

Ludwig Hoegner

**Automatische Texturierung von Fassaden
aus terrestrischen Infrarot-Bildsequenzen**

München 2014

**Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck**



Automatische Texturierung von Fassaden aus terrestrischen Infrarot-Bildsequenzen

Vollständiger Abdruck
der von der Ingenieur fakultät Bau Geo Umwelt
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigten Dissertation

von

Dipl.-Inf. Ludwig Hoegner

München 2014

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck

Adresse der Deutschen Geodätischen Kommission:



Deutsche Geodätische Kommission

Alfons-Goppel-Straße 11 • D – 80 539 München

Telefon +49 – 89 – 23 031 1113 • Telefax +49 – 89 – 23 031 -1283 / - 1100

e-mail hornik@dgfi.badw.de • <http://www.dgk.badw.de>

Prüfungskommission

Vorsitzender: Univ.-Prof. Dr.-Ing. habil. Christian Große

Prüfer der Dissertation: 1. Univ.-Prof. Dr.-Ing. Uwe Stilla
2. Univ.-Prof. Dr.-Ing. habil. Helmut Mayer,
Universität der Bundeswehr München

Die Dissertation wurde am 07.11.2013 bei der Technischen Universität München eingereicht
und durch die Ingenieur fakultät Bau Geo Umwelt am 19.05.2014 angenommen.

Diese Dissertation ist auf dem Server der Deutschen Geodätischen Kommission unter <<http://dgk.badw.de/>>
sowie auf dem Server der Technischen Universität München unter
<<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20140528-1178212-0-3> > elektronisch publiziert

© 2014 Deutsche Geodätische Kommission, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen.

Kurzfassung

Thermische Inspektionen von Gebäuden mit Infrarot(IR)-Kameras erfolgen traditionell in der Bildgeometrie. Eine 3D Georeferenzierung oder objektbezogene Speicherung der abgebildeten Objektoberflächen erfolgt nicht. Die Arbeit beschäftigt sich mit der Herstellung des dreidimensionalen Raumbezugs und der objektbezogenen Speicherung der abgebildeten Fassadenflächen unter Verwendung von 3D Gebäudemodellen. Gebäudemodelle werden heutzutage sowohl für die Erfassung des Bestands in standardisierter Form (CityGML) in Geoinformationssystemen als auch für Planung, Bau und Betrieb in Bauwerksinformationssystemen (BIM) verwendet. Durch die Herstellung eines 3D Raumbezugs wird es möglich, komplexe Gebäude thermographisch zu untersuchen, die in einzelnen Aufnahmen nicht vollständig erfasst werden können. Ebenso lassen sich dann Daten verschiedener Aufnahmezeitpunkte und Sensoren vergleichen und kombinieren. Weiterhin ergibt sich auch die Möglichkeit, Objekte in den Infrarottexturen zu detektieren und in die 3D Struktur des Gebäudemodells zu integrieren.

Als Eingangsdaten dienen neben den Modellen Bildsequenzen von Fassaden, die von einem Fahrzeug mit einer IR-Kamera aufgenommen werden. Ungünstige Sichtbedingungen und Verdeckungen in innerstädtischen Bereichen erlauben in Einzelbildern keine direkte Zuordnung von Fassadenflächen zu korrespondierenden Bildbereichen mit ausreichender Genauigkeit. Daher werden in dieser Arbeit Verfahren beschrieben und untersucht, die durch eine gemeinsame Auswertung von Bildern einer Folge eine genauere Zuordnung ermöglichen. Die einzelnen Verarbeitungsschritte von der Aufnahme der Sequenz bis zur Auswertung der Textur werden behandelt und bewertet. Die Vorgehensweise zur geometrischen Kalibrierung der IR-Kamera wird vorgestellt und die Qualität der Kalibrierung bestimmt.

Um eine möglichst gute Verknüpfung der IR-Bilder durch homologe Punkte zu erreichen, werden Verfahren zur Extraktion von markanten Punkten (Förstner-Punkte und SIFT-Features) verglichen. Da die Anzahl und Qualität der markanten Punkte abhängig von der verwendeten Szene ist, werden für gleiche Szenenausschnitte die Ergebnisse der Bilder einer IR-Kamera mit denen einer Videokamera verglichen. Anschließend werden zwei Verfahren zur Orientierung der Bildsequenzen und Koregistrierung mit dem 3D Gebäudemodell vorgestellt. Beim ersten Verfahren wird zunächst ein Szenenmodell auf Basis der relativen Orientierung rekonstruiert und anschließend mit dem Gebäudemodell koregistriert. Beim zweiten Verfahren wird das Vorwissen des Gebäudemodells in die Orientierung der Bildsequenz integriert, um so die Orientierungen der Bilder der Sequenz an das Gebäudemodell anzupassen. Diese Verfahren zeigt gegenüber dem ersten Verfahren für die meisten Szenen eine höhere Genauigkeit der Orientierungsparameter.

Zur Texturierung der Fassaden wird ein Verfahren vorgestellt, das aus der IR-Bildfolge ein entzerrtes Bildmosaik erstellt. Schließlich werden Ansätze zur Analyse der Fassadentexturen vorgestellt, wie z.B. die Bestimmung der Temperaturdifferenz zwischen zwei Aufnahmezeitpunkten und die Detektion von Leckagen und Heizungsrohren.

Die experimentellen Ergebnisse zeigen, dass sich die für Bilder aus dem sichtbaren Spektrum entwickelten Verfahren auch auf Bilder im thermischen Infrarotspektrum übertragen lassen, wenn auch mit geringerer Genauigkeit der Koregistrierung.

Abstract

Thermal inspection of buildings with infrared (IR) cameras is traditionally done in image geometry. A 3d georeferencing or object based storing of recorded object surfaces is not carried out. This work deals with the creation of the 3 dimensional space reference and the object related storage of the recorded facade surfaces using 3d building models. Today, building models are used for acquisition of existing buildings in standard data format (CityGML) in geoinformation systems as well as in planning, construction, and service in building information systems (BIM). The generation of 3d space references allows the thermal survey of complex buildings even if they are not completely visible in a single image. Additionally, data from different recordings and sensors can be compared and combined. Furthermore, the possibility is given to detect objects in infrared textures and integrate them into the 3d structure of the building model.

The input data includes not only models but image sequences of facades recorded from a vehicle using an IR camera. Bad viewing conditions and occlusions in dense urban areas do not allow a direct georeferencing in single images of facade planes and corresponding image areas with an appropriate accuracy.

Thus, in this theses methods are described and investigated to allow an accurate matching by combined analysis of a sequence of images. From the image recordings to texture analysis, all steps are discussed and evaluated.

The strategy to geometric calibration of IR cameras is discussed and the quality of the calibration is evaluated.

An optimal join of the IR images using homologues points is achieved by comparing different strategies to extract interest points (Foerstner points and SIFT features). As the number and quality of the features is strongly depending of the scene used, the same scene is also processed with images recorded with a video camera and the results are compared.

Then, two strategies for image sequence orientation and coregistration with the 3d building model are presented. The first strategy reconstructs a scene model using the relative orientation of the sequence followed by the coregistration with the building model. The second strategy integrates the preknowledge of the building model and the orientation of the image sequence in one step to adopt the image orientation to the building model. This strategy leads to higher accuracy of the orientation parameters for most of the scenes compared to the first strategy.

A method for texturing facades is presented that generates a rectified image mosaic from the IR image sequence. Finally, approaches for facade texture analysis are presented including temperature change over time and detection of leakages and heating pipes.

Experimental results show that methods developed for images of the visual spectrum can be transferred to images from the infrared spectrum, but with lower accuracy in the coregistration.

Inhaltsverzeichnis

Kurzfassung	3
Abstract	5
Inhaltsverzeichnis	7
Liste der verwendeten Abkürzungen	11
Abbildungsverzeichnis	13
Tabellenverzeichnis	15
1 Einleitung	17
1.1 Motivation	17
1.2 Stand der Forschung	19
1.2.1 Orientierung von Einzelbildern über Passpunkte	19
1.2.2 Orientierung mehrerer Bilder über homologe Punkte	20
1.2.3 Extraktion homologer Punkte	23
1.2.4 Dichte Tiefenschätzung	24
1.2.5 Analyse von Infrarotbildern	24
1.2.6 Integration weiterer Datenquellen	25
1.2.7 CityGML	25
1.3 Ziele der Arbeit	26
1.4 Aufbau der Arbeit	27
2 Grundlagen	29
2.1 Thermische Infrarotstrahlung	29
2.1.1 Physikalische Grundlagen	29
2.1.2 Detektoren für thermische Infrarotstrahlung	31
2.1.3 Thermografie im Bauwesen	34
2.2 Geometrische Kalibrierung der Kamera	35
2.3 Extraktion von homologen Punkten	36
2.3.1 Förstner-Operator	37
2.3.2 SIFT	38
2.4 Relative Orientierung in einer Bildsequenz	39
3 Koregistrierung von Infrarotbildern mit Gebäudemodellen	43
3.1 Geometrische Kalibrierung der Infrarotkamera	44
3.1.1 Untersuchung zu Messmarken	44
3.1.2 Bestimmung der inneren Orientierung und der Verzeichnungsparameter	46

3.1.3	Alternative Kalibrierung über Passpunkte	48
3.2	Zweistufiges Verfahren mit relativer Orientierung	48
3.2.1	Relative Orientierung der Bildsequenz	48
3.2.2	Absolute Orientierungen der Bilder der Sequenz	50
3.3	Einstufiges Verfahren mit Gebäudemodell	52
3.3.1	Orientierung von Bildpaaren	53
3.3.2	Orientierung der Bildsequenz	57
4	Texturextraktion aus Infrarotbildsequenzen	61
4.1	Texturextraktion aus Bildsequenzen mit konstanter Blickrichtung	61
4.1.1	Projektion der Einzelbilder	62
4.1.2	Koregistrierung von Teiltexturen	65
4.2	Auswertung von Fassadentexturen	66
4.2.1	Auswertung von Einzeltexturen	66
4.2.2	Gemeinsame Auswertung mehrerer Texturen	67
4.2.3	Detektion von Leckagen	69
5	Experimente	71
5.1	Beschreibung der Testszenarien	71
5.2	Verwendete Datensätze	72
5.2.1	Infratec VarioCam hr 384×288	72
5.2.2	FLIR SC3000 320×240	74
5.2.3	Merlin MID 320×256	74
5.2.4	Verwendete Bildsequenzen	75
5.2.5	3D Gebäudemodell	76
6	Ergebnisse	79
6.1	Ergebnisse der Kamerakalibrierung	79
6.1.1	Kalibrierung der VarioCam hr mit Testfeldern	79
6.1.2	Kalibrierung der FLIR SC3000 und MerlinMID mit Fassadenpunkten	80
6.2	Qualität der homologen Punkte	83
6.3	Ergebnisse des zweistufigen Verfahrens	87
6.3.1	Genauigkeit der relativen Orientierung	87
6.3.2	Restfehler bei der Koregistrierung von Punktwolke und 3D Modell	89
6.4	Ergebnisse des einstufigen Verfahrens	91
6.5	Ergebnisse der Texturextraktion	92
6.5.1	Texturen aus dem zweistufigen Orientierungsverfahren	92
6.5.2	Texturen aus dem einstufigen Orientierungsverfahren	93
6.6	Kombination von Texturen unterschiedlicher Bildsequenzen	94
6.6.1	Auswertung mehrerer Texturen unterschiedlicher Blickrichtung	94
6.6.2	Auswertung mehrerer Texturen unterschiedlicher Aufnahmezeitpunkte	97
6.6.3	Auswertung von Texturen unterschiedlicher Kamerasysteme	98
6.7	Leckagedetektion in Infrarottexturen	99
7	Bewertung	101
7.1	Einschränkungen durch die Eigenschaften des Infrarotspektrums	101
7.2	Geometrische Kalibrierung der Kamera	102

7.3	Zweistufiges und einstufiges Orientierungsverfahren	102
7.4	Texturextraktion	104
7.5	Texturkombination	105
7.6	Leakagedetektion	106
8	Zusammenfassung und Ausblick	107
8.1	Zusammenfassung	107
8.2	Ausblick	108
8.3	Danksagung	109
	Literaturverzeichnis	111
	A Bildsequenzen	119
	Lebenslauf	123

Liste der verwendeten Abkürzungen

Abkürzung	Beschreibung	Seite
TIR	Thermisches Infrarot	17
GPS	Ground Positioning Systeme	18
INS	Integrated Navigation Systeme	18
SLAM	Simultaneous Location and Mapping	20
SFM	Structure From Motion	20
RANSAC	RANdom SAMple Consensus	21
UAV	Unmanned Aerial Vehicle	22
KLT	Kanade-Lucas-Tomasi Feature Tracker	22
SGM	Semi Global Matching	24
SIFT	Scale Invariant Feature Transform	23
CityGML	City Geography Markup Language	25
NIR	Nahes Infrarot	30
MWIR	MidWave InfraRed – Mittleres Infrarot	30
LWIR	LongWave InfraRed – Fernes Infrarot	30
DoG	Difference of Gaussian Filter	38
GRIC	Geometric Robust Information Criterion	48
TUM	Technische Universität München	71

Abbildungsverzeichnis

2.1	Vergleich Abstrahlung Gebäudefassade im Mittleren und Langwelligen Infrarot	31
3.1	Standard Reflektormarke für sichtbares Spektrum (Simmler [2009])	45
3.2	Reflektierende Marken im IR (Simmler [2009])	45
3.3	Aktiv strahlende Marken im IR (Simmler [2009])	46
3.4	Testfeldanordnungen Kalibrierung innen und außen (Simmler [2009])	47
3.5	Überblick zweistufigen Orientierungsverfahren	49
3.6	Überblick zum einstufigen Orientierungsverfahren	53
4.1	Geometrische Auflösung Einzelbildtextur bei schräg vorwärts blickender Aufnahme	64
4.2	Geometrische Auflösung Fassadentextur bei schräg vorwärts blickender Aufnahme	65
4.3	Unterschiedliche Verdeckungen bei unterschiedlicher Blickrichtung	69
4.4	Ablauf der Leckagedetektion	70
5.1	Luftbild der TUM (Google)	72
5.2	Kameraaufbau Messfahrzeug	73
5.3	Thermalbilder Infratec Variocam hr	73
5.4	Thermalbilder FLIR SC3000	74
5.5	Vergleich Thermalbilder Merlin MWIR und FLIR SC3000 LWIR	75
5.6	Überblick über die aufgenommenen Fassaden	76
5.7	3D Gebäudemodell der TUM untexturiert	77
6.1	Aufnahme aktives Innentestfeld (Simmler [2009])	79
6.2	Aufnahme reflektierendes Außentestfeld (Simmler [2009])	80
6.3	Kameraaufbau FLIR SC3000 und MerlinMID für die Passpunktkalibrierung	81
6.4	Fassadenbilder für Kalibrierung der FLIR SC3000	82
6.5	Beispielsequenz Feature Tracking	85
6.6	Überlagertes Bildpaar nach Feature Tracking	86
6.7	Relativer Kamerapfad und 3D Punktwolke einer Sequenz in Modellkoordinaten	88
6.8	Koregistrierung Punktwolke 3D Modell	90
6.9	Koregistrierte Punktwolke	90
6.10	Schrägluftbild mit Blick auf die Ostfassade	92
6.11	3D Ansicht der extrahierten Texturen aus dem zweistufigen Orientierungsverfahren	93
6.12	3D Ansicht des Gebäudemodells mit IR Texturen aus dem zweistufigen Verfahren	93
6.13	Extrahierte Texturen aus Einzelbildern	94
6.14	Aus einer Bildsequenz extrahierte Textur einer Fassade	95
6.15	Texturierte Fassaden mit dem einstufigen Verfahren	95
6.16	Texturen aus Vorwärts- und Rückwärtssicht	95
6.17	Differenztextur der Vorwärts- und Rückwärtssicht	96

6.18	Texturen aus Vorwärtssicht abends und am Morgen	97
6.19	Differenztextur der Abkühlung über die Nacht	98
6.20	Überlagerung IR-Textur und RGB-Textur	98
6.21	Leckagedetektion: Ausmaskierung von Fassadenobjekten	99
6.22	Markierte Leckagestellen einer Fassade	100
A.1	Auszug Sequenz Ost	119
A.2	Auszug Sequenz West	120
A.3	Auszug Sequenz Sued	121
A.4	Auszug Sequenz Nord	122

Tabellenverzeichnis

6.1	Kalibrierergebnisse der Infratec VarioCam hr	81
6.2	Kalibrierergebnisse der FLIR SC3000 und der MerlinMID	83
6.3	Vergleich SIFT Features und Förstner Punkte	86
6.4	Genauigkeit der relativen Orientierung	87
6.5	Genauigkeit der absoluten Orientierung	89
6.6	Genauigkeit der direkten Orientierung	91

1 Einleitung

1.1 Motivation

Die Extraktion von Gebäudefassaden und Fassadenelementen hat in den letzten Jahren zu immer feiner detaillierten 3D Modellen und Texturen geführt. Zum einen geht es um eine möglichst effiziente und ansprechende Visualisierung (Müller et al. [2006]), zum anderen um die Extraktion von Fassadenelementen zur Verfeinerung der Gebäudegeometrie. Reznik & Mayer [2008] und Hohmann et al. [2009] nutzen optische Bilddaten zur Extraktion von Elementen wie Fenstern oder Gauben. Neben der reinen Geometrie ist eine Objekterkennung und damit eine Interpretation notwendig. Für die Objekterkennung entscheidend sind Unterschiede in der Strahlungsintensität in den beobachteten spektralen Bändern.

Infrarotkameras nehmen elektromagnetische Strahlung in dem für Menschen unsichtbaren thermischen infraroten Bereich (TIR) auf. Die gemessene Strahlungsintensität hängt in diesem Spektrum nicht von der reflektierten Strahlung an der Oberfläche eines Objektes ab, sondern von der inneren Energie, die ein Objekt in Form seiner Temperatur beinhaltet. Damit können andere Eigenschaften von Objekten erkannt werden, die im sichtbaren Spektrum verborgen bleiben. Bilder im thermischen Infrarot werden heute in unterschiedlichen Maßstäben und für die verschiedensten Anwendungszwecke eingesetzt. Bietet die Satellitenbeobachtung durch ein weites Blickfeld die Möglichkeit, große Flächen in einer Aufnahme zu beobachten, so ist die Auflösung der Satellitendaten für eine detaillierte Untersuchung einzelner Gebäude oder Straßenzüge nicht ausreichend. Satellitenbilder dienen z. B. zur Erkennung von Brandherden (Siegert et al. [2004]), Vegetation (Quattrochi & Luvall [1999]) oder zur Detektion von urbanen Wärmeinseln (Lo & Quattrochi [2003]). Flugzeuggetragene Infrarot-Systeme werden z.B. zur Detektion von stehenden Fahrzeugen (Stilla & Michaelsen [2002]; Hinz & Stilla [2006]) und bewegten Fahrzeugen (Kirchhof & Stilla [2006]) eingesetzt. Fernwärmeleitungen werden mit flugzeuggetragenen Infrarotkameras auf Leckagen untersucht (Koskeleinen [1992]). Terrestrische Infrarotkameras nehmen z.B. die Fassadenabstrahlung von Bauwerken im thermischen Infrarot (TIR) auf, um mit ihnen das Wärmeverhalten von Gebäuden für Wärmepässe bestimmen zu können (Klingert [2005]). Sie haben den Blick auf die Fassaden, nicht aber auf die Dächer, die aus Flugdaten ergänzt werden müssen. Dabei entstehen Probleme durch Teilverdeckung und den geringen Abstand zum Messobjekt z.B. in engen Straßenschluchten, was eine Kombination von Einzelbildern erfordert.

Aufnahmen von Gebäuden im thermischen Infrarot werden zumeist zeitaufwändig von einzelnen Standorten aufgenommen (Fouad & Richter [2012]). Dabei sind verschiedene Faktoren zu berücksichtigen, die einen Einfluss auf die gemessene Strahlungsintensität haben. Neben der äußeren Bedingungen wie Sonneneinstrahlung, Tageszeit, Wind und Temperatur sind dies die Innentemperatur der Räume, der Emissionsgrad der beobachteten Fläche, Spiegelungen anderer Objekte oder des kalten Himmels an Fenstern und metallischen Oberflächen. Ebenso strahlen benachbarte Objekte sich gegenseitig an, was sie aufwärmt und die Strahlungsintensität erhöht, ohne, dass daraus auf einen hohen Wärmeverlust geschlossen werden kann. Größere Gebäudeteile müssen auf Grund der geringen optischen Auflösung der Kameras durch mehrere Bilder erfasst

werden. Die gewonnenen Abbildungen werden in der Regel visuell interpretiert und Dokumentationen erfolgen in der aufgenommenen Bildgeometrie ohne einen echten 3D Bezug innerhalb einer Geodatenbasis. Klingert [2005] wendet auf den Infrarotaufnahmen klassische Bildverarbeitungsalgorithmen an, um warme oder kalte Stellen zu identifizieren. Er bleibt aber bei der Interpretation im Bildraum. In diesem Fall ist eine radiometrische Kalibrierung notwendig, um Strahlungsintensität messen und Temperaturen ableiten zu können. Eine geometrische Kalibrierung ist nicht notwendig und wird daher von den meisten Herstellern handelsüblicher Kamerasysteme auch nicht vorab durchgeführt.

Das fehlen von Informationen über die Orientierung der Kamera zum Aufnahmezeitpunkt und die Interpretation der Infrarotbilder im Bildraum schränken die Möglichkeiten der Auswertung stark ein. So ist es nicht möglich, größere oder komplexere Gebäude in ihrer Gesamtheit zu untersuchen, da nur getrennt aufgenommene Einzelbilder vorliegen. Ebenso ist eine automatische Auswertung von Bildern zu unterschiedlichen Aufnahmezeitpunkten nicht möglich, da Bildpunkte in einem Bild keine Entsprechung im anderen Bild zugeordnet werden kann. Außerdem ist eine gemeinsame Auswertung von Bildern aus verschiedenen Spektralbereichen nicht möglich. Ohne Orientierung der Bilder ist auch die Bestimmung von 3D Koordinaten für in den Bildern gefundene Objekte wie Fenster, Heizungsleitungen oder Leckagen nicht möglich.

Die Kombination mehrerer Bilder eines Objekts aus dem sichtbaren Spektrum durch eine direkte Georeferenzierung mit Passpunkten (Haralick et al. [1989]; Horaud et al. [1989]; Haralick et al. [1994]; Quan & Lan [1999]; Triggs [1999]; Zhi & Tang [2002]) oder die relative Orientierung der Bilder zueinander mit homologen Punkten (Haralick et al. [1994]; Hartley & Zisserman [2004]; Lepetit et al. [2004]; Nistér [2004]; Mayer [2007]; Pollefeys et al. [2008]; Heinrichs et al. [2008]; Reznik & Mayer [2008]; Ripperda [2009]; Becker [2010]) ist ein bereits gut erforschtes Feld, bei dem neben der Rekonstruktion der Geometrie und der Texturen auch die Extraktion der Geometrie von Fassadenelementen untersucht wird (Reznik & Mayer [2008]; Ripperda [2009]; Becker [2010]; Tuttas & Stilla [2013]). Die Bestimmung der äußeren Orientierung an einem vorhandenen Gebäudemodell erfolgt in Einzelbildern über eine Kantenzuordnung, in Bildpaaren über eine Flächenzuordnung. Da die Aufzeichnung der äußeren Orientierung einer Kamera mittels GPS mit Fehlern behaftet ist, können die Bilddaten nicht ohne vorherige Koregistrierung mittels Passpunkten an einem gegebenen Gebäudemodell kombiniert werden. Insbesondere bei Aufnahmen in urbanen Gebieten vom Boden aus ist abhängig von der aufgenommenen Szene und dem Detailgrad der vorhandenen Gebäudedaten eine direkte Zuordnung zwischen Bildern und Modell nicht möglich. In diesen Fällen werden Bildsequenzen zuerst relativ zueinander orientiert und anschließend in ein globales Koordinatensystem überführt. Für die Orientierung in Weltkoordinaten werden GPS/INS Systeme eingesetzt.

Im thermischen infraroten Spektrum gibt es bislang kaum Untersuchungen, diese Verfahren anzuwenden. Bilder im thermischen infraroten Spektrum zeigen bezüglich Pixelzahl und geometrischer Auflösung, aber auch in der Sichtbarkeit von Objekten ein anderes Verhalten wie Bilder im sichtbaren Spektrum. Da thermische Infrarotstrahlung nicht durch Reflexion an der Objektoberfläche entsteht, sondern durch die Wärmeenergie des beobachteten Objektes, zeigen Strukturen häufig keine starken Kanten oder markanten Punkte und sind daher schwerer zu lokalisieren. Durch die im Vergleich zu Bildern aus dem sichtbaren Spektrum geringere geometrische Auflösung wird die Genauigkeit gefundener Intensitätsunterschiede weiter verschlechtert. Auch sind in einem Bild gefundene Intensitätsunterschiede aus einem anderen Blickwinkel oder bereits nach kurzer Zeit häufig nicht mehr zu erkennen.

In dieser Arbeit soll daher untersucht werden, mit welchen zu erwartenden Genauigkeiten für typische urbane Szenarien bei der Rekonstruktion der äußeren Orientierung von Kamerastandorten, sowie bei der Rekonstruktion von 3D Geometrie und der Extraktion und Interpretation

von Texturen im thermischen Infrarot zu rechnen ist. Es werden nach einer Untersuchung über die Genauigkeit der geometrischen Kalibrierung von Kameras im thermischen Infrarot zunächst mit dem Förstner-Operator und dem SIFT-Operator zwei gängige Vertreter für die Detektion von markanten Punkten verglichen. Die Anzahl, Genauigkeit und Korrektheit der markanten Punkte hat einen entscheidenden Einfluss auf die Genauigkeit der relativen Orientierung der Bildsequenz und der anschließenden Koregistrierung mit dem gegebenen Gebäudemodell. Für die Bestimmung der relativen Orientierung und die Koregistrierung mit dem Gebäudemodell werden ein zweistufiges und ein einstufiges Verfahren verglichen. Auf Basis der Koregistrierung erfolgt die Extraktion der Texturen aus den Einzelbildern und ihre Kombination zu vollständigen Fassadentexturen. Zum Abschluss wird eine Reihe von Untersuchungen auf diesen Texturen durchgeführt, darunter die Überlagerung von Texturen und die Detektion von potenziellen Wärmelecks.

1.2 Stand der Forschung

Bei der automatischen Texturierung von Gebäudemodellen kann grob zwischen zwei Vorgehensweisen unterschieden werden. Zum einen werden vorhandene Geometrien sowohl für die Orientierung der Kameras wie auch die Extraktion der Texturen genutzt (Haralick et al. [1989, 1994]; Horaud et al. [1989]; Quan & Lan [1999]; Triggs [1999]; Zhi & Tang [2002]). Hier findet eine Zuordnung von Strukturen der Geometrie des Modells mit Strukturen, die aus den Bilddaten extrahiert werden statt. Hierfür kommen Liniensegmente, Passpunkte oder Flächen der Szene in Frage. Der Vorteil dieser Herangehensweise liegt darin, dass das Gebäudemodell global referenziert vorliegt und bereits mit topologischen und semantischen Informationen ausgestattet ist. Der Nachteil besteht jedoch darin, dass die Qualität des Gebäudemodells einen großen Einfluss auf die Qualität der Texturextraktion hat. Die korrekte Zuordnung von Liniensegmenten oder Punkten setzt eine annähernd bekannte äußere und innere Orientierung der Kamera voraus. Liegen diese Daten vor, so können aus einem einzelnen Bild Texturen erzeugt werden.

Zum anderen können aus Bilddaten unter Kenntnis der relativen Orientierung der Bilder die Gebäudegeometrie erzeugt und die Texturen extrahiert werden (Agarwal et al. [2009]; Fitzgibbon & Zisserman [1998]; Frahm et al. [2010]; Hartley [1997a]; Hartley & Zisserman [2004]; Heinrichs et al. [2008]; Longuet-Higgins [1981]; Maybank [1993]; Mayer [2007]; Mayer et al. [2012]; McGlone [2004]; Newcombe et al. [2011]; Nistér [2004]; Pollefeys et al. [2008]; Stewenius et al. [2005]). Hierfür werden zumeist homologe Punkte in den Bildern gesucht und damit die Bilder über eine Kombination aus Vorwärts- und Rückwärtsschnitt verknüpft. Dabei werden neben der relativen Orientierung der Bilder auch 3D Punkte der beobachteten homologen Punkte geschätzt. Die Prozessierung teilt sich in vier Schritte auf: Bestimmung homologer Punkte in den einzelnen Bildern, Bestimmung der relativen Orientierung – oft inklusive Selbstkalibrierung der Kamera (Agarwal et al. [2009]; Longuet-Higgins [1981]; Maybank [1993]; Mayer [2007]) –, Extraktion von 3D Koordinaten, Extraktion der Texturen. Diese Herangehensweise hat den großen Vorteil, dass die Texturen unmittelbar mit der Geometrie zusammen passen, da sie gleichzeitig aus denselben Eingangsdaten erzeugt werden. Im Umkehrschluss ist erzeugte und mit den Intensitätswerten der Bilder eingefärbte Punktwolke auf die in den Bildern sichtbaren Teile des Gebäudes beschränkt. Es fehlt der Punktwolke ein globaler Bezug und topologische und semantische Information.

1.2.1 Orientierung von Einzelbildern über Passpunkte

Die Orientierung von Einzelbildern über Passpunkte hat sich in der Luftbildphotogrammetrie bewährt und ist vielfach für Anwendungen im Nahbereich adaptiert worden. Dabei wird einem 2D Bildpunkt ein bekannter 3D Passpunkt zugewiesen. Die Verknüpfung von mindestens drei 2D Bildpunkten mit 3D Objektpunkten über die Kollinearitätsgleichungen erlaubt die Berechnung

der äußeren Orientierung der Kamera. Einen Überblick über Verfahren, die Einzelbilder über drei Passpunkte orientieren bieten Haralick et al. [1994], die insbesondere die numerische Stabilität der Verfahren bewerten. Gemeinsam ist diesen Verfahren, dass sie aus drei 2D Bildpunkten, deren 3D Objektkoordinaten im 3D Raum bekannt sind, die Kameraposition zurückrechnen. Im allgemeinen Fall besitzt die Positionsbestimmung aus diesen drei Punkten vier mathematisch richtige Lösungen. Horaud et al. [1989] schlagen vor, einen vierten Passpunkt hinzuzufügen, um die richtige der vier möglichen Lösungen zu bestimmen. Die Verallgemeinerung dieses Verfahrens auf n Passpunkte ist von Quan & Lan [1999] formalisiert worden. Sie führen ein lineares Gleichungssystem ein, um im allgemeinen Fall eine eindeutige Lösung zu finden. Zhi & Tang [2002] schlagen vor, eine direkte lineare Transformation (DLT) zu verwenden, die zur Bestimmung einer eindeutigen Lösung mittels symbolischer Vervollständigung um Gleichungssysteme ohne Einschränkung des Nullraums bestimmt. Triggs [1999] führt eine Generalisierung der DLT für die Bestimmung der äußeren und der inneren Orientierung der Kamera aus einem Einzelbild mit vier oder fünf bekannten 3D Punkten ein. Haralick et al. [1989] schlagen hierfür auch iterative Verfahren vor.

1.2.2 Orientierung mehrerer Bilder über homologe Punkte

In den letzten 20 Jahren sind verschiedenste Ansätze entwickelt worden, die sich mit dem Problem der gemeinsamen Bestimmung der inneren und äußeren Orientierung aus Bilddaten ohne Szenenwissen beschäftigen. Dabei kann man die Verfahren anhand der verwendeten Bilddaten in drei Gruppen einteilen: zum einen die echten Stereoverfahren mit zwei Kameras und räumlichem Vorwärtsschnitt, zum anderen Verfahren mit einer einzelnen Kamera, die in einer zeitlichen Sequenz die Szene von unterschiedlichen Standorten und/oder unterschiedlichen Blickwinkeln aufnimmt, und schließlich Verfahren mit ungeordneten Sammlungen von Bildern, die mit unterschiedlichen Kameras zu unterschiedlichen Zeiten aufgenommen worden sind und über keine bekannten Parameter der inneren und äußeren Orientierung verfügen (Agarwal et al. [2009]; Frahm et al. [2010]). Verfahren, die mit einer Kamera aus unterschiedlichen Positionen eine Abfolge von Bildern aufnehmen, verbinden die 2D Bilder über eine relative Orientierung und erzeugen anschließend eine Szene in 3D Modellkoordinaten (Hartley & Zisserman [2004]). Solche Verfahren werden auch als SLAM – Simultaneous Location And Mapping – oder SFM – Structure From Motion – bezeichnet und auch in der Augmented Reality eingesetzt, z.B. bei Parallel Tracking and Mapping (Klein & Murray [2007]) und Dense Tracking and Mapping (Newcombe et al. [2011]). Der Vorteil der relativen Orientierung besteht darin, dass sie für die Szenenrekonstruktion kein Vorwissen über die Szene benötigt und die Bilder nur über die Analyse des 2D Bildraumes der einzelnen Bilder orientiert werden. Für die Bestimmung der relativen Orientierung werden nur 2D Bildpunkte verwendet und deren Korrespondenzen in anderen Bildern wiederum nur im Bildraum gesucht, ohne, dass ein Übergang in den 3D Raum stattfindet. Erst im Anschluss an die Bestimmung der relativen Orientierung in 3D Modellkoordinaten findet eine 3D Szenenrekonstruktion statt. Die relative Orientierung ist also in unbekannten Szenen einsetzbar oder in Fällen, in denen die Bilder nur einen kleinen Ausschnitt einer Szene zeigen, aus dem direkt keine Positionsbestimmung der Kamera möglich ist, weil zu wenige bekannte Punkte und Liniensegmente oder nur Flächenteile sichtbar sind.

Longuet-Higgins [1981] beschreibt ein Verfahren, das ohne bekannte 3D Koordinaten von Passpunkten zwei Bilder auf Basis homologer Punkte relativ zueinander orientiert. Der Acht-Punkt-Algorithmus berechnet über die sogenannte Fundamentalmatrix die relative Orientierung des zweiten Bildes zum ersten. Da die Fundamentalmatrix auch die Informationen über die innere Orientierung und Verzeichnung der Kameras der beiden Bilder enthält, benötigt es keine kalibrierten Bilder. Liegen für die Bilder bereits Kameramatrizen mit den Kalibrierdaten vor, so vereinfacht sich die Fundamentalmatrix zur essentiellen Matrix. Da die Fundamentalmatrix nur bis auf einen Skalierungsfaktor genau bestimmt wird, werden für die Berechnung acht Punktkor-

respondenzen in beiden Bildern benötigt. Der Suchraum für korrespondierende 2D Bildpunkte im zweiten Bild wird durch die Epipolargeometrie beschrieben (Hartley [1997a]). Aus der essentiellen Matrix können die Rotationsmatrix und der Translationsvektor des zweiten Bildes relativ zum ersten bestimmt werden.

Weiterentwicklungen des Acht-Punkt-Algorithmus kommen mit sieben Punktpaaren aus, lassen dafür aber bis zu drei mathematisch korrekte Lösungen zu (Maybank [1993]; Hartley & Zisserman [2004]). Aus der berechneten relativen Orientierung der Kameras kann eine 3D Szenenrekonstruktion erfolgen. Die Weltkoordinaten eines Punktes in zwei Bildern lassen sich, bis auf den Skalierungsfaktor, durch eine inverse Projektion bestimmen.

Im kalibrierten Fall kann die Anzahl der Unbekannten und damit der benötigten Punktpaare für die Bestimmung der essentiellen Matrix der relativen Orientierung auf fünf Punktpaare reduziert werden (Nistér [2004]). Auch hier ist die essentielle Matrix nur bis auf den Skalierungsfaktor bestimmt. Fünf Paare korrespondierender, homologer Punkte beider Bilder bilden ein Gleichungssystem mit einem Polynom 10. Grades. Für jedes Bildpaar ergeben sich vier algebraisch korrekte Lösungen für die Rotationsmatrix und den Translationsvektor des zweiten Bildes. Die geometrisch richtige Lösung wird bestimmt, indem geprüft wird, ob alle Punktpaare in den 3D-Raum projiziert vor beiden Kameraebenen liegen. Um eine bessere Robustheit der Ergebnisse zu erreichen, wird die Auswahl der fünf Punktpaare zufällig nach dem RANSAC-Verfahren (Fischler & Bolles [1981]) vorgenommen und mehrfach wiederholt. Ein Hypothesen-Test findet falsche Punktpaare und sortiert sie aus (Torr & Murray [1997]; Zhang [1998]). In einer Bildsequenz aus vielen Einzelbildern wandert das Verfahren immer ein Bild weiter. Bereits akzeptierte Punktpaare werden in der Sequenz weiter verwendet solange sie sichtbar sind. Gleichzeitig wird nach neuen Punktpaaren gesucht. Dadurch kann der Hypothesen-Test auf ganze Bildfolgen ausgeweitet werden und erlaubt die Kombination ganzer Sequenzen, die über die Punktpaare und ihre geschätzten 3D Koordinaten miteinander verbunden sind.

Die Verknüpfung in Sequenzen wird bei Nistér von zwei auf drei Bilder ausgebaut. Dabei wird der sogenannte Trifokal-Tensor eingeführt (Hartley [1997b]). Durch die Berechnung der relativen Orientierung über drei anstatt zwei Bilder lässt sich die Qualität der Schätzung deutlich verbessern (Fitzgibbon & Zisserman [1998]; McGlone [2004]; Stewenius et al. [2005]). Die relative Orientierung über den Trifokaltensor ist nur bis auf einen Skalierungsfaktor bestimmt. Der Nutzen des Trifokaltensors besteht darin, ein in zwei Bildern gefundenes Punktpaar im dritten Bild zu überprüfen. Bei einer Fundamentalmatrix bzw. essentiellen Matrix ist dies nicht möglich, da ein Punkt in einem Bild auf eine Epipolarlinie im anderen Bild abgebildet wird. Der Trifokaltensor erweitert die Epipolargeometrie von zwei auf drei Bilder. Für ein Bildpaar liegen die homologen Punkte des ersten Bildes auf den korrespondierenden Epipolarlinien des zweiten Bildes. Bei einem dritten Bild liegt der homologe Punkt im Schnittpunkt der beiden Epipolarlinien, die durch die homologen Punkten im ersten und zweiten Bild aufgespannt werden. Daraus ergibt sich ein eindeutiges Ergebnis, sofern die homologen Punkte nicht in der Epipolarebene liegen, die durch die drei Projektionszentren der Bilder festgelegt ist. Damit ist der Trifokaltensor gerade im Hinblick auf eine Bildsequenz mit konstanter Blickrichtung und Bewegung gegenüber dem rein auf der Fundamentalmatrix bzw. essentiellen Matrix basierenden Ansatz im Vorteil.

Weiterentwicklungen dieses Verfahrens in Bezug auf Bildsequenzen existieren von Mayer [2007]; Mayer et al. [2012], Pollefeys et al. [2008] und Heinrichs et al. [2008].

Mayer [2007] adaptiert Nistér's Algorithmus zur Extraktion und Texturierung von Fassaden aus verschiedenen Ansichten. Aus den 3D Punkten der extrahierten Punktpaare schätzt er ebene Polygonflächen der Fassaden und extrahiert zusammen mit der geschätzten äußeren Orientierung der Kamera bezüglich der Vertikalen Texturen für die erzeugten Polygone. Bei ausreichend großer Aufnahmebasis können aus einer Sequenz von Bildern auch Fassadendetails wie Fenster extrahiert

werden (Reznik & Mayer [2008]). In Mayer et al. [2012] wird dieses Verfahren weiter verfeinert und um ein Semi-Global-Matching (Hirschmüller [2008]) zur Erzeugung dichter Punktwolken erweitert. Bartelsen & Mayer [2010] zeigen die Anwendung dieses Verfahrens für flugzeuggetragene Sequenzen von unbemannten Luftfahrzeugen (UAVs).

Heinrichs et al. [2008] kombinieren den Trifokal-Tensor einer Bildfolge mit echtem Stereo aus drei Kameras, die fest auf einem Träger montiert und gemeinsam kalibriert sind. Jede Kamera berechnet für sich über den Trifokal-Tensor aus drei Bildern ihre relative Orientierung. Die Bilder zweier Kameras werden anschließend zusammengeschaltet und die Koordinaten der Punktpaare verglichen. Daraus ergibt sich eine genauere Tiefenschätzung der 3D Punkte und ein genauerer Kamerapfad. Die dritte Kamera dient zur Kontrolle. Das System erreicht insgesamt Genauigkeiten von wenigen Millimetern für eine variable Stereobasis zwischen $1/2$ und einem Meter.

Pollefeys et al. [2008] kombinieren Nistér's Algorithmus (Nistér [2004]) mit einem Plane-Sweeping Stereo Algorithmus für die Tiefenschätzung (Yang & Pollefeys [2003]). Durch die Tiefenschätzung aus dem Plane-Sweeping kann ein grobes Umgebungsmodell nahe Echtzeit erstellt werden. Die dafür nötigen Kamerapositionen werden entweder aus der relativen Bewegung von Features in Bildern gewonnen (Nistér et al. [2006]) oder aus einem Kalman Filter (Grewal & Andrews [2008]) mittels GPS/INS Daten.

Bei der Betrachtung von Gebäuden liegen die meisten markanten Punkte auf den Fassaden und befinden sich damit oft annähernd in einer Ebene. Diese Punktkonstellation beeinflusst die Eindeutigkeit der relativen Orientierung. Die Rekonstruktion der inneren Orientierung der Kamera aus Punkten einer Ebene ist nicht möglich (Maybank [1993]), weil es für jede Kameraposition eine innere Orientierung und Rotationsmatrix gibt, die eine Homographie zwischen der Objektebene und der Bildebene beschreibt. Ist die innere Orientierung bekannt, ergeben sich zwei Lösungen für die essentielle Matrix (Longuet-Higgins [1986]; Maybank [1993]), es sei denn, die Basislinie der Kameras steht senkrecht auf der Ebene. In diesem Fall gibt es eine Lösung. Sonst kann über die Bedingung, dass die 3D-Punkte aller Punktpaare vor den Kameras liegen müssen die richtige Lösung ermittelt werden.

Über Homographien können Ebenen in Bildpaaren geschätzt und die Kamerapositionen durch eine Veränderung der Orientierung der Ebenen zueinander verbessert werden (Hartley & Zisserman [2004]). Dies führt insbesondere ohne bekannte innere Orientierung der Kamera zu Eindeutigkeitsproblemen, die von Pollefeys et al. [2002] untersucht werden. Die Homographie wird besonders in Echtzeitanwendungen und Echtzeit nahen Anwendungen zur Bildorientierung eingesetzt, weil sie zumeist niedrigeren Berechnungsaufwand gegenüber den Punkt orientierten Verfahren hat. Ein Beispiel dafür ist das Tracking-Verfahren der Universität von Oxford von Simon et al. [2000]. Im ersten Bild wird manuell eine Referenzebene aus vier Punkten definiert. Weiteres Szenenwissen ist nicht nötig. Die definierte Ebene wird in jedem folgenden Bild der Sequenz wieder gesucht und für die Positionsschätzung verwendet. Wenn ganze Bildsequenzen vorliegen, können mehrere Bilder für die Positionsschätzung verwendet werden. Gibson et al. [2003] detektieren markante Punkte mittels Harris-Corner-Detector Harris & Stephens [1988] in einem Bild und suchen diese Punkte in so vielen Bildern einer Sequenz wie möglich mit einem Kanade-Lucas-Tomasi Feature Tracker (KLT-Tracker) (Shi & Tomasi [1994]). Da die genaueste Tiefenschätzung dann erreicht werden kann, wenn die Bewegung der Punkte zwischen zwei Bildern eher groß ist, werden Bildpaare gesucht, die eine möglichst große Parallaxe aufweisen und gleichzeitig genügend gemeinsame Punktpaare besitzen. Aus diesen Bildern wird nicht nur die relative Orientierung der Kamera in jedem Bild bestimmt, sondern es werden auch lokale 3D Koordinaten der gemeinsamen Bildpunkte berechnet. Anschließend wird eine Bündelblockausgleichung über alle rekonstruierten Bildpaare durchgeführt. Dadurch werden für die gesamte Szene exakte 3D Punkte und die

Kameraorientierung rekonstruiert. Die Weiterentwicklung dieses Systems hat zur kommerziellen Software PFTrackTM geführt.

1.2.3 Extraktion homologer Punkte

Alle Verfahren zur relativen Bildorientierung benötigen homologe Punkte in den zu orientierenden Bildern. Hierbei haben sich sogenannte Interest-Operatoren als geeignet herausgestellt, die Punkte extrahieren, die in ihrer lokalen Umgebung markant sind und eindeutig beschrieben werden können. Sie müssen also eine möglichst große lokale Grauwertänderung aufweisen. Der Interest-Wert kann durch eine statistische Differenzbildung in der lokalen Umgebung wie beim Moravec-Operator (Moravec [1977]) bestimmt werden. Durch einen Schwellwert werden anschließend die Bildpunkte mit den höchsten Interestwerten als markante Punkte ausgewählt.

- Der Harris-Corner-Detektor (Harris & Stephens [1988]) ersetzt die im Moravec-Operator verwendeten Gradientensummen durch partielle Ableitungsfunktionen der Bildfunktion in x- und y-Richtung. Diese Ableitungen werden in einer Autokorrelationsmatrix kombiniert. Hieraus lässt sich bestimmen, ob der Punkt an der Stelle (x,y) ein markanter Punkt, ein Punkt auf einer geraden Kante oder ein Punkt in einer homogenen Fläche ist. Außerdem erhält jeder Punkt einen Interestwert, der seine Stärke angibt. Für den Harris-Corner-Detektor ergibt sich in Kombination mit einem Gauß-Filter einen Multiskalen-Ansatz. Dabei wird der Strukturtensor auf verschiedenen Bildskalen (Lindeberg & Garding [1997]) genutzt. Der Harris-Corner-Detektor ist invariant gegenüber Translation, Rotation und uniformer Skalierung, nicht aber gegenüber projektiven und affinen Transformationen.
- Auch der Förstner-Operator (Förstner & Gülch [1987]) verwendet den Strukturtensor der partiellen Ableitungen der Bildfunktion. Er nutzt die Eigenschaft aus, dass die Inverse der Autokorrelationsmatrix der Kovarianzmatrix entspricht, die die Genauigkeit des Interest-Punktes beschreibt. Große Gradienten in der Autokorrelationsmatrix bedeuten kleine Varianzen in ihrer Inversen. Aus den Eigenwerten lassen sich die Achsen einer Fehlerellipse bestimmen. Ein guter Interest-Punkt liegt dann vor, wenn seine Fehlerellipse möglichst klein und möglichst rund ist. Eine längliche Fehlerellipse lässt auf eine Kante schließen, große Ellipsen auf homogene Regionen. Der Förstner-Operator ist wenig rauschempfindlich. Die Bestimmung der Fehlerellipse kann mit Subpixel-Genauigkeit erreicht werden. Wie der Harris-Corner-Detektor ist der Förstner-Operator invariant gegenüber Translation, Rotation und uniformer Skalierung, nicht aber gegenüber projektiven und affinen Transformationen.
- Die skaleninvariante Merkmalstransformation (SIFT) (Lowe [2004]; Läbe & Förstner [2006]; Brown & Lowe [2007]) berechnet aus einem Bild eine große Anzahl von Merkmalspunkten. Sie sind wie die des Förstner-Operators invariant gegenüber Translation, Rotation, Skalierung und teilweise gegen Beleuchtungsänderungen. Darüber hinaus sind die Punkte robust gegenüber geometrischen Verzerrungen wie sie bei der perspektivisch verzerrten Bildern vorkommen. Die Suche nach Merkmalspunkten orientiert sich anders wie bei den vorangegangenen Operatoren nicht an Kanten im Bild, die in der 1. Ableitung detektiert werden, sondern arbeitet regionenbasiert mit der Detektion von „Blobs“ im Skalen-Raum (Lindeberg [1998]). Ein Bild wird dafür in eine Bildpyramide mit verschiedenen Auflösungen umgerechnet und es werden jeweils mit einem Difference-of-Gaussian-Filter die 2. Ableitungen gebildet. Als Merkmalspunkte werden diejenigen Pixel gesucht, die ein lokales Maximum oder Minimum in ihrer Achternachbarschaft und in den benachbarten Skalen der Bildpyramide darstellen. Da deutlich zu viele Extrema entstehen, um eine gute Zuordnung zwischen zwei oder mehr Bildern herstellen zu können, müssen anschließend instabile und unsichere Punkte mit geringem Kontrast oder auf Kanten entfernt werden. Im Unterschied zum

Harris-Corner-Detektor oder Förstner-Operator wird hier nicht die erste Ableitung, sondern die Hessematrix verwendet. Mit der Hessematrix zweiter Ordnung können die Eigenwerte der Hauptkrümmung auf jedem Bild der Bildpyramide bestimmt werden und aus ihnen der sogenannte Formfaktor ermittelt werden. Er ist minimal, wenn die Krümmungen gleich sind. Ein hoher Formfaktor weist darauf hin, dass die Krümmungen unterschiedlich sind, der Punkt also auf einer Kante liegt. Ähnliche Eigenschaften wie SIFT zeigt auch der SURF-Algorithmus (Bay et al. [2008]), der ebenfalls auf verschiedenen Bildauflösungen arbeitet, aber die Gradienten der Nachbarschaft über Haar-Wavelets beschreibt.

Geht man davon aus, dass die aufgenommenen Bilder in einer Sequenz mit hoher Taktfrequenz und annähernd gleich bleibendem Blickwinkel aufgenommen werden, so ist die Änderung der Szenen zwischen zwei Bildern und damit die Änderung eines markanten Punktes meist sehr gering (Kirchhof & Stilla [2006]). Schaffalitzky & Zissermann [2002] zeigen, wie die für die Suche nach homologen Punkten wichtige Überlappung zweier Bilder sowohl für geordnete wie ungeordnete Sequenzen bestimmt werden kann, um zu entscheiden, in welchen Bildern nach homologen Punkten gesucht werden soll. Es bietet sich dann eine flächenbasierte Zuordnung über ein Kreuzkorrelationsverfahren an. Insbesondere bei Infrarotaufnahmen, die meist einen geringen Detailgrad zeigen, findet dieses Verfahren Anwendung, zumal es Echtzeit fähig ist.

1.2.4 Dichte Tiefenschätzung

Die aus den Verfahren zur Bildorientierung resultierenden 3D Punktwolken der Szene beinhalten nur die relativ berechneten 3D Koordinaten der homologen Punkte. Von Hirschmüller [2008] wird ein heute in der photogrammetrischen 3D Rekonstruktion immer weiter verbreiteter Ansatz zur dichten Tiefenschätzung, der als Semi-Global-Matching (SGM) bezeichnet wird, vorgestellt und auf Basis der relativen Orientierung für alle Bildpunkte, die in mehreren Bildern auftreten, 3D Koordinaten bestimmt. Hirschmüller [2008] verwendet für die Zuordnung von Bildpunkten eine Kostenfunktion mit einer (semi-)globalen Glättungsfunktion. Leberl et al. [2010] zeigen die Kombination von Structure-from-Motion-Verfahren mit Semi-Global-Matching für die Generierung von texturierten Gebäudemodellen aus Luftbildsequenzen. Mayer et al. [2012] nutzen das Semi-Global-Matching als Erweiterung der 3D Rekonstruktion aus ungeordnete terrestrische und flugzeuggetragene Bildsequenzen.

1.2.5 Analyse von Infrarotbildern

Eine geometrische Kalibrierung mit aktiven und passiven Infrarotmarken wurde am Fachgebiet Photogrammetrie und Fernerkundung von Simmler [2009] mit photogrammetrischer Software durchgeführt. Es wurden Bildhauptpunktlage, Kammerakonstante und radialsymmetrische Verzeichnung bestimmt. Ähnlich haben Luhmann et al. [2010] und Lagüela et al. [2011] die Genauigkeiten der geometrischen Kalibrierung von Infrarotkameras untersucht. Gehrke et al. [2013] rekonstruieren manuell aus einer Serie von Infrarotaufnahmen Gebäude mit Texturen im thermischen Infrarot. Sie weisen insbesondere auf die Probleme bei der Interpretation dieser Texturen hin, da die Stärke der Abstrahlung im thermischen Infrarot von vielen Faktoren abhängt und eine eindeutige Interpretation oft nur intuitiv oder mit Umgebungs- und Witterungswissen möglich ist. Borrmann et al. [2013] zeigen erste Versuche der Erzeugung von Laserpunktwolken von Gebäudefassaden, die mit thermischen Intensitätswerten aus Infrarotbildern angereichert werden. Dabei werden beide Systeme gemeinsam auf einem Träger montiert und mit einer bekannten Szene kalibriert und koregistriert. Die Koregistrierung geschieht über die Punktwolken des Laserscanners und eine aus den Infrarotbildern mittels der Software Bundler erzeugten Punktwolke. Die Kombination von einfachen Gebäudemodellen und Infrarotbildern von Satelliten und Flugzeugen wird

bei Nicol & Wong [2005] eingesetzt, um den Einfluss der Bebauung auf urbane Wärmeinseln zu untersuchen. Dabei spielt neben Material und Höhe auch die Ausrichtung von Straßenschluchten als Windkorridore eine entscheidende Rolle.

