

Die Registrierung der hochauflösenden Bilder basiert wie die Registrierung der Panoramen auf der merkmalsbasierten Bildregistrierung. Im Unterschied zur Registrierung der Panoramen spielen die Intensitäten des Laserscanners keine Rolle, da die Punktwolken mit bereits registrierten Panoramen als Grundlage verwendet werden. Die Merkmals-Paarung liefert ein gutes Ergebnis, da die verwendeten Darstellungen Farbinformationen besitzen.
Eine Ausreißer-Erkennung filtert die Merkmals-Paarungen nach der Bildregistrierung durch eine berechnete Homographie zwischen den Bildern. Die verbleibenden Merkmals-Paare besitzen eine maximale Abweichung von 10 Pixel. Die verbleibende Abweichung der Merkmals-Paare und die Auswirkung auf die berechnete Projektionsmatrix wird in diesem Abschnitt betrachtet.
Die verwendeten Fotos zur Auswertung sind in Abb. 5.2 mit Zuordnung der gebrauchten Bezeichnung zu sehen. Bild 1 und 3 sind Überblicksfotos mit einer Auflösung von 4000×6500 Pixel. Das Bild 2 ist ein Detail-Foto. Die Auflösung beträgt 4000×4000 Pixel.

Für jedes Foto ist eine Registrierung durchgeführt worden. Die Resultate sind die bestimmte Projektionsmatrix P und die 3D-2D-Korrespondenzen (X_i, x_i). Die Auswertung untersucht

den Rückprojektionsfehler, der für eine Projektionsmatrix P bestimmt, mit welcher Abweichung die 3D-2D-Korrespondenzen erfüllt werden.

$$E_i = \|x_i - P \cdot X_i\|^2 \tag{5.2}$$

Die folgende Tabelle zeigt für jede Spalte das Verhältnis in den Fehlerklassen, in die Korrespondenzen mit einem Fehler $\leq \varepsilon$ fallen.

	$\varepsilon = 2px$	$\varepsilon = 3px$	$\varepsilon = 5px$	$\varepsilon = 10px$
Bild 1	12/125	30/125	78/125	122/125
Bild 2	2/62	4/62	8/62	23/62
Bild 3	5/31	7/31	12/31	28/31

In der Tabelle ist zu beobachten, dass die Gesamtanzahl der Merkmals-Paarungen abnimmt. In Bild 1 ist eine große Menge von Figuren enthalten. Die Figuren sind ebenfalls klar zu erkennen. Die Merkmals-Paarung wird durch diese Faktoren begünstigt.
In Bild 2 ist nur ein Ausschnitt aus Bild 1 dargestellt. Die Anzahl der möglichen Paarungen verringert sich dadurch. Die höhere Auflösung des Ausschnitts ermöglicht die Detektion von weiteren Merkmalen. Diese Merkmale beschreiben aber Unregelmäßigkeiten der Felsoberfläche. Wegen der Uneindeutigkeit tragen diese Merkmale selten zu weiteren Paarungen bei.
Die geringe Anzahl der Paarungen in Bild 3 besitzt andere Gründe. Das Bild zeigt an der linken Seite eine helle Stelle, die durch die künstliche Beleuchtung erzeugt wurde. Auf der linken Seite ist ebenfalls ein Schattenwurf zu erkennen. Diese Beleuchtungseffekte sind in den Panoramen nicht vorhanden. Die Paarung wird dadurch erschwert.
Ein weiteres Problem entsteht durch den sichtbaren Riss der Felswand in der Mitte. Aus verschiedenen Perspektiven ergeben sich sehr unterschiedliche Darstellungen, die die Merkmals-Paarung beeinflussen.

