Karlsruher Institut für Technologie

Institut für Mess- und Regelungstechnik

Automatische Berechnung von Referenzmessungen für Target Tracking Systeme mittels stereoskopischer Bildverarbeitung

Master-Arbeit
von

Piotr Orzechowski

Institut für Mess- und Regelungstechnik
Karlsruher Institut für Technologie

Erstgutachter:	Prof. Dr.-Ing. Christoph Stiller
Zweitgutachter:	Prof. Dr.-Ing. Jürgen Beyerer
Betreuender Mitarbeiter:	Dipl.-Ing. Eike Rehder

Datum: 21. Dezember 2015

Kurzfassung

Bei der Entwicklung von hochautomatisierten Fahrerassistenzsystemen steigen die Ansprüche an die Umfeldwahrnehmung enorm, da nur die zuverlässige Objekterkennung und -verfolgung eine umfassende Situationsanalyse ermöglicht. Aktuelle Forschungsarbeiten sind jedoch auf einen echtzeitfähigen Online-Filteransatz im Fahrzeug ausgelegt und stoßen dabei an ihre Grenzen. In dieser Arbeit wird ein akausales Offline-Verfahren entwickelt, welches aus hochauflösenden Stereobildern einer bewegten Referenzkamera eine bestmögliche Schätzung der Bewegung und Struktur der zu verfolgenden Objekte automatisch berechnet. Hierzu werden aus den Ausreißerlandmarken einer Eigenbewegungsschätzung zunächst Hypothesen kleinster Objektsegmente erzeugt. Diese Segmenthypothesen werden unter der Annahme starrer Objekte mithilfe des Bündelausgleichsverfahrens und eigens entwickelten Metriken auf Plausibilität geprüft. Segmente, die diesen Test bestehen, setzen sich mit hoher Wahrscheinlichkeit ausschließlich aus Landmarken desselben Objekts zusammen. Diese Landmarken werden daraufhin basierend auf der Überlappung der verbliebenen Segmente zu Objekten aggregiert. Anschließend wird die Relativbewegung und Struktur der Objekte jeweils, erneut unter Verwendung des Bündelausgleichs, bestmöglich rekonstruiert. Zuletzt wird die Oberfläche der Objekte mittels Alpha Shapes approximiert. Die Leistungsfähigkeit des Verfahrens wird anhand umfangreicher Testdaten demonstriert.

Abstract

In the realm of developing advanced driver assistance systems the environment perception requirements are increasing tremendously as comprehensive situation analysis depends on reliable object detection and tracking. However, current research on this topic is focused on real-time capable online filters for the use within the car and is now reaching its limit. This thesis presents an acausal offline method to automatically compute the best possible estimate of the motion and structure of moving obstacles using high-resolution stereo images of a moving reference camera. For this purpose, outlier landmarks of a self-localization are used to create hypotheses of smallest object segments. These hypotheses are tested for plausibility under the assumption of rigid objects using the bundle adjustment method and metrics specifically developed for this task. Segments passing the test consist most likely of landmarks of one single object. Their landmarks are then being aggregated to objects based on the overlapping of these passing segments. Afterwards, the relative movement and 3D structure is estimated for each object using bundle adjustment again. In a final step the objects' surfaces are approximated with alpha shapes. The performance of the proposed method is shown on extensive data sets.

Eigenständigkeitserklärung

Ich versichere hiermit wahrheitsgemäß, die Arbeit bis auf die dem Aufgabensteller bereits bekannten Hilfsmittel selbständig angefertigt, alle benutzten Hilfsmittel vollständig angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, den 21. Dezember 2015

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einleitung

1.1 Motivation

Spätestens seit den medienwirksamen Auftritten von Googles automatisch fahrenden Versuchsfahrzeugen [25][32] ist autonomes Fahren nach öffentlicher Wahrnehmung bereits in greifbarer Nähe. Tatsächlich forschen alle großen Automobilhersteller, unter anderem Daimler, BMW und Audi, sowie ihre Zulieferer an selbst fahrenden Fahrzeugen [17]. Auch der vergleichsweise junge Hersteller Tesla [23][27] und womöglich Apple [37] beschäftigen sich mit der Thematik.

Vollautomatisierte Systeme, die die vielfältigen Herausforderungen des Stadtverkehrs, auf Landstraßen und Autobahnen bei jeder Witterung, Tageszeit und Verkehrsdichte meistern, werden allerdings erst in den kommenden Jahrzehnten Serienreife erlangen. Bis dahin finden die hierfür notwendigen Technologien aber bereits heute als Fahrerassistenzsysteme schrittweise Einzug in den Alltag.

Darunter fallen zum einen lang bewährte Assistenzsysteme auf Stabilisierungsebene, wie das Antiblockiersystem (ABS) und das Elektronische Stabilitätsprogramm (ESP). Zum anderen sind mittlerweile zahlreiche sogenannte „fortgeschrittene Fahrerassistenzsysteme“ (engl. „advanced driver assistance systems“) auf Bahnführungs- und Navigationsebene in Serienfahrzeugen anzutreffen. Insbesondere sind hier Einparkassistenten, die adaptive Fahrgeschwindigkeitsregelung (engl. „adaptive cruise control“, kurz ACC) und Querführungsassistenten (bspw. „Lane Assist“ von VW) zu nennen.

Motiviert werden die Entwicklungen insbesondere über den erhofften Sicherheitsgewinn. Allein in Deutschland verunglückten im Jahr 2014 392.912 Personen an Verkehrsunfällen, davon 3.377 mit tödlichen Folgen. Im Jahre 1970 hatte Deutschland über $5\frac{1}{2}$ mal so viele und 1990 noch mehr als doppelt so viele Verkehrstote zu beklagen, trotz der seither steigenden Anzahl an jährlichen Kfz-Neuzulassungen [10]. Dieser dramatische Rückgang an Todesfällen ist unter anderem auf die aktiven Sicherheitsfunktionen ABS, ESP, ACC und weitere fortgeschrittene Fahrerassistenzsysteme zurückzuführen.

Doch auch der ökonomische Gesichtspunkt automatisierter Fahrzeuge ist nicht zu vernachlässigen. Insbesondere in der Transportbranche erhofft man sich eine Reduzierung der „total

cost of ownership“ (u.A. durch Kraftstoffeinsparung und Senkung der Wartungskosten), eine effizientere Nutzung der vorhandenen Infrastruktur und auf lange Sicht sicherlich auch die Einsparung von Personalkosten. Vor kurzem erst hat Daimler eine bundesweite Zulassung seines neuesten teil-autonomen LKWs für die Erprobung des sogenannten „Highway Piloten“ erhalten [1].

Fortgeschrittene Fahrerassistenzsysteme beruhen in der Regel auf der Auswertung vielfältiger Sensorik, darunter Ultraschall-, Radar, LIDAR- und Kamerasensorik. Letztere ist besonders beliebt, da Kameras klein, günstig und dennoch hochwertige Sensoren sind, deren Messungen (also Bilder) zudem vom Menschen intuitiv nachvollziehbar sind.

Beim Übergang von teil- zu hochautomatisierten Systemen steigen die Ansprüche an die Umfeldwahrnehmung enorm. Neben der Erkennung von Fahrspur und Verkehrsführung ist die Wahrnehmung anderer Verkehrsteilnehmer von wesentlicher Bedeutung. Nur die zuverlässige Objekterkennung und -verfolgung ermöglicht eine umfassende Situationsanalyse.

Heutige Erprobungsmethoden für solche Target Tracking Systeme stoßen jedoch an ihre Grenzen. In dieser Arbeit wurde daher ein Verfahren entwickelt und evaluiert, welches aus hochauflösenden Stereobildern eine Referenzmessung der zu verfolgenden Fahrzeuge automatisch berechnet. Der Fokus lag hierbei auf der offline Bildauswertung, so dass Messungen und Detektionen aller Zeitschritte gemeinsam ausgewertet und eine global optimale Schätzlösung geliefert wird.

1.2 Einordnung der Arbeit

Aktuelle Forschungsarbeiten zur Objektverfolgung sind stark auf einen Online-Filteransatz im Fahrzeug zugeschnitten und können systembedingt keine zukünftigen Messungen zur aktuellen Zustandsschätzung verwenden. Des Weiteren erlaubt eine Realzeitumsetzung, die in aktuellen Arbeiten angestrebt wird, nur die Verwendung von Schätzverfahren niedriger Komplexität (z.B. rekursive Bayes'sche Schätzer wie das Kalman-Filter).

Sivaraman und Trivedi [31] geben einen ausführlichen Überblick über den Stand der Technik seit 2005, weswegen hier auf eine wiederholte Zusammenfassung verzichtet wird. Zu Forschungsergebnissen vor 2005 sei zudem auf Sun, Bebis und Miller [34] verwiesen. Nichts desto trotz sollen die Veröffentlichungen, die diese Arbeit beeinflusst haben, nicht unerwähnt bleiben.

Allen voran sei die Arbeit von Lenz u. a. [22] genannt, die eine Methode zur Objekterkennung mit einer Stereokamera mithilfe eines dünn besetzten Szenenflusses (engl. sparse scene flow) beschreibt. Das Verfahren extrahiert und assoziiert aus den Bildsequenzen charakteristische Landmarken, deren Beobachtungen sich nicht durch die Eigenbewegung

der Kamera erklären lassen (sog. Ausreißer) und somit mit hoher Wahrscheinlichkeit zu bewegten Objekten gehören oder Rauschen darstellen. Diese Landmarken werden über eine Delaunay-Triangulation mit im Bildraum benachbarten Landmarken zu einem Graphen verbunden, dessen Knoten jeweils eine Landmarke repräsentieren. Dabei werden nur solche Knoten mit einer Kante verbunden, deren zwei Landmarken einen ähnlichen Szenenfluss aufweisen. Somit können die Landmarken, durch eine anschließende Bestimmung von Zusammenhangskomponenten im Graphen, einzelnen Objekten zugeordnet werden.

Kitt, Ranft und Lategahn [21] verwenden ebenfalls dünn besetzte Ausreißer der Eigenbewegungsschätzung, gruppieren die Landmarken jedoch mithilfe eines hierarchischen Clusteringverfahrens anhand ihrer rekonstruierten 3D Position zu Objekthypothesen. Anschließend wird die Bewegung jedes detektierten Objekts mittels erweitertem Kalman-Filter geschätzt.

Der Ansatz von Barth und Franke [6] beruht wie [22] und [21] lediglich auf der Annahme starrer Körper und arbeitet ebenso mit dünn besetzten Ausreißern der Eigenbewegungsschätzung. Allerdings schätzen sie zusätzlich zur Objektposition und -bewegung die Form der detektierten Objekte. Hierzu rekonstruierten sie objektfeste 3D Punktewolken der jeweils assoziierten Landmarken in einem weiteren Filter.

1.3 Zielsetzung

Wie bereits erwähnt, sind alle im letzten Abschnitt genannten Arbeiten für einen Online-Filteransatz im Fahrzeug ausgelegt, so dass sie den damit verbundenen Nachteilen unterliegen. Bei der Erprobung von Target Tracking Systemen sind allerdings Anforderungen wie Echtzeitfähigkeit und Kausalität nicht zwingend gegeben. Die Messungen beziehungsweise Resultate der zu erprobenden Systeme können zunächst mit allen zugehörigen Sensordaten aufgezeichnet und erst in einem Offline-Schritt auf leistungsfähiger Hardware evaluiert werden. Daher soll der Stand der Technik in dieser Arbeit durch das Aufheben beider Beschränkungen erweitert werden.

Das Ziel dieser Arbeit war also ein akausales Offline-Verfahren zu entwickeln, um

- bewegte Objekte in Stereobildern einer wiederum bewegten Kamera zu erkennen,
- diese zeitlich zu tracken und
- ihre Relativbewegung und Struktur bestmöglich zu rekonstruieren.

Das Problem kann also als „simultaneous tracking and reconstruction“ (STAR) bezeichnet werden. Als Eingangsdaten dienen assoziierte Beobachtungen dünn besetzter Ausreißer der Eigenbewegungsschätzung. Das Verfahren darf dabei auf der Annahme starrer Körper

beruhen, jedoch nicht auf einer Klassifikation der Verkehrsteilnehmer auf Grund gelernter Modelle, um eine möglichst weite Bandbreite auftretender Objekte behandeln zu können.

Diese Masterarbeit wurde im Rahmen einer Kooperation des Instituts für Mess- und Regelungstechnik des Karlsruher Instituts für Technologie und der Atlatec GmbH angefertigt. Die Atlatec GmbH ist ein 2014 gegründetes Start-up-Unternehmen, das sich mit der Kartierung und Lokalisierung mittels Kamerasystemen beschäftigt. Sie bietet ihren Kunden aus den Entwicklungsabteilungen der Automobilbranche Lösungen zur Erprobung ihrer Fahrerassistenzsysteme an.

1.4 Übersicht

Zu Beginn soll eine kurze Übersicht über die Gliederung dieser Arbeit gegeben werden.

Kapitel 2 führt in die notwendigen theoretischen Grundlagen, wie Schätztheorie, Abbildungsgeometrie, Struktur- und Bewegungsrekonstruktion und das zu Grunde liegende Landmarkentracking, ein.

In Kapitel 3 wird das entwickelte Verfahren im Detail vorgestellt. Dieses setzt sich aus den zwei unabhängigen Verarbeitungsschritten, der Objektaggregation und Objektrekonstruktion, zusammen.

Kapitel 4 stellt die durchgeführten Experimente und deren Ergebnisse anhand zahlreicher Grafiken und Bilder vor.

Kapitel 5 schließt die Arbeit mit einer Zusammenfassung und einem Ausblick auf mögliche Weiterentwicklungen ab.

2 Grundlagen

Bevor im nachfolgenden Kapitel die erarbeitete Methodik vorgestellt wird, sollen die hierfür notwendigen theoretischen Grundlagen zusammengefasst werden.

Zunächst stellt Abschnitt 2.1 die Methode der kleinsten Fehlerquadrate vor, welches eines der fundamentalen Schätzverfahren und damit das wichtigste mathematische Werkzeug der folgenden Verfahren bildet.

In Abschnitt 2.2 werden Grundlagen der Abbildungsgeometrie, insbesondere das Lochkameramodell, eingeführt.

Daraufhin wird in Abschnitt 2.3 das sogenannte Bündelausgleichsverfahren zur Struktur- und Bewegungsrekonstruktion aus einer Vielzahl bewegter Kamerabilder vorgestellt.

Abschnitt 2.4 erläutert letztlich summarisch das Landmarkentracking samt Eigenbewegungsschätzung, das insbesondere die Eingangsdaten der erarbeiteten Methodik in Kapitel 3 prägt.

2.1 Schätztheorie

Die mathematischen Herleitungen dieses Abschnittes orientieren sich an Nocedal und Wright [28]. Zusätzliche Quellen werden an gegebener Stelle erwähnt.

2.1.1 Least-Squares Verfahren

Optimierungsprobleme lassen sich in der Regel als Minimierungsprobleme formulieren:

$$\min_{\mathbf{x}} \; f(\mathbf{x}). \tag{2.1}$$

Es sei $\mathbf{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ eine stetige Abbildung und $\mathbf{x}^*$ die Lösung dieses Problems. Eine häufige Anwendung ist die Parameterschätzung eines mathematischen Modells $y = \phi(t; \mathbf{x})$, mit Modellparametern $\mathbf{x}$, zur Approximation von Messdaten (t_j, y_j) mit $j = 1 \ldots m$. Dabei wird der Schätzfehler, also die Abweichung des Modells von den Messdaten, minimiert. Die

Wahl von $t_j, y_j \in \mathbb{R}$ erlaubt im folgenden saubere Herleitungen, welche allerdings auch auf mehrdimensionale Optimierungsprobleme übertragen werden können.

Unter den zahlreichen numerischen Optimierungsmethoden ist das Verfahren der kleinsten Fehlerquadrate (engl. „least squares“) besonders verbreitet, da es bei unkorreliertem und normalverteiltem Fehlerverhalten einen Maximum-Likelihood Schätzer bildet. Bei diesem Verfahren wird die Summe der Fehlerquadrate

$$f(\mathbf{x}) = \frac{1}{2}\sum_{j=1}^{m} r_j^2(\mathbf{x}) \tag{2.2}$$

minimiert. $r_j(\mathbf{x}) = y_j - \phi(t_j; \mathbf{x})$ wird hierbei als Residuum bezeichnet und $m \geq n$ angenommen. Fasst man die Residuen r_j zu einem Residuenvektor $\mathbf{r} : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{m}}$,

$$\mathbf{r}(\mathbf{x}) = (r_1(\mathbf{x}), r_2(\mathbf{x}), \ldots, r_m(\mathbf{x}))^T, \tag{2.3}$$

zusammen, kann das Fehlerfunktional f als Euklid'sche Norm geschrieben werden:

$$f = \frac{1}{2}\|\mathbf{r}(\mathbf{x})\|_2^2. \tag{2.4}$$

Mithilfe dieser Formulierung lässt sich der Gradient von $f(\mathbf{x})$ mit der Jacobimatrix von $\mathbf{r}$

$$\mathbf{J}(\mathbf{x}) = \left(\frac{\partial r_j}{\partial x_i}\right)_{\substack{j=1\ldots m \\ i=1\ldots n}} \tag{2.5}$$

ausdrücken als

$$\nabla f(\mathbf{x}) = \sum_{j=1}^{m} r_j(\mathbf{x})\nabla r_j(\mathbf{x}) = \mathbf{J}(\mathbf{x})^T\mathbf{r}(\mathbf{x}). \tag{2.6}$$

Daraus ergibt sich folgende Hessematrix

$$\nabla^2 f(\mathbf{x}) = \sum_{j=1}^{m} \nabla r_j(\mathbf{x})\nabla r_j(\mathbf{x})^T + \sum_{j=1}^{m} r_j(\mathbf{x})\nabla^2 r_j(\mathbf{x}) \tag{2.7a}$$

$$= \mathbf{J}(\mathbf{x})^T\mathbf{J}(\mathbf{x}) + \sum_{j=1}^{m} r_j(\mathbf{x})\nabla^2 r_j(\mathbf{x}). \tag{2.7b}$$

Für den Fall linearer Residuen r_j, reduziert sich das Minimierungsproblem (2.2) auf die Lösung der Normalengleichung

$$\mathbf{J}^T\mathbf{J}\mathbf{x}^* = -\mathbf{J}^T\mathbf{r}. \tag{2.8}$$

Diese lässt sich durch die Cholesky Zerlegung von $\mathbf{J}^T\mathbf{J}$, die QR Zerlegung von $\mathbf{J}$ oder die Singulärwertzerlegung von $\mathbf{J}$ lösen. Letztere ist die robusteste, allerdings auch aufwändigste, unter den drei Methoden. Für eine ausführlichere Diskussion ihrer Vor- und Nachteile sei der Leser auf Nocedal und Wright [28] verwiesen.

2.1.2 Nichtlineare Least-Squares Verfahren

Nicht-lineare Optimierungsprobleme werden in der Regel durch iterative Verfahren gelöst. Dabei gibt es zwei grundlegende Strategien, um von $\mathbf{x}_k$ zu $\mathbf{x}_{k+1}$ zu iterieren:

Line-search Als „line-search“ wird die Strategie bezeichnet, zunächst eine Richtung $\mathbf{q}_k$ zu wählen und anschließend ausgehend von $\mathbf{x}_k$ entlang dieser Richtung ein $\mathbf{x}_{k+1}$ mit niedrigerem Funktionswert $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ zu suchen. Für die Konvergenz des Verfahrens ist hierbei entscheidend, eine geeignete Schrittweite α_k zu wählen:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{q}_k. \tag{2.9}$$

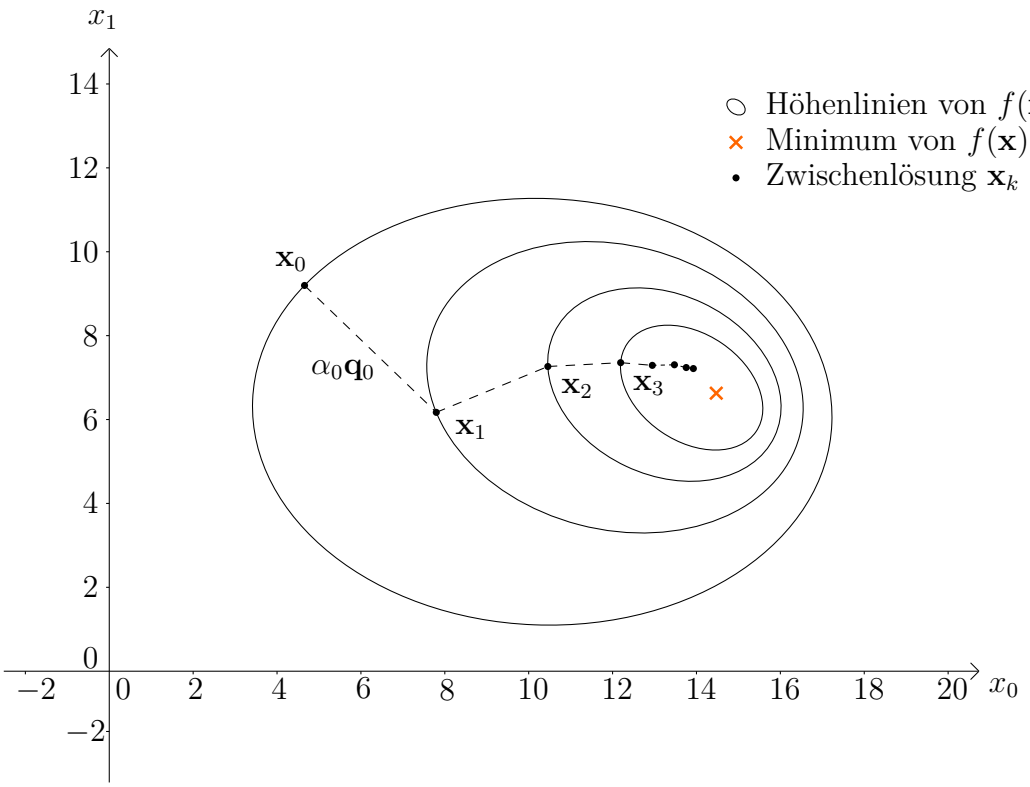
Die line-search Strategie wird in Abbildung 2.1(a) am Beispiel eines Gradientenabstiegs veranschaulicht. Hier ist zu erkennen, dass die Schrittweite zu Beginn groß und dann schrittweise kleiner gewählt wird, um eine schnelle Konvergenzrate zu erreichen aber gleichzeitig eine Oszillation um das Minimum herum zu vermeiden. Ein reiner Gradientenabstieg kann dennoch schlechte Konvergenzraten aufweisen, weshalb später die effizientere Newton-Methode vorgestellt wird.

Trust-region Bei der „trust-region“ Strategie wird die zu minimierende Funktion f in einer Umgebung von $\mathbf{x}_k$ durch ein Modell m_k angenähert. $\mathbf{x}_{k+1}$ wird nun als Minimum des Modells innerhalb der gewählten Umgebung gesetzt:

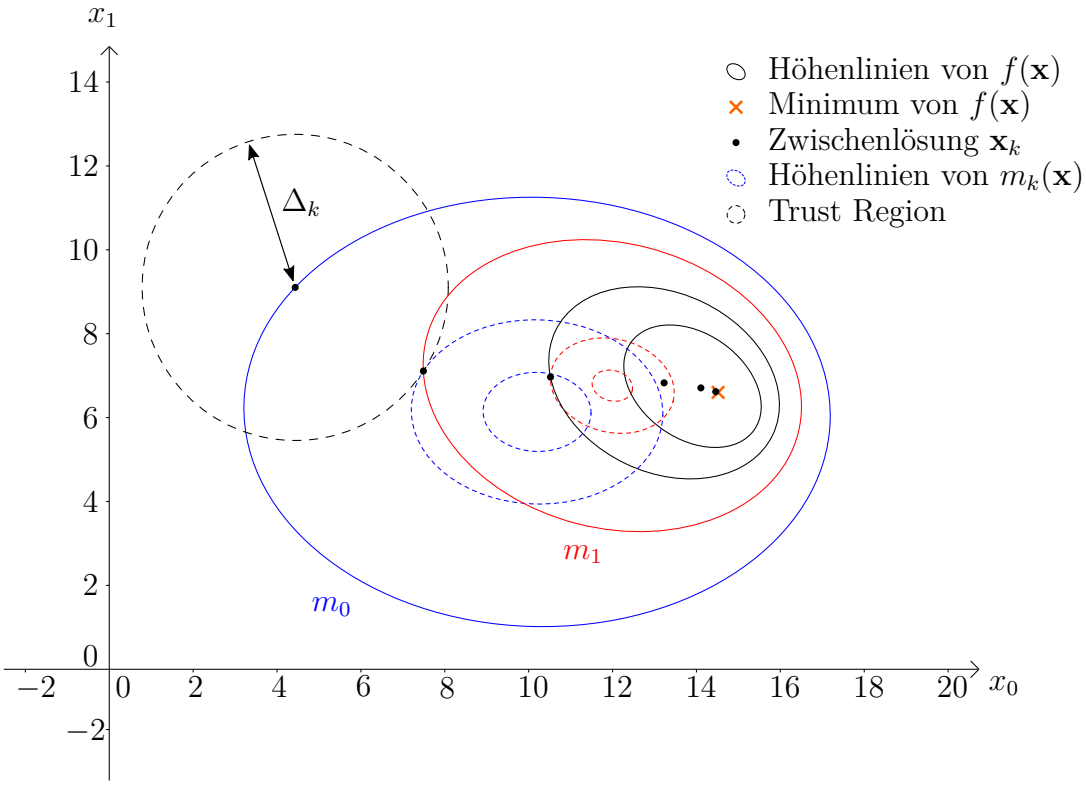
$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{q}_k; \quad \mathbf{q}_k = \arg\min_{\mathbf{q}} \left\{ m_k(\mathbf{x}_k + \mathbf{q}) \right\}. \tag{2.10}$$

Die sogenannte „trust-region“ um $\mathbf{x}_k$ wird üblicherweise durch einen Radius Δ_k festgelegt, so dass $\|\mathbf{q}_k\|_2 \leq \Delta_k$ gelten soll. Für m_k wird häufig eine quadratische Funktion verwendet, die sich aus dem Gradienten und der Hessematrix von f leicht bestimmen und minimieren lässt. Abbildung 2.1(b) veranschaulicht die Strategie.

Beide Strategien haben gemein, dass die Wahl des Startwertes $\mathbf{x}_0$ die Konvergenzrate stark beeinflussen kann, sowie bei Vorhandensein mehrerer Minima auch bestimmt, welches lokale Minimum gefunden wird.



(a) Line-search



(b) Trust-region

Abbildung 2.1 – Iterationsstrategien im Vergleich

2.1.2.1 Gauß-Newton

Die intuitive Schrittrichtung für iterative Minimierungsverfahren, ist die des negativen Gradienten

$$\mathbf{q}_k = -\nabla f_k, \tag{2.11}$$

wobei von nun an die Kurzschreibweise $f_k = f(\mathbf{x}_k)$ verwendet wird. Wie bereits erwähnt, ist die Newton-Methode mit

$$\mathbf{q}_k = -\nabla^2 f_k^{-1} \nabla f_k \tag{2.12}$$

dem reinen Gradientenabstieg auf Grund deutlich besserer Konvergenzraten zu bevorzugen. Ist die Hessematrix $\nabla^2 f(\mathbf{x})$ positiv definit und die Schrittlänge $\alpha = 1$, dann konvergiert die Newton-Methode quadratisch. In vielen Fällen ist die Bestimmung der Hessematrix aber zu rechenaufwendig, so dass sogenannte Quasi-Newton-Methoden diese nur approximieren.

Das Gauß-Newton Verfahren zur Lösung nicht-linearer Least-Squares Probleme ist nun eine solche Quasi-Newton-Methode mit line-search Strategie. Ausgehend von Gleichung (2.7) wird die Hessematrix approximiert durch

$$\nabla^2 f_k \approx \mathbf{J}_k^T \mathbf{J}_k, \tag{2.13}$$

womit sich die Iterationsrichtung zu

$$\mathbf{J}_k^T \mathbf{J}_k \mathbf{q}_k = -\mathbf{J}_k^T \mathbf{r}_k \tag{2.14}$$

ergibt. Diese kann im Hinblick auf Gleichung (2.8) als Lösung des linearen Least-Squares Problems

$$\min_{\mathbf{q}} \ \tfrac{1}{2} \left\| \mathbf{J}_k \mathbf{q} + \mathbf{r}_k \right\|_2^2 \tag{2.15}$$

gesehen und mithilfe der bereits erwähnten Methoden der QR- oder SVD Zerlegung gelöst werden.

Die Schrittweite α_k kann nun, wie der Name „line-search“ andeutet, experimentell angepasst

werden, bis $f(\mathbf{x}_k + \alpha_k \mathbf{q}_k)$ hinreichend kleiner ist als $f(\mathbf{x}_k)$. Die Wolfe Ungleichungen

$$f(\mathbf{x}_k + \alpha_k \mathbf{q}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f_k^T \mathbf{q}_k \tag{2.16a}$$

$$\nabla f(\mathbf{x}_k + \alpha_k \mathbf{q}_k)^T \mathbf{q}_k \geq c_2 \nabla f_k^T \mathbf{q}_k \tag{2.16b}$$

mit $0 < c_1 < c_2 < 1$ können dabei behilflich sein, eine gute Konvergenzrate zu erreichen. Hierfür legt (2.16a) die Bedingung für hinreichend starken Abstieg fest, während (2.16b) zu kleine Schritte vermeidet.