1.2.6 Integration weiterer Datenquellen

Trennt man die Erstellung des Gebäudemodells von der Extraktion der Texturen, so können für die Erstellung eines Gebäudemodells weitere Quellen genutzt werden. Solche Modelle können durch klassische Vermessung oder mittels Luftbildphotogrammetrie stereoskopisch (Krauss et al. [2007]) aufgenommen werden. Während bei einem aufwendig klassisch vermessenen Gebäude die Genauigkeit im Bereich von Millimetern liegt, ist diese bei der Rekonstruktion aus Stereoluftbildern insbesondere in der Höhe von Gebäuden deutlich ungenauer. Darüber hinaus ist es möglich, 3D-Punktwolken von Gebäudefassaden durch Laserabtastung (Rottensteiner et al. [2005]; Sithole & Vosselman [2005]; Hebel & Stilla [2009]) oder terrestrisch aufgenommene Bilder (Nistér [2000]; Mayer [2007]; Pollefeys et al. [2008]; Heinrichs et al. [2008]) zu erfassen. Hier können auch kleine Details der Fassade wie Fenster geometrisch erfasst und in das Modell integriert werden. Für die Extraktion von Fassaden aus luftgetragenen Laserdaten (LiDAR) greifen Hebel & Stilla [2009] auf Aufnahmen in Schrägsicht aus mehreren Richtungen zurück. Die entstandenen Punktwolken der einzelnen Überflüge werden koregistriert, indem in den einzelnen Punktwolken Flächen gesucht und eine Zuordnung dieser Flächen durchgeführt wird.

Liegen die innere und äußere Orientierung der Kamera und ein ausreichend detailliertes Gebäudemodell vor, so kann eine direkte Zuordnung der Bilder über die Kanten eines Polygonmodells durchgeführt werden. Ein Beispiel für ein solches Verfahren stellt das Tracking-Verfahren der EPFL in Lausanne dar (Lepetit et al. [2004]). Es bezieht Vorwissen eines Polygonmodells in die Positionsbestimmung mit ein. Eine virtuelle Kamera macht Aufnahmen des Polygonmodells. Diese Bilder werden gespeichert und mit den Aufnahmen der echten Kamera verglichen. Aus der bekannten Position der virtuellen Kamera wird die Position der echten Kamera abgeleitet. Aus den gespeicherten Ansichten werden die ähnlichsten zum aktuellen Bild herausgesucht und die Position der Kamera zwischen ihnen interpoliert. Dadurch liegt die Kameraposition im Koordinatensystem des Polygonmodells vor. Eine solche direkte Zuordnung für Thermalbilder wird für Luftbilder im mittleren thermischen Infrarot (MWIR) in Iwaszczuk et al. [2012] vorgeschlagen. Dort wird für jede Fassade aus einem einzelnen Luftbild, das die gesamte Fassade zeigt, die entsprechende Textur extrahiert. Bevor die Zuordnung der Modellkanten und der Kanten des Bildes durchgeführt werden kann, werden die äußere und innere Orientierung wie von Stilla et al. [2009] vorgeschlagen über Passpunkte am Gebäude verbessert. Eine Übertragung dieser direkten Zuordnung von Einzelbildern und dem Modell ist für terrestrische Thermalbilder in Hoegner et al. [2007] untersucht worden. Es zeigt sich, dass die sichtbaren Teile eines Gebäudes meist nicht ausreichen, um eine zufriedenstellende Zuordnung mit den Ecken und Kanten des 3D Modells durchzuführen.

1.2.7 CityGML

Als Standard für die Beschreibung von Gebäudemodellen hat sich das Format CityGML (Gröger et al. [2010]) durchgesetzt, das unterschiedliche Detailgrade von Gebäuden und Stadtszenen innerhalb einer Datenbank definiert. Der deutlichste Vorteil von CityGML gegenüber klassischen Speicherformaten wie vrml oder x3d, die lediglich Geometrie speichern, liegt in der Integration von Topologie und Semantik. Eine Wand ist nicht einfach ein Polygon, sondern gehört zu einem Gebäudeteil, der wiederum zum Gesamtgebäude gehört. Ein Fenster ist ein Teil dieser Wand. Diese Topologie muss manuell oder automatisch erzeugt werden (Nagel et al. [2009]). Werden nachträglich Objekte hinzugefügt, muss deren Topologie vom Elternelement abgeleitet werden. Im Falle der Extraktion von Geometrie aus einer Fassadentextur ist das Elternelement durch die

Fassade definiert, der die Textur zugeordnet ist. Die Semantik erlaubt die Attributierung von Objekten. Einer Fassade können Eigenschaften zugewiesen werden wie z.B. Durchschnittstemperatur. Es können auch zeitliche Attribute eingesetzt werden. So kann ein Gültigkeitszeitraum für Geometrie oder Texturen definiert werden.

1.3 Ziele der Arbeit

Terrestrische Aufnahmen von Fassaden werden, zumindest in urbanen Gebieten, meist mit relativ kurzen Abständen von der Straße aus aufgenommen. Dadurch entsteht das Problem, dass ganze Fassaden oft nicht in einem einzelnen Bild abgebildet sind. Darüber hinaus sind vorhandene Gebäudemodelle, die in dieser Arbeit sowohl als Referenz im Objektkoordinatensystem verwendet werden wie auch als Struktur in einem Geoinformationssystem, geometrisch einfach gehalten. Neben den Fassadenkanten und der Dachform sind kaum geometrische Elemente modelliert. Alle feineren Teilobjekte wie Türen und Fenster und erst recht Fassadenelemente wie Vorsprünge oder Fallrohre sind, wenn überhaupt nur als Texturen vorhanden. Das hat zur Folge, dass eine direkte Georeferenzierung der einzelnen Bilder über Passpunkte am Gebäude in der Regel nicht möglich ist. Deshalb wird in dieser Arbeit die Strategie verfolgt, die einzelnen Bilder zunächst untereinander zu orientieren, um die gesamte Bildsequenz dann dem Gebäudemodell zuordnen zu können.

Die vorliegende Arbeit stellt im Hinblick auf diese Anforderungen einen Schritt hin zur automatisierten Auswertung dar: Über ein 3D Gebäudemodell können mehrere Aufnahmen einer radiometrisch und geometrisch kalibrierten Infrarotkamera zu großflächigen Texturen zusammengesetzt und mit einem Raumbezug abgelegt werden. Über Verfahren der Bildverarbeitung lokalisierte Objekte wie Fenster und Leckagen erhalten ebenfalls einen Raumbezug, so dass es möglich wird, diese mit Detektionen aus dem Innenbereich zu vergleichen, um eine manuelle Interpretation des Wärmeflusses und möglicher Ursachen für Undichtigkeiten zu erleichtern. Wetterdaten können als Metadaten einer Aufnahmesequenz mit gespeichert werden und die helfen, die Vergleichbarkeit von Aufnahmen verschiedener Zeitpunkte zu verbessern.

Das Ziel dieser Arbeit ist die Integration und Bewertung von Methoden zur Orientierung von Bildsequenzen zur automatischen Extraktion und Analyse von Texturen und ihre Anpassung an die speziellen Eigenschaften von Bildern, die im thermischen Infrarotspektrum aufgenommen werden.

- Hierfür ist zu untersuchen, ob und mit welchen Genauigkeiten die geometrische Kalibrierung von thermischen Infrarotkameras auf Basis vorhandener Verfahren der Photogrammetrie durchgeführt werden kann.
- Es soll bewertet werden, ob und unter welchen Voraussetzungen mit welchen Genauigkeiten aus der Bildverarbeitung bekannte Operatoren zur Extraktion markanter Punkte – hier am Beispiel von Förstner-Punkten und SIFT-Features – auf thermische Infrarotbilder angewendet werden können. Hierbei soll untersucht werden, welche Zahl an markanten Punkten bei verschiedenen Szenarien detektiert und in einer Sequenz verfolgt werden können und ob sie damit als Basis für die Orientierung von Bildsequenzen geeignet und ausreichend genau bestimmbar sind.
- Es sollen zwei Verfahren zur Orientierung von Bildsequenzen entwickelt und bewertet werden. Im zweistufigen Verfahren soll zunächst die Genauigkeit einer relativen Orientierung ohne Szenenwissen untersucht werden. Anschließend wird die Qualität einer Koregistrierung

der relativ orientierten Bildsequenz mit einem gegebenen Gebäudemodell vorgestellt und bewertet. Für das einstufige Orientierungsverfahren wird eine Integration des Vorwissens aus Kamerapfad und Szene in die Orientierung der Bildsequenz entwickelt und bewertet. Beide Verfahren sollen anschließend verglichen werden.

- Auf Basis der orientierten Bildsequenzen sollen für verschiedene Szenarien Fassadentexturen extrahiert und bewertet werden.

Um eine sowohl großflächige als auch detailreiche Anreicherung der Gebäudedatenbasis zu erhalten, sind Probleme der automatischen Zuordnung und Kombination von Bilddaten zu 3D Vektordaten und der durchzuführenden Objektextraktion zu lösen. So erschwert z.B. der kleine und niedrig aufgelöste Bildausschnitt einer Infrarotkamera die Zuordnung auf ausgedehnten Strukturen.

Zur Zuordnung von Bilddaten zu Gebäudeteilen werden Verfahren untersucht, die die besonderen Eigenschaften von Infrarotbildern bezüglich der Objekterscheinung berücksichtigen. Mit diesen Verfahren werden aus den Bildern Strukturen extrahiert und mit denen der Gebäude verglichen. Es ist zum einen von ungenauen Projektionsparametern auszugehen, die z.B. zu einem Versatz führen, und zum andern von einer unvollständigen Zuordnung, da nicht alle Bildstrukturen eine Entsprechung im Gebäudemodell und umgekehrt haben.

Zur Extraktion von Signaturen werden Objektstrukturen und -attribute aus den Bilddaten bestimmt werden. Es wird untersucht, welche Strukturen und Attribute besser direkt aus den Infrarotbildsequenzen gewonnen werden können und welche besser aus flächenbezogenen, entzerrten Texturen, die aus den gemittelten und projizierten Bildsequenzen erzeugt und häufig aus mehreren Einzelbildern zusammengesetzt werden. Es werden Konzepte entwickelt, wie Bilddaten unterschiedlicher Infrarotsensoren fusioniert, unterschiedlicher Aufnahmerichtungen (z.B. bodenbasiert und luftgestützt) kombiniert, unterschiedlicher Zeitpunkte interpoliert und unvollständiger Erfassung extrapoliert werden können.

1.4 Aufbau der Arbeit

Im Anschluss an die Einleitung werden in Kapitel 2 die mathematischen Grundlagen für die Verarbeitung der Bilddaten von der geometrischen Kalibrierung, über die Extraktion homologer Punkte bis zur relativen Orientierung von Bildsequenzen erläutert.

Zur Erreichung der genannten Ziele wird in Kapitel 3 eine Verarbeitungskette zur Orientierung von Bildsequenzen aus dem thermischen Infrarotspektrum entwickelt. Im einzelnen gliedert sich die Verarbeitung in folgende Schritten:

- Bestimmung der inneren Orientierung: Es wird nachgewiesen, dass Verfahren zur optischen Kalibrierung von Kameras zur Bestimmung der inneren Orientierung auf thermische Infrarotkameras (TIR Kameras) übertragen werden können. Dies ist die Voraussetzung sowohl für die Bestimmung der relativen Orientierung wie auch der Texturextraktion aus Bildern.
- Automatische Extraktion von markanten Punkten: Für die automatische Bestimmung der relativen Orientierung von Bildern müssen markante Punkte automatisch bestimmt werden. Dazu werden gängige Verfahren untersucht und hinsichtlich ihrer Anwendbarkeit für thermische Infrarotbilder bewertet.
- Bestimmung der relativen und absoluten Orientierung: Für die Bestimmung der relativen Orientierung werden die zuvor extrahierten markanten Punkte eingesetzt und zusammen mit

der inneren Orientierung die relative Orientierung von Bildsequenzen bestimmt. Durch Koregistrierung mit dem vorhandenen Gebäudemodell und gemessenen Kameraorientierungen wird die Bildsequenz in ein globales Koordinatensystem überführt. Es wird nachgewiesen, dass unter der Annahme, dass die Punkte 1 und 2 erfüllt sind, auch der Übergang von 2D Bildkoordinaten in 3D Objektkoordinaten möglich ist. Auch hier findet eine Bewertung der erreichten Genauigkeit statt.

- Neben dem zweistufigen Orientierungsverfahren wird ein Verfahren vorgestellt, dass die relative Orientierung der Bildsequenz und die Koregistrierung mit dem 3D Gebäudemodell in einem einstufigen Prozess kombiniert.

Auf Basis der koregistrierten Bildsequenz wird in Kapitel 4 ein auf die spezielle Aufnahmegeometrie angepasstes Verfahren zur Extraktion von Fassadentexturen vorgestellt. Dabei werden zwei Verarbeitungsschritte erläutert:

- Extraktion von Texturen aus Bildsequenzen: Durch die Herstellung des Bezuges zwischen Bildern im thermischen Infrarot und gegebenen Modell können Texturen aus den Bilddaten extrahiert werden. Auch hier wird die Übertragbarkeit vorhandener Verfahren für Bilder im sichtbaren Licht auf Bilder im thermischen Infrarotspektrum überprüft.
- Analyse von Texturen im thermischen Infrarot: Es werden Verfahren zur Bildanalyse untersucht, um in Infrarottexturen heiße ('hot spots') und kalte ('cold spots') Stellen auf Fassaden zu lokalisieren. Die im Bildraum der Textur gefundenen Merkmale werden anschließend in den 3D-Objektraum übertragen und dem vorhandenen Modell hinzugefügt. Dies wird beispielhaft an der Detektion von Fenstern und Heizungsleitungen erläutert. Außerdem werden neben geometrischen Informationen thermale Eigenschaften der Fassade extrahiert, hier zum Beispiel die Fassadentemperatur. Zuletzt werden Kombinationen von Texturen derselben Fassade untersucht, um Informationen über zeitliche Veränderungen zu generieren.

Kapitel 5 beschreibt die gewählte Testumgebung, sowie die verwendeten Kameras und die gewählte Aufnahmegeometrie. In Kapitel 6 werden die Ergebnisse der Kamerakalibrierung, die Genauigkeiten der Extraktion der homologen Punkte durch Förstner-Operator und SIFT-Features, sowie die Qualität des zweistufigen und des einstufigen Orientierungsverfahren und die daraus resultierende Texturextraktion bewertet. Im Anschluss werden Ergebnisse der Objektdetektion und Texturkombination erläutert.

Kapitel 7 bewertet die vorgestellten Ergebnisse und erläutert Vorteile und Einschränkungen der gewählten Verfahren. Kapitel 8 gibt einen Ausblick auf mögliche weitere Untersuchungen im Kontext von Bildsequenzen und Texturen im thermischen Infrarotspektrum.

2 Grundlagen

In diesem Kapitel geht es um die Grundlagen, auf denen die vorzustellenden Orientierungsverfahren aufbauen. Dazu gehört neben den besonderen Bedingungen von Thermalaufnahmen (Abschnitt 2.1) und der Kalibrierung der Kameras (Abschnitt 2.2) die Auswahl von Detektoren für markante Punkte (Abschnitt 2.3), die verwendet werden, um Bildpaare und Bildsequenzen relativ zueinander zu orientieren (Abschnitt 2.4).

Bestehende Verfahren zur relativen Orientierung von Bildsequenzen und Extraktion von Texturen sind auf hochaufgelöste Bilder im sichtbaren Spektrum ausgerichtet. Daher ist es zunächst wichtig, die Eigenschaften thermischer Infrarotstrahlung zu untersuchen, um zu klären, in wie weit klassische Verfahren aus dem sichtbaren Spektrum übertragbar sein können.

2.1 Thermische Infrarotstrahlung

2.1.1 Physikalische Grundlagen

Im sichtbaren Spektrum hängt die von einem Objekt ausgesandte Strahlung von dessen Oberflächeneigenschaften ab, die verschiedene Wellenlängen des Spektrums unterschiedlich stark reflektieren und so einen Farb- und Helligkeitseindruck erzeugen. Die von einem Sensor gemessenen Intensitätswerte in verschiedenen Spektralbereichen hängen daher von den Reflexionseigenschaften und der Beleuchtung eines Objektes ab. Die Strahlungsintensität im thermischen Infrarotspektrum hingegen hängt von der inneren Energie eines Körpers (seiner Temperatur) und seinem Emissionsgrad ab. Jeder Körper strahlt abhängig von der in ihm gespeicherten Wärmeenergie elektromagnetische Strahlung aus. Je höher die Temperatur ist, desto energiereicher und kurzwelliger wird die Verteilung der Strahlung. Diesen Zusammenhang beschreibt die Planck'sche Strahlungsformel (Planck [1923]). Ab einer bestimmten Temperatur strahlt ein Objekt auch im sichtbaren Spektrum. Gleichzeitig nimmt die gemessene Strahldichte in jedem Wellenlängenbereich, in dem ein Objekt strahlt, mit der Temperatur zu. Das Wien'sche Verschiebungsgesetz (Wien [1896]) beschreibt diesen Zusammenhang für einen idealisierten schwarzen Körper, bei dem davon ausgegangen wird, dass sein Emissionsgrad 100% beträgt. Das bedeutet, dass die Temperatur eines Objektes anhand der Strahldichte in einem Wellenlängenintervall unter der Voraussetzung bestimmt werden kann, dass sein Emissionskoeffizient bekannt ist. Es bedeutet darüber hinaus, dass zwei Objekte in einer Thermalaufnahme nur dann unterscheidbar sind, wenn sie auf Grund einer unterschiedlicher Temperatur oder eines unterschiedlichen Emissionsgrades eine unterschiedliche Strahldichte haben.

Im Infrarot Spektrum liegen die Emissionsgrade fast aller Materialien bei über 90%. Das heißt, dass sich, anders als im sichtbaren Spektrum, Körper nicht durch ihre Oberflächenbeschaffenheit unterscheiden, sondern hauptsächlich durch ihre Temperatur. Eine Ausnahme bilden z.B. blank polierte Metallflächen und Glas. Sie wirken spiegelnd, haben also eine hohe Reflexion und eine geringe Absorption. Damit ist auch ihre Emission entsprechend gering, so dass im Infrarot solche Flächen hauptsächlich die Umgebung widerspiegeln. Eine ähnliche Einschränkung ist für Glas

zu machen. Es ist fast durchsichtig, hat also einen hohen Transmissionsgrad. Dadurch ist die Absorption geringer und mit ihr auch die Emission.

Der sichtbare Bereich des elektromagnetischen Spektrums beim Menschen liegt bei Wellenlängen zwischen $0,38\ \mu\text{m}$ für Blau bis $0,78\ \mu\text{m}$ für Rot. Bei einer Temperatur von $5700\ ^\circ\text{C}$, der Temperatur der Sonnenoberfläche, empfinden Menschen die Mischung der Frequenzen als Weiß. Unterhalb des roten sichtbaren Spektrums bei $0,78\ \mu\text{m}$ beginnt das Infrarotspektrum. Die Atmosphäre besitzt abhängig von der Wellenlänge eine unterschiedlich starke Transmission bzw. Absorption. Das bedeutet, dass unterschiedliche Spektralbereiche unterschiedlich stark durch die Atmosphäre gedämpft werden. Daher wird das infrarote Spektrum in drei Bereiche gegliedert, in denen die Atmosphäre einen hohen Transmissionsgrad aufweist, das nahe Infrarot (NIR) von $1\text{--}3\ \mu\text{m}$, das mittlere Infrarot (MWIR) von $3\text{--}5\ \mu\text{m}$ und das ferne Infrarot von $5\text{--}1000\ \mu\text{m}$. Der Bereich zwischen $3\ \mu\text{m}$ und $14\ \mu\text{m}$ ist für Wärmebildgeräte besonders wichtig und wird als thermisches Infrarot (TIR) bezeichnet, wobei das Teilintervall von $7\text{--}14\ \mu\text{m}$ als langwelliges Infrarot (LWIR) bezeichnet wird. Für thermische Infrarotmessungen wird sowohl der Bereich zwischen $3\text{--}5\ \mu\text{m}$ wie auch der Bereich zwischen $7\text{--}14\ \mu\text{m}$ verwendet. Zu beachten ist lediglich, dass im Bereich von $3\text{--}5\ \mu\text{m}$ Oberflächen noch einen reflektierenden Anteil des Sonnenlichts beinhalten, im Bereich von $7\text{--}14\ \mu\text{m}$ hingegen kaum Sonnenlicht auf die Oberfläche trifft und daher die von einem Objekt ausgesandte Strahlung nur die Eigenstrahlung auf Grund der eigenen Wärme beinhaltet. Die atmosphärische Dämpfung ist neben der Wellenlänge auch von der Entfernung abhängig. Für den Nahbereich mit Abständen von wenigen Metern kann der atmosphärische Einfluss aber vernachlässigt werden.

Objekte unterscheiden sich im Infrarot wie in einem Graustufenbild nur über unterschiedliche Intensitäten und sind umso besser voneinander zu trennen, je höher ihr Kontrast ist. Der Kontrast zweier Objekte unterschiedlicher Temperatur ΔT in einem Bild ergibt sich aus dem Unterschied der Strahlungsleistung ΔL der beiden Objekte im Wellenlängenbereich des Detektors. Der Kontrast wird daher definiert als Quotient $\Delta T/\Delta L$ (Zissler [1993]). Dieser Quotient ist jedoch für verschiedene Temperaturen und Wellenlängenbereiche nicht linear. Ein höherer Quotient bedeutet einen besseren Kontrast bei gleichem Temperaturunterschied. Während für normale Umgebungstemperaturen um $300\ \text{K}$ die Steigung im Wellenlängenbereich von $8\text{--}10\ \mu\text{m}$ größer ist, ist dies beim Vergleich eines Körpers mit $300\ \text{K}$ und einem Körper mit $600\ \text{K}$ im Wellenlängenbereich von $3\text{--}5\ \mu\text{m}$ der Fall. Das heißt, dass das Anwendungsgebiet entscheidet, in welchem Spektralbereich der verwendete Sensor die Strahldichte messen sollte. Für Gebäudefassaden kann von einer Temperatur im Bereich $300\ \text{K}$ ausgegangen werden. Daher wird hier üblicherweise der Spektralbereich von $7\text{--}14\ \mu\text{m}$ verwendet.

Es kann zwischen warmen und kalten Strahlungsquellen unterschieden werden, die sich warm bzw. kalt von ihrer Umgebung abheben. Warme Strahlungsquellen sind z.B. Lampen, spezielle Laser oder die Sonne. Eine kalte Strahlungsquelle ist beispielsweise ein wolkenloser kalter Himmel. Der Einfluss der Strahlung der Sonne mit einer Oberflächentemperatur von $6000\ \text{K}$ reicht bis ins Infrarotspektrum. Im Wellenlängenbereich von $3\text{--}5\ \mu\text{m}$ ist sie zwar um 2 Größenordnungen geringer als im sichtbaren Bereich, bestimmt aber einen großen Teil der Strahlung im mittleren Infrarot. In manchen Szenen verstärkt das unterschiedlich stark reflektierte Tageslicht den Kontrast im Infrarot, in anderen Szenen löscht es den Unterschied aus. Es findet im mittleren Infrarot also eine direkte Beeinflussung der Szene durch die Sonneneinstrahlung statt. Im langwelligeren Infrarot von $8\text{--}14\ \mu\text{m}$ liegt die Strahlungsleistung der Sonne um 4 Größenordnungen niedriger gegenüber dem Maximum. Der direkte Einfluss dieses Spektralbereichs ist dementsprechend in der Szene sehr gering. Durch den Energieeintrag in anderen Spektralbereichen, vor allem im sichtbaren Spektrum und im nahen Infrarot, wärmt sich ein Körper in der Sonne auf. Die erhöhte Temperatur führt zu einer erhöhten Abstrahlung im thermischen Infrarot.

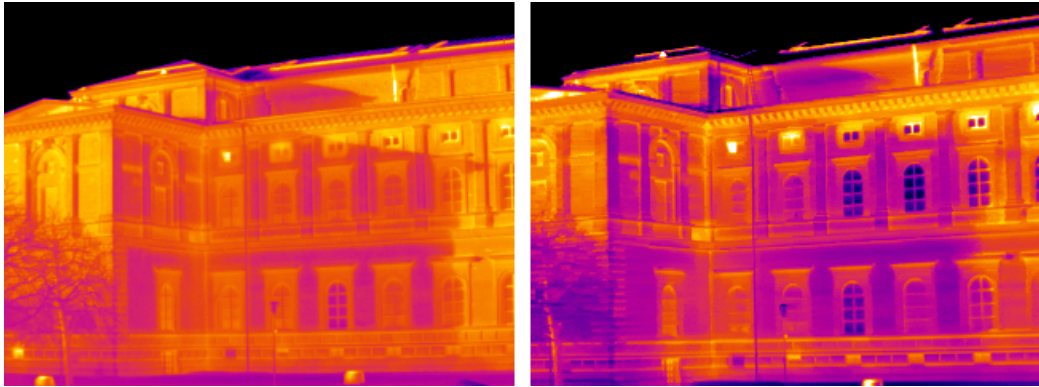


Abbildung 2.1: Aufnahme der Alten Pinakothek in München. links: Infrarotaufnahme im mittleren Infrarot, deutlich zu sehen der Schlagschatten der Sonne; rechts: Dieselbe Aufnahme im langwelligigen Infrarot. Zu sehen ist hier die unterschiedliche Aufwärmung der Fassade in der Sonne und im Schatten. Oben schwarz: der kalte Himmel

Abbildung 2.1 zeigt zwei Aufnahmen derselben Szene zum selben Aufnahmezeitpunkt. Im linken Bild ist der starke Schlagschatten des Sonnenlichts im mittleren Infrarot zu erkennen. Im rechten Bild ist diese Grenze im langwelligigen Infrarot verwischt. Sie zeigt nicht die Reflexion des Sonnenlichts, sondern die niedrigere Oberflächentemperatur im Schatten an. In urbanen Szenen treten häufig Straßenlaternen, Zimmerbeleuchtungen und Scheinwerfer von Fahrzeugen auf. Ihre Strahlungscharakteristik entspricht maximal 3000 K und liegt damit nach dem Stefan-Boltzmann'schen Gesetz weit unterhalb der der Sonne. Solche Objekte gelten als Punktstrahler, wenn ihre Abbildung kleiner als ein Detektorelement wird. Die eintreffende Strahlungsleistung nimmt dann mit dem Quadrat der Entfernung ab. Auf der anderen Seite stehen Flächenstrahler, die mehrere Detektorelemente abdecken. Ihre Strahlungsleistung auf einem Detektorelement bleibt konstant.

2.1.2 Detektoren für thermische Infrarotstrahlung

Für Infrarotstrahlung gelten dieselben optischen Gesetze wie für elektromagnetische Strahlung im sichtbaren Spektrum, also die Gesetze für Linsen, Spiegel, Bündelbegrenzung und Beugung. Die einfallende Strahlung wird über ein Linsensystem auf einen Detektor gelenkt. Physikalisch bedingt gibt es dennoch einige kleinere Unterschiede (Fischer [1992]).

Geometrie

Im Infraroten Spektrum müssen andere Materialien für Linsen verwendet werden als in sichtbarem Spektrum. Während der Brechungsindex dort bei 1,5 bis 2,0 liegt, basieren infrarot-durchlässige Optiken auf Materialien wie Germanium und Zinksulfid mit einem Brechungsindex von 4,0 bzw. 2,2. Diese höheren Brechungszahlen sorgen für kleinere Abbildungsfehler, erhöhen aber die unerwünschte Reflexion an der Linsenoberfläche. Das erfordert eine größere Sorgfalt bei der Reflexionsverminderung durch Vergütungsschichten und führt zu einer größeren Empfindlichkeit der Objektive. Diese Eigenschaften bestimmen auch die technischen Grenzwerte für die optische Auflösung (Stöbl [1993]). Bei einer Wellenlänge von $10\text{ }\mu\text{m}$ und einer Eintrittspupille vom 200 mm ergibt sich eine Winkelauflösung von 0,06 mrad oder $0,0035^\circ$. Um zu verhindern, dass das Beugungsbild eines Pixels auch die Nachbapixel beeinflusst, wird die Mindestkantenlänge eines Pixels auf $31,8\text{ }\mu\text{m}$ festgelegt. Auf Grund der Begrenzung der Detektorgröße folgt daraus eine technische Beschränkung der geometrischen Auflösung eines Matrixdetektors. Damit hinken Detektoren im thermischen Infrarot mit einer Auflösung von 320×240 Pixeln bis zu 1024×768 Pixel den Kame-

ras im sichtbaren Spektrum mit mehreren Megapixeln deutlich hinterher. Aus oben genannten Gründen kommen neben Matrixdetektoren auch Zeilendetektoren zum Einsatz, bei denen über ein Prisma oder einen Spiegel das Bild zeilenweise erzeugt wird. Diese Systeme erreichen zwar eine deutlich höhere geometrische Auflösung, haben aber durch die zeilenweise Aufnahme eine deutlich geringere zeitliche Aufnahmefrequenz.

Radiometrie

Neben der optischen geometrischen Auflösung spielt auch die radiometrische Auflösung eine wichtige Rolle. Dabei unterscheidet man zwischen der eigentlichen radiometrischen Auflösung, die die kleinste messbare Strahlungsdifferenz angibt, und der absoluten Messgenauigkeit der Strahlungsleistung. Die radiometrische Auflösung wird vor allem durch äußere elektromagnetische Strahlungseinflüsse (z.B. Eigenstrahlung von Kamera und Detektor) und elektrische Störeinflüsse (Rauschen) begrenzt. Die Störquellen des Detektors selbst erzeugen Rauschspannungen am Detektorausgang (Rogatto [1993]). Thermisches Rauschen entsteht durch das Schwingen von Atomen und ihren Elektronen auch bei tiefen Temperaturen. Die Bewegungen der Ladungsträger sind nur im Mittel gerichtet und konstant. Hier spricht man von Schrotrauschen. Außerdem kann die Anzahl der Ladungsträger im Detektor schwanken. Darüber hinaus beeinflusst $1/f$ -Rauschen die Messwerte, das mit der Frequenz abnimmt. Um den Einfluss des Rauschens auf das Messergebnis zu minimieren, gibt es drei wesentliche Ansätze. Bei bekannten Rauschverhalten kann eine nachgelagerte Signalverarbeitung über digitale Filter das Rauschen minimieren. Zum zweiten kann über eine Kühlung des Detektors der Störeinfluss freier Ladungen reduziert werden. Die dritte Möglichkeit ist eine längere Integrationszeit.

Die absolute Messgenauigkeit ist abhängig von den äußeren Bedingungen wie atmosphärischer Dämpfung, von den Emissionsgraden der beobachteten Objekte und von der Umrechnung des aufgezeichneten Strahlungsflusses im Detektor in elektrische Signale. Die Emissionsgrade werden für Fassaden mit 95-97 % angenommen. Die Kamera selbst muss radiometrisch kalibriert werden (Mermelstein et al. [2000]). Hierfür werden zwei Schwarzkörper unterschiedlicher, bekannter Temperatur verwendet und aus den gemessenen Strahlungsleistungen sowie deren Differenz eine Kalibrierung für die Temperaturbestimmung durchgeführt. Des Weiteren werden die produktionsbedingten Inhomogenitäten der Empfindlichkeit der einzelnen Detektorelemente korrigiert.

Detektortypen

Neben der Frage, ob als Detektor ein Zeilenscanner oder ein Matrixdetektor verwendet wird, unterscheiden sich die Detektoren auch im Typ. Sie werden in zwei Gruppen eingeteilt: Thermische Detektoren, die über eine Temperaturänderung des Detektors die Strahlungsleistung des Elements bestimmen, und photonische Detektoren, die direkt die eintreffenden Ladungen messen. Thermische Detektoren bestimmen die Strahlungsleistung am Sensor indirekt über die Abhängigkeit physikalischer Eigenschaften von der Temperatur. Thermische Detektoren können nach Vincent [1990] in folgende Klassen eingeteilt werden: Bolometer nutzen die Abhängigkeit des elektrischen Widerstands eines Metalls oder Halbleiters von der Temperatur. Pyroelektrische Detektoren messen die Veränderung der elektrischen Ladung, die durch die Veränderung der Polarisation beim Aufwärmen oder Abkühlen entsteht. Diese Systeme erreichen eine thermische Auflösung von 0,03 Kelvin. Die Erwärmung eines Detektorelements beeinflusst auch benachbarte Elemente, so dass kleine, intensive Punktstrahler, die eigentlich nur ein Detektorelement bestrahlen, als verschmierter Fleck anstatt als Punkt auftreten können. Die Elemente ungekühlter Detektoren benötigen eine größere Fläche von etwa $40 \times 40 \mu\text{m}$, was wiederum die Auflösung im Vergleich zu gekühlten Detektoren reduziert. Durch den Wegfall der aktiven Kühlung sind die Geräte jedoch deutlich mobiler, da die Geräte ohne aktive Kühlung kompakter gebaut werden können.

Photonische Detektoren messen die elektrische Ladung von auftreffenden Photonen. Sie sind daher schneller und effizienter als thermische Detektoren. Als Photodetektor dient ein Halbleiter. Die Energiezufuhr durch die einfallende Strahlung hebt ein Elektron auf eine höhere Bahn. Das ändert die elektrischen Eigenschaften des Halbleiters in Abhängigkeit von der Strahlungsenergie. Es existieren zwei verschiedene Methoden, diese Änderung zu messen. Der Halbleiter kann als Photowiderstand verwendet werden. In diesem Fall ändert sich durch die Strahlung der Widerstand des Detektors. Um diese Änderung zu messen, fließt ein Messstrom durch den Detektor. Dieser wärmt das Element zusätzlich auf. Somit wird eine höhere Kühlung notwendig. Ein Photowiderstand reagiert recht träge und ist daher nur für relativ niedrige Frequenzen einsetzbar. Die andere Möglichkeit ist die Verwendung der Halbleiter als Photodiode. Sie reagieren schneller und beeinflussen die umgebenden Detektorelemente nicht. Auch die innere Wärmeenergie des Detektors kann Elektronen vorübergehend auf eine höhere Bahn heben und so eine Rauschspannung am Detektor erzeugen. Diese Systeme müssen daher gekühlt werden, um den Einfluss, den fremde Photonen, die von anderen Detektorelementen oder der Kamera selbst ausgestrahlt werden, zu reduzieren. Diese Detektoren können mit einer kleinsten messbaren Temperaturdifferenz von 0,01 Kelvin eine höhere thermische Auflösung erreichen als die Bolometer und kommen bei entsprechender Kühlung auch mit deutlich kürzeren Integrationszeiten aus, da direkt einzelne Photonen gemessen werden und nicht auf den indirekten Effekt durch die Temperaturänderung gewartet werden muss.

Das Auslesen der Messwerte der einzelnen Elemente der Matrix ist abhängig von der Integrationszeit. Um die Messung der Intensitätswerte am Sensor möglichst genau zu machen, ist eine möglichst lange Integrationszeit zu wählen. Dem gegenüber steht der Wunsch nach einer hohen Bildaufnahmefrequenz oder einer kurzen Belichtungszeit für dynamische Szenen. Bei Zeilenscannern werden die Detektorelemente über das Bild geschoben, wobei versucht wird, durch ein Überlappen der aufgenommenen Bereiche eine Unterabtastung und Bildung von Alias-Effekten möglichst zu vermeiden. Bei einer Detektormatrix wird zwischen CCD (Charged-Coupled-Device) und CMOS (Complementary metal-oxide-semiconductor) unterschieden. CCD-Sensoren bestehen aus lichtempfindlichen Photodioden: Ein Detektorelement besteht aus einem dotierten Halbleiter und einer darüber liegenden isolierten transparenten Schicht eines elektrischen Leiters. Dabei werden durch die Energie der einfallenden Photonen Elektronen aus dem Halbleiter gelöst. Dadurch entstehen negative und positive Ladungen. Proportional zur einfallenden Strahlungsleistung werden dabei elektrische Ladungen in den einzelnen Detektorelementen gesammelt. Bei einer Überbelichtung kann es passieren, dass Ladungen eines Elements auf ein anderes Element überspringen. Man spricht dann von Blooming. Nach der Belichtung werden die Ladungen von Sensorelement zu Sensorelement wie in einem Puffer elementweise weitergeschoben und nacheinander ausgelesen. CMOS-Sensoren basieren ebenfalls auf einer Photodiode. Hierbei wird aber die Photodiode direkt mit Feldeffekttransistoren gekoppelt. Die einfallende Strahlung bewirkt eine proportionale Verringerung der Spannung an der Photodiode. Nach der Belichtung wird dieser Spannungswert ausgelesen. Dabei erfolgt das Auslesen einer Bildmatrix zeilenweise. Gegenüber CCD-Sensoren besteht dabei der Vorteil, dass direkt das Spannungssignal jedes Pixels ausgelesen wird, ohne die Ladungen durch eine Detektorzeile verschieben zu müssen. Hierdurch wird der Blooming-Effekt deutlich reduziert. Der Nachteil besteht darin, dass pro Detektorelement eine im Vergleich zu CCD große Fläche für nicht lichtempfindliche Elektronik verwendet wird und daher die Lichtempfindlichkeit bei gleicher Größe eines Detektorelements geringer ist. Zudem tritt durch das zeilenweise Auslesen der sogenannte Rolling-Shutter Effekt auf. Er führt bei schnellen Bewegungen dazu, dass Objekte verzerrt abgebildet werden. Eine senkrechte Linie erscheint dann unter Umständen als schräge Treppenlinie.

2.1.3 Thermografie im Bauwesen

Infrarotkameras haben sich im Bereich der zerstörungsfreien Prüfung (DIN 54190), im Bereich elektrischer Anlagen (DIN 54191) und der Gebäudeprüfung (DIN EN 13187) bewährt. Einen Spezialfall stellt die aktive Thermografie (DIN 54192) dar, bei der das beobachtete Objekt aktiv mit Wärmestrahlung beleuchtet wird, um unterschiedliche Energieaufnahmen auf Oberflächen sowie Temperaturänderungen zu untersuchen. Fouad & Richter [2012] definieren im „Leitfaden Thermografie im Bauwesen“ die wesentlichen Anforderungen und Möglichkeiten der Thermografie für die Gebäudeuntersuchung. Dabei geht es insbesondere um die Interpretation der aufgenommenen Intensitäten.

Der Wärmeverlust an der Oberfläche einer Fassade hängt von den drei physikalischen Effekten Konvektion, Verdunstung und Abstrahlung ab, die wiederum von verschiedenen Einflüssen wie Material, Temperaturdifferenz zwischen Innen- und Außenbereich, Außentemperatur und Wind sowie von benachbarten Objekten abhängen. Dabei beschreibt die Konvektion den Wärmefluss innerhalb eines Objektes. Die Verdunstung beschreibt die Abkühlung einer Oberfläche durch den Wechsel von z.B. Wasser in den energiereicheren gasförmigen Aggregatzustand. Die Abstrahlung schließlich beschreibt die durch den Emissionskoeffizienten eines Materials festgelegte temperaturabhängige Abstrahlung von Wärmeenergie als elektromagnetische Strahlung. So kann eine Betonmauer dieselbe Temperatur aufweisen wie eine Verbunddämmung. Durch den deutlich geringeren Emissionsgrad der Verbunddämmung gegenüber den Beton ist jedoch die Abstrahlung und damit der Wärmeverlust deutlich geringer. Dem Emissionskoeffizienten entsprechend absorbiert jedes Objekt auch einfallende Energie aus der Umgebung. Bei einer geringen Temperaturdifferenz zwischen einer Fassade und der Umgebung kühlt die Fassade daher nur langsam und gleichmäßig aus. Eine hohe Temperaturdifferenz führt zu einer rascheren Abkühlung der Fassade. Ein Baum vor einer Fassade hingegen bestrahlt das hinter ihm liegende Gebäude. Die Abkühlung fällt geringer aus als bei einer frei stehenden Fassade. Wind fördert die Auskühlung von Gebäuden, weil er die aufgewärmte Luft an der Oberfläche eines Gebäudes mitnimmt und so die Temperaturdifferenz an der Oberfläche der Fassade erhöht. Ein klarer Himmel begünstigt zusätzlich die Abkühlung, wohingegen bei bedecktem Himmel eine deutlich geringere erhöhte Abkühlung auftritt. Neben Witterungsbedingungen spielt auch die Geometrie der Szene eine Rolle bei der Interpretation. So strahlen sich benachbarte Fassaden an. Das kann dazu führen, dass sie eine höhere Abstrahlung zeigen als sie dem eigentlichen Wärmeverlust entspricht. Dieser Effekt tritt auch in Innenecken von Gebäuden z.B. in Innenhöfen auf. Dort scheint eine Leckage in einer Ecke vorzuliegen, die in Wirklichkeit nur auf die gegenseitige Bestrahlung zweier Wände zurückzuführen ist. Ein ähnlicher Effekt tritt an Fenstersimsen oder Dachüberständen auf. Dort ist es die aufsteigende, an der Fassade aufgewärmte Luft, die sich unter Vorsprüngen staut und dort zu einer erhöhten gemessenen Intensität führt. Es ist also keinesfalls klar, ob am Dachüberstand eine schlechte Isolierung mit starkem Wärmeverlust vorliegt, was natürlich auch sein kann. Hinzu kommen Materialien mit geringen Emissionskoeffizienten wie Glas, Metall und bestimmte lackierte Dachziegel. Sie zeigen überwiegend die Umgebungsstrahlung, was je nach Ausrichtung der Fläche und Blickrichtung der Kamera der Boden, ein anderes Objekt oder der Himmel sein kann. Hinzu kommen Fälle von hinterlüfteten Fassaden oder Dächern, wo der Wärmeverlust von der Kamera einzig am Austrittspunkt der Luft festgestellt wird, obwohl dort gar nicht die Ursache des Wärmeverlustes liegt.

Häufig lässt sich nur durch den Vergleich von Messungen der Fassaden von außen und in den Räumen die Ursache tatsächlich eingrenzen. Als Beispiel sei hier die Untersuchung zum Wärmeverlust eines Rollladenkastens genannt. Die Untersuchung von außen zeigt eine erhöhte Abstrahlung im Bereich des Kastens. Diese Abstrahlung kann eine Kombination aus Konvektion durch eine Kältebrücke, angestauter aufsteigender warmer Luft und gegenseitiger Bestrahlung von Flächen

sein. Die Untersuchung derselben Fassade von innen heraus zeigt am Rollladenkasten eine tiefere Temperatur an als an der umgebenden Wand. Es liegt also eine Kältebrücke vor.

Aus diesen Randbedingungen ergeben sich die notwendigen Informationen, um auf Basis von Thermalaufnahmen eine Schadensermittlung durchführen zu können:

- Auf Grund der Abhängigkeit der gemessenen Temperatur von der aufgezeichneten Strahlungsleistung muss die verwendete Infrarotkamera radiometrisch kalibriert werden, um Temperaturwerte bestimmen zu können.
- Die Wetterdaten zum Zeitpunkt der Aufnahme müssen bekannt sein, um ihren Einfluss auf die Abstrahlung der Fassade berücksichtigen zu können.
- Die Geometrie und Materialien des Gebäudes müssen bekannt sein, um über die Emissionskoeffizienten der verschiedenen Baustoffe von der gemessenen Strahlungsleistung auf die Temperatur schließen zu können.
- Aufnahme müssen das gesamte Gebäude abdecken.
- Außen- und Innenaufnahmen müssen gemeinsam ausgewertet werden, um

2.2 Geometrische Kalibrierung der Kamera

Handelsübliche Infrarotkameras werden bereits mit einer radiometrischen Kalibrierung ausgeliefert, die sowohl die Inhomogenitäten der Detektorelemente und das Rauschen des Detektors und der Elektronik, als auch die Temperaturkalibrierung beinhalten (Mermelstein et al. [2000]). Da der Schwerpunkt dieser Arbeit auf der Texturierung liegt und nicht auf der möglichst genauen Schätzung der Temperaturen, wird auf eine detaillierte Diskussion der radiometrischen Kalibrierung verzichtet.

Die Parameter der geometrischen Kalibrierung werden meist von den Herstellern nicht mitgeliefert oder können nur als grobe Näherung dienen. Vor einer Aufnahmereihe mit daher eine geometrische Kalibrierung mit den entsprechend festzuhaltenden Kameraeinstellungen neu durchgeführt werden. Die Beschreibung optischer Kamerasysteme folgt dem idealisierten Modell einer Lochkamera (Bachler [2001]) mit der Bildhauptpunktlage $H = (x_0; y_0)$ und der Kamerakonstante c_k . In digitalen Kameras dient die Pixelmatrix des Sensors als Referenzsystem für die Bildkoordinaten. Liegen genug Bilder einer Szene aus unterschiedlichen Positionen und Blickwinkeln vor, kann die innere Orientierung durch eine Bündelblockausgleichung (Kraus [2004]) mit Selbstkalibrierung bestimmt werden. Dabei werden 3D Objektpunkte $(X_i; Y_i; Z_i)$ und ihre korrespondierenden homologen Bildpunkte $(x_i; y_i)$ in den Bildern als Beobachtungen in eine Ausgleichung eingeführt. Als Unbekannte werden die äußeren Orientierungen der n Bilder $(X_0; Y_0; Z_0; \omega_0; \phi_0; \kappa_0)$ und die Parameter der inneren Orientierung $(x_0; y_0; c_k)$ eingeführt. Als Funktionalmodell dienen die Kollinearitätsgleichungen, wobei die a_{ij} die Elemente der aus den Winkeln $(\omega_0; \phi_0; \kappa_0)$ bestimmten Rotationsmatrix R sind.

Für optische Abbildungs- oder Verzeichnungseffekte kann das Kameramodell auf einen Satz von Parametern A_n erweitert werden, die Einflüsse wie radialsymmetrische Verzeichnungen oder Scherungen berücksichtigen und oben in der Verzeichnungskorrektur dx'_{rad} und dy'_{rad} zusammengefasst sind (Gleichung 2.1). Ebner [1976] schlägt hierfür einen Satz von 12 Parametern vor, der die Verzeichnungen an 9 Punkten des Bildes in einer Bündelblockausgleichung korrigiert. Grün [1978] definiert sogar 44 Parameter, die 25 Punkte korrigieren. Unter der Annahme einer qualitativ hochwertigen Optik liegt der überwiegende Einfluss der Abbildungsfehler bei der

radial-symmetrischen Verzeichnung (Steger et al. [2007]). Sie beschreibt die radiale Verzeichnung in einem Abstand $r = \sqrt{x'^2 + y'^2}$ zum Bildhauptpunkt für die unkorrigierten Bildkoordinaten (x', y') . Darüber hinaus ist die Verzeichnung von der Fokussierung und damit der Kamerakonstante abhängig. Sie wird daher für mehrere Entfernungseinstellungen gemessen und dazwischen näherungsweise interpoliert. Die radial-symmetrische Verzeichnung wird durch eine Reihenentwicklung (Luhmann [2003]; Godding [2006]) beschrieben (Gleichung 2.2).

$$\begin{aligned} x_i - x_0 &= c_x \cdot \frac{a_{11}(X_i - X_0) + a_{21}(Y_i - Y_0) + a_{31}(Z_i - Z_0)}{a_{13}(X_1 - X_0) + a_{23}(Y_1 - Y_0) + a_{33}(Z_1 - Z_0)} + dx'_{rad} \\ y_i - y_0 &= c_y \cdot \frac{a_{12}(X_1 - X_0) + a_{22}(Y_i - Y_0) + a_{32}(Z_i - Z_0)}{a_{13}(X_1 - X_0) + a_{23}(Y_i - Y_0) + a_{33}(Z_i - Z_0)} + dy'_{rad} \end{aligned} \quad (2.1)$$

$$\begin{aligned} dr_{rad} &= A_1 \cdot (r^3 - r_0^2 r) + A_2 \cdot (r^5 - r_0^4 r) + A_3 \cdot (r^7 - r_0^6 r) + \dots \\ dx'_{rad} &= \frac{dr_{rad}}{r} \cdot x' \\ dy'_{rad} &= \frac{dr_{rad}}{r} \cdot y' \end{aligned} \quad (2.2)$$

Dieses Modell (Gleichung 2.1 und 2.2) wird in Luhmann et al. [2010] auch auf Infrarotkameras angewandt und auf seine Genauigkeit hin untersucht. Godding [2006] schlägt weitere Korrekturen für radialsymmetrische und tangential Verzeichnungen sowie die Korrektur affiner Verzeichnungseffekte vor.

2.3 Extraktion von homologen Punkten

Für den Einsatz in Bildsequenzen sind automatische Verfahren notwendig, um markante, wohl definierte Punkte, wie sie an Ecken oder isolierten Punkten auftreten, in einem Bild finden zu können und korrespondierenden Punkten in einem anderen Bild zuordnen zu können. Man spricht von homologen Punkten, wenn in mindestens zwei Bildern 2D Bildpunkte einander zugeordnet werden, die denselben 3D Objektpunkt abbilden. Solche Punkte werden über Interestoperatoren bestimmt, die nach Förstner & Gülch [1987] folgende Kriterien erfüllen sollen:

- Ein markanter Punkt soll sich deutlich von seiner lokalen Umgebung unterscheiden.
- Ein markanter Punkt soll gegenüber geometrischen und radiometrischen Änderungen invariant sein.
- Ein markanter Punkt soll robust gegenüber Rauschen sein.
- Ein markanter Punkt soll eindeutig beschreibbar sein und damit selten sein.
- Ein markanter Punkt soll über seine Beschreibung interpretierbar sein.

Interestoperatoren liefern neben der Position eines markanten Punktes einen sogenannten Interestwert, der die Auffälligkeit oder Güte des markanten Punktes angibt. Sie lassen sich grob in die Gruppe der Kanten- und Eckendetektoren und die Gruppe der regionsbasierten Blob-Detektoren unterteilen. Zu den kantenbasierten Operatoren zählen der Moravec-Operator (Moravec [1977]), der Harris-Corner-Detektor (Harris & Stephens [1988]), der Förstner-Operator (Förstner & Gülch

[1987]) oder der Kanade-Lucas-Tomasi (KLT) Detektor (Shi & Tomasi [1994]). Ein Vertreter der regionsbasierten Operatoren ist die Skalen-Invariante Merkmals-Transformation (SIFT) (Lowe [2004]; Brown & Lowe [2007]; Labe & Forstner [2006]).

Auf der Basis der in den Bildern einer Sequenz extrahierten markanten Punkte sollen Punktkorrespondenzen gefunden werden. Hierfur gibt es zwei grundsatzliche Ansatze. In der Luftbildphotogrammetrie (Krauss et al. [2007]) geht man davon aus, dass zwischen zwei benachbarten Aufnahmen einer Sequenz hauptsachlich eine Translation stattfindet und die beobachtete Ebene annahernd eben ist. In diesem Fall kann man ein lokales Bildfenster um einen markanten Punkt herum ausschneiden und in einer Korrelation als Maske uber das zweite Bild schieben. Die Position im zweiten Bild mit dem hochsten Korrelationskoeffizienten ist die wahrscheinlichste Position des markanten Punktes im zweiten Bild. Es ist moglich, durch Vorwissen, wie z.B. die ungefahre Lage der Epipolarlinien, den Suchraum im zweiten Bild deutlich einzuschranken. Dieses Verfahren hat zwei Vorteile. Zum einen ist die Korrelation ein sehr schnelles Berechnungsverfahren, zum anderen benotigt man uber die markanten Punkte hinaus keine weitere Zusatzinformation. Sie ist auch in der Lage, mit anderungen der Helligkeit und des Kontrasts umzugehen. Die Korrelation scheitert jedoch bei groen Skalenanderungen oder Rotationen und bei projektiven und affinen Transformationen. Der normierte Korrelationskoeffizient lasst eine einfache Aussage uber die Gute einer Punktkorrespondenz zu. Die Korrelation eignet sich damit in Fallen mit gleichbleibenden Rotationswinkeln der Kamera und einer Aufnahmefrequenz, bei der in zwei benachbarten Bildern hauptsachlich Translationen der markanten Punkte auf ebenen Flachen vorkommen, wie z.B. in einer Videosequenz aus einem fahrenden Fahrzeug oder einem Flugzeug heraus.

Im Gegensatz dazu hat sich in der Nahbereichsphotogrammetrie ein Verfahren etabliert, bei dem markante Punkte, die fast immer in Form von kontrastreichen Markern in der Szene vorgegeben werden, in ihrer lokalen Umgebung durch abstrahierte Eigenschaften beschrieben werden. Diese Eigenschaften mussen so gewahlt werden, dass sie die Punkte moglichst eindeutig beschreiben und dabei moglichst invariant gegenuber radiometrischen und geometrischen Veranderungen sind.

Fur die folgenden Untersuchungen werden als Detektoren fur markante Punkte exemplarisch der Forstner-Operator (Abschnitt 2.3.1) als Vertreter der Gradientenverfahren und SIFT (Abschnitt 2.3.2) als Blob-Detektor verwendet. SIFT stellt fur extrahierte markante Punkte einen Deskriptor bereit, uber den eine Korrespondenzsuche in benachbarten Bildern durchgefuhrt werden kann.