Die Genauigkeit der Merkmals-Paarungen zeigt sich in den unterteilten Fehlerklassen. Die verwendeten 3D-2D-Korrespondenzen weichen deutlich von einer exakten Lösung ab. Der Grund liegt in der vereinfachten Berechnung der 3D-2D-Korrespondenzen. Für einen Bildpunkt wird der nächste Nachbar aus der projizierten Punktwolke gewählt. Bei großen Fotos ist die 3D-Punktdichte sehr gering und es treten große Abweichungen auf.
Die Qualität der Registrierung ist ausreichend, um orthographische Projektionen zu erstellen. Die lineare Korrektur ist in der Lage diesen Fehler auszugleichen. Eine Reduktion des Fehlers ist durch die Verwendung einer Punktwolke mit höherer Punktdichte möglich und steht für weitere Anwendungen gegebenenfalls zu Verfügung.

5.2 Analyse der orthographischen Darstellungen

An dieser Stelle werden die Ergebnisse der orthographischen Projektionen vorgestellt. Zur Überprüfung der Qualität wurden eine Darstellung mit Textur und eine Darstellung auf Grund der Farbinformation der Punktwolke erstellt. Beide Abbildungen sind zum Vergleich übereinandergelegt. Die Übereinanderlegung ist im Schachbrettmuster-Stil durchgeführt. Das Bild wird in kleine Felder zerteilt, die gemäß eines Schachbrettes abwechselnd die unterschiedlichen Bilder zeigen. Diese Übereinanderlegung zeigt gut kleine Abweichungen in der Registrierung an den Übergängen zwischen zwei Feldern. Die Ergebnisse der Registrierung der

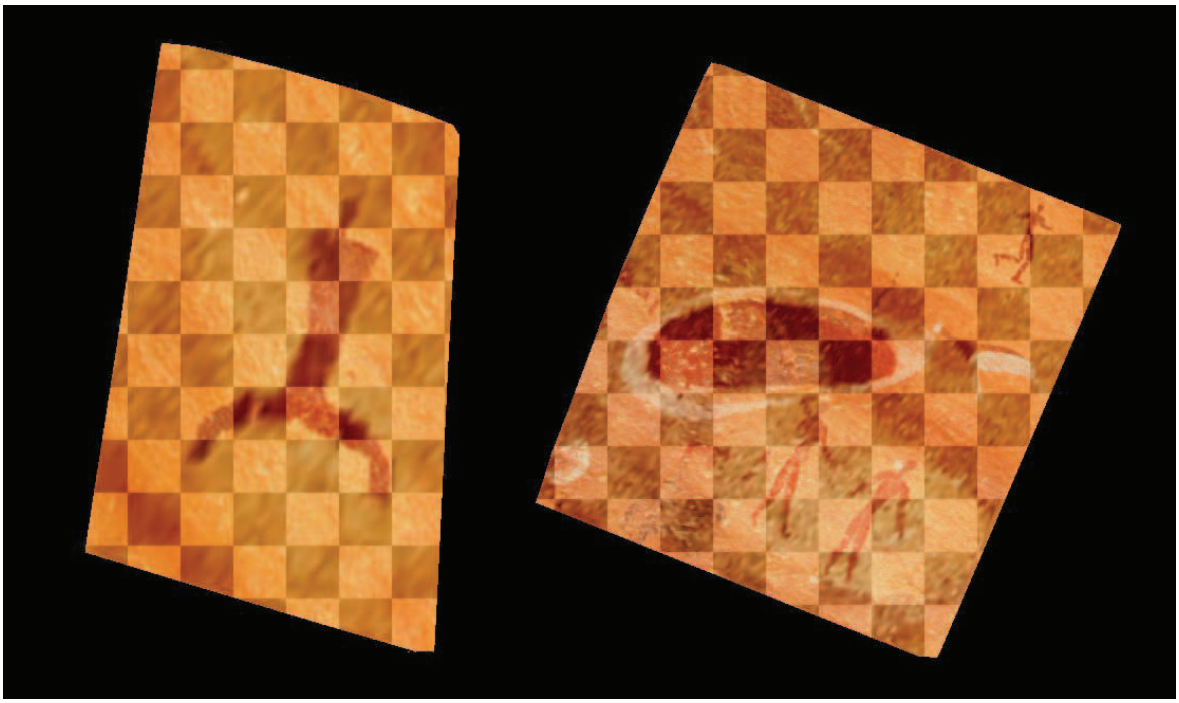

Abbildung 5.3: Diese Abbildung zeigt zwei erstellte orthographische Darstellungen aus Standpunkt 2. Die Grundlage der Registrierung war die farbige Punktwolke.