Das Gauß-Newton Verfahren konvergiert besonders schnell, wenn $\mathbf{J}^T\mathbf{J}$ den zweiten Term in Gleichung (2.7) dominiert, so dass die Näherung (2.13) sehr genau ist. Dies ist insbesondere in der Nähe der Lösung $\mathbf{x}^*$ und bei kleinen Residuen r_j der Fall.

2.1.2.2 Levenberg-Marquardt

Die Levenberg-Marquardt Methode verwendet, wie auch das Gauß-Newton Verfahren, Gleichung (2.13) als Approximation der Hessematrix von f. Allerdings verwendet sie zur Iteration die trust-region Strategie. Somit ist analog zu Gleichung (2.15) das zu lösende Teilproblem

$$\min_{\mathbf{q}} \ \|\mathbf{J}_k \mathbf{q} + \mathbf{r}_k\|_2^2, \quad \text{mit } \|\mathbf{q}\|_2 \leq \Delta_k \tag{2.17}$$

und trust-region Radius $\Delta_k > 0$.

Erfüllt die Gauß-Newton Lösung $\mathbf{q}_k^{\mathrm{GN}}$ aus Gleichung (2.15) die trust-region Bedingung $\|\mathbf{q}\|_2 \leq \Delta_k$, so löst diese trivialerweise auch Gleichung (2.17). Liegt $\mathbf{q}_k^{\mathrm{GN}}$ allerdings außerhalb der trust-region, wird die Fehlerfunktion f quadratisch mit

$$m_k(\mathbf{q}) = \tfrac{1}{2}\|\mathbf{r}_k\|^2 + \mathbf{q}^T \mathbf{J}_k^T \mathbf{r}_k + \tfrac{1}{2}\mathbf{q}^T \mathbf{J}_k^T \mathbf{J}_k \mathbf{q} \tag{2.18}$$

modelliert und an ihrer Stelle m_k unter der trust-region Bedingung minimiert. Die Schrittweite entspricht in diesem Fall dem gewählten trust-region Radius $\|\mathbf{q}_k\| = \Delta_k$.

Die optimale Schrittweite und die davon abhängige Iterationsrichtung wird iterativ, abhängig von der Übereinstimmung des Modells m_k und der Fehlerfunktion f, angepasst. Hierzu dient das Verhältnis von tatsächlicher zu prädizierter Reduktion des Funktionswertes

$$\tau = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{q}_k)}{m_k(\mathbf{0}) - m_k(\mathbf{q}_k)} \tag{2.19}$$

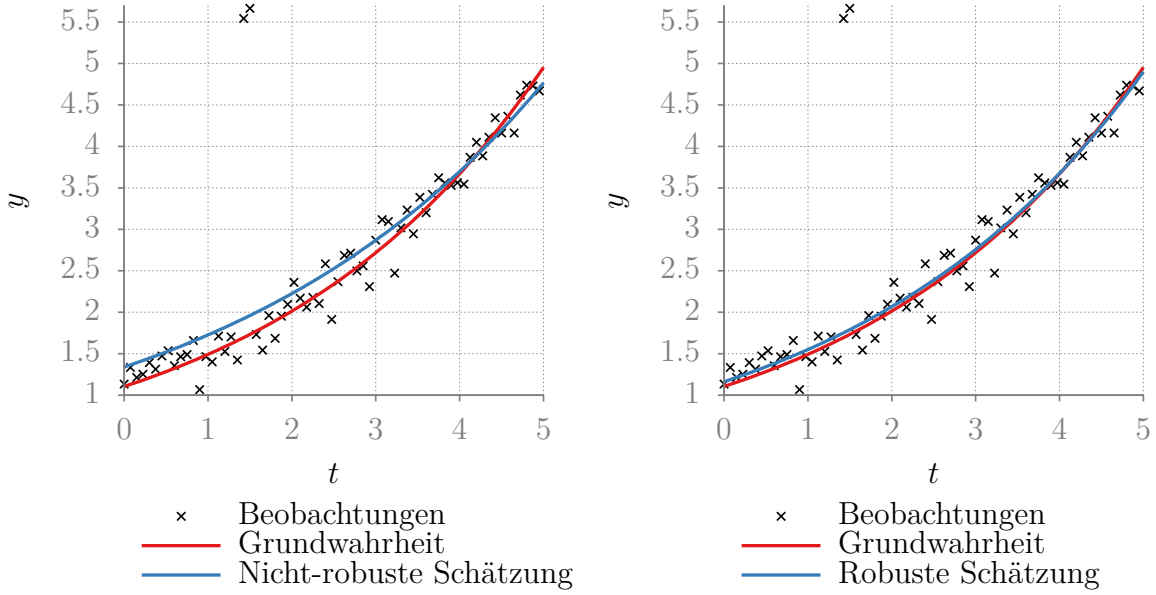


Abbildung 2.2 – Gewöhnliche und robuste LS-Schätzung bei Messungen mit Ausreißern, Daten und Codebeispiel übernommen aus [2]

als Maß dieser Übereinstimmung. Da sich $m_k(\mathbf{q}_k)$ aus der Minimierung von m_k ergibt, ist der Nenner immer positiv. τ kann somit nur für $f(\mathbf{x}_k + \mathbf{q}_k) > f(\mathbf{x}_k)$ negativ sein. In diesem Falle ist aber keine Reduktion des Funktionswertes erfolgt und der Schritt mit kleinerem Δ_k zu wiederholen. Für positives τ wird $f(\mathbf{x}_k + \mathbf{q}_k)$ für $\mathbf{x}_{k+1}$ übernommen. Liegt τ dabei nahe 1, so stimmt das Modell mit der Funktion sehr gut überein, so dass $\Delta_{k+1} > \Delta_k$ gewählt werden kann. Für τ nahe Null, wird der trust-region Radius verringert, in allen anderen Fällen bleibt er unverändert.

Für eine ausführliche Beschreibung des Algorithmus und Verfahren zur Minimierung von m_k sei der Leser erneut auf Nocedal und Wright [28] verwiesen.

Die Levenberg-Marquardt Methode wird in Kapitel 3 für den Bündelausgleich (siehe auch Abschnitt 2.3) bei der Objektaggregation und -rekonstruktion verwendet.

2.1.3 Robustes Least-Squares Verfahren

Das größte Problem bei der Anwendung des Least-Squares Verfahrens mit realen Messdaten ist die starke Anfälligkeit des Schätzers gegenüber Ausreißern, also solchen Messungen, die sehr starkes nicht-gauß'sches Rauschen aufweisen. Abbildung 2.2 verdeutlicht dieses Verhalten im Vergleich zu einer robusten Schätzung. Bei der gewöhnlichen LS-Schätzung führen einige wenige Ausreißer zu einem merklichen Bias der Schätzlösung, während die Lösung einer robusten LS-Schätzung nur unwesentlich verfälscht wird.

Ausreißer können beispielsweise aus korrupten Daten, falscher Versuchsdurchführung oder Fehlassoziationen entstehen. Insbesondere bei featurebasierten Bildverarbeitungsmethoden wie dem Bündelausgleich (siehe Abschnitt 2.3.2) treten immer wieder Fehlassoziationen auf. Daher ist es zwingend notwendig in solchen Anwendungen auf robuste Least-Squares Verfahren zurückzugreifen. Triggs u. a. [38] geben einen guten Einblick in die zugrundeliegende Theorie und liefern die Grundlage dieses Abschnitts.

Die Idee der robusten Least-Squares Verfahren besteht darin, Ausreißer als solche zu identifizieren und die zugehörigen Messungen gar nicht oder nur mit niedrigerem Gewicht in die Schätzung einfließen zu lassen. Zunächst müssen die einzelnen Residuen r_j also zu Residuenblöcken $\mathbf{r}_{j'}$ zusammengefasst werden, um abzubilden, welche Messungen jeweils logisch zusammen gehören. Eine Beobachtung in einem Kamerabild hat beispielsweise zwei logisch zusammenhängende Messungen (u, v) (siehe auch Abschnitt 2.2). Ergibt sich für eine solche Beobachtung, dass sie nicht plausibel ist (weil ihr bspw. eine Fehlassoziation zugrunde liegt), so sollen beide zugehörigen Messungen u und v verworfen oder geschwächt werden.

Um Ausreißer gänzlich aus der Schätzung zu entfernen, werden diese zum Beispiel nach einer Initialschätzung mithilfe eines Entscheidungskriteriums als solche identifiziert und die Schätzung anschließend ohne die assoziierten Messungen wiederholt. Die Zuverlässigkeit dieses Vorgehens lässt sich deutlich steigern, wenn es iterativ mit einem immer strengerem Kriterium wiederholt wird [21].

Eine Alternative oder auch Ergänzung ist die Modifikation der Kostenfunktion, so dass Messungen mit außergewöhnlich hohem Residuum schon während der Lösung des Least-Squares Problems geschwächt werden:

$$\min_{\mathbf{x}} \ \frac{1}{2} \sum_{j'} \rho_{j'} \left(\|\mathbf{r}_{j'}(\mathbf{x})\|^2 \right). \tag{2.20}$$

Die Verlustfunktion $\rho_{j'}(s)$ kann hierbei eine beliebige monoton steigende Funktion mit $\rho_{j'}(0) = 0$ und $\frac{\partial}{\partial s}\rho_{j'}(0) = 1$ sein. Allerdings sollte die Wahl von $\rho_{j'}(s)$ so motiviert werden, dass die resultierende Fehlerfunktion $f'_{j'}(\mathbf{x}) = \frac{1}{2}\rho_{j'}\left(\|\mathbf{r}_{j'}(\mathbf{x})\|^2\right)$ die Fehlerverteilung in Form einer negativen Log-Likelihood realistisch modelliert.

Abbildung 2.3 veranschaulicht $f'_{j'}(\mathbf{x})$ unter Verwendung von keiner und zwei beispielhaften Verlustfunktion. Die Erste wurde von Triggs u. a. [38] für featurebasierte Beobachtungen, die gelegentliche Fehlassoziationen aufweisen, vorgeschlagen. Hierbei wird die Fehlerverteilung als Gauß-Verteilung mit einem gleichverteiltem Hintergrundrauschen der Ausreißer modelliert:

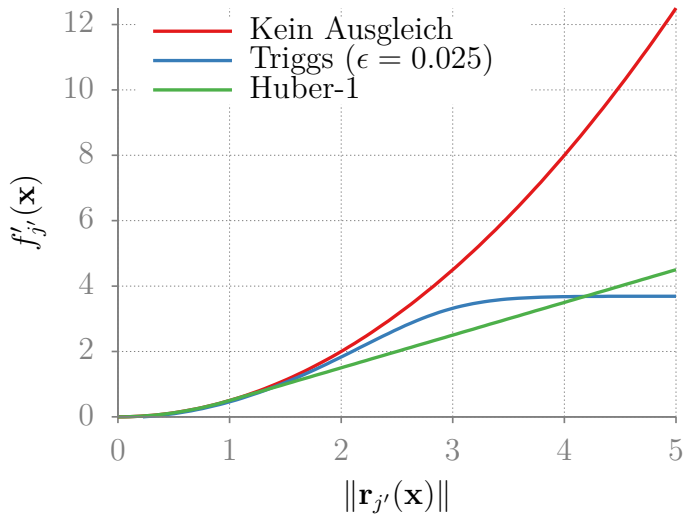


Abbildung 2.3 – Verlustfunktion $\rho_{j'}$

$$\rho_{j'}(s) = -2 \ln \left(e^{-\frac{1}{2}s} + \epsilon \right). \tag{2.21}$$

ϵ parametrisiert dabei die Ausreißerfrequenz.

Die Zweite in Abbildung 2.3 dargestellte Verlustfunktion, der sogenannte Huber-k-Schätzer, wurde von Huber [19] vorgeschlagen und ist deutlich effizienter zu bestimmen:

$$\rho_{j'}(s) = \begin{cases} s & \text{für } s \leq k \\ 2\sqrt{k|s|} - k & \text{für } s > k. \end{cases} \tag{2.22}$$

Der Huber-k-Schätzer wird beim Bündelausgleichsverfahren in Abschnitt 3.2 für eine robuste Objektrekonstruktion verwendet.

2.2 Abbildungsgeometrie

Für die folgenden Herleitungen wurden die Lehrbücher von Szeliski [36], Hartley und Zisserman [18], Stiller, Bachmann und Geiger [33] und Azad, Gockel und Dillmann [5] verwendet. Weitere Quellen werden an gegebener Stelle aufgeführt.

2.2.1 Lochkamera Modell

Die Bildentstehung, also der Abbildungsprozess eines kontinuierlichen 3D Punktes einer beobachteten Umgebung auf einen diskreten 2D Bildpunkt, lässt sich im Grundprinzip durch das erweiterte Lochkameramodell, in Abbildung 2.4 veranschaulicht, erklären.

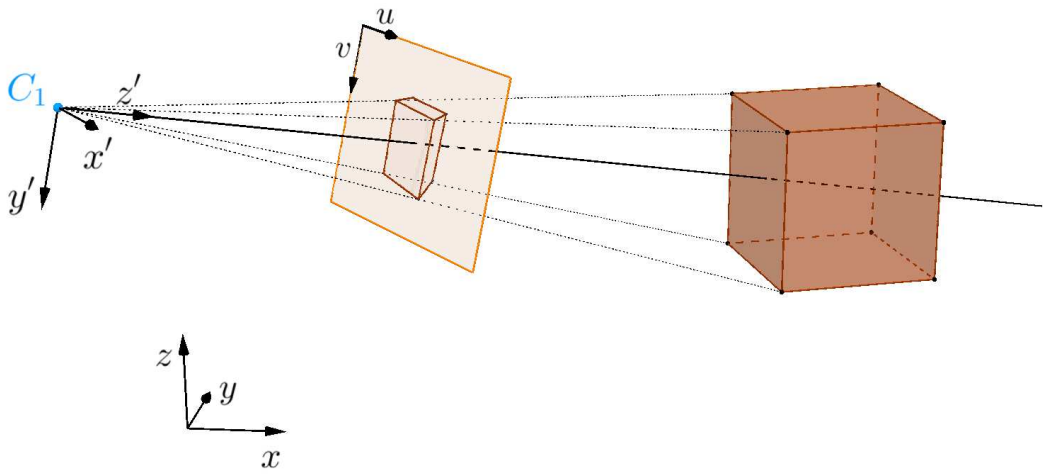


Abbildung 2.4 – Lochkameramodell

Zunächst seien die verwendeten Koordinatensysteme (KS) definiert:

Bildkoordinatensystem Zweidimensionales KS (u, v) mit Ursprung in der linken oberen Bildecke. Die u-Achse zeigt nach rechts, die v-Achse nach unten. Die Einheiten sind diskrete Pixel.

Kamerakoordinatensystem Dreidimensionales KS (x', y', z') mit Ursprung im optischen Zentrum (Brennpunkt) der Kamera. Die x- und y-Achse sind parallel zur u- und v-Achse des zugehörigen Bildkoordinatensystems. Folglich zeigt die z-Achse in Blickrichtung (rechtshändisches KS). Die Werte werden kontinuierlich in Metern angegeben.

Weltkoordinatensystem Dreidimensionales Bezugs-KS (x, y, z) mit beliebigem, aber weltfestem Ursprung. Die Werte werden kontinuierlich in Metern angegeben.

Bei der Beobachtung eines Punktes $\mathbf{x}' = (x', y', z')$ im Kamera-KS, wird dieser durch die Lochblende der Kamera perspektivisch auf einen Bildpunkt $\mathbf{w} = (u, v)$ projiziert. Mathematisch lässt sich diese Projektion in homogenen Koordinaten wie folgt beschreiben (die Tilde über einer Variable, bei $\tilde{\mathbf{w}}$ oder $\tilde{\mathbf{x}}$, weist dabei auf homogene Koordinaten hin):

$$\begin{pmatrix} u\,z' \\ v\,z' \\ z' \end{pmatrix} = \begin{pmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \tag{2.23}$$

$$\tilde{\mathbf{w}} = \mathbf{K}\,\mathbf{x}'. \tag{2.24}$$

Dabei beschreibt $\mathbf{K}$ die Kameramatrix mit den sogenannten intrinsischen Parametern. Hierzu zählen unter anderem c_u und c_v, die den Bildhauptpunkt beschreiben. Die achsenbezogenen Brennweiten f_u und f_v lassen sich aus der Brennweite f und dem Pixelabstand des Sensors in horizontaler bzw. vertikaler Richtung Δu, Δv berechnen:

$$f_u = \frac{f}{\Delta u}; \quad f_v = \frac{f}{\Delta v}. \tag{2.25}$$

Der Scherungsparameter s beschreibt eine mögliche Schrägstellung des Sensors, der idealerweise exakt rechtwinklig zur optischen Achse angebracht ist. Häufig wird dieser Parameter allerdings zu Gunsten einer genaueren Kalibrierung der verbleibenden Parameter vernachlässigt ($s = 0$).

Die extrinsischen Parameter einer Kamera erklären die Lage der Kamera zum Weltkoordinatensystem. Ein Punkt in Weltkoordinaten $\mathbf{x}$ kann somit mittels einer einfachen Rotation $\mathbf{R}$ und Translation $\mathbf{t}$ in Kamerakoordinaten $\mathbf{x}'$ überführt werden:

$$\mathbf{x}' = \mathbf{R}\,\mathbf{x} + \mathbf{t}. \tag{2.26}$$

Die Projektion eines Punktes in Weltkoordinaten $\mathbf{x}$ auf einen Bildpunkt $\mathbf{w}$ lässt sich also abgekürzt formulieren als

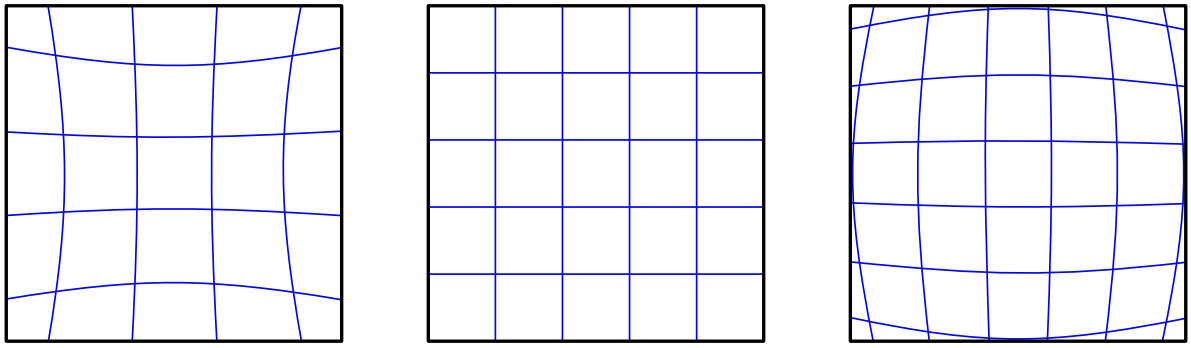
$$\tilde{\mathbf{w}} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\,\tilde{\mathbf{x}}, \tag{2.27}$$

mit der Projektionsmatrix $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$.

2.2.2 Verzeichnung

Das erweiterte Lochkameramodell (2.27) ist für die Herleitung des Bündelausgleichsverfahrens zur Szenen- / Objektrekonstruktion ausreichend, vernachlässigt allerdings zahlreiche Effekte von Verzeichnung bis zu chromatischer Aberration. Während die Verzeichnung im folgenden kurz erläutert werden soll, sei der Leser bezüglich fortgeschrittenerer Kameramodelle auf die Literatur verwiesen [36][18].

Die Verzeichnung beschreibt den Effekt, dass reale optische Linsen das Bild verzerrend auf den Sensor projizieren. Abbildung 2.5 veranschaulicht die gängigsten Modelle radialer Verzeichnung (kissen- und tonnenförmig). Insbesondere Weitwinkel- und sogenannte Fisheye-Objektive weisen eine starke Verzeichnung auf, sodass diese oft in einem Vorverarbeitungsschritt ausgeglichen werden muss. Für eine mathematische Beschreibung wird zunächst das Bildebenenkoordinatensystem eingeführt, um Bildpunkte vor der Brennweitenskalierung, Scherung und Verschiebung bzgl. der Bildmitte zu beschreiben.



(a) kissenförmige Verzeichnung **(b)** ohne Verzeichnung **(c)** tonnenförmige Verzeichnung

Abbildung 2.5 – Radiale Verzeichnung

Bildebenenkoordinatensystem Zweidimensionales KS (u', v') mit Ursprung im Schnittpunkt der optischen Achse mit der Bildebene. Beide Achsen sind parallel zu den Achsen des Bildkoordinatensystems.

Sei $\mathbf{w}' = (u', v')$ ein auf die Bildebene projizierter Punkt $\mathbf{x}$ und $\mathbf{w}^*$ der Bildpunkt nach radialer Verzeichnung $L(r')$, abhängig von dem Abstand des Punktes zur optischen Mitte $r' = \sqrt{u'^2 + v'^2}$. Dann wird der verzerrte Bildpunkt durch anschließende Anwendung der Kameramatrix bestimmt:

$$\tilde{\mathbf{w}}' = [\mathbf{R}|\mathbf{t}]\,\tilde{\mathbf{x}} \tag{2.28}$$

$$\mathbf{w}^* = L(r') \cdot \mathbf{w}' \tag{2.29}$$

$$\tilde{\mathbf{w}} = \mathbf{K}\tilde{\mathbf{w}}^*. \tag{2.30}$$

Die Verzeichnungsfunktion $L(r)$ lässt sich, motiviert durch eine Taylorentwicklung, in der Regel ausreichend gut durch Polynome niedriger Ordnung annähern, beispielsweise:

$$L(r) = 1 + \kappa_1 r^2 + \kappa_2 r^4. \tag{2.31}$$

Für Fisheye-Objektive reicht diese Näherung allerdings nicht aus. Daher haben Xiong und Turkowski [39] ein passenderes sogenanntes „äquidistantes“ Modell entwickelt (siehe auch [36]):

$$L(r) = s\,\frac{\operatorname{atan}(r)}{r}. \tag{2.32}$$

Der Faktor s hängt mit der Brennweite zusammen und wird ebenfalls bei der Kalibrierung (siehe Abschnitt 2.4) bestimmt.

Das für diese Arbeit verwendete Stereokamerasystem ist mit Fisheye-Objektiven ausgestattet, so dass in Kapitel 3 ein Kameramodell mit äquidistanter Verzeichnung, wie in Gleichung (2.32), verwendet wird.

2.3 Struktur- und Bewegungsrekonstruktion

Die folgenden Herleitungen orientieren sich an den Lehrbüchern von Szeliski [36] und Hartley und Zisserman [18]. Auf zusätzliche Quellen wird an entsprechender Stelle verwiesen.

2.3.1 Rekonstruktion von Punkten

Für die 3D Rekonstruktion der beobachteten Szene reicht ein einzelnes Bild nicht aus, da die Inverse der Abbildung (2.23)

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{K}^{-1} \begin{pmatrix} u\,z' \\ v\,z' \\ z' \end{pmatrix} \tag{2.33}$$

nicht eindeutig ist. Der Tiefenwert z' eines 3D Punktes in Kamerakoordinaten lässt sich aus einem einzelnen 2D Bildpunkt nicht rekonstruieren. Ein Mehrkamerasystem, beziehungsweise Bilder aus verschiedenen Perspektiven, kann diese Information hingegen unter Umständen liefern, so dass sich dieser Abschnitt der Stereo- bzw. Mehrkamerarekonstruktion widmet.

Es handelt sich also um das Problem, einen 3D Punkt, von nun an als Landmarke bezeichnet, aus einer Mehrzahl korrespondierender Bildpunkte und dazu bekannten Kamerapositionen zu rekonstruieren. Es seien mindestens zwei Kameras mit Index j gegeben, deren Kameramatrix $\mathbf{K}_j$ aus (2.23) sowie ihre Position $\mathbf{t}_j$ und Orientierung $\mathbf{R}_j$ bekannt sind. Außerdem bezeichne $\mathbf{w}_j$ die Beobachtung der Landmarke im Bild der Kamera j. Gesucht ist nun eine Schätzung für die Weltkoordinaten der Landmarke $\hat{\mathbf{p}}$, die diese Beobachtungen möglichst exakt erklärt.

Es gibt zahlreiche Möglichkeiten dieses Problem zu lösen (siehe insbesondere [18] und [36]). Hier soll jedoch eine Variante vorgestellt werden, die sich im Anschluss leicht zur Methodik des Bündelausgleichverfahrens erweitern lässt.

Betrachtet man zunächst die tatsächliche Position einer Landmarke $\bar{\mathbf{p}}$ und ihre Rückprojektion auf jedes Kamerabild $\bar{\mathbf{w}}_j = \mathbf{K}_j[\mathbf{R}_j|\mathbf{t}_j]\,\bar{\mathbf{p}}$, so werden die Beobachtungen $\mathbf{w}_j$ auf Grund von Korrespondenzfehlern, Pixelrauschen und anderen Fehlerquellen nicht exakt mit den Rückprojektionen übereinstimmen.

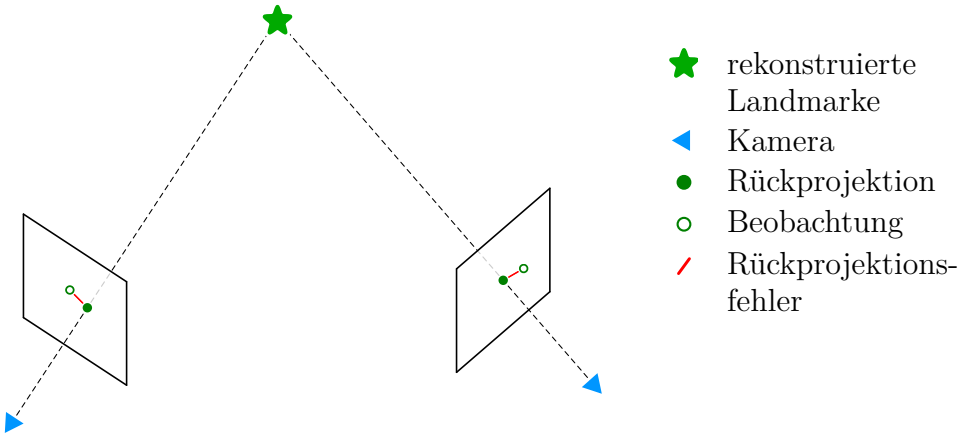


Abbildung 2.6 – Rekonstruktion von Punkten

Die durch die Beobachtungen definierten Sichtstrahlen schneiden sich also nicht im Punkt $\bar{\mathbf{p}}$, sondern liegen schief zueinander. Gesucht ist folglich eine solche Schätzung $\hat{\mathbf{p}}$, die $\bar{\mathbf{p}}$ so genau wie möglich annähert. Hierfür wird die Summe der quadratischen Rückprojektionsfehler $\sum_j \|\mathbf{w}_j - \hat{\mathbf{w}}_j\|^2$ mit einem iterativen Verfahren aus Abschnitt 2.1 minimiert:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \left\{ \sum_j \|\mathbf{w}_j - \mathbf{K}_j[\mathbf{R}_j|\mathbf{t}_j]\,\mathbf{p}\|^2 \right\}. \tag{2.34}$$

Abbildung 2.6 veranschaulicht das Verfahren schematisch.

Im einfachsten Falle einer idealen Stereokamera, sind zwei Kameras $j = \{0, 1\}$ mit identischen intrinsischen Parametern $\mathbf{K}_0 = \mathbf{K}_1$, gleicher Orientierung und lediglich einem seitlichen Versatz zueinander gegeben,

$$\mathbf{R}_0 = \mathbf{R}_1; \quad \mathbf{t}_1 = \mathbf{R}_0^{-1}(\mathbf{b} - \mathbf{t}_0), \tag{2.35}$$

wobei $\mathbf{b} = (b, 0, 0)$ als Baseline bezeichnet wird. Die optimale Lösung der stereoskopischen Rekonstruktion lässt sich sogar geschlossen bestimmen [18].

2.3.2 Bündelausgleich

Die zeitgleiche Rekonstruktion der 3D Positionen von m Landmarken, der Posen und gegebenenfalls auch intrinsischen Parameter von n Kameras wird in der Literatur als „structure from motion“ (SfM) oder auch „simultaneous localization and mapping“ (SLAM) bezeichnet. SfM bezeichnet dabei meist Offline-Verfahren, die vor allem im Bereich der Computer Vision verwendet werden, während SLAM nicht nur auf Kamerasensorik beschränkt ist und somit auch onlinebasierte Verfahren umfasst, die unter anderem in der Robotik populär sind.