2.3.1 Forstner-Operator

Anders als der Moravec-Operator (Moravec [1977]), der lokale Grauwertdifferenzen untersucht und deshalb als statistischer Interestoperator bezeichnet wird, gehort der Forstner-Operator (Forstner & Gulch [1987]) zur Gruppe der differentialgeometrischen Interestoperatoren. Er basiert auf der Annahme, dass ein markanter Punkt in einem Bild $B(x, y)$ dort vorliegt, wo starke Intensitatsanderungen in mehreren Richtungen vorliegen. An diesen Stellen zeigt die erste Ableitung der Bildfunktion B' groe Werte. Fuhrt man die partiellen Ableitungen b_x und b_y der Bildfunktion $B(x, y)$ ein, so erhalt man einen Strukturtensor H fur die Bildpunkte $b(i, j)$ in einer lokalen Umgebung um den Bildpunkt $b(x, y)$:

$$H_{x,y} = \begin{bmatrix} \sum_{(i,j) \in \Omega} b_x(i,j)^2 & \sum_{(i,j) \in \Omega} b_x(i,j)b_y(i,j) \\ \sum_{(i,j) \in \Omega} b_x(i,j)b_y(i,j) & \sum_{(i,j) \in \Omega} b_y(i,j)^2 \end{bmatrix} \quad (2.3)$$

Die inverse Matrix H^{-1} entspricht der Kovarianzmatrix, die die Genauigkeit des Interest-Punktes beschreibt. Das bedeutet, dass große Gradienten in A zu kleinen Varianzen in H^{-1} führen. Da die Eigenwerte λ_1 und λ_2 der Kovarianzmatrix H^{-1} die Achsen einer Fehlerellipse beschreiben, werden anschließend die beiden Parameter Größe w und Rundheit q der Fehlerellipse bestimmt:

$$w = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} = \frac{\det(H)}{\text{spur}(H)} \quad w > 0 \quad (2.4)$$

$$q = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 = \frac{4 \cdot \det(H)}{\text{spur}(H)^2} \quad 0 \leq q \leq 1 \quad (2.5)$$

Durch diese Eigenschaften erhält man für gefundene markante Punkte nicht nur eine Aussage über ihre Stärke, den Interestwert, sondern auch über die Art des gefundenen Punktes. Punkte auf Kanten werden im weiteren Verlauf nicht betrachtet, da sie in Folgebildern schwer eindeutig wiederzufinden sind. Die Größe w und die Form q der Fehlerellipse sind invariant gegenüber Translation und Rotationen und uniformen Skalierungen des Bildes. Eine Verschiebung oder Drehung ändert nichts an beiden Werten. Eine uniforme Skalierung wirkt auf alle gefundenen markanten Punkte gleich. Für projektive, affine Transformationen hingegen gilt dies nicht. Der Förstner-Operator eignet sich für Bilder, die keine großen Maßstabsunterschiede oder Rotationen aufweisen.

2.3.2 SIFT

Im Gegensatz zum Förstner-Operator arbeitet die skaleninvariante Merkmalstransformation (SIFT) (Lowe [2004]; Labe & Förstner [2006]; Brown & Lowe [2007]) nicht mit einem Strukturtensor der 1. Ableitung der Bildfunktion, sondern basiert auf der Hesse-Matrix der 2. Ableitung der Bildfunktion in einer geglätteten Bildpyramide.

In einem ersten Schritt wird aus dem Originalbild $B(x, y)$ ein Bildstapel $L(x, y, \sigma)$ erzeugt, wobei jedes Bild der Bildebene mit einer Gaußfunktion $G(x, y, \sigma)$ mit steigendem Glättungsfaktor σ geglättet wird. Je zwei aufeinander liegende geglättete Bilder $L(x, y, \sigma)$ und $L(x, y, k \cdot \sigma)$ werden voneinander abgezogen, um ein Bild der Extrema $D(x, y, \sigma)$ mit einem Difference-of-Gaussian-Operator (DoG) zu berechnen. Ein Punkt gilt dann als Extremum, wenn er in der Achter-Nachbarschaft seiner Glättungsebene und in der Neuner-Nachbarschaft der nächst höheren und niederen Glättungsebene, also insgesamt in der 26er-Nachbarschaft in seiner 3D Umgebung ein Maximum oder Minimum darstellt. Ein Punkt (x, y) in diesem Skalenraum wird auf der Ebene σ dargestellt als $D(x, y, \sigma)$. Auf diese Weise werden durch den Skalenraum hindurch markante Bildteile, sogenannte Blobs, lokalisiert. Dabei können an derselben Position im Bild mehrere Blobs in verschiedenen Skalenebenen auftreten, solange zwischen den Skalenebenen der beiden Blobs mindestens eine weitere Skalenebene ohne Maximum in der 26er-Nachbarschaft liegt.

Im nächsten Schritt wird die Position jedes detektierten Extremums subpixelgenau durch eine quadratische Taylorreihe des Skalenraums interpoliert (Lowe [2004]). Liegt das interpolierte Extremum mehr als 0,5 Pixel neben dem aus dem Skalenraum bestimmten, so liegt es näher am benachbarten Pixel als am aktuell berechneten und wird durch dieses ersetzt. Ebenfalls über die Taylorentwicklung wird der Kontrast eines Punktes im Skalenraum bestimmt und kontrastarme Punkte aussortiert. Um unsichere Punkte auf Kanten auszuschließen, wird die Hessematrix H aus den partiellen zweiten Ableitungen bestimmt, daraus der Formfaktor r berechnet und über einen Schwellwert Punkte auf Kanten aussortiert:

$$H(x, y) = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2.6)$$

$$r = 1 - \frac{\lambda_1 - \lambda_2^2}{\lambda_1 + \lambda_2} = \frac{4 \cdot \det(H)}{\text{spur}(H)^2} \quad (2.7)$$

$$\frac{\text{spur}(H)^2}{\det(H)} < \frac{(r + 1)^2}{r} \quad (2.8)$$

Lowe [2004] schlägt vor, Punkte mit $r > 10$ abzulehnen. Für die Bestimmung der Orientierung der markanten Punkte wird in einem Fenster um jeden markanten Punkt auf der niedrigsten Auflösungsstufe, auf der er detektiert wurde, in einer lokalen Umgebung die Gradientenrichtung und -stärke aller Pixel (x, y) aus den partiellen Ableitungen $\delta_x(x, y)$ und $\delta_y(x, y)$ der Bildpyramide $L(x, y, \sigma)$ berechnet und in einem Histogramm mit 10° Schrittweite eingetragen. Für jeden markanten Punkt werden alle Orientierungen gespeichert, deren Häufigkeit mindestens 80% der Orientierung mit der höchsten Häufigkeit erreichen. Die genaue Richtung wird über eine Parabel im Histogramm interpoliert. Die lokalen Maxima dieses Histogramms entsprechen den Hauptrichtungen der Gradienten für einen markanten Punkt. Dadurch ist die Invarianz gegenüber Verschiebungen, Skalierungen und Rotation erfüllt.

Aus dem Skalenbild der Auflösungsstufe, die der Größe des markanten Punktes am nächsten kommt wird nochmals ein Histogramm der Gradientenrichtung erstellt. Es nutzt ein 16×16 Pixel großes Fenster, das in 4×4 Unterfelder mit je 4×4 Pixeln unterteilt ist. In jedem dieser 16 Felder wird ein Histogramm der Gradientenstärke in einem 45° Raster erstellt. Daraus ergeben sich pro 4×4 Unterfeld 8 Werte. Bei 16 dieser Felder ergibt sich damit insgesamt ein Deskriptor mit 128 Einträgen, die radiometrisch robust normalisiert wird. Hierdurch wird eine substantielle Invarianz gegenüber Helligkeitsänderungen auf Grund von Beleuchtung und gegenüber geometrischen Projektionen erreicht. Da der Deskriptor als normierter Vektor jedes markanten Punktes aus dem Bild des Skalen-Raums mit der Auflösung des markanten Punktes erstellt ist, ist diese Beschreibung gegenüber kleinen affinen Transformationen unempfindlich. Bei einer Änderung der Blickrichtung um 50° ergibt sich im Allgemeinen noch eine erfolgreiche Zuordnung von markanten Punkten in zwei Bildern von 50% (Lowe [2004]). Bei höheren Änderungen nimmt die Erfolgsquote deutlich ab.

Im Vergleich zu anderen Operatoren für markante Punkte toleriert der SIFT-Algorithmus eine deutliche Änderung des Blickwinkels bis etwa 50° . Darüber hinaus bietet er auch bei Rotationen zwischen 35° und 45° und bei Skalierungen zwischen 2 und 2,5 bessere Ergebnisse wie andere Algorithmen (Mikolajczyk & Schmid [2005]). Der SIFT-Operator eignet sich daher auch für Fälle, in denen Aufnahme der selben Szene eine große Basis und Blickwinkeländerung aufweisen und ist auf Grund der geglätteten Bildpyramide auch deutlich besser in der Lage, mit unscharfen Kanten zurecht zu kommen.

2.4 Relative Orientierung in einer Bildsequenz

Bei der Orientierung der Bilder im Raum muss zunächst unterschieden werden zwischen Fällen, in denen Passpunkte im Objektraum vorliegen und Fällen, in denen die Orientierung allein aus homologen Punkten der Bilder bestimmt wird. Liegen Passpunkte vor, so kann aus drei Objektpunkten und ihren korrespondierenden Bildpunkten die Orientierung des Bildes bestimmt werden (Haralick et al. [1994]). Auf Grundlage eines vierten Punktes kann zwischen Konfigurationen der Kamera vor oder hinter dem Objekt unterschieden werden. Aus der Photogrammetrie ist dieses

Verfahren als räumlicher Rückwärtsschnitt bekannt, der bei einer Konfiguration mit mehr als den geforderten vier bekannten Objektpunkten über eine Bündelblockausgleichung die optimale Orientierung des Bildes auf Basis aller Beobachtungen bestimmt (Luhmann [2003]; Krauss et al. [2007]). Dieses Modell lässt sich auf Bildsequenzen ausweiten, die in einem Blickverband mit Passpunkten und Verknüpfungspunkten zwischen den einzelnen Bildern gemeinsam ausgewertet werden.

Bei der relativen Orientierung von zwei Bildern wird ihre relative Lage zueinander ohne bekannte 3D Passpunkte wiederhergestellt. Die Strahlen zweier homologer Bildpunkte (x_1, y_1) und (x_2, y_2) in zwei Bildern B_1 und B_2 schneiden sich in ihrem gemeinsamen Objektpunkt (X, Y, Z) . Dieser Punkt ist in einem lokalen Modellkoordinatensystem gegeben. Für die Beschreibung der relativen Orientierung eines Bildes zu einem anderen Bild werden fünf Parameter benötigt. Dabei wird zwischen der relativen Orientierung unabhängiger Bildpaare, die durch fünf Winkel $\phi_1, \phi_2, \kappa_1, \kappa_2$ und entweder ω_1 oder ω_2 beschrieben wird, und der relativen Orientierung durch Folgebildanschluss unterschieden, bei der zwei Translationen in y und z sowie drei Rotationen $\omega_2, \phi_2, \kappa_2$ bestimmt werden. Ein Skalierungsfaktor bleibt unbestimmt. Für die mathematische Repräsentation dieses Zusammenhangs wird entweder über die Kollinearitätsgleichungen ein räumlicher Vorwärts und Rückwärtsschnitt durchgeführt oder die relative Orientierung wird über die Epipolargeometrie beschrieben. Beim ersten Ansatz wird ein Zusammenhang über die Kollinearitätsgleichungen mittels räumlichem Vorwärtsschnitt zur Bestimmung der 3D Objektkoordinaten der homologe Punkte und mittels räumlichem Rückwärtsschnitt zur Bestimmung der relativen Orientierung hergestellt (Luhmann [2003]; Krauss et al. [2007]). Je nach Anzahl der vorhandenen Beobachtungen muss die innere Orientierung der Kamera bekannt sein oder kann im Rahmen einer Bündelblockausgleichung mit bestimmt werden.

Bei der automatischen Zuordnung von markanten Punkten in zwei oder mehr Bildern ist mit Fehlzusordnungen zu rechnen. Fehlzusordnungen sind abhängig von der Änderungen des Kamerastandortes und der Blickrichtung und der beobachteten Szene. Speziell auf Fassaden führen wiederholende Muster wie beispielsweise Fenster dazu, dass unterschiedliche markante Punkte ähnliche Beschreibungen haben und daher bei einer automatischen Zuordnung verwechselt werden können. Im Rahmen der Ausgleichung können bei geeignet gewähltem stochastischen Modell nach der ersten Iteration der Verbesserungsvektor und die Varianzen der Beobachtungen geprüft werden. Beobachtungen, also Punktpaare, mit hohen Verbesserungswerten im Verhältnis zu den Varianzen werden als Fehlzusordnungen eingestuft und entfernt. Eine andere Lösung ist das RANSAC-Verfahren (Fischler & Bolles [1981]). Hier wird eine minimale Zahl an benötigten Beobachtungen zufällig aus der Menge aller Beobachtungen gewählt und daraus das Ergebnis bestimmt. Dieses Ergebnis wird anschließend mit allen anderen Beobachtungen verifiziert. Je mehr Beobachtungen dieses zufällig gefundene Modell als Inlier stützen, umso besser ist die Lösung. Diese Zufallsauswahl wird abhängig vom Verhältnis der Inlier zu allen Punkten wiederholt. Am Ende wird die Lösung gewählt, die von den meisten Beobachtungen gestützt wird.

Die relative Orientierung kann auch über die Epipolargeometrie beschrieben werden (Longuet-Higgins [1981]; Hartley [1997a]). Die Fundamentalmatrix beschreibt den geometrischen Zusammenhang zwischen zwei Bildern mit acht Parametern, fünf der relativen Orientierung und drei Parameter der inneren Orientierung der Kamera. Ist die innere Orientierung bekannt, werden die verbliebenen fünf Parameter der relativen Orientierung durch die essentielle Matrix beschrieben. Durch die Epipolargeometrie wird darüber hinaus der Suchraum für einen korrespondierenden Punkt auf eine Epipolarlinie im anderen Bild begrenzt. Liegen die beobachteten Objektpunkte annähernd in einer Ebene, wird die Bestimmung der Fundamentalmatrix instabil. Bei der Betrachtung von Gebäuden liegen die meisten markanten Punkte auf den Fassaden und können sich damit annähernd in einer Ebene befinden. Diese Punktkonstellation beeinflusst die Eindeutigkeit

der relativen Orientierung, weil es für jede Kameraposition eine innere Orientierung und Rotationsmatrix gibt, die eine Homographie zwischen der Objektebene und der Bildebene beschreibt (Maybank [1993]). In diesen Fällen ist die innere Orientierung als Vorwissen nötig, um die essentielle Matrix berechnen und die Mehrdeutigkeiten der relativen Orientierung vermeiden zu können. Liegen für die Bestimmung der Fundamentalmatrix bzw. der essentiellen Matrix mehr als die benötigte Mindestzahl an homologen Punkten vor, so kann man entweder das überbestimmte Gleichungssystem in einer Ausgleichung lösen oder das RANSAC-Verfahren (Fischler & Bolles [1981]) einsetzen, das aus einer Minimalkonfiguration von zufällig gewählten Beobachtungen dann wiederholt eine Fundamentalmatrix bzw. essentielle Matrix bestimmt und anschließend prüft, wie viele Punktpaare durch die bestimmte relative Orientierung korrekt abgebildet werden, also auf der entsprechenden Epipolarlinie im jeweils anderen Bild liegen. Am Schluss wird die relative Orientierung ausgewählt, die durch die meisten Inlier gestützt wird. Ausgehend von dieser relativen Orientierung kann in einem weiteren Schritt eine Ausgleichung durchgeführt werden, bei der nur die Inlier als Beobachtungen berücksichtigt werden. Dieses Vorgehen erlaubt es im Gegensatz zur Ausgleichung mit allen homologen Punkten, Fehlzuordnungen auszusortieren, die andernfalls das Ergebnis der Ausgleichung negativ beeinflussen würden.

Diese Vorgehensweise kann auch beim Trifokalensor eingesetzt werden (Hartley [1997b]). Aus allen Beobachtungen wählt das RANSAC-Verfahren (Fischler & Bolles [1981]) einen minimalen Satz aus, um den Trifokalensor zu bestimmen. Anschließend wird geprüft, wie viele homologe Punkte in den jeweils zwei anderen Bildern auf den entsprechenden Epipolarlinien liegen. Für den Trifokalensor mit den meisten Inliern wird anschließend mit den Inliern als Beobachtungen eine Ausgleichung durchgeführt. Eine weitere Steigerung der Stabilität wird durch den 5-Punkt-Algorithmus von Nistér [2004] erreicht, der statt mit acht Punktpaaren lediglich mit fünf Punktpaaren arbeitet, unter der Voraussetzung, dass die Bilder bereits kalibriert vorliegen. Durch die Reduktion auf fünf Punktpaare erhöht sich die Wahrscheinlichkeit, eine Kombination aus fünf korrekten Punktpaaren im RANSAC-Verfahren (Fischler & Bolles [1981]) zu wählen deutlich.

3 Koregistrierung von Infrarotbildern mit Gebäudemodellen

Obwohl für die in dieser Arbeit untersuchten Szenarien ein Gebäudemodell vorliegt, ist auf Grund des geringen geometrischen Detailgrades des Modells und der zu erwartenden kleinen Bildausschnitte davon auszugehen, dass eine direkte Orientierung der Einzelbilder allein am vorhandenen Modell nicht möglich ist. Daher soll eine relative Orientierung der Bilder stattfinden. Hierbei sollen zwei unterschiedliche Herangehensweisen verglichen werden, die im folgenden Kapitel vorgestellt werden.

Zum einen soll in einem zweistufigen Verfahren die Bildsequenz zunächst relativ orientiert und anschließend mit dem bestehenden Modell koregistriert werden. Hierfür wird zunächst basierend auf dem Trifokalensor (Hartley [1997b]) eine initiale relative Orientierung von sich überlappenden Bildtriplets durchgeführt, die anschließend in einer Bündelblockausgleichung zusammengefasst werden. Dieses Verfahren entspricht dem Vorgehen, das in Mayer [2007] beschrieben ist. Im Gegensatz zu Mayer [2007] und Heinrichs et al. [2008] wird aber anschließend kein polygonales Fassadenmodell aus den resultierenden 3D Punkten erstellt, auf das dann die Bilder der Sequenz projiziert werden. Die resultierenden 3D Punkte der homologen Bildpunkte werden in der zweiten Stufe segmentiert und in Ebenen eingeteilt, die dann in einer Bündelblockausgleichung an die Flächen des bestehenden Gebäudemodells angepasst werden. Anschließend erfolgt auf Basis der so korrigierten Kameraorientierungen die Projektion der Bilder auf das Gebäudemodell. Dieses Verfahren wird in Abschnitt 3.2 näher erläutert.

Ein zweites Verfahren verwendet das Vorwissen über das vorhandene Gebäudemodell und die innere und äußere Orientierung der Kamera in einem einstufigen Ausgleichungsverfahren. Hierfür wird zunächst in Bildpaaren die Fundamentalmatrix bestimmt und so Fehlzuordnungen eliminiert. Anschließend wird jedes Bildpaar in einer Bündelblockausgleichung mit den aufgezeichneten äußeren Orientierungen als Beobachtungen und dem Gebäudemodell zusammengeführt. Die so korrigierten äußeren Orientierungen werden anschließend für die Projektion der Bilder auf das Gebäudemodell verwendet. Dieses Vorgehen wird in Abschnitt 3.3 beschrieben.

Um für eine Koregistrierung die Parameter der inneren Orientierung verwenden zu können, ist vorab eine geometrische Kalibrierung nötig. In diesem Kapitel werden die verwendeten Verfahren zur geometrischen Kalibrierung von thermischen Infrarotkameras (Abschnitt 3.1) sowie die Verfahren zur Koregistrierung von Gebäudemodell und Bildsequenz im zweistufigen (Abschnitt 3.2) und einstufigen (Abschnitt 3.3) Fall vorgestellt.

3.1 Geometrische Kalibrierung der Infrarotkamera

Das in diesem Abschnitt beschriebene Verfahren der geometrischen Kalibrierung von Infrarotkameras wird erstmals in Simmler [2009] vorgestellt. In Abschnitt 3.1.1 werden kurz die dort durchgeführten Untersuchungen zu Messmarken vorgestellt. Abschnitt 3.1.2 geht auf die Berechnung der Parameter der inneren Orientierung sowie der Verzeichnung ein. Das von Simmler [2009] genutzte Verfahren entspricht hier dem, das später von Luhmann et al. [2010] auf seine Genauigkeit hin untersucht worden ist.

3.1.1 Untersuchung zu Messmarken

Infrarot-Detektoren messen Intensitätsunterschiede der Strahlung eines bestimmten Wellenlängenbereiches (2.1). Das heißt, um Objekte voneinander unterscheiden zu können, müssen sie unterschiedlich stark strahlen, also eine unterschiedliche Temperatur oder einen deutlich abweichenden Emissionskoeffizienten besitzen. Klassische Messmarken an Wänden erfüllen diese Bedingung nicht. Sie sind fest an den Wänden angebracht und haben in der Regel dieselbe oder eine nur minimal von der Wand abweichende Temperatur. Auch die Emissionskoeffizienten im infraroten Spektrum unterscheiden sich kaum von dem einer Wand. Alternativ ist es möglich, die Reflexionseigenschaften bestimmter Materialien im infraroten Spektrum zu nutzen. Anstatt sich von der Temperatur des Hintergrundes abzuheben, kann eine Messmarke als Reflektor eingesetzt werden. In diesem Fall ist es notwendig, die Messmarken zu bestrahlen, um einen Kontrast zum Hintergrund herzustellen. Ein weiteres Problem liegt in der zeitlichen Veränderung. Eine einmal aufgewärmte Messmarke kühlt sich ohne Energiezufuhr während der Messreihe wieder ab und ist daher über die Einzelaufnahmen nicht stabil. Messmarken sollen außerdem eine klare abgegrenzte Form haben, die es ermöglicht, das Zentrum der Marke eindeutig zu bestimmen. Die Wärme verteilt sich jedoch in einem unscharfen Bereich. Eine genaue Eingrenzung einer Messmarke wird dadurch erschwert. Abbildung 3.1 zeigt eine klassische uncodierte Messmarke im sichtbaren und thermischen infraroten Spektrum. Die im sichtbaren Spektrum deutlich reflektierende Marke zeigt im thermischen infraroten Spektrum nur einen schwachen Kontrast.

Auf Grund der geringeren optischen Auflösung müssen die Messmarken größer ausgelegt sein wie bei einer Kamera im sichtbaren Spektrum mit gleichem Öffnungswinkel. Schließlich kann die Bestimmung der Parameter der inneren Orientierung für verschiedene Entfernungen bei verschiedenen Fokussierungen notwendig sein. Für die großflächige Aufnahme von Fassaden wird von einer konstanten Fokussierung ausgegangen. Es sind also Messmarken nötig, die sowohl für kurze wie lange Entfernungen zum Sensor geeignet sind. Es ergeben sich daher folgende Auswahlkriterien für geeignete Messmarken im Infraroten:

- Kontrast zum Hintergrund durch eine deutlich andere Strahlungsstärke oder Reflexion
- Zeitliche Stabilität der Marke zumindest über wenige Minuten
- Scharfe Abgrenzung der Marke in einer klaren geometrischen Form
- Größe der Messmarke im Bereich des optischen Auflösungsvermögens des Detektors
- Testmarken für Innen- und Außentestfeld

Verwendet man spiegelnde Materialien, so zeigen sie die Umgebung, in der gemessen wird. Wie bei einem normalen Spiegel hängt das reflektierte Bild vom Blickwinkel der Kamera ab und ist damit nicht stabil. Ein Strahler, der auf die Messmarke gerichtet wird, ist also nur aus einem bestimmten Blickwinkel der Kamera sichtbar. Aus anderen Blickwinkeln zeigt die Messmarke

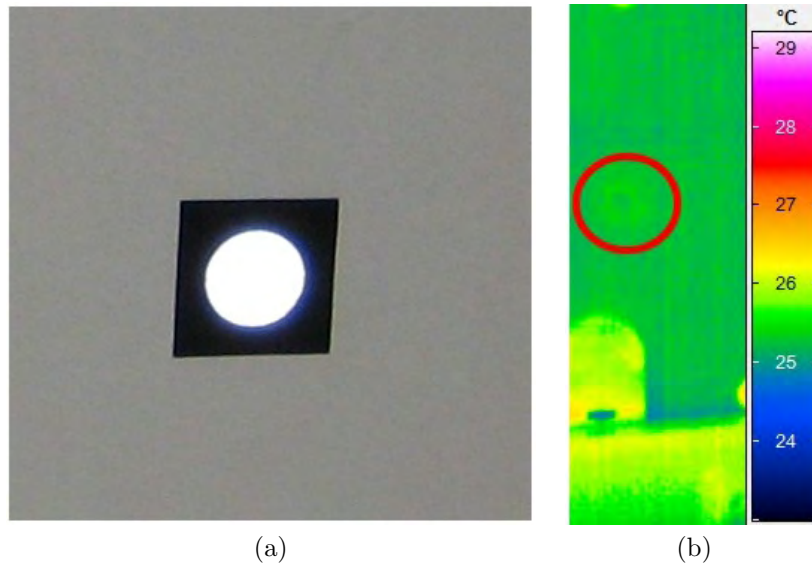


Abbildung 3.1: Standard Messmarke: a) im sichtbaren Spektrum, b) im Infrarot (roter Kreis). Der Kontrast im Infraroten reicht nicht aus, um die Messmarke genau zu bestimmen. Heller Fleck unten links: Ein schwarzer Tischtennisschläger (Simmler [2009])

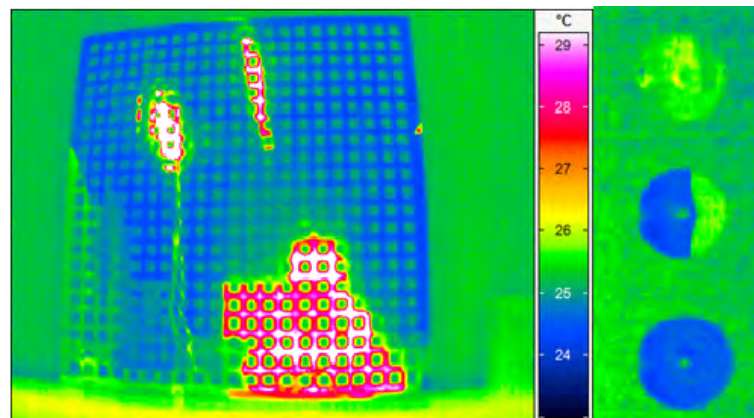


Abbildung 3.2: Reflektierende Infrarot Marken. Links: Ein Metallgitter spiegelt die Temperaturen der Szene abhängig vom Blickwinkel. Rechts: CDs reflektieren die Umgebung abhängig vom Blickwinkel (Simmler [2009])

die Umgebung, in Räumen also die umgebenden Wände, die normalerweise dieselbe Temperatur haben, wie die Wand, auf der die Messmarke befestigt ist. Abbildung 3.2 zeigt die Blickwinkelabhängigkeit der reflektierten Infrarotstrahlung. Das linke Bild zeigt einen Innenraum, der an einem Metallgitter reflektiert wird. Im rechten Bild zeigen auf dem Boden ausgelegte CDs die Umgebung einer Außenaufnahme, also Bäume oder Häuser oder je nach Blickwinkel der Kamera und Ausrichtung der Messmarken den Himmel. Da der Himmel im Infraroten immer kalt gegenüber der Umgebung erscheint, kann man sich diese Eigenschaft für ein Außentestfeld zu Nutze machen. Damit ein Testfeld auf Reflektoren von allen Aufnahmerrichtungen aus den Himmel reflektiert, müssen die Reflektoren flach auf dem Boden ausgebreitet werden. Außerdem darf in der Umgebung des Testfelds kein hohes Objekt stehen, das sich in den Reflektoren spiegeln könnte.

Das Aussehen reflektierender Messmarken ist also von der Umgebung abhängig. Alternativ kommen Messmarken zum Einsatz, deren eigene Temperatur sie vom Hintergrund abhebt. Einmalig aufgewärmte Objekte halten ihre Wärme im Allgemeinen nicht lange genug für eine Messreihe,

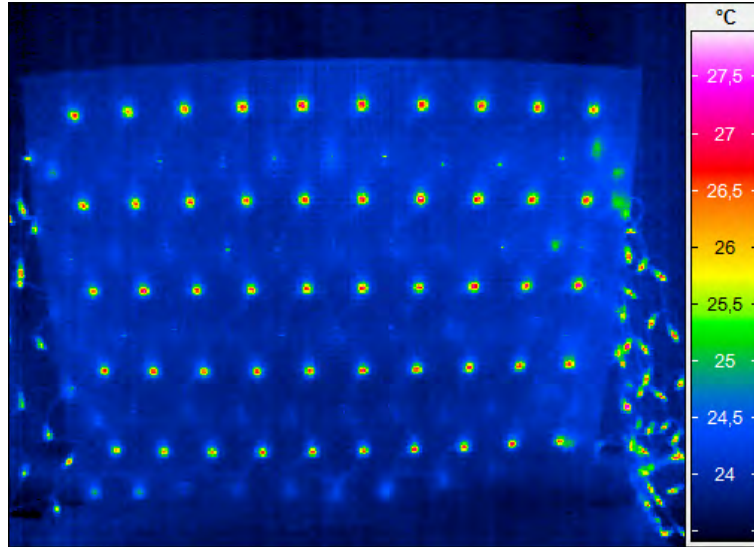


Abbildung 3.3: Aktiv strahlende Marken im Infraroten: Glühlampen eines Lichternetzes werden durch eine Lochrasterplatte gesteckt und photogrammetrisch eingemessen (Simmler [2009])

weil sie wieder abkühlen und ihre Wärme an den Hintergrund verlieren. Messmarke und Hintergrund nähern sich also in ihrer Temperatur an und damit verschlechtert sich der Kontrast und die Grenzen der Messmarke verschwimmen. Gleiches gilt auch für eine abgekühlte Messmarke. Geeignet als Messmarke sind aktive Strahler wie Glühlampen (Abb. 3.3). Sie heizen sich selbst auf und halten ihre Temperatur deutlich über der Umgebung. Damit sind sie über eine Messperiode hinweg stabil. Ihre Abstrahlung ist aus verschiedenen Blickwinkeln annähernd gleich. Die Messmarke kann also aus verschiedenen Blickwinkeln mit gleicher Intensität gesehen werden. Es liegt ein geschlossener deutlich abgegrenzter Körper vor. Die Wärmequelle kann als kugelförmig angenommen werden. Die Glühlampen sind in einem festen Raster angeordnet. Ein weiterer Vorteil der aktiv strahlenden Messmarke liegt darin, dass ihre Größe nicht so stark von der optischen Auflösung der Kamera abhängt. Während die passive Messmarke nicht mehr sichtbar ist, wenn sie unterhalb der optischen Auflösung liegt, überstrahlt die aktive Messmarke die Umgebung. Sie ist also auch dann noch sichtbar, wenn ihre Abbildung auf dem Sensor ein Pixel nicht ausfüllt.

3.1.2 Bestimmung der inneren Orientierung und der Verzeichnungsparameter

In Abschnitt 2.2 wurde darauf eingegangen, dass für eine erfolgreiche Kalibrierung eine möglichst große Anzahl an Messpunkten vorhanden sein sollte, um eine möglichst hohe Genauigkeit durch Redundanz zu erreichen. Da die innere Orientierung der Kamera mit der äußeren Orientierung über die Abbildungsgeometrie verbunden ist (Luhmann [2003]; Kraus [2004]), kann die innere Orientierung aus mehreren Einzelaufnahmen rekonstruiert werden. Simmler [2009] bestimmt die innere Orientierung (x_0, y_0, c_k) und die radialsymmetrische Verzerrung mit zwei Parametern A_1 und A_2 . Dieses Verfahren wird in der vorgelegten Arbeit verwendet, um für die Bündelblockausgleichung Näherungswerte für die innere Orientierung zu erhalten. Das genaue Vorgehen bei der Kalibrierung wird in Abschnitt 3.1 beschrieben. Für die Anordnung der Messmarken wird ein regelmäßiges Gitter verwendet. Dieses Gitter wird entweder senkrecht an einer Wand angebracht (Abb. 3.4a) oder waagerecht auf dem Boden ausgebreitet (Abb. 3.4b). Die Positionen der einzelnen Marken sind eingemessen und werden für die Bündelblockausgleichung als fehlerfreie Beobachtungen angenommen. Die Kamera nimmt die Szene aus verschiedenen Positionen auf. Für das senkrechte Testfeld wird von links, frontal und rechts jeweils von schräg unten und oben aufgenommen. Zusätzlich werden die oberen Aufnahmen mit auf dem Kopf stehender Kamera

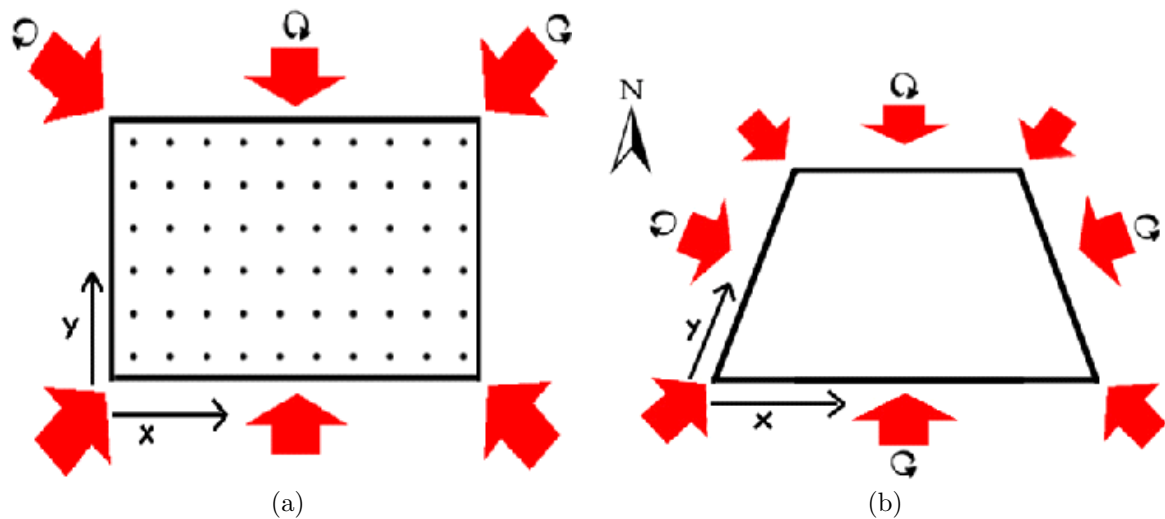


Abbildung 3.4: Testfeldanordnungen: a) Senkrecht test field mit Blickrichtungen von links, frontal und rechts jeweils von schräg oben und unten. Die drei oberen Aufnahmen wurden zusätzlich mit auf den Kopf gestellter Kamera wiederholt, b) Waagrecht test field. Die Kamera wandert von einem erhöhten Standpunkt aus um das Testfeld herum. Einzelne Standorte werden zusätzlich mit auf den Kopf gestellter Kamera wiederholt (Simmler [2009])

wiederholt. Die resultierenden neun Aufnahmen werden für die Ausgleichung verwendet. Bei einem waagerechten Testfeld am Boden blickt die Kamera von einem erhöhten Standpunkt aus auf das Testfeld und wandert in 45° Schritten um das Testfeld herum. Zusätzlich werden im 90° Rhythmus Aufnahmen mit auf dem Kopf stehender Kamera gemacht. Es ergeben sich hier 12 Aufnahmen für die Bündelblockausgleichung.

Das senkrechte Testfeld (Abb. 3.4a) erlaubt größere Variationen des Betrachtungswinkels, von sehr flach bis hin zu einer Frontaldraufsicht. Das waagerechte Testfeld ist beim Blickwinkel beschränkt auf die Höhe der Kamera über dem Boden. Da für die Bestimmung der inneren Orientierung die Fokussierung der Kamera eine Rolle spielt, müssen die Messungen für verschiedene Fokussierungen, also Objektstände durchgeführt werden. In der Photogrammetrie misst man normalerweise bei Kameras eine Fokussierung auf unendlich. Dafür ist ein mehrere Meter großes Testfeld nötig. In der klassischen Kamerakalibrierung werden hierbei Messmarken an einer Hauswand angebracht. Im Infraroten bedeutet das, dass solche Messmarken aktiven Strahler sein müssen, da reflektierende Messmarken wie bereits beschrieben die Umgebung reflektieren und daher Probleme mit dem Kontrast und der Stabilität der Marken aus verschiedenen Blickwinkeln auftreten. Möchte man dennoch reflektierende Messmarken verwenden, so ist ein flach auf dem Boden liegendes Testfeld besser geeignet. Die Messmarken reflektieren dann aus jeder Blickrichtung den kalten Himmel.

Um eine Vergleichbarkeit sowohl zwischen den einzelnen Durchläufen als auch zwischen den Testfeldern zu erlangen, werden die Messmarken beider Testfelder als Verknüpfungspunkte deklariert. Die vier Eckpunkte des Testfelds werden als Passpunkte definiert. Bei der Deklaration als Passpunkt wird festgelegt, dass die Messmarken mit einer definierten Genauigkeit die Soll-Punktkoordinaten treffen. Dadurch entsteht ein Zwang für das gesamte Netz der ausgeglichenen Punkte. Bei einem Verknüpfungspunkt werden die Soll-Punktkoordinaten in der Koordinatenberechnung nicht berücksichtigt. Stattdessen wird ein Strahlenschnitt erzeugt und die Koordinaten ohne Restriktionen bestimmt.

3.1.3 Alternative Kalibrierung über Passpunkte

Leider ist eine Testfeldkalibrierung nicht immer möglich. Insbesondere bei fest verbauten oder luftgetragenen Kamerasystemen kommt daher eine geometrische Kalibrierung aus den aufgenommenen Bildsequenzen selber in Betracht. Hierfür werden Passpunkte am 3D Gebäudemodell verwendet und die entsprechenden Bildpunkte manuell in einer Auswahl von Bildern der Sequenz markiert. Dieses Vorgehen ist im Detail in Stilla et al. [2009] für Luftbilder im mittleren thermischen Infrarot beschrieben. Für die terrestrischen Sequenzen ergibt sich hier wieder die Problematik, dass nur wenige Ecken des Gebäudemodells sichtbar sind. Für die Bestimmung der inneren Orientierung sind daher zusätzliche Einzelaufnahmen zu machen, die möglichst große Teile eines Gebäudes zeigen, um Ecken im Modell denen im Bild zuordnen zu können. Das mathematische Modell der Ausgleichung entspricht dem Vorgehen, das in Abschnitt 3.1.2 beschrieben ist, allerdings mit einer unregelmäßigen Verteilung der Passpunkte.

3.2 Zweistufiges Verfahren mit relativer Orientierung

Im zweistufigen Verfahren werden die Schritte der relativen Orientierung der Bilder und die Koregistrierung mit dem 3D Gebäudemodell getrennt behandelt. Die relative Orientierung findet ohne Vorwissen über die äußere Orientierung statt. Sie entspricht weitgehend dem Vorgehen, wie es von Mayer [2007]; Mayer et al. [2012] vorgeschlagen wird. Das ursprünglich zur relativen Orientierung mehrerer unsortierter Bilder mit teilweise großer Stereobasis und veränderte Blickwinkeln entwickelte Verfahren wird mit geringen Anpassungen an einzelne Parameter auf Sequenzen geordneter Bilder mit konstantem Blickwinkel angepasst. In dieser Arbeit wird lediglich der erste Teil des von Mayer [2007]; Mayer et al. [2012] vorgestellten Verfahrens verwendet, die relative Orientierung. Abbildung 3.5a zeigt die von Mayer [2007]; Mayer et al. [2012] übernommene relative Orientierung der Bildsequenz. Die anschließende Objekt- und Texturextraktion wird durch ein eigenes Verfahren ersetzt, das das Vorwissen über die Kamerapositionen zum Aufnahmezeitpunkt und das vorhandene Gebäudemodell nutzt. Die einzelnen Schritte sind in Abbildung 3.5b dargestellt. Zum einen erleichtert dieses Vorgehen die Schätzung der Kameraorientierungen innerhalb der Sequenz, zum anderen muss zusätzlich zur relativen Orientierung der Sequenz die Koregistrierung mit einem bereits vorhandenen Modell durchgeführt werden.

3.2.1 Relative Orientierung der Bildsequenz

Die Verknüpfung der Bilder einer Sequenz in einer relativen Orientierung nach Mayer [2007]; Mayer et al. [2012] lässt sich in eine Reihe von Teilschritten zerlegen (Abb. 3.5a). Im ersten Schritt werden in den Bildern markante Punkte detektiert. Hier können sowohl Förstner-Punkte (Förstner & Gülch [1987]) wie auch SIFT Features (Lowe [2004]) eingesetzt werden. Um die Zahl der Fehlzuordnungen zu reduzieren, wird in einem ersten Schritt über eine Kreuzkorrelation die lokale Umgebung der zugeordneten homologen Punkte verglichen. Dabei werden auch Rotationen mit berücksichtigt. Die relativen Koordinaten korrespondierender Punkte werden durch affine Kleinste-Quadrate-Zuordnung (Grün [1985]) verbessert und Genauigkeitseigenschaften bestimmt. Die folgende Berechnung der essentiellen Matrizen aller aufeinander folgenden Bilder und des kalibrierten Trifokaltensors (Nistér [2000]) nutzt das RANSAC-Verfahren (Fischler & Bolles [1981]) in Kombination mit dem Geometric Robust Information Criterion (GRIC) (Torr [1997]). Dabei wird wiederholt ein zufälliger Satz von homologen Punkten gewählt und daraus ein Modell berechnet. Dieses Modell wird mit den übrigen homologen Punkten verglichen und bewertet, wie gut diese zum Modell passen. Es wird nicht das Modell ausgewählt, das durch die meisten Punkte gestützt wird, sondern es werden die Punkte entsprechend ihrer Residuen gewichtet und das Modell mit der kleinsten Summe der Residuen ausgewählt. Es zeigt sich, dass diese Methode

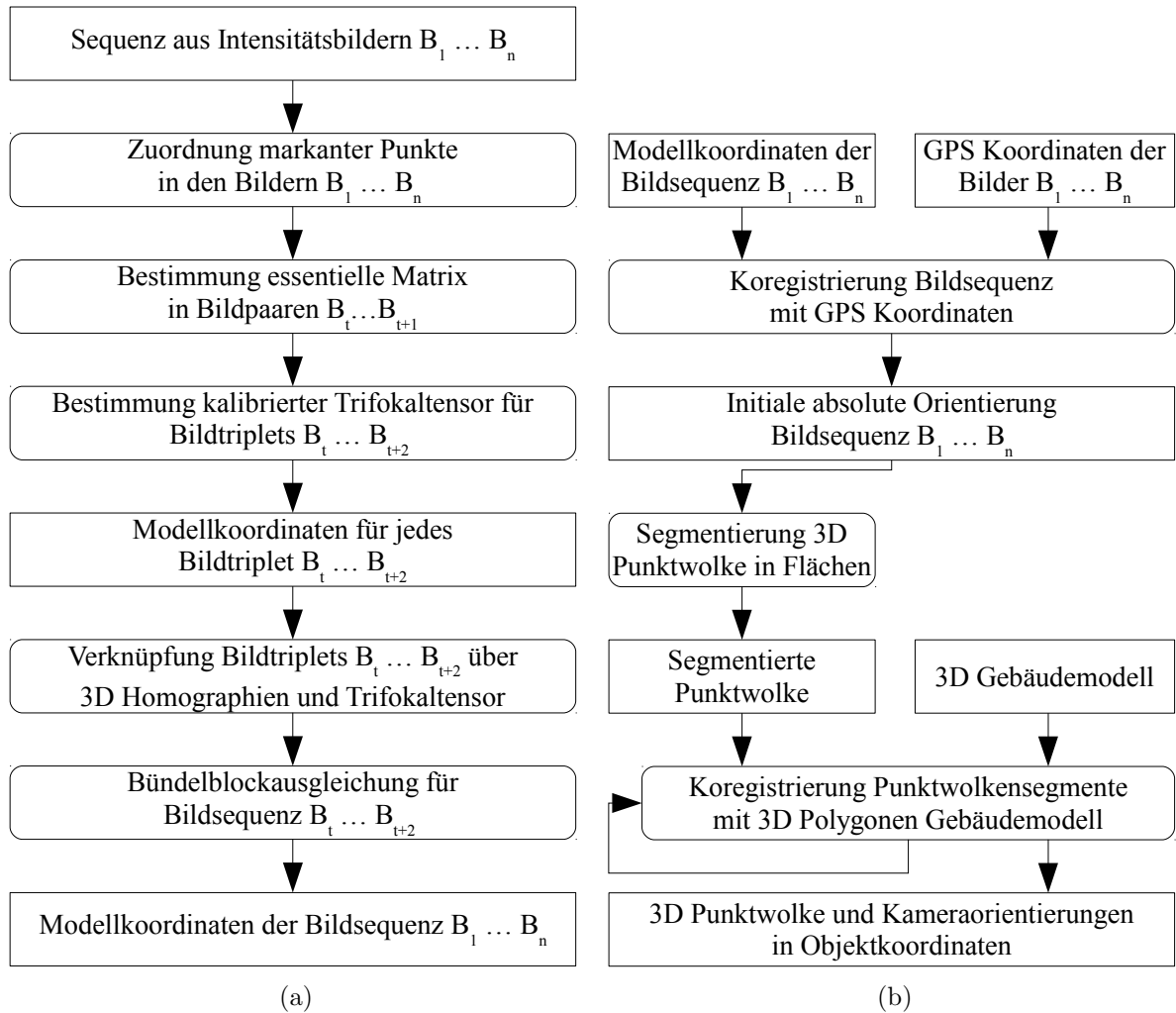


Abbildung 3.5: Überblick zweistufiges Orientierungsverfahren: a) Relative Orientierung bei Mayer [2007]. Nach der Detektion der SIFT Features werden aus der Bildsequenz überlappende Bildtriplets gebildet. Anschließend wird die gesamte Sequenz ausgeglichen, b) Bestimmung der absoluten Orientierung durch Bündelausgleichung mit GPS-Pfad und Gebäudemodell. Da die initialen Parameter der absoluten Orientierung auf den ungenauen GPS-Positionen basieren, wird der Koregistrierungsschritt iterativ durchgeführt, bis sich keine Verbesserung mehr ergibt.

insbesondere bei wiederkehrenden Mustern wie Fensterecken immer noch zu viele Fehlzuordnungen zulässt und falsche Modelle liefert. Zusätzlich führt Mayer et al. [2012] daher eine Maximum Likelihood Schätzung nach dem Vorschlag von Chum et al. [2003] durch, um die beste RANSAC Schätzung zu finden. Durch den Übergang von Bildpaar zu Bildtriplet wird die Schätzung des Modells insbesondere bei Szenen mit dominanten Ebenen deutlich besser. Anstelle einer relativen Orientierung wird als Modell der Trifokaltensor gewählt.

Trifokaltensoren werden für alle aufeinanderfolgenden Triplets bestimmt, also für $B_t \dots B_{t+2}$, $B_{t+1} \dots B_{t+3}$ usw. Um den Suchraum für den Trifokaltensor einzuschränken, werden zuerst über die Fundamentalmatrizen aufeinander folgende Bildpaare verknüpft. Mayer et al. [2012] begründen den Einsatz des Trifokaltensors mit der höheren Zuverlässigkeit. Während bei der Fundamentalmatrix lediglich ein Test auf den Abstand eines Modellpunktes zur Epipolarlinie möglich ist, kann mit dem Trifokaltensor die gefundene Konfiguration durch das dritte Bild tatsächlich überprüft werden. Durch die bereits angesprochene Überschneidung der Triplets in der Sequenz wird zusätzlich mehr Stabilität erreicht. Für die Trifokaltensoren wird eine robuste Bündelausgleichung

(Hartley & Zisserman [2004]; Mikhail et al. [2001]; Grün [1978]) mit allen im RANSAC-Verfahren als Inliern markierten Punkten ausgeführt, die die jeweilige essentielle Matrix bzw. den jeweiligen Trifokalensor stützen. Durch den Trifokalensor wird das Modell auch gegenüber dominierenden Ebenen im Bild, z.B. Fassaden robuster.

Die Kombination aller Bilder gelingt auf Grundlage der verschränkt bestimmten Bildtriplets. Zwei aufeinanderfolgende Bildtriplets haben zwei gemeinsame Bilder. Da das zweite und dritte Bild eines Triplets dem ersten und zweiten Bild eines Folgetriplets entsprechen, können zwei Triplets über eine 3D Homographie auf Grundlage der Projektionsmatrizen verbunden werden. Die Projektion der Punkte erfolgt in den Bildebenen mittels des Trifokalensors und Kleinst-Quadrat-Zuordnung. In Mayer [2007] wird statt einer sequenziellen Strategie, die mit dem ersten Bild beginnt und Bild für Bild der Sequenz ergänzt, eine hierarchische Strategie vorgeschlagen, in der zunächst je zwei benachbarte Triplets verbunden werden und anschließend hierarchisch je zwei benachbarte Sets von Triplets, bis alle Triplets in einem Set zusammengefasst sind. Dies hat neben der Parallelisierbarkeit vor allem den Vorteil, dass alle Sets, die kombiniert werden, gleich groß sind und somit die Varianzen und Residuen der homologen Punkte, die für die Gewichtung in der Bündelblockausgleichung verwendet werden, vergleichbar sind.

Dieses Verfahren zeigt sowohl auf dem Testdatensatz des Schlosses Ettlingen (Strecha et al. [2008]) mit Winkeländerungen und sich wiederholenden Mustern als auch für eine Sequenz aus einem Fahrzeug heraus (Mayer et al. [2012]) sehr hohe Genauigkeiten mit Rückprojektionsfehlern von 0,14 Pixeln und einer Abweichung in der geschätzten Bewegung von 0,1 % und 0,007° pro Bildpaar.

3.2.2 Absolute Orientierungen der Bilder der Sequenz

Für die Herstellung der absoluten Orientierung der Bildsequenz werden neben der Bildsequenz selbst die Daten der Kamerakalibrierung benötigt, die, ausgehend von bekannten Werten, in der Bündelblockausgleichung der relativen Orientierung bereits verfeinert wurden. Außerdem werden die GPS Positionsmessungen eingeführt. Da die Aufzeichnungsfrequenz des GPS-Empfängers nicht mit der Aufnahme Frequenz der Infrarotkameras übereinstimmen muss und gegebenenfalls keine Synchronisation vorliegt, werden die Positionen der Kamera zum Zeitpunkt der Aufnahme aller Bilder aus dem GPS Pfad auf Basis der vorhandenen Zeitstempel linear interpoliert. Der GPS Pfad wird als bereits gefiltert angenommen, beispielsweise mit einem Kalman-Filter (Sorenson [1970]). Es wird angenommen, dass ein 3D Gebäudemodell auf Polygonbasis existiert, das mit Hilfe der Bildsequenz texturiert werden soll. Es wird davon ausgegangen, dass dieses Modell im Level-of-Detail 2 nach der Definition von CityGML (Gröger et al. [2010]) vorliegt.

Die Bestimmung der absoluten Orientierung gliedert sich in drei Teilschritte, die in Abbildung 3.5b dargestellt sind. Zunächst findet eine grobe Orientierung durch eine Kleinst-Quadrat-Ausgleichung des Kamerapfades der relativen Orientierung und der interpolierten GPS Positionen statt. Hier werden die sieben Parameter der absoluten Orientierung $X_0, Y_0, Z_0, \omega_0, \phi_0, \kappa_0$ inklusive des Skalierungsfaktors s bestimmt. In Fällen, in denen eine Sequenz aus einem annähernd gerade verlaufenden Kamerapfad aufgenommen wird, liegen die Projektionszentren aller Beobachtungen annähernd auf einer Geraden. Bei einer Ähnlichkeitstransformation zwischen dem relativ orientierten Kamerapfad und dem gemessenen GPS Pfad kann daher der Drehwinkel des Pfades um die Achse in Bewegungsrichtung nicht eindeutig bestimmt werden. Geht man davon aus, dass die erzeugte Punktwolke aus der relativen Orientierung überwiegend das gegebene Gebäudemodell zeigt, so wird eine Koregistrierung aus Punktwolke und 3D Modell durchgeführt, bei der dieser Drehwinkel mitbestimmt wird. Im zweiten Schritt werden die aus der relativen Orientierung bestimmten 3D Punkte segmentiert und Flächen des Modells zugewiesen. Es folgt eine Kleinst-Quadrat-Ausgleichung über den Abstand der Punkte zu ihren zugewiesenen Gebäudeflächen. Die

Zuordnung und Abstandsminimierung wird iterativ wiederholt bis sich keine Änderungen mehr ergeben.

Der erste Schritt, die Koregistrierung des GPS Pfades mit dem relativ orientierten Kamerapfad der Bildsequenz, ist eine nötige Vorstufe, um eine Zuordnung zum Gebäudemodell herstellen zu können. Die Kameraorientierungen der Bildsequenz werden damit ins Objektkoordinatensystem überführt. Wie bereits erläutert ist die aus der GPS Messung stammende Position der Kameras zu ungenau, um sie direkt für eine Projektion der Bilder auf das 3D Modell zu verwenden. Eine Koregistrierung über die Kanten des Modells scheidet wegen der geringen Sichtbarkeit und Anzahl von Kanten in Fassadenaufnahmen ebenfalls aus. Zudem liegen keine Informationen über die Blickrichtung vor. Diese sind aber in der relativen Orientierung bestimmt worden. Zu bestimmen sind damit sieben Parameter der absoluten Orientierung, bestehend aus einem Translationsvektor t ($\Delta x, \Delta y, \Delta z$), einer Rotationsmatrix R , die die Winkel (ω, ϕ, κ) beinhaltet, und einem Skalierungsfaktor s . Als Beobachtungen werden die Positionen der Kameras der relativen Orientierung (X_{rt}, Y_{rt}, Z_{rt}) und die aus der GPS Messung abgeleiteten Positionen (X_{mt}, Y_{mt}, Z_{mt}) verwendet, die über ihre Zeitstempel t einander zugeordnet werden. Das Ziel ist die Minimierung des quadratischen Abstands zwischen den Kamerapositionen des GPS Pfades (X_{mt}, Y_{mt}, Z_{mt}) und mit t , R und s transformierten Kamerapositionen der relativen Orientierung, die dann absolut orientiert in Objektkoordinaten (X_{at}, Y_{at}, Z_{at}) vorliegen. Damit handelt es sich um eine räumliche Ähnlichkeitstransformation (Luhmann [2003]).