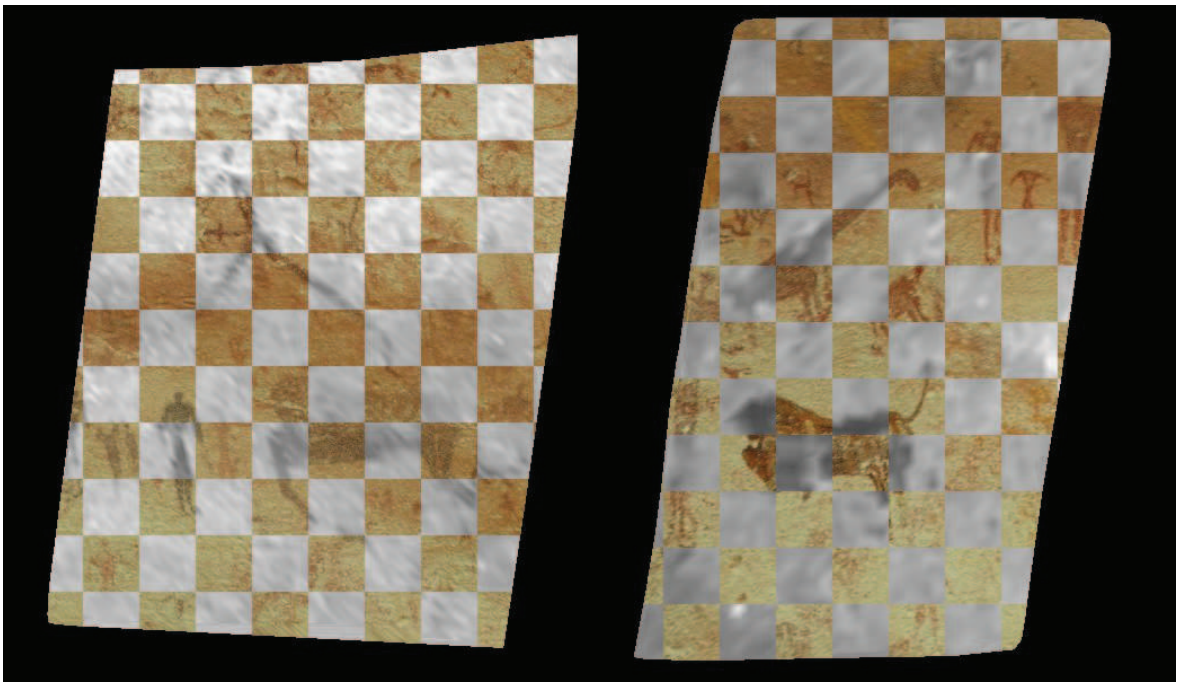

Abbildung 5.4: Diese Abbildung zeigt zwei erstellte orthographische Darstellungen aus Standpunkt 3. Die Grundlage der Registrierung war die Punktwolke mit Intensitätswerten.

Abbildung 5.5: Diese Abbildung zeigt zwei erstellte orthographische Darstellungen aus dem hochauflösenden Bild 1 und 3.

Panoramen mit Farbinformation sind in Abb. 5.3, die der Panoramen mit Intensitätswerten in Abb. 5.4 gezeigt. Die Ergebnisse der Registrierung der hochauflösenden Fotos sind in Abb. 5.5 dargestellt.