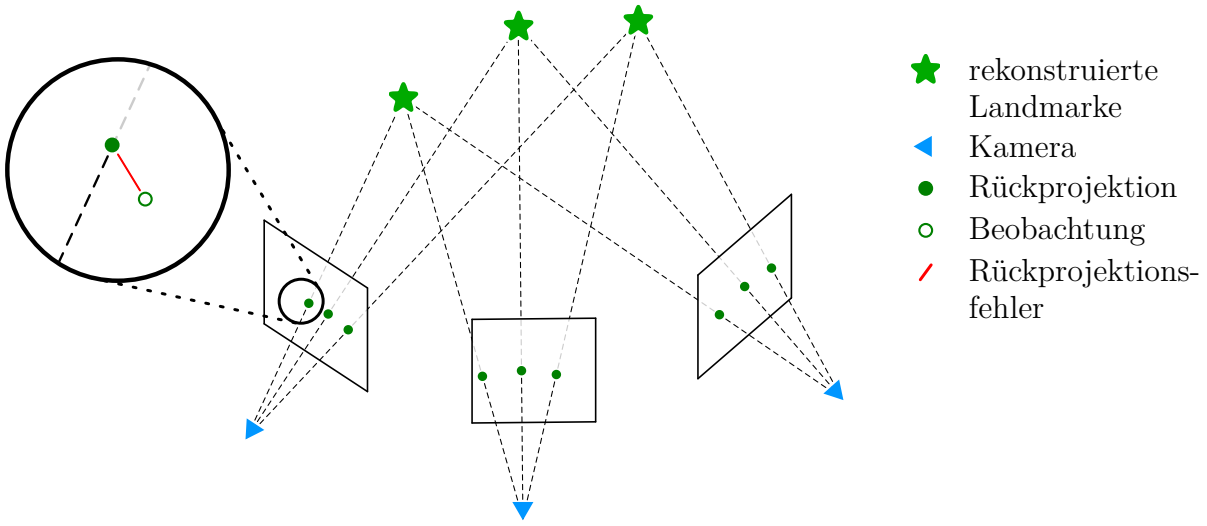


Abbildung 2.7 – Bündelausgleich

Abbildung 2.7 stellt das Problem exemplarisch dar. Es sind drei Kameras und drei Landmarken so angeordnet, dass die Landmarken in allen drei Kamerabildern beobachtet werden. Ziel ist es die Landmarkenpositionen und Kameraposen bestmöglich zu schätzen. Die in SfM verwendete Methodik ist als „Bündelausgleichsverfahren“ (kurz „Bündelausgleich“, engl. „bundle adjustment“) bekannt und minimiert, analog zur zuvor hergeleiteten Punktrekonstruktion, den quadratischen Rückprojektionsfehler, der in der Vergrößerung rot hervorgehoben ist.

Ausgehend von Gleichung 2.34 kann man die Minimierungsfunktion also leicht zum Bündelausgleich erweitern, indem mehrere Landmarken berücksichtigt, die Kameraposen und -matrizen im Parametervektor $\boldsymbol{\lambda}$ eingebunden werden und außerdem nicht jede Landmarke in jedem Kamerabild beobachtet sein muss:

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} \left\{ \sum_{(i,j)\in\Psi} \|\mathbf{w}_{ij} - \mathbf{K}_j[\mathbf{R}_j|\mathbf{t}_j]\ \mathbf{p}_i\|^2 \right\}. \tag{2.36}$$

Der zusätzliche Index i verweist auf die jeweilige Landmarke, deren Sichtbarkeit in jedem Bild j durch Ψ bestimmt wird. Abweichend von Abbildung 2.7 ist die Beobachtbarkeit aller Landmarken in allen Kamerabildern keine Voraussetzung des Bündelausgleichsverfahrens, vereinfacht dort allerdings die Darstellung. Der Parametervektor $\boldsymbol{\lambda}$ setzt sich nun aus den Landmarkenpunkten, Kameraposen und -matrizen zusammen:

$$\boldsymbol{\lambda} = (\mathbf{p}_0, \ldots, \mathbf{p}_{m-1}, \boldsymbol{\omega}_0, \ldots, \boldsymbol{\omega}_{n-1}, \mathbf{t}_0, \ldots, \mathbf{t}_{n-1}, \mathbf{k}_0, \ldots, \mathbf{k}_{n-1})^T, \tag{2.37}$$

wobei eine Minimalparametrierung in Form der Angle-Axis Repräsentation einer Orientierung $\boldsymbol{\omega}_j = \theta_j \vec{\mathbf{n}}_j$ mit Rotationsachse $\vec{\mathbf{n}}_j$ und -winkel θ_j, statt der redundanten Rotations-

matrizen $\mathbf{R}_j$, und für die intrinsischen Kameraparameter $\mathbf{k} = (f_u, f_v, s, c_u, c_v)^T$ gewählt wurde.

Beim SfM aus Bildern einer bewegten Stereokamera vereinfacht sich der Parametervektor enorm, da zum einen nur noch zwei ($\mathbf{k}_0$ und $\mathbf{k}_1$) statt n Kameramatrizen geschätzt werden müssen. Zum anderen halbiert sich die Anzahl der zu schätzenden Kameraposen, da beide Kameras fest zueinander verbaut sind (siehe Baseline in Gleichung 2.35). Außerdem ist es üblich zunächst eine gute intrinsische Kalibrierung mithilfe von Kalibriermustern oder einer leicht validierbaren Aufzeichnung in einer featurereichen statischen Umgebung durchzuführen, bevor die Struktur- und Bewegungsrekonstruktion schwieriger Sequenzen in Angriff genommen wird.

Das in Kapitel 3 entwickelte Verfahren basiert auf Eingangsdaten, die mit Hilfe einer solchen Stereokamera gewonnen wurden. Deren intrinsische Kalibrierungen $\mathbf{k}_0$ und $\mathbf{k}_1$ werden im Zuge der in Abschnitt 2.4 erläuterten Selbstkalibrierung gewonnen. Das Bündelausgleichsverfahren wird für die Objektaggregation und -rekonstruktion in Abschnitt 3.1 beziehungsweise 3.2 verwendet.

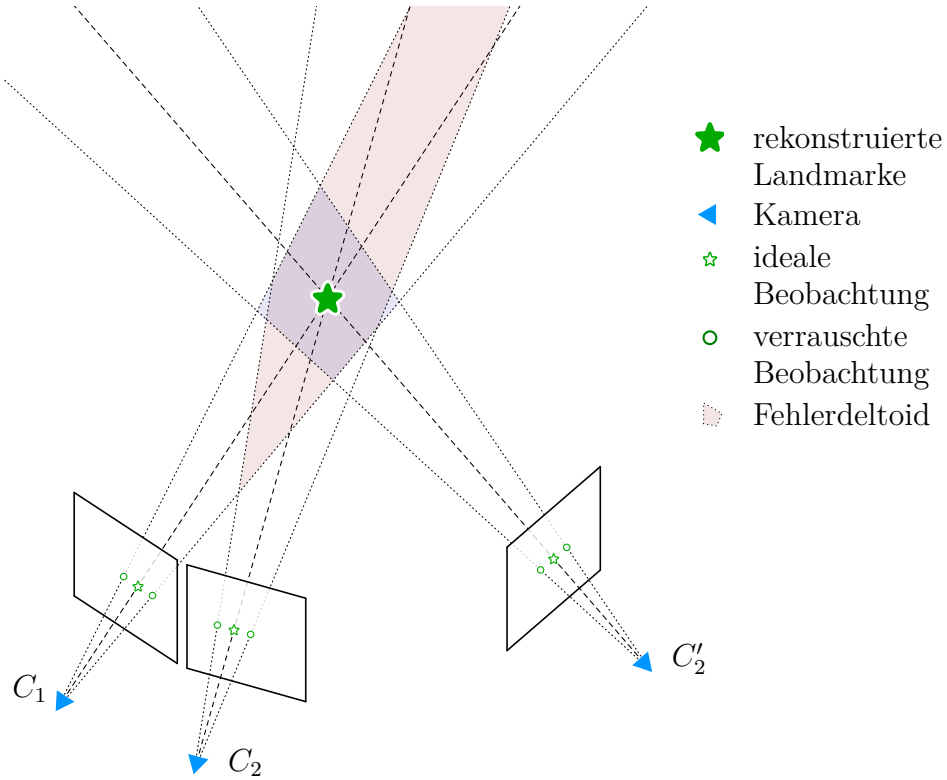


Abbildung 2.8 – Tiefenunsicherheit

2.3.3 Unsicherheit der 3D Rekonstruktion

Eine Schwierigkeit bei der Strukturrekonstruktion aus Kamerabildern ist, zuverlässige Tiefeninformation der rekonstruierten 3D Punkte zu gewinnen. Die Problematik wird in Abbildung 2.8 anhand eines Zweikamerasystems verdeutlicht. Ausgehend von den Rückprojektionen der grünen Landmarke in die Bilder der Kameras C_1, C_2 und C_2' wird Quantisierungsrauschen durch Verschiebung der idealen Beobachtungen $\bar{\mathbf{w}}_j$ um jeweils ein Pixel nach $\mathbf{w}_{j1}$ und $\mathbf{w}_{j2}$ simuliert. Die Sichtstrahlen der Beobachtungen $\mathbf{w}_{11}$ und $\mathbf{w}_{12}$ aus Kamera C_1 und $\mathbf{w}_{21}$ und $\mathbf{w}_{22}$ aus Kamera C_2 legen das rot unterlegte Fehlerdeltoid fest. Wie bereits Matthies und Shafer [26] feststellen, lässt sich dieses mit einem Gauß'schen Fehlermodell nur näherungsweise abbilden, weshalb an dieser Stelle darauf verzichtet wird.

Dennoch wird deutlich: Ist die Baseline, also der Abstand der beiden Kameras C_1 und C_2, gering, so weist die Rekonstruktion in der Tiefe eine hohe Unsicherheit auf. Bei einer größeren Baseline und somit stärkeren Rotation der Kameras um die Landmarke herum, wie zwischen C_1 und C_2', ist die Tiefenunsicherheit hingegen wesentlich geringer. Dies wird durch das blaue kürzere Fehlerellipsoid einleuchtend.

Szeliski und Kang [35] untersuchen diese Tiefenunsicherheit auch für Mehrkamerasysteme und stellen fest, dass mindestens drei oder mehr Bilder sowie eine starke Rotation notwendig sind, um eine genaue Rekonstruktion zu ermöglichen. Sie zeigen zudem folgende Möglichkeit, die Unsicherheit indirekt aus der Kovarianzmatrix der LS-Schätzung zu bestimmen. Angenommen Beobachtungen im Bild unterliegen Gauß'schem Rauschen, kann die Unsicherheit mithilfe der inversen Kovarianzmatrix, oder auch Fisher-Information $\mathbf{A}$, berechnet werden:

$$\mathbf{A} = \mathbf{J}(\hat{\mathbf{x}})^T \mathbf{J}(\hat{\mathbf{x}}). \tag{2.38}$$

Hierbei sei $\mathbf{J}(\hat{\mathbf{x}})$ die Jacobimatrix des Residuenvektors $\mathbf{r}$ der Schätzlösung $\hat{\mathbf{x}}$ (siehe (2.5)). $\mathbf{A}$ ist also mit der Approximation der Hessematrix (2.13) identisch und somit bei Verwendung eines Gauß-Newton- oder Levenberg-Marquardt-Schätzers bereits bekannt. Wenn auch die intrinsischen Kameraparameter bereits bestimmt wurden und der Parametervektor $\boldsymbol{\lambda}$ die Form

$$\boldsymbol{\lambda} = \left(\mathbf{p}_0, \ldots, \mathbf{p}_{m-1}, \mathbf{m}_0, \ldots, \mathbf{m}_{n-1}\right)^T \tag{2.39}$$

mit Strukturparametern $\mathbf{p}_i$ und Bewegungsparametern $\mathbf{m}_j = (\boldsymbol{\omega}_j, \mathbf{t}_j)$ hat, dann hat $\mathbf{A}$ die Form

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{A}_p & \mathbf{A}_{pm} \\ \hline \mathbf{A}_{pm}^T & \mathbf{A}_m \end{array} \right) \tag{2.40}$$

$\mathbf{A}_p$ und $\mathbf{A}_m$ sind dabei Blockdiagonalmatrizen mit

$$\mathbf{A}_{p_i} = \sum_j \nabla_{\mathbf{p}} r_{ij}(\mathbf{x}) \nabla_{\mathbf{p}} r_{ij}(\mathbf{x})^T \tag{2.41}$$

$$\mathbf{A}_{m_j} = \sum_i \nabla_{\mathbf{m}} r_{ij}(\mathbf{x}) \nabla_{\mathbf{m}} r_{ij}(\mathbf{x})^T \tag{2.42}$$

auf den Diagonalen, wobei $\nabla_{\boldsymbol{\chi}} f$ den Gradienten von f nach $\boldsymbol{\chi}$ beschreibt. $\mathbf{A}_{pm}$ ist dagegen auf Grund von Kreuztermen vollbesetzt.

Um nun eine Aussage über die Unsicherheit einer geschätzten Landmarkenposition zu treffen, muss die zugehörige Blockmatrix $\mathbf{A}_{p_i}$ untersucht werden. Young und Chellappa [40] bestimmen beispielsweise eine Cramer-Rao Schranke aus den Inversen der Diagonaleinträge. Szeliski und Kang [35] stellen allerdings fest, dass die Cramer-Rao Schranke vergleichsweise schwach ausfallen kann, insbesondere wenn $\mathbf{A}$ singulär oder nahezu singulär ist. Daher schlagen sie eine Eigenwertanalyse der Blockmatrizen vor. Kleine Eigenwerte deuten dabei auf eine hohe Unsicherheit, verschwindende Eigenwerte sogar auf eine Mehrdeutigkeit hin.

Ähnlich motiviert kann die Kovarianzmatrix $\mathbf{C}$ durch Invertieren von $\mathbf{A}$ bestimmt und anschließend die Eigenwerte der Blockmatrizen $\mathbf{C}_{\mathbf{p}_i}$ untersucht werden. Die Invertierung ist zwar rechenaufwendig, weshalb sie meist vermieden wird. Diese Variante hat jedoch den Vorteil, dass sie der Hauptkomponentenanalyse (engl. „principal component analysis“) entspricht und somit die resultierenden Eigenwerte- und vektoren eine konkrete räumliche Bedeutung haben. Der Eigenvektor $\boldsymbol{\nu}_{\mathbf{p}_i,\max}$ zum größten Eigenwert $\lambda_{\mathbf{p}_i,\max}$ von $\mathbf{C}_{\mathbf{p}_i}$ beschreibt nämlich die Richtung der größten Unsicherheit und $\lambda_{\mathbf{p}_i,\max}$ ihre Varianz (in Metern).

Somit sind Verfahren gegeben, mit denen die Schätzlösung bezüglich einzelner Landmarken zusätzlich zum verbleibenden Residuum bezüglich ihrer Unsicherheit bewertet werden kann. Die Bewertung der Kameraposen wird von Szeliski und Kang [35] vermieden, da die Fehlerfortpflanzung von kombinierten Rotationen und Translationen wenig untersucht ist. Daher werden auch in dieser Arbeit hauptsächlich nur die Strukturparameter hinsichtlich ihrer Unsicherheit bewertet.

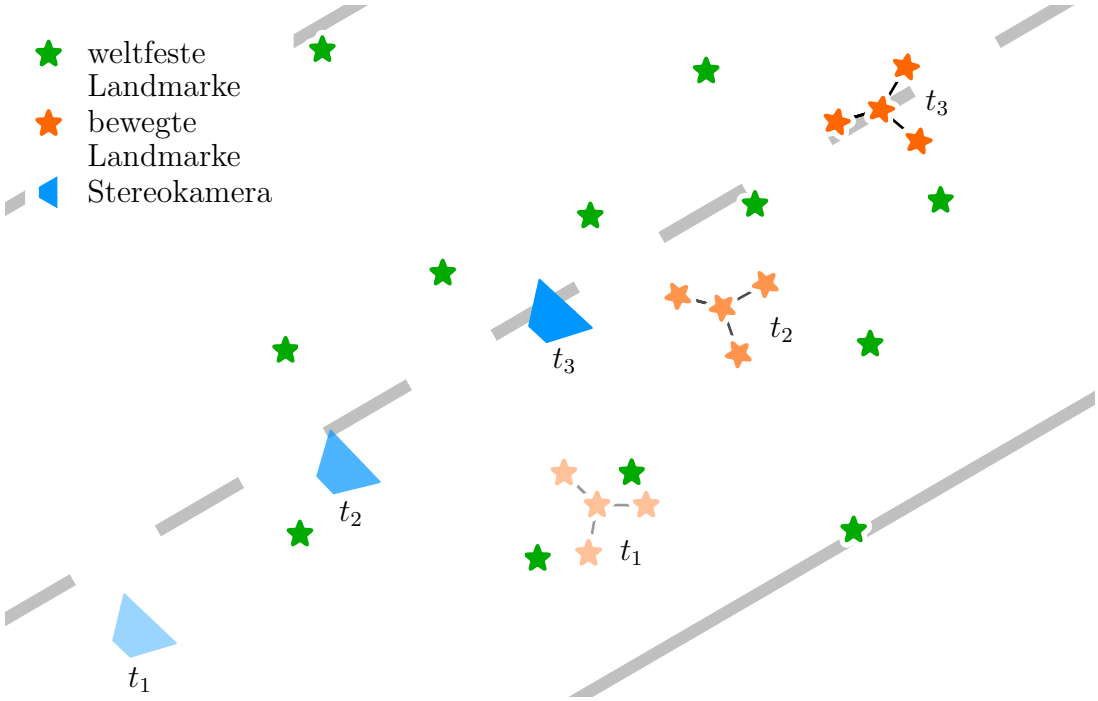


Abbildung 2.9 – Eigenbewegungsschätzung und Kartierung

2.4 Eingangsdaten

Die Eingangsdaten dieser Arbeit ergeben sich aus der Eigenbewegungsschätzung und Kartierung dünn besetzter Bildmerkmale, die seitens Atlatec realisiert wurde.

Für dieses Landmarkentracking bildet das Unternehmen zunächst die hochauflösenden Stereobilder in einem Entropisierungsschritt mittels Histogrammstreckung von 12-bit Graustufen auf effizienter handhabbare 8-bit ab.

In den entropisierten Bildern werden mithilfe eines spektralen Featuredeskriptors Bildmerkmale extrahiert und dann mit den Bildmerkmalen der anderen Kamera sowie der beiden zeitlich folgenden Bildern assoziiert, wie von Geiger, Ziegler und Stiller [14] beschrieben. Somit weist jede Landmarke mindestens vier Bildkorrespondenzen auf und wird auf Grund des zeitlichen Trackings als sogenanntes „Tracklet“ bezeichnet.

Zuletzt wird die Position der weltfesten Landmarken und die Eigenbewegung der Stereokamera mittels Bündelausgleich rekonstruiert. Bewegte Landmarken werden hierbei als Ausreißer der Schätzung erkannt und als solche markiert.

Die intrinsische Kalibrierung der beiden Kameras $\mathbf{k}_0$ und $\mathbf{k}_1$, sowie deren seitlicher Versatz b (siehe Gleichung (2.35)), wird zuvor separat in einer weitestgehend statischen und featurereichen Umgebung ebenfalls mittels Bündelausgleich bestimmt.

Abbildung 2.9 veranschaulicht einen Straßenabschnitt in Vogelperspektive mit zahlreichen weltfesten Landmarken, einer bewegten Stereokamera und einem bewegten Objekt mit

vier Landmarken, die als Ausreißer markiert wurden. Diese zeitlich getrackten Ausreißer der Eigenbewegungsschätzung bilden die Eingangsdaten dieser Arbeit.

3 Methodik

Bevor die erarbeitete Methodik im Detail vorgestellt wird, soll hier eine kurze Systemübersicht gegeben werden. Abbildung 3.1 veranschaulicht die wesentlichen Verarbeitungsschrit-

(a) Eingangsdaten

(b) Erkannte Objekte

(c) Rekonstruierte Objekte

Abbildung 3.1 – Systemüberblick

te.

Im ersten Schritt, dem sich Abschnitt 3.1 widmet, werden die Tracklets der Ausreißerlandmarken der Eigenbewegungsschätzung, dargestellt in Abbildung 3.1(a), zu Objekten aggregiert. Diese Aufgabe ist besonders schwierig, da keinerlei Information darüber vorliegt, wie viele Objekte in der Szene vorhanden sind, räumlich getrennte Objekte sich im Bildbereich durchaus überlappen können und zudem viele Störungen die Eingangsdaten prägen. Solche Landmarken, die aus Bildrauschen entstehen oder zu keinem bewegten Verkehrsteilnehmer gehören (wie bspw. Wolken oder im Wind schwingende Bäume), werden also herausgefiltert und die übrigen zu räumlich getrennten Objekten aggregiert, wie in Abbildung 3.1(b) zu sehen. Eine besondere Herausforderung ist dabei, sich im Bild überlappende Objekte klar voneinander zu trennen, da das Verschmelzen zweier Objekte zu schwerwiegenden Schätzfehlern bei der anschließenden Objektrekonstruktion führen kann. Zudem wird dies allein unter der Annahme starrer Körper und somit klassenfrei realisiert. Um also den Eindruck einer lernbasierten Objekterkennung mittels klassischer Bildverarbeitungsmethoden zu vermeiden, wird für diesen Verarbeitungsschritt statt dessen der Begriff „Objektaggregation“ verwendet. Dieser Abschnitt bildet sowohl vom Umfang als auch im Hinblick auf den wissenschaftlichen Beitrag den Hauptbestandteil dieser Arbeit.

Im zweiten Verarbeitungsschritt, in Abschnitt 3.2 erläutert, wird die Relativbewegung und Struktur jedes einzelnen Objekts mittels Bündelausgleichs rekonstruiert. Hierbei fallen weitere rauschende Bildmerkmale dank einer robusten Least-Squares Schätzung heraus. Aus den so gewonnen 3D Positionen der Landmarken eines Objekts wird zusätzlich die Oberflächenstruktur des Objekts zu Visualisierungszwecken approximiert. Abbildung 3.1(c) zeigt das Ergebnis der Struktur- und Oberflächenrekonstruktion als Rückprojektion im Bild.

3.1 Objektaggregation

Um die Tracklets zu Objekten zu gruppieren, wird untersucht, welche im Bild benachbarten Landmarken zu dem selben Objekt gehören. Hierzu wird aus der Annahme starrer Objekte zunächst abgeleitet, dass ein Teilsegment bestehend aus wenigen Landmarken desselben Objekts ebenfalls starr sein muss. Diese Folgerung wird ausgenutzt, um das Assoziationsproblem über die Starrheitshypothese zu lösen.

Zu Beginn wird eine initiale Nachbarschaft der Bildmerkmale durch Triangulation erzeugt und aus benachbarten Landmarken kleinste Objektsegmente, die in dieser Arbeit „Atome“ genannt werden, gebildet. Anschließend wird für jedes Atom untersucht, ob sich seine Beobachtungen unter der Starrheitshypothese des Teilsegments erklären lassen. Zuletzt

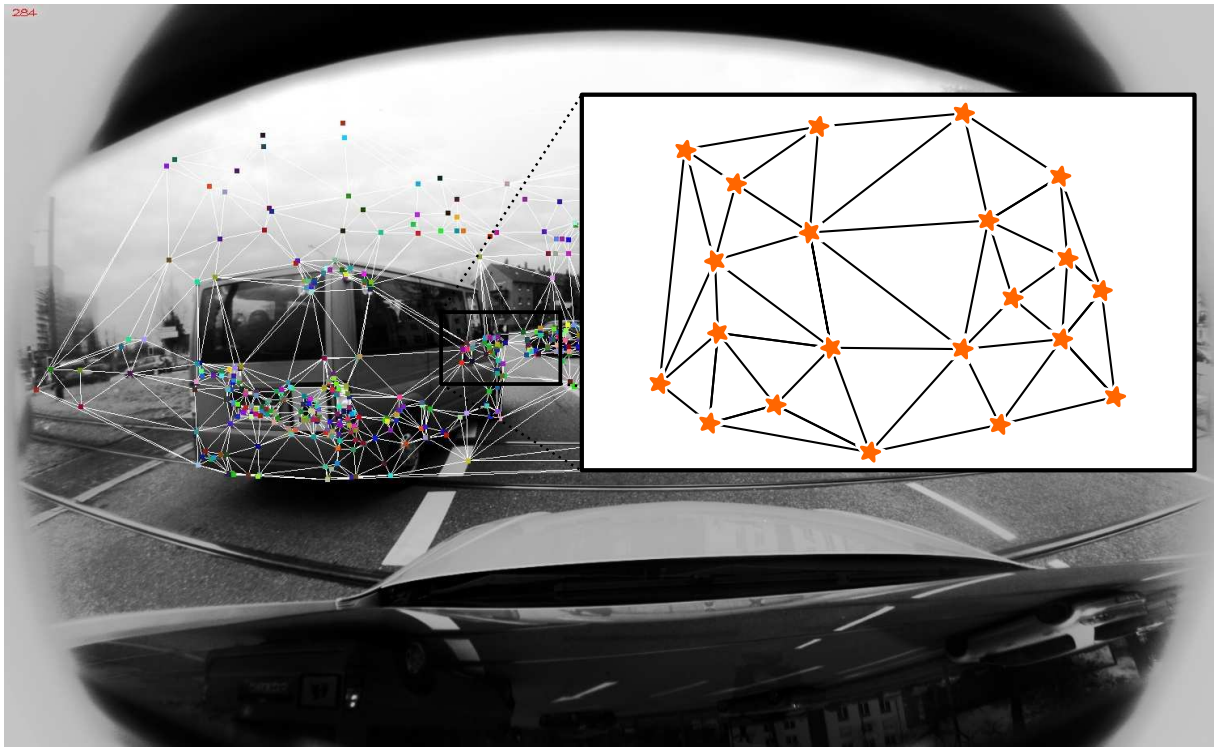


Abbildung 3.2 – Delaunay-Triangulation mit schematischer Darstellung

werden alle Atome, für die sich die Hypothese bestätigen konnte, mit ihren benachbarten starren Atomen zu Objekten zusammengefügt.

3.1.1 Atombildung

Die Nachbarschaftsbeziehung der Bildmerkmale eines Bildes wird mittels Delaunay-Triangulation bestimmt. Dabei wird der Bildraum in Dreiecke zerteilt, deren Kanten jeweils zwei Bildmerkmale verbinden. Wie später im folgenden Teilabschnitt deutlich wird, können sehr stumpfe Dreiecke für den Hypothesentest problematisch werden. Daher erzeugt die Delaunay-Triangulation insofern gut konditionierte Dreiecke, da sie den dualen Graphen des von den Bildmerkmalen bestimmten Voronoi-Graphen realisiert [9]. Dadurch wird der kleinste Winkel jedes Dreiecks maximiert, wodurch besonders stumpfe Dreiecke vermieden werden. Abbildung 3.2 veranschaulicht die Triangulation sowohl im Kamerabild, als auch schematisch.

Da die Triangulation in jedem Bild durchgeführt wird und die Landmarken zeitlich getrackt sind, entsteht somit nicht nur eine Nachbarschaft zwischen den Landmarken eines Zeitschrittes, sondern zwischen allen jemals benachbarten Landmarken.

Anhand der nun ermittelten Nachbarschaft werden Atome, also kleinste potentielle Teilsegmente der Objekte generiert. Hierbei bildet jede Landmarke mit seinen zu einem beliebigen

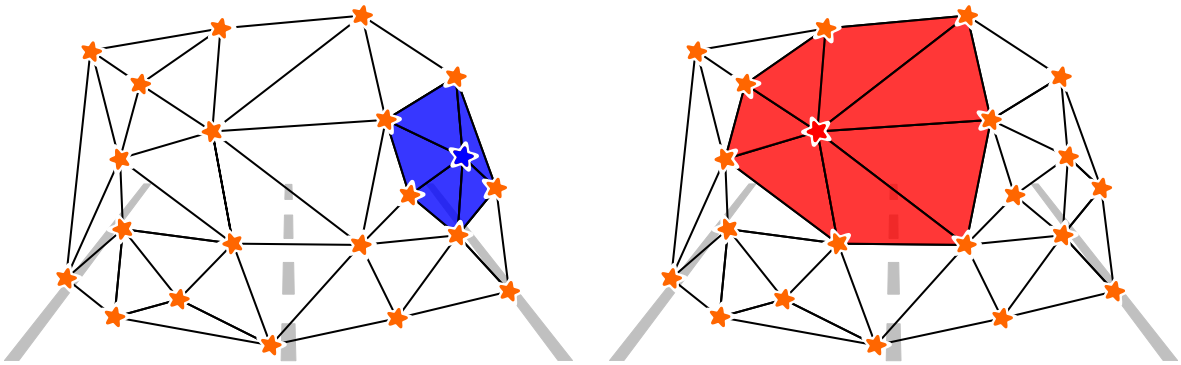


Abbildung 3.3 – Beispielatome im Bildbereich

Zeitpunkt direkt benachbarten Landmarken ein Atom. Im Schnitt besteht ein Atom somit in den Datensätzen von Kapitel 4 aus etwa 7 und maximal aus bis zu 25 Landmarken, deren Tracklets sich über durchschnittlich 7,47 Frames erstrecken. Abbildung 3.3 skizziert zwei Beispielatome im Bildbereich. Ein Objekt wird später aus zahlreichen zusammenhängenden Atomen bestehen.

Ein Atom ist zu jedem Zeitschritt (auch „Frame“ genannt) beobachtbar, in welchem mindestens drei ihrer Landmarken Beobachtungen aufweist. Da die jeweiligen Landmarken eines Atoms allerdings unterschiedlich lang beobachtet werden, werden in den ersten und letzten Frames des Atoms nur wenige seiner Landmarken beobachtet. Für den Hypothesentest im nachfolgenden Teilabschnitt wird es sinnvoll sein, die Atome auf Frames zu kürzen, in denen eine Mindestanzahl ihrer Landmarken beobachtet werden. In Kapitel 4 werden somit mindestens fünf Landmarken pro Frame eines Atoms gefordert. Atome, die daraufhin nur noch sehr wenige Frames oder von vornherein zu wenige Landmarken aufweisen, die also nicht genügend Beobachtungen für eine aussagekräftige Analyse haben, werden gänzlich verworfen.