Diese erste Transformation überführt die Bildsequenz ins Objektkoordinatensystem und in eine ungefähr korrekte Orientierung zum 3D Gebäudemodell. Für die Extraktion der Texturen ist weniger die exakte Bestimmung der absoluten Kameraposition an sich wichtig, sondern die genaue Orientierung relativ zum Gebäudemodell. Daher wird im zweiten und dritten Schritt der Kamerapfad an das Gebäudemodell angepasst. Der zweite Schritt segmentiert die aus der Bildsequenz extrahierten 3D Objektpunkte der homologen Bildpunkte. Dazu werden sie in Punkte gruppiert, die auf einer Fläche liegen. Für diese Gruppierung werden verschiedene Annahmen getroffen:

- Fassaden bestehen weitgehend aus planaren Strukturen. Benachbarte Punkte auf einer Fassade haben deshalb dieselbe Normalenrichtung.
- Fassaden sind senkrecht stehende Flächen. Die Normalenvektoren der Punkte müssen daher annähernd waagerecht sein.
- Fassaden treffen an Kanten zusammen. An starken Änderungen der Normalenrichtung liegt folglich eine Kante zwischen Fassadenelementen vor.

Nachdem ein 3D Punkt keine Normalenrichtung besitzt, muss diese zunächst aus der lokalen Nachbarschaft eines Punktes ermittelt werden. Hebel & Stilla [2009] schlagen hierfür eine Hauptkomponentenanalyse vor. Dabei werden für jeden Punkt p_i die Punkte p_{ij} in seiner lokalen Umgebung innerhalb eines vorgegebenen Radius betrachtet und ihr Schwerpunkt $\bar{p}_i$ bestimmt. Bestimmt man nun die Kovarianzmatrix C_i von p_i aus

$$C_i = (p_{i0} - \bar{p}_i, \dots, p_{in} - \bar{p}_i) \circ (p_{i0} - \bar{p}_i, \dots, p_{in} - \bar{p}_i)^T \quad (3.1)$$

so ergeben sich die Eigenwerte λ_1 , λ_2 und λ_3 mit ihren zugehörigen Eigenvektoren v_1 , v_2 und v_3 . Nach einer Normierung der Eigenwerte mit der Anzahl an berücksichtigten Nachbarpunkten sortiert man die Eigenwerte mit ihren Vektoren nach aufsteigender Größe. Ist nun λ_1 nahe Null

und sind λ_2 und λ_3 deutlich größer, so liegt der Punkt p_i auf einer planaren Fläche, deren Normalenvektor durch v_1 gegeben ist. Diese Prozedur wird für alle Punkte der Punktwolke wiederholt. Punkte, die die Planaritätsbedingung nicht erfüllen, werden verworfen. Benachbarte Punkte, die eine ähnliche Normalenrichtung aufweisen, werden zu einer Flächenhypothese zusammengefasst. Während bei Hebel & Stilla [2009] über ein RANSAC Verfahren Flächen bestimmt werden, ist dies in dem hier beschriebenen Fall nicht notwendig. Aus dem Gebäudemodell sind die Fassadenflächen bereits bekannt.

Für die Zuordnung einer Punktmenge zu einer Fassade werden drei Kriterien getestet:

- Die Orientierung der Modellfläche muss mit der Orientierung des Punktwolkensegments übereinstimmen. Für die Modellfläche ist die Normalenrichtung bekannt. Für das Punktwolkensegment ist es der Mittelwert der Normalenvektoren, die für die Punkte des Segments bestimmt wurden.
- Der Abstand zwischen der Modellfläche und dem Punktwolkensegment muss unterhalb eines Schwellwerts liegen. Er ergibt sich aus dem mittleren Abstand der Punkte von der Modellfläche entlang des Normalenvektors.
- Die Modellfläche darf nicht kleiner sein als das Punktwolkensegment. Hierfür wird die Länge und Höhe der Punktwolke entlang ihrer beiden Hauptachsen bestimmt. Die Hauptachsen ergeben sich aus den Mittelwerten der bestimmten Eigenvektoren v_2 und v_3 aller Punkte des Segments. Diese Maße werden mit den Maßen des Fassadenelements verglichen.

Für die Koregistrierung ergeben sich damit in der Ausgleichung als Beobachtungen die Ebenengleichungen der Fassadenelemente sowie die 3D Punkte der Punktwolkensegmente. Als Unbekannte treten die sieben Parameter der absoluten Orientierung auf (Translationsvektor t , der Rotationsmatrix R und Skalierungsfaktor s). Minimiert wird der quadratische Fehler zwischen den 3D Punkten der Punktwolke und ihrer Projektion auf die Fassadenflächen entlang der Normalenvektoren der Fassaden.

3.3 Einstufiges Verfahren mit Gebäudemodell

Im Gegensatz zum zweistufigen Verfahren sollen im einstufigen Verfahren das Vorwissen der Kameraposition aus dem GPS Pfad und das Vorwissen der Szene aus dem 3D Gebäudemodell direkt in die Orientierung der Bildsequenz eingebracht werden. Folgende Randbedingungen sind dabei zu beachten:

- Die Bildsequenz liegt als zeitlich sortierte Abfolge von Bildern mit annähernd konstanter Blickrichtung vor. Diese Bedingung ist gegeben, wenn die Aufnahme mit einer fest montierten Kamera von einem Fahrzeug aus durchgeführt werden. Durch diese Einschränkung erhöht sich die Zahl der korrekten Zuordnungen homologer Punkte, da der Suchraum bei einer annähernd linearen Bewegung deutlich eingeschränkt werden kann.
- Die GPS Positionen sind für jedes Bild aus dem GPS Pfad interpoliert worden. Ihre beschränkte Genauigkeit wird im stochastischen Modell durch entsprechende Gewichte ausgedrückt.
- Die Blickrichtung der Kamera zum Aufnahmezeitpunkt liegt nicht als Messung vor, sondern wird ausgehend von einer initialen Bestimmung aus der Bewegungsrichtung bestimmt. Auch hier müssen wegen der verbleibenden Unsicherheiten die Gewichte im stochastischen Modell entsprechend gesetzt werden.

- Die innere Orientierung der Kamera ist bekannt.

Das einstufige Verfahren schätzt in einer iterativen Bündelblockausgleichung direkt die äußere Orientierung der Bildsequenz und bestimmt 3D Objektpunkte. Beginnend mit einem Bildpaar wird eine Bündelblockausgleichung durchgeführt, deren Ergebnis für die nächste Iteration jeweils um ein Bild erweitert wird. Neben den aus einem räumlichen Vorwärtsschnitt geschätzten Anfangswerten der Objektkoordinaten wird das Vorwissen der Gebäudefassaden als fiktive Beobachtungen eingebracht. Viele Einzelbilder und Bildpaare zeigen fast ausschließlich Teile einer Fassadenebene und können daher für sich genommen nicht in allen Freiheitsgraden der äußeren Orientierung verbessert werden. Daher wird ein Bildpaar, das den Anfang eines Gebäudes oder markante und im Modell modellierte Fassadensprünge zeigt, für die initiale Schätzung verwendet.

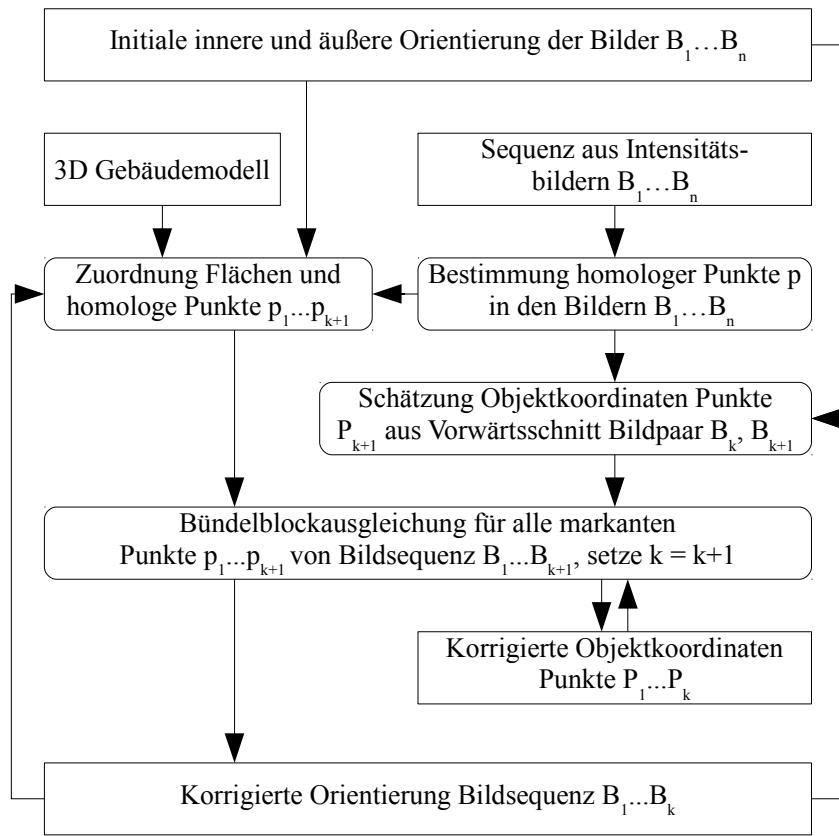


Abbildung 3.6: Überblick zum einstufigen Orientierungsverfahren. Bestimmung Objektkoordinaten aus homologen Punkten mit initialer äußerer Orientierung. Die Bündelblockausgleichung minimiert den Abstand der homologen Bildpunkte und der in die Bildebenen zurückprojizierten Objektpunkte. Bildpaare werden an die Sequenz angehängt durch die gemeinsame äußere Orientierung eines Bildes und bereits bekannter Objektkoordinaten homologer Punkten.

3.3.1 Orientierung von Bildpaaren

In der Bündelblockausgleichung wird mit Hilfe der Kollinearitätsgleichungen (Gl. 2.1) der Abstand der rückprojizierten 3D Objektpunkte zu den ursprünglichen 2D Bildpunkten minimiert. Hierfür werden auf Basis der initialen inneren und äußeren Orientierung aus den homologen Punkten der Bildpaare über einen räumlichen Vorwärtsschnitt initiale 3D Objektkoordinaten der unbekannten Objektpunkte geschätzt, die in der Ausgleichung zurück in die Bildebenen projiziert und mit ihren homologen Punkten verglichen werden. Das Schema der Verarbeitungsschritte ist in Abbil-

dung 3.6 dargestellt. Als Beobachtungen werden die homologen Bildpunkte (x_{ik}, y_{ik}) verwendet, wobei i die fortlaufende Nummerierung aller homologer Punkte eines Objektpunktes (X_i, Y_i, Z_i) und $k \in [1, n]$ die fortlaufende Nummerierung der Bilder $(B_1 \dots B_n)$ der Sequenz darstellen. Gegeben sind außerdem Näherungswerte für die äußeren Orientierungen $(X_{0k}, Y_{0k}, Z_{0k}, \omega_k, \phi_k, \kappa_k)$ der Bilder $(B_1 \dots B_n)$ sowie die innere Orientierung der Kamera aus Bildhauptpunkt (x_0, y_0) , Kamerakonstante (c_x, c_y) und den Verzeichnungsparametern (A_1, A_2) . Gegeben ist auch das 3D Gebäudemodell als Polygonmodell. Das Modell dient zur Formulierung von Nebenbedingungen als fiktive Beobachtungen in der vermittelnden Ausgleichung. Ein 3D Objektpunkt (X_i, Y_i, Z_i) der Fassade muss auf der Fassade liegen. Definiert man die Ebene der Fassade über die Hesse-Normalform

$$\mathbf{r} \cdot \mathbf{n} - d = 0 \quad (3.2)$$

so beschreibt $\mathbf{n}$ den normierten Normalenvektor der Ebene und d den Abstand der Ebene zum Ursprung des Koordinatensystems. Ein Punkt P liegt dann auf der Ebene, wenn sein Ortsvektor $\mathbf{r} = (X_i, Y_i, Z_i)^T$ die obige Gleichung der Hesse-Normalform erfüllt. Damit geht eine Fassadenebene mit ihrem Normalenvektor $\mathbf{n} = (n_0, n_1, n_2)^T$ und ihrem Abstand d zum Ursprung in die Beobachtungen mit ein. Der Vektor $\mathbf{r}$ entspricht den unbekannten Objektkoordinaten (X_i, Y_i, Z_i) . Damit ist mit Gleichung 3.2 direkt die Zusatzbedingung, dass Punkte auf einer Fassade liegen müssen, als fiktive Beobachtung definiert.

Ist die Hesse-Normalform einer Fassadenebene bekannt, kann der Schnittpunkt (X_{ik}, Y_{ik}, Z_{ik}) dieser Ebene mit dem Projektionsstrahl eines Bildpunktes (x_{ik}, y_{ik}) bestimmt werden. Hierfür nutzt man die Geradengleichung

$$g : \mathbf{r} = \mathbf{r}_0 + \lambda \cdot \mathbf{u}, \quad (3.3)$$

wobei $\mathbf{r}_0$ der Ortsvektor und $\mathbf{u}$ der Richtungsvektor ist. Als Ortsvektor wird das Projektionszentrum der Kamera gewählt: $\mathbf{r}_0 = (X_0, Y_0, Z_0)^T$. Der Richtungsvektor ergibt sich aus dem normierten Vektor zwischen Bildpunkt (x, y) in Objektkoordinaten und dem Projektionszentrum. Alle Punkte $\mathbf{r}$, für die es ein λ gibt, das die Gleichung erfüllt, liegen auf dieser Geraden. Setzt man beide Formeln gleich und löst nach λ auf, so ergibt sich

$$\lambda = \frac{d - \mathbf{r}_0 \cdot \mathbf{n}}{\mathbf{u} \cdot \mathbf{n}}. \quad (3.4)$$

Dieses λ in die Geradengleichung eingesetzt ergibt den Schnittpunkt (X_{ik}, Y_{ik}, Z_{ik}) als Vektor $\mathbf{r}$.

Wären die innere und äußere Orientierung sowie die homologen Punkte und das Gebäudemodell fehlerfrei, so fielen die rückprojizierten Objektpunkte (X_i, Y_i, Z_i) mit den homologen Bildpunkten (x_{ik}, y_{ik}) aus Bild k und (x_{ik+1}, y_{ik+1}) aus Bild $k + 1$ zusammen. In der Realität wird das nicht der Fall sein:

- Die Bildpunkte (x_{ik}, y_{ik}) und (x_{ik+1}, y_{ik+1}) weisen eine beschränkte Genauigkeit auf oder können sogar eine Fehlzuordnung darstellen.
- Die innere Orientierung $(x_0, y_0, c_x, c_y, A_1, A_2)$ liegt aus der Kalibrierung nur näherungsweise vor.
- Die äußere Orientierung $(X_{0k}, Y_{0k}, Z_{0k}, \omega_k, \phi_k, \kappa_k)$ ist ebenfalls nur begrenzt genau gemessen (X_{0k}, Y_{0k}, Z_{0k}) oder nur initial geschätzt und interpoliert $(\omega_k, \phi_k, \kappa_k)$.
- Das Gebäudemodell verfügt nur über eine begrenzte Genauigkeit und einen begrenzten Detailgrad. So sind Vorsprünge oder Fenster nicht modelliert.

- Nicht alle Objekte, die in der Bildsequenz sichtbar sind, sind Bestandteil des 3D Modells, z.B. parkende Autos oder Bäume.

Als Unbekannte werden eingeführt die 3D Objektpunkte (X_i, Y_i, Z_i) sowie die zu verbessern- den Parameter der äußeren Orientierung $(X_{0k}, Y_{0k}, Z_{0k}, \omega_k, \phi_k, \kappa_k)$ und der inneren Orientierung $(x_0, y_0, c_x, c_y, A_1, A_2)$.

Die Schätzwerte für die Parameter der äußeren Orientierung ergeben sich aus dem interpolierten GPS Pfad und den vom ersten Bild der Sequenz aus geschätzten Blickwinkel. Die Parameter der inneren Orientierung und der radialsymmetrischen Verzeichnung werden aus der vorab durchgeführten Kalibrierung entnommen. Für die Anfangsschätzung der Objektpunkte (X_i, Y_i, Z_i) wird ein Strahlenschnitt der beiden Abbildungsstrahlen der homologen Bildpunkte (x_{ik}, y_{ik}) und (x_{ik+1}, y_{ik+1}) durch die Projektionszentren der beiden Bilder k und $k+1$ durchgeführt. Aus der Bestimmung des Schnittpunktes der beiden Abbildungsstrahlen mit dem Modell sind die Parameter der beiden Geradengleichungen (Gl. 3.3) bereits bekannt. Für einen gesuchten Objektpunkt P sei nun $P' = X'_0 + \lambda' u'$ und $P'' = X''_0 + \lambda'' u''$ mit $X'_0 + \lambda' u'$ und $X''_0 + \lambda'' u''$ den Geradengleichungen der Abbildungsstrahlen zweier homologer Punkte p' und p'' . Die Objektpunkte P' und P'' sind nach der Orthogonalitätsbedingung dann die beiden Punkte auf den Abbildungsstrahlen mit dem kürzesten Abstand, wenn gilt:

$$\begin{aligned}
 (P' - P'') \cdot u' &= 0 \\
 (P' - P'') \cdot u'' &= 0 \\
 (u'^T \cdot u') \lambda' - (u'^T \cdot u'') \lambda'' + u'^T (X'_0 - X''_0) &= 0 \\
 (u''^T \cdot u') \lambda' - (u''^T \cdot u'') \lambda'' + u''^T (X'_0 - X''_0) &= 0
 \end{aligned} \tag{3.5}$$

Aus der Berechnung von λ' und λ'' folgen die Koordinaten der Objektpunkte P' und P'' . Deren Mittelung ergibt die Anfangsschätzung der Koordinaten des Objektpunktes $P = (X_i, Y_i, Z_i)$.

Als Schätzwert für die Objektkoordinaten (X_i, Y_i, Z_i) könnte im Prinzip auch der Schnitt eines Abbildungsstrahles mit dem 3D Modell verwendet werden. Dann ist jedoch der Startwert von seiner Schätzung her nur an einem Bild aufgehängt, anstatt am Bildpaar. Außerdem würde dieses Vorgehen bei Objektpunkten, die nicht auf einer Fassade liegen, zu falschen Startwerten in der Ausgleichung führen. Für die Zuordnung der Objektpunkte (X_i, Y_i, Z_i) zu einer Fläche des Modells, zu der der Abstand minimiert werden soll, muss zunächst die entsprechende Fläche bestimmt werden. Hierfür wird der Schnitt aus den beiden Abbildungsstrahlen eines Bildpaares mit dem Modell genutzt. Schneiden beide Abbildungsstrahlen dieselbe Fläche des Modells, so sind die homologen Punkte ein guter Kandidat für einen Objektpunkt. Schneiden die Strahlen unterschiedliche Flächen, so kann das drei Ursachen haben:

- Der Objektpunkt liegt nah an einer Kante. Die beiden Abbildungsstrahlen werden sich im Laufe der Iterationsschritte annähern und schließlich dieselbe Fläche schneiden. Wird die Anzahl der Bilder erhöht, so stabilisieren sich die Objektkoordinaten.
- Bei den homologen Punkten handelt es sich um eine Fehlzurordnung. In diesem Fall wird ihr Abstand sich nicht kontinuierlich verringern. Bei der Hinzunahme weiterer Bilder ergeben sich im Falle weiterer Fehlzurordnungen weitere Objektpunkthypothesen mit großem Abstand und eventuell auf anderen Flächen. Es kann aber auch der Fall eintreten, dass einer der beiden homologen Punkte korrekte Entsprechungen in weiteren Bildern findet. Dann bildet sich ein Cluster von Objektpunkthypothesen, die über die Iterationsschritte hinweg ihren Abstand zueinander verringern.

- Der Objektpunkt liegt nicht auf einer der Flächen des Modells, sondern auf einem verdeckenden Objekt. Dies kann ein Baum oder ein Fahrzeug vor dem Gebäude sein. In diesem Fall bleiben über die Iterationen hinweg die Abstände auch bei der Hinzunahme weiterer Bilder erhalten. Allerdings verringert sich der Mindestabstand der beiden Abbildungsstrahlen im Laufe der Iterationen und der Objektpunkt kann für die Orientierung der Bildsequenz verwendet werden. Dafür ist er insbesondere deshalb gut geeignet, weil er nicht in der dominanten Fassadenebene liegt.

Liegen beide Schnittpunkte auf derselben Ebene, so wird diese Ebene in der Ausgleichung für die Abstandsminimierung mit dem Objektpunkt aus dem räumlichen Vorwärtsschnitt der beiden Abbildungsstrahlen verwendet.

Für das Funktionalmodell liegen somit alle Beobachtungen und Schätzwerte der Unbekannten vor. Der Zusammenhang von Bildpunkten (x_i, y_i) , der äußeren Orientierung $(X_0, Y_0, Z_0, \omega, \phi, \kappa)$ und inneren Orientierung $(x_0, y_0, c_x, c_y, A_1, A_2)$ und den unbekannte Objektpunkten (X_i, Y_i, Z_i) wird über die Kollinearitätsgleichungen (Gl. 2.1) beschrieben, die hier mit zwei Parametern A_1 und A_2 der radialsymmetrischen Zeichnung erweitert sind:

$$\begin{aligned}
 x_i - x_0 &= c_x \cdot \frac{a_{11}(X_i - X_0) + a_{21}(Y_i - Y_0) + a_{31}(Z_i - Z_0)}{a_{13}(X_1 - X_0) + a_{23}(Y_1 - Y_0) + a_{33}(Z_1 - Z_0)} + dx'_{rad} \\
 y_i - y_0 &= c_y \cdot \frac{a_{12}(X_1 - X_0) + a_{22}(Y_1 - Y_0) + a_{32}(Z_1 - Z_0)}{a_{13}(X_1 - X_0) + a_{23}(Y_1 - Y_0) + a_{33}(Z_1 - Z_0)} + dy'_{rad} \\
 dr_{rad} &= A_1 \cdot (r^3 - r_0^2 r) + A_2 \cdot (r^5 - r_0^4 r) \\
 dx'_{rad} &= \frac{dr_{rad}}{r} \cdot x' \\
 dy'_{rad} &= \frac{dr_{rad}}{r} \cdot y'
 \end{aligned} \tag{3.6}$$

Die Funktion dr_{rad} beschreibt den Einfluss der radialsymmetrischen Verzeichnung abhängig vom Abstand r eines Bildpunktes zum Bildhauptpunkt (x_0, y_0) . Für die Zusatzbedingungen, dass die Objektpunkte auf einer Fassadenebene liegen müssen, wird Gleichung 3.2 verwendet, wobei $\mathbf{r}_i = (X_i, Y_i, Z_i)$ gilt.

Das stochastische Modell geht von unkorrelierten Bildkoordinaten (x_i, y_i) der homologen Punkte aus. Ihre Genauigkeit wird für Förstner-Punkte und SIFT-Features aus ihren Parametern entnommen. Für Förstnerpunkte wird der Kehrwert der Fläche der Fehlerellipse verwendet: je kleiner die Fehlerellipse, umso größer das Gewicht. Bei SIFT-Features wird der Ähnlichkeitswert der homologen Punkte genutzt. Die Parameter der Fassadenebenen werden als fehlerfrei angenommen. Die Gewichte der Parameter der inneren Orientierung ergeben sich aus den Standardabweichungen der Kamerakalibrierung. Für die äußere Orientierung eines Bildpaars werden die gefilterten Werte des GPS Pfades als Näherungswerte der Unbekannten eingesetzt. Sie kommen daher nicht als Beobachtungen vor und sind damit nicht im stochastischen Modell enthalten. Für die als fiktive Beobachtung formulierte Nebenbedingung (Gl. 3.2) müssen ebenfalls Gewichte ermittelt werden. Da nur für Punkte, die auch tatsächlich auf der Ebene liegen, der sie zugeordnet werden, die Bedingung aus Gleichung 3.2 erfüllt ist, müssen zunächst die Punkte aussortiert werden, für die das nicht zutrifft. Hierfür werden zwei Kriterien angesetzt. Zum einen wird auf den kleinsten Abstand der Abbildungsstrahlen (Gl. 3.5) zurückgegriffen, da dieser eine Aussage darüber zulässt, wie gut die Schätzung der 3D Objektkoordinate ist. Zum anderen wird der ermittelte Abstand des Objektpunktes von der Ebene verwendet. Je weiter ein Punkt von seiner Ebene entfernt liegt, umso unwahrscheinlicher ist es, dass er auf dieser Ebene liegt. Da es möglich ist, dass zu Beginn der Ausgleichung alle Punkte einen großen Abstand zur Ebene haben, weil

die Anfangsschätzung der äußeren Orientierung z.B. mit einer großen Verschiebung behaftet ist, wird der Gewichtsanteil aus dem Abstand zur Ebene normiert über alle Abstände. Damit erhalten Objektpunkte mit geringerem Abstand zu Beginn der Ausgleichung ein größeres Gewicht.

3.3.2 Orientierung der Bildsequenz

Im vorangegangenen Abschnitt wurde das Grundprinzip des gewählten Ausgleichungsansatzes erläutert. In diesem Abschnitt geht es um die Erweiterung auf Bildsequenzen. Für die Kombination von Bildsequenzen können im wesentlichen drei Ansätze unterschieden werden. Zunächst können die Sequenzen, wie in der Luftbildphotogrammetrie üblich, über Passpunkte verknüpft werden (Luhmann [2003]). Hier werden alle Bilder und alle Beobachtungen gemeinsam ausgewertet. Dieser Ansatz wird jedoch bei einer großen Anzahl von Bildern sehr rechenintensiv. Alternativ können die Bilder sequentiell abgearbeitet werden. Eine bereits ausgeglichene Sequenz wird um ein Bild erweitert, das bestmöglich an die bestehende Sequenz angeschlossen wird. Dieses Verfahren kommt bei bewegten Systemen häufig zum Einsatz. Mayer et al. [2012] schlagen für Bildsequenzen ein hierarchisches Verfahren vor, bei dem benachbarte Gruppen von Bildern iterativ zu größeren Gruppen zusammengefasst werden bis die gesamte Bildfolge ausgeglichen ist. Dieser Ansatz hat neben der Effizienz den Vorteil, dass er alle Bilder mit ihren Beobachtungen ungefähr gleich stark gewichtet.

Für die Bildsequenzen, die im weiteren Verlauf dieser Arbeit untersucht werden, wird der sequentielle Ansatz gewählt. Der erste Ansatz scheidet aus, da auf den Fassaden keine Passpunkte vorhanden sind. Fassadenkanten sind nur teilweise und nicht in voller Länge sichtbar, können daher nicht für eine Koregistrierung verwendet werden. Da mit dem GPS Pfad und dem Gebäudemodell Zusatzinformationen zur Verfügung stehen, ist der Einfluss von Fehlern in den ersten Bildern im Vergleich zu einer Sequenz, bei der rein über relative Orientierung ohne Vorwissen gearbeitet wird, sehr klein. Des weiteren erzielen Bilder, in denen mehrere Fassaden sichtbar sind, eine höhere Genauigkeit, weil die 3D Objektpunkte dann in verschiedenen Ebenen angeordnet sind und dies eine Korrektur der äußeren Orientierung erleichtert. Bei Bildern, die nur eine Fassade zeigen, liegen fast alle 3D Objektpunkte in einer Ebene. Bei solchen Bildern ist es schwer, eine eindeutige Lösung für eine Korrektur der äußeren Orientierung zu bestimmen, da eine Verschiebung der Punktwolke auf der Fassadenebene das Ergebnis der Ausgleichung nicht beeinflusst.

Wenn man davon ausgeht, dass das erste Bild einer Sequenz den Anfang eines Gebäudes zeigt, so kann man davon ausgehen, dass mehrere Fassaden zumindest teilweise sichtbar sind, die entweder vom selben Gebäude oder von Nachbargebäuden stammen. Dadurch ist in der Bündelblockausgleichung für solche Bilder unter der Verwendung des Vorwissens eine höhere Genauigkeit und Stabilität zu erwarten wie für Bilder, in denen fast ausschließlich eine Fassadenfläche zu sehen ist.

Zu den Orientierungsparametern eines Bildes tragen alle benachbarten Bilder der Sequenz bei, mit denen das Bild homologe Punkte hat. Ermittelt man für die gesamte Sequenz zuerst alle homologen Punkte, so ergibt sich daraus, welches Bild von welchen anderen Bildern direkt abhängig ist. Dieses Wissen wird genutzt, um die Bündelblockausgleichung zu vereinfachen. Die aus Ausgleichungssicht beste Lösung ergibt sich, wenn die gesamte Bildsequenz, in der eine bestimmte Fassade zu sehen ist, zusammen ausgeglichen wird. Geht man davon aus, dass Bilder am Anfang der Sequenz genauer orientierbar sind wie in der Sequenzmitte, weil am Anfang eine weitere Fassadenebene zu sehen ist und die 3D Objektpunkte sich nicht überwiegend in einer Ebene befinden, so wird eine stabilere Anfangskonfiguration erreicht, wenn die Ausgleichung ausgehend vom ersten Bildpaar sequenziell durchgeführt wird. Sind alle Bilder einer Fassade sequentiell angehängt, wird die gesamte Bildsequenz einer Fassade gemeinsam ausgeglichen.

Das Schema, das in Abschnitt 3.3.1 erläutert wurde, wird auf alle Bilder der Sequenz erweitert, die homologe Punkte mit dem ersten Bild teilen. Dabei ergeben sich folgende Erweiterungen im Funktionalmodell:

- Für bereits aus dem Bildpaar bestimmte Objektpunkte (X_i, Y_i, Z_i) werden die ermittelten Koordinaten als Schätzwerte für die unbekannten Objektkoordinaten in der Ausgleichung verwendet. Durch das zusätzliche Einbringen als Beobachtungen ist es möglich, die Objektpunkte zu gewichten. Das erlaubt es, Punkte, die nicht auf einer Fassadenebene liegen oder im vorangegangenen Schritt aus falschen homologen Punkten erstellt wurden und eines entsprechenden Restfehler aufweisen, mit einem niedrigeren Gewicht zu versehen. Die bereits verbesserten beobachteten Bildpunkte (x_{ik}, y_{ik}) und (x_{ik+1}, y_{ik+1}) des Bildpaares sowie die nicht verbesserten beobachteten Bildpunkte (x_{ik+2}, y_{ik+2}) des dritten Bildes werden als Beobachtungen eingeführt, soweit das dritte Bild über einen homologen Punkt des Objektpunktes (X_i, Y_i, Z_i) verfügt.
- Für neu zu bestimmende Objektpunkte, die sich aus homologen Punkten des zweiten und dritten Bildes ergeben, werden die Bildpunkte als Beobachtungen gewählt und die Objektkoordinaten als Unbekannte, deren Schätzwert aus Gleichung 3.5 bestimmt wird.
- Die äußeren Orientierungen $(X_{0k}, Y_{0k}, Z_{0k}, \omega_k, \phi_k, \kappa_k, X_{0k+1}, Y_{0k+1}, Z_{0k+1}, \omega_{k+1}, \phi_{k+1}, \kappa_{k+1})$ aller bereits ausgeglichenen Bilder werden sowohl als Beobachtungen wie als Unbekannte eingeführt. Für die äußere Orientierung $(X_{0k+2}, Y_{0k+2}, Z_{0k+2}, \omega_{k+2}, \phi_{k+2}, \kappa_{k+2})$ des neu hinzugefügten Bildes werden die Näherungswerte des GPS Pfades verwendet.
- Die Anzahl der Parameter der inneren Orientierung bleibt gleich, da sie über den Verlauf einer Sequenz als konstant angenommen werden und bei der Hinzunahme weiterer Bilder erneut verbessert werden.
- Die Nebenbedingungen für die Lage der Objektpunkte auf einer Fassadenebene bleiben erhalten und werden auf neue Objektpunkte ausgedehnt. Für bereits ausgeglichene Objektpunkte wird das Gewicht aus dem Kehrwert des Abstands zur Fassadenebene bestimmt. Für neue Punkte wird es wie im Falle eines Bildpaares geschätzt. Die Zuordnung zu einer Fassade wird für alle Punkte wie im Bildpaar beschrieben in jeder Iterationsstufe ausgeführt.

Liegen aus mehr als zwei Bildern Beobachtungen (x_{ik}, y_{ik}) eines unbekannten Objektpunktes (X_i, Y_i, Z_i) vor, so stellt sich die Frage, welche Beobachtungen für die initiale Positionsschätzung verwendet werden sollen. Zwar gibt es die Möglichkeit, über einen analytischen Ansatz alle Abbildungsstrahlen zu schneiden, dies kommt jedoch bereits einer kleinen Ausgleichung sehr nahe und ist für die Schätzung eines Startwertes zu umfangreich. Die Genauigkeit der Schätzung der 3D Objektkoordinaten hängt von der Basis der verwendeten Bilder ab. Bei einer kleinen Basis tritt ein schleifender Schnitt auf, das heißt, dass der Winkel zwischen den Abbildungsstrahlen sehr klein wird und dadurch die Standardabweichung in Richtung der Abbildungsstrahlen steigt. Daher bietet sich eine große Basis für eine genauere Schätzung der Objektkoordinaten an. Zugleich muss dabei aber die Genauigkeit der äußeren Orientierung berücksichtigt werden. Diese ist zu Beginn der Ausgleichung unbekannt und wird ihrerseits für das stochastische Modell geschätzt. Ausgehend vom ersten Bildpaar, das in Abschnitt 3.3.1 ausgeglichen worden ist, darf man bei einer geordneten Bildsequenz davon ausgehen, dass bei einem dritten Bild, das hinzugefügt wird, die meisten homologen Punkte sowohl im ersten wie im zweiten Bild ebenfalls vorkommen. Damit existiert für ihre Objektkoordinaten bereits eine Schätzung in Form des Ergebnisses aus der Orientierung des Bildpaares. Dieses Ergebnis wird als Schätzwert in die Ausgleichung der erweiterten Sequenz eingebracht. Lediglich für homologe Punkte, die ausschließlich das zweite und das dritte

Bild verbinden, muss eine Schätzung durchgeführt werden. Dies geschieht analog zum Bildpaar durch einen räumlichen Vorwärtsschnitt. Dabei wird für das zweite Bild die korrigierte äußere Orientierung aus der bisherigen Sequenz verwendet. Für das dritte Bild wird die initiale Schätzung aus dem GPS Pfad verwendet. Für die innere Orientierung wird das Ergebnis der bisherigen Ausgleichung eingesetzt. Für die neu hinzugekommenen unbekannten Objektpunkte werden sowohl die Objektkoordinaten (Gl. 3.5) wie auch die zugehörige Modellfläche (Gl. 3.4) bestimmt.

Dieser Ablauf wird iterativ wiederholt (Abb. 3.6), bis alle Bilder, die dieselbe Fassade zeigen, über eine Ausgleichung verbunden sind. Für die nächste in der Sequenz sichtbare Fassade, werden die Bilder und ihre Beobachtungen und Unbekannten entfernt, in denen die nächste Fassade nicht zu sehen ist. Es bleiben die bereits ausgeglichenen Bilder, die die nächste Fassade zeigen. Hier werden ebenfalls iterativ alle Bilder ergänzt, die diese Fassade zeigen und bisher nicht in der Ausgleichung enthalten sind.

Dieses Vorgehen bietet zwei Vorteile. Zum einen wird die Ausgleichung nicht unnötig komplex, indem Beobachtungen, die keinen direkten Einfluss auf das Ergebnis haben, weggelassen werden. Zum anderen ist das Ziel der Ausgleichung nicht die Suche nach dem besten Pfad der gesamten Sequenz, sondern die Suche nach der besten Projektion auf die einzelnen Fassaden. Durch die Beschränkung der Beobachtungen auf diejenigen, die die Fassade zeigen, richtet sich das Ergebnis der Ausgleichung maßgeblich nach der entsprechenden Fassade. Gleichzeitig bleibt durch die Überschneidung bei Bildern, die zwei oder mehrere Fassaden zeigen, eine Stützung der Ausgleichung durch bereits verbesserte äußere Orientierungen und berechnet Objektpunkte erhalten. Ein Nachteil dieser Vorgehensweise liegt allerdings darin, dass die innere Orientierung nicht über die gesamte Bildsequenz festgehalten wird, sondern nur jeweils für die Bilder, die eine bestimmte Fassade zeigen. Somit werden Fehler in der relativen Orientierung teilweise auf die innere Orientierung geschoben, die in Wahrheit für die gesamte Sequenz konstant ist.

4 Textureextraktion aus Infrarotbildsequenzen

Im Anschluss an die in Kapitel 3 beschriebene Bestimmung der Inneren und Äußeren Orientierung der Bilder einer geordneten Sequenz von Thermalaufnahme wird in diesem Kapitel die Extraktion der Fassadentexturen auf Basis der durchgeführten Orientierung erläutert. Verfahren wie das Semi-Global-Matching (Hirschmüller [2008]) nutzen im Anschluss an eine mit markanten Punkten durchgeführte Äußere Orientierung zwei oder mehr Bilder, um dort gleichzeitig auf Basis aller Bildpixel eine dichte 3D Punktwolke zu erzeugen. Sie berechnen die Intensitätswerte der 3D Objektpunkte aus allen an der Berechnung seiner Position beteiligten Pixeln. Ein solch aufwendiges Verfahren ist hier nicht notwendig, da mit dem 3D Gebäudemodell und der bekannten inneren und äußeren Orientierung direkt ein 3D Objektpunkt für jedes in den Bildraum projizierte Pixel bestimmt werden kann. Damit ist eine Extraktion der Textur einer Fassade aus einem Einzelbild möglich. In Abschnitt 4.1 wird zunächst die Extraktion der Texturen aus Einzelbildern (Abschnitt 4.1.1) und anschließend die Kombination dieser Einzeltexturen zur gesamten Fassadentextur (Abschnitt 4.1.2) erläutert. Die Auswertung der Texturen in Abschnitt 4.2 gliedert sich in zwei Gruppen. Zum einen werden Ansätze diskutiert, wie Verfahren der Digitalen Bildverarbeitung genutzt werden können, um 3D referenzierte oder semantische Informationen von Fassaden zu extrahieren (Abschnitt 4.2.1). Zum anderen wird diskutiert, wie mehrere extrahierte Texturen derselben Fassade genutzt werden können, um zeitliche Verläufe, Aufnahmen aus unterschiedlichen Blickrichtungen, Aufnahmen verschiedener Kameras und Aufnahmen aus unterschiedlichen Wellenlängenbereichen (Abschnitt 4.2.2) zu kombinieren und damit einen Mehrwert gegenüber einer manuellen Beurteilung von Einzelaufnahmen zu generieren.

4.1 Textureextraktion aus Bildsequenzen mit konstanter Blickrichtung

Für das hier vorgeschlagene Verfahren gelten folgende Randbedingungen, die teilweise bereits durch die Verarbeitungsschritte in Kapitel 3 erfüllt sind:

- Die innere und äußere Orientierung der Bilder ist bekannt. Die äußere Orientierung liegt im selben Koordinatensystem wie das Gebäudemodell vor.
- Bei der Bildsequenz, die zur Extraktion der Texturen verwendet wird, sind die Aufnahmen entlang eines Pfades sortiert.
- Der Blickwinkel der Kamera bleibt über die Sequenz hinweg annähernd konstant.

- Die Fassade wird nur mit einem Bildstreifen aufgenommen, der entweder die gesamte Höhe der Fassade abdeckt oder zumindest den Teil, der als Textur modelliert werden soll.
- Einer Fassadentextur ist genau eine Bildgröße zugeordnet.

Die erste Bedingung ist durch die bereits durchgeführte Orientierung in Kapitel 3 erfüllt. Auf Grund der beschränkten Messgenauigkeit und der Ausgleiche verbleiben dennoch kleinere Restfehler. Die Projektion der Einzelbilder auf das Gebäudemodell wird in Abschnitt 4.1.1 dargestellt. Die Korrektur der verbliebenen Abbildungsfehler wird in Abschnitt 4.1.2 im Rahmen der Kombination der Teiltexturen der Einzelbilder zur Gesamttextur einer Fassade erläutert. Die zweite Bedingung erfordert eine Sequenz von Bildern, deren zeitliche Abfolge einem Abfahren oder Abscannen der Fassade entspricht. Die dritte Bedingung fordert, dass das Kamerasystem fix auf einem Träger montiert ist und keine Schwenkbewegungen durchgeführt werden. Natürlich kommt es bei einer realen Aufnahme auf Grund der Bewegung des Trägers zu kleinen Schwankungen. Diese wirken sich jedoch auf die prinzipielle Gültigkeit der getroffenen Annahme nicht aus. Die vierte Bedingung schließlich schränkt die Extraktion auf einen Bildstreifen ein. Das heißt, eine Fassadentextur wird aus einem Bildstreifen einer Kamera erzeugt. Die Texturen werden nicht aus Teilbildern mehrerer Kameras zusammengesetzt. Die drei zuletzt genannten Bedingungen erlauben es, von einer sortierten Abfolge von Bildern auszugehen. Die letzte Bedingung dient der Vereinfachung der Texturkombination. Bei der Projektion der Bilder auf die Fassade handelt es sich um eine geometrische Transformation, auf die eine Interpolation der Intensitätswerte erfolgt. In diesem Schritt wird sichergestellt, dass anschließend alle Teiltexturen in derselben Bildgröße in Pixeln vorliegen und jeweils die sichtbaren Bereiche der Fassade in der Textur mit Werten besetzt sind. Dadurch erspart man sich bei der Kombination der Texturen eine erneute Umrechnung und Interpolation. Der Nutzen dieser Bedingungen wird in Abschnitt 4.1.2 ersichtlich. Es ergeben sich dadurch deutliche Vereinfachungen bei der Kombination der Teiltexturen der einzelnen Bilder.

Die genannten Bedingungen werden durch eine Kamera, die auf einem Fahrzeug montiert ist, das sich entlang einer Straße bewegt, erfüllt. Für die Frage des festen Blickwinkels dieser Kamera ergeben sich drei grundsätzliche Konfigurationsmöglichkeiten:

- Die Kamera blickt in Nadirsicht frontal auf die Fassade.
- Die Kamera blickt seitlich vorwärts schauend.
- Die Kamera blickt seitlich rückwärts schauend.

Die Blickrichtung der Kamera hat einen direkten Einfluss auf die Kombination der Teiltexturen aus den Einzelbildern und wird in Abschnitt 4.1.2 genauer erläutert.

4.1.1 Projektion der Einzelbilder

Unter der Annahme, dass innere und äußere Orientierung bekannt sind und die äußere Orientierung im selben Koordinatensystem vorliegt wie das zu texturierende Modell, kann eine Texturextraktion aus einem Einzelbild mittels Projektion erfolgen. Dafür wird im Allgemeinen eine indirekte geometrische Transformation durchgeführt, bei der das Bild, dessen Pixelwerte zu bestimmen sind, in das Ursprungsbild, aus dem die Intensitätswerte berechnet werden sollen, projiziert wird (Luhmann [2003]). Die Fassadentextur ist über ihre 2D Bildkoordinaten mit der 3D Fassadenebene verbunden. Diese ist wiederum über ihre 3D Objektpunkte, die in Kapitel 3 bestimmt wurden, mit dem Bildkoordinatensystem der Kamerabilder verbunden. Da es sich bei der Fassadenebene um eine planare Fläche handelt, kann eine Projektion zwischen Bildkoordinaten

der Kamera und Bildkoordinaten der Textur über eine ebene Entzerrung bzw. Homographie erfolgen. Es sind vier korrespondierende Punkte in beiden Bildern nötig, um die acht Parameter der Entzerrung zu bestimmen (Luhmann [2003]). Bei der indirekten geometrischen Transformation werden die Pixel der Textur in das Kamerabild projiziert. Hierfür werden die vier Eckpunkte der Textur verwendet.

Sei die der Kamera zugewandte Vorderseite einer Fassade definiert durch die Richtung des Normalenvektors und die Fassadentextur festgelegt auf eine Größe von $m \times n$ Pixeln. Seien des Weiteren die Pixelkoordinaten (r, c) der Eckpunkte des rechteckigen Fassadenpolygons zu $P_{lo} = (0, 0)$, $P_{lu} = (m - 1, 0)$, $P_{ru} = (m - 1, n - 1)$, $P_{ro} = (0, n - 1)$ sowie die 3D Objektkoordinaten (X_p, Y_p, Z_p) für alle vier Eckpunkte der Fassade gegeben. Dann können über die Kollinearitätsgleichungen (Gl. 3.6) die Eckpunkte des Fassadenpolygons ins Kamerabild projiziert werden. Damit existieren vier Punktkorrespondenzen zwischen der Fassadentextur und dem Kamerabild, aus denen die ebene Entzerrung bzw. Homographie bestimmt werden kann. Diese wird benutzt, um für alle Pixel der Fassadentextur die Pixelkoordinaten im Kamerabild zu ermitteln und dort den Intensitätswert des Fassadenpixels durch bilineare Interpolation zu bestimmen. Für Pixel des Texturbildes, für die keine vier Nachbarn für eine Interpolation zur Verfügung stehen, weil sie am Bildrand oder außerhalb liegen, wird der Wert im Texturbild auf NaN gesetzt, um zu kennzeichnen, dass dieser Teil der Fassadentextur nicht im Eingangsbild sichtbar ist.

Betrachtet man für eine solche Projektion die geometrische Auflösung der Fassadentextur, so spielt neben dem Basis-Höhen-Verhältnis, gegeben durch den Abstand der Fassade von der Kamera einerseits und der Brennweite der Kamera und der Anzahl und Größe der Sensorelemente andererseits, der Blickwinkel der Kamera auf die Fassade eine große Rolle. Im Nadirfall blickt die Kamera in Richtung des Normalenvektors der Kamera. Somit ist die geometrische Auflösung der Textur konstant. Bei einer schrägen Sicht ist dies nicht der Fall. Die geometrische Auflösung ist in verschiedenen Bereichen der Textur unterschiedlich. Ein Vorteil der Schrägsicht ergibt sich aber, wenn Fassaden nicht nur parallel, sondern auch senkrecht zur Fahrtrichtung stehen. In diesem Fall sieht eine 90° seitwärts blickende Kamera die senkrecht zur Fahrtrichtung stehenden Fassaden fast nicht. Darüber hinaus ergibt sich die Frage der besten erreichbaren geometrischen Auflösung. In den Vorbedingungen für die Texturextraktion ist festgesetzt worden, dass eine Fassade nur mit einem Bildstreifen aufgenommen wird. Um die gesamte Fassade abbilden zu können, muss sie also in ihrer gesamten Höhe in einen Bildstreifen passen. Damit ergibt sich im optimalen Fall, in dem die Kamera exakt die Fassade abbildet, die geometrische Auflösung aus der Fassadenhöhe geteilt durch die Anzahl der Zeilen des Kamerasensors. Wählt man einen Blickwinkel schräg vorwärtsschauend, so dass am rechten Bildrand die Fassade in ihrer vollen Höhe sichtbar ist, so ist dort die geometrische Auflösung ebenso hoch wie im Nadirfall. Je weiter man zum linken Bildrand kommt, umso höher wird die geometrische Auflösung im Bild bei einem kleiner werdenden Bildausschnitt. Die Veränderung der geometrischen Auflösung ist beispielhaft für einen schräg vorwärts gerichteten Blickwinkel in Abbildung 4.1 dargestellt. Dabei repräsentiert weiß die höchste Auflösung im Bild, schwarz die niedrigste. Rot bedeutet, dass für diese Pixel der Fassadentextur aus diesem Bild kein Wert berechnet werden konnte, weil dieser Bereich der Fassade im Bild nicht sichtbar ist.

Diese schräg vorwärts blickende Aufnahmegeometrie erhöht also die geometrische Auflösung der Fassadentextur auf Kosten des Blickfeldes. Geht man davon aus, dass bei einer Vorbeifahrt an der Fassade eine Sequenz von Bildern aufgenommen wurde, so zeigt jedes einzelne dieser Bilder die in Abbildung 4.2 a gezeigte Auflösungsverteilung. Liegen die Bilder in einer geordneten Sequenz vor, so verschiebt sich zwischen einem Bild und seinem Nachfolger der sichtbare Bereich einer Fassade und damit entsprechend auch der sichtbare Teil der Fassadentextur. Abbildung 4.2 b verdeutlicht, dass in diesem Fall der im Bild 2 sichtbare Bereich einer Fassade in der zweiten

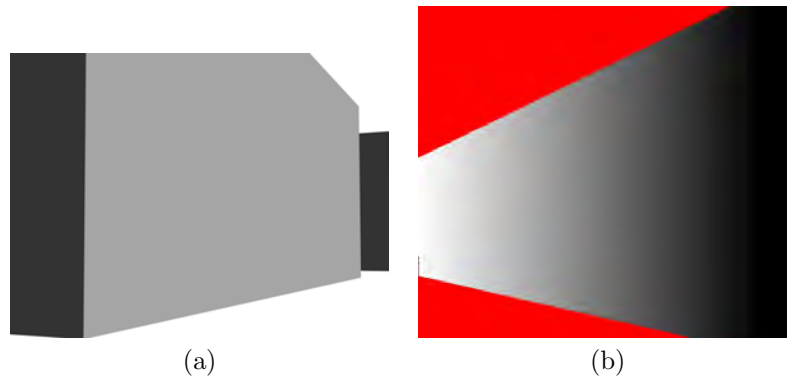


Abbildung 4.1: Schema der geometrischen Auflösung einer Einzelbildtextur bei schräg vorwärts blickender Aufnahmerichtung: a) Blick auf eine Fassade des Gebäudemodells schräg nach vorne blickend. b) Auflösungsänderung über die Fassade; weiß: höchste Auflösung, schwarz: niedrigste Auflösung, rot: nicht gesetzter Wert für im Bild nicht sichtbare Bereiche der Fassade

Textur eine höhere geometrische Auflösung verglichen mit demselben Bereich der ersten Textur hat, die aus Bild 1 extrahiert wurde. Da alle Teiltexturen einer Fassade, die aus Einzelbildern extrahiert wurden, auf dasselbe Texturraster der Gesamttextur der Fassadenfläche projiziert werden, kann folgende Annahme getroffen werden: Jedes Pixel einer Teiltextur, das mit einem Wert belegt ist, hat eine höhere geometrische Auflösung als der an dieser Stelle aus vorangegangenen Teiltexturen gespeicherte Wert. Außerdem belegt die neu hinzugenommene Teiltextur neue, bislang nicht sichtbare Bereiche der Fassadentextur mit Werten. Geht man von einer ausreichend hohen Aufnahmefrequenz der Sequenz aus, so kann man vereinfacht eine Auflösungsverteilung annehmen, wie sie in Abbildung 4.2 c dargestellt ist. Es ergibt sich für eine Gesamttextur einer Fassade, dass ihre geometrische Auflösung damit mindestens der einer Nadiraufnahme entspricht. Die höchste geometrische Auflösung besitzt die Textur im Bereich der Aufnahmehöhe der Kamera, bei einer Kamera in einem Fahrzeug also bei ungefähr 1-2 Metern Höhe, bei einer Kamera an einem Ausleger oberhalb des Fahrzeugs bei 2-3 Metern oder mehr. Nach oben und unten nimmt die geometrische Auflösung kontinuierlich ab.

Anzumerken ist, dass diese Auflösungsverteilung nur für Texturen gilt, die parallel zur Fahrtrichtung der Kamera, im Regelfall also entlang der Straße angeordnet sind. In Fällen, in denen Fassaden schräg oder quer zur Fahrtrichtung stehen, sind sie entweder in einer Schrägsichtaufnahme gar nicht zu sehen oder die Kamera bewegt sich im Laufe der Sequenz auf die Fassade zu. Damit nimmt die geometrische Auflösung ebenfalls zu bei einem kleiner werdenden Ausschnitt der Fassade. Das Konzept des Überschreibens der bisherigen Textur mit den sichtbaren Bereichen der neuen Textur bleibt also uneingeschränkt erhalten.

Die schräge Aufnahmegeometrie erlaubt einerseits die Aufnahme von Fassaden, die quer zur Fahrtrichtung stehen und führt zu unterschiedlichen geometrischen Auflösungen in verschiedenen Bereichen der Textur. Andererseits entstehen dadurch auch Verdeckungen. Bei der Eigenverdeckung verdeckt eine Fassade eine andere desselben Gebäudeteils. Bei der Fremdverdeckung verdeckt ein Gebäudeteil einen anderen. Dieses Problem ist nur dadurch zu reduzieren, dass man neben einer Aufnahmeserie in schräg vorwärts gerichteter Blickrichtung eine weitere in schräg rückwärts blickender Aufnahmerichtung durchführt. Wenn man das Auflösungsschema aus Abbildung 4.2 betrachtet, so kann dieses Schema leicht auf eine rückwärts blickende Kamera angepasst werden. Die Bilder der Sequenz werden immer noch in Fahrtrichtung von links nach rechts aufgenommen. Durch den Blick nach schräg hinten zeigt nun aber nicht mehr der linke Teil eines Bildes einen kleinen Fassadenausschnitt bei großer geometrischer Auflösung, sondern der rechte Teil. Um die Teiltexturen der Sequenz also optimal übereinander zu legen, wird die Sequenz jetzt

in umgekehrter Reihenfolge gebildet. Die vollständige Textur wird beginnend mit der Teiltextur des zuletzt aufgenommenen Bildes gebildet und arbeitet sich in der Sequenz nach vorne. Am Ende steht eine Auflösungsverteilung wie in Abbildung 4.2 c gezeigt.