In der folgenden Tabelle sind die Einflüsse der linearen Korrektur bei der Erstellung der einzelnen Fotos aufgelistet. Die Verschiebung in horizontale Richtung wird als *du*, die vertikale Verschiebung als *dv* bezeichnet. Die gemessene Einheit ist Pixel.

Abbildung	du	dv
Abb. 5.3 (links)	0	0
Abb. 5.3 (rechts)	0	0
Abb. 5.4 (links)	-1	-4
Abb. 5.4 (rechts)	0	5
Abb. 5.5 (links)	1	5
Abb. 5.5 (rechts)	-1	0

In den Werten dieser Tabelle ist ersichtlich, dass die Registrierung der Panoramen mit den farbigen Punktwolken ein wesentlich genaueres Ergebnis liefert als die Registrierung mit den Intensitätswerten. Eine lineare Korrektur ist bei den farbigen Punktwolken nicht notwendig. Die Korrektur der Drehachse setzt genaue Positionen der Merkmale voraus. Bei der Registrierung mit Intensitätswerten schlägt sich dieser Umstand negativ aus. In der Tabelle sind die Verschiebungen in der y-Richtung dominant und zeigen, dass die Drehachse nicht korrigiert werden konnte.

Kapitel 6

Abschlussbetrachtung

Die Sicht der letzten Kapitel hat sich auf die Problemstellung und deren Lösung beschränkt. Die Ansätze und die dadurch erreichten Ergebnisse sind vorgestellt worden. In diesem Kapitel wird die Arbeit in einen globalen Kontext gestellt. Die Anwendbarkeit und die Grenzen der Ansätze werden diskutiert. Verschiedene Ausblicke und mögliche Erweiterungen für weitere Anwendungen werden betrachtet.

6.1 Anwendbarkeit und Grenzen

Die Entwicklung der Lösungsansätze ist in erster Linie mit dem Ziel verfolgt worden, die Problemstellung zu lösen. Die gegebenen Daten haben eine spezielle Charakteristik und Eigenschaften, die bei der Entwicklung berücksichtigt und ausgenutzt worden sind. Die Charakteristik und die Eigenschaften werden im Blick auf eine allgemeinere Anwendbarkeit betrachtet.
Das Problem der Registrierung der sphärischen Panoramen besteht im Allgemeinen aus der Bestimmung einer unbekannten Translation. Für die vorliegenden Daten funktioniert der gefundene Lösungsansatz, eine sphärische Darstellung der Punktwolke mit den Panoramen zu registrieren. In der experimentellen Auswertung hat sich aber gezeigt, dass die Registrierung auf Grund der Intensitätswerte der Punktwolke lediglich wegen einer speziellen Eigenschaft der Daten möglich war, nämlich wegen des eindeutigen Übergangs zwischen Felswand und der helleren Umgebung.
Im Allgemeinen wird die Registrierung auf Grund von Intensitätswerten nur möglich sein, wenn das Intensitätsbild eine ähnliche Darstellung zu den Grauwerten der Panoramen zeigt. Die verwendeten SURF-Merkmale sind für die Registrierung zwischen Laser-Intensitäten und Farbwerten nicht geeignet, da die Merkmale durch die Intensitätsverläufe beschrieben werden und keine lineare Abhängigkeit zwischen diesen Werten besteht.
Eine Registrierung auf Grund einer kolorierten Punktwolke zeigt in der experimentellen Auswertung gute Ergebnisse. Eine mögliche Anwendung könnte die automatische Texturierung einer kolorierten Punktwolke sein.
Das Problem der Registrierung der hochauflösenden Bilder besitzt eine Verallgemeinerung, nämlich die Registrierung eines beliebigen Bilds in einer 3D-Szene. Der Lösungsansatz nutzt aus, dass die vorliegenden 3D-Daten im Groben eine Oberfläche beschreiben. Das allgemeine