In Abbildung 3.4 sind alle zum dargestellten Zeitpunkt beobachtbaren Atome eingezeichnet. Es wird insbesondere die hohe Anzahl an Atomen, ihre im Vergleich zu den Objekten geringe Größe und ihre starke gegenseitige Überlappung deutlich.

3.1.2 Hypothesen- und Plausibilitätstest

Für jedes Atom soll nun die Starrheitshypothese überprüft werden, mit der Absicht solche Atome herauszufiltern, die ausschließlich aus Landmarken eines einzigen Objekts bestehen. Zu diesem Zweck wird das in Abschnitt 2.3.2 erläuterte Bündelausgleichsverfahren für jedes Atom angewandt. Die Initialwerte der Landmarkenpositionen werden dabei durch eine Stereorekonstruktion im ersten Frame des Atoms bestimmt, während für die initialen Werte der Relativbewegung des Atoms die Kameraposen aus der Eigenbewegungsschätzung

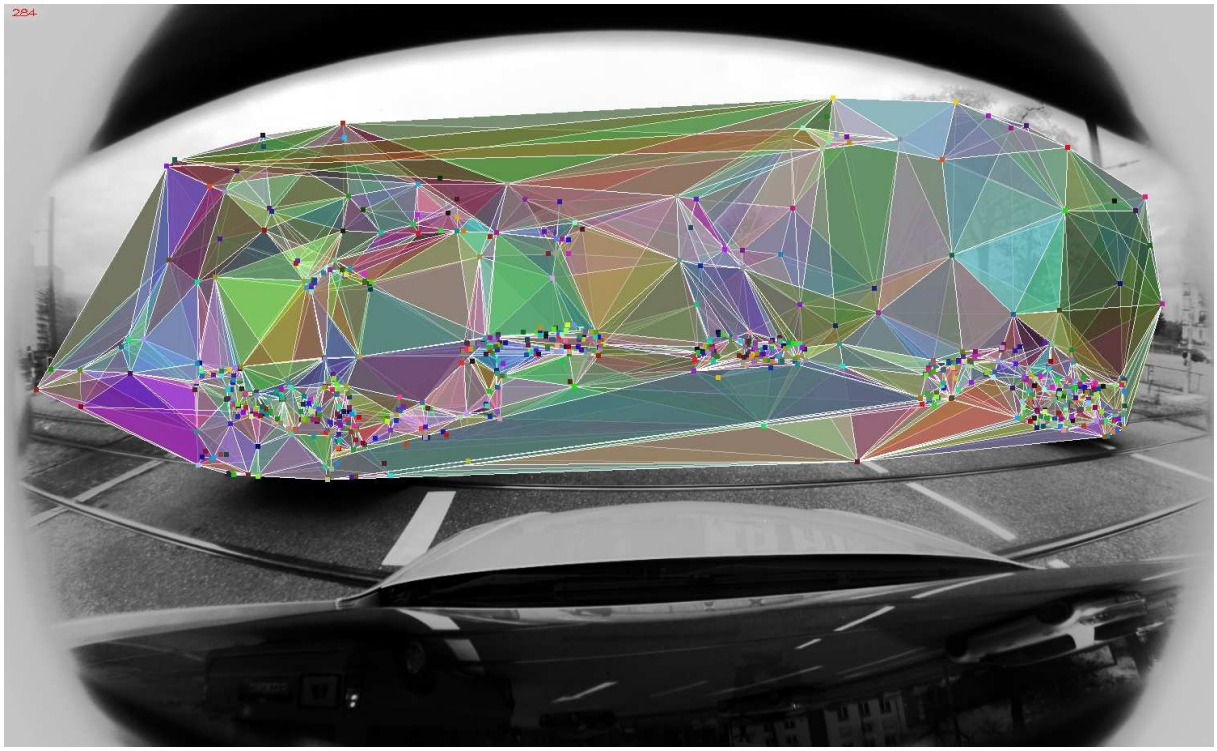


Abbildung 3.4 – Alle zum gewählten Zeitpunkt beobachtbaren Atome

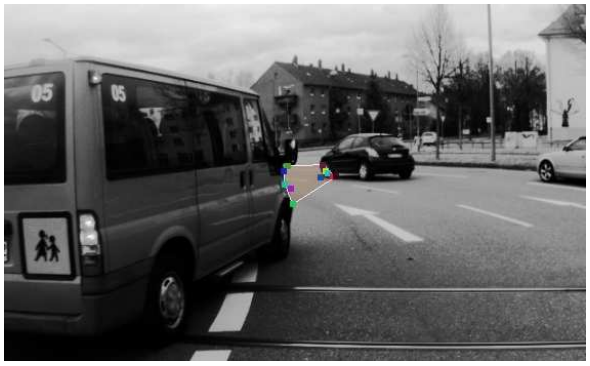
übernommen wird.

Da ein Bündelausgleich von unbewegten Landmarken ausgeht, sollte für starre Atome also eine Lösung mit geringem Schätzfehler gefunden werden, während sich die Beobachtungen objektübergreifender Atome nur schwer erklären ließen. Wie sich später zeigen wird, trifft diese Vermutung nicht immer zu, so dass zusätzlich zu den intuitiven Schätzmetriken wie Schätzfehler und -unsicherheit weitere Plausibilitätskriterien entwickelt und untersucht wurden.

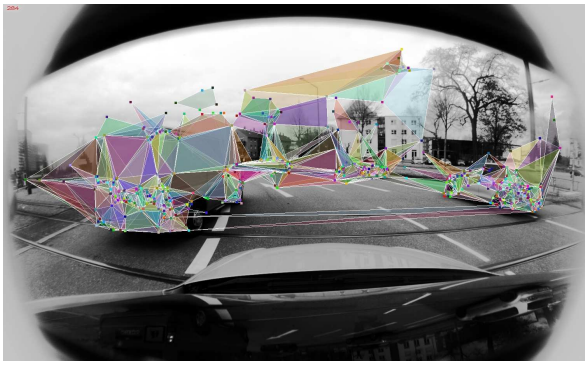
Bei objektübergreifenden Atomen beeinflussen hauptsächlich Ausreißerlandmarken die Schätzlösung so, dass diese von starren Atomen unterscheidbar werden. Daher ist der Einfluss von Ausreißern auf die Schätzung zunächst erwünscht. In diesem Schritt ist es somit wichtig, noch keine Methoden der robusten Schätzung oder Ausreißerdetektion, wie in Abschnitt 2.1.3 diskutiert, anzuwenden. Diese werden erst in Abschnitt 3.2 zum Einsatz kommen.

Die untersuchten Kriterien basieren auf unterschiedlichen Metriken, die nun einzeln vorgestellt werden. Jedes Kriterium für sich wird der zuverlässigen Unterscheidung starrer von nicht-starren Atomen nicht gerecht. Abbildung 3.5 veranschaulicht allerdings wie einige Kriterien miteinander kombiniert das Problem lösen. Jene entwickelte Kriterien, die sich als ungeeignet erwiesen haben, werden ebenfalls erläutert und ihre Schwächen begründet.

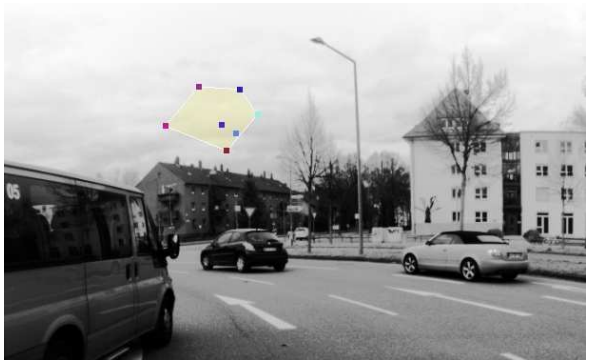
Rückprojektionsfehler Das naheliegendste Starrheitskriterium nutzt die in Glei-



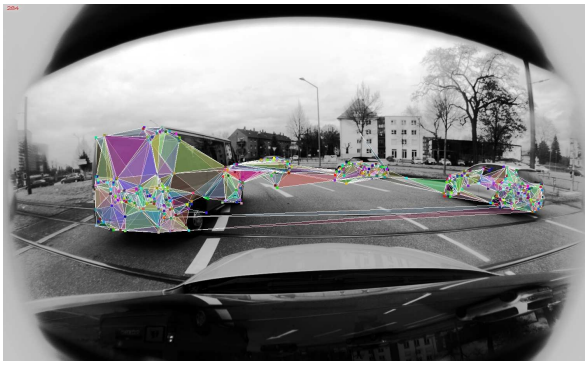
(a) Atom mit hohem Rückprojektionsfehler



(b) Gefiltert nach Rückprojektionsfehler

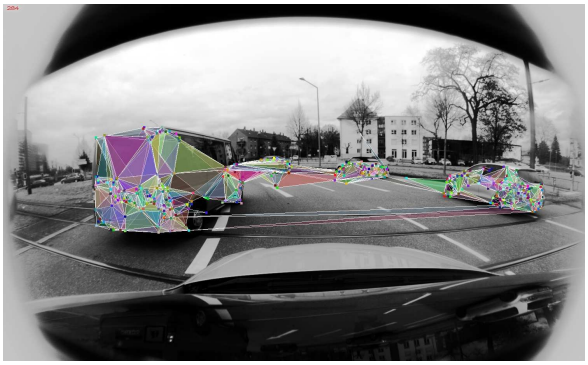


(c) Atom mit hoher Unsicherheit der Landmarkenposition



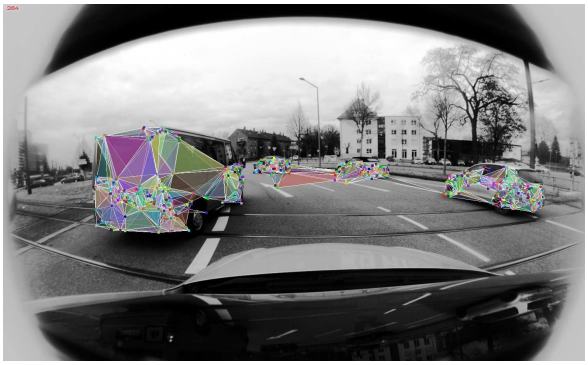
(d) Zusätzlich nach Unsicherheit der Landmarkenposition gefiltert

(e) Atom mit hoher Unsicherheit der Kameraposition



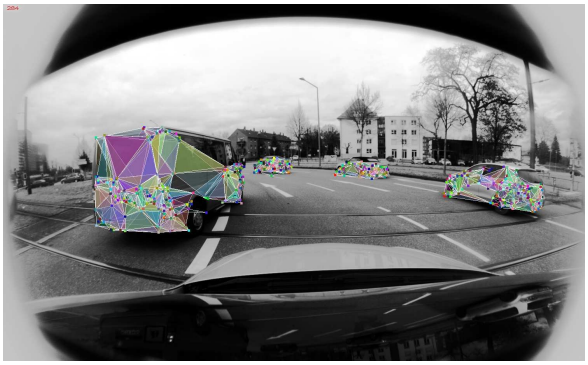
(f) Zusätzlich nach Unsicherheit der Kameraposition gefiltert

(g) Atom mit hoher Axialität



(h) Zusätzlich nach Axialität gefiltert

(i) Atom mit hoher Ausdehnung



(j) Zusätzlich nach Ausdehnung gefiltert

Abbildung 3.5 – Hypothesen- und Plausibilitätstest mit dem Filterergebnis auf der linken Seite und jeweils einem Beispielatom im rechten Bild

chung 2.36 verwendete Kostenfunktion des Bündelausgleichs, also den quadratischen Rückprojektionsfehler, als Metrik (vgl. Gleichung (2.4)):

$$f_2 := \frac{1}{2} \sum_{(i,j)\in\Psi} \|\mathbf{w}_{ij} - \hat{\mathbf{w}}_{ij}\|^2 = \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|_2^2. \tag{3.1}$$

Besonders einleuchtend ist, dass nicht-starre Atome, deren Landmarken, wie in Abbildung 3.5(a), zu mehreren Objekten gehören oder zum Großteil stark verrauscht sind, wesentlich mehr Beobachtungen mit großem Residuum und somit einen insgesamt höheren quadratischen Rückprojektionsfehler aufweisen müssten.

Bezeichnet man die Menge aller Atome mit $\mathcal{A}$, kann also dieser erste Hypothesentest

$$\hat{\mathcal{A}}_{\text{starr},1} = \{a \in \mathcal{A} \mid f_2(a) \leq f_{2,\max}\} \tag{3.2}$$

die Menge der starren Atome $\mathcal{A}_{\text{starr}}$ annähern. Damit wird bereits, wie in Abbildung 3.5(b) zu sehen ist, eine große Zahl nicht-starrer Atome herausgefiltert. Allerdings reicht dieses Kriterium selbst, wie schon angedeutet, nicht aus, um nicht-starre Atome zuverlässig herauszufiltern. Insbesondere Atome mit nur einer objektfremden Landmarke oder ausschließlich verrauschten Hintergrundlandmarken, wie solchen auf Wolken, Gebäuden oder der Fahrbahn, werden durch dieses Kriterium nicht vollständig erfasst.

Die Parametrierung eines geeigneten Wertes für $f_{2,\max}$ und die noch folgenden Grenzwerte wird in Abschnitt 3.1.4 erläutert.

Viele Atome, die überwiegend aus Landmarken eines Objekts und beispielsweise nur einer Landmarke eines anderen Objekts bestehen (nachfolgend als „fast-starre Atome“ bezeichnet), weisen einen erstaunlich unauffälligen quadratischen Rückprojektionsfehler auf. Dies rührt daher, dass die Residuen dieser einen Landmarke in der ganzen Residuensumme

trotz Quadrierung wenig ins Gewicht fallen, wenn das Atom besonders viele korrekte Beobachtungen hat (bspw. bei vergleichsweise vielen Landmarken oder Frames).

Einige sogenannte „quasi-weltfeste Landmarken“ liegen beispielsweise zwar auf einer weltfesten Oberfläche (z.B. der Fahrbahn), sind aber dennoch Ausreißer der Eigenbewegungsschätzung (siehe Abschnitt 2.4), weil sie sich zeitweise unter anderem mit den Fahrzeugschatten mitbewegen. Fast-starre Atome, die solche quasi-weltfesten Landmarken beinhalten, sind besonders schwierig zu identifizieren, weil ein Großteil der Beobachtungen, vor allem in den Frames in denen sich diese Landmarke mitbewegt, durchaus plausibel erscheinen.

Ähnlich problematisch sind Landmarken, die aus Assoziationsfehlern entstehen und sich zunächst auf Objekt A mitbewegen, nach einigen Frames aber auf Objekt B überspringen und sich dann mit diesem mitbewegen. Fast-starre Atome, die aus Landmarken des Objekts A und einer solchen fehlassoziierten Landmarke bestehen, erscheinen starr, bis auf die wenigen Frames, in welchen sich jene Landmarke auf Objekt B mitbewegt. Aus dem selben Grund erscheinen ebenso Atome mit Landmarken des Objekts B und dieser fehlassoziierten Landmarke nahezu starr. Somit besteht die Gefahr, dass diese Atome nicht als nicht-starre Atome erkannt und daher nicht verworfen werden, sondern in die Komponentenbildung (siehe Abschnitt 3.1.3) einfließen und durch die gemeinsame Landmarke beide Objekte zu einem verschmelzen.

Ausgehend von Gleichung (3.1) und dieser Beobachtung sollen in zwei zusätzlichen Kriterien Residuen, deren Komponenten vereinzelt ungewöhnlich hoch sind, mittels einer p-Norm stärker gewichtet werden:

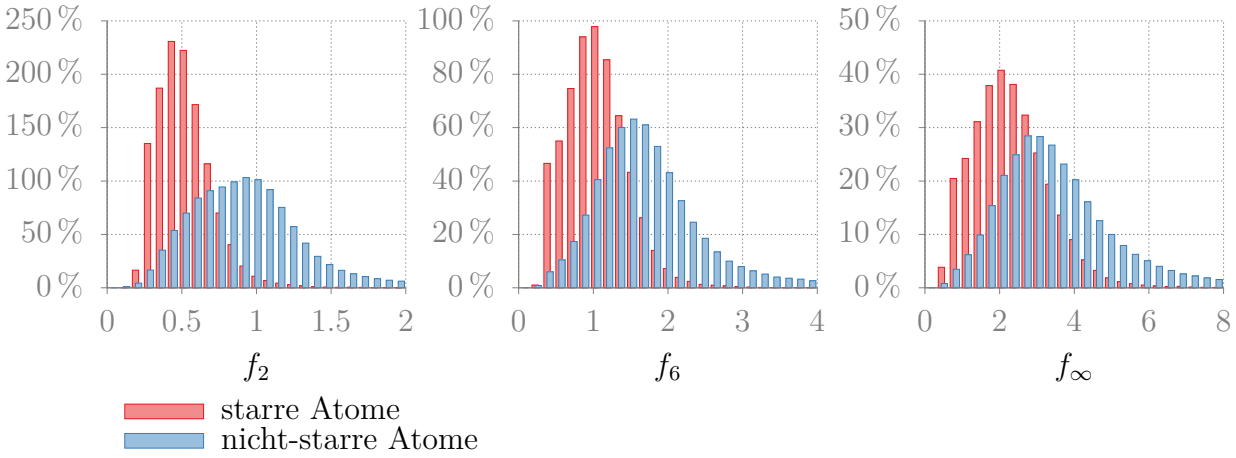
p-Norm Vektornormen der Form

$$\|\mathbf{v}\|_p = \left(\sum_i |v_i|^p \right)^{\frac{1}{p}} \tag{3.3}$$

mit reelem Skalar $p \geq 1$ werden als p-Normen bezeichnet und sind eine Verallgemeinerung der euklidischen Norm [4]. Eine p-Norm mit $p > 2$ gewichtet große Residuen stärker als die euklidische Norm, so dass sie für das gesuchte Kriterium als Metrik verwendet werden kann:

$$f_p := \frac{1}{\#\Psi} \sum_{(i,j)\in\Psi} \|\mathbf{w}_{ij} - \hat{\mathbf{w}}_{ij}\|_p^p = \frac{1}{\#\Psi} \|\mathbf{r}(\mathbf{x})\|_p^p. \tag{3.4}$$

Die Normalisierung mit der Anzahl an Beobachtungen $\#\Psi$ erlaubt es, Atome mit vielen Beobachtungen besser mit solchen mit wenigen Beobachtungen vergleichen zu können.



(a) Euklidische Norm des Residuenvektors **(b)** 6-Norm des Residuenvektors **(c)** Maximumsnorm des Residuenvektors

Abbildung 3.6 – Relative Häufigkeitsdichte der normbasierten Starrheitskriterien im Vergleich (annotierte Atome aus $\mathcal{A}$)

Maximumsnorm Die Maximumsnorm ist, wie auch die euklidische und Summennorm, eine der wichtigsten Spezialfälle der p-Norm und entsteht aus dem Grenzwert $p \to \infty$. Sie ist somit durch

$$\|\mathbf{v}\|_\infty = \max_i |v_i| \tag{3.5}$$

definiert [4]. Hieraus lässt sich folgendes Kriterium ableiten, das dem betragsmäßig größten Residuum entspricht:

$$f_\infty := \|\mathbf{r}(\mathbf{x})\|_\infty = \max_{(i,j)\in\Psi} |r_{ij}(\mathbf{x})|. \tag{3.6}$$

Leider erreichte keines der beiden Kriterien $f_p(a) \leq f_{p,\max}$ (mit $p > 2$) und $f_\infty(a) \leq f_{\infty,\max}$ in den Datensätzen aus Kapitel 4 das angestrebte Ziel fast-starre Atome effektiv zu entfernen. Zudem ist, verglichen mit dem quadratischen Rückprojektionsfehler, keines davon besser dazu geeignet, starre von nicht-starren Atomen zu trennen, wie in den Histogrammen in Abbildung 3.6 deutlich wird.

Dort ist die relative Häufigkeitsdichte starrer im Vergleich zu nicht-starren Atomen nach den Metriken f_2, f_6 (als Beispielinstanz von f_p) und f_∞ aufgetragen. Wie zu erkennen ist, überlappen sich die beiden Verteilungen unter Verwendung der f_6- oder f_∞-Metrik stärker als unter f_2 und weisen somit einen größeren Klassifikationsfehler auf. Die Zuordnung der Atome in starre und nicht-starre Atome wurde für diese Statistik auf einer Sequenz von 1500 Stereobildern manuell vorgenommen (siehe Abschnitt 3.1.4).

Wegen ihrer Untauglichkeit wurden beide Kriterien nach f_p und f_∞ zurückgestellt und

statt dessen geeignetere Kriterien gesucht.

Der unauffällig geringe Rückprojektionsfehler solcher fast-starren Atome kann nur durch eine schlechte Beobachtbarkeit der entsprechenden Landmarkenpositionen erklärt werden. Somit liegt die Vermutung nahe, dass diese Landmarkenpositionen eine große Unsicherheit der Rekonstruktion aufweisen.

Unsicherheit der Landmarkenpositionen Wie bereits in Abschnitt 2.3.3 ausgeführt, kann die Kovarianz $\mathbf{C}$ der Least-Squares Schätzung aus der Fisher-Information (2.38) bestimmt werden, welche der bereits berechneten Hesseapproximation des Gauß-Newton- oder Levenberg-Marquardt-Schätzers entspricht. Für die Unsicherheit der Landmarkenpositionen sind hiervon lediglich die Blockmatrizen $\mathbf{C}_{\mathbf{p}_i}$ der Kovarianz relevant. Für jede Landmarke wird schließlich mittels Eigenwertanalyse von $\mathbf{C}_{\mathbf{p}_i}$ die Varianz $\lambda_{\mathbf{p}_i,\max}$ entlang ihrer ersten Hauptachse bestimmt, womit das Maximum dieser Landmarkenvarianzen als weitere Metrik definiert werden kann:

$$\breve{\sigma}_p := \max_i \{\lambda_{\mathbf{p}_i,\max}\}\,; \qquad \text{mit } \lambda_{\mathbf{p}_i,\max} \text{ max. EW von } \mathbf{C}_{\mathbf{p}_i}. \tag{3.7}$$

Tatsächlich erweist sich die maximale Unsicherheit der Landmarkenposition $\breve{\sigma}_{\mathbf{p}}$ als gutes Kriterium, um zahlreiche Atome herauszufiltern, die verrauschte Hintergrundlandmarken, insbesondere solche auf weit entfernten Gebäuden oder Wolken, wie beispielsweise in Abbildung 3.5(c), beinhalten. Wendet man den aktualisierten Hypothesentest

$$\hat{\mathcal{A}}_{\text{starr},2} = \{a \in \mathcal{A} \mid f_2(a) \leq f_{2,\max} \;\&\; \tag{3.8a}$$

$$\breve{\sigma}_{\mathbf{p}}(a) \leq \breve{\sigma}_{\mathbf{p},\max}\} \tag{3.8b}$$

an, so verbleiben fast ausschließlich objektfeste Landmarken, die allermeisten starren Atome und noch einige Atome, die zwei Objekte verbinden oder quasi-weltfeste Landmarken aufweisen. Die verbleibenden Atome $\hat{\mathcal{A}}_{\text{starr},2}$ sind in Abbildung 3.5(d) eingezeichnet.

Analog wurde auch die Unsicherheit der Relativbewegung, sprich der Kameraposition im Bündelausgleich, als zusätzliches Kriterium untersucht. Es zeigt sich zwar nicht so effektiv wie $\breve{\sigma}_{\mathbf{p}}$, identifiziert allerdings einige fast-starre Atome, wie jenes in Abbildung 3.5(e).

Unsicherheit der Relativbewegung Trotz der in Abschnitt 2.3.3 erwähnten fehlenden theoretischen Interpretation von Kovarianzen von Posen, konnte experimentell bestätigt werden, dass eine Betrachtung der Unsicherheit der translatorischen Relativbewegung einige fast-starre Atome entfernen kann ohne dabei fälschlicherweise zu viele starre Atome herauszufiltern. Das Kriterium wird analog zur Unsicherheit der Landmarkenpositionen mittels Eigenwertanalyse der Blockmatrizen $\mathbf{C}_{\mathbf{t}_i}$ (vgl. Gleichung (2.40)) bestimmt, wobei die Unsicherheit der Rotation außer Acht gelassen

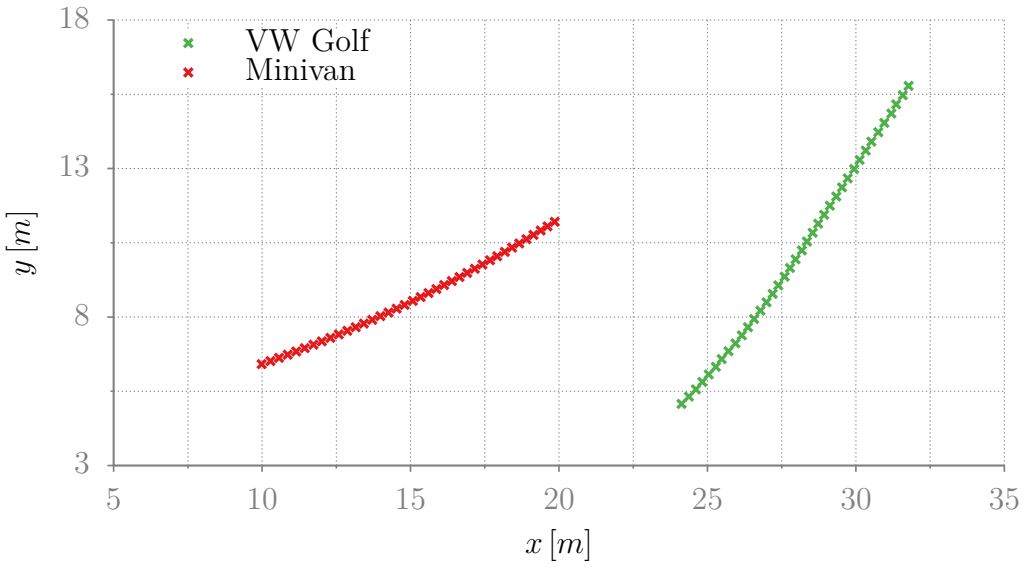


Abbildung 3.7 – Trajektorien des Kleinbusses und VW Golf, im Zeitintervall des Atoms aus Abbildung 3.5(g)

wird:

$$\breve{\sigma}_t := \max_i \{\lambda_{\mathbf{t}_i,\max}\}; \quad \text{mit } \lambda_{\mathbf{t}_i,\max} \text{ max. EW von } \mathbf{C}_{\mathbf{t}_i}. \tag{3.9}$$

Somit ergibt sich der erweiterte Hypothesentest

$$\hat{\mathcal{A}}_{\text{starr},3} = \{a \in \mathcal{A} \mid f_2(a) \leq f_{2,\max} \;\&\; \tag{3.10a}$$

$$\breve{\sigma}_{\mathbf{p}}(a) \leq \breve{\sigma}_{\mathbf{p},\max} \;\&\; \tag{3.10b}$$

$$\breve{\sigma}_{\mathbf{t}}(a) \leq \breve{\sigma}_{\mathbf{t},\max}\}, \tag{3.10c}$$

dessen Ergebnis in Abbildung 3.5(f) dargestellt ist. Offensichtlich sind noch viele räumlich ausgedehnte sowie einige langgestreckte nicht-starre Atome in $\hat{\mathcal{A}}_{\text{starr},3}$ verblieben. Zudem sind mit den bisher untersuchten Kriterien die Möglichkeiten der Residuen- und Kovarianz-basierten Schätzmetriken ausgeschöpft, so dass weitere Plausibilitätstests basierend auf einer gründlichen Untersuchung der verbliebenen Atome entwickelt wurden.

Hierfür sollen zunächst solche langgestreckten nicht-starren Atome, wie in Abbildung 3.5(g), näher betrachtet werden. Jenes Atom besteht aus einer Landmarke des linken Kleinbusses und sieben Landmarken des, zwei Fahrspuren rechts davon entfernten, VW Golf. Der Golf folgt seiner Fahrspur, während der Kleinbus die Spur wechselt, so dass sich beide Fahrzeuge mit einem Winkel von knapp 30° zueinander bewegen. Zudem fährt der Golf mit etwas höherer Geschwindigkeit als der Kleinbus. Zur Veranschaulichung sind vorausgreifend die rekonstruierten Trajektorien der beiden Fahrzeuge im Zeitintervall, in welchem das betrachtete Atom beobachtbar ist, in Abbildung 3.7 dargestellt.

Trotz der zwar ähnlichen, aber dennoch erkennbar unterschiedlichen Trajektorien der beiden Objekte und somit auch Landmarken, weist dieses Atom sehr niedrige Rückprojektionsfehler f_2 und Unsicherheiten $\breve{\sigma}_{\mathbf{p}}$ und $\breve{\sigma}_{\mathbf{t}}$ auf. Die scheinbar plausible Schätzung der Struktur und Relativbewegung des Atoms über mehr als 30 Frames hinweg, ist erst durch diese spezifische Anordnung der Landmarken trotz der Annahme, dass die Landmarken zueinander unbewegt sind, möglich. Eine Analyse der durch den Bündelausgleich rekonstruierten Atomstruktur, wie in den projektiven Abbildungen 3.8, liefert eine Erklärung.