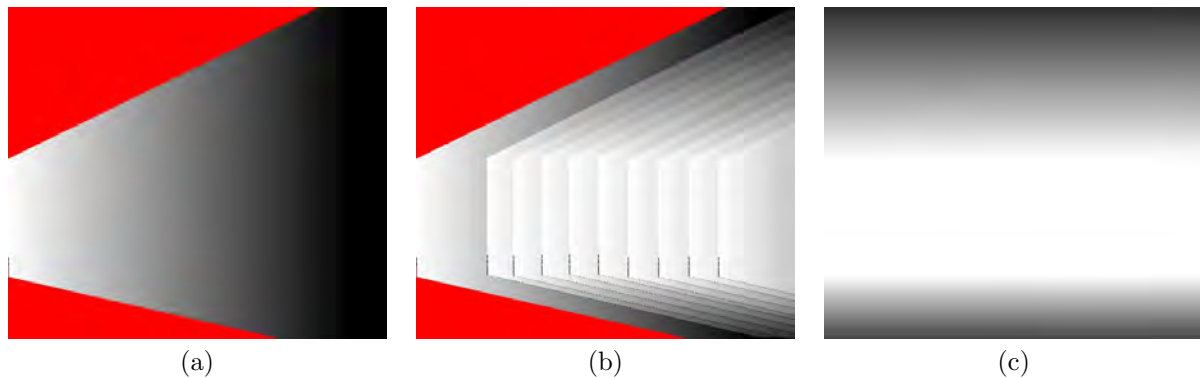


Abbildung 4.2: Schema der geometrischen Auflösung einer zusammengesetzten Fassadentextur bei schräg vorwärts blickender Aufnahmerichtung, weiß: höchste geometrische Auflösung, schwarz: niedrigste geometrische Auflösung, rot: nicht sichtbare Teile der Fassade. a) Einzelbildtextur mit nach rechts abnehmender geometrischer Auflösung bei größer werdendem Blickfeld. b) Überlagerung der geometrischen Auflösungen mehrerer Bilder einer Sequenz: Jede Folgetextur sieht nur Bereiche, in denen sie eine höhere Auflösung hat als die bereits vorhandenen Teiltexturen. c) Geometrische Auflösung der zusammengesetzten Fassadentextur: Die geometrische Auflösung ist auf der Höhe der Kamera am höchsten und nimmt nach oben und unten ab. Insgesamt ist die geometrische Auflösung höher als bei einer Bildsequenz in Nadiraufnahme.

4.1.2 Koregistrierung von Teiltexturen

Die in Abschnitt 4.1.1 erläuterten Zusammenhänge zur geometrischen Auflösung der zusammengesetzten Fassadentextur gelten nur unter der Bedingung, dass die Teiltexturen aus fehlerfrei projizierten Kamerabildern erzeugt werden. Trotz der Bündelblockausgleichung in Kapitel 3 können kleinere Restfehler bei der Projektion erhalten bleiben. Diese betreffen sowohl die Rotation der Textur, wie auch die Position und die Skalierung. Auf Grund der Bündelblockausgleichung wird davon ausgegangen, dass es sich hierbei nur um kleine Abweichungen handelt, die durch lineare Transformationen behoben werden können.

Zwei Teiltexturen, die aus zwei aufeinander folgenden Bildern einer Sequenz aufgenommen wurden, zeigen eine große Überlappung. Die Verschiebung des Sichtfeldes zwischen zwei Bildern und damit auch ihren Texturen ist als gering anzusehen. Es kann daher davon ausgegangen werden, dass markante Bildregionen in einer Textur in der anderen ebenfalls vorkommen und eine ähnliche Erscheinung und geometrische Auflösung haben. Daher kann eine Zuordnung zwischen zwei aufeinander folgenden Teiltexturen anhand einer einfachen Korrelation erfolgen. In der vorhandenen Textur werden Patches gebildet und mit Patches in der hinzuzufügenden Textur verglichen. Da der Suchraum auf Grund der erfolgten Projektion als sehr klein anzunehmen ist, können hier auch für wiederkehrende Muster wie Fensterreihen eindeutige Korrespondenzen gefunden werden. Für die hinzuzufügende Teiltextur werden aus vier Patches, die in der vorhandenen und der hinzuzufügenden Textur gefunden wurden und räumlich möglichst an den vier Ecken des gemeinsamen sichtbaren Bereiches liegen sollen, die Parameter für eine Ebene Entzerrung bestimmt, mit der eine verbliebene geometrische Verzerrung behoben und die Teiltextur an die bestehende angefügt wird. Die neue Teiltextur überschreibt mit allen Pixelwerten, die besetzt sind, die bestehende Fassadentextur, da, wie in Abschnitt 4.1.1 erläutert, jedes sichtbare Pixel einer Folgetextur eine höhere geometrische Auflösung hat, wie die bereits zusammengesetzte Textur. Die in der neuen Teiltextur nicht belegten Pixel bleiben unverändert erhalten, wodurch sich die Auflösungsverteilung wie in Abbildung 4.2 c ergibt.

4.2 Auswertung von Fassadentexturen

Bei den erzeugten Fassadentexturen handelt es sich um einkanalige Intensitätsbilder des thermischen infraroten Spektrums. Unterschiedliche Intensitäten können die in Kapitel 2.1 beschriebenen Ursachen haben. So können Kanten, Ecken und Linien einen Hinweis auf Heizungsleitungen oder Fenster geben. Gerade Fenster sind im Hinblick auf Thermaltexturen interessant. Glas zeigt im infraroten Spektrum nicht seine Eigenstrahlung, sondern überwiegend die Reflexion der umliegenden Infrarotstrahlung. Daher sollen Fenster für eine Beurteilung der Fassade detektiert und beispielsweise bei der Berechnung der Fassadentemperatur oder der Suche nach Heizungsleitungen ausgeschlossen werden. Unterschiedliche Materialien führen zu unterschiedlichen Abstrahlungen. Ein Fenstersturz aus Beton hebt sich deutlich von einer Ziegelmauer ab, selbst wenn die gesamte Mauer verputzt ist.

Für die Auswertung von Thermaltexturen kommen im Prinzip alle Verfahren der Digitalen Bildanalyse in Frage, die als Eingangsdaten auf Intensitätswerten basieren. Zum einen gibt es Pixel oder Regionen basierte Segmentierungsverfahren, die flächige Regionen unterschiedlicher Intensität segmentieren. Zum anderen können über Kantenoperatoren linienartige Strukturen und Intensitätssprünge detektiert werden.

Detektierte Primitive werden z.B. mit Hilfe von Grammatiken zu Objekten kombiniert und schließlich als 3D Geometrie dem Gebäudemodell hinzugefügt. Da diese Verfahren meist auf den Primitiven und nicht auf den Intensitätsbildern arbeiten, können sie auf Bilder im thermischen Infrarot übertragen werden. Eine Methode der Fensterdetektion basierend auf der Kombination von Primitiven auf Basis eines Gestaltsystems ist detailliert in Michaelsen et al. [2012] beschrieben.

In dieser Arbeit werden Segmentierungsverfahren eingesetzt, um zunächst Fenster zu detektieren und anschließend Heizungsleitungen bzw. Leckagen. Die 3D Referenzierung der Fassadentexturen erlaubt eine gemeinsame Auswertung mehrerer Fassadentexturen. Mit algebraischen Transformationen werden Aufnahmen zu verschiedenen Zeitpunkten genutzt, um über eine Differenzbildung Temperaturänderungen zu detektieren. Dadurch erzeugte Differenztexturen werden dann wiederum für die Suche nach Segmenten oder Primitiven genutzt. Die Kombination von vorwärts und rückwärts blickenden Texturen hilft bei der Identifikation von Strukturen, die nicht in der Fassadenebene liegen, da eine Differenztextur hier Abweichungen zeigt. Eine Überlagerung einer Thermaltextur mit einer Textur aus dem sichtbaren Spektrum wird verwendet, um die Genauigkeit der Projektion der Thermaltextur zu ermitteln.

Weitere Anwendungsmöglichkeiten, die in dieser Arbeit jedoch nicht behandelt werden, liegen in der Kombination von thermalen Texturen aus dem mittleren und langwelligen thermischen Infrarotspektrum. Durch den Vergleich der Intensitätswerte lässt sich die Schätzung von Emissionskoeffizienten und der Temperatur verbessern. Schließlich ist es durch die Umrechnung in einheitliche Fassadentexturen möglich, Texturen, die mit verschiedenen Kamerasystemen desselben Wellenlängenbereichs aufgenommen wurden, zu vergleichen.

4.2.1 Auswertung von Einzeltexturen

Die Untersuchung einer Einzeltextur kann zwei verschiedene Ziele verfolgen. Zum einen geht es um die Lokalisation und Identifikation von Objekten, zum anderen um die Ableitung von Attributen einer Fassadentextur. Bei den extrahierten Thermaltexturen handelt es sich um einkanalige Intensitätsbilder. Strukturen an der Oberfläche wie Fenster oder Simse zeigen ebenso wie Strukturen unter der Oberfläche, beispielsweise Materialwechsel oder Heizungsleitungen, eine deutliche Intensitätsänderung gegenüber der Fassadenfläche. Auf der Fläche selbst hingegen liegt eine eher homogene Intensität vor, die lediglich durch Wind oder benachbarte Objekte leicht variiert.

Regionen basierte Segmentierungsverfahren erzeugen beginnend mit initialen, zufällig ausgewählten Startpunkten gemäß eines Homogenitätskriteriums in einem iterativen Prozess Pixelregionen. Hier wird beispielsweise das Kriterium Differenz des Intensitätswertes zum Mittelwert eingesetzt. Da Startpunkte nicht zwingend repräsentativ für eine Region sein müssen, insbesondere, wenn man Rauschen mit einbezieht, ist die Differenz zum Startpunkt im Vergleich störanfällig. Die Differenz zum Nachbapixel hat Schwierigkeiten bei weichen Intensitätsübergängen. Ein weiterer untersuchter Ansatz ist das Split-und-Merge-Verfahren. Das gesamte Bild wird in vier gleiche Quadranten unterteilt. Erfüllt ein Quadrant das Homogenitätskriterium, stoppt das Verfahren für diesen Quadranten. Ist das Homogenitätskriterium nicht erfüllt, wird die Aufteilung in vier Teilquadranten iterativ fortgesetzt. Nach jedem Teilungsschritt wird für homogene Quadranten geprüft, ob benachbarte Quadranten zusammengefasst werden können. Für Heizungsleitungen ergeben sich langgezogene Regionen, für Fenster runde und rechteckige Regionen. Die Unterscheidung kann anhand des Verhältnisses der beiden Hauptachsen der segmentierten Region erfolgen.

Reznik & Mayer [2008] greifen für Texturen im sichtbaren Spektrum auf eine Objektbibliothek von Fenstern zurück, die nach einem Markov-Chain-Monte-Carlo Modell (MCMC) mit der beobachteten Fassade verglichen werden. Dabei gilt die Annahme einer gewissen Ähnlichkeit zwischen den Elementen der Bibliothek und den Fenstern der echten Fassade. Eine solche Bibliothek für Thermalaufnahme scheitert am Erscheinungsbild der Fenster im infraroten Spektrum und im speziellen Fall der hier vorliegenden Texturen an der wechselnden geometrischen Auflösung. Ein anderer Ansatz sucht nicht die Fenster direkt, sondern geometrische Primitive, die über eine Grammatik zu Fenstern und Fensterreihen zusammengesetzt werden. Ein solcher Ansatz wird in Michaelsen et al. [2012] verfolgt. Als Beispiele dienen die mit den in dieser Arbeit vorgestellten Verfahren extrahierten Thermaltexturen. Hierbei wird über eine maskierte Korrelation nach Ecken in der Textur gesucht und diese Ecken zu Fenstern kombiniert. Dabei zeigt sich das Problem der Definition von Kriterien für Fenster und Fensterelemente, die über die gesamte Textur bei unterschiedlicher Erscheinung der Fenster und unterschiedlicher geometrischer Auflösung innerhalb einer Textur gültig sind. Ein weiteres Problem sind in diesem Verfahren Fehldetektionen, weil an Leckagen Ecken von Fenstern gefunden werden.

Für die Leckagedetektion sind insbesondere Heizungsrohre interessant. Diese zeichnen sich als segmentweise lineare Strukturen ab. Daher wird hier eine Detektion von Kanten mittels des Canny Kantendetektors (Canny [1986]) durchgeführt. Lange schmale, parallele Linienpaare werden als Heizungsleitungen interpretiert. Kurze Linienpaare mit größerem Abstand als Fenster.

Sind die Objekte auf der Fassade detektiert, so werden ihre Pixel maskiert. Die verbliebene, um Störobjekte bereinigte Fassade wird verwendet, um die Durchschnittstemperatur der Fassade zu ermitteln. Für eine Infrarotkamera ist auf Basis ihrer radiometrischen Kalibrierung eine Umrechnung ihrer gemessenen Strahlungsleistung in einem Bildpunkt zur Temperatur des beobachteten Objektes möglich. Aus diesem Bild lässt sich damit nach der Entfernung der Störobjekte die Fassadentemperatur durch einfache Mittelung berechnen. Dabei wird ein fester Emissionswert von 96% für massives Mauerwerk angenommen, der für verschiedene Mauerwerkstypen repräsentativ ist und eine ausreichende Näherung für die Temperaturberechnung darstellt.

4.2.2 Gemeinsame Auswertung mehrerer Texturen

Da bei der Extraktion der Texturen ein einheitlicher Bildmaßstab in Pixel pro Meter auf den Fassaden verwendet wird, haben alle Texturen einer Fassade denselben Bildmaßstab. Die verschiedenen Intensitätswerte eines Pixels aus den verschiedenen Texturen werden zu einem Vektor kombiniert und gemeinsam ausgewertet. Hierfür müssen die radiometrischen Werte der verschiedenen Sequenzen kalibriert sein. In der radiometrischen Kalibrierung von Infrarotkameras werden

meist bereits Korrekturwerte angebracht, die es erlauben, bei Aufnahmen im selben Wellenlängenbereich die kalibrierten Intensitätswerte zu verwenden.

Wenn man davon ausgeht, dass man mit derselben Kamera zwei Vorbeifahrten an einer Fassade in einem zeitlichen Abstand vornimmt, so zeigt die pixelweise Differenz der beiden Texturen die Änderung der Abstrahlung zwischen beiden Sequenzen, also ein Aufwärmen oder Abkühlen. Diese Differenzbildung ist insbesondere für zwei Anwendungen sinnvoll. In einer einzelnen Textur treten Objekte nicht immer deutlich hervor. Im Laufe eines Tages haben sich zwei unterschiedliche Materialien ähnlichen Emissionsgrades auf eine gleiche Temperatur erwärmt. Beide kühlen jedoch unterschiedlich schnell aus. Ein Differenzbild zeigt dann beide Objekt voneinander getrennt. Eine Abkühlung über die Nacht ist auch ein Hinweis auf die ungefähre Energieabstrahlung eines Gebäudes. Für eine genaue Bewertung sind allerdings weitergehende Informationen über das Gebäude wie Material, Dämmung, Heizung nötig. Vergleicht man zwei Thermaltexturen mit einem größeren zeitlichen Abstand, so kann beispielsweise der Erfolg einer Fassadendämmung bewertet werden, indem über die Differenzbildung eine qualitative Aussage über die erfolgte Dämmwirkung gegeben werden kann.

Gerade wenn Aufnahme mit einigem zeitlichen Abstand wiederholt werden, ist keineswegs garantiert, dass dasselbe Kamerasystem für die Aufnahmen zur Verfügung steht. Unter der Annahme, dass die Größe der Fassadentextur gleich gewählt ist und beide Kameras radiometrisch kalibriert sind, kann auch in diesem Fall ein einfaches Differenzbild über die in der Größe und Radiometrie genormte Textur gebildet werden.

Auf Grund der in dieser Arbeit vorgeschlagenen Schrägsicht der Kamera wird in Teilen der Fassadentextur eine höhere geometrische Auflösung als in anderen Bereichen erreicht (Abb. 4.2). Zugleich ergibt sich durch die Schrägsicht ein Abschattungsproblem (Abb. 4.3). Für eine möglichst vollständige Extraktion von Texturen komplexerer Gebäude wird daher eine zweite Sequenz mit schräg nach hinten blickender Kamera hinzugenommen. Für Fassaden parallel zur Fahrtrichtung liegen zwei Fassadentexturen vor. Für Objektpunkte in der Fassadenebene liegen die korrespondierenden Bildpunkte beider Texturen übereinander (Abb. 4.3a). Für Fassaden quer zur Fahrtrichtung ergibt sich dadurch, dass sie genau in einer der beiden Sequenzen aufgenommen werden (Abb. 4.3b). Für Vorsprünge und Nischen ergeben sich in jeder Textur Abbildungsfehler (Abb. 4.3c), da die Punkte auf die Fassadenebene abgebildet werden, in Wirklichkeit aber vor oder hinter der Ebene liegen. Bei einer Differenzbildung beider Texturen entstehen daher an Positionen mit Abbildungsfehlern auf Grund der Schrägsicht und Projektionsfehler auf die Fassadenebene Restwerte, bei den Pixeln in der Fassadenebene ohne Abbildungsfehler ergibt die Differenz einen Wert nahe 0. Da aus beiden Einzeltexturen die Intensitätswerte bekannt sind, lässt sich aus dem Restwert bestimmen, ob ein Punkt vor oder hinter der Fassadenebene liegt. In Abbildung 4.3c erkennt man, dass in beiden Bildern jeweils nur ein Teil der Nische abgebildet wird. Kombiniert man beide Bilder, so dass die Pixel, die die Fassadenebene zeigen, korrekt übereinander liegen, so sind die Pixel der Nische in beiden Bildern versetzt. In der kombinierten Ansicht fällt bei einer Fensternische die Fensterfläche der Vorwärtssicht auf den Rand der Nische in der Rückwärtssicht und umgekehrt. Damit lassen sich Objekte auf der Fassade identifizieren und erhalten sogar eine zusätzliche Information. Allerdings funktioniert diese Methode nicht bei Abweichungen von der Fassadenebene, die sich waagerecht durch die Fassade ziehen, wie beispielsweise bei einem durchlaufenden Mauervorsprung, der die oberen von den unteren Stockwerken trennt. Hier kommt es zwar ebenfalls zu einem Versatz. Da die Intensitätswerte in Versatzrichtung jedoch konstant sind, bleibt bei einer Differenzbildung kein Restbetrag übrig.

Da für alle Fassaden eine einheitliche Pixelzahl pro Meter und damit für eine Fassade eine einheitliche Maske vorgegeben ist, ergibt sich die Möglichkeit, Texturen verschiedener Wellenlängenbereiche zu kombinieren. Eine Kombination von Texturen, die aus Bildern von verschiedenen

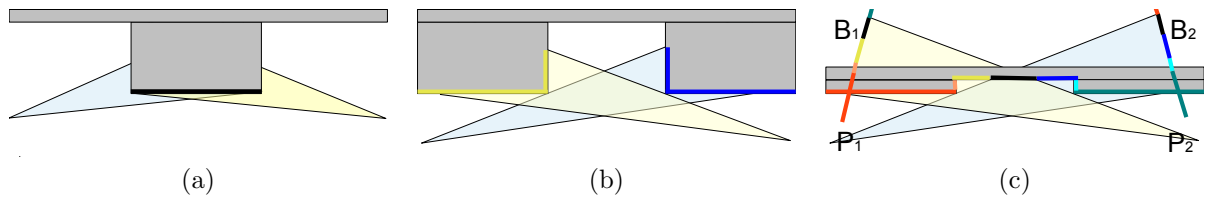


Abbildung 4.3: Unterschiedliche Verdeckungen bei unterschiedlicher Blickrichtung: a) Vorwärts- und Rückwärtssicht auf eine Fassade parallel zur Straße. Die Fassade wird zweimal aufgenommen. b) Vorwärts- und Rückwärtssicht auf eine Fassade mit zurückgesetzten Elementen quer zur Straße. Die quer zur Straße liegenden Fassaden werden nur einmal aufgenommen. Außerdem tritt auf den zurückversetzten Fassaden Verdeckung auf. c) Vorwärts- und Rückwärtssicht an einem Fenster. In beiden Bildern werden auf Grund der projektiven Abbildung unterschiedliche Teile des Fensters abgebildet. Bei einer Überlagerung der beiden Bilder kommt es zu einem Versatz im Bereich der Fensternische.

Sensoren verschiedener Wellenlängen (z.B. RGB, MWIR, LWIR) erstellt wurden, eröffnet die Möglichkeit einer Multispektralklassifikation. In dieser Arbeit wird diese Kombination exemplarisch mit zwei Wellenlängenbereichen des thermischen Infrarots gezeigt. Ebenfalls möglich ist eine Kombination mit einer Fassadentextur aus dem sichtbaren Spektrum. Hierauf wird in dieser Arbeit bei der Bewertung der Abbildungsgenauigkeit der erzeugten Thermaltexturen auf Grundlage einer Textur aus dem sichtbaren Spektrum, die als Ground-Truth dient, kurz eingegangen.

4.2.3 Detektion von Leckagen

Neben der Untersuchung des Temperaturverlaufs über die Zeit oder der Bestimmung der Fassadentemperatur ermöglichen Thermaltexturen die Detektion von Objekten. Dies gilt insbesondere für Objekte, die im normalen sichtbaren Spektrum verborgen bleiben oder deren Detektion durch die zusätzliche Information der Thermaltextur verbessert werden kann. Für Kameras im sichtbaren Spektrum nicht erkennbar sind beispielsweise Heizungsrohre, Leckagen oder Materialübergänge. Sie zeigen im Infrarot einen hohen Kontrast bzw. eine hohe Intensitätsänderung und sind damit über intensitätsbasierte Segmentierungsverfahren detektierbar. Leider zeigen auch Fenster ein solches Verhalten. Sie treten als helle oder dunkle Flächen auf, je nachdem, wie hoch der Anteil reflektierter Strahlung ist. Das Erscheinungsbild von Fenstern unterliegt auf Grund der Vielzahl an Rahmenformen und -größen und unterschiedlichen Unterteilungen der Fensterfläche großen Variationen. Wie auch die Rahmen strahlt das Glas selbst zwar entsprechend seiner Temperatur und seines Emissionsgrades, jedoch wird bei Glas dieser Effekt durch Reflexionen überlagert.

Hier soll ein einfacherer Ansatz verfolgt werden. Leckagen befinden sich in der Fassade, ihre Abbildung also auf der Fassadenebene. Daher sind ihre Intensitätswerte in zwei Texturen, die aus Aufnahmen unterschiedlicher Blickrichtung stammen, an derselben Position in den Bildern. Geometrische Fassadenelemente wie Fenster hingegen befinden sich meist nicht in der Fassadenebene. Bei ausreichend großer Basis zwischen den Aufnahmen und geändertem Blickwinkel werden diese Objekte daher in zwei Texturen versetzt abgebildet (Abb. 4.3c). Maskiert man alle Pixel aus, die in einer Differenztextur hohe Werte haben, so verbleiben Pixel auf der Fassadenebene und Pixel aus Bereichen mit homogener Temperatur. Verbleibende Bereiche mit Intensitäten, die von der normalen Fassadenabstrahlung abweichen, sind Kandidaten für Leckagen. Der Ablauf der Detektion ist in Abbildung 4.4 dargestellt. Für die vorgeschlagene Konfiguration einer vorwärts und einer rückwärts blickenden Aufnahme ist diese Annahme gegeben. Bestimmt man die Differenztextur beider Texturen, ergeben sich für Punkte auf der Fassadenebene Differenzen nahe Null. Für Objekte, die nicht auf der Fassadenebene liegen und daher versetzt projiziert werden, ergeben sich große Differenzen. Anschließend wird das Differenzbild in eine binäre Maske umgewandelt, mit der die Bereiche der Textur ausgeblendet werden, in denen Objekte auf der Fassade

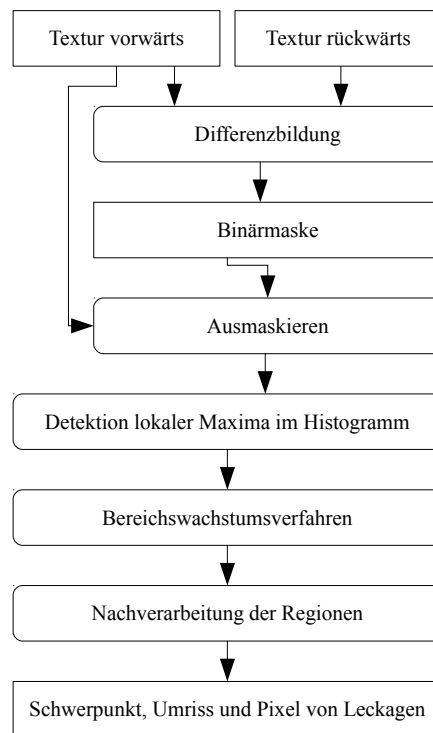


Abbildung 4.4: Ablauf der Leckagedetektion. Nach einer Differenzbildung der Texturen aus Vorwärts- und Rückwärtssicht wird für die Pixel mit hohen Restwerten eine Maske erzeugt, um sie von der weiteren Verarbeitung auszuschließen. Auf den verbliebenen Pixeln findet ein Bereichswachstumsverfahren statt. Regionen, deren Intensität sich von der des absoluten Maximums im Histogramm unterscheidet, werden als Regionen von Leckagen detektiert.

liegen. Über morphologische Operatoren wird die Maske verbessert. Hierzu wird eine Dilatation durchgeführt und anschließend ein Closing. Die so entstandene Binärmaske wird als zusätzliche Eingabe für die Segmentierung der Textur gewählt.

Alle Pixel, die in der Binärmaske markiert sind, werden von der weiteren Verarbeitung ausgenommen. Für die übrigen Pixel wird über ein Histogramm die Intensität der Fassadenfläche ermittelt. Unter der Annahme, dass der Großteil der Fassade nicht von einer Leckage bedeckt ist, ergibt sich die durchschnittliche Fassadentemperatur aus dem absoluten Maximum im Histogramm. Daneben werden sich einzelne lokale Maxima sammeln. Pixel mit Intensitätswerten im Bereich dieser lokalen Maxima sind Kandidaten für Leckagen. Ein Bereichswachstumsverfahren basierend auf den lokalen Grauwertdifferenzen liefert schließlich eine Segmentierung der gesamten Textur, ausgenommen der ausmaskierten Pixel. Die Segmente, deren Mittelwert nah am ermittelten absoluten Maximum des Histogramms liegen, werden zur Fassadenfläche zusammengefasst. Die übrigen Segmente werden als mögliche Leckagen markiert.

Für die segmentierten Leckagen werden Umriss und Schwerpunkt bestimmt. Der Umriss ergibt sich aus einer binären morphologischen Konturbildung, bei der vom Originalsegment die Erosion des Segments abgezogen wird. Der Schwerpunkt ergibt sich aus der Mittelung der Koordinaten der Pixel des Segments. Über die Texturkoordinaten werden Schwerpunkt und Umriss im 3D Gebäudemodell gespeichert und erhalten so über die Fassade eine 3D Referenz. Weitere Schritte wie die Bestimmung eines Polygons für die Segmente werden hier nicht weiter verfolgt. Als Nebenprodukt fällt die um Störungen von Objekten und Leckagen bereinigte Fassadentemperatur als Maximum des zuvor bestimmten Histogramms der Textur ab.

5 Experimente

5.1 Beschreibung der Testszenarien

Für ein geeignetes Testszenario sind zunächst verschiedene Randbedingungen zu definieren:

- Neben einem ausreichend genauen 3D Gebäudemodell für die Texturextraktion ist es notwendig, die Fassaden von einem Fahrzeug aus möglichst frei von Fremdverdeckungen einsehen zu können.
- Die Fassaden sollen außerdem von der Länge und Höhe so gewählt sein, dass sie von einer einzelnen Aufnahme nicht erfasst werden können und somit das Verfahren der Orientierung der Bildsequenz überhaupt notwendig ist.
- Es werden gleichermaßen lange homogene Fassaden untersucht wie auch Fassaden, die aus mehreren versetzten Elementen bestehen, so dass es zu einer Eigenverdeckung verschiedener Gebäudeteile kommt.
- Da für die Koregistrierung der Bildsequenz und des Modells GPS Koordinaten verwendet werden, wird außerdem ein Szenario benötigt, das sowohl Bereiche mit guter GPS Signalqualität beinhaltet, wie auch Bereiche mit Sichtbarkeits- und Reflexionseinschränkungen. Zur Kontrolle wird der Kamerapfad zusätzlich über eine Tachymetermessung verfolgt.
- Für das Gebäude sollen Ground Truth Daten für die Position von detektierten Fassadenelementen vorhanden sein oder im Rahmen der Qualitätsuntersuchung ermittelt werden können.
- Das Objekt muss zu Zeiten beobachtet werden, die günstig für Thermalaufnahmen sind, d.h. während der Heizperiode, wenn kalte Nächte auftreten. Schnee und Regen sollten vermieden werden. Um Verdeckungen durch Vegetation einzuschränken, wird ein Zeitpunkt ohne Belaubung gewählt.
- Für die Untersuchung der zeitlichen Veränderungen sind Aufnahmen kurz nach Sonnenuntergang und kurz vor Sonnenaufgang durchzuführen.
- Für die Untersuchungen zur Kombination von Texturen sollen zwei Wellenlängenbereiche im mittleren und langwelligen thermischen Infrarot aufgezeichnet werden. Darüber hinaus wird über eine zusätzliche Videokamera eine RGB Farbbildsequenz aufgezeichnet. Die Fassadenvorbeifahrten sind jeweils in derselben Fahrtrichtung mit schräg vorwärts und schräg rückwärts blickender Kamera durchzuführen.

Auf der Basis dieser Anforderungen wurde das Stammgelände der Technischen Universität München (TUM) als Testgebiet ausgewählt. Abbildung 5.1 zeigt das Stammgeländes der TUM,

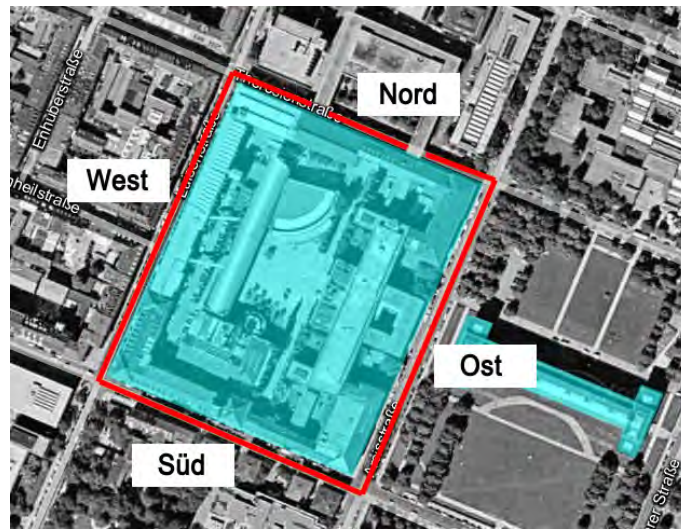


Abbildung 5.1: Luftbild des Stammgeländes der TUM (Google). Während der Gebäudekomplex an der Nord-, West- und Südseite von dichter Bebauung umgeben ist, liegt auf der Ostseite eine Grünanlage, in deren Mitte lediglich die Alte Pinakothek steht. Die als Testdaten verwendeten Gebäudeteile sind blau unterlegt. Die mit dem Fahrzeug abgefahrenen Kamerapfade sind in rot markiert. Es wurde jeweils vorwärts und rückwärts blickend aufgenommen.

das an der Nord- West-, und Südseite an dichte Bebauung angrenzt und eine in mehreren Bauabschnitten erstellte, aber weitgehend ebene Fassaden zeigt. Auf der Ostseite ist die Fassaden in mehrere versetzte Elemente verschiedener Struktur unterteilt. Gegenüber befindet sich eine Grünanlage, die lediglich durch die Alte Pinakothek unterbrochen wird, die im Modell rechts zu sehen ist. Die Gebäude, für die Testdaten aufgenommen werden, sind farblich in blau unterlegt. Die gefahrenen Kamerapfade sind rot markiert. Jede Strecke wurde vorwärts und rückwärts blickend nach Sonnenuntergang aufgenommen. Die rechte Strecke zwischen dem blau markierten Hauptgelände und der auf der Wiese stehenden Pinakothek wurde zusätzlich nochmals am darauffolgenden Morgen vor Sonnenaufgang aufgenommen.

5.2 Verwendete Datensätze

In diesem Abschnitt werden die Spezifikationen der in für die Experimente verwendeten Kameras vorgestellt. Für die Thermalaufnahmen werden in dieser Arbeit drei Kamerasysteme verwendet. Zum einen kommt eine VarioCam hr von Infratec zum Einsatz. Mit dieser Kamera werden Fassadenbilder in Nadirsicht aufgenommen. Diese Kamera wird über ein Testfeld kalibriert. Zum anderen kommen die in einem Messfahrzeug verbauten Kameras FLIR SC3000 für das thermische langwellige Infrarotspektrum und MerlinMID für das thermische mittlere Infrarotspektrum zum Einsatz. Diese beiden Kameras sind zusammen mit einer Videokamera in einem Messausleger eines Kleintransporters verbaut, der über einen GPS Empfänger verfügt (Abb. 5.2). Am Kamerahalter ist zusätzlich ein Prisma zur Verfolgung mit einem Tachymeter angebracht.

5.2.1 Infratec VarioCam hr 384×288

Bei der Infratec VarioCam hr handelt es sich um eine Bolometer Kamera, die im Spektralbereich von $7\text{-}12\mu\text{m}$ arbeitet. Mit einer Detektorgröße von 384×288 Bildpunkten bei einer thermalen Auflösung von $0,08\text{ K}$ ist sie die Kamera mit der größten Bildauflösung. Diese Kamera wird eingesetzt, um Aufnahmen der Fassaden in Nadirsicht aufzuzeichnen. Bild 5.3 zeigt ein Bildpaar aus der in Nadirsicht aufgenommenen Sequenz. Die Daten werden als einkanalige Intensitätsbilder

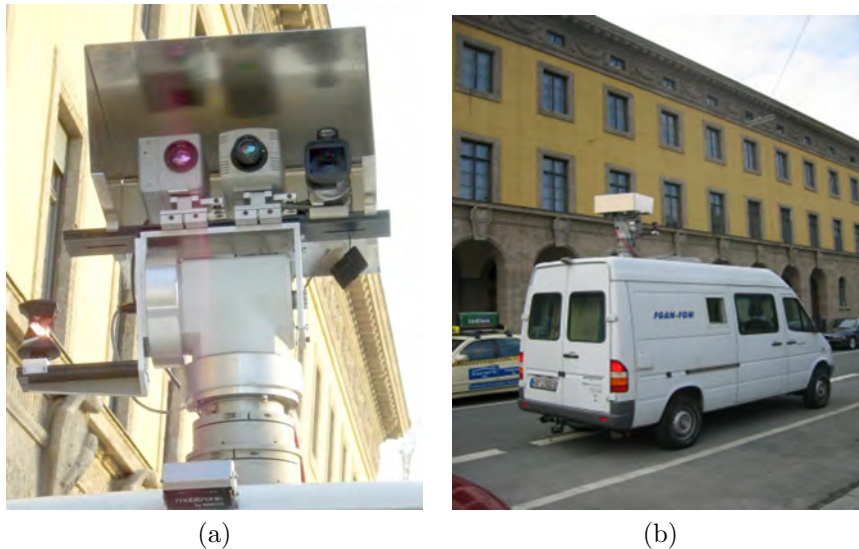


Abbildung 5.2: Kameraaufbau des Messfahrzeugs. (a) Anordnung der Kameras auf dem Ausleger: FLIR SC3000 (links), MerlinMID (mitte), Videokamera (rechts). Am linken Rand ist ein Prisma zu sehen, das vom Tachymeter angemessen wird. (b) Messfahrzeug bei einer Testfahrt mit schräg vorwärts blickendem Kamerasystem.

aufgenommen und sind hier über eine Lookup-Table eingefärbt, um bei der visuellen Betrachtung einen besseren Kontrast zu erzielen. Man erkennt deutlich die Fenster, die in dieser Perspektive das gegenüberliegende Bauwerk spiegeln. Mit leicht erhöhter Temperatur heben sich auch die Heizungsleitungen in der Außenwand ab. Außerdem erkennt man einen Sims zwischen dem Erdgeschoss und dem ersten Stock, der kalt ist und einen Wärmestau der aufsteigenden Luft verursacht sowie die auf der Fassade angesetzten Steinplatten mit den Emblemen der Fakultäten. Die Kamera ist mit einem 45° Weitwinkelobjektiv ausgestattet, um die Aufnahme der gesamten Fassadenhöhe in Nadirsicht zu ermöglichen. Entsprechend groß sind die radialsymmetrischen Verzeichnungen, die in den noch unkorrigierten Originalbildern deutlich zu sehen sind. Die Aufnahmen wurden Anfang Dezember zwischen 8:00 Uhr und 9:00 Uhr aufgenommen.

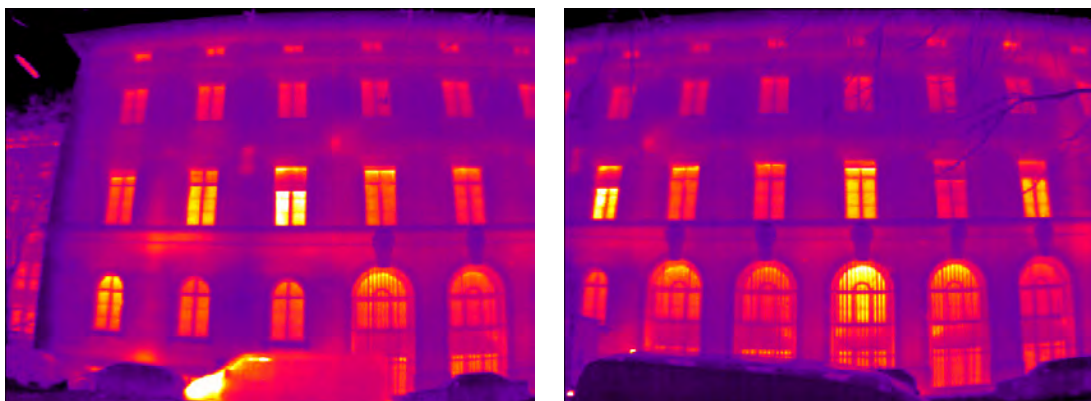


Abbildung 5.3: Thermalbilder Infratec Variocam hr. Zwei überlappende Bilder in Nadirsicht auf die Fassade. Die einkanaligen Intensitätsbilder sind für eine bessere visuelle Beurteilung eingefärbt. Gut zu sehen sind die Fensterflächen, die die gegenüber liegende Fassade spiegeln und die Heizungsleitungen, die sich warm (rot) abzeichnen. Ebenfalls zu sehen sind zwischen der unteren und mittleren Fensterreihe der kühle (dunkel) vorstehende Sims und die aufgesetzten Steinplatten mit den Emblemen der Fakultäten.

5.2.2 FLIR SC3000 320×240

Die FLIR SC3000 ist im Gegensatz zur Infratec VarioCam mit einem Quantendetektor ausgestattet, misst also direkt die einfallenden Photonen. Die Kamera mit ihrem 320×240 Pixel Detektorchip mit einer Empfindlichkeit im Bereich von 8-9 μ m muss daher gekühlt werden und erreicht eine thermale Auflösung von 0,02 K. Bild 5.4 zeigt zwei Bilder aus zwei Sequenzen, die mit vorwärts (5.4 a) und rückwärts (5.4 b) gerichteter Kamera aufgenommen wurden. Die einkanaligen Intensitätsbilder sind zur besseren Sichtbarkeit eingefärbt. Im Gegensatz zur Infratec Kamera erkennt man hier die unterschiedliche Abstrahlung des Erdgeschosses und der Obergeschosse. Dort sind unterschiedliche Steine verbaut worden. Insbesondere im unteren Mauerwerk sind sogar die Fugen zwischen den Steinen zu sehen. Durch die Schrägsicht sieht man von den Fensterflächen selbst nur wenig. Sie bilden sich lediglich als schmale Streifen in den unteren beiden Fensterreihen ab. Dafür sieht man die Fensternischen und den Wärmestau über den Fenstern. Der Sims über den Erdgeschoss hebt sich ebenso wie die Fakultätseembleme wieder vom Hintergrund ab. Die FLIR SC3000 ist mit einem Teleobjektiv bestückt. Es treten kaum radialsymmetrische Verzeichnungen auf. Die Heizungsleitungen sind auch hier erkennbar. Die Aufnahmen erfolgten Mitte April zwischen 6:00 Uhr und 7:00 Uhr morgens.

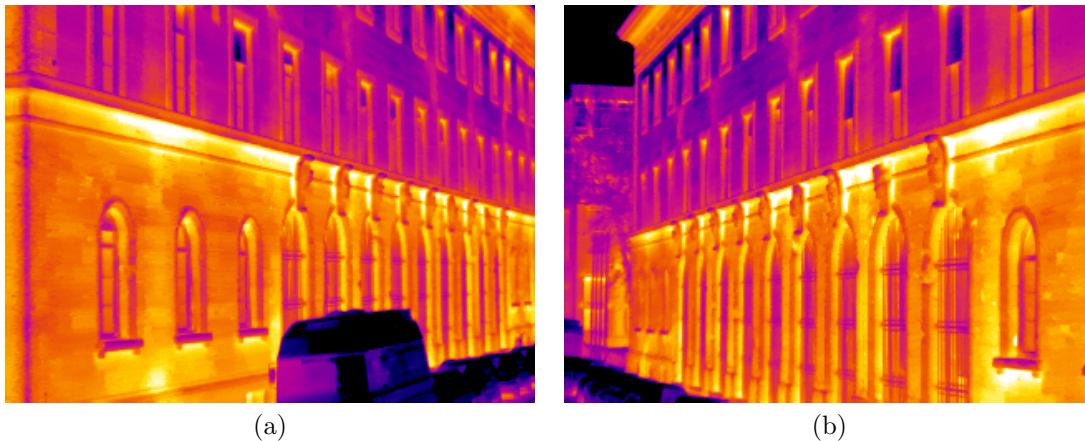


Abbildung 5.4: Thermalbilder FLIR SC3000. Einkanalige Intensitätswerte in Farbskala umgerechnet. (a) Schräg vorwärts blickendes Einzelbild. (b) Schräg rückwärts blickendes Einzelbild. Die beiden unterschiedlichen Materialien im Erdgeschoss und den Obergeschossen sind gut sehen, ebenso die Fenster und Heizungsrohre. Der Sims erscheint dunkler mit einem Wärmestau auf der Unterseite. Die Steinplatten mit den Emblemen der Fakultäten heben sich ebenfalls ab.

5.2.3 Merlin MID 320×256

In Abbildung 5.5 ist ein gleichzeitig aufgenommenes Bildpaar der MerlinMID Kamera aus dem mittleren thermischen Infrarot (5.5 a) und der FLIR SC3000 aus dem langwelligen thermischen Infrarot (5.5 b) zu sehen. Beide einkanaligen Intensitätsbilder sind mit derselben Tabelle in Farbwerte umgerechnet. Die MerlinMID ist mit ihrer Bildgröße von 320×256 Pixeln ebenfalls eine Kamera mit Quantendetektor, der im Bereich von 3-5 μ m arbeitet und gekühlt wird. Ihre thermale Auflösung liegt wie bei der FLIR SC3000 bei 0,02 K. Die MerlinMID ist ebenfalls mit einem Teleobjektiv ausgestattet und radialsymmetrische Verzeichnungen treten kaum auf. Beide Kameras zeigen ein ähnliches Bild des Haupteingangs der TUM. Deutlich zu sehen sind die Fensterstürze der zugemauerten Fenster im Erdgeschoss, sowie der Sturz über dem Eingang. Die Heizungsleitungen sind deutlich zu sehen und laufen direkt an den beiden kleinen Fenstern entlang. Ein Unterschied zeigt sich bei der Betrachtung der Fenster. Durch den schrägen Blickwinkel sieht man in beiden Fällen die Fensternische und auf der Scheibe eine Reflexion der Nische und

eine Spiegelung der Objekte auf der gegenüber liegenden Straßenseite. Während im langwelligen thermischen Infrarot hier der kalte Himmel zu sehen ist, zeigt das mittlere thermische Infrarot bereits eine erhöhte Abstrahlung. Dies kommt durch die Aufnahme in der Morgendämmerung. Die Sonne strahlt im mittleren Infrarot und diese Strahlung im mittleren Infrarot ist auf der nach Osten ausgerichteten Fassade bereits messbar.

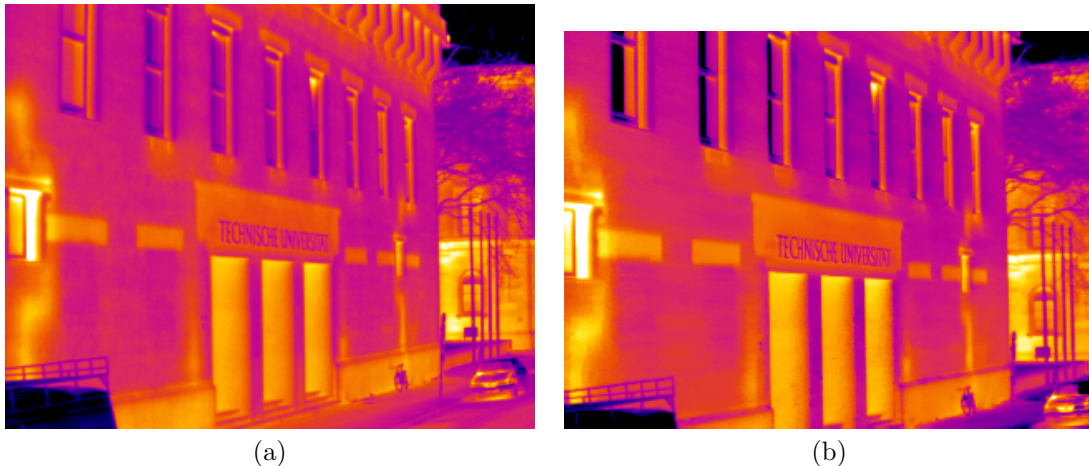


Abbildung 5.5: Vergleich der Thermalbilder zweier Spektralbereiche. (a) MerlinMID Infrarotkamera im mittleren thermischen Infrarot 3-5µm (b) FLIR SC3000 Infrarotkamera im langwelligen thermischen Infrarot 8-9µm. Haupteingang der TUM. Als helle warme Balken scheinen alte Fensterstürze und der Träger über den Eingang durch, ebenso die Heizungsleitungen. Der Unterschied in beiden Bildern ist recht gering, da die Aufnahme am Morgen vor Sonnenaufgang erfolgt ist.

5.2.4 Verwendete Bildsequenzen

Mit dem Trägersystem aus SC3000, MerlinMID und Videokamera (Abb. 5.2) wurden jeweils vorwärts und rückwärts blickend die vier Außenseiten des TUM Gebäudekomplexes aufgezeichnet. Das Fahrzeug ist dabei immer in derselben Richtung einen vorgegebenen Pfad gefahren.

Abbildung 5.6 zeigt je einen Ausschnitt der vier verschiedenen Außenseiten aus einer mit der FLIR SC3000 aufgenommenen Sequenz in vorwärts zeigender Blickrichtung. In Abbildung 5.6 a ist die Ostseite des Gebäudekomplexes zu sehen, die aus einem zurückgesetzten durchgehenden Gebäude und drei kleineren, direkt an der Straße gebauten Gebäuden besteht. Hier werden die besten Testergebnisse erwartet, da die kleinen Fassaden unabhängig voneinander extrahiert werden und durch die Freifläche auf der gegenüberliegenden Seite ein guter GPS Empfang gewährleistet ist, der zu einer guten Anfangsschätzung der äußeren Orientierung führt. Abbildung 5.6 b zeigt die Westseite des Komplexes. Hier befinden sich drei Fassadenabschnitte derselben Tiefe, wobei der erste Abschnitt auf Grund seiner komplexen Fassadentextur als problematisch anzusehen ist. Die Fenster und Vorsprünge bilden ein regelmäßiges Gitter, das leicht zu Fehlzuordnungen führen kann, wenn die äußere Orientierung und damit die relative Bewegung der Interestpunkte nicht ausreichend genau geschätzt werden können. Hinzu kommt, dass durch die enge Häuserschlucht die GPS Position deutlich schlechter bestimmt ist als auf der Ostseite. Abbildung 5.6 c zeigt eine Aufnahme der Südseite. Diese besitzt wie die Westseite eine durchgehende Fassadenebene, die in mehrere kleine Fassaden aufgeteilt ist. Durch die feinere Gliederung der Fassade ist die Gefahr von Fehlzuordnungen als geringer anzusehen als an der Westfassade. Die GPS Position kann ausreichend gut für die Anfangsschätzung der Äußeren Orientierung bestimmt werden. Die Nordfassade (Abbildung 5.6 d) besteht aus mehreren Fassadenteilen in derselben Fassadenebene. Zwei Brücken über die Straße hinüber zum Nordgelände führen zu Störungen in der Verfolgung der

Merkmalspunkte und der Flächenzuordnung. Für die Infratec Variocam wurden nur Aufnahmen der Ostfassade durchgeführt. Ein Überblick über die Sequenzen findet sich in Anhang A.

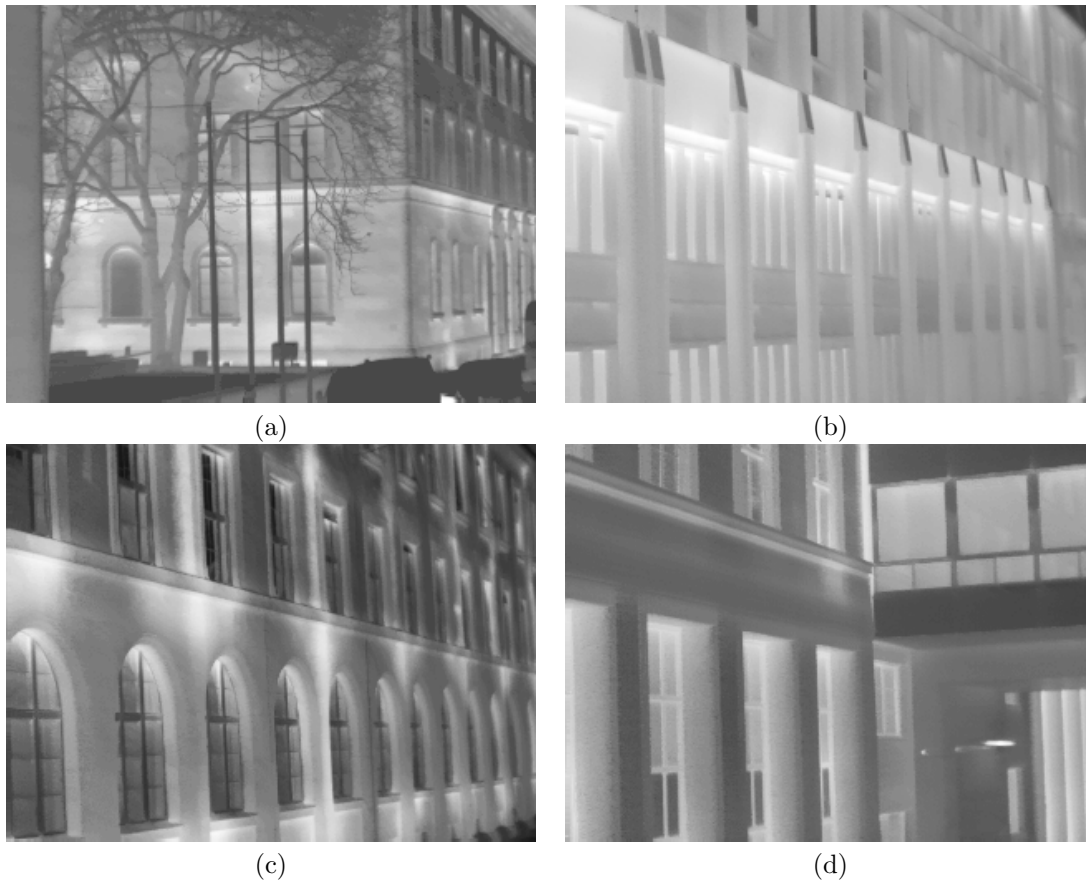


Abbildung 5.6: Überblick über die aufgenommenen Fassaden. Beispielbilder der SC3000 vorwärts blickend für alle vier untersuchten Fassaden a) Ostfassade: Mehrere Gebäudeteile unterschiedlicher Tiefe strukturieren die Szene. b) Westfassade: Drei Gebäudeabschnitte gleicher Tiefe unterschiedlicher Bauart, moderner Teil mit streng wiederholender Fassadenstruktur. c) Südfassade: Verschiedene kleinere Gebäudeteile derselben Tiefe. Fassadenstruktur ähnlich Ostseite. d) Nordfassade: Mehrere Gebäudeteile verschiedener Bauart derselben Tiefe. Besonderheit sind hier die beiden Brücken, die das untersuchte Hauptgebäude mit dem Nordgelände verbinden. Ein Überblick über die Sequenzen findet sich in Anhang A.

5.2.5 3D Gebäudemodell

Als 3D Gebäudemodell steht ein Polygonmodell zur Verfügung, das im CityGML Format im Level-of-Detail (LOD) 2 vorliegt. Das Modell ist mittels der Software Erdas Imagine aus einem Stereoluftbild erstellt worden. Es sind alle Außenfassaden modelliert, die zur Straße hin zeigen. Die Genauigkeit der Fassade in der X-Y-Ebene liegt bei knapp 1/2 Meter. Grund dafür ist die Erzeugung aus dem Stereoluftbild. Dort wird die Dachaußenkante zum Boden verlängert, der Dachüberstand also nicht modelliert. Die Genauigkeit in der Höhe liegt bei knapp einem Meter. Damit liegt das Gebäudemodell in dem Genauigkeitsbereich, den übliche vorhandene 3D Modelle besitzen. Die eigentliche Fassadenkante am Übergang zum Dach ist häufig kaum erkennbar. Die Dachkante zwischen dem Dachüberstand und dem Himmel zeichnet sich jedoch deutlich ab. Sie wird in der Generierung des Gebäudemodells als Fassadenoberkante verwendet. Im Bild werden dort etliche Punkte auftreten, nicht aber am Übergang von Fassade zu Dachüberstand. Die Bodenkante der Fassade ist meist durch Fahrzeuge verdeckt, also ohnehin nicht sichtbar. Damit ist auch der Fußpunkt der Texturen verdeckt. Geometrische Details der Fassaden sind nicht model-

liert. Jede Fassade besteht damit nur aus einem Polygonzug, der die Fassadenfläche beschreibt. Das Modell liegt georeferenziert in UTM Koordinaten vor. Darüber hinaus sind für ausgewählte Fassaden manuell RGB-Texturen erstellt worden. Diese dienen später als Ground Truth für die Bewertung der Genauigkeit der automatisch extrahierten Thermaltexturen.