Problem reduziert sich auf die Registrierung eines Bilds auf einer Oberfläche.
Der Lösungsansatz setzt eine bereits texturierte Punktwolke voraus. Eine kolorierte Punktwolke kann als Eingabe genügen, solange die Auflösung der Punktwolke zur Erstellung eines Abbilds ausreicht. Der Benutzer wählt einen Bereich auf der Oberfläche zur Registrierung aus. Ein Vorschlag, diese Interaktion zu vermeiden, wird im nächsten Abschnitt gegeben. Die Registrierung zeigt unter diesen Voraussetzungen akzeptable Ergebnisse.
Die Vereinfachung des allgemeinen Problems gibt die Möglichkeit, eine Blickrichtung senkrecht zur Oberfläche zu wählen. Dadurch wird es möglich, ein Referenzbild der 3D-Daten zu erstellen und eine Bildregistrierung durchzuführen. Die verwendeten SURF-Merkmale zeigen gute Ergebnisse, auch wenn die Blickrichtung von der festgelegten Wahl abweicht. Eine Wahl der Blickrichtung im allgemeinen Problem ist schwierig und schränkt die Anwendbarkeit ein.

6.2 Ausblicke und Erweiterungsmöglichkeiten

In diesem Abschnitt wird abschließend betrachtet, inwieweit die Lösungsansätze erweitert und verbessert werden können.
Das entwickelte System liefert als Ausgabe ein Bild der orthographischen Projektion der registrierten Daten. Die Zwischenergebnisse der Registrierungen könnten auch für eine alternative Weiterverarbeitung von Interesse sein.
Die erkannten Fehlerquellen (Kap. 5) können durch Erweiterungen teilweise beseitigt werden, so dass die lineare Korrektur vor dem Rendern entfallen kann. Die Ersetzung der Nächsten-Nachbarn-Suche für 3D-2D-Korrespondenzen durch eine Interpolation kann beispielsweise eine Verbesserung der Genauigkeit der Registrierung liefern, ohne dass die Auflösung der Punktwolke erhöht werden muss. Eine Anwendung könnte die Erstellung eines Modells mit einer sehr hochauflösenden Textur sein.
Ein gutes Ergebnis zeigt die Registrierung der Panoramen mit den kolorierten Punktwolken, so dass die texturierten Punktwolken für weitere Anwendungen zur Verfügung stehen. Das System kann durch eine geeignete Exportierung erweitert werden. Die Punktwolken mit Intensitätswerten des Laserscanners sind für das Projekt weniger relevant, da kolorierte Punktwolken bereits vorliegen.
Die automatische Registrierung der hochauflösenden Bilder wäre eine interessante Erweiterung, da dadurch die Benutzer-Interaktion zur Auswahl des Bildbereichs entfallen würde. Eine Menge von Referenzbildern könnte dazu von der Oberfläche erstellt werden. Eine Registrierung mit der Menge an Referenzbildern setzt jedoch eine Heuristik zur Entscheidung voraus, ob gleiche Ausschnitte zu sehen sind. Eine weitere Aufbereitung der Daten wäre zusätzlich notwendig, um die Punkte der Oberfläche von den Punkten des Bodens zu trennen.

Anhang A

Verwendete Software

Bei der Umsetzung dieser Arbeit wurde die Funktionalität von vorhandenen Implementierungen genutzt. Die einsetzten Bibliotheken sind mit Verwendungszweck in der folgenden Liste vermerkt. Zur Erstellung dieses Dokuments wurde ein Template [Gockel 08] verwendet.

1. Die Bibliothek OpenCV [Bradski 00] ist genutzt worden. Die Funktionalität der Merkmalserkennung, des Template-Matchings und die grundlegenden Funktionen zum Arbeiten mit Bildern und Matrizen sind verwendet worden.
2. Die Verarbeitung durch die Rendering-Pipeline erfolgt durch die OpenGL-Bibliothek. Die Rendering-Pipeline übernimmt das Erstellen der Darstellungen aus dem texturierten Mesh.
3. Die Bibliothek Levmar [Lourakis l] implementiert das Levenberg-Marquardt-Verfahren zur nichtlinearen Minimierung. Die Minimierung ist bei der Berechnung der Projektionsmatrix und der Korrektur der Drehachse zum Einsatz gekommen.
4. Zur Erstellung der Benutzeroberfläche ist Qt-Bibliothek eingesetzt worden.