Beim Vergleich von Abbildung 3.8(a) und 3.8(b) fällt auf: Dadurch dass die sieben Landmarken des rechten Fahrzeugs sehr nah beieinander, aber sehr weit von der Landmarke des linken Fahrzeugs liegen, lässt sich eine starke Bewegung der linken Landmarke über eine kleine Rotation des Atoms erklären, ohne dass sich dabei die Rückprojektionen der restlichen Landmarken signifikant verändern. Die in Abschnitt 2.3.3 diskutierte Tiefenunsicherheit führt in solchen Fällen folglich zu einer starken Rotationsunsicherheit der Atome. Abbildung 3.8(c) zeigt zur Verdeutlichung der 3-dimensionalen Struktur das selbe Atom aus einer anderen Perspektive.

Dies soll mit Abbildung 3.9 schematisch untermauert werden, wo die Szene aus der Draufsicht dargestellt ist. Unter der Annahme, dass das Atom starr ist, führt die axiale Form des Atoms bei der Rotation dazu, dass die Landmarken des grünen Fahrzeugs ihre Position nur leicht ändern, während die Landmarke des roten Fahrzeugs eine große Distanz zurücklegt. Somit hat diese Rotation nur eine geringe Auswirkung auf die Rückprojektionen der Landmarken und führt zu einem insgesamt kleinen Rückprojektionsfehler, trotz nicht starrem Atom.

Da dieser Effekt vor allem bei sehr langgestreckten Atomen zu beobachten ist, wird im ersten Plausibilitätskriterium die rekonstruierte Atomstruktur auf besonders starke Axialität hin geprüft.

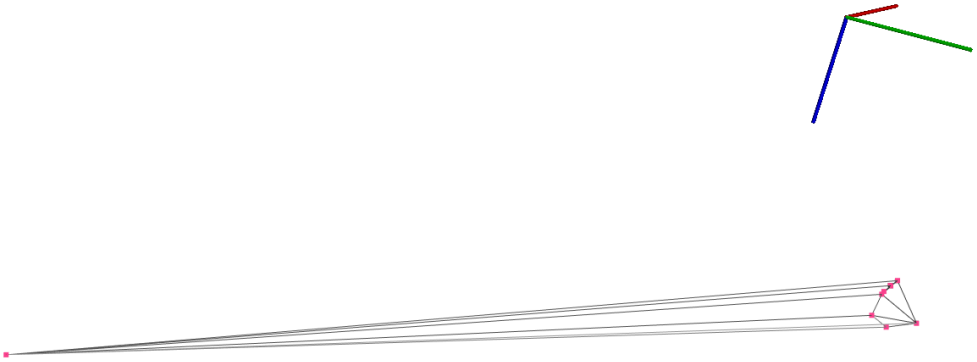
Axialität Die Axialität, oder auch Achsigkeit, eines Atoms wird aus ihrer Strukturrekonstruktion, also den Positionen ihrer Landmarken mittels Hauptkomponentenanalyse [20] bestimmt. Hierzu wird die Kovarianz der räumlichen Verteilung der Landmarkenpositionen $\mathbf{C}_{\mathbf{p}}$ bestimmt (nicht zu verwechseln mit der Kovarianz der Landmarkenpositionsschätzung $\mathbf{C}_{\mathbf{p}_i}$). Aus den Landmarkenpositionen $\mathbf{p}_i$ und ihrem Mittelwert

$$\boldsymbol{\mu}_{\mathbf{p}} = \frac{1}{m} \sum_{i=0}^{m-1} \mathbf{p}_i \tag{3.11}$$

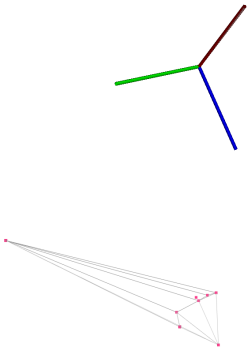
kann die Matrix der zentrierten Landmarkenpositionen berechnet werden:

$$\mathbf{P} = (\mathbf{p}_0 - \boldsymbol{\mu}_{\mathbf{p}}, \mathbf{p}_1 - \boldsymbol{\mu}_{\mathbf{p}}, \ldots, \mathbf{p}_{m-1} - \boldsymbol{\mu}_{\mathbf{p}})^T. \tag{3.12}$$

(a) Stark axiales Atom

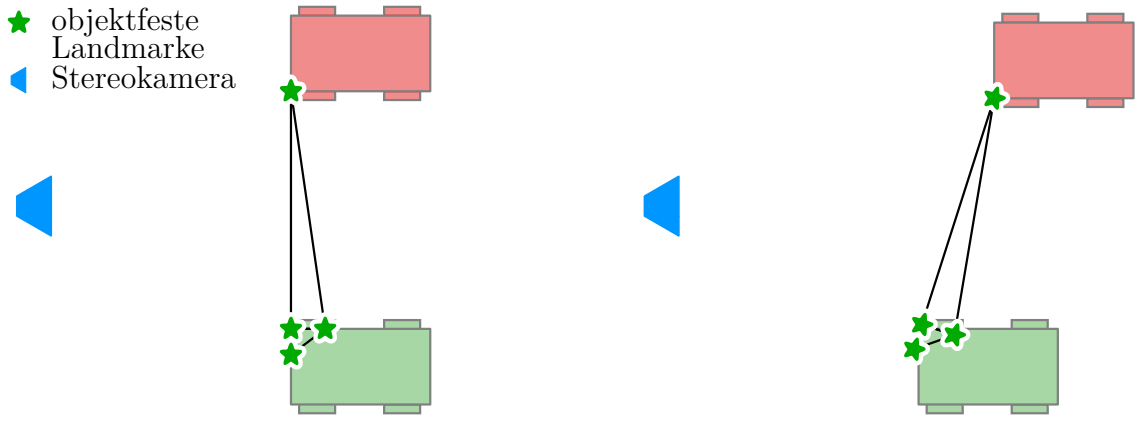


(b) nach kleinster Drehung



(c) aus einer anderen Perspektive

Abbildung 3.8 – Stark axiales Atom aus unterschiedlichen Blickwinkeln



(a) Die starke Bewegung der oberen Landmarke ... **(b)** kann durch eine kleine Rotation des starren Atoms erklärt werden.

Abbildung 3.9 – Tiefenunsicherheit der Landmarken führt zu Rotationsunsicherheit der Atome

Die Kovarianz $\mathbf{C_p}$ ergibt sich dann aus

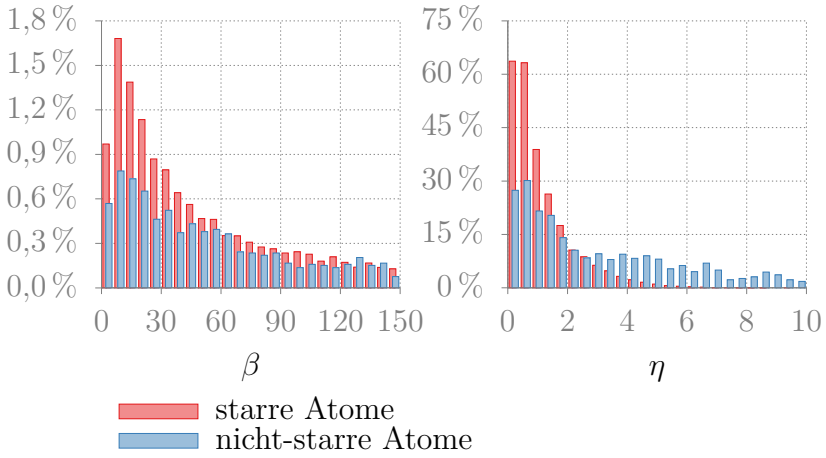
$$\mathbf{C_p} = \frac{1}{m-1}\,\mathbf{P}^T\mathbf{P}. \tag{3.13}$$

Die Hauptkomponenten werden schließlich, ähnlich wie für $\breve{\sigma}_p$, aus den Eigenwerten und -vektoren von $\mathbf{C_p}$ bestimmt. Die Axialität eines Körpers sei nun als Verhältnis des betragsmäßig größten zum zweitgrößten Eigenwert, also der Ausdehnung entlang der ersten Hauptachse relativ zur Ausdehnung entlang der zweiten Hauptachse, definiert:

$$\alpha = \frac{\lambda_{\mathbf{p}_i,\max}}{\lambda_{\mathbf{p}_i,\mathrm{med}}}; \qquad \text{mit } \lambda_{\mathbf{p}_i,\max} > \lambda_{\mathbf{p}_i,\mathrm{med}} > \lambda_{\mathbf{p}_i,\min} \text{ EW von } \mathbf{C_p}. \tag{3.14}$$

In der Praxis eignet sich die Axialität sehr gut als Kriterium, um etwa die Hälfte aller verbliebenen nicht-starren Atome zu eliminieren, wie in Abbildung 3.5(h) deutlich wird. Somit sei der Hypothesentest von $\hat{\mathcal{A}}_{\mathrm{starr},3}$ um den α-basierten Plausibilitätstest erweitert zu:

$$\hat{\mathcal{A}}_{\mathrm{starr},4} = \{a \in \mathcal{A} \mid f_2(a) \leq f_{2,\max} \;\&\; \tag{3.15a}$$
$$\breve{\sigma}_{\mathbf{p}}(a) \leq \breve{\sigma}_{\mathbf{p},\max} \;\&\; \tag{3.15b}$$
$$\breve{\sigma}_{\mathbf{t}}(a) \leq \breve{\sigma}_{\mathbf{t},\max} \;\&\; \tag{3.15c}$$
$$\alpha(a) \leq \alpha_{\max}\}. \tag{3.15d}$$



(a) Planarität der Atome **(b)** Ausdehnung der Atome

Abbildung 3.10 – Relative Häufigkeitsdichte der Plausibilitätskriterien nach Planarität und Ausdehnung der Atome (annotierte Atome aus $\hat{\mathcal{A}}_{\text{starr},4}$)

Nun sind nur noch einige großflächige Atome in $\hat{\mathcal{A}}_{\text{starr},4}$ verblieben. Um insbesondere solche großflächigen und zugleich flachen Atome, wie in Abbildung 3.5(i), zu behandeln, wurde analog zur Axialität auch die Planarität der Atome als Kriterium untersucht.

Planarität Die Planarität eines Atoms wird, wie bereits schon die Axialität, aus der Strukturrekonstruktion des Atoms mittels Hauptkomponentenanalyse bestimmt. Jedoch wird die Planarität als Verhältnis des betragsmäßig größten zum kleinsten Eigenwert definiert:

$$\beta = \frac{\lambda_{\mathbf{p}_i,\max}}{\lambda_{\mathbf{p}_i,\min}}; \qquad \text{mit } \lambda_{\mathbf{p}_i,\max} > \lambda_{\mathbf{p}_i,\text{med}} > \lambda_{\mathbf{p}_i,\min} \text{ EW von } \mathbf{C}_\mathbf{p}. \tag{3.16}$$

Zwar können anhand der Planarität β viele verbliebene nicht-starre Atome gerausgefiltert werden, allerdings trifft dieses Kriterium auch auf viele starre Atome zu. Dies wird aus dem Histogramm 3.10(a), das die relative Häufigkeitsdichte der verbliebenen Atome aus $\hat{\mathcal{A}}_{\text{starr},4}$ bezüglich ihres Planarfaktors β aufträgt, ersichtlich. Immerhin bestehen die Karosserien von Fahrzeugen häufig aus ebenen Oberflächensegmenten, besonders deutlich beim Kleinbus in Abbildung 3.1(a). Die Auswirkungen eines Kriteriums $\beta(a) < \beta_{\max}$ auf starre Atome ist so gravierend, dass die Objekte dadurch bei der Komponentenbildung (siehe Abschnitt 3.1.3) in mehrere Teile zerfallen. Daher wurde dieses Kriterium nicht in den Hypothesen- und Plausibilitätstest eingebunden.

Statt dessen werden die nicht-starren Atome in $\hat{\mathcal{A}}_{\text{starr},4}$ anhand ihrer räumlichen Ausdehnung gefiltert.

Räumliche Ausdehnung Die räumliche Ausdehnung eines Atoms wird aus ihrer Struk-

turrekonstruktion als maximaler euklidischer Abstand zweier seiner Landmarken zueinander definiert:

$$\eta = \max_{i \neq i'} \|\mathbf{p}_i - \mathbf{p}_{i'}\|. \tag{3.17}$$

Das Histogramm in Abbildung 3.10(b) belegt, dass sich die Ausdehnung gut dazu eignet, weitere nicht-starre Atome zu verwerfen.

Es sei betont, dass die Einschränkung der Ausdehnung der Atome nicht einem Ausschluss großer Objekte gleich kommt. Vielmehr ergibt sich daraus die Forderung, dass große Objekte ausreichend dicht mit Landmarken belegt sein müssen, um erkannt zu werden. Diese Forderung erweist sich als hinnehmbar, solange $\eta_{\max}$ nicht sehr klein gewählt wird. Große Objekte, die nur wenige Landmarken aufweisen, sind in der Regel sehr weit entfernt, so dass sie die übrigen Tests (v.A. $\breve{\sigma}_{\mathbf{p}}(a) \leq \breve{\sigma}_{\mathbf{p},\max}$) ohnehin nicht bestehen würden. Somit kann der Hypothesen- und Plausibilitätstest von $\hat{\mathcal{A}}_{\text{starr},4}$ bedenkenlos um die Ausdehnung $\eta(a)$ erweitert werden:

$$\begin{aligned}
\hat{\mathcal{A}}_{\text{starr},5} = \{a \in \mathcal{A} \mid\ & f_2(a) \leq f_{2,\max} \ \& && \text{(3.18a)}\\
& \breve{\sigma}_{\mathbf{p}}(a) \leq \breve{\sigma}_{\mathbf{p},\max} \ \& && \text{(3.18b)}\\
& \breve{\sigma}_{\mathbf{t}}(a) \leq \breve{\sigma}_{\mathbf{t},\max} \ \& && \text{(3.18c)}\\
& \alpha(a) \leq \alpha_{\max} \ \& && \text{(3.18d)}\\
& \eta(a) \leq \eta_{\max}\}. && \text{(3.18e)}
\end{aligned}$$

Der Ausdehnungsfaktor $\eta(a)$ zeigt sich als besonders mächtiges Kriterium, um nicht-starre Atome zu identifizieren. Außerdem ist es das einzige Kriterium, das in der Lage ist, objektübergreifende aber starre Atome herauszufiltern. Bewegen sich nämlich zwei Objekte im Beobachtungsintervall eines Atoms, das beide Objekte verbindet, gleich (im Sinne ihrer Geschwindigkeit und Bewegungsrichtung), so ist dieses objektübergreifende Atom tatsächlich starr, aber für die Komponentenbildung ebenso unerwünscht, wie alle anderen objektübergreifenden Atome.

In einem lernbasierten Verfahren könnte die Ausdehnung als Tuningparameter betrachtet und vermutlich so scharf eingestellt werden, dass sie einige der bisherigen Kriterien sogar überflüssig machen würde. Allerdings ist der optimale Schwellwert dann sehr stark von den Eigenschaften der Eingangsdaten, wie der Merksmalsdichte, abhängig. Da in dieser Arbeit allerdings ein Verfahren ohne Lernmethoden, die unter anderem viele manuell annotierte Trainingsdaten erfordern, entwickelt werden sollte, werden die Parameter der einzelnen Kriterien, insbesondere $\eta_{\max}$, physikalisch motiviert gewählt. Somit ist für die Verarbeitung eines neuen Datensatzes eine im Vergleich zu lernbasierten Verfahren nur geringfügige bis keine Anpassung bereits bestimmter Parameter erforderlich. Abschnitt 3.1.4 beschäftigt

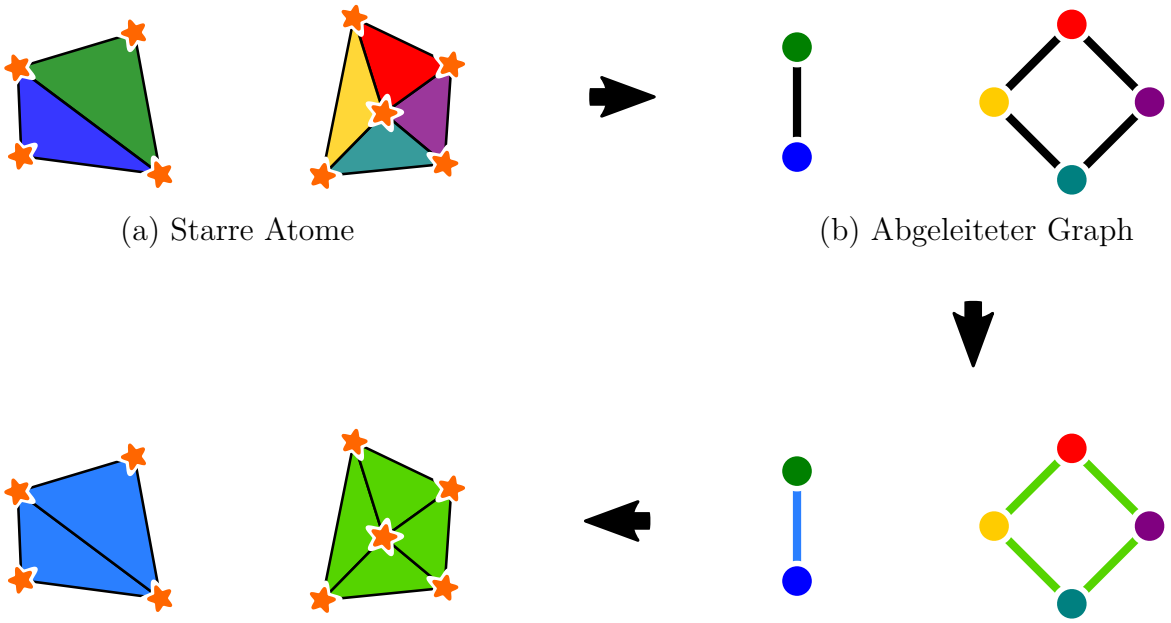
sich ausführlicher mit der Parametrierung der Kriterien.

Zwar enthält $\hat{\mathcal{A}}_{\text{starr},5}$ immer noch vereinzelte nicht-starre oder objektübergreifende Atome, bildet nun aber eine sehr zufriedenstellende Grundlage, um den Großteil der Objekte voneinander zu isolieren, wie in Abbildung 3.5(j) zu erkennen ist. Daher soll für die folgenden Verarbeitungsschritte $\hat{\mathcal{A}}_{\text{starr}} = \hat{\mathcal{A}}_{\text{starr},5}$ als beste Näherung der Menge starrer Atome verwendet werden.

3.1.3 Komponentenbildung

Alle Atome aus $\hat{\mathcal{A}}_{\text{starr}}$ für die sich also bestätigen konnte, dass sie mit hoher Wahrscheinlichkeit aus Landmarken eines starren Objekts bestehen, werden nun zu Objekten zusammengefügt. Haben zwei starre Atome gemeinsame Landmarken, folgt aus der Starrheitshypothese der Objekte, dass die Landmarken beider Atome nicht nur innerhalb der Atome selbst, sondern auch zu den Landmarken des anderen Atoms unbewegt sind. Somit ist die Vereinigung beider Atome, also die Vereinigung aller ihrer Landmarken, ebenfalls starr. Demnach können alle benachbarten Atome zu starren Objekten zusammengefügt werden.

Hierzu sollen die Atome noch einmal in der vereinfachten Darstellung in Abbildung 3.11(a) betrachtet werden, wobei diesmal jedes Atom zur besseren Übersicht aus nur drei Landmarken besteht. Es sind sechs Atome dargestellt, welche visuell sofort zwei Objektgruppen zugeordnet werden können. Diese Zusammengehörigkeit der Atome lässt sich rechnerge-



(a) Starre Atome (b) Abgeleiteter Graph

(c) Zusammenhangskomponenten (d) Aggregierte Objekte

Abbildung 3.11 – Aggregation der Landmarken zu Objekten

Abbildung 3.12 – Ergebnis der Objektaggregation

stützt besonders einfach in einem Graphen, der die Nachbarschaftsbeziehung der Atome repräsentiert, ermitteln. Hierzu wird für jedes Atom ein Knoten im Graphen erzeugt und Atome, die mindestens eine gemeinsame Landmarken teilen, mit einer Kante verbunden (siehe Abbildung 3.11(b)).

Mittels der Methode der Zusammenhangskomponenten (engl. „connected components“) [3] lassen sich im Graphen zusammenhängende Komponenten, also maximale Teilgraphen deren Knoten paarweise durch einen Pfad verbunden sind, in linearer Zeit ermitteln (siehe Abbildung 3.11(d)). Für jede Komponente werden schließlich die Landmarken ihrer Atome zu einem Objekt aggregiert, wie schematisch in Abbildung 3.11(c) und im Kamerabild 3.12 dargestellt.

Wie in Abschnitt 3.1.1 erwähnt, sind zwei Atome auch dann miteinander benachbart, wenn sie zwar zu keinem Zeitpunkt beide beobachtbar sind, sie aber mindestens eine gemeinsame Landmarke und somit ein gemeinsames Tracklet teilen, da sich dessen Beobachtungsintervall demnach mit denen der beiden Atome überlappt. Dies soll anhand Abbildung 3.13 veranschaulicht werden. Dort ist dargestellt, zu welchen Zeitpunkten jede Landmarke beobachtet wird und welche Landmarken benachbart sind. Dabei sind die Landmarken bzw. ihre Tracklets mit einer eindeutigen Farbe gekennzeichnet und die in diesem Intervall beobachteten Atome markiert. Aus der Starrheitshypothese und der Tatsache, dass eine Landmarke nur zu einem Objekt gehören kann, folgt, dass Atom A und B durch die

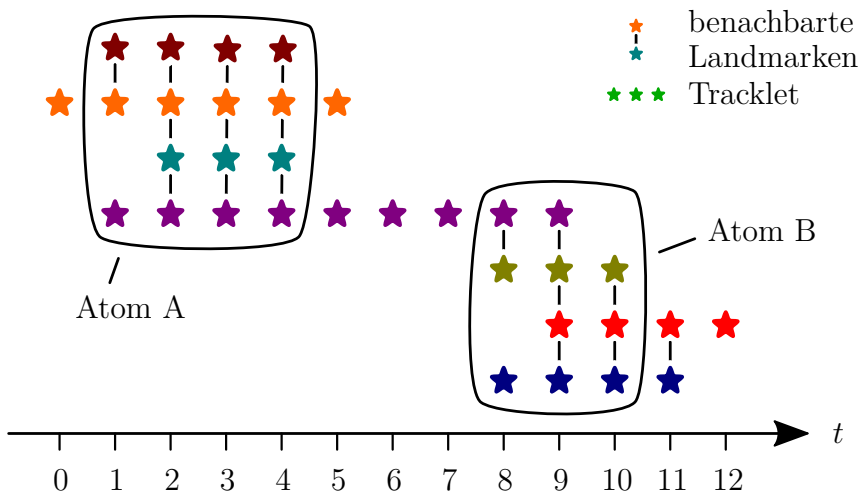


Abbildung 3.13 – Tracking der Landmarken führt zu zeitlicher Nachbarschaft von Atomen

gemeinsame violette Landmarke benachbart sind, obwohl deren Beobachtungsintervalle sich nicht überlappen.

Diese zeitliche Nachbarschaft hat schwerwiegende Folgen:

Auf der einen Seite ermöglicht sie erst eine zuverlässige Verfolgung der Objekte, trotz Teilverdeckungen (auch Selbstverdeckungen) und Phasen, in denen das Objekt wegen zu weniger Landmarken nicht beobachtbar ist.

Auf der anderen Seite resultiert daraus, dass zwei Objekte bereits dann als eine Komponente erkannt, also zu einem Objekt verschmolzen werden, sobald ein Atom des einen Objekts auch nur eine Landmarke mit einem Atom des anderen Objekts teilt. Dies ist insbesondere bei Landmarken, die auf Grund einer Fehlassoziation von einem Objekt auf ein anderes springen, leicht der Fall.

Somit ist die Komponentenbildung sehr sensibel gegenüber nicht verworfenen nicht-starren Atomen. Umso wichtiger ist also ein korrekter Hypothesen- und Plausibilitätstest der Atome.

3.1.4 Parametrierung

Die im zuvor erläuterten Hypothesen- und Plausibilitätstest letztlich verwendeten Parameter wurden für jedes Kriterium zunächst theoretisch motiviert und anschließend unter anderem mithilfe annotierter Daten experimentell bestimmt.

Die Ausgangswerte der Parameter wurden anhand vorhandenen Modellwissens gewählt. So wird der quadratische Rückprojektionsfehler f_2 auf Grund von Pixelrauschen in der

Größenordnung einzelner Pixel liegen, während eine akzeptable Unsicherheit der Landmarkenpositionen $\breve{\sigma}_{\mathbf{p}}$ nur wenige Meter betragen sollte. Die Unsicherheit der Relativbewegung $\breve{\sigma}_{\mathbf{t}}$ lässt sich hingegen auf Grund ihrer Abhängigkeit von der relativen Orientierung schlecht abschätzen (siehe Abschnitt 2.3.3 zur Diskussion über die Interpretation der Kovarianz von Kameraposen). Misst man Länge, Breite und Höhe eines Atoms entlang seiner drei Hauptachsen, können Atome, die über 100 mal länger als breit sind, also eine Axialität α von mehr als 100 aufweisen, durchaus als unplausible Segmenthypothesen betrachtet werden. Ebenso können Atome, mit einer Ausdehnung η von mehreren Metern, als höchstwahrscheinlich objektübergreifend eingestuft werden. Die so motivierten Ausgangswerte der Parametrierung sind in Tabelle 3.1 aufgeführt.

Kriterium	Parameter	Ausgangswert
Maximaler quadratischer Rückprojektionsfehler	$f_{2,\max}$	1,0 [px^2]
Maximale Unsicherheit der Landmarkenposition	$\breve{\sigma}_{\mathbf{p},\max}$	2,0 [m]
Maximale Unsicherheit der Relativbewegung	$\breve{\sigma}_{\mathbf{t},\max}$	-
Maximale Axialität	$\alpha_{\max}$	100,0
Maximale Ausdehnung	$\eta_{\max}$	5,0 [m]

Tabelle 3.1 – Ausgangswerte für die Parametrierung der Hypothesen- und Plausibilitätstests

Ausgehend von diesen Initialwerten wurden die Parameter experimentell eingestellt. Hierzu wurde eine graphische Benutzeroberfläche entwickelt, die die Bildmerkmale und Atome einer ausgewählten Sequenz wie in Abbildung 3.5 visualisiert. Über Schieberegler können die Parameter verstellt und deren Auswirkung interaktiv anhand der gefilterten Atome untersucht werden. Spezifische Atome, die sich als besonders problematisch herausstellen, können markiert und ihre Metriken ausgegeben werden. Zudem ist die ganze Sequenz per Mausrad abspielbar, womit sich ein nützliches Instrument zur schnellen Parametrierung ergibt.

Für die Feineinstellung der jeweiligen Parameter wurden zusätzlich manuell annotierte Daten, die jede Landmarke einer Teilsequenz mit einem Objekt assoziieren, zu Hilfe genommen. Die Annotationen wurden im Wesentlichen durch Einkreisen der einzelnen Objekte in jedem Bild gewonnen. In einem ersten Versuch wurden diese Annotationen dazu verwendet, um Atome, deren Landmarken nur einem Objekt zugeordnet sind, als starr zu markieren und als solche zur Evaluierung der Parameter des Hypothesen- und Plausibilitätstests zu verwenden. Dies hat sich allerdings als leicht irreführend herausgestellt, da es zahlreiche Atome gibt, deren Landmarken alle innerhalb einer solchen markierten Objekthülle liegen, jedoch mindestens eine davon stark verrauscht ist oder eine Reflexion in der Objektoberfläche darstellt und diese Atome damit korrekterweise als nicht-starr markiert werden müssten. Somit führte dies zu vielen als falsch negativ bewerteten Atomen, die genau genommen richtig negative darstellten.

Statt der Bewertung des Starrheitstests für Atome ist also eine geeignete Bewertung der Objekterkennung wesentlich zielführender. Daher werden die Annotationen zur Berechnung einer Grundwahrheit der Objektaggregation verwendet, welche mit der berechneten Aggregation verglichen wird. Hierzu werden die Bounding Boxen der Objekte im Bildbereich bestimmt und die Assoziation mit der von Bernardin und Stiefelhagen [7] vorgestellten „multiple object tracking accuracy“ (MOTA) Metrik evaluiert. Diese beschreibt die Fehlerfreiheit der Assoziation und dient als Kostenfunktion eines Gauß-Newton Verfahrens, das die Parameter hinsichtlich der Grundwahrheit optimiert. Die Ableitung dieser Kostenfunktion wird dabei numerisch mit der Ridders Methode [29] approximiert.