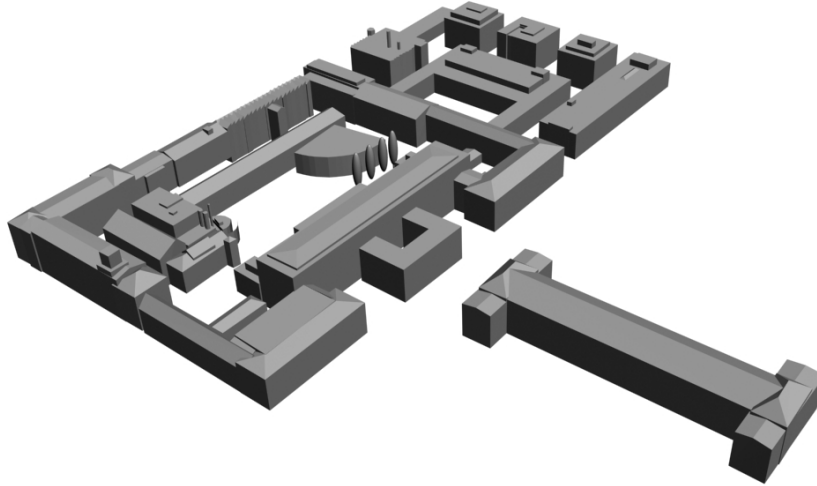


Abbildung 5.7: 3D Gebäudemodell des Stammgeländes der TUM ohne Texturen. Während der Gebäudekomplex an der Nord-, West- und Südseite von dichter Bebauung umgeben ist, liegt auf der Ostseite eine Grünanlage, in deren Mitte die Alte Pinakothek (das langgezogene Gebäude rechts) steht.

6 Ergebnisse

6.1 Ergebnisse der Kamerakalibrierung

Für die Bestimmung der inneren Orientierung werden drei Szenarien vorgestellt, zum einen ein senkrechtes Innentestfeld mit aktiven Strahlern als Messmarken für einen kurzen Objekt-Abstand (Abb. 6.1) und ein waagerechtes Außentestfeld mit reflektierenden Messmarken für eine Fokussierung auf unendlich (Abb. 6.2). Zum anderen wird für die fest am Messfahrzeug verbauten Kameras die innere Orientierung aus Passpunkten auf der Fassade bestimmt.

6.1.1 Kalibrierung der VarioCam hr mit Testfeldern

Die Testfeldkalibrierungen wurden mit der Variocam HR von Zeiss/Infratec (siehe Kap. 5.2.1) mit einer optischen Auflösung des Detektors von 384×288 Pixeln und einem 45° Germanium-Objektiv durchgeführt. Die radiometrische Auflösung des Detektors liegt bei 0,08 Kelvin. Für beide Testfelder heben sich die Messmarken deutlich vom Hintergrund ab. Die Messpunkte des Innentestfelds haben einen Abstand von etwa 12 cm. Die exakten Abstände sind mit einer kalibrierten Kamera vermessen worden und können daher als fehlerfreie Beobachtungen angenommen werden. Für das Außentestfeld wurde ein Abstand von etwa 100 cm für die Messmarken festgesetzt. Auch hier wurden die Koordinaten mit einer kalibrierten Messkamera ermittelt.

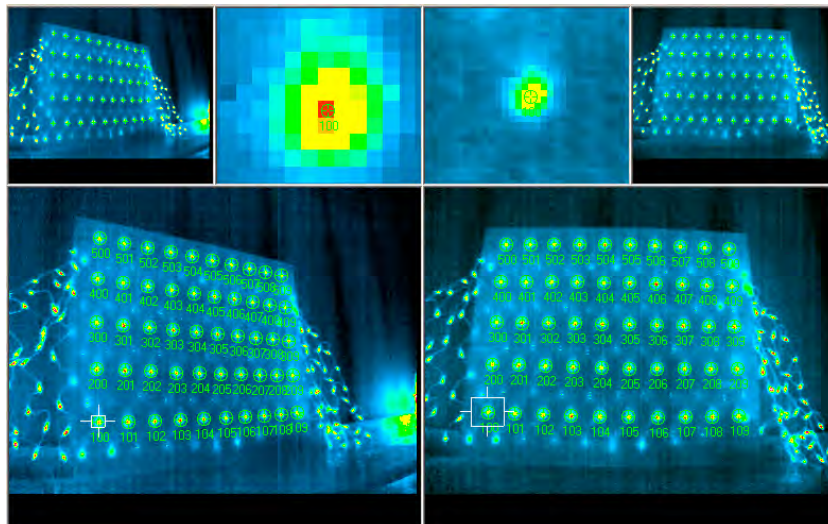


Abbildung 6.1: Aufnahme des Innentestfelds mit aktiven Strahlern. Die Nummerierung zeigt die erfolgreiche Zuordnung der zuvor eingemessenen Messpunkte (Simmler [2009])

Für beide Testfelder erfolgt zunächst ein Durchlauf ohne Schätzung der Parameter, um die Standardabweichung der Ausgleichung zu berechnen und so einen Vergleichswert der Ausgangssituation zu erzeugen. Die Standardabweichung σ_0 der Ausgleichung verbessert sich durch die

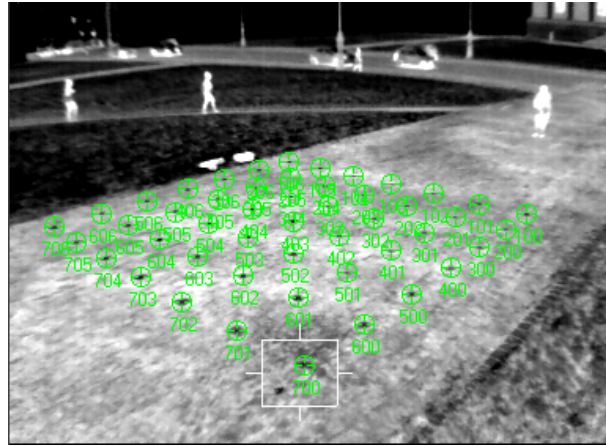


Abbildung 6.2: Aufnahme des Außentestfelds aus reflektierenden CDs. Die Nummerierung zeigt die erfolgreiche Zuordnung der zuvor eingemessenen Messpunkte (Simmler [2009])

Einbeziehung der Kammerkonstanten c_k , des Bildhauptpunktes (x_0, y_0) und der radialsymmetrischen Verzeichnung (A_1, A_2) beim Innentestfeld von $17,9 \mu\text{m}$ auf $10,5 \mu\text{m}$ und beim Außentestfeld von $33,3 \mu\text{m}$ auf $14,4 \mu\text{m}$. Bei einer Pixelgröße der Infrarotkamera von $35 \mu\text{m}$ ist damit Subpixelgenauigkeit erreicht. Tabelle 6.1 zeigt die Genauigkeit der Bestimmung der Parameter der Inneren Orientierung für beide Testfelder. Für die aktiven Messmarken (Abb. 6.1) sind dabei alle Parameter bis auf A_2 signifikant bestimmbar. Für die reflektierenden Messmarken ergeben die Einschränkungen bei den möglichen Blickwinkeln sowie die ungenauere Bestimmung der Messmarken größere Standardabweichungen bei den verschiedenen Parametern. Allerdings zeigt sich, dass die radial-symmetrische Verzeichnung nur in sehr geringem Maße von der Fokussierung abhängig ist und sowohl für den Nahbereich wie auch für die Fokussierung auf unendlich in derselben Größenordnung liegt. Die mittlere Passpunktgenauigkeit erreicht für das innentestfeld (Abb. 6.1) trotz der geringen optischen Auflösung unter Einbeziehung aller Parameter in der Ausgleichung Werte von $0,46 \text{ mm}$ und $0,675 \text{ mm}$ in x- und y-Richtung. In der z-Achse, die als Tiefenachse im Bild definiert ist, liegt die Genauigkeit auf Grund der flachen Messfelder bei $5,544 \text{ mm}$. Dieser Wert ließe sich durch ein Messfeld mit verschiedenen Tiefen der Messmarken weiter verbessern. Für gemessene Objektpunkte ergeben sich mittlere Genauigkeiten von $0,635 \text{ mm}$, $0,607 \text{ mm}$ und $0,1046 \text{ mm}$ in x, y und z-Richtung. Für die Fokussierung auf unendlich (Abb. 6.2) liegen die Genauigkeiten der Objektpunkte zwischen 4 und 5 mm in x, y und z-Richtung.

Die Ergebnisse zeigen, dass eine Bestimmung der Parameter der Inneren Orientierung auch für Infrarotkameras möglich ist, trotz ihrer geringeren optischen Auflösung und des schwächeren Kontrasts im Vergleich zu optischen Messmarken. Es zeigt sich, dass die Genauigkeiten für die innere Orientierung aus Kamerakonstante c und Bildhauptpunkt $H = (x_0, y_0)$ für beide Versuche in einer ähnlichen Größenordnung liegt und die Parameter signifikant bestimmbar sind. Für die Verzeichnungsparameter A_1 und A_2 sind die Größenordnungen sowohl der Verzeichnungsparameter selbst als auch ihrer Standardabweichungen ebenfalls weitestgehend in derselben Größenordnung. Leider ist die Signifikanz der Verzeichnungsparameter im Allgemeinen deutlich schwächer als die der inneren Orientierung.

6.1.2 Kalibrierung der FLIR SC3000 und MerlinMID mit Fassadenpunkten

Für die FLIR SC3000 und die MerlinMID wird die Bestimmung der inneren Orientierung auf Grundlage von Passpunkten auf der Fassade realisiert. Abbildung 6.3 zeigt den Aufbau des Ka-

Anzahl verbesserter Parameter		Innentestfeld		Außentestfeld	
		3	5	3	5
Standardabweichung σ_0 [μm]		15,6831	10,4508	32,8034	14,4130
Mittlere Passpunktgenauigkeit [mm]	x	1,521	0,46	5,0	3,1
	y	0,857	0,675	6,9	1,8
	z	6,762	5,544	1,1	0,4
Kammerkonstante [mm]	c	12,1631	12,5015	12,4733	12,4818
	σ_c	0,0479	0,0348	0,0244	0,0112
Bildhauptpunktlage [mm]	x_0	-0,2952	0,1254	-0,0944	0,0002
	σ_{x_0}	0,0254	0,0231	0,0210	0,0094
	y_0	0,1361	0,2397	-0,0170	0,0048
	σ_{y_0}	0,0260	0,0180	0,0296	0,0130
Radialsymmetrische Verzeichnung [mm]	A_1	-	$-2,1327 \times 10^{-3}$	-	$-2,0259 \times 10^{-3}$
	σ_{A_1}	-	$1,1681 \times 10^{-4}$	-	$7,7746 \times 10^{-5}$
	A_2	-	$-6,1212 \times 10^{-6}$	-	$-5,6166 \times 10^{-6}$
	σ_{A_2}	-	$2,0861 \times 10^{-6}$	-	$1,6452 \times 10^{-6}$

Tabelle 6.1: Kalibrierergebnisse der Infratec VarioCam hr: Verbesserungen und Standardabweichungen mit Kammerkonstante, Bildhauptpunktlage ohne (3 Parameter) und mit (5 Parameter) radialsymmetrischer Verzeichnung

merasystems mit Blick auf das TUM Gebäude. Am Gebäude selbst sind mehrere markante Punkte als Passpunkte eingemessen worden.



Abbildung 6.3: Kameraaufbau FLIR SC3000 und MerlinMID für die Passpunktkalibrierung

Die Messung der Passpunkte ist mit einigen Einschränkungen verbunden. Da das Fahrzeug nur auf der Straße fahren kann, sind die möglichen Blickrichtungen und Sichtfelder eingeschränkt. Durch die Schrägsicht ist die geometrische Auflösung in den Bildern nicht konstant. Damit variiert auch die Genauigkeit, mit der die Bildkoordinaten der Passpunkte bestimmt werden können. Es stehen nur Bilder aus der Vorwärtssicht und der Rückwärtssicht zur Verfügung. Einzelbilder der Sequenz zeigen nur einen kleinen Ausschnitt einer Fassade und daher auch nur wenige Passpunkte. Zur Verbesserung der Schätzung wird daher zusätzlich eine Aufnahme von einem festen Standort aus verwendet, die die vollständige Fassade zeigt. Sie ist über die Passpunkte mit allen Bildern aus der Vorwärts- und Rückwärtssicht verbunden. In Abbildung 6.4 a ist der Bildausschnitt der FLIR SC3000 auf einen Gebäudeteil zu sehen. Nimmt man den entsprechenden Teil des 3D Modells hinzu, so sieht man, dass die Szene über mehrere Ebenen und Tiefen verfügt. Durch eine ausreichende Anzahl an eingemessenen Fensterecken auf beiden Fassaden des Gebäudeteils entsteht ein Passpunktfeld mit verschiedenen Tiefen. Durch die geringe geometrische Auflösung der Thermalaufnahme in Verbindung mit der geringen Pixeldichte auf der Fassade von 5-10 Pixeln pro Meter ist die Festlegung der Bildkoordinaten für die Passpunkte auf der Fassade entsprechend

ungenau. In Abbildung 6.4 sind die ausgewählten Bilder für die Kalibrierung der SC3000 dargestellt. Neben der Aufnahme der gesamten Fassade (6.4 a) werden Bilder verwendet, die die quer zur Fahrtrichtung angeordnete Fassade zeigen (6.4 b). Die Passpunkte auf dieser Fassade sind nur in der vorwärts blickenden Sequenz sichtbar. Passpunkte auf der Fassade parallel zur Straße sind sowohl in der Vorwärtssicht (6.4 c) wie auch in der Rückwärtssicht (6.4 d) vorhanden.

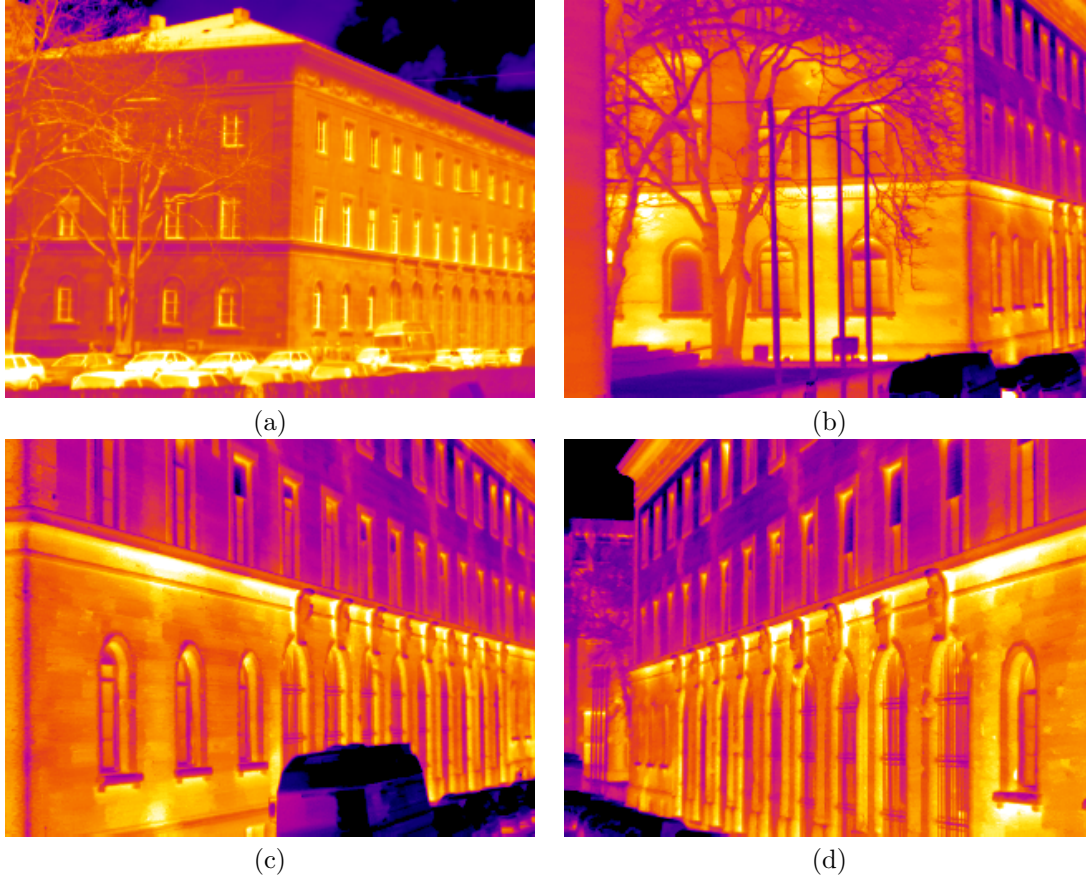


Abbildung 6.4: Fassadenbilder für Kalibrierung der FLIR SC3000. (a) Aufnahme vom festen Standpunkt aus. Man sieht die beiden Fassaden und die Fensterreihen, deren Koordinaten bekannt sind. (b) Einzelbild aus einer Sequenz, die die quer zur Straße liegende Fassade zeigt. (c+d) Einzelbilder der vorwärts und rückwärts blickenden Sequenz mit der parallel zur Straße liegenden Fassade

Die Berechnungsschritte gleichen denen für die VarioCam hr. Die feste Position der Kameras für die Aufnahme der Totalen ist im Koordinatensystem der Passpunkte gemessen worden. Dadurch wird trotz der geringen Anzahl an Bildern sichergestellt, dass bei der Ausgleichung die Kamerakonstante in Verbindung mit dem Abstand stabil gehalten wird. Für die anderen Aufnahmen werden die gespeicherten GPS Positionen als Näherung verwendet. Die Standardabweichung σ_0 der Ausgleichung liegt bei Einbeziehung der Kammerkonstanten c_k , des Bildhauptpunktes (x_0, y_0) und der radialsymmetrischen Verzeichnung (A_1, A_2) bei $12,6 \mu\text{m}$ für die SC3000 und $12,9 \mu\text{m}$ für die MerlinMID. Tabelle 6.2 zeigt die Genauigkeit der Bestimmung der Parameter der Inneren Orientierung für beide Kameras. Dabei ist zu erkennen, dass Kamerakonstante c signifikant bestimmbar ist. Für die Bildhauptpunktlage $H = (x_0, y_0)$ wird lediglich eine schwache Signifikanz mit einem Faktor von 1 bis 3 erreicht. Allerdings liegen die Standardabweichungen deutlich niedriger als für die VarioCam hr (Tabelle 6.1). Der Verzeichnungsparameter A_1 ist ebenfalls schwach signifikant bestimmt. Dort liegen die absoluten Werte der Parameter und die Standardabweichungen ebenfalls deutlich, um bis zu 3 Zehnerpotenzen unter den Werten der VarioCam hr. Der zweite Verzeichnungsparameter A_2 konnte hingegen nicht signifikant bestimmt werden. Auch seine Wer-

te lagen um mehrere Potenzen unter denen der VarioCam hr. Daraus kann geschlossen werden, dass die radialsymmetrische Verzeichnung für die SC3000 und die MerlinMID keine relevanten Einfluss auf die Abbildung hat. Das stützt die bereits durch visuelle Kontrolle getroffene Annahme, dass sowohl die SC3000 wie die MerlinMID durch die Wahl des Teleobjektives im Rahmen der Messgenauigkeit praktisch verzeichnungsfrei abbilden.

Für alle drei Kameras gilt, dass weitere Verzeichnungsparameter nicht bestimmt werden, da ihre Werte nicht signifikant bestimmbar sind.

Anzahl verbesserter Parameter		FLIR SC3000		MerlinMID	
		3	5	3	5
Standardabweichung σ_0 [μm]		12,6	11,8	12,9	12,4
Mittlere Passpunktgenauigkeit [mm]	X	24,6	24,2	24,5	24,1
	Y	24,8	24,2	24,5	24,0
	Z	24,5	24,1	24,3	24,0
Kammerkonstante [mm]	c	50,7662	50,721	52,5853	52,6144
	σ_c	0,1265	0,1073	0,1196	0,1082
Bildhauptpunktlage [mm]	x_0	-0,0146	-0,00833	0,0254	0,00713
	σ_{x_0}	0,00447	0,00354	0,0105	0,00704
	y_0	0,00926	0,00654	0,0102	0,00832
	σ_{y_0}	0,00283	0,00318	0,00964	0,00586
Radialsymmetrische Verzeichnung [mm]	A_1	-	$-5,5743 \times 10^{-9}$	-	$-1,7427 \times 10^{-9}$
	σ_{A_1}	-	$3,4361 \times 10^{-9}$	-	$7,7746 \times 10^{-8}$

Tabelle 6.2: Kalibrierergebnisse der FLIR SC3000 und der MerlinMID: Verbesserungen und Standardabweichungen mit Kammerkonstante, Bildhauptpunktlage ohne (3 Parameter) und mit (5 Parameter) radialsymmetrischer Verzeichnung

6.2 Qualität der homologen Punkte

Die Verteilung, Anzahl und Genauigkeit der homologen Punkte hat einen direkten Einfluss sowohl auf die relative Orientierung im zweistufigen Verfahren als auch auf die absolute Orientierung mit Nebenbedingungen. Daher soll vor der Betrachtung der 3D Rekonstruktion die Qualität der homologen Punkte sowohl bei Förstner-Punkten wie auch SIFT-Features untersucht werden. Für die Bewertung der Qualität der homologen Punkte werden vier Kriterien herangezogen:

- Punktdichte: Anzahl gefundener Interest-Punkte pro Bild
- Korrektheit: Anzahl der korrekt zugeordneten homologen Punkten eines Bildpaares
- Stabilität: Zeitintervall, in dem ein Interest-Punkt verfolgt werden kann
- Ähnlichkeit: Ähnlichkeit der homologen Punkte, angegeben durch die Summe der quadratischen Differenzen (SSD)

Diese Kriterien hängen von mehreren Faktoren ab. Einen großen Einfluss hat die beobachtete Szene. Je mehr Struktur die Szene aufweist, umso mehr Interest-Punkte werden gefunden. Je höher das Auflösungsvermögen der Kamera ist, umso feinere Details können gesehen werden, die wiederum die Zahl der Interestpunkte erhöhen und zu einer genaueren Lokalisation führen. Die Bewegung zwischen zwei Bildern hat einen Einfluss auf die Anzahl korrekt zugeordneter homologer Punkte und auf die Anzahl der Bilder, über die eine Verfolgung eines Interest-Punktes möglich ist. Die Ergebnisse der vier Qualitätskriterien hängen also in besonderer Weise vom beobachteten Szenario ab. Daher wird als Vergleich dasselbe Szenario wie es die Infrarotkamera sieht, mit einer

normalen dreikanaligen RGB-Videokamera aufgezeichnet. Zur Untersuchung des Einflusses des Auflösungsvermögens werden die Bilder der RGB-Kamera sowohl in ihrer originalen Auflösung wie auch in einer reduzierten Auflösung betrachtet, die dem Auflösungsvermögen der Infrarotkamera entspricht. Auf beiden Auflösungsstufen werden ebenfalls Interestpunkte detektiert und durch die Bildsequenz verfolgt. Das Ergebnis der Detektion und Verfolgung wird für die Infrarotkamera und die beiden Auflösungen der Videokamera verglichen. Für die Bewertung werden exemplarisch mehrere Sequenzen ausgewählt, die mit der FLIR SC3000 aufgenommen wurden. Abbildung 6.5 zeigt für zwei Aufnahmezeitpunkte einer Sequenz jeweils das Infrarotbild (6.5a) und das RGB Bild (6.5b). Für die Suche nach den Interestpunkten werden die originalen einkanaligen Intensitätsbilder der Infrarotkamera verwendet. Für die Videokamera werden die originalen dreikanaligen Bilder entsprechend Formel 6.1 in einkanalige Grauwertbilder umgewandelt. Die leichten Unterschiede in der Szene sind der Anordnung der Kameras auf dem Träger geschuldet.

$$B_{\text{grau}}(x, y) = 0,299 \cdot B_{\text{rot}}(x, y) + 0,587 \cdot B_{\text{gruen}}(x, y) + 0,114 \cdot B_{\text{blau}}(x, y) \quad (6.1)$$

In Tabelle 6.3 ist ein Auszug aus den Vergleichen von Förstner-Punkten und SIFT Features dargestellt. Zwar unterscheiden sich für unterschiedliche Szenen die jeweiligen Zahlen leicht, der Trend ist jedoch in allen Sequenzen der gleiche. Deshalb sind hier die Durchschnittswerte über alle getesteten Sequenzen aufgetragen. Wie zu erwarten werden in der Videosequenz mit 768×576 Pixeln im Schnitt deutlich mehr Interestpunkte gefunden als in den kleineren Bildern. Die Anzahl korrekt zugeordneter homologer Punkte nimmt mit der Aufnahmefrequenz zu. Der Bildinhalt ändert sich in diesem Fall nicht so stark und der Suchraum wird stärker eingeschränkt. Im Vergleich des niedrig aufgelösten Videobildes mit den Thermalbildern zeigt sich, dass in den Thermalbildern deutlich weniger Interestpunkte bei denselben Parametern gefunden werden. Sowohl die Kantenstärke wie der Kontrast sind dort deutlich schwächer ausgeprägt. Auch sind bei weitem nicht so viele kleine Details zu sehen. Die Genauigkeit und Zuverlässigkeit der gefundenen homologen Punkte ist zwar etwas schlechter als im Videobild, liegt aber in derselben Größenordnung. Auffällig ist, dass bei den Videosequenzen deutlich mehr Förstner-Punkte gefunden werden als SIFT Features, was auf den hohen Detailgrad der Textur der Fassade zurückzuführen ist. Die SIFT Features können aber besser zugeordnet werden. Im Infrarotbild hingegen finden beide Detektoren annähernd gleich viele markante Punkte. Bei der Zuordnung zeigt sich, dass der Förstner-Operator bei hohen Bildfrequenzen und entsprechend kleinen Änderungen in der Szene gerade im Infrarotbild eine deutlich höhere Anzahl korrekter homologer Punkte aufweist.

Abbildung 6.6 zeigt das Ergebnis einer Koregistrierung des in Abbildung 6.5 vorgestellten Bildpaares auf Basis der mit SIFT gefundenen homologen Punkte. Verwendet worden sind hier nur solche Punkte, die der Fassade parallel zur Straße zugeordnet worden sind, um deren Textur zu extrahieren. Das erste Bild in rot ist dabei der Master. Das zweite Bild in blau wird mit den homologen Punkten koregistriert. Man kann dort bereits den in Abschnitt 4.1.1 angesprochenen auf Grund der Schrägsicht schiefen Bildbereich sehen, der auf der linken Seite eine hohe geometrische Auflösung bei kleinem Blickfeld und rechts eine niedrige geometrische Auflösung bei großem Blickfeld zeigt. Für SIFT Features ergibt sich die SSD (Ähnlichkeitswert) aus den einzelnen Komponenten der Featurevektoren der homologen Punkte. Man sieht für die Videobilder kaum eine Verbesserung durch die dichtere Bildfolge. Bei den Thermalbildern ist die SSD deutlich größer und die dichtere Bildfolge, die zu kleineren Unterschieden im Bildinhalt führt, bringt eine deutliche Verbesserung (Abb. 6.3).

Die Gründe dafür sind in der Berechnung der SIFT Features zu finden. Die Anzahl der Stufen der Bildpyramide hängt von der Auflösung des Eingangsbildes ab. Liegen ausreichend Auflösungsstufen vor, so können in verschiedenen Stufen mehrere Punkte an derselben Stelle detektiert wer-

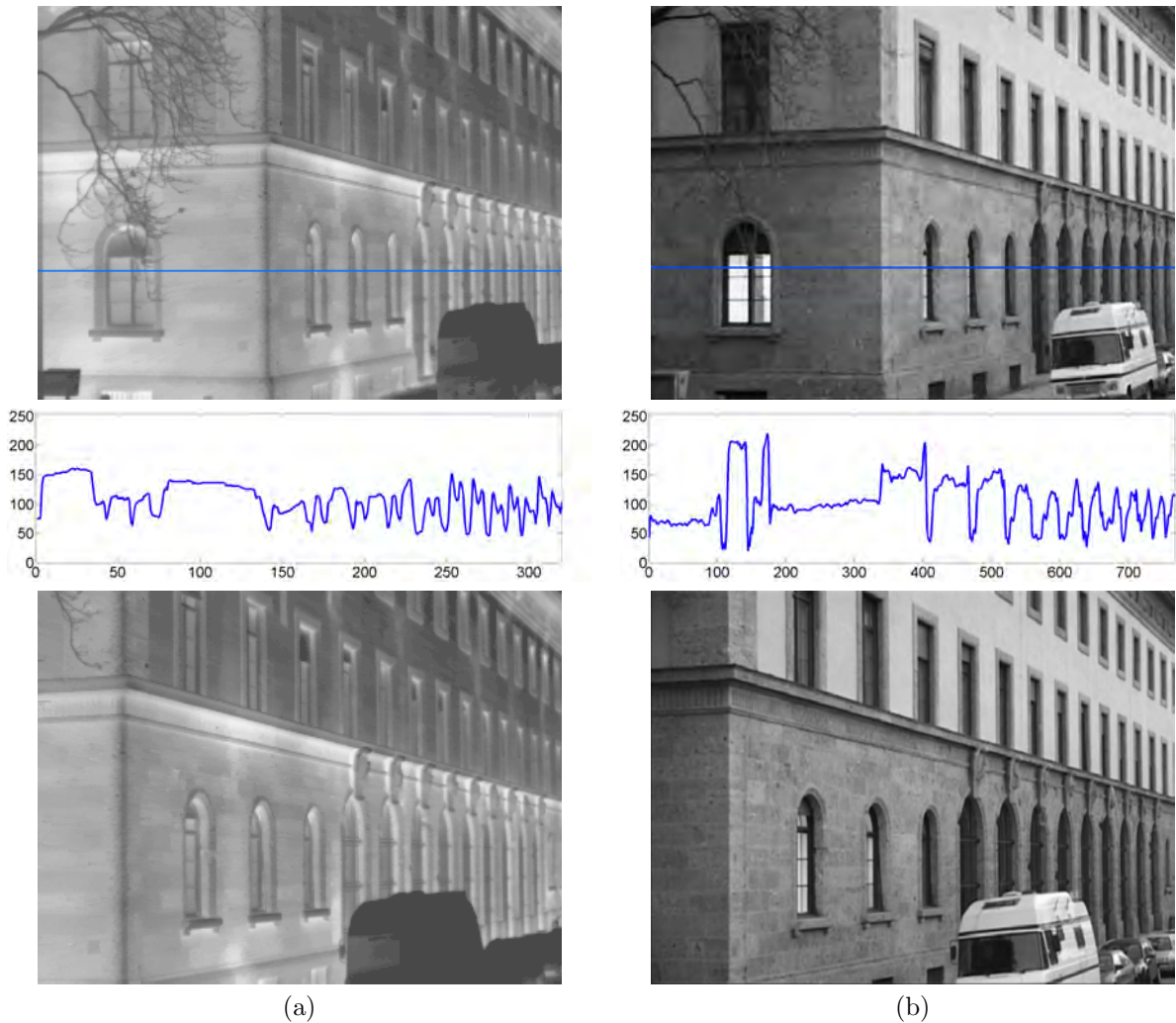


Abbildung 6.5: Beispielsequenz Feature Tracking. Beispielbilder aus einer Sequenz. Unterschiede in der sichtbaren Szenen ergeben sich aus der Anordnung auf dem Kameraträger. Die Profillinien entsprechen jeweils der blau markierten Zeile der oberen Bilder. a) Aufnahmen der FLIR SC3000 mit in 8 bit Grauwerte umgerechneten Intensitätswerten. Bildgröße 320×240 Pixel. b) Aufnahmen der RGB-Videokamera umgewandelt in 8 bit Grauwerte. Bildgröße 768×576 Pixel.

den (Abschnitt 2.3.2). Die niedrige Auflösung der Thermalaufnahme reduziert die Anzahl der Stufen der Bildpyramide und damit die Anzahl gefundener Interestpunkte. Außerdem sind die Grauwertübergänge im Thermalbild deutlich schwächer und verwaschener als im Videobild (siehe Profillinien in Abb. 6.5). Um ein ausreichend markantes lokales Maximum in der 26er Umgebung der Bildpyramide zu finden, wird aus der Bildpyramide ein niederes Auflösungslevel gewählt. Damit wird eine größere lokalen Umgebung für die Berechnung des Featurevektors verwendet. Diese größere Umgebung ändert sich durch die Fortbewegung der Kamera auf Grund der nicht konstanten geometrischen Auflösung der Schrägsicht stärker als eine kleine Umgebung. Somit ist der Unterschied der Featurevektoren bei großen Bildabständen im Thermalbild höher als im Videobild. Abbildung 6.6 zeigt die Koregistrierung zweier Einzelbilder auf eine bestimmte Fassadenfläche. Als Farbdarstellung ist ein Anaglyphen Farbschema gewählt. Dabei ist das erste Bild in rot eingefärbt, das zweite Bild in cyan. Bereiche, die in beiden Bildern zu sehen sind, zeigen die korrekte Farbe (RGB im RGB Bild, Grauwerte im Infrarotbild). Man erkennt in den überlagerten Bereichen, dass die Bilder auf die Fassadenebene ohne Farbränder und damit ohne sichtbare Fehler projiziert sind. Der rötliche Bereich ist nur im ersten Bild zu sehen gewesen und daher in

	SIFT			Förstner		
	Video 768	Video 320	IR 320	Video 768	Video 320	IR 320
# Interest Punkte	1246,2	531,4	168,4	1474,9	513,0	201,2
Aufnahmefrequenz: 1 Bild pro Sekunde						
# Korrekte Homologe Punkte	287,2	124,7	31,9	118,4	49,2	31,1
Trackingdauer in Sekunden	11,6	11,6	10,4	11,6	11,6	11,2
Ähnlichkeit (SSD)	0,006	0,011	0,025	0,021	0,023	0,028
Aufnahmefrequenz: 5 Bilder pro Sekunde						
# Korrekte Homologe Punkte	683,3	244,6	129,2	431,2	187,4	168,4
Trackingdauer in Sekunden	11,8	11,8	10,5	11,8	11,8	11,3
Ähnlichkeit (SSD)	0,005	0,009	0,012	0,006	0,011	0,014

Tabelle 6.3: Vergleich SIFT Features und Förstner Punkte. Durchschnittswerte über mehrere Sequenzen. Während das höher aufgelöste Videobild deutlich bessere Werte zeigt, sind die Unterschiede zwischen den beiden gleich großen Video- und IR-Bildern vor allem bei 5 Bildern pro Sekunde zumindest für einige Kriterien gering. Die Zahl der korrekten homologen Punkte steigt mit der Bildwiederholrate an. Dadurch verbessert sich auch die Ähnlichkeit.

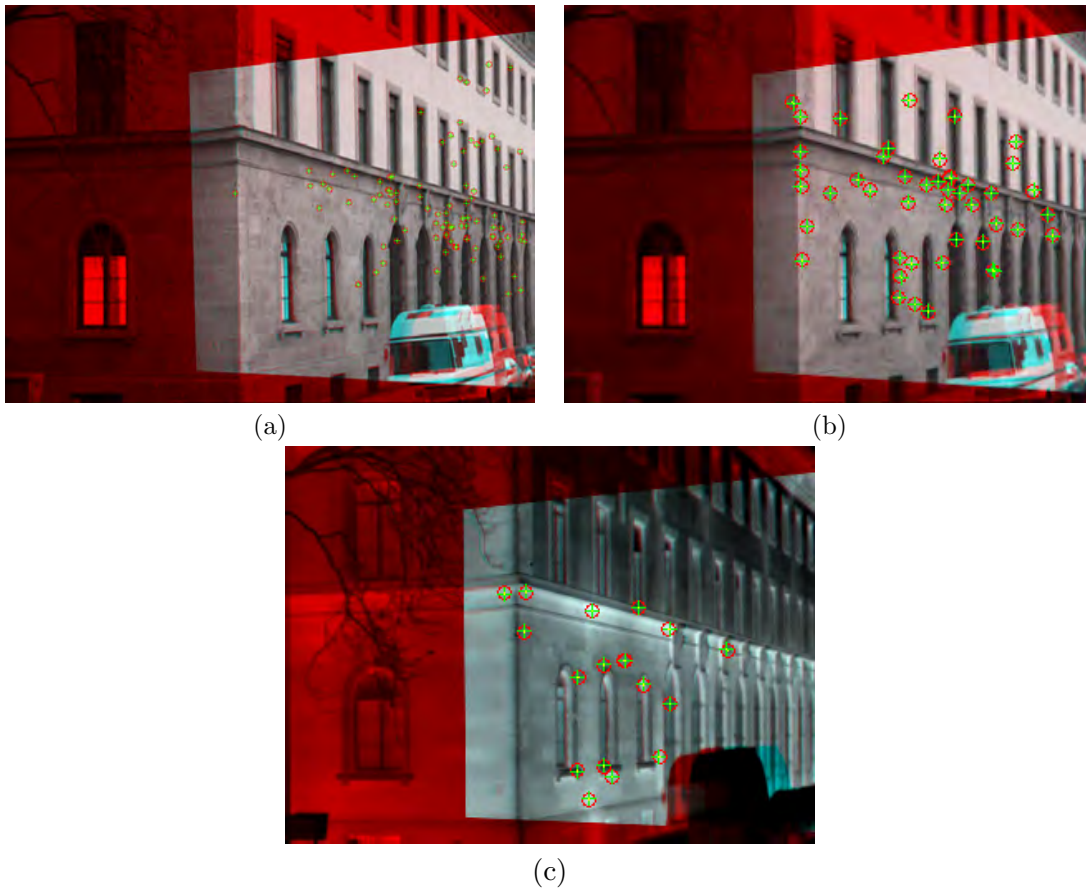


Abbildung 6.6: Überlagertes Bildpaar nach Feature Tracking. Korrespondierende SIFT-Features in Grün mit roten Kreis markiert für einen Aufnahmefrequenz von 1 Sekunde. Rot: erstes Bild. Blau: zweites Bild koregistriert zum ersten. (a) Videokamera mit Originalbildgröße 768×576 Pixel. (b) Videokamera mit reduzierter Bildgröße 320×240 Pixel. (c) SC3000 mit Originalbildgröße 320×240 Pixel.

Falschfarben dargestellt. Bei dem Fahrzeug vor der Fassade erkennt man rote und cyan Muster, die durch den perspektivischen Versatz des Fahrzeugs in beiden Bildern entstehen, da das Fahrzeug nicht auf der Projektionsebene der Fassade liegt. Die homologen Punkte in beiden Bildern sind als rote Kreise und grüne Kreuze dargestellt. Auch sie liegen auf der Fassadenebene direkt

übereinander. Die erzielbare Genauigkeit im Hinblick auf eine 3D Rekonstruktion wird in den folgenden beiden Abschnitten für das zweistufige Verfahren (Abschnitt 6.3) und das einstufige Verfahren (Abschnitt 6.4) näher erläutert.

6.3 Ergebnisse des zweistufigen Verfahrens

Das in Abschnitt 3.2 beschriebene zweistufige Verfahren führt zunächst eine relative Orientierung der Bildsequenz durch, in der für die gefundenen homologen Punkte 3D Punkte als Modellkoordinaten der relativen Orientierung bestimmt werden. Anschließend werden die Kameraorientierungen und die 3D Modellpunkte mit dem GPS Pfad und dem 3D Gebäudemodell koregistriert. Zur Bewertung der Genauigkeit der relativen Orientierung wird der Fehler der zurück in die Bilder der Sequenz projizierten 3D Modellpunkte bestimmt. Je genauer die relative Orientierung bestimmt ist, umso kleiner ist der mittlere quadratische Rückprojektionsfehler. Dieser Fehler ist es auch, der in der Bündelblockausgleichung minimiert wird. Für den zweiten Schritt, die Koregistrierung, werden der mittlere quadratische Abstand der 3D Objektpunkte zu den 3D Modellflächen bestimmt.

6.3.1 Genauigkeit der relativen Orientierung

Als Ergebnis der Bündelblockausgleichung der relativen Orientierung einer Bildsequenz werden einige statistische Maße ausgegeben, die in Tabelle 6.4 zusammengestellt sind. Es werden folgende Maße verglichen: Der Varianzfaktor a posteriori (σ_0^2), die Quadratsumme der Verbesserungen der Beobachtungen (vPv^T) und die größte (v_{max}) und durchschnittliche Verbesserung (v_{mean}), der größte (w_{max}), kleinste (w_{min}) und durchschnittliche Restfehler (w_{mean}) der ins Bild zurückprojizierten 3D Punkte zu ihren beobachteten homologen Punkten, sowie die Standardabweichungen der Parameter der relativen Orientierung. Diese Werte werden aus Ausschnitten den vorwärts blickenden Bildsequenzen der FLIR SC3000 für alle vier Szenarien (Abb. 5.6) für zwei verschiedene Aufnahmefrequenzen mit 5 Bildern pro Sekunden und einem Bild pro Sekunde ermittelt. Die Vergleichswerte für das einstufige Orientierungsverfahren befinden sich in Tabelle 6.6.

	Ostfassade		Westfassade		Südfassade		Nordfassade*	
	5 fps	1 fps	5 fps	1 fps	5 fps	1 fps	5 fps	1 fps
$\sigma_{xy}[\mu\text{m}]$	0.3857	1.793	12.74	16.36	0.9486	2.833	14.71	17.83
$\sigma_{XYZ}[\text{m}]$	0.06855	0.2867	0.3627	0.6459	0.1068	0.5812	0.3815	0.7046
$\sigma_{X0}[\text{m}]$	0.1285	0.2006	0.2512	0.3107	0.08374	0.2934	0.3486	0.4073
$\sigma_{Y0}[\text{m}]$	0.04055	0.1674	0.1856	0.3084	0.2731	0.6265	0.4502	0.6897
$\sigma_{Z0}[\text{m}]$	0.02063	0.1702	0.05744	0.2052	0.06718	0.2663	0.0815	0.4833
$\sigma_{\omega0}[\text{gon}]$	0.03016	0.03417	0.05230	0.06711	0.6114	0.817	1.029	1.165
$\sigma_{\phi0}[\text{gon}]$	0.06945	0.1280	0.2041	1.572	0.09485	0.1674	0.5176	0.8814
$\sigma_{\kappa0}[\text{gon}]$	0.2259	0.5935	0.4583	0.7139	0.5834	1.418	1.923	2.067

Tabelle 6.4: Genauigkeit der relativen Orientierung mit den Parametern: Standardabweichung der beobachteten Bildkoordinaten (σ_{xy}), Standardabweichung der unbekannten Objektkoordinaten (σ_{XYZ}), Standardabweichungen der äußeren Orientierungen (σ_{X0} , σ_{Y0} , σ_{Z0} , $\sigma_{\omega0}$, $\sigma_{\phi0}$, $\sigma_{\kappa0}$) der Bilder. * Für die Nordfassade konnte keine durchgehende Sequenz berechnet werden. Die Zahlen beziehen sich auf die berechenbare Teilsequenz.

Zu erkennen ist, dass die Genauigkeiten für die höhere Aufnahmefrequenz von 5 fps im Regelfall besser sind. Zwar ist für ein einzelnes Bildpaar die Aufnahmebasis hier kleiner, die homologen Punkte können jedoch besser verfolgt werden und sind in mehr Bildern vorhanden. Das erhöht die Anzahl der Beobachtungen für die Ausgleichung und damit die Redundanz, die wiederum die Standardabweichung drückt. Auf der anderen Seite wird mit einer kleinen Basis der räumliche

Vorwärtsschnitt ein schleifender Schnitt, d.h., die Entfernung des Schnittpunktes zur Basislinie der Kamera wird ungenauer. Bei einer kleinen Basis wird darüber hinaus auch die Berechnung von essentieller Matrix und Trifokaltensor numerisch instabiler. Bei 5 fps ist für die im Szenario gewählte Fahrzeuggeschwindigkeit die beste Änderung des Szeneninhalts erreicht, bei der ein guter Kompromiss aus Stereobasis auf der einen Seite und Anzahl homologer Punkte auf der anderen Seite gefunden ist. Man erkennt einen deutlichen Unterschied in den Standardabweichungen zwischen der Ost- und Südfassade auf der einen und der West- und Nordfassade auf der anderen Seite. Während die Ost- und die Südfassade eine abwechslungsreiche Geometrie und Textur besitzen, die für die Suche nach homologen Punkten gut geeignet ist, hat die Westfassade zu Beginn eine Fassade mit einer wiederholenden Textur und sehr flacher Geometrie. Dies erschwert die Suche nach korrekten homologen Punkten und die genaue Bestimmung der Orientierungsparameter. Auf der Nordfassade spielt die Verdeckung der Fassade durch die Brücke eine entscheidende Rolle. Solange die Kamera auf die Brücke zufährt, wird die Berechnung durch die Brückenfläche quer zur Fahrtrichtung gestützt. Erreicht die Kamera die Brücke, fällt diese Stützung weg und die zuvor verdeckte Fassade wird sichtbar, für die es kaum homologe Punkte aus früheren Bildern gibt. Die Folge ist, dass die Berechnung entweder komplett abbricht oder aber mit wenigen Punkten und einem entsprechenden Fehler weiterläuft. Der große Abstand der Bilder ist hier nachteilig, weil dann am Übergang von der Brücke zur Fassade die wenigen Bilder ausgelassen werden, die ausreichend homologe Punkte beider Flächen enthalten.

Als Ergebnis der Bündelausgleichung über die gesamte Sequenz ergeben sich im Modellkoordinatensystem des ersten Bildpaares der Sequenz die relativen äußeren Orientierungen der übrigen Bilder und eine Korrektur der inneren Orientierung der Kamera. Außerdem erhält man aus den homologen Punkten der Bilder die korrespondierenden 3D Modellkoordinaten im Modellkoordinatensystem. Die relativen Orientierungen und 3D Modellpunkte werden anschließend verwendet, um über eine Koregistrierung mit dem gemessenen GPS Pfad und dem gegebenen Gebäudemodell den Kamerapfad in das globale Objektkoordinatensystem zu überführen. Abbildung 6.7 zeigt eine 3D Punktwolke der homologen Punkte einer Bildsequenz. In grün sind die geschätzten Kamerapositionen dargestellt. Zu sehen ist die Punktwolke aus einer Sequenz der Ostseite, zu der auch die in Abbildung 5.4 gezeigte Fassade gehört. Man erkennt, dass sich die Mehrzahl der Punkte an den Fensterreihen und im Bereich der Gitter im Erdgeschoss gruppieren.

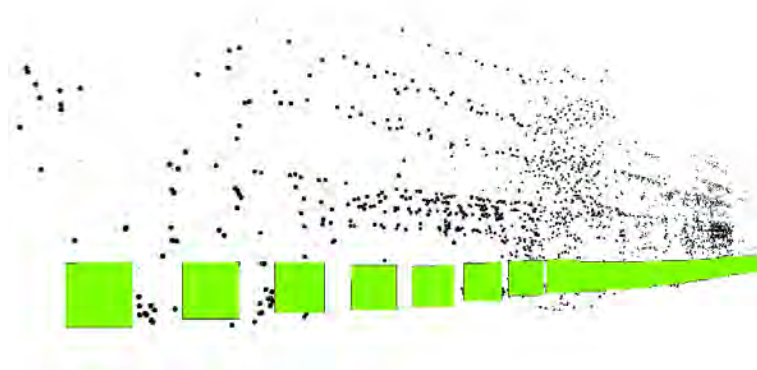


Abbildung 6.7: Relativer Kamerapfad und 3D Punktwolke einer Sequenz in Modellkoordinaten. In grün sind die geschätzten Kamerapositionen dargestellt, in schwarz die 3D Punkte der homologen Punkte der Bildsequenz. Die Punkte sind entlang der Fensterreihen gruppiert.

6.3.2 Restfehler bei der Koregistrierung von Punktwolke und 3D Modell

Die Überführung von geschätztem Kamerapfad und Punktwolke in das Objektkoordinatensystem des 3D Gebäudemodells läuft in zwei Stufen ab. Zunächst findet eine grobe absolute Orientierung über eine Koregistrierung der relativ orientierten Projektionszentren der Bilder mit den GPS Positionen der Kamera statt. Da bei den hier verwendeten Szenarien entlang gerader Straßen gefahren wurde, bleibt der Rollwinkel in Fahrtrichtung unbestimmt. Um dies auszugleichen, wird die Szene in den Modellkoordinaten der ersten Kamera der Sequenz so gedreht, dass die 3D Punkte in vertikalen Ebenen angeordnet sind. Diese Annahme ist zulässig, da die Punkte auf vertikal ausgerichteten Fassaden liegen. Anschließend findet die Koregistrierung mit dem 3D Gebäudemodell statt. Tabelle 6.5 zeigt für die vier Szenarien die jeweils erreichten Genauigkeiten der absoluten Orientierung ($X_0, Y_0, Z_0, \omega, \phi, \kappa$) und des Maßstabs s sowie den mittleren Abstand d_{mean} der 3D Punkte von ihrer korrespondierenden Fassadenfläche. Punkte, die keiner Fassade zugewiesen werden, sind in d_{mean} nicht berücksichtigt.

	Ostfassade	Westfassade	Südfassade	Nordfassade*
σ_{X_0} [m]	0.0483	0.0851	0.0714	(0.258)*
σ_{Y_0} [m]	0.0495	0.0834	0.0713	(0.265)*
σ_{Z_0} [m]	0.133	0.212	0.185	(0.637)*
σ_{ω} [gon]	0.00458	0.00537	0.00483	(0.0263)*
σ_{ϕ} [gon]	0.00472	0.00493	0.00506	(0.0307)*
σ_{κ} [gon]	0.00612	0.0214	0.0174	(0.0441)*
s	0.00278	0.0106	0.00946	(0.273)*
d_{mean}	0.0364	0.0253	0.0238	(0.0461)*

Tabelle 6.5: Genauigkeit der absoluten Orientierung. *Für die Nordfassade konnte keine durchgehende Sequenz berechnet werden. Die Werte beziehen sich daher auf das erste Teilstück bis zur Brücke.

Die Ergebnisse zeigen, dass trotz der Koregistrierung der Punktwolke mit dem 3D Gebäudemodell Standardabweichungen im Bereich von wenigen Zentimetern vorkommen. Interessant ist, dass die Standardabweichungen für die absolute Orientierung für die Ostfassade kleiner sind als für die übrigen Fassade und, dass der mittlere Abstand der 3D Punkte von ihrer Fassade aber größer ist als bei der West- und Südfassade. Der Grund liegt in der Struktur der Ostfassade. Die Unterteilung in mehrere kleine Gebäudeteile führt dazu, dass es nicht nur eine Fassadenebene entlang der Straße gibt, auf der alle Punkte liegen, sondern auch quer zur Straße angeordnete Fassaden. Das führt zum einen dazu, dass die Orientierungsparameter genauer bestimmt werden können, da die optimale Anordnung der 3D Punkte auf den Fassaden durch die unterschiedlich orientierten Fassaden eingeschränkt wird. Zum anderen führt es auch zu größeren Widersprüchen bei der Anordnung der Punkte. Bereits kleine Abweichungen in der Skalierung führen dazu, dass die Punkte nicht mehr auf ihren Fassaden liegen. Darüber hinaus treten durch Bäume Artefakte auf, die nicht vollständig entfernt werden können. Abbildung 6.8 zeigt für die Ostfassade aus den Punktwolken geschätzte Ebenen und das 3D Gebäudemodell nach der Überführung in das Objektkoordinatensystem aber vor der Koregistrierung der Punktwolke mit dem 3D Modell.

Hier ist aus den Punkten, die einer Fassade zugewiesen werden eine Fläche (blau) geschätzt worden, deren Größe sich nach dem Umriss der zu ihr gehörenden Punkte richtet. Man erkennt eine leichte Verkipfung (Abb. 6.8 a) der Punkte gegenüber den Modell. Dieser Einfluss ist auf kleine Fehler in der Bündelblockausgleichung zurückzuführen. Hier werden die Verbesserungen auf alle Parameter der inneren und relativen Orientierung aufgeteilt. Zu sehen ist auch, dass die Flächen von Größe und Abstand her nicht perfekt auf das Gebäudemodell passen. Daher rühren die Restfehler beim Abstand der 3D Punkte vom Gebäudemodell. Der Fehler entlang der Fahrtrichtung (Abb. 6.8 b) hingegen kann gut korrigiert werden. Die koregistrierte Punktwolke der Sequenz der Ostfassade ist in Abbildung 6.9 dargestellt. Die kleinen Quadrate repräsentieren die

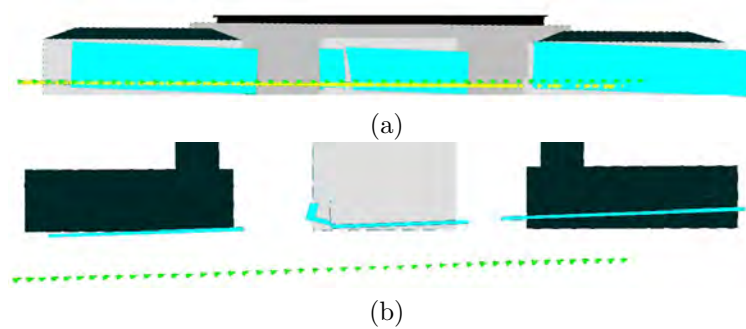


Abbildung 6.8: Koregistrierung Punktwolke 3D Modell. Blau: aus Punktwolken geschätzte Flächen. Grün: Kamerapositionen der verwendeten Bilder. Gelb: ursprünglicher GPS Pfad. a) Ansicht von vorne: zu sehen ist eine leichte Verkippung der aus den Punkten geschätzten Flächen gegenüber dem Modell. b) Ansicht von oben: Kamerapfad und Punktwolke laufen auf das Gebäude zu.

einzelnen 3D Punkte und die Ausrichtung der Fassade, zu der sie gehören. Die gelben Spitzen sind die korrigierten Kameraorientierungen, die blauen der GPS Pfad. Man erkennt, dass die Punkte nun in den Ebenen der Fassaden liegen. Am linken Rand liegen ein paar Punkte außerhalb. Hier konnte kein Skalierungsfaktor bestimmt werden, bei dem alle Punkte auf ihren Flächen liegen. Außerdem sind zwischen den drei Gebäudeteilen Punkte von Bäumen zu sehen.