Literaturverzeichnis

[Bay 06] Herbert Bay, Tinne Tuytelaars, L. Van Gool. SURF: Speeded Up Robust Features. Tagungsband: 9th European Conference on Computer Vision, Graz Austria, Mai 2006.

[Bradski 00] G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.

[Crow 84] Franklin C. Crow. Summed-area tables for texture mapping. Tagungsband: Computer Graphics (SIGGRAPH '84 Proceedings), Band 18, Seiten 207–212, 1984.

[Delponte 06] Elisabetta Delponte, Francesco Isgrò, Francesca Odone, Alessandro Verri. SVD-matching using SIFT features. Graph. Models, 68(5):415–431, Sept. 2006.

[Fischler 81] Martin A. Fischler, Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6):381–395, 1981.

[Gockel 08] Gockel. Form der wissenschaftlichen ausarbeitung. <http://wwwiaim.ira.uka.de/form-der-wissenschaftlichen-ausarbeitung>.

[Hartley 04] R. I. Hartley, A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518, zweite Auflage, 2004.

[Kuper 09a] Rudolph Kuper, Hans Leisen, Heiko Riemer, Frank Förster, Sabine Krause, Erik Büttner, Frank Darius, Karin Kindermann. Wadi sura field report season 2009-2. Website. Online verfügbar unter http://www.wadisura.phil-fak.uni-koeln.de/uploads/media/WADI_SURA_report_2_2009-2.pdf.

[Kuper 09b] Rudolph Kuper, Heiko Riemer, Frank Förster. Wadi sura field report season 2009-1. Website. Online verfügbar unter http://www.wadisura.phil-fak.uni-koeln.de/uploads/media/WADI_SURA_report_1_2009-1.pdf.

[Kuper 10] Rudolph Kuper, Hans Leisen, Heiko Riemer, Frank Förster, Sabine Krause, Jürgen Seidel. Wadi sura field report season 2010. Website. Online verfügbar unter http://www.wadisura.phil-fak.uni-koeln.de/uploads/media/WADI_SURA_report_3_2010-1.pdf.

[Levenberg 44] Kenneth Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. The Quarterly of Applied Mathematics, 2:164–168, 1944.

[Lourakis l] M.I.A. Lourakis. levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++. [web page] http://www.ics.forth.gr/~lourakis/levmar/. [Accessed on 31 Jan. 2005.].

[Lowe 04] David G. Lowe. Distinctive image features from Scale-Invariant keypoints. International Journal of Computer Vision, 60(2):91–110, Nov. 2004.

[Muja 09] Marius Muja, David G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. Tagungsband: In VISAPP International Conference on Computer Vision Theory and Applications, Seiten 331–340, 2009.

[Nister 03] D. Nister. Preemptive ransac for live structure and motion estimation. Ninth IEEE International Conference on, Seiten 199–206 vol.1, 2003.

[Nolting 99] Wolfgang Nolting. Klassische Mechanik, Band 1 der Reihe Grundkurs Theoretische Physik. Verlag Vieweg, Braunschweig/Wiesbaden, fünfte Auflage, 1999.

[Szeliski 06] Richard Szeliski. Image alignment and stitching: a tutorial. Found. Trends. Comput. Graph. Vis., 2:1–104, January 2006.

[Triggs 00] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, Andrew W. Fitzgibbon. Bundle adjustment - a modern synthesis. Tagungsband: Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, ICCV '99, Seiten 298–372, London, UK, 2000. Springer-Verlag.