Die so gewonnenen Parameter unterscheiden sich nicht mehr stark von den zuvor manuell bestimmten und können unter Umständen trotz zugrunde liegender Metrik für einen Anwender intuitiv weniger zufriedenstellend sein als die manuell ermittelten Parameter. Dies resultiert insbesondere daher, dass sich die MOTA Metrik aus drei gleichgewichteten Verhältnissen zusammensetzt: der Rate falsch positiv erkannter Objekte, also dem Verhältnis fehlerhaft erkannter Objekte zu insgesamt erkannten Objekten, der Rate der falsch negativ, also nicht erkannten, Objekte, und der Fehlassoziationsrate. Der Anwender würde diese Verhältnisse allerdings nicht unbedingt gleich gewichten, zumal das Verschmelzen zweier Objekte zu einer Hypothese für die Objektrekonstruktion wesentlich gravierendere Folgen hat, als der Zerfall eines Objektes in zwei Hypothesen. Beide Fälle sind allerdings nur indirekt durch die Raten dieser Metrik repräsentiert. Bei der Parametrierung ist also vielmehr ein gewünschtes Gleichgewicht zwischen dem Verschmelzen ähnlich bewegter Objekte und dem Zerfallen großer Objekte zu vielen Kleinen einzustellen.

Ein so oder manuell ermittelter Parametersatz liefert in der Regel zufriedenstellende Resultate und ist auf Datensequenzen mit unterschiedlichen Eigenschaften übertragbar. Seine Leistungsfähigkeit hängt allerdings auch von der Merkmalsdichte und anderen Eigenschaften der Eingangsdaten ab, die wiederum von der Sensorkonfiguration und vor allem der Qualität der Bilddaten selbst, also deren Auflösung, Farbtiefe und Belichtung, und den für das Landmarkentracking (siehe Abschnitt 2.4) gewählten Einstellungen ab. Eine Anpassung und weiteres Finetuning ermöglichen somit unter Umständen einige weitere quasi-weltfeste Landmarken auszuschließen, die ein oder andere Verschmelzung oder auch gegebenenfalls das Zerfallen eines Objektes zu verhindern. Um also die bestmögliche Performanz zu erreichen, können die Parameter nochmals je Sensorkonfiguration angepasst werden, wobei sie dann nur leicht unterschiedlich ausfallen. Die jeweils verwendeten Parameter sind in Kapitel 4 aufgeführt.

3.2 Objektrekonstruktion

In einem letzten Schritt wird aus den gefilterten und assoziierten Landmarken die Relativbewegung und Struktur der Objekte rekonstruiert. Da dies ohne jegliches Vorwissen über zu erwartende Formen und Bewegungsmodelle der Objekte, allerdings unter der Annahme starrer Objekte, erfolgen soll, wird hierfür wiederum auf das Bündelausgleichsverfahren zurückgegriffen. Anschließend wird in Abschnitt 3.2.2 die Objektoberfläche zur besseren Visualisierung mittels Alpha-Shapes approximiert.

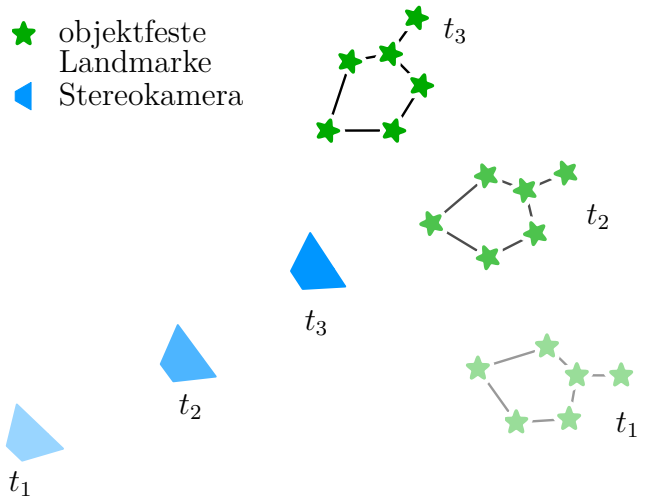
3.2.1 Schätzung von Objektstruktur und -bewegung

Zunächst werden ausgehend von dem Ergebnis der Komponentenbildung, wie bereits in Abbildung 3.12 veranschaulicht, die Relativbewegung jedes Objekts und dessen Struktur, das heißt die Positionen seiner Landmarken in Bezug zu einem Objektkoordinatensystem, bestmöglich geschätzt. Im Gegensatz zum Hypothesen- und Plausibilitätstest im vorigen Abschnitt dienen hierbei die Beobachtungen aller zu einem Objekt aggregierten Landmarken als Eingangsdaten des Bündelausgleichs in Gleichung (2.36). Damit werden alle verfügbaren Messungen eines Objekts in die Schätzung miteinbezogen und restliche Beobachtungen, insbesondere der Landmarken anderer Objekte, ausgeschlossen.

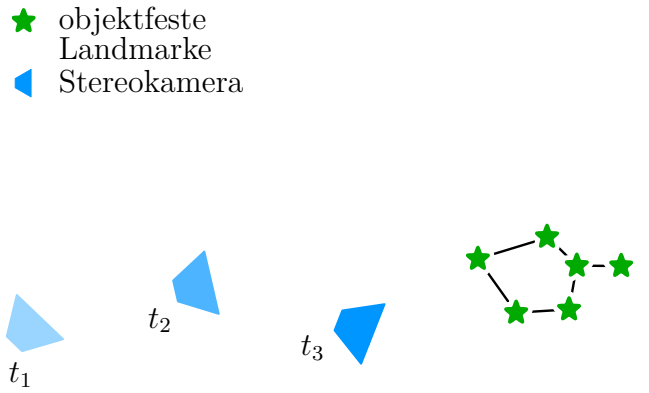
Wie bereits bei den Atomen, wird jedes Objekt auf ein Zeitintervall begrenzt, das eine Mindestanzahl an Beobachtungen je Frame aufweist. Damit soll verhindert werden, dass die ersten beziehungsweise letzten Frames eines Objekts, in welchen zu wenige Landmarken für eine aussagekräftige Analyse beobachtbar sind, die Schätzung merklich verschlechtern. Die Anzahl an Beobachtungen pro Frame ist hierfür ein zumeist ausreichendes, jedoch kein optimales Kriterium. Ein Objekt in großer Entfernung kann beispielsweise trotz vieler sichtbarer Landmarken nur schlecht rekonstruiert werden, während bei geringer Entfernung bereits eine wenige Landmarken zu einem präzisen Ergebnis führen. Die Entwicklung einer geeigneteren Metrik überstieg allerdings den zeitlichen Rahmen dieser Arbeit.

Da das Ergebnis der Objektaggregation noch einige Ausreißerlandmarken, unter anderem quasi-weltfeste Landmarken, beinhalten kann, werden in diesem Verarbeitungsschritt robuste Schätzverfahren für den Bündelausgleich verwendet. Zum einen wird die Huber-Funktion (2.22) als Anpassung der Kostenfunktion, wie in Gleichung (2.20), verwendet, um zu verhindern, dass solche Ausreißer die Schätzung dominieren.

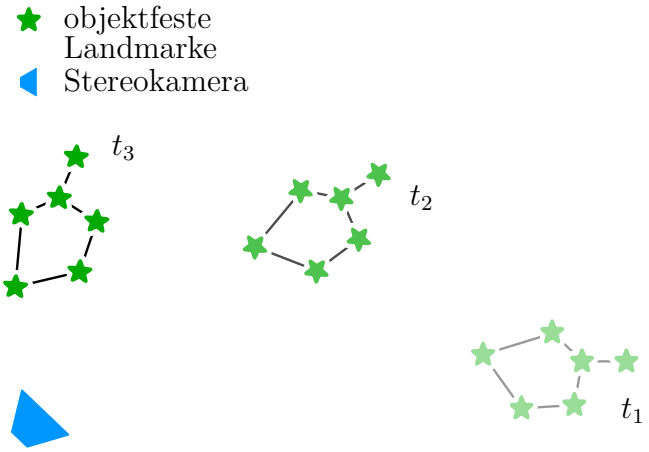
Zusätzlich werden Beobachtungen mit einem auffällig großen quadratischen Rückprojektionsfehler $f_{j'}(\mathbf{x}) > f_{\text{outlier}}$ (ohne Anwendung der Huber-Funktion), wie in Abschnitt 2.1.3 diskutiert, ausgeschlossen und der Bündelausgleich erneut bestimmt. Dieses Vorgehen wird mit immer kleinerem f_{outlier} iterativ wiederholt, so dass zum Schluss alle identifizierbaren



(a) Tatsächliche Bewegung der Kamera und eines Objekts im Weltkoordinatensystem



(b) Rekonstruierte Relativbewegung der Kamera und Objektstruktur



(c) Rekonstuierte Relativbewegung und Struktur des Objekts im Kamerakoordinatensystem

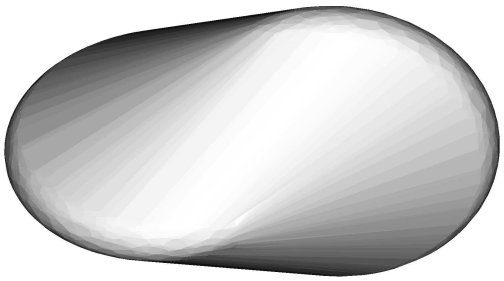
Abbildung 3.14 – Rekonstruktion der Relativbewegung der Kamera und Objektstruktur und Wechsel des Bezugskoordinatensystems

Abbildung 3.15 – Ergebnis der Struktur- und Bewegungsrekonstruktion

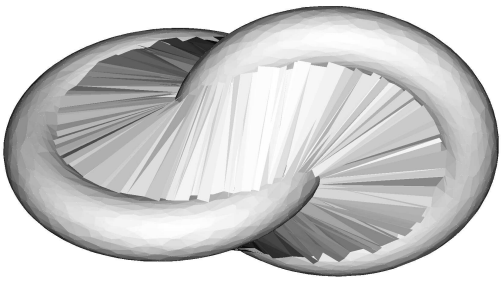
Ausreißer aus der Schätzung entfernt wurden. Dabei wird die Schätzlösung eines Iterationsschrittes als Startwert des Bündelausgleichs im darauf folgenden Schritt verwendet, womit die folgenden Iterationen nur noch einen Bruchteil der Zeit erfordern, um eine Lösung zu finden. Landmarken, die nach dem letzten Durchlauf insgesamt zu wenige akzeptierte Beobachtungen aufweisen, werden letztlich ganz verworfen.

Als Resultat des Bündelausgleichs erhält man neben der Position aller Landmarken, also der Struktur des Objekts, die Bewegung der Kamera relativ zu diesem Objekt beziehungsweise einem gemeinsamen Weltkoordinatensystem. Die Relativbewegung der Objekte bezüglich der Kamera selbst kann hieraus nun durch einen Wechsel des Bezugskoordinatensystems (siehe Gleichung (2.26)), wie in Abbildung 3.14 veranschaulicht, bestimmt werden. Abbildung 3.14(a) zeigt die bewegte Kamera und ein bewegtes Objekt zu drei Zeitpunkten schematisch in der Vogelperspektive des Weltkoordinatensystems. Die durch den Bündelausgleich rekonstruierte Objektstruktur und Relativbewegung der Kamera sind in Abbildung 3.14(b) dargestellt, während die daraus berechnete Objektbewegung im Kamerakoordinatensystem in Abbildung 3.14(c) zu sehen ist.

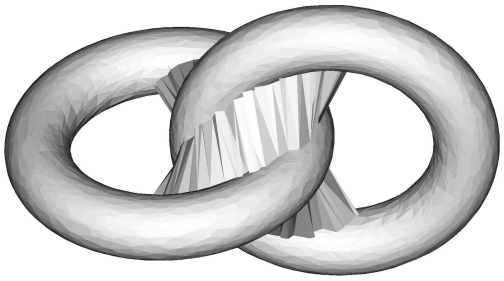
Das Ergebnis der Struktur- und Bewegungsrekonstruktion der Objekte ist in Abbildung 3.15 als Rückprojektion aller, auch zu diesem Zeitpunkt ursprünglich nicht sichtbaren, geschätzten Objektlandmarken dargestellt.



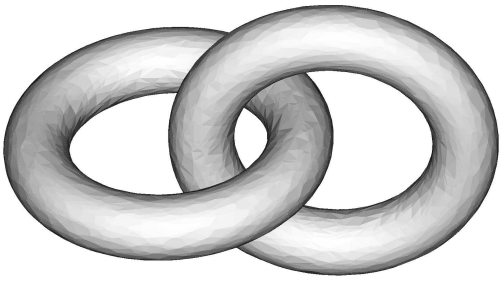
(a) Alpha shape mit alpha = ∞ entspricht der konvexen Hülle.



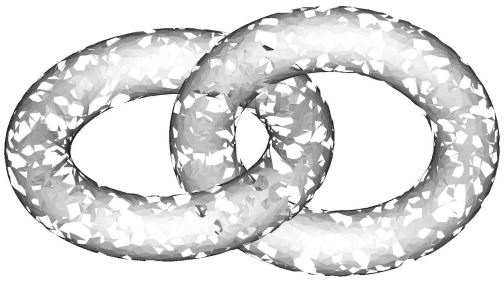
(b) alpha = 2



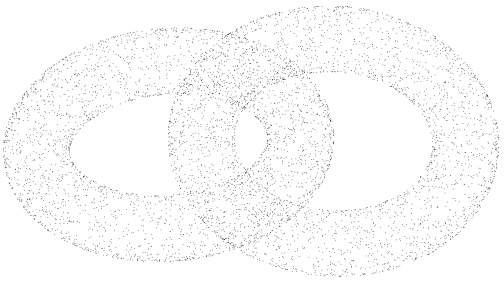
(c) alpha = 1.2



(d) alpha = 0.5



(e) alpha = 0.2



(f) Alpha shape mit alpha = 0 ergibt die ursprüngliche Punktwolke.

Abbildung 3.16 – Alpha shapes der Punktwolke zweier verschränkter Tori mit unterschiedlichen Werten für alpha. Der Innenradius der Tori beträgt 1, der Außenradius 4 (in der Einheit von alpha).

3.2.2 Oberflächenapproximation mit Alpha Shapes

In Abbildung 3.15 sind wesentlich mehr Rückprojektionen als Beobachtungen in diesem Frame (vgl. Abbildung 3.1(a)) sichtbar. Daraus wird ersichtlich, dass auf Grund der Akausalität der Schätzung, zu jedem Zeitschritt die vollständige Objektstruktur der Lösung zur Verfügung steht. In der Darstellung dieser Punktewolke, ob als Rückprojektion im Kamerabild oder einer perspektivischen Ansicht eines 3D Viewers, ist die 3D Struktur jedoch insbesondere in unbewegten Bildern nicht gut erkennbar. Daher ist es für eine aussagekräftige 3D Rekonstruktion erforderlich, aus der Punktwolke ein Oberflächenmo-

dell abzuleiten. In der Literatur werden häufig Bounding-Boxen zur Veranschaulichung verwendet [6, 22, 8]. Diese eignen sich hervorragend für eine vereinfachte Darstellung der Ausmaße und Orientierung eines Objekts, werden allerdings dem Detailreichtum des in dieser Arbeit rekonstruierten Strukturmodells nicht gerecht.

Um dieses also besser abzubilden, wird die Oberfläche jedes Objekts aus seinen Landmarken approximiert und anschließend als Dreiecksnetz im Kamerabild oder in einem geeigneten 3D Viewer dargestellt. Eine naheliegende Möglichkeit der Oberflächenapproximation ist, die konvexe Hülle der Punktwolke zu bestimmen. Zahlreiche Fahrzeugtypen, wie Lastkraftwagen mit Ladefläche, Kranwagen oder Cabriolets, haben allerdings keine konvexe Form. Auch auf den ersten Blick konvexe Fahrzeuge haben viele Elemente, die die Karrosserie um nicht-konvexe Formen ergänzen, beispielsweise die Seitenspiegel, eine Anhängerkupplung oder auch die Räder. Daher ist auch die konvexe Hülle eine zu starke Vereinfachung der Objektoberfläche.

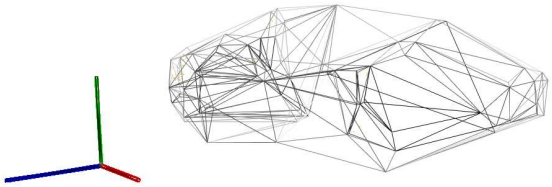
Die von Edelsbrunner, Kirkpatrick und Seidel [11] eingeführten sogenannten „alpha shapes“ bilden eine Verallgemeinerung der konvexen Hülle, die auch konkave Aushöhlungen ermöglicht und sind für die angestrebte Oberflächenapproximation sehr gut geeignet. Der reelle Parameter $0 \leq \alpha \leq \infty$ bestimmt dabei die maximale Krümmung der zugelassenen Aushöhlungen. Somit ist ein Alpha Shape mit $\alpha \to \infty$ identisch mit der konvexen Hülle der selben Punktmenge. Für kleiner werdendes α entstehen jedoch allmählich stärkere Aushöhlungen in der Oberfläche, bis in der Form sogar Tunnel entstehen, sie in mehrere Teile oder für $\alpha \to 0$ in alle Punkte zerfällt. Abbildung 3.16 veranschaulicht die alpha shapes der Punktwolke zweier verschränkter Tori mit verschiedenen Werten für alpha. Für eine mathematische Definition und ausführlichere Einführung der Alpha Shapes sei der Leser auf Edelsbrunner und Mücke [13] und Edelsbrunner [12] verwiesen.

Das mittels Alpha Shapes gewonnene Dreiecksnetz wird schließlich zum einen als Rückprojektion in das Kamerabild, wie in Abbildung 3.17, eingezeichnet. Zusätzlich können mit dem 3D Viewer der „PointCloud Library“ [30] Ansichten eines Objekts aus beliebigen Perspektiven erzeugt werden. Dabei wird sowohl die Darstellung des Dreiecksnetzes unterstützt, als auch ein 3D Rendering der Oberfläche. Abbildung 3.18 zeigt die approximierte Oberfläche eines Beispielobjekts (Mercedes-Benz CL-Klasse) im Kamerabild und dem 3D Viewer.

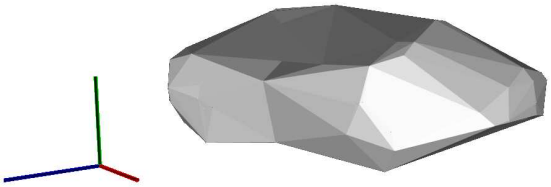
Abbildung 3.17 – Ergebnis der Oberflächenapproximation

(a) Pink markiertes Beispielobjekt von vorne rechts

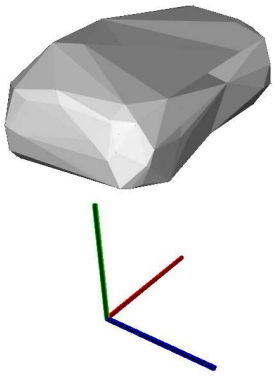
(b) und hinten rechts sichtbar.



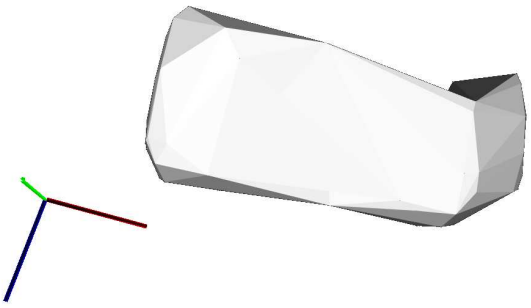
(c) Alpha Shape mit $\alpha = 1.5$ als Dreiecksnetz



(d) Flat-Shading der Alpha Shape Oberfläche



(e) Aus einer anderen Perspektive



(f) Sicht von oben

Abbildung 3.18 – Oberflächenapproximation für ein Beispielobjekt

4 Experimente

Die im vorigen Kapitel entwickelte Methode zur Lösung des „simultaneous tracking and reconstruction“ Problems wurde mit zahlreichen seitens Atlatec zur Verfügung gestellten Datensätzen getestet.

Die zur Aufzeichnung verwendete Sensorik wird in Abschnitt 4.1 kurz vorgestellt. Anschließend werden von den vorhandenen Atlatec Sequenzen zwei besonders interessante Teilsequenzen in Abschnitt 4.2 näher erläutert, bevor in Abschnitt 4.3 die Ergebnisse dieser Sequenzen gezeigt und daran die wesentlichen Stärken, wie auch Schwächen des Verfahrens aufgewiesen werden.

Zusätzlich war eine Evaluierung mit dem KITTI Benchmark vorgesehen. Abschnitt 4.4 erörtert warum sich die Daten als ungeeignet erwiesen haben.

4.1 Verwendete Sensorik

Atlatec verwendet für die Aufzeichnung von Bildsequenzen ein eigens entwickeltes Stereokamerasystem. Die in Abbildung 4.1 dargestellte „Atlabox“ beinhaltet neben zwei hochauflösenden Industriekameras mit Fisheye-Objektiven eine integrierte Festplatte und ist auf einfache Bedienbarkeit und schnelle Inbetriebnahme ausgelegt. Sie lässt sich mittels Saugnäpfen an ein beliebiges Testfahrzeug anbringen und die Aufzeichnung per Knopfdruck starten. Anschließend kann die Festplatte mit den gespeicherten Bildsequenzen der Box entnommen und die Aufzeichnungen verwertet werden. Die wichtigsten technischen Daten der Atlabox sind in Tabelle 4.1 aufgeführt.

Brennweite	$f_u = f_v = 670\,\text{px}$
Auflösung	1664 px x 1020 px
Horizontaler Öffnungswinkel	$\approx 100°$
Baseline	$b = 0{,}4\,\text{m}$
Bildrate	10 Hz

Tabelle 4.1 – Technische Daten der Atlabox

Abbildung 4.1 – Zur Aufzeichnung verwendete Atlabox auf einem Testfahrzeug

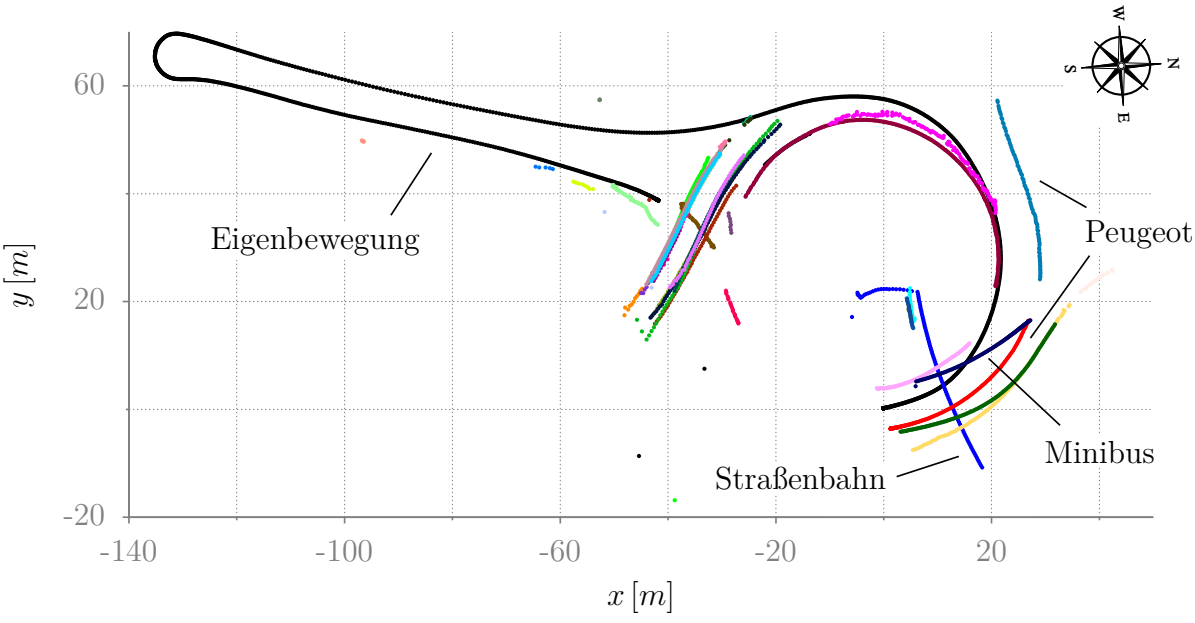
4.2 Ausgewählte Sequenzen

4.2.1 Karlsruhe Südstadt

Von den zahlreichen getesteten Atlatec Bildsequenzen sollen in diesem und im folgenden Abschnitt zwei Sequenzen im Detail vorgestellt werden. Die erste ist eine 1500 Frames lange Teilsequenz einer insgesamt 4500 Frames umfassenden Aufzeichnung in der Karlsruher Südstadt. Die Sequenz verläuft entlang der B10, zu großen Teilen am großen Kreisel an dem die Stuttgarter Straße, die B10 und die Wolfartsweierer Straße zusammen kommen. Ein Satellitenbild der befahrenen Umgebung ist in Abbildung 4.2(a) abgebildet. Es sind 40 bewegte Objekte, davon 33 Personenkraftwagen (PKWs), drei Fußgänger, zwei Minibusse, ein Lastkraftwagen (LKW) und eine Straßenbahn sichtbar.

In Abbildung 4.2(b) sind die rekonstruierten Trajektorien der Objekte, sowie die des aufzeichnenden Testfahrzeugs in eine Karte eingezeichnet. Die Sequenz beginnt im Stand an einer Ampel, an welcher zunächst eine Straßenbahn quer vorbei fährt. Darauf hin schaltet die Ampel auf Grün und sowohl das Testfahrzeug, wie auch weitere Verkehrsteilnehmer fahren in den Kreisverkehr ein. Nach einer knapp dreiviertel Umfahrung des Kreisverkehrs

(a) Satellitenbild der befahrenen Strecke in Karlsruhe [24]



(b) Eigenbewegung und Trajektorien der beobachteten Objekte der Karlsruher Testsequenz

Abbildung 4.2 – Karlsruher Testsequenz

biegt das Testfahrzeug in eine Seitenstraße ein, wendet dort und fährt wieder auf den Kreisverkehr zu. Dort bleibt es wegen des dichten Verkehrs stehen, während viele Fahrzeuge vorbeifahren. Abbildung 4.3 zeigt einige Eindrücke der Aufzeichnung mit eingezeichneter Strukturrekonstruktion der Objekte. Außerdem sind alle bereits in Kapitel 3 verwendeten Beispielbilder aus dieser Sequenz entnommen.

Die Sequenz weist 404 703 Atome auf, die im Schnitt jeweils über 10,36, jedoch bis zu 180 Frames am Stück beobachtbar sind und im Schnitt aus 6,89, jedoch bis zu 25 Landmarken bestehen. Einen genaueren Eindruck hierüber verschaffen die Histogramme in Abbildung 4.4. Wie in Abschnitt 3.1.1 angekündigt, werden für eine aussagekräftige Analyse Atome mit weniger als 4 Frames oder weniger als 5 Landmarken gänzlich verworfen. Im Rahmen der Objektrekonstruktion werden die Objekte auf ein Zeitintervall mit mindestens 10 Landmarken pro Frame gekürzt und Landmarken, die auf Grund dessen weniger als 4 Frames dieses Objekts belegen, ebenfalls verworfen (vgl. Abschnitt 3.2.1).

Somit bleiben immer noch 361 286 Atome übrig, für welche jeweils der Hypothesen- und Plausibilitätstest mithilfe des Bündelausgleichsverfahrens durchgeführt wird. Durch diese enorme Anzahl an Atomen im Vergleich zur Anzahl der Objekte nimmt die Objektaggregation in etwa $\frac{2}{3}$ der Gesamtrechenzeit in Anspruch. Auf einem Rechner mit Intel Core i7-5820K 6-Kern Prozessor und 16 GB Arbeitsspeicher benötigt die vorgestellte Methode zur Objektaggregation und -rekonstruktion für diese Karlsruher Teilsequenz insgesamt etwa 4 Stunden Rechenzeit.

Es wurden verschiedene Parametersätze gefunden, die dazu geeignet sind die bewegten Objekte einer Sequenz zu erkennen und ihre Bewegung, wie auch Struktur zufriedenstellend zu rekonstruieren. Die für die Karlsruher Sequenz besten und letztlich verwendeten Parameter des Hypothesen- und Plausibilitätstests und der Ausreißerdetektion sind in Tabelle 4.2 aufgeführt. Beim Vergleich der beiden Parametersätze wird deutlich, dass bei Verschärfung des Kriteriums maximaler Ausdehnung, andere Kriterien, wie die maximale Unsicherheit der Kameraposition und die maximale Axialität überflüssig werden können. Dennoch sind, wie in Abschnitt 3.1.4 diskutiert, generischere Parametersätze, wie der erste in Tabelle 4.2, zu bevorzugen, weil diese weniger Sequenz-spezifisch sind und somit besser übertragbar sind. Die robuste Objektrekonstruktion wurde in sieben Iterationen mit immer kleinerem Ausreißerschwellwert f_{outlier} bestimmt, wobei dieser bei allen Parametersätzen gleich gewählt wurde.