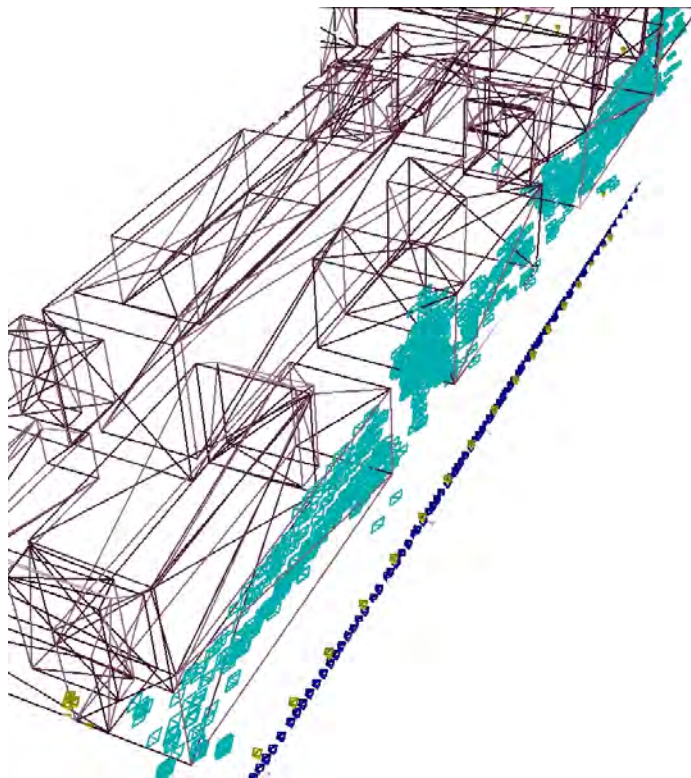


Abbildung 6.9: Koregistrierte Punktwolke. Cyan: 3D Punkte mit Orientierung der zugeordneten Fassadenfläche. Gelb: Korrigierte Kamerapositionen der verwendeten Bilder. Blau: GPS Pfad. Man erkennt am linken Rand einige Punkte, die außerhalb der Fassade liegen und zwischen den Gebäudeteilen Punkte, die zu Bäumen gehören.

In Tabelle 6.5 sind die Werte der Nordfassade speziell markiert. Hier konnte auf Grund der Verdeckungen durch die Brücken kein durchgehender Pfad für die relative Orientierung ermittelt

werden. Daher sind hier nur Werte für den Abschnitt vom Fassadenbeginn bis zur ersten Brücke aufgeführt.

6.4 Ergebnisse des einstufigen Verfahrens

Die in Abschnitt 6.3 gezeigten Ergebnisse liefern einen Hinweis darauf, dass die Orientierung über die 3D Objektpunkte und ihre homologen Bildpunkte eine deutlich höhere Genauigkeit ermöglicht als im zweistufigen Verfahren durch die Koregistrierung mit dem Modell erreicht werden kann. Für das einstufige Verfahren, in dem das Vorwissen über den Kamerapfad als Anfangsschätzung der äußeren Orientierung und die Nebenbedingung der Fassadenpunkte eingebracht werden, ergeben sich die in Tabelle 6.6 aufgeführten Werte. Es werden folgende Werte verglichen: der Varianzfaktor a posteriori (σ_0^2), die Quadratsumme der Verbesserungen der Beobachtungen (vPv^T) und die höchste (v_{max}) und durchschnittliche Verbesserung (v_{mean}), der größte (w_{max}), kleinste (w_{min}) und durchschnittliche Restfehler (w_{mean}) der ins Bild zurückprojizierten 3D Punkte zu ihren beobachteten homologen Punkten, sowie die Standardabweichungen der Parameter der äußeren Orientierung. Darüber hinaus wird der mittlere Abstand d eines Punktes zur Fassade eingeführt.

	Ostfassade		Westfassade		Südfassade		Nordfassade	
	5 fps	1 fps	5 fps	1 fps	5 fps	1 fps	5 fps	1 fps
$\sigma_{xy}[\mu\text{m}]$	0.3534	1.611	1.529	9.119	0.9185	2.994	1.405	4.389
$\sigma_{XYZ}[\text{m}]$	0.05944	0.2754	0.1910	0.5549	0.09583	0.5472	0.1748	0.2095
$\sigma_{X0}[\text{m}]$	0.04454	0.2090	0.06776	0.2982	0.04969	0.3479	0.03390	0.3697
$\sigma_{Y0}[\text{m}]$	0.02294	0.1392	0.01750	0.1784	0.1005	0.6016	0.1972	0.6765
$\sigma_{Z0}[\text{m}]$	0.009547	0.1600	0.03065	0.1963	0.06421	0.2459	0.04909	0.4213
$\sigma_{\omega0}[\text{gon}]$	0.02323	0.09799	0.3148	0.4839	0.5981	0.8034	0.6936	1.103
$\sigma_{\phi0}[\text{gon}]$	0.02196	0.1309	0.08770	0.1621	0.05319	0.1016	0.4930	0.767
$\sigma_{\kappa0}[\text{gon}]$	0.2016	0.2299	0.3806	0.6729	0.5991	0.9379	1.522	1.878

Tabelle 6.6: Genauigkeit der direkten Orientierung. Parameter: Standardabweichung der beobachteten Bildkoordinaten (σ_{xy}), Standardabweichung der unbekannten Objektkoordinaten (σ_{XYZ}), Standardabweichungen der äußeren Orientierungen (σ_{X0} , σ_{Y0} , σ_{Z0} , $\sigma_{\omega0}$, $\sigma_{\phi0}$, $\sigma_{\kappa0}$) der Bilder.

Im Vergleich zu Tabelle 6.4 zeigt sich, dass sich die Werte vor allem bei der hohen Bildwiederholrate deutlich verbessert haben. Der Einfluss des schleifenden Schnittes, der durch die geringe Aufnahmebasis entsteht, kann durch das Vorwissen des Gebäudemodells, auf dem der 3D Punkt zu liegen kommen soll, massiv reduziert werden. Standardabweichung der Objektpunkte σ_{XYZ} liegt in allen Fällen im Bereich einiger Zentimeter.

Für die gut strukturierte Ost- und Südfassade können bereits in einer relativen Orientierung sehr gute Werte bestimmt werden, so dass der Zugewinn durch die direkte Koregistrierung klein bleibt. Es ist hier im Gegenteil sogar so, dass durch die beschränkte Genauigkeit des Gebäudemodells und des GPS Pfades neue Fehlerquellen eingeführt werden. Auf der Ostfassade führen die 3D Punkte der Bäume außerdem zu einem leichten Offset in der Ausrichtung des Kamerapfades.

Bei der West- und der Nordfassade lässt sich eine deutliche Verringerung der Standardabweichungen feststellen. Durch die Vorgabe einer groben äußeren Orientierung kann das Problem der Fehlzuordnung in dem sich wiederholenden Fenstermuster der Westfassade vermieden werden. Durch das Gebäudemodell wird die Drift auf der langgezogenen Fassade reduziert. Durch die Fassadenebene des Modells werden auch Sprünge in den GPS Positionen teilweise ausgeglichen. Die Nordfassade schließlich zeigt für beide Bildwiederholraten einen deutlichen Qualitätsanstieg. Hier wirkt sich das Wissen über die Geometrie insbesondere der beiden Brücken deutlich auf die Genauigkeit aus. Es wird annähernd dieselbe Messgenauigkeit wie bei den anderen Fassaden

erreicht. Allerdings bleibt dort die Problematik einer hohen Verdeckung bestehen, die erhebliche Teile der Fassade betrifft.

6.5 Ergebnisse der Texturextraktion

Bei der Bewertung der Texturextraktion wird zunächst kurz auf die Extraktion aus dem zweistufigen Verfahren eingegangen. Anschließend wird die Extraktion aus dem einstufigen Verfahren untersucht. Es zeigt sich, dass die Extraktion im einstufigen Verfahren mit dem Vorwissen aus dem Gebäudemodell deutlich bessere Extraktionsergebnisse liefert. Abbildung 6.10 zeigt ein Schrägluftbild der Ostfassade. Diese Fassade wird im Folgenden genauer betrachtet.



Abbildung 6.10: Schrägluftbild mit Blick auf die Ostfassade

6.5.1 Texturen aus dem zweistufigen Orientierungsverfahren

Wie in Abbildung 6.8 erläutert ergeben sich aus der Bestimmung der Punktwolke ohne Vorwissen Abweichungen in der Geometrie der geschätzten Ebenen, die mit dem 3D Gebäudemodell koregistriert werden. Dadurch entstehen auch bei der Extraktion der Texturen Abbildungsfehler. In Abbildung 6.11 ist die bereits mehrfach gezeigte Ostfassade zu sehen. Hier sind die um die 3D Punkte herum geschätzten Flächen eingezeichnet und texturiert. Man erkennt die leichte Verkipfung, die bereits in Abbildung 6.8 zu sehen ist. Außerdem fällt auf, dass der Anfang der Fassade links abgeschnitten ist. Der Verkipfungseffekt verstärkt sich im Laufe der Sequenz. Bei der rechten Fassade ist bereits eine sehr deutliche Verkipfung erkennbar. Bei der mittleren Fassade ergibt sich zusätzlich das Problem, dass das oberste Stockwerk nach vorne aus der Fassadenebene heraus versetzt ist und daher nicht als Teil der Fassade erkannt wird.

Verwendet man für die Extraktion der Texturen nicht die aus der Punktwolke geschätzten Ebenen, sondern das gegebene 3D Gebäudemodell und den korrigierten Kamerapfad, so lässt sich der Effekt der Verkipfung reduzieren, wie Abbildung 6.12 zeigt. Hier sind die Ostfassade und die Südfassade mit den jeweils automatisch extrahierten Texturen der Fassadenteile zu sehen, wie sie im Modell vorgegeben sind. Bei der Ostfassade erkennt man im Unterschied zu Abbildung 6.11, dass keine Verkipfung der Texturen mehr vorhanden ist. Deutlich zu sehen ist aber, dass die Texturen abgeschnitten sind. Der Textur an der Gebäudeecke fehlen die ersten drei Fensterspalten. Im Verlauf der Sequenz sieht man, dass die oberen Stockwerke immer stärker abgeschnitten werden. Dieser Fehler hängt mit der bereits erwähnten Verkipfung zusammen. Sie ergibt sich, da in der relativen Orientierung die Verbesserungen auf alle Parameter aufgeteilt werden und durch die lange Sequenz mit vielen Bildern sich bereits kleinste Abweichungen auswirken.

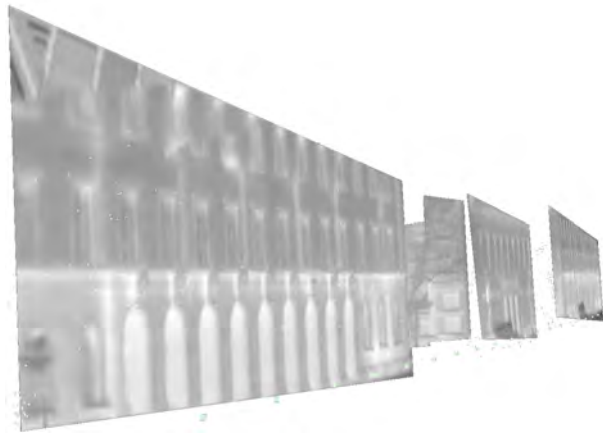


Abbildung 6.11: 3D Ansicht der extrahierten Texturen aus dem zweistufigen Orientierungsverfahren

Bei der Südfassade wird das obere Stockwerke ebenfalls abgeschnitten. Auch hier werden einige Fensterspalten abgeschnitten. Für das Gebäude am linken Rand und den Turm in der Mitte wurde überhaupt keine Fassade erzeugt, weil die Gebäude in der Extraktion übersprungen worden sind. Das Gebäude am linken Rand ist nur in wenigen Bildern zu Beginn der Sequenz zu sehen. Eine gute relative Orientierung für die Extraktion der Texturen ist somit nicht vorhanden. Beim Turm hingegen ist der Fehler auf die Projektion zurückzuführen. Da nur ein kleiner Teil des Turmes in der Sequenz sichtbar ist, schlägt die Berechnung der Texturkoordinaten fehl.

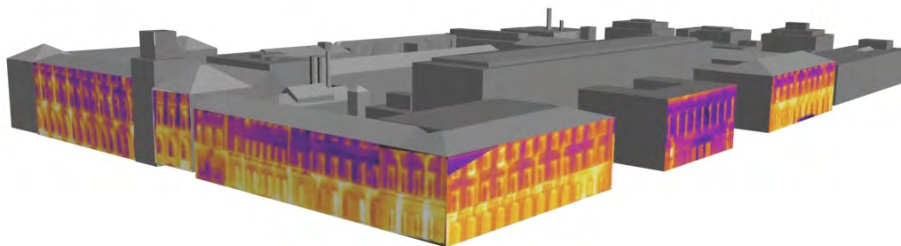


Abbildung 6.12: 3D Ansicht des Gebäudemodells mit IR Texturen aus dem zweistufigen Verfahren

6.5.2 Texturen aus dem einstufigen Orientierungsverfahren

Das einstufige Orientierungsverfahren passt direkt den Kamerapfad an die Fassaden an. Die Ausgleicheung wird jeweils auf die Bilder beschränkt, die eine bestimmte Fassade zeigen und die Projektionsparameter somit jeweils für diese optimiert. Abbildung 6.13 zeigt zwei Texturen derselben Fassade, die aus zwei Bildern derselben Sequenz extrahiert worden sind. Beide Bilder wurden schräg vorwärts blickend aufgenommen. Gut zu sehen ist, dass durch die Schrägsicht das Sichtfeld nicht quadratisch ist. Die geometrische Auflösung nimmt nach rechts hin ab und das Blickfeld wird größer. Vor der Fassade ist ein Transporter zu sehen, der einen Teil der unteren Fensterreihe verdeckt.

Die Überlagerung der Texturen aus den beiden Bildern zeigt bereits nur Abweichungen im Bereich von ein bis zwei Pixeln. Verzerrungen treten praktisch nicht auf. Über eine Koregistrierung mittels Interessenspunkten werden verbliebene Abbildungsfehler zwischen den Teiltexturen beseitigt (Abb. 6.14 a). In Rot und Grün sind die Verknüpfungspunkte dargestellt, die für die Anpassung der beiden Teiltexturen verwendet worden sind. Beginnend mit der ersten Textur werden die Fol-

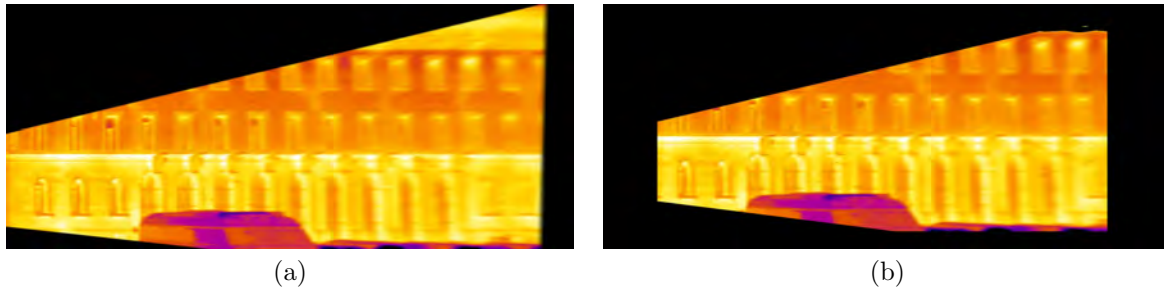


Abbildung 6.13: Extrahierte Texturen aus zwei schräg vorwärts blickenden Bildern. Deutlich zu sehen ist das nach rechts zunehmende Blickfeld bei gleichzeitig abnehmender geometrischer Auflösung. Schwarze Bildpunkte zeigen Fassadenteile, die im Bild nicht zusehen sind.

getexturen aus der Sequenz kopiert. Jedes Folgebild überschreibt mit den Pixeln, für die es einen gültigen Intensitätswert gibt, die bisherige Textur. Bei genauer Betrachtung wird deutlich, dass es nach rechts hin zu einer leichten Verzerrung der Textur kommt. Dies ist auf den verbliebenen Restfehler in der Schätzung der Blickrichtung der Kamera zurückzuführen. Da dieser Fehler jedoch für alle Bilder einer Sequenz besteht, werden die Bilder überlagert, ohne dass dieser Fehler die Überlagerung beeinflusst, aber auch, ohne dass der korrigiert wird. Beim Überschreiben der vorhandenen kombinierten Textur mit einer Folgetextur wird der Fehler immer weiter nach rechts hinausgeschoben. Lediglich in der obersten Fensterreihe bleibt er durch kleine Verzerrungen sichtbar.

Als Ergebnis erhält man die in Abbildung 6.14 b zu sehende Gesamttextur der Fassade. Zu erkennen ist die bis auf kleinere Sprünge in der obersten Fensterreihe störungsfreie Kombination der Einzeltexturen. Die in den Einzelbildern sichtbare Dachunterkante ist auf die Fassade projiziert. Dies liegt an dem Gebäudemodell, das aus einem Stereoluftbild erzeugt wurde und keine Dachüberstände beinhaltet. Zu sehen ist auch, dass der Kleinbus vor der Fassade nicht vollständig entfernt werden konnte. In der oberen Fensterreihe und der auf die Fassade projizierten Dachunterkante erkennt man eine leichte Wellenbewegung. Diese ist darauf zurückzuführen, dass korrespondierende Punkte zwischen Teiltexturen auf Grund der höheren Auflösung und besseren Sichtbarkeit in mehreren Bildern hauptsächlich in den unteren beiden Fensterreihen gefunden werden. Daher wird die Anpassung der Texturen auch für diese beiden Reihen optimiert. Dabei ergibt die Kreuzkorrelation einen durchschnittlichen Wert von 0.9988 für die lokalen Umgebungen der Verknüpfungspunkte. Der mittlere Skalierungsfaktor bei der Anpassung liegt bei 1.01 und der mittlere Korrekturwinkel bei 49.058 mgon. Die Projektion der aufeinanderfolgenden Einzeltexturen einer Sequenz ist im einstufigen Verfahren also sehr stabil, was auf die Ausgleichung über alle Bilder einer Fassade aus einer Sequenz zurückzuführen ist.

In Abbildung 6.15 ist die Textur auf dem Gebäudemodell zu sehen. Daneben ist die ebenfalls extrahierte Textur der Seitenwand dargestellt. Die schwarzen Bereiche waren in keinem Bild der Sequenz zu sehen, da sie durch das Nachbargebäude verdeckt waren. Der Baum vor der Fassade wurde daher mit den Bildern auf die Textur projiziert.

6.6 Kombination von Texturen unterschiedlicher Bildsequenzen

6.6.1 Auswertung mehrerer Texturen unterschiedlicher Blickrichtung

Die Texturen, die aus der vorwärts und der rückwärts blickenden Sequenz erzeugt werden, können unter zwei Bedingung verwendet werden, um die Struktur der Fassade zu untersuchen. Zunächst darf der zeitliche Abstand der Aufnahme nicht zu groß werden, damit keine Temperaturände-

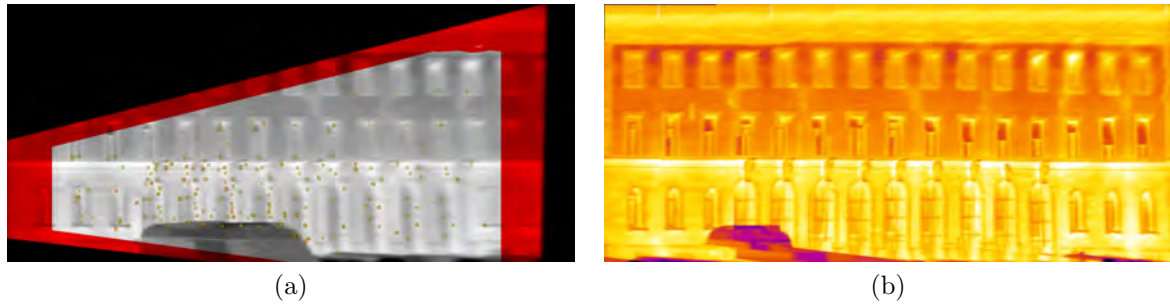


Abbildung 6.14: Aus einer Bildsequenz extrahierte Textur einer Fassade. a) Koregistrierte Überlagerung der beiden Teiltexturen aus Abbildung 6.13. Grün: Interessenspunkte im ersten Bild (rot gefärbt). Rot: Interessenspunkte im zweiten Bild. b) Textur der Frontseite mit Ausrichtung zur Straße. Zu dieser Textur gehören auch die Texturen aus Abbildung 6.13.

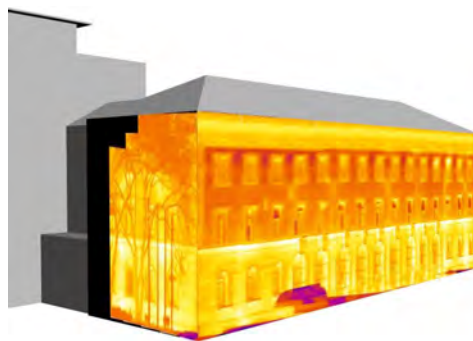


Abbildung 6.15: Texturierten Fassade. Texturen der Frontfassade und der seitlichen Fassade, die teilweise durch andere Gebäudeteile und teilweise durch einen Baum verdeckt ist.

rungen auf der Fassade auftreten. Zum anderen müssen beiden Sequenzen mit dem annähernd gleichen Kamerapfad aufgenommen worden sein, d.h. entweder gleichzeitig aus einer Vorbeifahrt oder dicht nacheinander auf dem selben Weg in der selben Fahrtrichtung, aber mit unterschiedlicher Blickrichtung. Kombiniert man die vorwärtsblickende (Abb. 6.16 a) und die rückwärts blickende (Abb. 6.16 b) Textur, so kann daraus ein Differenzbild erstellt werden.

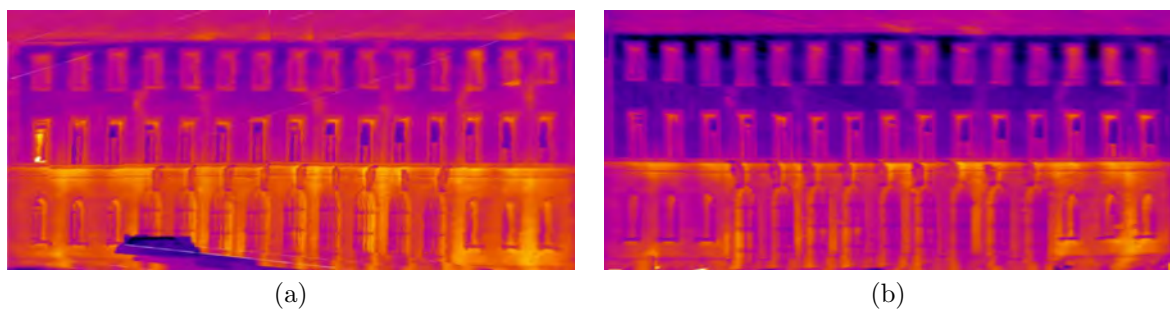


Abbildung 6.16: Texturen aus a) Vorwärts- und b) Rückwärtssicht. Zu erkennen ist der Höhenversatz zwischen beiden Texturen, der durch die unterschiedliche Blickrichtung und Verdeckung der Fassade (Kleintransporter) entsteht.

Technisch bedingt konnten die Texturen nicht gleichzeitig aufgenommen werden, sondern mit einem geringen zeitlichen Versatz, der zu einer leichten Abkühlung der Fassade geführt hat. Diese Abkühlung für die Rückwärtssicht wurde für die Erzeugung der Differenztextur herausgerechnet, um eine gleichzeitige Aufnahme der beiden Texturen zu simulieren. Neben den durch die unterschiedliche Blickrichtung bedingten Unterschieden in der Sichtbarkeit von zurückgesetzten

und hervorstehenden Fassadenelementen zeigen sich leichte Unterschiede in der Abbildung der Textur. So ist der Dachüberstand in Abbildung 6.16 a deutlich breiter auf die Fassadenebene abgebildet als in Abbildung 6.16 b. Bei genauerer Betrachtung erkennt man, dass die gesamte Fassadentextur leicht nach oben verschoben ist. Ursache dafür ist die Tatsache, dass auf Grund der parkenden Fahrzeuge die Bodenkante der Fassade nie zu sehen ist und deswegen der sichtbare Bereich der Fassade nicht bis zum Boden reicht. Kleinere Fehler können auch aus Abweichungen des Kamerapfades durch die zeitlich versetzte Aufnahme entstehen. Abhängig vom Blickwinkel und der geschätzten Höhe der Kamera über dem Boden wird daher der jeweils sichtbare Teil der Fassade auf die gesamte Fassadenhöhe im Modell abgebildet. In der Vorwärtssicht (Abb. 6.16 a) wird in Blickrichtung ein großer Teil der Fassade durch den in der Textur sichtbaren Transporter verdeckt. Der Bodenkante der Fassade dadurch weiter oben auf der echten Fassade gesetzt, als das in der Rückwärtssicht der Fall ist. Hier verdeckt der Transporter die Fassade nicht. Verstärkt wird dieser Effekt dadurch, dass beginnend mit dem ersten Bild, in dem eine Fassade zu sehen ist, das Anhängen der Bilder in der Ausgleichung geschieht. Weil der Transporter am Anfang der Bildsequenz zu sehen ist, zieht sich diese Abweichung durch die gesamte Verbesserung des Kamerapfades. Auch das Zusammensetzen der vollständigen Fassadentextur läuft auf diese Weise ab. Die erste Teiltextur wird als Referenz verwendet und alle Folgetexturen an ihr ausgerichtet. Damit bleibt der Höhenversatz in der zusammengesetzten Textur erhalten.

Die Differenztextur von Vorwärts und Rückwärtssicht ist in Abbildung 6.17 a ohne Korrektur dargestellt. In hellen Bereiche ist die Intensität der Textur aus der Vorwärtssicht höher, in dunklen Bereichen die Intensität der Textur aus der Rückwärtssicht. Zu sehen ist der Versatz am oberen Ende und ein schwarz-weiß Übergang an der Kante zwischen Erdgeschoss und erstem Obergeschoss. Hier ist zusätzlich ein Höhenversatz des Simses vorhanden. Derselbe Effekt kann in der mittleren Fensterreihe beobachtet werden. Durch eine Kreuzkorrelation beider Bilder kann ein großer Teil des Höhenversatzes korrigiert werden. Die so verbesserte Überlagerung ist in Abbildung 6.17 b zu sehen. Zur besseren Visualisierung sind die absoluten Beträge der Änderung dargestellt.

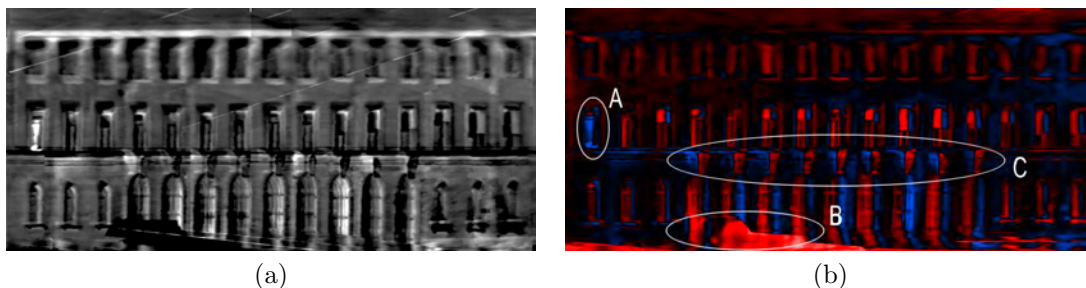


Abbildung 6.17: Differenztextur der Vorwärts- und Rückwärtssicht. a) Differenz ohne Korrektur. Zu sehen sind helle und dunkle horizontale Balkenpaare am Dachansatz, der mittleren Fensterreihe und dem Sims zwischen Erdgeschoss und erstem Obergeschoss, verursacht durch den Höhenversatz der beiden Texturen. b) Differenztextur nach Korrektur. Schwarz: konstante Intensität, Rot: höhere Intensität in Rückwärtsblick, Blau: höhere Intensität in Vorwärtsblick. A) Zwischen beiden Fahrten geschlossenes Fenster. b) Fahrzeug vor der Fassade, das falsch auf die Ebene abgebildet wird. C) Steinplatten mit Fakultätsemlen, die aus der Fassadenebene herausstehen.

Das linke Fenster der mittleren Reihe ist geöffnet und zeigt eine höhere Temperatur als die Fassade. Da dies nur in der ersten Textur der Fall ist, gibt es folglich an dieser Stelle eine große Temperaturabnahme. Weiterhin zu sehen ist der bereits angesprochene Transporter, der ebenfalls nur in der Vorwärtssicht sichtbar ist. Da das Fahrzeug kälter als die Fassade ist, scheint die Wand hier wärmer zu werden. Die Fenster liegen nicht in der Fassadenebene, sondern sind deutlich zurückversetzt. Auf Grund der Schrägsicht sieht die Kamera jeweils nur eine Innenseite

der Fensternische. Diese hat eine andere Temperatur als die Fenster selber, die als Spiegel wirken, und die Fensterrollladen, die teilweise heruntergelassen sind. Die Fenster mit Rollläden sind gut erkennbar an dem Rot-Blauen Balken. Bei den vergitterten Fenstern im Erdgeschoss kommt noch eine Hintergrunderwärmung aus dem Gebäude hinzu, die in einer Fahrtrichtung durch das Fenster auf die Kamera scheint. Am deutlichsten ist die Differenz der beiden Texturen bei den oberhalb der unteren Fensterreihe angebrachten Steintafeln. Diese stehen aus der Fassade hervor und sind kühler als diese. Unter ihnen hält sich jedoch die Wärme der Fassade, was die Tafeln von unten scheinbar leuchten lässt. Durch die Geometrie verdecken die Tafeln in den beiden Texturen einen Teil der Fassade. Der rötliche und bläuliche Schimmer auf der Fassadenfläche kommt durch den Zeitversatz der Aufnahmen und die Schätzung der mittleren Abkühlung zu Stande und kann hier ignoriert werden.

6.6.2 Auswertung mehrerer Texturen unterschiedlicher Aufnahmezeitpunkte

Die Kombination zweier Texturen derselben Blickrichtung zu unterschiedlichen Aufnahmezeitpunkten erlaubt zweierlei Untersuchungen. Im Beispiel aus Abschnitt 6.6.1 ist der Höhenversatz zwischen der vorwärts und der rückwärts blickenden Aufnahmegeometrie diskutiert worden. Dass dieser Versatz hauptsächlich auf den Blickwinkel und die unterschiedliche Verdeckung zurückzuführen ist, zeigt Abbildung 6.18. Hier sind die beiden Texturen der abendlichen (6.18 a) und morgendlichen (6.18 b) Fahrt entlang derselben Fassade in derselben Blickrichtung gegenübergestellt. Man erkennt, dass es zu keinem Höhenversatz kommt. Die Position und Orientierung der Kamera auf dem Fahrzeug blieb zwischen beiden Fahrten unverändert. Auch der Kleintransporter, der einen Teil des unteren Bereichs der Fassade verdeckt, steht unverändert. Daraus lässt sich schließen, dass der Höhenfehler, der in Abschnitt 6.6.1 angesprochen wird, in der Tat mit der Blickrichtung und wechselnden Verdeckung zu tun hat, die Genauigkeit der Texturextraktion für dieselbe Aufnahmekonfiguration aber relativ hoch ist.

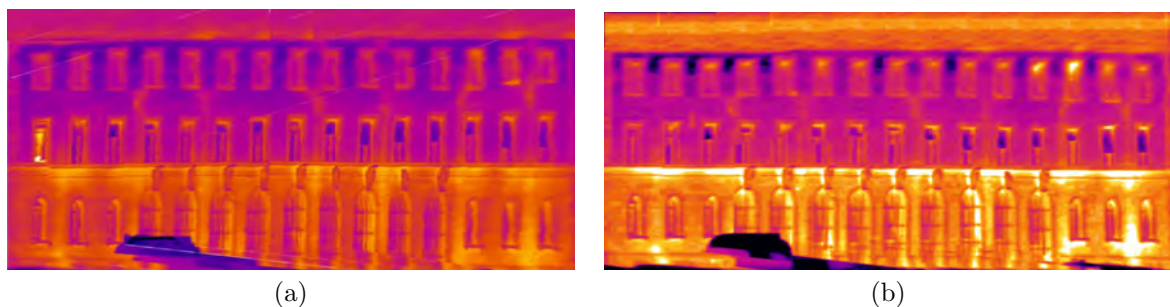


Abbildung 6.18: Texturen aus Vorwärtssicht a) abends und b) am Morgen. Die Umrechnung der Intensitätswerte in die Farbtabelle ist für beide Bilder jeweils so angepasst, dass die Details sichtbar sind. Man sieht, dass beide Texturen geometrisch sehr ähnlich sind. Das spricht für eine gute Wiederholgenauigkeit des Verfahrens.

Diese Beobachtung lässt sich weiter untermauern, indem man für diese beiden Texturen eine Differenztextur bildet (Abb. 6.19). Die Farbkodierung gibt die Temperaturabnahme über Nacht in Grad Celsius an. Die Temperaturen sind aus den Grauwerten über die bekannte radiometrische Kalibrierung bestimmt worden. Gegenüber Abbildung 6.17 fällt auf, dass keine kalt-warm Kanten an Tiefensprüngen auftreten, da beide Texturen von derselben Richtung aus aufgenommen worden sind. Zu sehen ist die deutliche Abkühlung aller Fassadenteile. Man erkennt das geöffnete Fenster und auch die Störung durch das parkende Fahrzeug im Bild unten. Das Fahrzeug ist über Nacht deutlich stärker ausgekühlt wie die Fassade und hat daher eine hohe Differenz in der Intensität. Das gleiche gilt für das Fenster, das bei der Abendfahrt geöffnet war und daher den warmen Innenraum des Raumes und eine Person im Fenster zeigt. Bei der morgendlichen Fahrt war das

Fenster geschlossen und zeigte daher dieselbe Intensität wie die Nachbarfenster. Man erkennt, dass die Rahmen der Fenster sowie die auf die Fassade projizierte Dachunterseite und der Sims zwischen Erdgeschoss und erstem Obergeschoss gut übereinander liegen. Die Bestimmung der Kameraorientierung aus der Bildsequenz mit Hilfe des Gebäudemodells liefert also wiederholbare Genauigkeiten.



Abbildung 6.19: Differenztextur der Abkühlung über die Nacht. Skala angegeben in Grad Celsius, berechnet aus den Grauwerten der Texturen und der radiometrischen Korrektur der Kamera

6.6.3 Auswertung von Texturen unterschiedlicher Kamerasysteme

Unterschiedliche Kamerasysteme unterscheiden sich meist im Wellenlängenbereich, der Bildgröße, dem Sichtfeld und/oder der radiometrischen Auflösung. Am interessantesten ist der Vergleich mit einer RGB-Textur, die mit einer herkömmlichen Kamera aufgenommen wurde. Abbildung 6.20 a zeigt eine RGB-Textur derselben Fassade, für die in den vorangegangenen Abschnitten IR-Texturen erzeugt worden sind. Durch Anwendung eines Difference-of-Gaussian (DoG) Filters werden die Kanten der RGB-Textur bestimmt. Projiziert man diese Kanten in die IR-Textur, ergibt sich ein Eindruck von der Positionsgenauigkeit der extrahierten IR-Textur (Abb. 6.20 b). Die RGB Textur ist manuell erzeugt und über Passpunkte auf der Fassade koregistriert. Die extrahierten Kanten können also als fehlerfrei angenommen werden. Es zeigt sich, dass die Positionen der Fenster im Allgemeinen sehr gut mit den extrahierten Kanten übereinstimmen. Man erkennt an der unteren Fensterreihe, dass die Thermaltextur im Vergleich zur RGB-Textur nach unten gestreckt ist. Auch hier ist wie bei der Koregistrierung der vorwärts und rückwärts blickenden Aufnahmen die Verdeckung des Fassadensockels durch Fahrzeuge eine Ursache. Bereits bei der Orientierung der Bildsequenz führt die Verdeckung dazu, dass die Bildpunkte aus diesem Bereich nicht der Fassadenfläche zugeordnet werden.

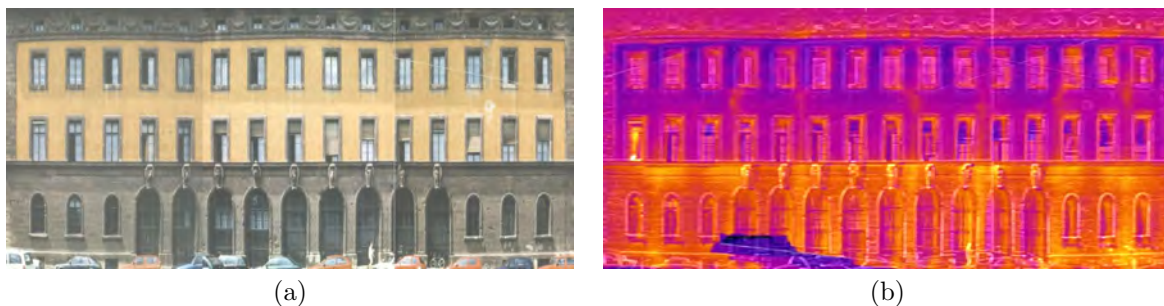


Abbildung 6.20: Überlagerung IR-Textur und RGB-Textur. a) RGB-Textur b) IR-Textur überlagert mit extrahierten Kanten der RGB-Textur. Man erkennt eine gute Übereinstimmung der Fensterpositionen und des Simses. Die untere Fensterreihe ist im Thermalbild gegenüber der RGB-Textur nach unten verzogen.

6.7 Leakedetektion in Infrarottexturen

Für die Leakedetektion (Abschnitt 4.2.3) werden Texturen aus vorwärts und rückwärts blickenden Aufnahmen benötigt. Diese wurden lediglich für die Ostfassade ausgewertet. Abbildung 6.21 zeigt die Verarbeitungsschritte zur Erstellung der Maske für die auszuschließenden Bereiche der Textur. Es ist festzustellen, dass die Maskierung für Fenster höherer geometrischer Auflösung besser funktioniert (Abb. 6.21 b). Dies hat damit zu tun, dass die niedrig aufgelösten Fenster der obersten Etage sehr unscharf und verschwommen abgebildet werden. Dadurch sind trotz eines Versatzes zwischen beiden Texturen die verbleibenden Grauwertdifferenzen relativ gering. Umgekehrt verbleibt trotz Koregistrierung beider Texturen ein Versatz im Sims zwischen Erdgeschoss und erstem Obergeschoss, der im Differenzbild zu sehen ist. Ebenfalls zu sehen sind die Fakultätseembleme, die aus der Fassade herausstehen. Nach der Umwandlung des Differenzbildes in eine Binärmaske (Abb. 6.21 c) wird über ein Bereichswachstum und morphologische Operatoren die endgültige Maske (Abb. 6.21 d) erstellt. Man sieht, dass einige Fenster der obersten Reihe nicht vollständig ausmaskiert werden. Das Sims und die Fakultätseembleme sind in der Maske enthalten. Im Bereich der großen Fenster im Erdgeschoss hat das Bereichswachstum dazu geführt, dass die gesamte Fläche zwischen den Fenstern ebenfalls ausmaskiert wurde. Hier ist allerdings mit bloßem Auge in Abbildung 6.21 a ebenfalls keine Unterscheidung zwischen potentiellen Heizungsleitungen und den Fenstern möglich.

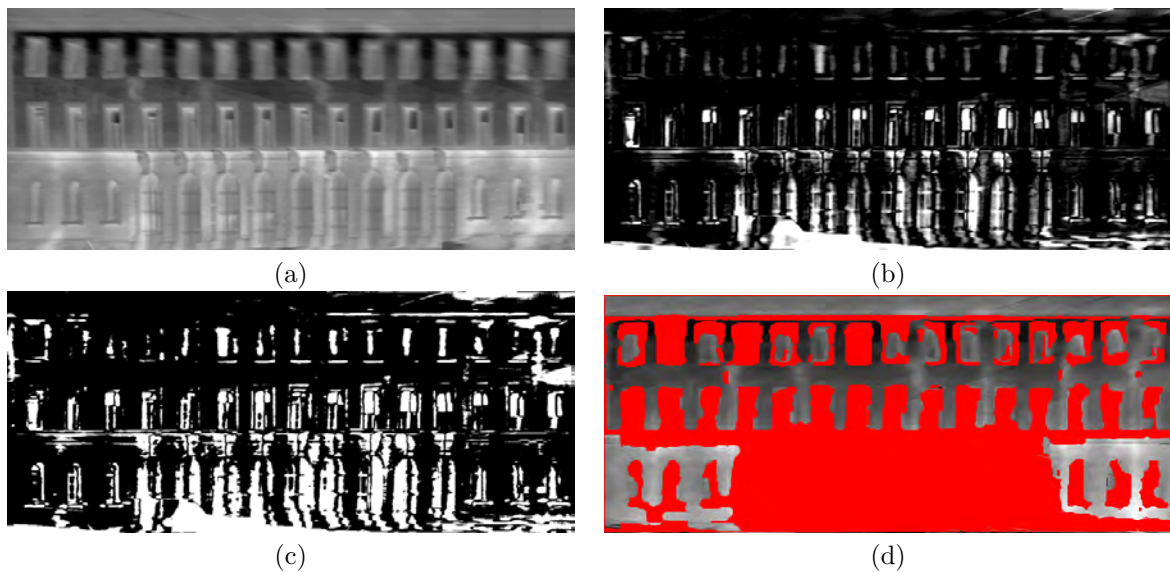


Abbildung 6.21: Leakedetektion: Ausmaskierung von Fassadenobjekten. a) Fassadentextur aus vorwärts blickender Fahrt. Fenster, Simse und Leckagen sind gut zu sehen. Die Fenster heben sich im oberen Teil hell, im unteren dunkel ab. b) Differenzbild der Texturen aus Vorwärts und Rückwärtssicht. Fenster, Simse und Objekte treten hervor. Die Fassadenfläche ist fast völlig schwarz. c) Binärmaske der in (b) gezeigten Differenztextur. d) Endgültige Maske nach einem Bereichswachstum. Der untere Bereich der großen Fenster ist vollkommen ausmaskiert. Bei einigen Fenstern der obersten Reihe sind nur Teile gefunden worden.

Abbildung 6.22 zeigt das Ergebnis der Segmentierung der Textur bezüglich möglicher Leckagen unter Ausschluss der ausmaskierten Teile. Man sieht, dass alle Fenster ausgespart wurden, auch diejenigen, die in der Maske zuvor nicht vollständig ausmaskiert worden sind. Der Grund hierfür ist ihr zu geringer Kontrast zur Fassadenfläche. Man erkennt in gelb, dass sechs der sieben vorhandenen Steigleitungen der Heizung in den oberen Stockwerken gefunden wurden, eine allerdings recht schwach. Die nicht gefundene Steigleitung an der linken Fensterreihe zeigt keinen erkennbaren Kontrast und wurde daher nicht gefunden. Auch die vier sichtbaren Steigleitungen im Erdgeschoss im Bereich der äußeren Fenster links und rechts wurden detektiert, die drei inne-

ren Steigleitungen zwischen den großen Fenstern hingegen nicht. Dort war der gesamte Bereich ausmaskiert und wurde daher nicht untersucht. Daneben sind zwischen ein paar Fenstern der mittleren Reihe kleinere Leckagen detektiert worden, die in der Textur keinen hohen Kontrast zeigen. Diese punktförmigen Leckagen treten an den Schnittkanten der Teiltexturen auf, aus denen die Fassadentextur zusammengesetzt ist. Sie können durch ein nachgelagertes morphologisches Öffnen entfernt werden.

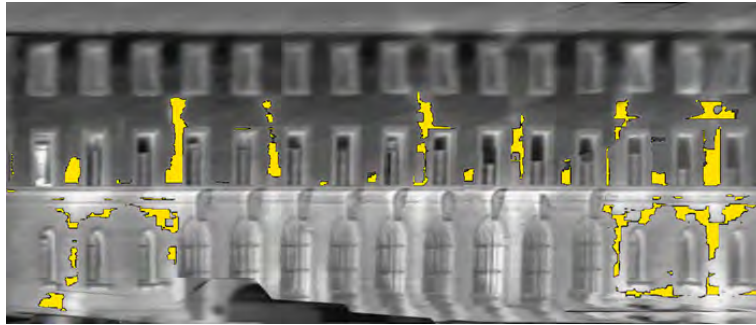


Abbildung 6.22: Markierte Leckagestellen einer Fassade. Fenster und Embleme wurden ausgespart. Leckagen am Rand von Fenstern sind wegen der Ausmaskierung (Abb. 6.21) nicht gefunden worden. Die Heizungsleitungen und Leckagestellen sind in gelb dargestellt.

7 Bewertung

7.1 Einschränkungen durch die Eigenschaften des Infrarotspektrums

Die Arbeiten zur geometrischen Kalibrierung von thermischen Infrarotkameras (Abschnitt 3.1), zur Extraktion von Interestpunkten zur relativen Orientierung von Bildsequenzen (Abschnitt 3.2) und ihrer Koregistrierung mit einem Gebäudemodell, zur direkten Koregistrierung von Bildsequenzen mit einem 3D Gebäudemodell (Abschnitt 3.3) und zur Extraktion von Texturen aus thermischen Infrarotbildsequenzen haben gezeigt, dass Verfahren, die sich für Bildsequenzen im sichtbaren Licht bewährt haben, prinzipiell auf thermische Infrarotbildsequenzen übertragen werden können. Es haben sich jedoch im Verlauf der Untersuchungen drei Eigenschaften der Thermalbilder als limitierende Faktoren der erzielbaren Genauigkeiten und auch der grundsätzlichen Anwendbarkeit bekannter Algorithmen herausgestellt.

Zum einen führt die geringe Bildauflösung zu einer Beschränkung der erzielbaren Messgenauigkeit. Dies zeigt sich bei der Detektion von Messmarken und Passpunkten bei der Kalibrierung (Abschnitt 3.1). Hier konnten für den Laborfall neben der Brennweite c und der Lage des Bildhauptpunktes (x_0, y_0) zwei Verzeichnungsparameter (A_1, A_2) bestimmt werden (Abschnitt 6.1.1). Bei der Kalibrierung ausschließlich aus Passpunkten (Abschnitt 6.1.2) ist eine signifikante Bestimmung nur der drei Parameter der inneren Orientierung (c, x_0, y_0) möglich. Die Verzeichnungsparameter können hingegen nicht signifikant bestimmt werden.

Neben der Bildauflösung spielt die Radiometrie der Infrarotkameras eine entscheidende Rolle. Es können nur Objekte unterschieden werden, deren Abstrahlung im infraroten Spektrum sich unterscheidet. Dies kann entweder auf Grund unterschiedlicher Emissionskoeffizienten oder unterschiedlicher Objekttemperatur geschehen. Da die Emissionskoeffizienten der meisten beobachteten Objekte sehr ähnlich sind, können Objekte meist nur durch Temperaturunterschiede getrennt wahrgenommen werden. Die Ausbreitung von Wärme in massiven Fassaden macht aber eine scharfe Abgrenzung von Objekten sehr schwierig. Tabelle 6.3 zeigt, dass selbst bei gleicher Bildauflösung in den Thermalbildern deutlich weniger Interestpunkte gefunden werden als im Grauwertbild der Videokamera. Dies ist auf schwache Kontrastwerte zurückzuführen. Die Werte in Tabelle 6.3 basieren auf der Verwendung derselben Parameter für alle Bilder. Bei einer Reduktion der Schwellwerte für detektierte Interestpunkte steigt zwar die Zahl der gefundenen Punkte, allerdings nimmt die Zahl der korrekten Zuordnungen kaum zu. Gefundene homologe Punkte hingegen zeigen sich über die Sequenz hinweg ähnlich stabil wie bei den Grauwertbildern der Videokamera.

Die dritte Einschränkung des thermischen Infraroten ist die radiometrische Stabilität. Dieses Problem zeigt sich bei der Kalibrierung mit aufgewärmten Marken, die ihre Erscheinung ständig ändern und damit geometrisch nicht stabil sind. Auch macht es sich insbesondere bei der Koregistrierung von Texturen bemerkbar, die zu unterschiedlichen Zeitpunkten aufgenommen wurden. Während ein bestimmter Interestpunkt sich in einem Bild hell abhebt, kann er in einem ande-

ren Bild dunkler als seiner Umgebung sein. Führt man eine Suche nach korrespondierenden SIFT Features auf Bildern unterschiedlicher Aufnahmezeitpunkte durch, so bricht die Zuordnungsquote rapide ein, da die Feature Vektoren, die aus lokalen Grauwertdifferenzen bestimmt werden, sich zu stark unterscheiden. Damit fällt einer der großen Vorteile von SIFT, die Robustheit gegenüber radiometrischen Veränderungen, hier weg. Die Annahme, dass ein Pixel, das heller ist als seine Nachbarn, relativ zu ihnen immer heller sein wird, trifft auf Thermalaufnahmen häufig nicht zu.

7.2 Geometrische Kalibrierung der Kamera

Für die Kamerakalibrierung sind zwei unterschiedliche Herangehensweisen untersucht worden: Zum einen eine Testfeldkalibrierung (Abschnitt 6.1.1) wie sie von Simmler [2009] und Luhmann et al. [2010] beschrieben wird, zum anderen eine Kalibrierung aus einer aufgenommenen Szene heraus mit Passpunkten (Abschnitt 6.1.2), wie sie für die Kalibrierung einer Infrarotkamera für Luftbilder von Stilla et al. [2009] durchgeführt wird.

Es zeigt sich, dass die geringe Bildauflösung und die radiometrischen Beschränkungen bei der Auswahl von Testmarken und dem Messen von Passpunkten dazu führen, dass die Genauigkeit der Bestimmung der Parameter der inneren Orientierung und der Verzeichnungsparameter deutlich hinter den Werten einer photogrammetrischen Kamera zurückbleibt. Da Zielmarken thermisch stabil über eine Aufnahmesequenz sein müssen, kommen nur aktiv strahlende Marker in Betracht oder Marker, die ohne Bestrahlung eine relativ warme oder kalte Umgebung reflektieren. Die erreichbaren Genauigkeiten für die innere Orientierung sind ausreichend, um sie als Startwerte in der Bündelblockausgleichung der relativen Orientierung der Sequenzen verwenden zu können, als fehlerfrei bestimmt dürfen sie aber nicht angenommen werden. Die zwei bestimmbareren Verzeichnungsparameter stellen ebenfalls nur eine Anfangsnäherung dar.

Bei der Kalibrierung auf Grundlage von Passpunkten in einer Fassadenbildsequenz zeigt sich, dass die Kalibrierung allein aus einer Sequenz mit einer konstanten Blickrichtung heraus verhältnismäßig instabil ist. Zwar können in einer Bündelblockausgleichung Parameter für die innere Orientierung bestimmt werden. Diese sind jedoch nicht signifikant. Dies ist vor allem darauf zurückzuführen, dass nur ein kleiner Gebäudeteil zu sehen ist und die meisten Punkte auf einer Ebene liegen. Diese Ebene wird schräg vorwärts und rückwärts blickend aufgenommen. Durch die Hinzunahme der rückwärts blickenden Bildsequenz entsteht trotz der unterschiedlichen Blickwinkel nur eine geringe Verbesserung. Bei der Koregistrierung der Texturen aus Vorwärtssicht und Rückwärtssicht (Kap. 6.6.1) hat sich bereits gezeigt, dass die Rate der gefundenen homologen Punkte extrem niedrig ist, da die radiometrische Beschreibung der Punkte sich bei einer so hohen Veränderung des Blickwinkels bei der niedrigen Bildauflösung deutlich unterscheiden. Die Hinzunahme eines Bildes, das von einem festen Standort aus die gesamte Fassade mit den Passpunkten zeigt, stabilisiert die Verknüpfung der einzelnen Bilder beider Sequenzen deutlich. Die Zahl der gefundenen Verknüpfungspunkte erhöht sich dadurch deutlich.

7.3 Zweistufiges und einstufiges Orientierungsverfahren

Die beiden untersuchten Orientierungsverfahren unterscheiden sich in der grundsätzlichen Herangehensweise. Das zweistufige Orientierungsverfahren (Abschnitt 3.2) nutzt für die relative Orientierung der Bilder zueinander kein vorhandenes Szenenwissen. Die gemessenen Parameter der äußeren Orientierung werden ebenso wie das Gebäudemodell nur für die Überführung des rein aus den Bildern bestimmten Kamerapfades in das globale Objektkoordinatensystem verwendet. Die kleine Stereobasis und der konstante Blickwinkel einer Bildsequenz führen zu ungenauen Anfangsschätzungen der 3D Koordinaten der homologen Punkte. Während bei der relativen Ori-

entierung der Trifokaltensor die Stabilität durch drei Bilder erhöht, wird im direkten Orientierungsverfahren (Abschnitt 3.3) das bekannte 3D Modell als Referenz eingeführt. Der Vorteil des Trifokaltensors liegt darin, dass er auch für Punkte, die nicht auf der Fassadenebene liegen, die Schätzung der 3D Koordinaten überprüfen und verbessern kann. Dennoch bleibt bei einer kleinen Stereobasis ein gewisser Spielraum für die 3D Koordinaten bestehen. Bei der direkten Orientierung ist das 3D Modell fix. Fehlerhafte Zuordnungen können schnell entdeckt werden. Allerdings muss in jedem Iterationsschritt der Ausgleichung die Gewichtung der Nebenbedingung für alle 3D Punkte angepasst werden, damit Punkte, die nicht auf der Fassadenebene liegen, die äußere Orientierung nicht zu stark beeinflussen. Ist die initiale äußere Orientierung jedoch zu weit von den tatsächlichen Werten entfernt, passiert es, dass das Verfahren keine Lösung findet, da zu wenige 3D Punkte der richtigen Fassade zugeordnet werden können. In diesen Fällen verbessert sich die Orientierung nicht mit jedem Iterationsschritt, sondern bleibt bei einer grob falschen Schätzung hängen.