4.2.2 München Schleißheimer Straße

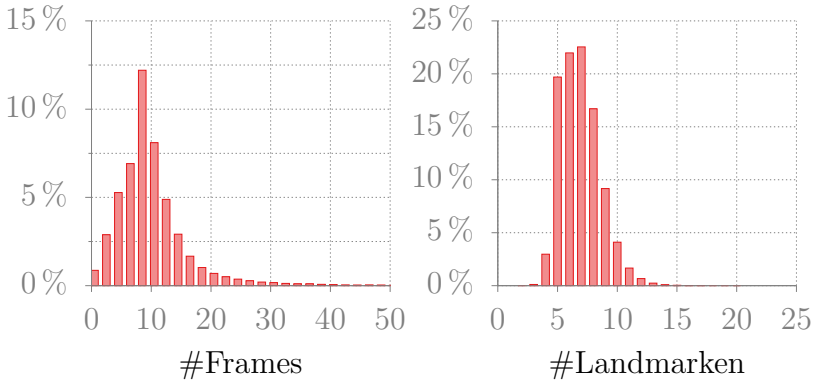
Die zweite ausgewählte Sequenz wurde entlang der Schleißheimer Straße in München aufgezeichnet und umfasst insgesamt 10 000 Frames. Wie auch die Karlsruher Aufzeichnung,

(a) Straßenbahn zu Beginn der Sequenz

(b) Abbiegevorgang in die Seitenstraße

(c) Zurück am Kreisverkehr

Abbildung 4.3 – Beispielbilder aus der Karlsruher Sequenz



(a) Anzahl an Frames pro Atom **(b)** Anzahl an Landmarken pro Atom

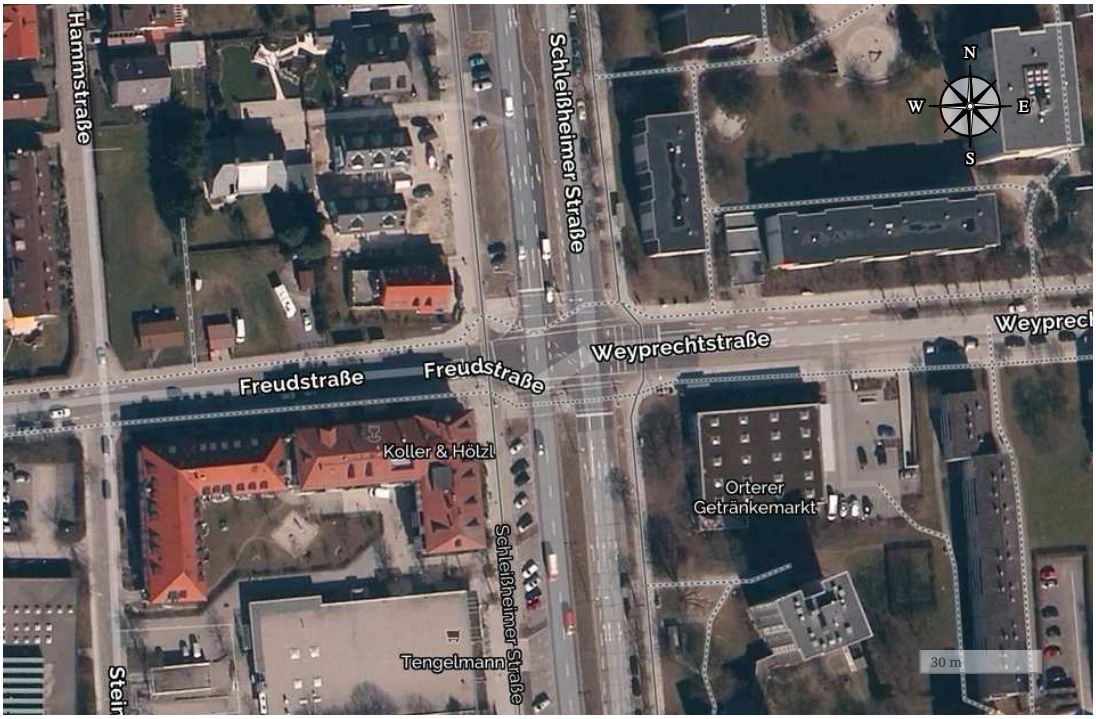
Abbildung 4.4 – Relative Häufigkeitsdichte der Atome bzgl. ihrer Anzahl an Frames bzw. Landmarken

Kriterium	Parameter	Satz 1	Satz 2
quadratischer Rückprojektionsfehler	$f_{2,\max}$	0,65 [px^2]	0,65 [px^2]
Unsicherheit der Landmarkenposition	$\breve{\sigma}_{\mathbf{p},\max}$	2,5 [m]	2,0 [m]
Unsicherheit der Kameraposition	$\breve{\sigma}_{\mathbf{t},\max}$	70	-
Axialität	$\alpha_{\max}$	80,0	-
Ausdehnung	$\eta_{\max}$	3,5 [m]	2,9 [m]
Ausreißerschwellwert	f_{outlier}	{10, 7,5, 5,0, 2,5, 1,5, 1,2, 1,0}	

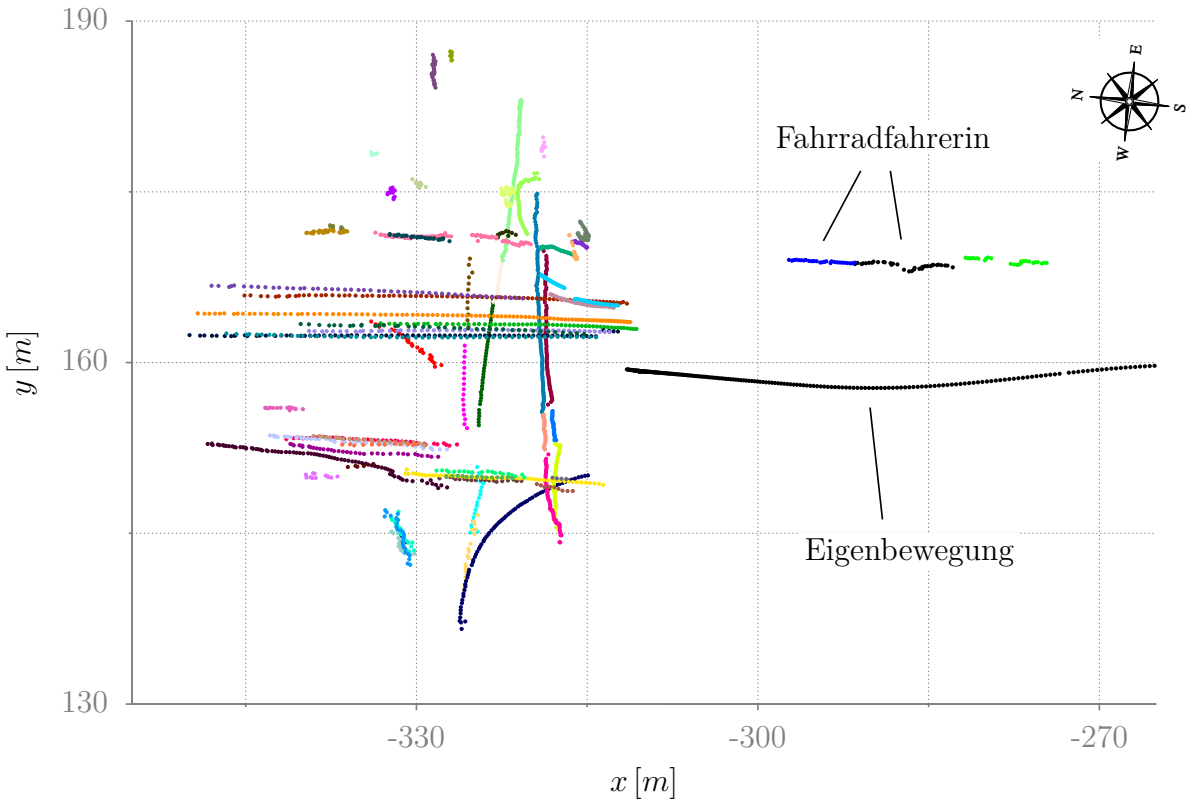
Tabelle 4.2 – Für die Karlsruher Sequenz verwendete Parameter des Hypothesen- und Plausibilitätstests der Objektaggregation sowie der Ausreißerdetektion der Objektrekonstruktion

entstand diese zur Tageszeit bei leicht bewölkter, aber dennoch trockener Witterung. Für die Bildverarbeitung sind dies optimale Voraussetzungen, da die Umgebung hell beleuchtet ist, allerdings keine Gefahr der Überbelichtung durch direkte oder auch reflektierte Sonnenstrahlung besteht. Zudem beinhaltet die Aufzeichnung eine interessante Teilsequenz an einer vielbefahrenen Kreuzung von 950 Frames. Darin nähert sich das aufzeichnende Versuchsfahrzeug der Kreuzung und ordnet sich an vorderster Stelle in die Linksabbiegerspur ein, während der Querverkehr freie Fahrt hat. Anschließend schaltet die Ampel um, so dass nur der nicht-kreuzende Längsverkehr auf Grün geschaltet ist, das Versuchsfahrzeug aber noch halten muss. Ein Satellitenbild dieser Kreuzung ist in Abbildung 4.5(a) abgebildet.

Während dieser vergleichsweise kurzen Sequenz können vielfältige bewegte Objekte beobachtet werden. Darunter sind 23 PKWs, hiervon zwei Cabriolets, zwei Geländelimousinen und ein Erlkönig, sechs Fahrradfahrer, vier Fußgänger, vier Kleintransporter und ein Sattelzug. In sechs weiteren Teilsequenzen der Münchner Aufzeichnung sind außerdem ein Autotransporter, ein Lastzug und ein Tieflader, der zwei Arbeitsbühnen transportiert, zu



(a) Satellitenbild der vielbefahrenen Kreuzung in München [24]



(b) Eigenbewegung und Trajektorien der beobachteten Objekte der Münchner Testsequenz

Abbildung 4.5 – Münchner Testsequenz

sehen. Im folgenden Kapitel wird gezeigt, dass auch solche exotischeren Objekte problemlos erkannt und rekonstruiert werden können, solange sie im Beobachtungsintervall in sich starr sind und in den Eingangsdaten ausreichend viele Landmarken aufweisen. Abbildung 4.5(b) visualisiert die Eigenbewegung sowie die rekonstruierten Trajektorien der beobachteten Objekte in der Kreuzungssequenz.

Eine besondere Herausforderung dieser Bilddaten ist, dass die Fahrbahn einen bepflanzten Mittelstreifen mit hohem Gras aufweist. Somit ist die Erkennung des Gegenverkehrs bereits im Schritt des Featuretrackings stark gestört, so dass Fahrzeuge des Gegenverkehrs, falls sie hoch genug sind, lediglich einige Landmarken am oberen Teil ihrer Karosserie aufweisen. Einen ebenso dramatischen Effekt haben zahlreiche Pfosten von Verkehrsschildern und Ampeln am Fußgängerübergang, ebenfalls im Bereich des Mittelstreifens. Dies ist in Abbildung 4.5(b) erkennbar: Die Trajektorien des Gegenverkehrs enden bereits im Kreuzungsbereich, statt auf der Höhe der Kamera, wo sie das Blickfeld tatsächlich verlassen. Das hohe Gras und die Pfosten sind außerdem im Kamerabild 4.6(b) und 4.6(c) zu sehen.

Ausgehend von den Parametern der Karlsruher Sequenz wurde, wie in Abschnitt 3.1.4 beschrieben, ein für die Münchner Aufzeichnung angepasster Parametersatz für die Objektaggregation ermittelt. Dieser weicht leicht von dem der Karlsruher Sequenz ab, da für diese Aufzeichnung eine andere Sensorkonfiguration, insbesondere ein anderes Versuchsfahrzeug und abweichende Einstellungen des Landmarkentrackings, verwendet wurden. Die Werte liegen allerdings in der selben Größenordnung. Der maximale tolerierte Rückprojektionsfehler als Ausreißerkriterium beim Bündelausgleich der Objektrekonstruktion wurde, ebenso wie die Mindestanzahl an Frames und Landmarken pro Atom, sowie pro Objekt, unverändert übernommen. Tabelle 4.2 listet die wichtigsten Parameter nochmals auf.

Kriterium	Parameter	Satz 1
quadratischer Rückprojektionsfehler	$f_{2,\max}$	0,48 [px^2]
Unsicherheit der Landmarkenposition	$\breve{\sigma}_{\mathbf{p},\max}$	2,5 [m]
Unsicherheit der Kameraposition	$\breve{\sigma}_{\mathbf{t},\max}$	-
Axialität	$\alpha_{\max}$	10,0
Ausdehnung	$\eta_{\max}$	3,0 [m]
Ausreißerschwellwert	f_{outlier}	{10, 7,5, 5,0, 2,5, 1,5, 1,2, 1,0}

Tabelle 4.3 – Für die Münchner Sequenzen verwendete Parameter des Hypothesen- und Plausibilitätstests der Objektaggregation sowie der Ausreißerdetektion der Objektrekonstruktion

(a) Zwei Fahrradfahrer zu Beginn der Kreuzungssequenz

(b) Grünphase für den Querverkehr

(c) Pfosten und hohes Gras erschweren die Sicht auf den entgegenkommenden Verkehr

Abbildung 4.6 – Beispielbilder aus der Münchner Kreuzungssequenz

4.3 Ergebnisse

Mit der in Kapitel 3 entwickelten Methode konnten Objekte in anspruchsvollen Bildsequenzen zuverlässig detektiert, verfolgt und ihre Bewegung, sowie Struktur, rekonstruiert werden. Bereits in den vorhergehenden Abschnitten wurden einige Ergebnisse in Form von rekonstruierten Objekttrajektorien und -oberflächen vorausgegriffen. In diesem Abschnitt sollen jedoch die wesentlichen Vor- und Nachteile der erarbeiteten Methodik diskutiert und zusammengefasst werden. Diese resultieren vor allem daraus, dass die Objektaggregation allein auf der Hypothese starrer Objekte beruht und das gesamte Verfahren akausal arbeitet.

4.3.1 Generalisierbarkeit

Die Annahme, dass alle Objekte starr sind, erweist sich als sehr zielführend und scheitert nur in wenigen Situationen. Sie ermöglicht die Erkennung vielfältiger Objekte, von Radfahrern, über verschiedene Arten von PKWs, bis hin zu Nutzfahrzeugen, wie gewöhnlichen LKWs, Autotransportern oder Kranwagen, und Straßenbahnen. In Abbildung 4.6(b) wurden beispielsweise zwei Fahrradfahrer, zwei PKWs und ein Kleintransporter korrekt erkannt und rekonstruiert. Abbildung 4.3(a) zeigt eine rekonstruierte Straßenbahn, deren Strukturmodell in Abbildung 4.7 nochmals in zwei, von den ursprünglichen Kameraperspektiven abweichenden, Perspektiven dargestellt ist.

Die Starrheitshypothese ist außerdem so generisch, dass sie nicht nur die Erkennung vielfältiger, sondern auch bisher unbeobachteter Objekttypen ermöglicht. Die Parameter der Münchner Sequenzen wurden allein anhand der Kreuzungsszene gewonnen, führen in den restlichen Teilsequenzen jedoch auch zur erfolgreichen Objekterkennung und -

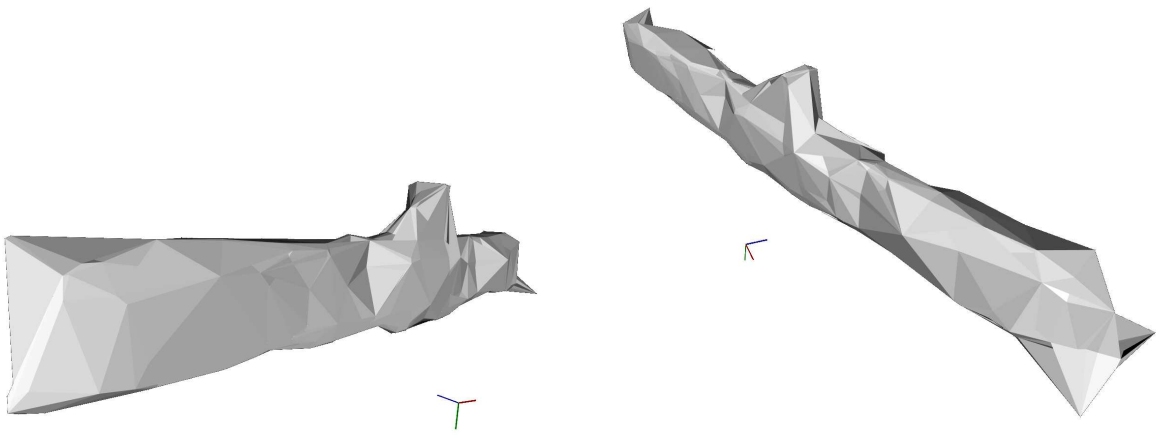


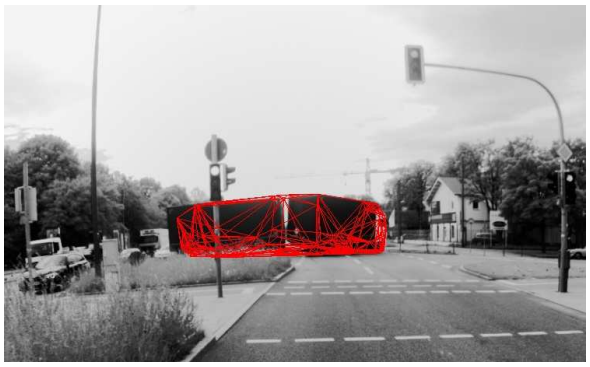
Abbildung 4.7 – Rekonstruierte Struktur der Straßenbahn aus der Karlsruher Sequenz

(a) Originalbild mit einem Tieflader

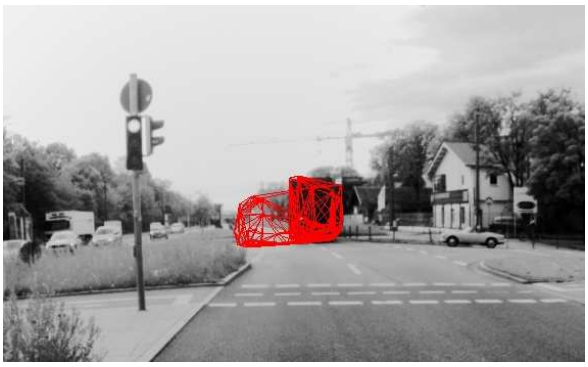
(b) Rekonstruktion des Tiefladers samt Ladung

(c) Originalbild mit einem Autotransporter

(d) Rekonstruktion des Autotransporters



(e) Als ein Objekt aggregierter Lastzug



(f) Anhänger sind mit der Zugmaschine allerdings nicht starr verbunden

(g) Zwei zu einem Objekt verschmolzene Fußgänger

(h) Nicht erkannter Fußgänger

Abbildung 4.8 – Beispielbilder zur Generalisierbarkeit des Verfahrens

rekonstruktion von Objektklassen, die in der Kreuzungsszene nicht aufgetreten sind. Darunter fällt zum Beispiel der Tieflader in Abbildung 4.8(a) und 4.8(b) und der Autotransporter in Abbildung 4.8(c) und 4.8(d). Dies ist ein großer Vorteil gegenüber lernbasierten Verfahren.

Für Objekte, die an sich nicht starr oder nicht vollständig starr sind, funktioniert die Objektaggregation allerdings nur fehlerhaft bis gar nicht. Solche Objekte sind in den untersuchten Aufzeichnungen vor allem Fußgänger oder Fahrzeuge mit Anhängern.

Abbildung 4.8(e) und 4.8(f) zeigt die Rekonstruktion eines Lastzugs, also eines LKWs mit Anhänger. Sowohl der LKW als auch der Anhänger sind in sich starr, allerdings ist der Anhänger nicht starr, sondern über eine Anhängerkupplung, die eine Drehung des Anhängers zulässt, mit dem LKW verbunden. Somit ist die Kombination der beiden zum Lastzug nicht vollständig starr. Im gezeigten Beispiel wurde der Lastzug zwar richtig erkannt, allerdings als ein einzelnes Objekt aggregiert und somit eine starre Struktur geschätzt. Diese stimmt für die Zeit des Abbiegemanövers des Lastzugs zwar noch näherungsweise, allerdings nicht mehr für die Geradeausfahrt, in der LKW und Anhänger gleich ausgerichtet sind.

Fußgänger bewegen in der Regel Arme, Beine, Kopf und teilweise auch den Oberkörper. Somit sind diese offensichtlich nicht starr, sondern bestehen lediglich aus fast-starren Gliedern. Dadurch ist eine Erkennung im Rahmen der Objektaggregation zwar manchmal noch möglich, eine korrekte Struktur- und Bewegungsschätzung allerdings nicht. Die beiden Fußgänger in Abbildung 4.8(g) wurden beispielsweise noch als ein Objekt erkannt, weil sie nebeneinander mit gleicher Geschwindigkeit und Richtung laufen (siehe auch Abschnitt 4.3.5), während der Herr in Abbildung 4.8(h) gar nicht erkannt wurde. Somit ist das Verfahren nicht zur Fußgängererkennung und -rekonstruktion geeignet.

4.3.2 Detailreiche Modelle trotz hohen Entfernungen

Eine der wichtigsten Stärken des Verfahrens resultiert aus der Akausalität. Für die Schätzung von Struktur und Bewegung eines Objekts werden nämlich alle verfügbaren Messungen verwendet, so dass im Anschluss zu jedem Zeitpunkt der Sequenz das vollständige, bestmögliche Strukturmodell zur Verfügung steht.

Dies ist insbesondere bei Teilverdeckungen der Objekte (siehe Abschnitt 4.3.3) und großen Entfernungen wertvoll. Denn wenn ein Fahrzeug aus großer Entfernung, wie in Abbildung 4.9(a), näher kommt, belegt es zunächst nur einen kleinen Teil des Bildes, so dass kausale Verfahren hier nur eine schlechte Schätzung erreichen können. Wird dasselbe Objekt allerdings auch bei geringer Entfernung, also mit höherer Auflösung, wie in Abbildung 4.9(b) beobachtet, kann bei der Objektrekonstruktion ein genaues und detailreiches

Strukturmodell geschätzt werden. Dieses steht dann ab der ersten Beobachtung zur Verfügung, also auch zum Zeitpunkt der großen Entfernung.

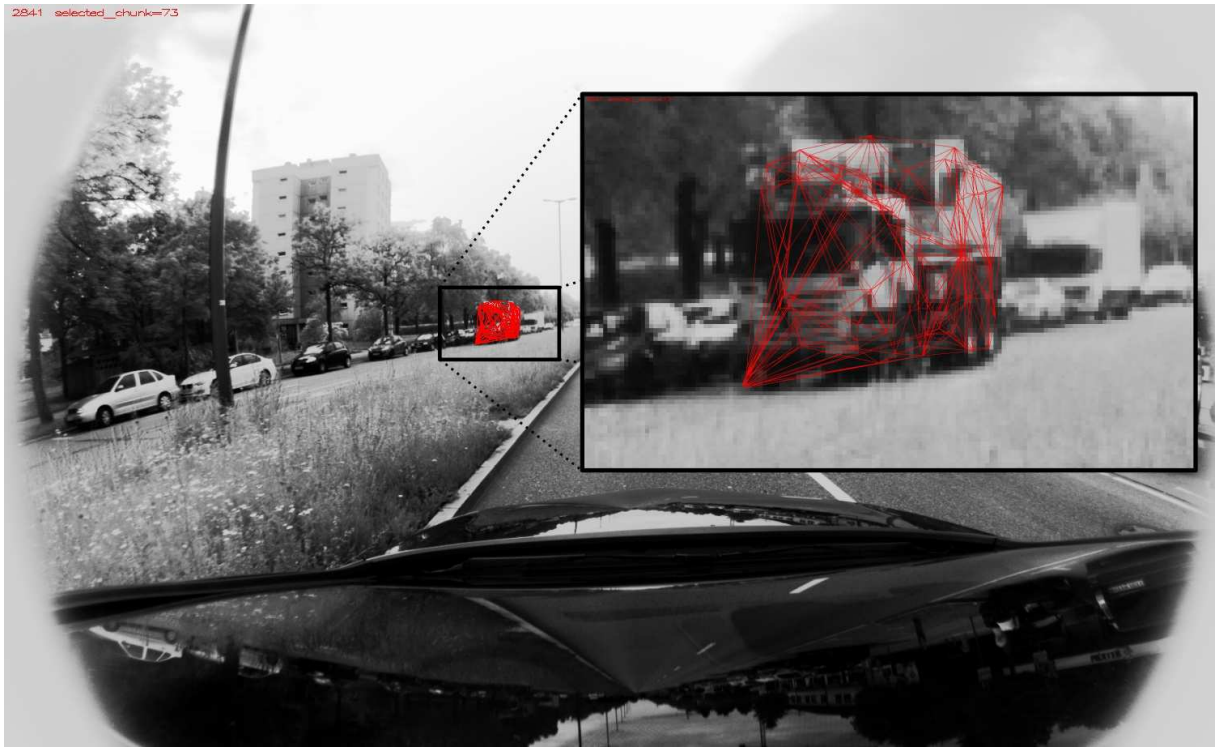
4.3.3 Einfluss von Verdeckungen

Ähnlich wie im vorigen Abschnitt bringt die Akausalität auch einen großen Vorteil bei Teilverdeckungen von Objekten. Ist ein Objekt nur teilweise sichtbar, weil es von einem anderen Objekt verdeckt wird, oder aus dem Blickfeld ragt, steht zu diesem Zeitpunkt dennoch das vollständige Strukturmodell zur Verfügung. Wurde das Objekt nämlich zu einem anderen Zeitpunkt zu größeren Teilen beobachtet, ist diese Information auch zu allen anderen Zeitpunkten verfügbar. Dies wird in Abbildung 4.10(a) besonders deutlich, wo nur noch die Heckklappe des VW Golf sichtbar ist, jedoch das Strukturmodell in voller Fahrzeuglänge dargestellt werden kann, weil der Golf zuvor in voller Länge beobachtet wurde. Abbildung 4.10(b) zeigt zudem die Rückprojektion der Strukturrekonstruktion zweier Fahrzeuge, die teilweise von Radfahrern verdeckt sind.

Genaugenommen verdeckt ein Objekt sich auch immer teilweise selbst, so dass zu keinem Zeitpunkt das Objekt von allen Seiten beobachtbar ist. Dreht sich ein Objekt allerdings stark relativ zur Kamera, so wird das rekonstruierte Strukturmodell eine Rundumsicht des Objekts repräsentieren. Dies wird an der rekonstruierten Objektstruktur des Kleintransporters aus Abbildung 4.3(c) in den Abbildungen 4.10(c)-(e) deutlich. Der Kleintransporter fährt, aus Kameraperspektive, von links nach rechts am Versuchsfahrzeug vorbei. Dadurch wird er zu Beginn seitlich von vorne beobachtet, wobei das Heck verdeckt ist. Während der Kleintransporter rechts aus dem Kreisverkehr herausfährt, wird er seitlich von hinten beobachtet, wobei nun die Front verdeckt ist. Insgesamt beinhaltet das rekonstuierte Strukturmodell alle beobachteten Seiten, also sowohl Front, rechte Seite als auch das Heck.

Während Teilverdeckungen keine Schwierigkeiten für die Objektrekonstruktion bereiten, führt eine vollständige Verdeckungen eines Objekts dazu, dass es nach der Verdeckung als neues Objekt behandelt wird, das Tracking hier also versagt. Dies liegt daran, dass das Landmarkentracking nicht darauf ausgelegt, ist Bildmerkmale mit Landmarken zu assoziieren, die im vorhergehenden Frame nicht mehr sichtbar sind. Dies tritt also beispielsweise dann auf, wenn ein Fahrzeug das Versuchsfahrzeug kreuzt, so dass die Sicht auf ein anderes Fahrzeug zeitweise verdeckt wird.

In der Karlsruher Sequenz verdeckt zum Beispiel der Minibus, sichtbar in Abbildung 3.17, beim Fahrspurwechsel den blau markierten Peugeot 207 für 23 Frames. An der Trajektorie des Peugeots in Abbildung 4.2(b) ist die zeitweise Verdeckung anhand der Lücke, sowie die Neuassoziation anhand des Farbwechsels, zu erkennen. An dieser Stelle sei angemerkt, dass die Farben der Trajektorien nicht mit den verwendeten Farben in den Kamerabildern



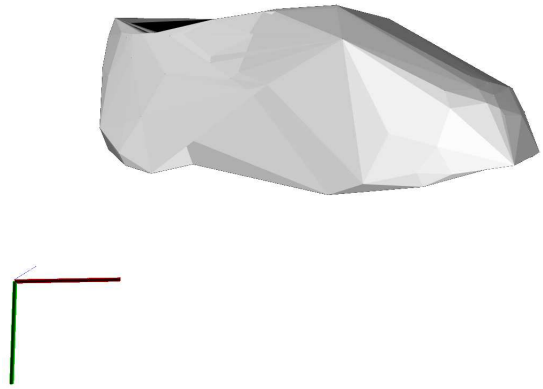
(a) Detailliertes Strukturmodell eines Kranwagens in großer Entfernung ...

(b) auf Grund späterer Beobachtungen in geringer Entfernung

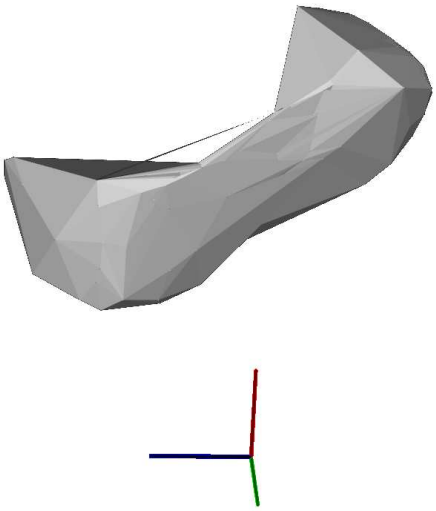
Abbildung 4.9 – Vollständiges Strukturmodell dank Akausalität ab der ersten Beobachtung verfügbar

(a) Strukturmodell mit voller Fahrzeuglänge verfügbar, obwohl nur die Heckklappe sichtbar ist

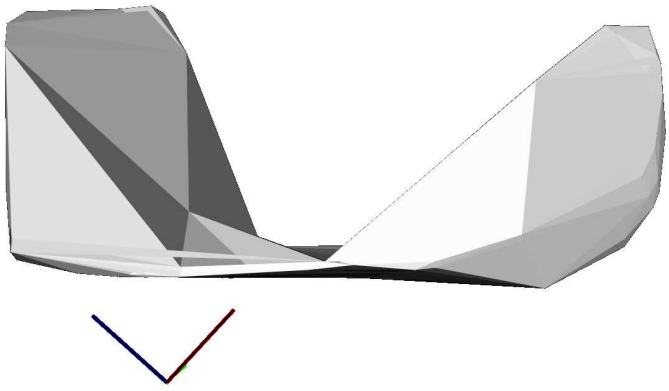
(b) Durch Fahrradfahrer zum Teil verdeckte Fahrzeuge



(c) Aus frontal-seitlicher Sicht ist die Rückseite des Kleintransporters verdeckt



(d) Kleintransporter aus anderer Perspektive



(e) Strukturmodell des Kleintransporters beinhaltet Heck, rechte Seite und Front

Abbildung 4.10 – Umfangreiche Strukturmodelle trotz Teil- oder Selbstverdeckungen

übereinstimmen. In den Trajektorienplots hat jedes jemals in der Sequenz erkannte Objekt eine eindeutige Farbe, während die Farben der Rückprojektionen in den Kamerabildern nur im untersuchten Kamerabild eindeutig sind.

Es kann jedoch sogar eine Teilverdeckung, die das ganze Objekt passiert und somit jede Landmarke des Objekts kurzzeitig verdeckt, dazu führen, dass die Objektverfolgung an dieser Stelle abbricht und das Objekt statt dessen als ein Neues getrackt wird. Dies passiert allerdings nicht, sobald sich jenes Objekt so schnell bewegt oder das verdeckende Objekt so dünn ist, dass mindestens eine der Landmarken in keinem Kamerabild verdeckt wird. Das erfolgreiche Tracking der einen Landmarke resultiert dann in einem erfolgreichen Objekttracking.

In der Münchner Sequenz ist eine Neuassoziation durch Teilverdeckung beispielsweise zu Beginn zu beobachten, wo die grün markierte Fahrradfahrerin aus Abbildung 4.6(a) hinter zwei Bäumen vorbei fährt. Der erste Baum ist noch zu dünn, um Probleme zu bereiten. Beim zweiten Baum findet allerdings ein Assoziationswechsel statt, der in ihrer Trajektorie in Abbildung 4.5(b) zu erkennen ist.

4.3.4 Empfindlichkeit gegen Fehlassoziationen von Landmarken

Wie in Abschnitt 3.1.2 diskutiert, können Landmarken mit einer Fehlassoziation, also solche Landmarken, denen zuerst ein Bildmerkmal eines Objekts A und nach einigen Frames ein Bildmerkmal eines anderen Objekts B zugeordnet wird, schnell dazu beitragen, dass beide Objekte während der Objektaggregation zu einem einzigen Objekt verschmelzen. Solche Fehlassoziationen kommen zwar nicht häufig allerdings regelmäßig vor und führen zu einer deutlich schlechteren Struktur- und Bewegungsrekonstruktion.

Abbildung 4.12 zeigt die einzige Fehlassoziation der Karlsruher Sequenz auf, die in der Verschmelzung zweier Objekte resultierte. Beide Fahrzeuge durchqueren seitlich und

Abbildung 4.11 – Zwei durch eine Fehlassoziation verschmolzene Fahrzeuge

(a) Die markierte Landmarke liegt zunächst auf dem rechten Fahrzeug ...

(b) ein Frame später wird ein Bildmerkmal des linken Fahrzeugs der selben Landmarke zugeordnet.

Abbildung 4.12 – Fehlassoziation einer Landmarke

längs zueinander versetzt den Kreisverkehr, als die rot markierte Landmarke vom rechten Fahrzeug auf das linke Fahrzeug zu springen scheint. Nicht alle Atome, die diese Landmarke beinhalten, konnten durch den Hypothesen- und Plausibilitätstest herausgefiltert werden, so dass beide Fahrzeuge als ein größeres Objekt rekonstruiert werden, wie in Abbildung 4.11 zu sehen ist. Es wird deutlich, dass die Strukturrekonstruktion sehr ungenau geworden ist, während die Bewegungsschätzung gut ausgefiel. In Abbildung 4.2(b) ist im linken Teil des Kreisels nämlich keine auffällig unstetige Trajektorie vorhanden.

4.3.5 Verschmelzen ähnlich bewegter Objekte

Die Gefahr, dass benachbarte Objekte im Schritt der Objektaggregation zu einem Objekt verschmelzen, besteht nicht nur bei Fehlassoziationen der Landmarken. Sie tritt ebenso bei Objekten auf, die sich ähnlich, also mit nahezu gleicher Geschwindigkeit und Richtung, fortbewegen. Atome, die aus Landmarken solch zweier Objekte bestehen, erscheinen starr, da sich jede ihrer Landmarken ähnlich bewegt.

Liegen zudem beide Objekte nah beieinander, haben diese Atome eine unauffällige Form und Größe. Damit können Atome, die ähnlich bewegte und räumlich nahe Objekte verbinden, durch den Hypothesen- und Plausibilitätstest nur schwer als objektübergreifendes Atom identifiziert werden. Dies führt zwangsläufig zur Verschmelzung der beiden Objekte und damit verbunden zu einer schlechteren Strukturrekonstruktion, einer zumeist fehlerhaften Oberflächenapproximation und zu einer schlechteren Bewegungsschätzung.

Da jedes Objekt in der Regel einige Atome hat, die nur für wenige Frames beobachtbar sind, kann ein objektübergreifendes Atom also schon dann starr und plausibel erscheinen, wenn sich beide Objekte innerhalb dieses vergleichsweise kurzen Intervalls ähnlich und nah beieinander bewegen. Dies wird in Abbildung 4.13 deutlich, wo das violett markierte

(a) Zwei verschmolzene Fahrzeuge mit getrennten Oberflächen

(b) Zwei verschmolzene Fahrzeuge mit fehlerhaft approximierten Oberflächen

Abbildung 4.13 – Verschmelzen ähnlich bewegter Objekte

Fahrzeugpaar in 4.13(a) noch nah beieinander liegt, das linke Fahrzeug allerdings das Rechte bis in 4.13(b) knapp überholt hat.

An diesem Beispiel wird allerdings auch ersichtlich, dass die Oberflächenapproximation mittels Alpha Shapes zwei getrennte Oberflächen erzeugen kann. Diese Erkenntnis sollte für zukünftige Arbeiten verwendet werden, um verschmolzene Objekte im Nachhinein zu trennen und nochmals separat zu rekonstruieren. Aus Abbildung 4.11 und dem blau markierten Fahrzeugpaar in Abbildung 4.13(b) wird jedoch auch klar, dass die Trennung mittels Alpha Shapes nicht zwingend gelingt.

4.3.6 Zerfallen homogener Objekte

Um das Verschmelzen von Objekten zu verhindern, kann ein Anwender schnell dazu geneigt sein, „schärfere“ Parameter für den Hypothesen- und Plausibilitätstest zu wählen. Wie in Abschnitt 3.1.4 bereits diskutiert, kann dies allerdings dazu führen, dass kleine Objekte mit wenigen Landmarken gar nicht mehr erkannt werden. Große, besonders dünn besetzte Objekte können zudem in mehrere kleinere Objekthypothesen zerfallen. Dieser Effekt kann vor allem bei großen Objekten mit homogener Oberflächentextur, wie beispielsweise LKWs mit unbedruckten Planen, auftreten.

In den untersuchten Datensequenzen ist jedoch noch ein weiterer, in Abbildung 4.14 dargestellter, Fall zu beobachten, in dem ein Objekt in zwei Hypothesen zerfällt. Das betroffene Taxi hat zwar gewöhnliche Ausmaße eines PKWs und wird in nächster Nähe beobachtet, womit es eigentlich ausreichend viele Landmarken zur erfolgreichen Objektaggregation hat. Allerdings sind viele der detektierten Bildmerkmale im mittleren Bereich durch Verspiegelungen verrauscht, so dass alle dortigen Atome als nicht-starr eingestuft und verworfen wurden und das Taxi somit in zwei Objekthypothesen zerfallen ist.

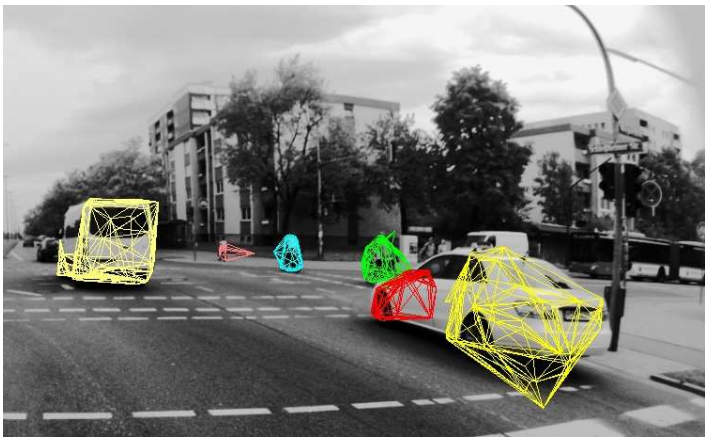


Abbildung 4.14 – Ein in zwei Objekte zerfallenes Fahrzeug

4.4 KITTI

Der KITTI Datensatz [16, 15] ist einer der aktuellsten und herausfordernsten Datensätze und Benchmarks für die Entwicklung und Evaluierung der Bildverarbeitung mobiler Robotik und Fahrerassistenzsysteme. Er besteht aus den Daten zweier Graustufen- und zweier Farbkameras, eines Velodyne Laserscanners und eines hochpräzisen Lokalisierungssystems, das GPS, GLONASS und eine inertiale Messeinheit mit RTK Korrektursignalen kombiniert.

Durch eine semiautomatisch erstellte Grundwahrheit des optischen Flusses, Verwendung der Lokalisierungsinformation als Grundwahrheit für visuelle Odometrie, manuell erstellte 3D Annotationen der Laserscans und Evaluierungsmetriken für jede dieser Aufgaben, ergibt sich ein umfassender Benchmark aus 289 Graustufen- und Farbbildpaaren für Stereomatching und Optischen Fluss, 22 Stereosequenzen über 39,2 km für visuelle Odometrie und 50 Stereosequenzen mit über 200 000 Objektannotationen für Objekterkennung und -verfolgung.

Für die vorliegende Arbeit war somit der Tracking Datensatz und Benchmark besonders interessant und eine Evaluierung des Verfahrens mit anderen als den zur Entwicklung verwendeten Atlatec Sequenzen in Aussicht. Es hat sich allerdings herausgestellt, dass die Bilddaten aus zweierlei Gründen unverhältnismäßig schlechte Ergebnisse lieferten.

Zum einen haben die KITTI Bilder, bereits unter Berücksichtigung einer Region of Interest, nur etwa $\frac{1}{3}$ (0,5 Mpx) der Auflösung der Atlatec Daten. Dies führt im Wesentlichen zu erheblich höherer Unsicherheit der Rekonstruktion von einzelnen Atomen, sowie ganzen Objekten. Zum anderen sind die KITTI Bilddaten leicht überbelichtet, was entweder an einer fehlerhaften Abbildung von 12 auf 8 Bit Farbtiefe oder schlechten Belichtungseinstellungen liegen kann. Jedenfalls hat dies zur Folge, dass die Objektoberflächen deutlich an Detail verlieren und daher häufig zu homogene Flächen aufweisen, um gute Bildmerkmale zu erzeugen. In Kombination mit der niedrigeren Auflösung, weist jedes Objekt somit wesentlich weniger und verrauschtere Landmarken und folglich auch weniger Atome auf.

Im Vergleich zu den Atlatec Daten, bestanden die Objekte folglich nur noch aus etwa fünf bis zehn statt hunderten Atomen. Dies hat die Objektaggregation sowie die Struktur- und Bewegungsrekonstruktion stark gestört.

Abgesehen von der Qualität der Daten, ist der KITTI Objekterkennungs- und verfolgungs Benchmark auf eine Erkennung aller sichtbaren Objekte ausgelegt und bewertet die eingereichten Ergebnisse dementsprechend unter anderem anhand ihrer Erkennungsrate. Da das in dieser Arbeit vorgestellte Verfahren allerdings systembedingt nur bewegte Objekte erkennt, wäre eine Bewertung durch diesen Benchmark irreführend gewesen.

Wegen dieser Gründe wurde letztlich auf eine Anpassung an und Evaluierung mit den KITTI Daten zu Gunsten weiterer Verfahrensoptimierungen verzichtet.

5 Zusammenfassung und Ausblick

In dieser Arbeit wurde ein akausales Offline-Verfahren zur Lösung des STAR („simultaneous tracking and reconstruction“) Problems entwickelt.

- Dieses erkennt, ausgehend von assoziierten Beobachtungen dünn besetzter Ausreißerlandmarken einer Eigenbewegungsschätzung, bewegte Objekte in Stereobildern einer wiederum bewegten Kamera.
- Die erkannten Objekte werden dabei zeitlich verfolgt.
- Zudem wird ihre Relativbewegung und Struktur unter Verwendung aller verfügbaren Messungen rekonstruiert.

Das Verfahren besteht aus zwei Verarbeitungsschritten, welche wie folgt funktionieren.

Für die Objekterkennung wurden aus den Ausreißerbeobachtungen einer Eigenbewegungsschätzung zunächst Hypothesen kleinster, als Atome bezeichneter, Objektsegmente erzeugt. Hierfür wurden die Atome aus im Bild benachbarten Landmarken erzeugt, wobei die Nachbarschaft mittels Delaunay-Triangulation bestimmt wurde. Anschließend wurde die Hypothese aufgestellt, dass ein Atom - falls es tatsächlich ein Segment eines Objekts darstellt - genauso wie das Objekt selbst in sich starr sein muss. Für jedes Atom wurde daraufhin ein Bündelausgleich gelöst, um die Bewegung und Struktur der Atome zu bestimmen. Dieses Schätzergebnis wurde in fünf hierfür entwickelten Metriken verwendet, um festzustellen, ob es sich bei dem untersuchten Atom tatsächlich um ein Objektsegment handelt, oder das Atom aus Landmarken mehrerer Objekte oder gar verrauschten Hintergrundlandmarken besteht. Atome, die diesen Hypothesen- und Plausibilitätstest nicht bestanden, wurden verworfen. Die verbliebenen Atome wurden schließlich, unter Zuhilfenahme einer Komponentenbildung in Graphen, mit allen Atomen gruppiert, mit welchen sie Landmarken teilen, und ergaben so die erkannten Objekte.

Für jedes Objekt wurde im Schritt der Objektrekonstruktion ein Bündelausgleich mit den Beobachtungen aller Landmarken des Objekts unter Verwendung eines robusten Least-Squares Schätzers ermittelt. Als Resultat liegt eine bestmögliche Schätzung der Relativbewegung und Struktur der Objekte, in Form von 3D Punktwolken in Objektkoordinaten und den relativen Kameraposen, vor. Zusätzlich wurde eine Approximation der Objektoberflächen mittels Alpha Shapes erzeugt.

Durch die Verwendung der Hypothese starrer Körper und den damit ermöglichten Verzicht auf ein klassenbasiertes Vorgehen, wurde die Erkennung einer weiten Bandbreite auftretender Objektklassen möglich. Zudem erkennt die Methode auch solche Objektklassen, die in den zur Parametrierung verwendeten Datensätzen nicht beobachtet wurden. In den Testsequenzen wurden Fahrradfahrer, verschiedene Arten von PKWs, unter anderem Cabriolets und Geländelimousinen, Nutzfahrzeuge, wie Kleintransporter, gewöhnliche LKWs, Autotransporter oder Kranwagen, und Straßenbahnen erkannt. Fußgänger wurden leider nur unzuverlässig erkannt, weil sie nicht starr sind.

Dank der Akausalität des Verfahrens ist eine bestmögliche Bewegungs- und Strukturschätzung zu jedem Zeitschritt der Sequenz verfügbar. Bei der Erprobung von Target Tracking Systemen ist dies insbesondere bei Verdeckungen und großen Entfernungen eine wesentliche Überlegenheit gegenüber kausalen Verfahren. In zahlreichen Situationen der Testsequenzen war ein umfassendes Strukturmodell trotz Teilverdeckungen verfügbar. Bei sich aus der Ferne nähernden Objekten konnte außerdem bereits von der ersten Beobachtung an ein detailliertes Strukturmodell bereitgestellt werden.

Es bieten sich folgende Möglichkeiten an, um die beobachteten Schwächen des Verfahrens zu beheben:

- Um Fußgänger zuverlässiger zu erkennen, sollte in der Vorverarbeitung eine höhere Featuredichte erreicht werden. Somit stiege die Wahrscheinlichkeit, dass zumindest der Oberkörper einer Person als starres Segment erkannt wird, weil es Atome gibt, die sich nur aus Landmarken des Oberkörpers zusammensetzen.
- Die Objektaggregation sollte robuster gegen fehlassoziierte Landmarken gestaltet werden. Hierfür wäre beispielsweise denkbar bei der Komponentenbildung im Graphen n-fach zusammenhängende Komponenten zu fordern. Somit würden statt einer einzigen erst n gemeinsame Landmarken dazu führen, dass zwei Atome verbunden werden.
- Eine andere Möglichkeit zur robusten Objektaggregation wäre, eine iterative Komponentenbildung umzusetzen. Ausgehend von vielen Startknoten würde man die Atome zu Molekülen wachsen lassen und in jeder Iteration einen Hypothesentest durchführen. Somit würden nur Moleküle zu Komponenten weiterwachsen, die keine objektfremden Landmarken aufweisen und somit keine verschmolzenen Komponenten entstehen.
- Die Methode der Alpha Shapes könnte dafür verwendet werden, zu erkennen, ob während der Objektaggregation zwei Objekte verschmolzen sind. Ist dies der Fall, liefert das erzeugte Alpha Shape die notwendigen Informationen, um beide Objekte zu trennen, so dass diese anschließend nochmals als separate Objekte geschätzt

werden können.

- Objekte, die in mehrere Teile zerfallen sind, können unter Umständen auf Grund ihrer räumlichen Überlappung als zusammenhängendes Objekt identifiziert werden. So könnten diese nachträglich verbunden und als gemeinsames Objekt rekonstruiert werden.
- Die rekonstruierte Bewegung der Objekte kann zudem dafür verwendet werden, um Trackingfehler bei Verdeckungen oder dem Stillstand von Objekten zu korrigieren. Sind für ein Fahrzeug also in solchen Fällen mehrere Objekthypothesen entstanden, können diese zusammengefügt werden.

Die in dieser Arbeit entwickelte Methode zur Objekterkennung und -rekonstruktion eignet sich zur automatischen Berechnung von Referenzmessungen für Target Tracking Systeme. Durch die Akausalität, die klassenfreie Objekterkennung und die verwendeten nichtlinearen Schätzverfahren bietet sie wesentliche Vorteile gegenüber bisher untersuchten online Filterverfahren.

Literatur

[1] Daimler AG. *Mercedes-Benz Actros mit Highway Pilot - Weltpremiere auf öffentlichen Straßen*. Pressemitteilung. Stuttgart, 10. Feb. 2015. URL: http://media.daimler.com/dcmedia/0-921-614341-49-1852861-1-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0.html (besucht am 16. 11. 2015).

[2] Sameer Agarwal, Keir Mierle u. a. *Ceres Solver*. URL: http://ceres-solver.org.

[3] Martin Aigner. *Graphentheorie*. Springer Studium Mathematik - Bachelor. Springer Fachmedien Wiesbaden, 2015. ISBN: 978-3-658-10322-4 978-3-658-10323-1. DOI: 10.1007/978-3-658-10323-1_2. URL: http://link.springer.com/book/10.1007/978-3-658-10323-1.

[4] Hans Wilhelm Alt. *Lineare Funktionalanalysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. ISBN: 978-3-642-22260-3 978-3-642-22261-0. URL: http://link.springer.com/10.1007/978-3-642-22261-0.

[5] Pedram Azad, Tilo Gockel und Rüdiger Dillmann. *Computer Vision: principles and practice*. Elector International Media BV 2008, 4. Apr. 2008. ISBN: 978-0-905705-71-2.

[6] A. Barth und U. Franke. „Tracking oncoming and turning vehicles at intersections“. In: *2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC). Sep. 2010, S. 861–868. DOI: 10.1109/ITSC.2010.5624969.

[7] Keni Bernardin und Rainer Stiefelhagen. „Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics“. In: *J. Image Video Process.* 2008 (Jan. 2008), 1:1–1:10. ISSN: 1687-5176. DOI: 10.1155/2008/246309. URL: http://dx.doi.org/10.1155/2008/246309.

[8] R. Danescu, F. Oniga und S. Nedevschi. „Modeling and Tracking the Driving Environment With a Particle-Based Occupancy Grid“. In: *Intelligent Transportation Systems, IEEE Transactions on* 12.4 (Dez. 2011), S. 1331–1342. ISSN: 1524-9050. DOI: 10.1109/TITS.2011.2158097.

[9] Mark de Berg u. a. „Voronoi Diagrams“. In: *Computational Geometry*. Springer Berlin Heidelberg, 2008, S. 147–171. ISBN: 978-3-540-77973-5 978-3-540-77974-2. DOI: 10.1007/978-3-540-77974-2_7. URL: http://link.springer.com/chapter/10.1007/978-3-540-77974-2_7.

[10] destatis. *Statistisches Jahrbuch 2015*. Wiesbaden: Statistisches Bundesamt, Okt. 2015. ISBN: 978-3-8246-1037-2. URL: www.destatis.de/jahrbuch.

[11] H. Edelsbrunner, D. Kirkpatrick und R. Seidel. „On the shape of a set of points in the plane“. In: *Information Theory, IEEE Transactions on* 29.4 (Juli 1983), S. 551–559. ISSN: 0018-9448. DOI: 10.1109/TIT.1983.1056714.

[12] Herbert Edelsbrunner. *A Short Course in Computational Geometry and Topology*. SpringerBriefs in Applied Sciences and Technology. Springer International Publishing, 2014. ISBN: 978-3-319-05956-3 978-3-319-05957-0. DOI: 10.1007/978-3-319-05957-0_6. URL: http://link.springer.com/book/10.1007/978-3-319-05957-0.

[13] Herbert Edelsbrunner und Ernst P. Mücke. „Three-dimensional Alpha Shapes“. In: *ACM Trans. Graph.* 13.1 (Jan. 1994), S. 43–72. ISSN: 0730-0301. DOI: 10.1145/174462.156635. URL: http://doi.acm.org/10.1145/174462.156635.

[14] A. Geiger, J. Ziegler und C. Stiller. „StereoScan: Dense 3d reconstruction in real-time“. In: *2011 IEEE Intelligent Vehicles Symposium (IV)*. 2011 IEEE Intelligent Vehicles Symposium (IV). Juni 2011, S. 963–968. DOI: 10.1109/IVS.2011.5940405.

[15] Andreas Geiger, Philip Lenz und Raquel Urtasun. „Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite“. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012.

[16] Andreas Geiger u. a. „Vision meets Robotics: The KITTI Dataset“. In: *International Journal of Robotics Research (IJRR)* (2013).

[17] Stefan Grundhoff. „Wie geht es weiter mit dem Zukunftsthema Autonomes Fahren“. In: *heise Autos* (24. März 2015). URL: http://heise.de/-2583520 (besucht am 16. 11. 2015).

[18] Richard. I. Hartley und Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Second. Cambridge University Press, 2004. ISBN: 0-521-54051-8.

[19] Peter J. Huber. „Robust Estimation of a Location Parameter“. In: *The Annals of Mathematical Statistics* 35.1 (März 1964), S. 73–101. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177703732. URL: http://projecteuclid.org/euclid.aoms/1177703732.

[20] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer New York, 2002. ISBN: 978-0-387-95442-4 978-0-387-22440-4. DOI: 10.1007/0-387-22440-8_1. URL: http://link.springer.com/book/10.1007/b98835.

[21] B. Kitt, B. Ranft und H. Lategahn. „Detection and tracking of independently moving objects in urban environments“. In: *2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC). Sep. 2010, S. 1396–1401. DOI: 10.1109/ITSC.2010.5625265.

[22] P. Lenz u. a. „Sparse scene flow segmentation for moving object detection in urban environments“. In: *2011 IEEE Intelligent Vehicles Symposium (IV)*. 2011 IEEE Intelligent Vehicles Symposium (IV). Juni 2011, S. 926–932. DOI: 10.1109/IVS.2011.5940558.

[23] Eric Mack. *Elon Musk: Tesla Will Be First With Autonomous Driving; Admits To Apple Meeting*. Forbes. 19. Feb. 2014. URL: http://www.forbes.com/sites/ericmack/2014/02/19/elon-musk-tesla-will-be-first-with-autonomous-driving-admits-to-apple-meeting/ (besucht am 16. 11. 2015).

[24] *MapQuest*. URL: http://www.mapquest.com/.

[25] John Markoff. „Google Cars Drive Themselves, in Traffic“. In: *The New York Times* (9. Okt. 2010). ISSN: 0362-4331. URL: http://www.nytimes.com/2010/10/10/science/10google.html (besucht am 14. 11. 2015).

[26] L. Matthies und S.A. Shafer. „Error modeling in stereo navigation“. In: *IEEE Journal of Robotics and Automation* 3.3 (Juni 1987), S. 239–248. ISSN: 0882-4967. DOI: 10.1109/JRA.1987.1087097.

[27] Tesla Motors. *Your Autopilot has arrived*. 14. Okt. 2015. URL: http://www.teslamotors.com/blog/your-autopilot-has-arrived (besucht am 16. 11. 2015).

[28] Jorge Nocedal und Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006. ISBN: 978-0-387-30303-1 978-0-387-40065-5. DOI: 10.1007/978-0-387-40065-5.

[29] C. J. F. Ridders. „Accurate computation of F'(x) and F'(x) F"(x)“. In: *Advances in Engineering Software (1978)* 4.2 (1. Apr. 1982), S. 75–76. ISSN: 0141-1195. DOI: 10.1016/S0141-1195(82)80057-0. URL: http://www.sciencedirect.com/science/article/pii/S0141119582800570.

[30] Radu Bogdan Rusu und Steve Cousins. „3D is here: Point Cloud Library (PCL)“. In: *IEEE International Conference on Robotics and Automation (ICRA)*. Shanghai, China, 9.–13. Mai 2011.

[31] S. Sivaraman und M.M. Trivedi. „Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis“. In: *IEEE Transactions on Intelligent Transportation Systems* 14.4 (Dez. 2013), S. 1773–1795. ISSN: 1524-9050. DOI: 10.1109/TITS.2013.2266661.

[32] Jack Stewart. „Google is to start building its own self-driving cars“. In: *BBC News* (28. Mai 2014). URL: http://www.bbc.com/news/technology-27587558 (besucht am 14. 11. 2015).

[33] Christoph Stiller, Alexander Bachmann und Andreas Geiger. „Maschinelles Sehen“. In: *Handbuch Fahrerassistenzsysteme*. Springer.

[34] Zehang Sun, G. Bebis und R. Miller. „On-road vehicle detection: a review“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.5 (Mai 2006), S. 694–711. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2006.104.

[35] R. Szeliski und Sing Bing Kang. „Shape ambiguities in structure from motion“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.5 (Mai 1997), S. 506–512. ISSN: 0162-8828. DOI: 10.1109/34.589211.

[36] Richard Szeliski. *Computer Vision: Algorithms and Applications*. 1st. New York, NY, USA: Springer-Verlag New York, Inc., 2010. ISBN: 1-84882-934-5 978-1-84882-934-3.

[37] Edward Taylor und Alexei Oreskovic. „Apple studies self-driving car, auto industry source says“. In: *Reuters* (Sat Feb 14 21:31:55 UTC 2015). URL: http://www.reuters.com/article/2015/02/14/us-apple-autos-idUSKBN0LI0IJ20150214 (besucht am 16. 11. 2015).

[38] Bill Triggs u. a. „Bundle Adjustment — A Modern Synthesis“. In: *Vision Algorithms: Theory and Practice*. Hrsg. von Bill Triggs, Andrew Zisserman und Richard Szeliski. Lecture Notes in Computer Science 1883. Springer Berlin Heidelberg, 20. Sep. 1999, S. 298–372. ISBN: 978-3-540-67973-8 978-3-540-44480-0. DOI: 10.1007/3-540-44480-7_21. URL: http://link.springer.com/chapter/10.1007/3-540-44480-7_21.

[39] Yalin Xiong und K. Turkowski. „Creating image-based VR using a self-calibrating fisheye lens“. In: *, 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. Proceedings.* , 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. Proceedings. Juni 1997, S. 237–243. DOI: 10.1109/CVPR.1997.609326.

[40] Gem-Sun Young und R. Chellappa. „Statistical analysis of inherent ambiguities in recovering 3-D motion from a noisy flow field“. In: *, 10th International Conference on Pattern Recognition, 1990. Proceedings.* , 10th International Conference on Pattern Recognition, 1990. Proceedings. Bd. i. Juni 1990, 371–377 vol.1. DOI: 10.1109/ICPR.1990.118131.