Bei der Kombination der Bildtriplets in der relativen Orientierung wird als zusätzliche Bedingung die horizontale Anordnung der Punktwolken eingeführt. Diese Nebenbedingung soll das Vorwissen, dass Fassaden beobachtet werden, allgemein einbringen. Da sich zwischen zwei Triplets immer zwei Bilder überlappen, ist dieses Verfahren auch ohne Vorwissen stabil. Durch die kleine Basis zwischen den Aufnahmen können jedoch kleine Abweichungen in den geschätzten Parametern der äußeren Orientierung entstehen und sich über die gesamte Sequenz aufschaukeln. Die Verbesserungen im Rahmen der Ausgleichung werden entsprechend der zu Beginn bekannten bzw. ermittelten Genauigkeiten auf alle Parameter verteilt. Bei der direkten Koregistrierung werden Bildpaare verknüpft. Diese Verknüpfung über nur ein gemeinsames Bild zweier benachbarter Bildpaare wäre ohne das Gebäudemodell zu instabil. Durch die vorgegebene initiale äußere Orientierung und das Gebäudemodell schaukeln sich Fehler insbesondere bei den Winkeln nicht so stark auf wie bei der relativen Orientierung. Insbesondere die Verkippungen, die in Abbildung 6.11 zu sehen sind, können durch das Vorwissen deutlich reduziert werden. Das einstufige Verfahren ist sehr davon abhängig, dass die ersten Bilder korrekt zusammengefügt werden. Der Bildverband wird mit zunehmender Bildzahl stabiler. Durch das sukzessive Anfügen haben Bilder am Anfang der Sequenz mehr Gewicht als Bilder am Ende. Das führt auch zu der in den Abbildungen 6.16 und 6.17 beschriebenen Verschiebung bzw. Stauchung der Textur durch die verdeckte Bodenkante.

Der Einfluss des Gebäudemodells auf die Stabilität äußert sich in zweierlei Hinsicht. Beim einstufigen Verfahren werden die 3D Punkte auf die bekannten Fassadenebenen gezogen. Das reduziert Fehlereinflüsse in der äußeren Orientierung wie Verkippungen. Bilder, in denen mehrere Fassaden zu sehen sind, sind für beide Verfahren von großer Bedeutung. In diesen Bildern lässt sich die Orientierung der Kamera deutlich exakter bestimmen. Es müssen nicht nur alle projizierten Punkte auf einer Ebene liegen, sondern verteilt auf zwei oder mehr Ebenen. Dadurch sinkt die Zahl der möglichen korrekten Konfigurationen der äußeren Orientierung und die Genauigkeit der geschätzten Parameter wird verbessert. Eine Fassade mit mehreren, im 3D Modell vorhandenen Fassadenteilen verschiedener Orientierung verbessert somit direkt die Bestimmung der äußeren Orientierung. Auf der anderen Seite führen solche Fassadenanordnungen häufig zu Verdeckungen und erschweren die Extraktion der Texturen.

Entscheidend für beide Verfahren ist eine ausreichend genaue äußere Orientierung. Das zweistufige Verfahren verwendet die äußere Orientierung zur Überführung der Modellkoordinaten in globale Objektkoordinaten, um anschließend eine Koregistrierung von Punktwolke und 3D Gebäudemodell durchzuführen. Die Orientierung der Bildsequenz selbst ist von der äußeren Orientierung unabhängig, da dieses Vorwissen nicht verwendet wird. Es zeigt sich, dass Höhenfehler im GPS Pfad, wie sie in den verschiedenen Sequenzen vorkommen, keinen Einfluss auf das Ergebnis haben. Hier steigt nur die Anzahl der nötigen Iterationen bei der Koregistrierung an. Lagefehler

hingegen können zu Fehlern in der Projektion führen. Da die extrahierten 3D Punkte nicht auf den Kanten der Fassaden liegen, ist die von den 3D Punkten eingeschlossene Fassadenfläche etwas kleiner als das 3D Polygon des Modells. Die äußere Orientierung hat also entlang der parallel zur Straße laufenden Fassaden etwas Spiel. Durch quer stehende Fassaden wird diese Unsicherheit deutlich eingeschränkt. Ist die Position hier jedoch ungenau, so wird die gesamte Punktwolke verschoben. Durch die Interpolation und Glättung des Kamerapfades kommt es darüber hinaus zu kleinen Fehlern im Abstand zum Gebäudemodell. Im einstufigen Verfahren geht der interpolierte und geglättete Kamerapfad dagegen in die Ausgleichung der Bildsequenz ein. Dadurch können Orientierungsfehler an das Modell angepasst minimiert werden. Auf der anderen Seite führen grobe Fehler bei der äußeren Orientierung dazu, dass die Ausgleichung unter Umständen abbricht. Generell kann festgestellt werden, dass das zweistufige Verfahren mit der relativen Orientierung die Orientierung aller Bilder zueinander optimiert, während das einstufige Verfahren mit der direkten Koregistrierung die Bilder pro Fassade optimiert. Dieser Unterschied erklärt auch, warum das zweistufige Verfahren bei der Nordfassade versagt und das einstufige ein Ergebnis liefert. Die Optimierung rein basierend auf einer relativen Orientierung scheitert, sobald nicht genug homologen Punkte vorhanden sind, um eine stabile Schätzung durchzuführen. Die gesamte Sequenz ist nur so stabil wie die schwächste Verbindung. Beim einstufigen Verfahren stützt das Vorwissen des Modells die korrekte Zuordnung von homologen Punkten und die Bestimmung ihrer 3D Koordinaten. Außerdem findet die Optimierung stückweise für jede Fassade einzeln statt. Schlägt das Verfahren für eine Fassade fehl, so kann es für andere Fassaden der Sequenz trotzdem erfolgreich sein.

Das zweistufige Verfahren mit der relativen Orientierung der Bildsequenz und anschließenden Koregistrierung auf das 3D Modell ist dann zu bevorzugen, wenn keine oder sehr ungenaue Daten über die äußere Orientierung vorliegen. Allerdings ist die erreichbare Genauigkeit ohne das Vorwissen geringer (Tab. 6.4) als bei der Einbindung des Gebäudemodells in die Orientierung der Bildsequenz (Tab. 6.6). Das einstufige Verfahren, das die äußere Orientierung und das 3D Modell in die Orientierung der Bildsequenz einbezieht, liefert unter der Bedingung, dass die äußere Orientierung hinreichend genau gegeben ist, bessere Ergebnisse.

7.4 Texturextraktion

Die Texturextraktion findet durch eine Projektion der Einzelbilder einer Bildsequenz auf die Fassaden des Gebäudemodells statt. Im Gegensatz zu Verfahren, bei denen über gemeinsame Kanten im Modell und den Bildern im Bildraum nochmals koregistriert werden kann (Stilla et al. [2009]), ist dies auf Grund des geringen Blickfelds und des geringen Detailgrades der verwendeten Modelle hier nicht möglich. Die aus der Orientierung der Bildsequenzen gewonnenen verbesserten Parameter der inneren und äußeren Orientierung werden direkt für die Projektion der Einzeltexturen aus den Einzelbildern der Sequenz verwendet. Die Genauigkeit der Projektion kann daher nur indirekt bestimmt werden. Zum einen werden die Einzeltexturen der Einzelbilder zur kompletten Fassadentextur zusammengesetzt. Die Koregistrierung dieser Einzeltexturen zeigt, dass die Projektion über eine ausgeglichene Sequenz von Bildern stabil ist und lediglich kleinere Korrekturen vorgenommen werden müssen, die auf kleine Bewegungen des Kamerasystems zurückzuführen sind, die nicht im GPS Pfad aufgezeichnet wurden.

Ein Vergleich von Texturen unterschiedlicher IR Sequenzen untereinander und mit einer als Referenz verwendeten RGB-Textur zeigt, dass bei der Texturextraktion vor allem zwei Fehler auftreten. Zum einen führen Verdeckungen des Fassadensockels durch parkende Fahrzeuge dazu, dass es am unteren Ende der Fassade keine 3D Punkte gibt. Dadurch wird die Kameraposition in der Höhe niedriger eingeschätzt als sie in Wirklichkeit ist. In der Folge wird bei der Projektion des Modells in die Einzelbilder der Bereich der Bilder, der die Fahrzeuge zeigt, abgeschnitten. Die

Textur beginnt am unteren Rand der Fassade mit dem sichtbaren Fassadenteil, der dadurch nach unten verschoben ist. Da zumeist parkende Fahrzeuge in langen Reihen am Gebäude entlang stehen, wird dieser Fehler auch bei der Kombination der Einzelbilder zur Sequenz nicht korrigiert. Hinzu kommt, dass bei der Kombination der Teiltexturen die Teiltextur des ersten Bildes die Position aller Folgebilder vorgibt. Dieses Problem lässt sich nur lösen, wenn die Höhe des Kamerasystems über dem Fußpunkt der Fassade genau bekannt ist. Die notwendige Genauigkeit ist aber mit üblichen GPS Systemen insbesondere in Straßenschluchten nicht zu erreichen.

Die zweite Fehlerquelle ergibt sich durch das Gebäudemodell. In 3D Gebäudemodellen sind zumeist keine Dachüberstände modelliert. Daher wird die von der Straße aus sichtbare Dachunterkante auf die Fassade projiziert. Da die Fassade durch verdeckte Bereiche am unteren Ende abgeschnitten und dadurch nach unten verschoben ist, werden diese Teile am oberen Ende als vermeintliche Fassadenelemente der Textur hinzugefügt. Dieses Problem ist nur lösbar durch einen höheren Detailgrad des Gebäudemodells, in dem solche Vorsprünge enthalten sind.

Durch die Aufnahmegeometrie bedingt ist die geometrische Auflösung der Texturen nicht konstant. Dies vereinfacht die Kombination der Teiltexturen, da jede Folgetextur bei der Vorwärts- und jede Vorgängertextur bei der Rückwärtssicht in den Bereichen der Textur, die in der jeweiligen Teiltextur sichtbar sind, eine höhere geometrische Auflösung hat als die bereits zusammengesetzten Teiltexturen. Dadurch entstehen an manchen Stellen sichtbare Kanten beim Übergang des Detailgrades. Diese Kanten sind abhängig von der Aufnahmefrequenz und dem Abstand zweier benachbarter Bilder. Ist der Abstand kleiner, so sind zwei Texturen ähnlicher und der Unterschied der geometrischen Auflösung geringer. Die Kante am Übergang ist schwächer als bei einem großen Abstand mit einer starken Änderung der geometrischen Auflösung am Übergang von einer Teiltextur zur nächsten. Natürlich könnte man den Übergang auch wie bei der Bildung von Orthophotomosaiken weich gestalten oder Kanten für den Übergang suchen. Da es hier aber nicht um eine optisch 'schöne' Textur geht, sondern um eine möglichst hoch aufgelöste Darstellung, ist hier auf solche Vorgehensweisen verzichtet worden.

7.5 Texturkombination

Die Kombination von Texturen verschiedener Bildsequenzen ist in Abschnitt 6.6 anhand einiger Beispiele demonstriert worden. Die Kombination der Texturen aus vorwärts und rückwärts blickenden Aufnahmen liefert die Möglichkeit, Objekte, die nicht in der Fassadenebene liegen, zu erkennen. Da sie geometrisch an der falschen Stelle der Fassade abgebildet werden, müssen solche Elemente für die Bestimmung von Temperatur oder Abstrahlung entfernt werden. Sie enthalten häufig falsche, für die Fassade nicht repräsentative Elemente, z.B. Fenster oder Simse, wie sie in Abbildung 6.17 deutlich zu erkennen sind. Ob die Kombination von Vorwärts- und Rückwärtssicht ausreicht, um eine 3D Rekonstruktion dieser Elemente auf der Fassadenfläche durchzuführen, ist fraglich. Zum einen ist die Stereobasis zwar sehr groß, auf der anderen Seite tritt aber das Problem der automatischen Zuordnung der korrespondierenden Punkte auf. Eine genauere Untersuchung hierzu und eine geeignete Anpassung der Aufnahmekonfiguration kann Teil zukünftiger Forschungsarbeit sein. Die unterschiedliche Erscheinung derselben geometrischen Objekte spielt auch eine Rolle bei der Koregistrierung von Texturen aus verschiedenen Blickrichtungen. Da unterschiedliche Blickrichtungen unterschiedliche Bereiche verdecken, treten die beschriebenen Effekte beim Versatz der Texturen auf der Fassade unterschiedlich stark auf. Je nach Struktur der Fassade kann eine automatische Koregistrierung damit an der Suche nach korrespondierenden Punkten scheitern.

Die Kombination von Texturen aus Sequenzen, die zu unterschiedlichen Zeiten aufgenommen wurden, führt direkt zur Untersuchung dynamischer Prozesse von Aufheizung und Abkühlung.

Bei der Kombination zweier Texturen, die aus Sequenzen gleicher Blickrichtung aufgenommen wurden, zeigt sich, dass die Schätzung der Fassadenebene annähernd die gleichen Ergebnisse liefert. Im Gegensatz zur vorwärts/rückwärts Kombination tritt kein erkennbarer Versatz in den Texturen auf, der vor einer gemeinsamen Auswertung korrigiert werden müsste. Allerdings trifft diese Aussage nur dann zu, wenn sich die verdeckten Bereiche zwischen den beiden Aufnahmen nicht entscheidend ändern. Zwar ist es egal, ob vor dem Gebäude dieselben Autos stehen, ein größerer Transporter oder Lastwagen verfälscht hingegen nicht nur die Textur an der entsprechende Stelle der Fassade, sondern kann die Schätzung der Fassadenfläche beeinflussen und wie bereits beschrieben zu einer anderen Positionierung der Textur auf der Fassade führen. Ist eine Koregistrierung durchzuführen, dann ist es wiederum abhängig von den äußeren Bedingungen, ob markante Punkte der einen Textur auch in der anderen Textur auftreten. So können in einer Sommeraufnahme Heizungsleitungen nicht zu sehen sein und die Fenstersimse kühler sein als die Fassade. Im Winterbild hingegen treten die Heizungsleitungen deutlich hervor und auch die Simse sind auf Grund der Heizkörper im Innenraum plötzlich wärmer als die umgebende Fassadenfläche.

Die Überlagerung mit den detektierten Kanten der RGB-Textur (Abb. 6.20) zeigt den Höhenfehler der automatisch extrahierten Infrarottexturen. Man sieht aber auch, dass die Position entlang der Längsseite der Fassade sehr gut übereinstimmt. Es stellt sich hier die Frage, inwiefern die vorhandene RGB-Textur als Ersatz für geometrische Details des Gebäudemodells für eine Koregistrierung mit der Infrarotbildsequenz verwendet werden kann.

7.6 Leckagedetektion

Die Leckagedetektion zeigt grundsätzlich die Möglichkeit, aus den Thermaltexturen Objekte über Segmentierungsverfahren zu extrahieren. Die Erstellung der Masken für Bereiche, die nicht untersucht werden sollen, erfolgt ohne Vorwissen oder Interpretation. Eine Detektion der Fenster, wie sie von Reznik & Mayer [2008] oder Michaelsen et al. [2012] vorgeschlagen wird, würde deutlich exaktere Masken liefern. Summiert man für jede Zeile und jede Spalte die maskierten Pixel, so ergibt sich eine Häufigkeitsverteilung, die verwendet werden könnte, um über eine Grammatik die Extraktion von Fenstern zu unterstützen. Für die reine Detektion von Leckagestellen ist das nicht nötig gewesen. Die Fensterdetektion ist nicht Teil dieser Arbeit und die verschiedenen Methoden der Detektion bieten alleine genug Material für umfangreiche Untersuchungen.

8 Zusammenfassung und Ausblick

8.1 Zusammenfassung

Das Ziel dieser Arbeit ist die Umsetzung von Verfahren der Photogrammetrie zur Extraktion und Interpretation von Fassadentexturen aus Bildsequenzen aus dem thermischen Infrarot. Im Rahmen der Untersuchungen wurden die in Abschnitt 1.3 aufgezeigten Fragestellungen näher betrachtet:

- (1) Die Bestimmung der Parameter der inneren Orientierung für thermische Infrarotkameras ist grundsätzlich möglich. Die erzielten Ergebnisse liegen bezüglich ihrer Genauigkeit zwar deutlich hinter den Genauigkeiten, die photogrammetrische Kamerasysteme erreichen, sie sind aber ausreichend, um als initiale Werte für eine Bündelausgleichung verwendet zu werden. Neben einer Kalibrierung über einen Messaufbau wurde auch die Kalibrierung anhand von in einer aufgenommenen Szene ermittelten Passpunkten untersucht. Dieses Vorgehen ist insbesondere für eingebaute Systeme in Fahrzeugen von Vorteil und erlaubt eine Kalibrierung vor einer Messfahrt.
- (2) Mit dem Förstner-Operator in Verbindung mit einer Korrelation und dem SIFT Algorithmus wurden zwei Verfahren zur Detektion und zum Vergleich von Interestpunkten untersucht. Die Ergebnisse zeigen, dass sowohl Gradienten basierte Verfahren wie der Förstner-Operator als auch Blob basierte Verfahren wie SIFT auf thermische Infrarotbilder übertragen werden können. Bedingt durch schwächere und weichere Kantenübergänge und weniger Details ist die Anzahl der gefundenen Interestpunkte allerdings deutlich niedriger als bei einem Bild im sichtbaren Licht bei gleicher Bildgröße und geometrischer Auflösung. Es zeigt sich, dass die Zuordnung der homologen Punkte innerhalb einer Sequenz stabil ist. Bei größeren Änderungen des Blickwinkels und größerem zeitlichen Versatz treten jedoch zunehmend Schwierigkeiten in der Zuordnung auf. Dies ist darauf zurückzuführen, dass Veränderungen in der Abstrahlung im Infrarotspektrum von Blickwinkel und Temperatur abhängen. Damit greifen Modelle, bei denen die Beschreibung der Interestpunkte auf lokale Intensitätsverhältnisse zurückgreifen, in diesen Fällen nicht mehr uneingeschränkt.
- (3) Es wurden zwei Verfahren zur Orientierung von Bildsequenzen vorgestellt. Im ersten Verfahren wird anhand von homologen Punkten über die Verknüpfung von Bildtriplets zunächst eine relative Orientierung der Bildsequenz hergestellt. Anschließend wird diese Bildsequenz über den gemessenen GPS Pfad mit dem gegebenen Gebäudemodell koregistriert. Es zeigt sich, dass dieses Verfahren für gut strukturierte Fassaden gute Ergebnisse liefert, dass aber bei langgezogenen Fassaden kleine Fehler in der Schätzung der Orientierungsparameter zu stärkeren Abweichungen bei der späteren Berechnung der Texturen führen können. Das zweite Verfahren nutzt die bekannten GPS Positionen und das gegebene Gebäudemodell direkt in der Orientierung der Bildsequenz. Dadurch kann das summieren von kleinen Orientierungsfehlern vermieden werden. Allerdings beeinflusst die beschränkte Genauigkeit der GPS

Messungen die Qualität der Ergebnisse. Daher ist dieses Verfahren nur dann geeignet, wenn ausreichend gute Anfangswerte der äußeren Orientierung gegeben sind.

- (4) Basierend auf einer schräg vorwärts und schräg rückwärts blickenden Kamera konnte gezeigt werden, dass Folgetexturen einer Sequenz praktisch ohne Informationsverlust übereinander kopiert werden können und so die komplette Fassadentextur erzeugt wird. Die resultierenden Texturen zeigen auf Grund der Schrägsicht keine konstante geometrische Auflösung. Am höchsten ist sie in der Sichtebene der Kamera und nimmt mit den Abstand zu dieser Sichte Ebene ab. Bei der Extraktion der Texturen zeigt sich, dass der Fußpunkt von Fassaden häufig durch Fahrzeuge verdeckt ist und daher im Bild zu hoch geschätzt wird. In der Folge sind die Texturen nach unten verschoben, anstatt auf die Fassade projizierte Fahrzeuge zu zeigen. Im Gegenzug führt die beschränkte Genauigkeit des verwendeten Modells dazu, dass die Unterseite des Dachüberstandes auf die Fassadenebene projiziert wird.
- (5) Es wurden verschiedene Ansätze zur Weiterverarbeitung der thermalen Fassadentexturen untersucht. Die Kombination von Texturen aus der Vorwärts- und Rückwärtssicht erlaubt die Detektion von Elementen, die außerhalb der Fassadenebene liegen und daher an falscher Stelle auf der Textur abgebildet sind. Die Kombination von Texturen aus Aufnahmen verschiedener Zeitpunkte ermöglicht die Untersuchung dynamischer Prozesse, wie z.B. die Aufheizung und Abkühlung oder die Untersuchung durchgeführter Sanierungsmaßnahmen. Es wurde eine mögliche Prozesskette vorgestellt, um in den Texturen Leckagen und Fenster zu detektieren. Weitergehende Untersuchung zur geometrischen Extraktion wurden in dieser Arbeit hingegen nicht durchgeführt.

Insgesamt wurde gezeigt, dass die in der Photogrammetrie gebräuchlichen Verfahren prinzipiell auf Bilder aus dem thermischen Infrarot übertragbar sind. Es wurden einige Einschränkungen und Besonderheiten erläutert und Untersuchungen zur Genauigkeit durchgeführt, die zeigen, dass auf Grund der im Vergleich geringen Bildauflösung und der Eigenschaften von Objekten im thermischen Infrarot die erzielbaren Genauigkeiten erheblich schlechter sind als im sichtbaren Spektrum.

8.2 Ausblick

In dieser Arbeit wurden einige grundsätzliche Fragestellungen zur erreichbaren Genauigkeit bei thermischen Infrarotaufnahmen für die 3D Extraktion und Texturierung untersucht. Die erzielten Genauigkeiten basieren auf den vorhandenen Daten. Für zukünftige Untersuchungen können weitere oder verbesserte Sensoren eingesetzt werden. Im Laufe der Arbeit sind bereits neue Infrarotkameras auf den Markt gekommen, die über eine größere Pixelzahl verfügen. Solche Kameras erlauben eine höhere geometrische Auflösung der Szene. Dadurch kann entweder der geometrische Detailgrad erhöht oder die Fassade in Nadirsicht aufgenommen werden. Hierdurch steigt die Zahl der detektierbaren Interestpunkte an und die Verknüpfung der Einzelbilder wird stabiler. Die radiometrische Auflösung der SC3000 Kamera steht mittlerweile bereits bei preisgünstigeren Kamerasystemen zur Verfügung, sodass die erzielte radiometrische Auflösung als realistisch für weiter Untersuchungen gelten darf.

Für die Bestimmung der inneren Orientierung und der Verzeichnungsparameter ist die Untersuchung weiterer Varianten von Messaufbauten notwendig. Denkbar sind hier reflektierende Marken, die über ein Infrarotblitzlicht beleuchtet werden und so keinen Aufwärmeeffekten unterworfen werden, die zu kleineren Abweichungen in der Schätzung der Markenmittelpunkte führen.

Insbesondere das einstufige Verfahren benötigt eine ausreichend gute äußere Orientierung als Startwert für die Bündelausgleichung. In den bisher durchgeführten Versuchen war aus technischen Gründen lediglich ein GPS System eingebaut. Die Blickwinkel wurden zum Sequenzbeginn einmalig gemessen und im Verlauf aus der Bewegungsrichtung abgeleitet. Der Einsatz eines INS Systems kann die Schätzung der Blickrichtung erheblich verbessern und gerade in Straßenschluchten mit hoher Verdeckung die Positionsbestimmung unterstützen. Ein zusätzliches Odometer zur Wegstreckenmessung könnte auch bei langem Ausfall des GPS die Drift des Systems reduzieren.

Die Koregistrierung der Bildsequenzen mit dem Gebäudemodell kann auf zweierlei Arten weiter verbessert werden. Zum einen wird durch eine detailreichere Gebäudegeometrie die Abbildung auf die Fassadenebenen verbessert, wenn beispielsweise der Dachüberstand korrekt modelliert ist. Zum anderen führt eine detaillierte Geometrie dazu, dass die Projektion des 3D Modells in den Bildraum der Kamera über mehr Kanten und Ecken verfügt, die eine Zuordnung mit Kanten und Ecken aus den Bilddaten erlauben. Es ist zu untersuchen, wie solche Zusatzdaten in die bestehende Bündelblockausgleichung eingeführt werden können. Alternativ zur verfeinerten Geometrie bietet es sich an, eine vorhandene Fassadentextur in den Bildraum der Kamera zu projizieren und deren Kanten mit den Infrarotbildern zu koregistrieren. Für beide Varianten sind die erzielbaren Genauigkeitsgewinne zu untersuchen. Es stellt sich außerdem die Frage, ob eine Anpassung der Gewichte des stochastischen Modells zu einer Steigerung der Qualität führen kann.

Insbesondere die Auswertung der Fassadentexturen ist in dieser Arbeit nur kurz angerissen worden. Hier bietet sich die Möglichkeit, durch Kombination von Texturen verschiedener Spektralbereiche einen deutlichen Gewinn bei der Extraktion von Fassadenobjekten und Fassadenparametern zu erzielen. Da für die meisten Materialien an Fassaden die genauen Emissionskoeffizienten unbekannt sind, lässt sich auch die Temperatur nicht exakt bestimmen. Aus der Satellitenfernerkundung sind Verfahren bekannt, aus dem Intensitätsverhältnis zweier Thermalkanäle über das Wiensche Verschiebungsgesetz (Wien [1896]) den Emissionsgrad zu schätzen und damit die Temperaturbestimmung verbessern. Die Kombination einer Textur aus dem sichtbaren Licht und einer Thermaltextur bietet die Möglichkeit, bestehende Verfahren zur Detektion von Fenstern zu erweitern. Während in Thermalaufnahmen die Detektion von Fenstern als Unregelmäßigkeiten auf der Fassade möglich ist, ist es in der hoch aufgelösten Textur des sichtbaren Lichts leichter, die exakte Geometrie eines Fensters zu ermitteln. Es zeigt sich, dass durch eine Segmentierung und anschließende Gruppierung Fensterreihen und Spalten gefunden werden können, die im weiteren Verlauf für die Extraktion von Fenstern oder die Verfeinerung der Maskierung für die Leckagedetektion verwendet werden könnten.

8.3 Danksagung

Für die Bereitstellung des Messfahrzeuges mit Infrarotkameras und der Unterstützung bei der Messung dankt der Autor Herrn Dr. Clement, Herrn Dr. Schwarz und Herrn Kremer von Forschungsinstitut für Optronik und Mustererkennung (FGAN-FOM), Ettlingen, heute Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung (IOSB), Ettlingen. Die vorgestellten Arbeiten sind durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen des Projektes „Enrichment and Multi-purpose Visualization of Building Models with Emphasis on Thermal Infrared Data“, unterstützt worden. Die Arbeit entstand in Zusammenarbeit mit den Mitgliedern des Leonhard Obermeyer Centers (LOC) - TUM Center of Digital Methods for the Built Environment.

Literaturverzeichnis

- Agarwal S, Snavely N, Simon I, Seitz S, Szeliski R (2009) Building rome in a day. In: *Proceedings of the International Conference on Computer Vision*
- Bachler T (2001) *Arbeiten mit der Camera obscura*. Stuttgart: Lindemanns. ISBN 3895062227.
- Bartelsen J, Mayer H (2010) Orientation of Image Sequences Acquired from UAVs and with GPS Cameras. *Surveying and Land Information Science*, 70 (3): 151–159.
- Bay H, Ess A, Tuytelaars T, van Gool L (2008) SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, 110 (3): 346–359.
- Becker S (2010) *Automatische Ableitung und Anwendung von Regeln für die Rekonstruktion von Fassaden aus heterogenen Sensordaten*. Ph.D. thesis Universität Stuttgart.
- Borrmann D, Houshiar H, Elseberg J, Nüchter A (2013) Vom Kombinieren von 3D-Modellen mit Farb- und Temperaturinformationen. *Photogrammetrie – Laserscanning – Optische 3D-Messtechnik. Beiträge der Oldenburger 3D-Tage 2013*, : 246–253.
- Brown M, Lowe D (2007) Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision*, 74 (1): 59–73.
- Canny J (1986) A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8 (6): 679–698.
- Chum O, Matas J, Kittler J (2003) Locally Optimized RANSAC. In: Michaelis B, Krell G (eds) *Proceedings of the DAGM 2003*, 2781.
- DIN 54190 (2011) *Zerstörungsfreie Prüfung - Thermografische Prüfung*.
- DIN 54191 (2009) *Zerstörungsfreie Prüfung – Thermografische Prüfung elektrischer Anlagen*.
- DIN 54192 (2010) *Zerstörungsfreie Prüfung – Aktive Thermografie*.
- DIN EN 13187 (1998) *Wärmetechnisches Verhalten von Gebäuden – Nachweis von Wärmebrücken in Gebäudehüllen – Infrarot-Verfahren*.
- Ebner H (1976) Self-calibrating bundle adjustment. Invited Paper. In: *Proceedings of the 13th Congress of the International Society for Photogrammetry*, 21 (13).
- Fischer R (1992) What’s so Different About IR Lens Design? In: *Proc. SPIE CR41*
- Fischler M, Bolles R (1981) Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24 (6): 381–395.
- Fitzgibbon A, Zisserman A (1998) Automatic Camera Recovery for Closed or Open Image Sequences. *Lecture Notes in Computer Science*, 1998 (1406).
- Fouad N, Richter T (2012) *Leitfaden Thermografie im Bauwesen*. Fraunhofer IRB Verlag, Stuttgart.

- Frahm J, Fite-Georgel P, Gallup D, Johnson T, Raguram R, Wu C, Jen Y, Dunn E, Clipp B, Lazebnik S, Pollefeys M (2010) Building rome on a cloudless day. In: Daniilidis K, Maragos P, Paragios N (eds) *Proceedings of the European Conference on Computer Vision*
- Förstner W, Gülch E (1987) A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. In: *Proceedings of the ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data*, 1987: 281–305.
- Gehrke R, Trabold D, Greiwe A, Abel J (2013) 3D-Modellierung eines Bauwerkes aus RGB- und Thermal-aufnahmen für das Facility Managment. *Photogrammetrie – Laserscanning – Optische 3D-Messtechnik. Beiträge der Oldenburger 3D-Tage 2013*, : 254–263.
- Gibson S, Hubbard R, Cook J, Howard T (2003) Interactive reconstruction of virtual environments from video sequences. *Computers & Graphics*, 27 (2): 293–301.
- Godding R (2006) Camera Calibration. *Handbook of Machine Vision*. 5, : 333–359.
- Grewal M, Andrews A (2008) *Kalman Filtering Theory and Practice*. John Wiley & Sons. ISBN 0470173661.
- Gröger G, Kolbe T, Czerwinski A, Nagel C (2010) *OpenGIS City Geography Markup Language (CityGML) Encoding Standard*, 1.0.0 edition.
- Grün A (1978) Experiences with self-calibrating bundle adjustment. In: *ACSM-ASP Convention, Washington 1978*
- Grün A (1985) Adaptive Least Squares Correlation: A Powerful Image Matching Technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14 (3).
- Haralick R, Joo H, Lee C, Zhuang X, Vaidya V, Kim M (1989) Pose estimation from correspondence point data. *SMC*, 19 (6): 1426–1446.
- Haralick R, Lee C, Ottenberg K, Nölle M (1994) Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem. *International Journal of Computer Vision*, 13 (3): 331–356.
- Harris C, Stephens M (1988) A combined corner and edge detector. In: *Proc. Fourth Alvey Vision Conference*: 147–151.
- Hartley R (1997a) In Defense of the Eight-Point Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (6): 580–593.
- Hartley R (1997b) Lines and Points in three Views and the Trifocal Tensor. *International Journal of Computer Vision*, 22 (2): 125–140.
- Hartley R, Zisserman A (2004) *Multiple View Geometry in Computer Vision*. Cambridge University Press. ISBN 0521540518.
- Hebel M, Stilla U (2009) Automatische Koregistrierung von ALS-Daten aus mehreren Schrägansichten städtischer Quartiere. *PFG Photogrammetrie – Fernerkundung – Geoinformation*, 2009 (3): 161–175.
- Heinrichs M, Hellwich O, Rodehorst V (2008) Robust Spatio-Temporal Feature Tracking. In: *International Archives of Photogrammetry and Remote Sensing*, 37 (B3a): 51–56.
- Hinz S, Stilla U (2006) Car detection in aerial thermal images by local and global evidence accumulation. *Pattern Recognition Letters*, 27 (4): 308–315.
- Hirschmüller H (2008) Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (2): 328–341.
- Hoegner L, Kumke H, Meng L, Stilla U (2007) Automatic extraction of textures from infrared image sequences and database integration for 3D building models. *PFG Photogrammetrie – Fernerkundung – Geoinformation*, 2007 (6): 459–468.

- Hohmann B, Krispel U, Havemann S, Fellner D (2009) Cityfit: High-Quality Urban Reconstruction by Fitting Shape Grammars to Images and Derived Textured Point Clouds. In: *Proceedings of the 3rd ISPRS International Workshop 3D-ARCH 2009*
- Horaud R, Conio B, Le Boulleux O, Lacolle B (1989) An Analytic Solution for the Perspective 4-Point Problem. *Computer Vision, Graphics and Image Processing*, 1989 (47): 33–44.
- Iwaszczuk D, Hoegner L, Schmitt M, Stilla U (2012) Line based matching of uncertain 3D building models with IR image sequences for precise texture extraction. *PFG Photogrammetrie – Fernerkundung – Geoinformation*, 2012 (5): 511–521.
- Kirchhof M, Stilla U (2006) Detection of moving objects in airborne thermal videos. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61: 187–196.
- Klein G, Murray D (2007) Parallel tracking and mapping for small AR workspaces. In: *International Symposium on Mixed and Augmented Reality*
- Klingert M (2005) Methoden der digitalen Bildverarbeitung in der Infrarotbauwerksthermografie. In: *Forum Bauinformatik*, 2005: 287–294.
- Koskeinen L (1992) Predictive maintenance of district heating networks by infrared measurement. In: *Proceedings SPIE*, 1682.
- Kraus K (2004) *Photogrammetrie: Band 1: Geometrische Informationen aus Photographien und Laserscanneraufnahmen*. De Gruyter. ISBN 3110177080.
- Krauss T, Lehner M, Reinartz P, Stilla U (2007) Modelling urban areas coarse and fast from high resolution stereo satellite images. In: *International Archives of Photogrammetry and Remote Sensing*, 36 (3/W49B): 77–82.
- Lagüela S, González-Jorge H, Armesto J, Arias P (2011) Calibration and verification of thermographic cameras for geometric measurements. *Infrared Physics & Technology*, 54: 92–99.
- Läbe T, Förstner W (2006) Automatic relative orientation of images. In: *Proceedings of the 5th Turkish-German Joint Geodetic Days*, 2006.
- Leberl F, Bischof H, Pock T, Irschara A, Kluckner S (2010) Aerial Computer Vision for a 3D Virtual Habitat. *IEEE Computer*, 43 (6): 24–31.
- Lepetit V, Pilet J, Fua P (2004) Point Matching as a Classification Problem for Fast and Robust Object Pose Estimation. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004 (2): 244–250.
- Lindeberg T (1998) Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30 (2): 77–116.
- Lindeberg T, Garding J (1997) Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D structure. *International Journal of Computer Vision*, 29 (15): 415–434.
- Lo C, Quattrochi D (2003) Land-use and land-cover change, urban heat island phenomenon, and health implications: A remote sensing approach. *Photogrammetric engineering and remote sensing*, 69 (9): 1053–1063.
- Longuet-Higgins H (1981) A computer algorithm for reconstruction of a scene from two projections. *Nature*, 203: 133–135.
- Longuet-Higgins H (1986) The Reconstruction of a Plane Surface from Two Perspective Projections. In: *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 227 (1249): 399–410.

- Lowe D (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60 (2): 91–110.
- Luhmann T (2003) *Nahbereichsphotogrammetrie: Grundlagen, Methoden und Anwendungen*. Wichmann. ISBN 3879073988.
- Luhmann T, Ohm J, Piechel J, Roelfs T (2010) Geometric Calibration of Therographic Cameras. *International Archives of Photogrammetry, Remote Sensing and Spacial Information Sciences*, 38 (5): 411–416.
- Maybank S (1993) *Theory of Reconstruction from Image Motion*. Springer Verlag. ISBN 0387555374.
- Mayer H (2007) 3D Reconstruction and Visualization of Urban Scenes from Uncalibrated Wide-Baseline Image Sequences. *PFG Photogrammetrie – Fernerkundung – Geoinformation*, 2007 (3): 167–176.
- Mayer H, Bartelsen J, Hirschmueller H, Kuhn A (2012) Dense 3D Reconstruction from Wide Baseline Image Sets. In: Dellaert F, Frahm JM, Pollefeys M, Leal-Taixe L, Rosenhahn B (eds) *Outdoor and Large-Scale Real-World Scene Analysis. 15th International Workshop on Theoretical Foundations of Computer Vision, Dagstuhl Castle, Germany, June 26 - July 1, 2011*, 7474: 285–304.
- McGlone C (2004) *Manual of Photogrammetry*. ASPRS. ISBN 1570830711.
- Mermelstein MD, Snail KA, Priest RG (2000) Spectral and radiometric calibration of midwave and long-wave infrared cameras. *Optical Engineering*, 39 (2): 347–352.
- Michaelsen E, Iwaszczuk D, Sirmacek B, Hoegner L, Stilla U (2012) Gestalt grouping on façade textures from IR image sequences: Comparing different production systemse. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 39 (B3): 303–308.
- Mikhail E, Bethel J, McGlone J (2001) *Introduction to Modern Photogrammetry*. John Wiley & Sons, New York.
- Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27: 1615–1630.
- Müller P, Wonka P, Haegler S, Ulmer A, Gool LV (2006) Procedural Modeling of Buildings. In: *ACM Transactions on Graphics*, 25 (3): 614–623.
- Moravec H (1977) Towards Automatic Visual Obstacle Avoidance. In: *Proceedings of the 5th International Joint Conference on Artificial Intelligence*: 584.
- Nagel C, Stadler A, Kolbe T (2009) Conceptual Requirements for the Automatic Reconstruction of Building Information Models from Uninterpreted 3D Models. In: *International Archives of Photogrammetry and Remote Sensing*, 34 (30).
- Newcombe R, Lovegrove S, Davison A (2011) DTAM: Dense tracking and mapping in real-time. In: *Proceedings of the International Conference on Computer Vision*
- Nicol J, Wong M (2005) Modeling urban environmental quality in a tropical city. *Landscape and Urban Planning*, 73.
- Nistér D (2000) Reconstruction From Uncalibrated Sequences with a Hierarchy of Trifocal Tensors. In: *Proceedings of the European Conference on Computer Vision*, 2000 (1): 649–663.
- Nistér D (2004) An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (6): 756–777.
- Nistér D, Naroditsky O, Bergen J (2006) Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23 (1): 3–20.

- Planck M (1923) *Die Ableitung des Strahlungsgesetzes, sieben Abhandlungen aus dem Gebiete der elektromagnetischen Strahlungstheorie*. Oswald's Klassiker der exakten Wissenschaften. Akademische Verlagsgesellschaft.
- Pollefeys M, Nistér D, Frahm J, Akbarzadeh A, Mordohai P, Clipp B, Engels C, Gallup D, Kim S, Merrell P, Salmi C, Sinha S, Talton B, Wang L, Yang Q, Stewénus H, Yang R, Welch G, Towles H (2008) Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision*, 78 (2-3): 143–167.
- Pollefeys M, Verbiest F, Gool LV (2002) Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery. In: *Proceedings of the ECCV 2002. Part II*, 2351: 837–851.
- Quan L, Lan L (1999) Linear n-point camera pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 (8): 774–780.
- Quattrochi D, Luvall J (1999) Thermal infrared remote sensing for analysis of landscape ecological processes: methods and application. *Landscape Ecology*, 14 (6): 577–598.
- Reznik S, Mayer H (2008) Implicit Shape Models, Self-Diagnosis, and Model Selection for 3D Facade Interpretation. *PFG Photogrammetrie – Fernerkundung – Geoinformation*, 2008 (3): 187–196.
- Ripperda N (2009) *Rekonstruktion von Fassadenstrukturen mittels formaler Grammatiken und Reversible Jump Markov Chain Monte Carlo Sampling*. Ph.D. thesis Leibniz Universität Hannover.
- Rogatto W (1993) Electro-Optical Components. In: Accetta J, Shumaker D (eds) *The Infrared and Electro-Optical Systems Handbook*. Bellingham, Washington (USA): ERIM Infrared Information Analysis Center, Ann Arbor, Michigan (USA), and SPIE Optical Engineering Press.
- Rottensteiner F, Trinder J, Clode S, Kubrik K (2005) Automated Delineation of Roof Planes from LIDAR Data. In: *International Archives of Photogrammetry and Remote Sensing*, 36 (3/W19): 221–226.
- Schaffalitzky F, Zissermann A (2002) Multi-view Matching for Unordered Image Sets, or How Do I Organize My Holiday Snaps? In: *Proceedings of the ECCV 2002. Part I*, 2350: 414–431.
- Shi J, Tomasi C (1994) Good Features to Track. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994: 593–600.
- Siegert F, Zhukov B, Oertel D, Limin S, Page S, Rielay O (2004) Peat fires detected by the BIRD satellite. *International Journal of Remote Sensing*, 25 (16): 3221–3230.
- Simmler C (2009) *Entwicklung einer Messanordnung zur geometrischen Kalibrierung von Infrarot-Kameras*. Bachelorarbeit der Technischen Universität München.
- Simon G, Fitzgibbon A, Zissermann A (2000) Markerless Tracking using Planar Structures in the Scene. In: *Proceedings of the International Symposium on Augmented Reality*, 2000: 120–128.
- Sithole G, Vosselman G (2005) Filtering of airborne laser scanner data based on segmented point clouds. In: *International Archives of Photogrammetry and Remote Sensing*, 36 (3/W19): 66–71.
- Sorenson H (1970) Least-Squares estimation: from Gauss to Kalman. *IEEE Spectrum*, 7: 63–68.
- Steger C, Ulrich M, Wiedemann C (2007) *Machine Vision Algorithms and Applications*. Weinheim: Wiley-VCH.
- Stewenius H, Schaffalitzky F, Nistér D (2005) How hard is 3-view triangulation really? In: *Proceedings of the Tenth IEEE International Conference on Computer Vision*, 2005 (1): 686–693.
- Stilla U, Kolecki J, Hoegner L (2009) Texture mapping of 3D building models with oblique direct georeferenced airborne IRimage sequences. In: *International Archives of Photogrammetry, Remote Sensing and Geospatial Sciences*, 38 (1-4-7/W5).

- Stilla U, Michaelsen E (2002) Detektion von Fahrzeugen in Videosequenzen von luftgetragenen thermischen Kameras. *Österreichische Zeitschrift für Vermessung und Geoinformation VGI*, 90 (3+4): 79–86.
- Stöbl W (1993) *Fourier Optik. Eine Einführung*. Springer-Verlag. ISBN 3540532870.
- Strecha C, von Hansen W, van Gool L, Fua P, Thoennessen U (2008) On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery. *Computer Vision and Pattern Recognition*, : 1–8.
- Torr P (1997) An Assessment of Information Criteria for Motion Model Selection. *Computer Vision and Pattern Recognition*, : 47–53.
- Torr P, Murray D (1997) The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix. *International Journal of Computer Vision*, 24 (3): 271–300.
- Triggs B (1999) Camera Pose and Calibration from 4 or 5 known 3D Points. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999 (1): 278–284.
- Tuttas S, Stilla U (2013) Reconstruction of façades in point clouds from multi aspect oblique ALS. *ISPRS Annals*, II-3/W3: 91–96.
- Vincent J (1990) *Fundamentals of Infrared Detector Operation and Testing*. John Wiley & Sons. ISBN 0471502723.
- Wien W (1896) Über die Energievertheilung im Emissionsspectrum eines schwarzen Körpers. *Annalen der Physik*, 294 (8): 662–669.
- Yang R, Pollefeys M (2003) Multi-resolution real-time stereo on commodity graphics hardware. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003: 211–217.
- Zhang Z (1998) Determining the Epipolar Geometry and its Uncertainty: a Review. *International Journal of Computer Vision*, 17 (2): 161–195.
- Zhi L, Tang J (2002) A Complete Linear 4-Point Algorithm for Camera Pose Determinatio. *MM Research Reprints*, 21: 239–249.
- Zissler G (1993) Sources of Radiation. In: Accetta J, Shumaker D (eds) *The Infrared and Electro-Optical Systems Handbook*. Bellingham, Washington (USA): ERIM Infrared Informaion Analysis Center, Ann Arbor, Michigan (USA), and SPIE Optical Engineering Press.

Appendices

Anhang A

Bildsequenzen



Abbildung A.1: Auszug Sequenz Ost: Aufnahme in Fahrtrichtung, Uhrzeit 22:30 Uhr

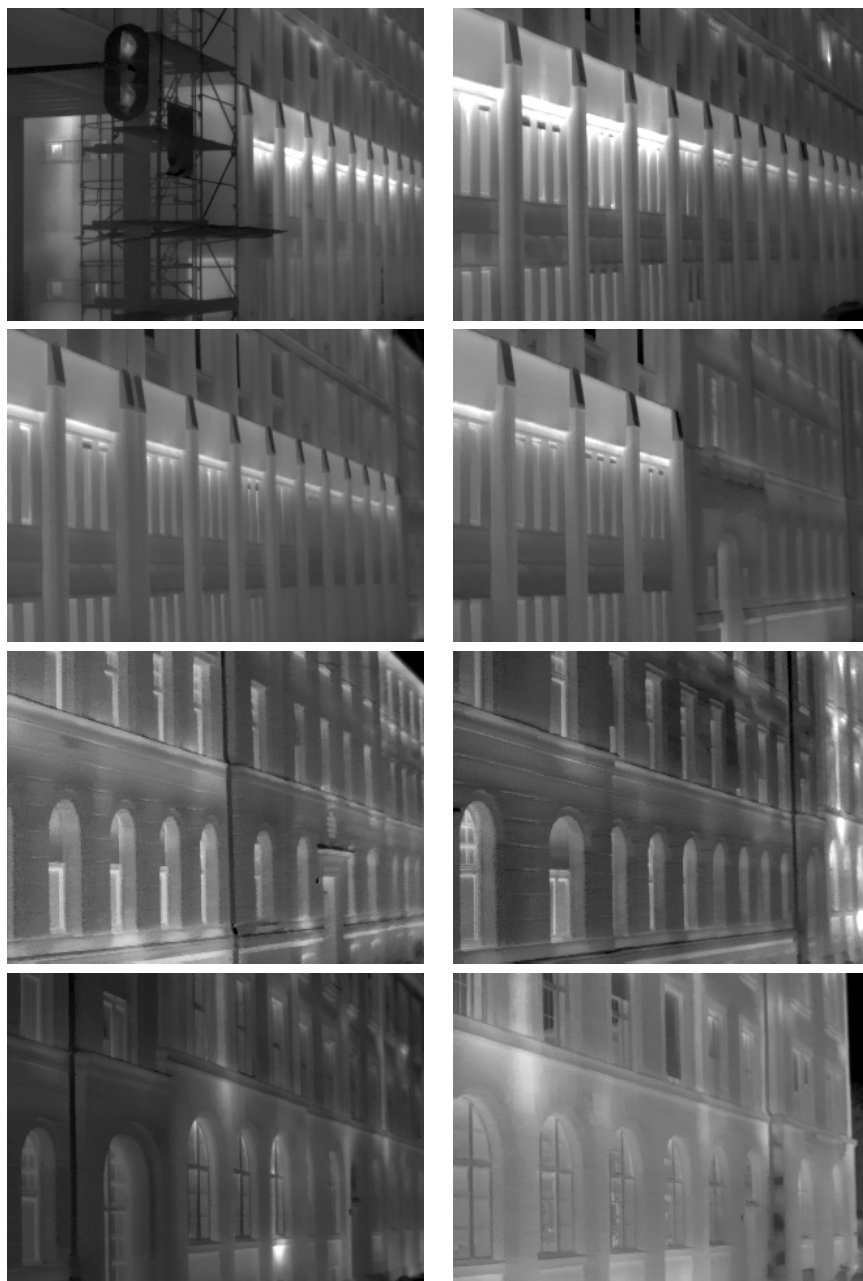


Abbildung A.2: Auszug Sequenz West: Aufnahme in Fahrtrichtung, Uhrzeit 22:00 Uhr

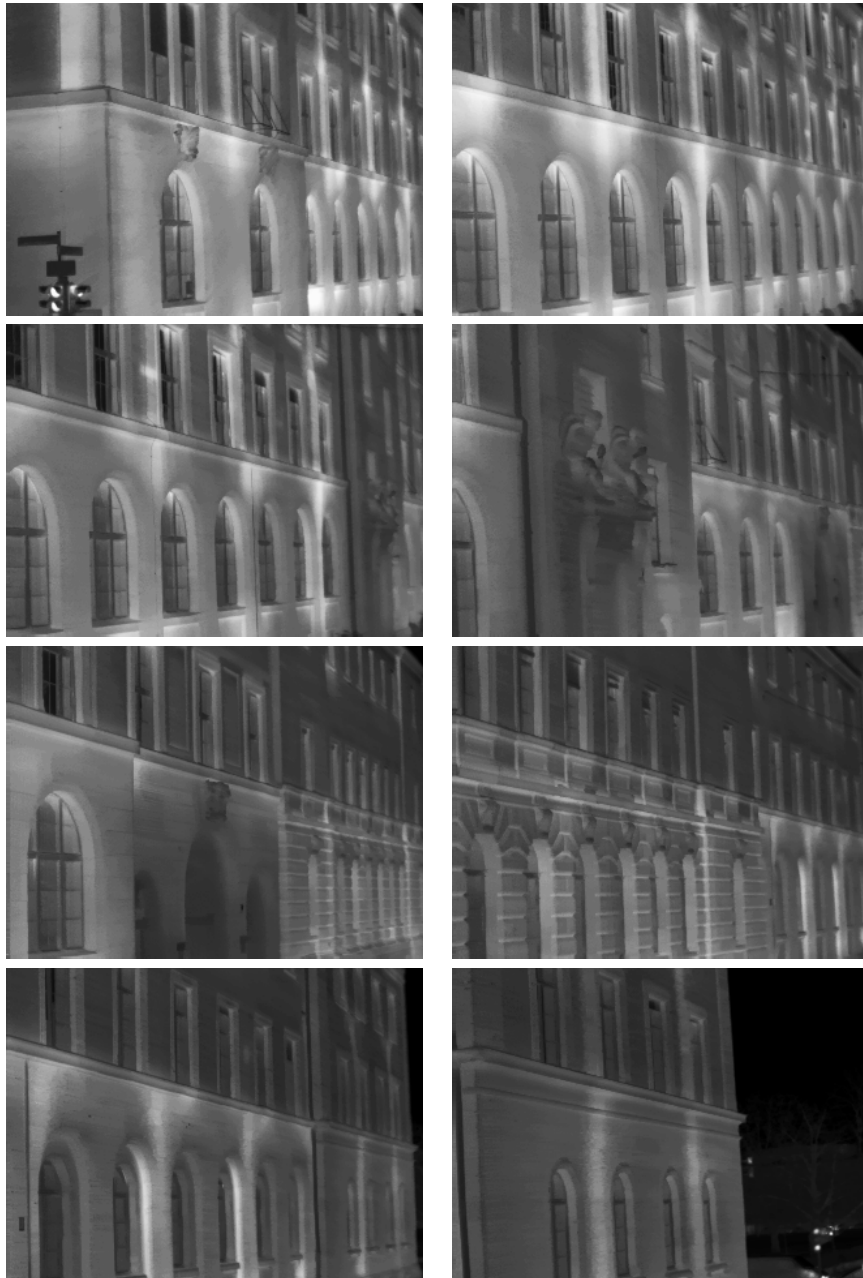


Abbildung A.3: Auszug Sequenz Sued: Aufnahme in Fahrtrichtung, Uhrzeit 22:15 Uhr



Abbildung A.4: Auszug Sequenz Nord: Aufnahme in Fahrtrichtung, Uhrzeit 22:45 Uhr

Lebenslauf

Name	Ludwig Hoegner
Geburtstag und -ort	31.08.1979 in München
Wohnort	Am Hartmannshofer Bächl 11 80997 München



Ausbildung/Tätigkeit

1986 – 1990	Grundschule an der Rotbuchenstraße, München
1990 – 1999	Städtisches Theodolindengymnasium, München
1999 – 2000	Zivildienst Arbeiterwohlfahrt München-Stadt e.V.
2000 – 2005	Studium der Informatik und Elektrotechnik an der Technischen Universität München
2006 – 2010	Wissenschaftlicher Mitarbeiter am Fachgebiet Photogrammetrie und Fernerkundung der Technischen Universität München
seit 2010	Akademischer Rat a.Z. am Fachgebiet Photogrammetrie und Fernerkundung der Technischen Universität München

